Electronic Journal of Statistics
Vol. 12 (2018) 1429–1467
ISSN: 1935-7524
https://doi.org/10.1214/18-EJS1427

High-dimensional robust precision matrix estimation: Cellwise corruption under ϵ-contamination

Po-Ling Loh

Departments of Electrical & Computer Engineering and Statistics
University of Wisconsin – Madison
1415 Engineering Drive
Madison, WI 53706, USA
e-mail: loh@ece.wisc.edu

and

Xin Lu Tan*

e-mail: xinlutan92@gmail.com

Abstract: We analyze the statistical consistency of robust estimators for precision matrices in high dimensions. We focus on a contamination mechanism acting cellwise on the data matrix. The estimators we analyze are formed by plugging appropriately chosen robust covariance matrix estimators into the graphical Lasso and CLIME. Such estimators were recently proposed in the robust statistics literature, but only analyzed mathematically from the point of view of the breakdown point. This paper provides complementary high-dimensional error bounds for the precision matrix estimators that reveal the interplay between the dimensionality of the problem and the degree of contamination permitted in the observed distribution. We also show that although the graphical Lasso and CLIME estimators perform equally well from the point of view of statistical consistency, the breakdown property of the graphical Lasso is superior to that of CLIME. We discuss implications of our work for problems involving graphical model estimation when the uncontaminated data follow a multivariate normal distribution, and the goal is to estimate the support of the population-level precision matrix. Our error bounds do not make any assumptions about the the contaminating distribution and allow for a nonvanishing fraction of cellwise contamination.

MSC 2010 subject classifications: Primary 62F35; secondary 62G35.
Keywords and phrases: Robust covariance matrix estimation, cellwise contamination, Kendall's tau, Spearman's rho, median absolute deviation.

Received October 2016.

1. Introduction

Covariance matrix estimation has long taken center stage in multivariate analysis [4]. The sample covariance estimator, which originates as the maximum likeli-

*This work was completed while the author was a Ph.D. student in the Department of Statistics, The Wharton School, University of Pennsylvania.

hood estimator under a multivariate normal model, is optimal in many respects: It is unbiased, consistent, efficient under various distributional assumptions, and easily computable. Despite its positive traits, however, the sample covariance matrix is also highly non-robust when data are contaminated. Hence, various procedures in robust statistics have been derived to obtain a covariance matrix estimator that behaves well even in the presence of contaminated data [36, 28].

In other areas of multivariate analysis, the precision matrix $\mathbf{\Omega}^* := (\mathbf{\Sigma}^*)^{-1}$ is of significant interest. Examples include computing Mahalanobis distances, linear discriminant analysis, and Gaussian graphical models. In the setting of graphical models, a random vector $\mathbf{X}$ is associated with an undirected graph $G = (V, E)$ that encodes conditional independence relations between components of $\mathbf{X}$ [39]. The vertex set V contains $\{1, \ldots, p\}$, while the edge set E consists of pairs (i, j), where $(i, j) \in E$ if X_i and X_j are connected by an edge. For each non-edge $(i, j) \notin E$, the variables X_i and X_j are conditionally independent given all other variables. When $\mathbf{X} \sim N(\boldsymbol{\mu}, \mathbf{\Sigma}^*)$, pairwise conditional independence holds if and only if $\mathbf{\Omega}^*_{ij} = 0$. Thus, recovering the support of the precision matrix is equivalent to graphical model selection. The aforementioned observations have been used for network reconstruction in many scientific fields, including genetics and neuroscience (e.g., see [65, 55] and the references cited therein). When the dimensionality p is small compared to the number of samples n, a reasonable method for robust precision matrix estimation could consist of computing a robust estimate of the covariance matrix and then taking a matrix inverse.

With the recent deluge of high-dimensional data, however, a need has arisen to devise high-dimensional analogs of classical procedures that are both computable and possess rigorous theoretical guarantees. Although several methods, notably the graphical Lasso (GLasso) [68, 6, 26] and the constrained ℓ_1-minimization for inverse matrix estimation (CLIME) [11] estimator, have been proposed for high-dimensional precision matrix estimation, robust estimation of high-dimensional precision matrices has only recently emerged in the literature. The GLasso and CLIME estimators tend to perform poorly under contaminated data, since they take as input the sample covariance matrix that is sensitive to even a single outlier.

Popular classical robust covariance estimators are applicable in settings where less than half the observation vectors are contaminated. Such an assumption is closely connected to the Tukey-Huber contamination model that underlies much of the existing robustness theory [61, 35]. In the Tukey-Huber model, a mixture distribution with a dominant nominal component (such as a multivariate normal distribution) and a minority unspecified component are posited, and each observation vector is either completely clean or completely contaminated. Classical robust covariance estimators then involve downweighting contaminated observations in order to reduce their influence. When the dimension p is large, however, the fraction of perfectly observed data vectors may be rather small: If all components of an observation vector had an independent chance of being contaminated, most observation vectors would be contaminated. Thus, downweighting an entire observation would waste the information contained in the clean components of the observation vector. This describes the setting of the

cellwise contamination model, which was developed by Alqallaf et al. [2]. It generalizes the classical Tukey-Huber contamination model, which may be viewed as a case of *rowwise* contamination of the data matrix, and is fairly realistic for applications involving measurement error in DNA microarray analysis [59] or dropout measurements in sensor arrays [57].

On the other hand, most existing approaches for robust covariance estimation focus on affine equivariance. These include the M-estimators [44], Minimum Volume Ellipsoid (MVE) and Minimum Covariance Determinant (MCD) estimators [50, 51], and Stahel-Donoho (SD) estimator [56, 21]. Although affine equivariance may be a desirable property under rowwise contamination, it is less appropriate in the setting of cellwise contamination, since linear combinations of observation vectors lead to a propagation of outliers [3]. In addition, the MVE, MCD, and SD estimators all require heavy computational effort, rendering them impractical for high-dimensional datasets. To deal with cellwise contamination, Van Aelst [62] proposed a modified SD estimator that adapts winsorization [36, 2] and a cellwise weighting scheme. Similar to the original SD estimator, however, computation is only feasible for small p. A recent approach by Agostinelli et al. [1] is capable of dealing with both rowwise and cellwise outliers. The procedure consists of two steps: (1) flagging cellwise outliers as missing values; and (2) applying a rowwise robust method to the incomplete data. However, computation is again infeasible in high dimensions. Other recent proposals for robust high-dimensional covariance matrix estimation include those suggested by Chen et al. [17] and Han et al. [31], but both methods treat different contamination models and are not suitable to handle data with cellwise contamination: Han et al. [31] study robust high-dimensional scatter matrix estimation when data are drawn from heavy-tailed distributions, and Chen et al. [17] study a method based on "matrix depth" designed for handling rowwise contamination that is computationally intractable in high dimensions. However, note that our proposed estimators are computationally feasible.

In fact, relatively few approaches exist for robust high-dimensional precision matrix estimation under any form of contamination. One method is supplied by the TLasso estimator of Finegold and Drton [25], which builds upon the GLasso and models the data as coming from the multivariate t-distribution, a long-tailed surrogate for the multivariate normal distribution. The "alternative multivariate t-distribution" is used to model a case where different coordinates of the distribution are obtained from the latent multivariate normal distribution using different weights. Although the TLasso demonstrates a higher degree of robustness than the GLasso under both rowwise and cellwise contamination in simulations, however, a theoretical analysis from the point of view of robust statistics has not been derived.

More recently, Oellerer and Croux [46] and Tarr et al. [58] proposed a promising new method for high-dimensional precision matrix estimation, designed specifically for cellwise contamination. The method consists of combining a robust covariance estimator that may be computed efficiently with a suitable high-dimensional precision matrix estimation procedure. Similar plug-in estimators based on rank-based correlation matrix estimates were previously proposed by

Liu et al. [42] and Xue and Zou [66] for model selection and parameter estimation in nonparanormal graphical models. However, an important difference in our present work is the model assumption under which we derive our error rates. Since we draw our motivation from robust statistics, our goal is to prove estimation error bounds under an ϵ-contaminated distribution. In contrast, Liu et al. [42] and Xue and Zou [66] focus on establishing consistency *when the observations are drawn cleanly* from a nonparanormal model, although they do explore the effect of rowwise outliers in their simulations. Other follow-up work [30, 29, 24, 23, 64, 5] again focuses on establishing statistical consistency under transformational or heavy-tailed variants of the high-dimensional Gaussian model. Instead, we adopt the framework of Oellerer and Croux [46] and Tarr et al. [58], who study the behavior of robust estimators when a fraction of the data are contaminated. A rigorous high-dimensional analysis from the point of view of statistical consistency is absent from this line of work—the goal of the present paper is to fill this gap.

Our main contributions are to derive statistical error bounds in elementwise ℓ_∞-norm for robust precision matrix estimation procedures according to the proposals of Oellerer and Croux [46] and Tarr et al. [58]. We study the setting of the cellwise ϵ-contamination model, where at most an ϵ fraction of entries in the data matrix are corrupted by outliers. Our work thus fuses two threads of research involving classical robust statistics and high-dimensional estimation in a novel and rigorous manner. The bounds we derive match standard high-dimensional bounds for uncontaminated precision matrix estimation, up to a constant multiple of ϵ. Furthermore, they are of a complementary nature to the theoretical results supplied by Oellerer and Croux [46], since we are primarily concerned with robustness as measured from the viewpoint of statistical consistency, rather than breakdown behavior. After the initial posting of our work, a nice minimax theory for statistical estimation under the ϵ-contamination model has been developed [17, 18]; we build upon some of these results to show that the error rates achieved by our estimators are minimax optimal.

More generally, our results reveal an interesting interplay between bounds for statistical error under ϵ-contamination and classical measures of robustness such as the influence function [27] and breakdown point [22]. Estimators with bounded influence have long been favored in classical robust statistics, as the rate of change in the statistical functional associated with the estimator is controlled when the nominal distribution is contaminated by an arbitrary point mass distribution. Our results show that a variety of bounded influence estimators, including Kendall's and Spearman's correlation coefficients, give rise to (inverse) covariance estimators with statistical error rates that depend linearly on the degree of contamination; the converse relationship may be seen to hold more generally as a result of our proof arguments. On the other hand, our discussion of the breakdown point of the precision matrix estimators, building upon the analysis of Oellerer and Croux [46], emphasizes the significant differences between the notions of breakdown point and statistical consistency. Whereas our analysis shows that the robust CLIME and GLasso procedures have comparable behavior from the point of view of high-dimensional statistical

consistency, the CLIME estimator has a substantially smaller breakdown point than the GLasso, due to its constrained feasibility region. Rather than advocating one measure of robustness over another, our discussion emphasizes the value of weighing different measures of robustness in selecting an appropriate estimator.

The remainder of the paper is organized as follows: Section 2 furnishes the mathematical background for the cellwise contamination model and the robust covariance and precision matrix estimators to be considered in the paper. Section 3 presents our main theoretical contributions, providing bounds on the statistical error of the covariance and precision matrix estimators under the cellwise contamination model, as well as concrete consequences in the presence of outliers and/or missing data. Section 4 provides a discussion of the breakdown point for the robust GLasso and CLIME estimators. In Section 5, we discuss the main steps of the proofs of our theorems. Section 6 contains simulation results that are used to validate the theoretical results of the paper. We conclude with a discussion in Section 7, including some avenues for future research.

Notation: For a vector $\mathbf{a} = (a_1, \ldots, a_p)^T \in \mathbb{R}^p$, we write $\|\mathbf{a}\|_1 = \sum_{i=1}^p |a_i|$ and $\|\mathbf{a}\|_2 = (\sum_{i=1}^p a_i^2)^{1/2}$ to denote the ℓ_1- and ℓ_2-norms of $\mathbf{a}$. For a matrix $\mathbf{A} = (a_{ij}) \in \mathbb{R}^{p\times q}$, we define the elementwise ℓ_1-norm $\|\mathbf{A}\|_1 = \sum_{i=1}^p \sum_{j=1}^q |a_{ij}|$, the elementwise ℓ_∞-norm $\|\mathbf{A}\|_\infty = \max_{1\le i\le p, 1\le j\le q} |a_{ij}|$, the Frobenius norm $\|\mathbf{A}\|_F = (\sum_{i=1}^p \sum_{j=1}^q a_{ij}^2)^{1/2}$, the spectral norm $|||\mathbf{A}|||_2 = \sup_{\|x\|\le 1} \|\mathbf{A}\mathbf{x}\|_2$, and the ℓ_1-operator norm $|||\mathbf{A}|||_1 = \max_{1\le j\le q} \sum_{i=1}^p |a_{ij}|$. We write $\mathbf{A} \succ 0$ (respectively, $\mathbf{A} \succeq 0$) to indicate that $\mathbf{A}$ is positive definite (respectively, positive semidefinite), in which case we denote by $\lambda_1(\mathbf{A}) \ge \lambda_2(\mathbf{A}) \ge \cdots \ge \lambda_p(\mathbf{A})$ the ordered eigenvalues of $\mathbf{A}$. We write $\mathbf{I}$ for the identity matrix and $\mathbf{0}$ for the vector of all zeros (the respective dimension of which will be clear from context). The binary operation $\otimes$ denotes the tensor product.

2. Background and problem setup

We begin with a description of the cellwise contamination model, followed by a rigorous formulation of the robust covariance and precision matrix estimators to be studied in our paper.

Following the notation of [2, 3], we write the cellwise contamination model in the following form:

$$\mathbf{X}_k = (\mathbf{I} - \mathbf{B}_k)\mathbf{Y}_k + \mathbf{B}_k\mathbf{Z}_k, \qquad \forall k = 1, \ldots, n. \tag{1}$$

Here, we observe the contaminated random vector $\mathbf{X}_k \in \mathbb{R}^p$. The unobservable random vectors $\mathbf{Y}_k, \mathbf{Z}_k$, and $\mathbf{B}_k$ are independent, and $\mathbf{Y}_k \sim G$ (a nominal distribution) and $\mathbf{Z}_k \sim H^*$ (an unspecified outlier generating distribution). Furthermore, $\mathbf{B}_k = \mathrm{diag}(B_{k1}, \ldots, B_{kp})$ is a diagonal matrix, where $B_{k1}, \ldots, B_{kp}$ are independent Bernoulli random variables with $P(B_{ki} = 1) = \epsilon_i$, for all $1 \le i \le p$.

When $\epsilon_1 = \cdots = \epsilon_p = \epsilon$, the probability of an observation vector having no contamination in any component is $(1-\epsilon)^p$, a quantity that decreases exponentially as the dimension increases. This probability goes below the critical value $1/2$ for $p \geq 14$ at $\epsilon = 0.05$, and for $p \geq 69$ at $\epsilon = 0.01$. Equation (1) is a special case of a more general model, where we allow other joint distributions for $B_{k1}, \ldots, B_{kp}$. For instance, if $B_{k1}, \ldots, B_{kp}$ were completely dependent (i.e., $P(B_{k1} = \cdots = B_{kp}) = 1$), we would obtain the rowwise contamination model. In that case, the probability of an observation vector being totally free of contamination would be $1-\epsilon$, which is independent of the dimension. Alqallaf et al. [3] also use the terms *fully independent contamination model (FICM)* and *fully dependent contamination model (FDCM)* to denote the cellwise and rowwise contamination settings, in order to distinguish the pattern of contamination across rows of the data matrix.

Throughout, we will work under the cellwise contamination model (1), and assume that G is a multivariate normal distribution $N(\boldsymbol{\mu}, \boldsymbol{\Sigma}^*)$. Our goal is to estimate the matrices $\boldsymbol{\Sigma}^*$ and $\boldsymbol{\Omega}^* = (\boldsymbol{\Sigma}^*)^{-1}$ from the (uncontaminated) normal component.

2.1. Covariance matrix estimation

When $\epsilon = 0$ (i.e., the data are uncontaminated), we may use the classical sample covariance matrix estimator $\tilde{\boldsymbol{\Sigma}}$, defined pairwise as

$$\tilde{\boldsymbol{\Sigma}}_{ij} = \frac{1}{n-1}\sum_{k=1}^{n}(X_{ki} - \bar{X}_i)(X_{kj} - \bar{X}_j), \qquad \forall 1 \leq i, j \leq p,$$

where $\bar{X}_i = \frac{1}{n}\sum_{k=1}^{n} X_{ki}$ and $\bar{X}_j = \frac{1}{n}\sum_{k=1}^{n} X_{kj}$. When $n \gg p$, the sample covariance is an efficient estimator for $\boldsymbol{\Sigma}^*$. However, when $\epsilon > 0$, the performance of $\tilde{\boldsymbol{\Sigma}}$ may be compromised depending on the properties of H^*: Under the cellwise contamination model, for $i \neq j$, we have

$$\begin{aligned}(\boldsymbol{\Sigma}_X^*)_{ij} &= (1-\epsilon_i)(1-\epsilon_j)\,(\boldsymbol{\Sigma}_Y^*)_{ij} + \epsilon_i\epsilon_j\,(\boldsymbol{\Sigma}_Z^*)_{ij} \\ &= (\boldsymbol{\Sigma}_Y^*)_{ij} - (\epsilon_i + \epsilon_j - \epsilon_i\epsilon_j)\,(\boldsymbol{\Sigma}_Y^*)_{ij} + \epsilon_i\epsilon_j\,(\boldsymbol{\Sigma}_Z^*)_{ij}\,.\end{aligned}$$

When no restrictions are placed on the covariance $\boldsymbol{\Sigma}_Z^*$ of the contaminating distribution, the elementwise deviations between $\boldsymbol{\Sigma}_X^*$ and $\boldsymbol{\Sigma}_Y^*$ (and consequently, also the sample covariance $\tilde{\boldsymbol{\Sigma}}_X := \tilde{\boldsymbol{\Sigma}}$ and $\boldsymbol{\Sigma}_Y^*$) will in general behave arbitrary badly. Furthermore, note that even when $\boldsymbol{\Sigma}_Z^*$ is constrained to lie in a space where the deviations between $\boldsymbol{\Sigma}_X^*$ and $\boldsymbol{\Sigma}_Y^*$ are suitably bounded, we would require the contaminating distribution to have properties such as sub-Gaussian tails in order to ensure consistency of the sample covariance estimator on the order of $\mathcal{O}\left(\sqrt{\frac{\log p}{n}}\right)$. When a procedure based on covariance estimation is used to estimate the precision matrix, the errors incurred during the covariance estimation step would propagate to the next step. For instance, this issue would

arise in using the CLIME or GLasso estimator. In contrast, our theory for robust covariance estimators will not require any assumptions on either $\mathbf{\Sigma}_Z^*$ or the tail behavior of the contaminating distribution.

To deal with cellwise contamination in the high-dimensional setting, we therefore take the pairwise approach suggested by Oellerer and Croux [46], where a robust covariance or correlation estimate is computed for each pair of variables. Early proposals of robust procedures are of this type [7, 47], where a coordinatewise approach is taken for robust estimation of location. In addition to having relatively low computational complexity, the pairwise approach is appealing because a high breakdown point of the pairwise estimators translates into a high breakdown point of the overall covariance matrix. For $1 \le i, j \le p$, we write

$$\mathbf{\Sigma}_{ij}^* = \sigma_i \sigma_j \boldsymbol{\rho}_{ij}, \tag{2}$$

where $\sigma_i = [\mathrm{Var}(X_{ki})]^{1/2}$, $\sigma_j = [\mathrm{Var}(X_{kj})]^{1/2}$, and $\boldsymbol{\rho}_{ij} = \mathrm{Corr}(X_{ki}, X_{kj})$. We will take suitable robust estimators of $\hat{\sigma}_i$, $\hat{\sigma}_j$, and $\hat{\boldsymbol{\rho}}_{ij}$, to obtain the covariance matrix estimator $\hat{\mathbf{\Sigma}}$, with (i, j) entry $\hat{\mathbf{\Sigma}}_{ij} = \hat{\sigma}_i \hat{\sigma}_j \hat{\boldsymbol{\rho}}_{ij}$.

To estimate σ_i, we consider the median absolute deviation from the median (MAD), a robust measure of scale. The MAD estimator was popularized by Hampel [27], who attributes the concept to Gauss. It has a breakdown point of 50%. Let $X_{(1),i} \le \cdots \le X_{(n),i}$ denote the ordered values of $X_{1i}, \ldots, X_{ni}$. The sample median $\hat{m}_i$ and the sample MAD $\hat{d}_i$ are defined, respectively, as $\hat{m}_i = X_{(k^*),i}$ and $\hat{d}_i = W_{(k^*),i}$, where $W_{ki} = |X_{ki} - \hat{m}_i|$, for all $k = 1, \ldots, n$, and $k^* = \lceil n/2 \rceil$. Expressed another way,

$$\hat{d}_i = \underset{1 \le k \le n}{\mathrm{median}} \left(\left| X_{ki} - \underset{1 \le \ell \le n}{\mathrm{median}}(X_{\ell i}) \right| \right). \tag{3}$$

We then estimate σ_i by $\hat{\sigma}_i = [\Phi^{-1}(0.75)]^{-1} \hat{d}_i$, where the constant $[\Phi^{-1}(0.75)]^{-1}$ is chosen in order to make the estimator consistent for σ_i at normal distribution. The population-level median of a distribution with cdf F is defined to be $m(F) := F^{-1}(0.5)$, where $F^{-1}(c) = \inf\{x : F(x) \ge c\}$, for $c \in [0, 1]$. Similarly, we may define the population-level MAD $d(F)$ to be the median of the distribution of $|X - m(F)|$, where X has cdf F.

To estimate $\boldsymbol{\rho}_{ij}$, we consider the classical nonparametric correlation estimators, Kendall's tau and Spearman's rho:

Kendall's tau: This statistic is given by

$$\boldsymbol{r}_{ij}^K = \frac{2}{n(n-1)} \sum_{k<\ell} \mathrm{sign}(X_{ki} - X_{\ell i}) \mathrm{sign}(X_{kj} - X_{\ell j}), \tag{4}$$

where $\mathrm{sign}(X) = 1$ if $X > 0$, $\mathrm{sign}(X) = -1$ if $X < 0$, and $\mathrm{sign}(0) = 0$.

Spearman's rho: This statistic is given by

$$\boldsymbol{r}_{ij}^S = \frac{\sum_{k=1}^n [\mathrm{rank}(X_{ki}) - (n+1)/2][\mathrm{rank}(X_{kj}) - (n+1)/2]}{\sqrt{\sum_{k=1}^n [\mathrm{rank}(X_{ki}) - (n+1)/2]^2 \sum_{k=1}^n [\mathrm{rank}(X_{kj}) - (n+1)/2]^2}}, \tag{5}$$

where $\mathrm{rank}(X_{ki})$ denotes the rank of X_{ki} among $X_{1i}, \ldots, X_{ni}$.

The population versions of the estimators are given, respectively, by

$$\boldsymbol{\rho}_{ij}^K = E[\text{sign}(X_{1i} - X_{2i})\text{sign}(X_{1j} - X_{2j})], \tag{6a}$$

$$\boldsymbol{\rho}_{ij}^S = 3E[\text{sign}(X_{1i} - X_{2i})\text{sign}(X_{1j} - X_{3j})]. \tag{6b}$$

When $\epsilon_1 = \cdots = \epsilon_p = 0$, we have $\mathbf{X}_k \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma}^*)$; in this case, it is known that [37, 38]

$$\boldsymbol{\rho}_{ij} = \sin\left(\frac{\pi}{2}\boldsymbol{\rho}_{ij}^K\right) = 2\sin\left(\frac{\pi}{6}\boldsymbol{\rho}_{ij}^S\right).$$

Hence, for asymptotic consistency at normal distribution, our estimator for $\boldsymbol{\rho}_{ij}$ is the transformed version of Kendall's tau and Spearman's rho, given by $\sin(\frac{\pi}{2}\boldsymbol{r}_{ij}^K)$ and $2\sin(\frac{\pi}{6}\boldsymbol{r}_{ij}^S)$, respectively. We then define as $\hat{\boldsymbol{\Sigma}}$ our robust covariance matrix estimator, with

$$\hat{\boldsymbol{\Sigma}}_{ij}^K = \hat{\sigma}_i\hat{\sigma}_j \sin\left(\frac{\pi}{2}\boldsymbol{r}_{ij}^K\right), \qquad \text{and} \qquad \hat{\boldsymbol{\Sigma}}_{ij}^S = 2\hat{\sigma}_i\hat{\sigma}_j \sin\left(\frac{\pi}{6}\boldsymbol{r}_{ij}^S\right). \tag{7}$$

2.2. Precision matrix estimation

A long line of literature exists for precision matrix estimation in the high-dimensional setting. We will focus our attention on sparse precision matrix estimation; i.e., $\boldsymbol{\Omega}^*$ contains many zero entries. In this section, we review two techniques, the GLasso and CLIME, which produce a sparse precision matrix estimator based on optimizing a function of the sample covariance matrix. As proposed by Oellerer and Croux [46] and Tarr et al. [58], these methods may easily be modified to obtained robust versions, where the sample covariance matrix estimator is simply replaced by a robust covariance estimator $\hat{\boldsymbol{\Sigma}}$ as described in the previous section.

The graphical lasso (GLasso) estimator [68, 6, 26] is defined as the maximizer of the following function:

$$\tilde{\boldsymbol{\Omega}} = \underset{\boldsymbol{\Omega}\succ 0}{\text{argmin}} \left\{\text{tr}(\tilde{\boldsymbol{\Sigma}}\boldsymbol{\Omega}) - \log\det(\boldsymbol{\Omega}) + \lambda\|\boldsymbol{\Omega}\|_1\right\}.$$

Here, $\lambda > 0$ is a tuning parameter that controls the sparsity of the resulting precision matrix estimator.

In this paper, we replace the sample covariance matrix $\tilde{\boldsymbol{\Sigma}}$ by the robust alternative $\hat{\boldsymbol{\Sigma}}$, and consider a variant where only the off-diagonal entries of the estimator are penalized:

$$\hat{\boldsymbol{\Omega}} = \underset{\boldsymbol{\Omega}\succ 0}{\text{argmin}} \left\{\text{tr}(\hat{\boldsymbol{\Sigma}}\boldsymbol{\Omega}) - \log\det(\boldsymbol{\Omega}) + \lambda\|\boldsymbol{\Omega}\|_{1,\text{off}}\right\}. \tag{8}$$

Note that although the program (8) is convex for any choice of $\hat{\boldsymbol{\Sigma}} \in \mathbb{R}^{p\times p}$, several state-of-the-art algorithms for optimizing the GLasso require the matrix $\hat{\boldsymbol{\Sigma}}$ to be positive semidefinite [26, 70, 34]. We will first derive statistical theory for the robust GLasso without a positive semidefinite projection step, and then discuss properties of the projected version in Section 4.

An alternative to the GLasso is the method of constrained ℓ_1-minimization for inverse matrix estimation (CLIME) proposed in Cai et al. [11]. The CLIME routine solves the following convex optimization problem by linear programming:

$$\tilde{\mathbf{\Omega}} = \underset{\mathbf{\Omega}\in\mathbb{R}^{p\times p}}{\operatorname{argmin}} \|\mathbf{\Omega}\|_1 \qquad \text{subject to} \qquad \|\tilde{\mathbf{\Sigma}}\mathbf{\Omega} - \mathbf{I}\|_\infty \le \lambda.$$

Note that here, no symmetry condition is imposed on $\mathbf{\Omega}$, and the solution is not symmetric in general. If a symmetric precision matrix estimate is desired, we may perform a post-symmetrization step on $\tilde{\mathbf{\Omega}} = (\tilde{\omega}^1_{ij})$ to obtain the symmetric matrix $\tilde{\mathbf{\Omega}}_{\text{sym}}$, defined by

$$\begin{aligned} &\tilde{\mathbf{\Omega}}_{\text{sym}} = (\tilde{\omega}_{ij}), \qquad \text{where} \\ &\tilde{\omega}_{ij} = \tilde{\omega}_{ji} = \tilde{\omega}^1_{ij}\mathbb{1}(|\tilde{\omega}^1_{ij}| \le |\tilde{\omega}^1_{ji}|) + \tilde{\omega}^1_{ji}\mathbb{1}(|\tilde{\omega}^1_{ij}| > |\tilde{\omega}^1_{ji}|). \end{aligned} \tag{9}$$

In other words, between $\tilde{\omega}^1_{ij}$ and $\tilde{\omega}^1_{ji}$, we pick the entry with smaller magnitude. Similar to the GLasso case, we will robustify the CLIME estimator by solving

$$\hat{\mathbf{\Omega}} = \underset{\mathbf{\Omega}\in\mathbb{R}^{p\times p}}{\operatorname{argmin}} \|\mathbf{\Omega}\|_1 \qquad \text{subject to} \qquad \|\hat{\mathbf{\Sigma}}\mathbf{\Omega} - \mathbf{I}\|_\infty \le \lambda, \tag{10}$$

and then apply post-symmetrization (9) to obtain the robust CLIME estimator $\hat{\mathbf{\Omega}}_{\text{sym}}$.

We remark that the same estimators (8) and (10), based on plugging in a robust rank-based surrogate of the correlation matrix, also appeared in Liu et al. [42] and Xue and Zou [66]. However, the focus of both papers was to derive consistency of the estimators under a nonparanormal model, rather than quantifying the effect of deviations from normality, which is the primary objective of the present paper.

3. Main results and consequences

We now provide rigorous statements of the main results of the paper. We begin by stating a series of meta-theorems that bound the elementwise ℓ_∞-norm error on the robust covariance and correlation matrices, as well as the precision matrix estimators. Although the arguments deriving Theorems 2 and 3 (on the CLIME and GLasso estimators, respectively) from Theorem 1 are not new to the high-dimensional statistics literature, Theorem 1 clearly isolates the conditions on the robust correlation and scale estimates that we require for our overall error bounds to hold. The main novelty of our work comes in Sections 3.2 and 3.3, where we show that the conditions in Theorem 1 indeed hold w.h.p. for specific robust estimators of interest.

3.1. Meta-theorems for covariance and precision estimation

We first list the key properties of our robust correlation and scale estimators that we require for our theory:

(i) The correlation and scale estimators satisfy the deviation bounds

$$\max_{1\le i,j\le p} |\hat{\boldsymbol{\rho}}_{ij} - \bar{\boldsymbol{\rho}}_{ij}| \le C_1\sqrt{\frac{\log p}{n}}, \text{ and} \tag{11a}$$

$$\max_{1\le i\le p} |\hat{\sigma}_i - \bar{\sigma}_i| \le C_1'\sqrt{\frac{\log p}{n}}. \tag{11b}$$

(ii) The correlation and scale estimators satisfy the bounded-influence inequalities

$$\max_{1\le i,j\le p} |\bar{\boldsymbol{\rho}}_{ij} - \boldsymbol{\rho}_{ij}| \le C_2\epsilon, \text{ and} \tag{12a}$$

$$\max_{1\le i\le p} |\bar{\sigma}_i - \sigma_i| \le C_2'\epsilon, \tag{12b}$$

Here, $\bar{\boldsymbol{\rho}}_{ij}$ and $\bar{\sigma}_i$ represent population-level quantities. (For instance, $\bar{\sigma}_i = E(\hat{\sigma}_i)$; however, it will be convenient in our analysis to show concentration to a slightly different quantity for the correlation estimators.) The inequalities (12a) and (12b) may be interpreted as "bounded-influence" type conditions on the correlation and scale estimators, enforcing that the population-level version of the estimator is affected only linearly under ϵ-contamination.

We also assume that the scale parameters of the uncontaminated distributions are bounded:

$$\max_{1\le i\le p} \sigma_i \le M_\sigma. \tag{13}$$

We will show in the next section that these inequalities are satisfied for estimators of interest such as the MAD scale estimator and the Kendall's tau or Spearman's rho correlation estimators, under the cellwise contamination model (1).

The following simple theorem provides error bounds for covariance matrix estimation. As pointed out by the reviewers, situations may arise when the end goal is correlation matrix estimation. In such settings, it is unnecessary to obtain scale estimates $\{\hat{\sigma}_i\}_{i=1}^p$, so we only need bounds on the behavior of the correlation estimators.

Theorem 1 (Covariance and correlation matrices)**.**

(i) Suppose the correlation estimator $\hat{\boldsymbol{\rho}}$ satisfies inequalities (11a) and (12a). Then

$$\|\hat{\boldsymbol{\rho}} - \boldsymbol{\rho}^*\|_\infty \le C_1\sqrt{\frac{\log p}{n}} + C_2\epsilon.$$

(ii) Suppose the covariance estimator $\hat{\boldsymbol{\Sigma}}$ is defined elementwise according to $\hat{\boldsymbol{\Sigma}}_{ij} = \hat{\sigma}_i\hat{\sigma}_j\hat{\boldsymbol{\rho}}_{ij}$, and inequalities (11b), (12b), and (13) also hold. Suppose $C_1'\sqrt{\frac{\log p}{n}} + C_2'\epsilon < M_\sigma$. Then

$$\left\|\hat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}^*\right\|_\infty \le 4M_\sigma^2\left(C_1\sqrt{\frac{\log p}{n}} + C_2\epsilon\right) + 3M_\sigma\left(C_1'\sqrt{\frac{\log p}{n}} + C_2'\epsilon\right).$$

The proof of Theorem 1 is straightforward and mostly involves applications of the triangle inequality. The proof is provided in Appendix B.1.

We now turn to the more complicated problem of obtaining statistical error bounds on the precision matrix estimators described in the previous section. We begin with the CLIME estimator. Consider the following uniformity class of matrices:

$$\mathcal{U}(q, s_0(p), M) = \left\{ \boldsymbol{\Omega} : \boldsymbol{\Omega} \succ 0, \|\boldsymbol{\Omega}\|_1 \leq M, \max_{1 \leq i \leq p} \sum_{j=1}^{n} |\omega_{ij}|^q \leq s_0(p) \right\}, \tag{14}$$

for $0 \leq q < 1$, where $\boldsymbol{\Omega} := (\omega_{ij}) = (\boldsymbol{\omega}_1, \ldots, \boldsymbol{\omega}_p)$. The following result provides an elementwise error bound on the estimation error between the CLIME output and the true precision matrix, provided the true precision matrix lies in the class (14) defined above. The proof of Theorem 2 follows Cai et al. [11]; the proof is provided in Appendix B.2 only for completeness.

Theorem 2 (CLIME estimator)**.** *Suppose* $\boldsymbol{\Omega}^* \in \mathcal{U}(q, s_0(p), M)$*. If* $\hat{\boldsymbol{\Omega}}$ *is the output of the CLIME estimator* (10) *with* $\lambda \geq M\|\hat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}^*\|_\infty$*, then*

$$\|\hat{\boldsymbol{\Omega}} - \boldsymbol{\Omega}^*\|_\infty \leq 4\lambda \left\|\boldsymbol{\Omega}^*\right\|_1.$$

Remark 1. *Clearly, the optimal choice of* λ *to minimize the estimation error bound in Theorem 2 is* $\lambda = C_1'' \sqrt{\frac{\log p}{n}} + C_2'' \epsilon$*, where* C_1'' *and* C_2'' *are the constant prefactors appearing in Theorem 1. In this case,*

$$\|\hat{\boldsymbol{\Omega}} - \boldsymbol{\Omega}^*\|_\infty \leq 4 \left\|\boldsymbol{\Omega}^*\right\|_1 \left(C_1'' \sqrt{\frac{\log p}{n}} + C_2'' \epsilon \right) \leq 4M \left(C_1'' \sqrt{\frac{\log p}{n}} + C_2'' \epsilon \right).$$

Hence, we see that the error bound on the precision matrix may be separated cleanly into two parts: A linear term capturing the effect of ϵ*-contamination, and an* $\mathcal{O}\left(\sqrt{\frac{\log p}{n}}\right)$ *term capturing the minimax rate of estimation even when* $\epsilon = 0$*.*

For the GLasso, we focus on precision matrices satisfying the following assumption:

Assumption 1 (Incoherence)**.** *There exists some* $0 < \alpha \leq 1$ *such that*

$$\max_{e \in S^c} \left\| \boldsymbol{\Gamma}^*_{eS} (\boldsymbol{\Gamma}^*_{SS})^{-1} \right\|_1 \leq 1 - \alpha, \tag{15}$$

where $\boldsymbol{\Gamma}^* := \boldsymbol{\Sigma}^* \otimes \boldsymbol{\Sigma}^*$ *and* $S = \operatorname{supp}(\boldsymbol{\Omega}^*)$ *is the true edge set.*

We then have the following result, which is stated in terms of the population-level quantities $\kappa_{\Sigma^*} = \left\|\boldsymbol{\Sigma}^*\right\|_1$ and $\kappa_{\Gamma^*} = \left\|(\boldsymbol{\Gamma}^*_{SS})^{-1}\right\|_1$, as well as d, the maximum number of nonzero elements in each row of $\boldsymbol{\Omega}^*$. The proof of Theorem 3 is contained in Appendix B.3.

Theorem 3 (GLasso estimator)**.** *Suppose Assumption 1 holds. If the regularization parameter satisfies*

$$\left\|\hat{\mathbf{\Sigma}} - \mathbf{\Sigma}^*\right\|_\infty \leq \frac{\alpha\lambda}{8}, \tag{16}$$

and

$$2\kappa_{\Gamma^*}\lambda\left(1+\frac{\alpha}{8}\right) \leq \min\left\{\frac{1}{3\kappa_{\Sigma^*}d}, \frac{1}{3\kappa_{\Sigma^*}^3\kappa_{\Gamma^*}d}\right\}, \text{ and } 48\kappa_{\Sigma^*}^3\kappa_{\Gamma^*}^2\left(\frac{\alpha}{8}+1\right)^2 d\lambda \leq \alpha, \tag{17}$$

then the GLasso estimator (8) *satisfies* $\text{supp}(\hat{\mathbf{\Omega}}) \subseteq \text{supp}(\mathbf{\Omega}^*)$*, and*

$$\|\hat{\mathbf{\Omega}} - \mathbf{\Omega}^*\|_\infty \leq 2\kappa_{\Gamma^*}\left(1+\frac{8}{\alpha}\right)\lambda.$$

Comparing the results of Theorems 2 and 3, we see that as in the traditional uncontaminated setting, the GLasso delivers slightly stronger guarantees, at the expense of more stringent assumptions. In particular, we will take $\lambda = C_1''\sqrt{\frac{\log p}{n}} + C_2''\epsilon$ (and assume for simplicity that $\epsilon = 0$), the GLasso requires the sample size to scale as $n \geq Cd^2 \log p$ in order for the conditions (17) to hold, whereas the CLIME requires $n \geq C' \left\|\mathbf{\Omega}^*\right\|_1^2 \log p$ in order to achieve consistency. When the parameter M defining the precision matrix class scales more slowly than d^2, the CLIME thus requires a weaker scaling. In addition, the GLasso result supposes Assumption 1, which posits an incoherence bound on submatrices of $\mathbf{\Gamma}^*$. On the other hand, Theorem 3 establishes that the $\text{supp}(\hat{\mathbf{\Omega}}) \subseteq \text{supp}(\mathbf{\Omega}^*)$ for the GLasso estimator, whereas Theorem 2 only guarantees consistency for the CLIME estimator in terms of ℓ_∞-norm. In the case of the CLIME estimator, however, the true support of $\mathbf{\Omega}^*$ may be obtained via thresholding, assuming the nonzero elements of $\mathbf{\Omega}^*$ are of the order $\Omega\left(\sqrt{\frac{\log p}{n}}\right)$.

Remark 2. *Focusing on the level of contamination ϵ in relation to the problem dimensions, note that Theorems 2 and 3 both imply an $\mathcal{O}\left(\sqrt{\frac{\log p}{n}}\right) + \mathcal{O}(\epsilon)$ error bound on the precision matrix estimator, under the corresponding assumptions. Hence, when $\epsilon \leq C\sqrt{\frac{\log p}{n}}$, the estimation error matches the error of the optimal precision matrix estimator in the uncontaminated case, up to a constant factor [49, 12]. Otherwise, condition* (17) *implies that $\epsilon = \mathcal{O}\left(\frac{1}{d}\right)$. Hence, although the level of contamination tolerated by the estimator decreases as the degree of the underlying graph increases, it is* not *required to decrease as n and p increase, as long as the ratio $\sqrt{\frac{\log p}{n}}$ remains fixed. Thus, the conclusions of Theorems 2 and 3 are truly high-dimensional. As in the case of the robust covariance matrix estimators, a nice feature is that when the data are uncontaminated ($\epsilon = 0$), the estimation error of the robust precision matrix estimator agrees with the optimal rate.*

Note that since the inverse of the correlation matrix has the same support as the precision matrix, we could also estimate $\mathrm{supp}(\boldsymbol{\Omega}^*)$ using the Kendall's or Spearman's correlation matrices $\hat{\boldsymbol{\rho}}^K, \hat{\boldsymbol{\rho}}^S$, defined by

$$\hat{\boldsymbol{\rho}}^K_{ij} = \sin\left(\frac{\pi}{2}\boldsymbol{r}^K_{ij}\right), \qquad \text{and} \qquad \hat{\boldsymbol{\rho}}^S_{ij} = 2\sin\left(\frac{\pi}{6}\boldsymbol{r}^S_{ij}\right), \tag{18}$$

respectively, as inputs to the CLIME (10) or GLasso (8). Indeed, Liu et al. [42] and Xue and Zou [66] proposed to plug in the correlation matrix estimators (18) into regularization routines for precision matrix estimation under the nonparanormal graphical model; in their case, the model under study is only identifiable up to centering and scaling, so a scale estimate is not necessary. However, note that the conditions imposed for support recovery would need to hold for the correlation matrix $\boldsymbol{\rho}^*$, rather than for the precision matrix $\boldsymbol{\Omega}^*$. In particular, a minimum signal strength requirement on $\boldsymbol{\rho}^*$ is stronger than the same requirement imposed on $\boldsymbol{\Omega}^*$, since the latter can scale inversely with the standard deviations of individual variables in the joint distribution. We have therefore chosen to focus our attention in this paper on the output of the CLIME and GLasso when applied to an estimate of the covariance instead of the correlation matrix.

3.2. Special cases: Robust estimators

We begin by considering the robust scale estimates $\{\hat{\sigma}_i\}$. In particular, we show that the MAD estimator satisfies the required error bounds. Define the expression

$$c(\sigma_i) = \frac{15}{64\sqrt{2\pi}\sigma_i}\exp\left(-\frac{(1.1\sigma_i + 0.5)^2}{2\sigma_i^2}\right), \qquad \forall 1 \le i \le p, \tag{19}$$

and let

$$c_{\min} := \min\{c(\sigma_i) : \sigma_i > 0\}.$$

We have the following result:

Proposition 1. *[MAD estimator] Suppose* $\epsilon := \max_{1\le i\le p}\epsilon_i \le \frac{1}{16}$. *Also suppose* $C' > \frac{1}{\Phi^{-1}(0.75)c_{\min}\sqrt{2}}$ *and* $\Phi^{-1}(0.75)C'\sqrt{\frac{\log p}{n}} < 1$. *Then with probability at least* $1 - 6p^{-\{2[\Phi^{-1}(0.75)]^2 C'^2 c^2_{\min}-1\}}$, *inequality* (11b) *holds with* $C_1' = C'$; *and inequality* (12b) *holds with* $C_2' = 7.2M_\sigma$, *where* $\bar{\sigma}_i = E(\sigma_i)$.

We now turn to correlation matrix estimators. Our first result provides a bound on the statistical error of the robust covariance estimator $\hat{\boldsymbol{\Sigma}}^K$ based on Kendall's tau correlations. Note that our result does *not* involve any assumptions on the contaminating distribution H. Thus, the distribution H may contain point masses, and we do not require a probability density function of H to even exist.

Corollary 1. *[Kendall's tau correlation] Under the cellwise contamination model* (1), *suppose inequality* (13) *is satisfied, and* $\epsilon = \max_{1\le i\le p}\epsilon_i \le 0.02$.

Let $C > \pi\sqrt{2}$ *and* $C' > \frac{1}{\Phi^{-1}(0.75)c_{\min}\sqrt{2}}$, *and suppose*

$$\max\left\{C\sqrt{\frac{\log p}{n}} + 26\pi\epsilon,\ C'\sqrt{\frac{\log p}{n}} + 7.2M_\sigma\epsilon\right\} \le 1, \tag{20}$$

and $\Phi^{-1}(0.75)C'\sqrt{\frac{\log p}{n}} < 1$. *Then with probability at least*

$$1 - 2p^{-\left(\frac{C^2}{\pi^2}-2\right)} - 6p^{-\{2[\Phi^{-1}(0.75)]^2C'^2c_{\min}^2-1\}},$$

the Kendall's tau correlation estimator $\hat{\boldsymbol{\rho}}^K$ *satisfies conditions* (11a) *and* (12a) *with*

$$C_1 = C(M_\sigma^2 + M_\sigma + 1) + C'(2M_\sigma + 1), \quad \text{and} \quad C_2 = 97M_\sigma^2 + 89M_\sigma + 82,$$

where $\bar{\boldsymbol{\rho}}_{ij} = \sin\left(\frac{\pi}{2}E(\boldsymbol{r}_{ij}^K)\right)$. *Hence, the GLasso and CLIME estimators based on* $\hat{\boldsymbol{\Sigma}}$ *yield precision matrix estimators satisfying the error bound*

$$\|\hat{\boldsymbol{\Omega}} - \boldsymbol{\Omega}^*\|_\infty \le C_1''\sqrt{\frac{\log p}{n}} + C_2''\epsilon.$$

Note that although the restriction $\epsilon \le 0.02$ may seem somewhat prohibitive, the proof of Theorem 1 reveals that the specific bound on ϵ is an artifact of the proof technique, and a more careful analysis would allow for a larger degree of contamination, at the expense of slightly looser constants C_1 and C_2, as long as ϵ is bounded by some constant in $[0, 1)$. The proof of Corollary 1 is provided in Section 5.1.

The next result is an analog of Corollary 1, derived for the robust covariance estimator $\hat{\boldsymbol{\Sigma}}^S$ based on Spearman's correlation coefficient. We assume that the ranks of variables between samples are distinct; note that this happens almost surely when the contaminating distribution has continuous density. The proof of Corollary 2 is provided in Section 5.2.

Corollary 2. *[Spearman's rho correlation] Under the cellwise contamination model* (1), *suppose the variable ranks are distinct. Also suppose inequality* (13) *is satisfied and* $\epsilon = \max_{1\le i\le p}\epsilon_i \le 0.01$. *Let* $C > 8\pi$ *and* $C' > \frac{1}{\Phi^{-1}(0.75)c_{\min}\sqrt{2}}$, *and suppose*

$$\max\left\{\frac{5C}{2}\sqrt{\frac{\log p}{n}} + 51\pi\epsilon,\ C'\sqrt{\frac{\log p}{n}} + 7.2M_\sigma\epsilon\right\} \le 1,$$

and the sample size satisfies $\Phi^{-1}(0.75)C'\sqrt{\frac{\log p}{n}} < 1$ *and* $n \ge \max\left\{15,\ \frac{144\pi^2}{C^2\log p}\right\}$. *Then with probability at least*

$$1 - 2p^{-\left(\frac{C^2}{32\pi^2}-2\right)} - 6p^{-\{2[\Phi^{-1}(0.75)]^2C'^2c_{\min}^2-1\}},$$

the Spearman's rho correlation estimator $\hat{\boldsymbol{\rho}}^S$ *satisfies the conditions* (11a) *and* (12a) *with*

$$C_1 = \frac{5C}{2}(M_\sigma^2 + M_\sigma + 1) + C'(2M_\sigma + 1), \quad \text{and} \quad C_2 = 175M_\sigma^2 + 168M_\sigma + 161.$$

Hence, the GLasso and CLIME estimators based on $\hat{\boldsymbol{\Sigma}}$ *yield precision matrix estimators satisfying the error bound*

$$\|\hat{\boldsymbol{\Omega}} - \boldsymbol{\Omega}^*\|_\infty \leq C_1'' \sqrt{\frac{\log p}{n}} + C_2'' \epsilon.$$

Remark 3. *The conclusion of Corollary 2 is very similar to that of Corollary 1, except for constants and an additional requirement on the size of* n*. However, note that when* $\frac{\log p}{n} = o(1)$*, implying the statistical consistency of the robust covariance estimator, the requirement* $n \geq \max\left\{15, \frac{144\pi^2}{C^2 \log p}\right\}$ *is essentially extraneous.*

It is worth noting that although the high-dimensional error bounds derived in Corollaries 1 and 2 are substantially different from the canonical measures analyzed in the robust statistics literature, our bounds are somewhat related to the notion of the influence function of an estimator. The influence function [27], defined at the population level, measures the infinitesimal change incurred by the statistical functional associated with an estimator when the underlying distribution is contaminated by a point mass. Thus, an estimator has a bounded influence function if the extent of the deviation in its functional representation due to contamination remains bounded, regardless of the location of the point mass. The error bounds appearing in Corollaries 1 and 2 also reveal that the extent to which the error deviation between the robust covariance estimator and the true covariance grows is bounded by a constant depending only on M_σ. The two notions do not match precisely; for instance, our theorems allow contamination by an arbitrary distribution rather than simply a point mass, and we are comparing finite-sample deviations of an estimator from $\boldsymbol{\Sigma}^*$ rather than population-level deviations of a statistical functional under a contaminated distribution. However, note that by sending $n \to \infty$ in the finite-sample bounds and taking the contaminating distribution to be a point mass, we may conclude that the influence function of the robust covariance estimator is bounded when deviations are measured in the elementwise ℓ_∞-sense. Furthermore, the arguments in our proofs (cf. Lemmas 8 and 9 in Appendix D) may be used to derive the fact that the corresponding correlation estimators have a bounded influence function, the precise forms of which appear in Croux and Dehon [19].

The framework of Theorem 1 enables us to extend our analysis to other natural robust candidates for $\hat{\boldsymbol{\Sigma}}$, composed of entrywise correlation and scale estimates. To illustrate this point, we mention several examples below:

- **Quadrant correlation estimator.** The quadrant correlation estimator is defined by

$$r_{ij}^{Q} = \frac{1}{n}\sum_{k=1}^{n} \operatorname{sign}\left(X_{ki} - \operatorname*{median}_{1\le\ell\le n} X_{\ell i}\right) \operatorname{sign}\left(X_{kj} - \operatorname*{median}_{1\le\ell\le n} X_{\ell j}\right),$$

and is also known to have bounded influence [54]. One can show that the quadrant correlation estimator also satisfies the inequalities (11a) and (12a) appearing in Theorem 1; the derivations are similar to those employed for Kendall's tau and Spearman's rho correlation, so we do not provide the details here.

- **Gnanadesikan-Kettenring estimator.** Tarr et al. [58] and Oellerer and Croux [46] also propose to use the following estimator for pairwise covariances: Noting that

$$\operatorname{Cov}(X,Y) = \frac{1}{4\alpha\beta}\left[\operatorname{Var}(\alpha X + \beta Y) - \operatorname{Var}(\alpha X - \beta Y)\right],$$

the proposal is to replace the variance estimator by a robust variance estimator (e.g., the square of the MAD estimator). The drawback of this estimator in comparison to the covariance estimators based on Kendall's tau and Spearman's rho is that the covariance estimator has a maximal breakdown point of 25% under cellwise contamination, since the argument in the variance involves a sum of variables, and any robust variance estimator has a maximal breakdown point of 50%. However, from the point of view of statistical consistency, the Gnanadesikan-Kettenring covariance estimator may be seen to perform equally well. Indeed, consider the covariance estimator

$$\frac{1}{4}\left(\hat{\sigma}^2_{(i,j),+} - \hat{\sigma}^2_{(i,j),-}\right), \tag{21}$$

where $\hat{\sigma}_{(i,j),+}$ is the (rescaled) MAD statistic computed from $\{X_{ki} + X_{kj} : 1 \le k \le n\}$, and $\hat{\sigma}_{(i,j),-}$ is analogously defined to be the MAD statistic computed from $\{X_{ki} - X_{kj} : 1 \le k \le n\}$. Then our derivations showing the consistency of the MAD estimator (cf. Lemmas 6 and 7, with minor modifications) can be used to derive the requisite deviation bounds for $\hat{\sigma}_{(i,j),+}$ and $\hat{\sigma}_{(i,j),-}$, from which we may conclude that the pairwise covariance estimator (21) satisfies the deviation bounds (11b) and (12b), as well.

- Q_n **estimator.** Finally, consider the Q_n scale estimator [52], defined by

$$Q_n = c\{|X_k - X_\ell| : k < \ell\}_{(k^*)},$$

where c is a constant factor and $k^* = \lceil \binom{n}{2}/4 \rceil$. The Q_n estimator is also known to have a bounded influence property for real-valued data. Since the Q_n estimator is also based on quantiles, essentially the same types of arguments used to derive MAD concentration (cf. Appendix C) may be used to establish the desired bounds (11b) and (12b) appearing in Theorem 1.

Finally, note that although Theorems 1 and 2 have been derived under the assumption that the uncontaminated data follow a normal distribution, the same

proof techniques may be applied to settings where the uncontaminated data are drawn from a different underlying distribution, as long as the uncontaminated distribution is suitably well-behaved. Since our primary goal is precision matrix estimation, we have focused only on the scenario where the uncontaminated data are drawn from a Gaussian distribution, in which case the structure of the precision matrix is of great interest in the statistical community.

3.3. Special cases: Contamination models

We now interpret the conclusions of our theorems in some concrete contamination settings of interest.

Constant fraction of outliers We first briefly discuss the most basic setting of cellwise contamination, to emphasize the generality of our results. Following the model (1), suppose each entry of the data matrix $\mathbf{X}$ is contaminated independently with probability ϵ. Furthermore, either all contaminated entries may be drawn independently from a fixed contaminating distribution, or the contaminated entries in each row may be drawn jointly from a fixed contaminating distribution. In each case, Corollaries 1 and 2, in conjunction with Theorems 2 and 3, provide elementwise error bounds on the robust precision matrix estimators constructed from the CLIME and GLasso. The strength of the theorems lies in the fact that we do not make any side assumptions about the outlier distribution; it may be heavy-tailed and/or contain point masses. Hence, whereas statistics such as the sample covariance and sample correlation will have slower rates of convergence due to a constant fraction of outliers drawn from an ill-behaved distribution, their robust counterparts are agnostic to the outlier distribution.

It is also important to note that the error bounds for covariance and precision matrix estimation continue to hold when $\epsilon > C\sqrt{\frac{\log p}{n}}$. The difference is that in such scenarios, the statistical error will be of the order $\mathcal{O}(\epsilon)$ rather than $\mathcal{O}\left(\sqrt{\frac{\log p}{n}}\right)$. However, the effect of an ϵ fraction of outliers nonetheless grows only linearly as a function of ϵ. This emphasizes the robustness properties of the covariance and precision matrix estimators studied in our paper.

Missing data Turning to a somewhat different setting, note that missing data may also be seen as an instance of cellwise contamination. In this model, data are missing completely at random (MCAR), meaning that the probability of missingness is independent of the location of the unobserved entry of the data matrix [40]. In other words, if we observe the matrix $\mathbf{X}^{\text{mis}}$ with missing entries, where the probability that an entry in column i is missing is equal to ϵ_i, we have

$$X_{ki}^{\text{mis}} = \begin{cases} Y_{ki}, & \text{with probability } 1-\epsilon_i, \\ \text{missing}, & \text{with probability } \epsilon_i, \end{cases} \tag{22}$$

where $\mathbf{Y}$ is the fully-observed matrix. Note that if we zero-fill the missing entries of $\mathbf{X}^{\text{mis}}$, the resulting matrix $\mathbf{X}$ exactly follows the cellwise contamination

model (1), with $\mathbf{Z}_k = \mathbf{0}$ for all k. The following result is an immediate consequence of our theorems:

Corollary 3. *Suppose data are drawn from the missing data model (22), and the matrix* $\mathbf{X}$ *is the zero-filled data matrix. Let* $\epsilon = \max_{1 \le i \le p} \epsilon_i$. *Under the same conditions as in Theorem 2, we have*

$$\|\hat{\mathbf{\Omega}} - \mathbf{\Omega}^*\|_\infty \le 4\lambda \left\|\mathbf{\Omega}^*\right\|_1,$$

for the robust CLIME estimator constructed from $\mathbf{X}$. *Under the same conditions as in Theorem 3, we have* $\operatorname{supp}(\hat{\mathbf{\Omega}}) \subseteq \operatorname{supp}(\mathbf{\Omega}^*)$ *and*

$$\|\hat{\mathbf{\Omega}} - \mathbf{\Omega}^*\|_\infty \le \kappa_{\Gamma^*} \left(1 + \frac{8}{\alpha}\right) \lambda,$$

for the robust GLasso estimator constructed from $\mathbf{X}$.

Note that the conclusion of Corollary 3 does not require the matrix $\mathbf{X}$ to be zero-filled for missing values; in fact, we could fill the missing entries with samples generated according to any distribution (as long as the distribution remains the same across rows). This is because the missing entries are treated as outliers. Of course, our bounds should only be interpreted up to constant factors, and filling missing entries in a strategic way, e.g., filling entries in column i with the mean $E(X_{ki})$, could lead to smaller estimation error in practice.

Rowwise contamination Although we have thus far assumed that data are contaminated according to a cellwise mechanism, we now show that the same results apply for rowwise contamination, as well. Recall that each row in the data matrix for the rowwise contamination model with contamination level ϵ is given by

$$\mathbf{X}_k = (1 - B_k)\mathbf{Y}_k + B_k \mathbf{Z}_k, \qquad \forall 1 \le k \le n, \tag{23}$$

where $\mathbf{Y}_k$ is the uncontaminated row vector, $\mathbf{Z}_k$ is the contamination vector, and $B_k \sim \text{Bernoulli}(\epsilon)$.

Although model (23) differs from model (1), all we need to do is verify the deviation bounds in Theorem 1 under this model. For instance, equation (34) in the proof of Corollary 1 may be replaced by the equation

$$(X_{ki}, X_{kj}) \overset{\text{i.i.d.}}{\sim} F_{ij} = (1 - \epsilon)\Phi_{\boldsymbol{\mu}_{\{i,j\}}, \mathbf{\Sigma}_{\{i,j\}}} + \epsilon H_{ij}, \qquad \forall 1 \le k \le n. \tag{24}$$

Equation (24) comes from the fact that the pair is either drawn jointly from a normal distribution with probability $1 - \epsilon$, or from the contaminating distribution with probability ϵ. Then the remainder of the argument follows as before, implying that the same conclusion of Corollary 1 applies. We therefore arrive at the following result:

Corollary 4. *Under the rowwise contamination model (23), the same conclusions as in Corollary 3 hold for the CLIME and GLasso estimators constructed from* $\mathbf{X}$.

We emphasize that the rowwise contamination model (23) is *not* in general a special case of the cellwise contamination model (1); rather, the proof techniques for analyzing the cellwise model may be used to handle the rowwise model, as well.

3.4. Optimality of bounds

Finally, we discuss the optimality of the bounds obtained in Theorems 2 and 3. We have the following result:

Theorem 4. *Suppose* $\frac{d^2 \log p}{n} = o(1)$. *There exist absolute constants* $C, c > 0$ *such that*

$$\inf_{\hat{\mathbf{\Omega}}} \sup_{\mathbf{\Omega}^* \in \mathcal{U}, H^*} P_{(\epsilon, \mathbf{\Omega}^*, H^*)} \left(\|\hat{\mathbf{\Omega}} - \mathbf{\Omega}^*\|_\infty \geq C \left(\sqrt{\frac{\log p}{n}} \vee \epsilon \right) \right) \geq c,$$

for all $\epsilon \in [0, 1]$.

In the statement of the theorem, we use $P_{(\epsilon, \mathbf{\Omega}^*, H^*)}$ to denote the probability taken with respect to data drawn from a cellwise ϵ-contaminated model, where the uncontaminated distribution is $N(0, (\mathbf{\Omega}^*)^{-1})$ and the contaminating distribution is H^*. The proof of Theorem 4 uses ideas from Chen et al. [17], and the main step is to demonstrate that the pair of multivariate normal distributions with precision matrices $\Theta_1 = I$ and $\Theta_2 = I + 2\epsilon 11^T$ have the property that the ϵ-contamination balls around the two distributions have nontrivial intersection, so the clean distributions cannot be distinguished from contaminated data. Full details are provided in Section 5.3.

As pointed out by a reviewer, the lower bound provided in Theorem 4 is very much a worst-case bound and may not be tight when the class of contaminated models is further restricted. For instance, in the case of missing data according to the MCAR model (cf. Section 3.3), the sample size is effectively reduced to $(1 - \epsilon)n$. Hence, we would expect the minimax rate to be $\sqrt{\frac{\log p}{(1-\epsilon)n}}$, which is smaller than the $\sqrt{\frac{\log p}{n}} \vee \epsilon$ rate derived in Theorem 4 for, say, a constant fraction of missing data. As Theorem 4 reveals, however, classes of ϵ-contaminated distributions exist for which the upper bounds are tight.

4. Breakdown point

Our last topic concerns the breakdown point of the estimators studied in this paper. As discussed in Donoho and Huber [22] and Hampel et al. [28], breakdown analysis concerns the *global* behavior of a procedure, under large departures from an assumed situation. On the other hand, the theoretical analysis of statistical consistency and efficiency are related to notions of infinitesimal robustness, and quantifies the *local* behavior of a procedure at or near the assumed situation. Donoho and Huber [22] draw an analogy between the fields

of material science and statistics, where the notions of stiffness (resistance of a material to displacements caused by a small load) and breaking strength (the amount of load required to make the material fracture) parallel those of the influence function and the breakdown point. Ideally, a procedure should perform well both locally and globally; optimizing either measure alone is unwise. Our key result of this section shows that although the GLasso and CLIME estimators both enjoy roughly the same statistical rate of estimation, the CLIME does *not* perform as well as the GLasso when the breakdown point is used to quantify the degree of robustness. In both cases, we use the theoretically optimal choice of regularization parameter $\lambda^* = C_1''\sqrt{\frac{\log p}{n}} + C_2''\epsilon$.

Our analysis of the GLasso estimator closely follows that of Oellerer and Croux [46]; however, since the specific precision matrix estimators analyzed in our paper differ slightly, we include the full argument for the sake of completeness. We define the finite-sample breakdown point of the precision matrix estimator under cellwise contamination to be

$$\epsilon_n(\hat{\mathbf{\Omega}}, \mathbf{X}) := \min_{1 \le m \le n} \left\{ \frac{m}{n} : \sup_{\mathbf{X}^m} D(\hat{\mathbf{\Omega}}(\mathbf{X}), \hat{\mathbf{\Omega}}(\mathbf{X}^m)) = \infty \right\}, \tag{25}$$

where

$$D(\mathbf{A}, \mathbf{B}) := \max\left\{ |\lambda_1(\mathbf{A}) - \lambda_1(\mathbf{B})|, |\lambda_p^{-1}(\mathbf{A}) - \lambda_p^{-1}(\mathbf{B})| \right\},$$

and $\mathbf{X}^m$ is a data matrix obtained from $\mathbf{X}$ by replacing at most m entries in each column by arbitrary elements. We also define the explosion finite sample breakdown point of a covariance matrix estimator as follows:

$$\epsilon_n^+(\mathbf{S}, \mathbf{X}) := \min_{1 \le m \le n} \left\{ \frac{m}{n} : \sup_{\mathbf{X}^m} |\lambda_1(\mathbf{S}(\mathbf{X})) - \lambda_1(\mathbf{S}(\mathbf{X}^m))| = \infty \right\} \tag{26}$$

(cf. Maronna and Zamar [45]). Note that the explosion breakdown point only accounts for maximum eigenvalues, whereas the overall covariance matrix estimator breaks down under explosion or *implosion* (i.e., arbitrarily small minimum eigenvalues). Also, the breakdown point under cellwise contamination is less than or equal to the breakdown point under rowwise contamination.

We will consider the breakdown behavior of a slightly tweaked version of the GLasso presented earlier. Consider the matrix

$$\check{\mathbf{\Sigma}}(\mathbf{X}) := \operatorname*{argmin}_{\mathbf{M} \succeq 0} \|\hat{\mathbf{\Sigma}} - \mathbf{M}\|_\infty, \tag{27}$$

where $\hat{\mathbf{\Sigma}} = \hat{\mathbf{\Sigma}}(\mathbf{X})$ is the robust covariance matrix estimator constructed from the data matrix $\mathbf{X}$. Let

$$\check{\mathbf{\Omega}}(\mathbf{X}) := \operatorname*{argmin}_{\mathbf{\Omega} \succ 0} \left\{ \operatorname{tr}(\check{\mathbf{\Sigma}}\mathbf{\Omega}) - \log\det(\mathbf{\Omega}) + \lambda \|\mathbf{\Omega}\|_{1,\text{off}} \right\} \tag{28}$$

be the corresponding GLasso estimator. Note that from a computational standpoint, the projection step (27) is important so that fast solvers for the GLasso program (28) may be applied [26]. Furthermore, the projection step (27) is

convex, and the additional computational time is negligible compared to the computation required for running the GLasso. We have the following result, proved in Section 5.4:

Theorem 5. *Consider the positive semidefinite version of the robust GLasso estimator* (28) *with regularization parameter* $\lambda = 2\lambda^*$. *Under the same conditions as in Theorem 3, we have* $\operatorname{supp}(\check{\mathbf{\Omega}}) \subseteq \operatorname{supp}(\mathbf{\Omega}^*)$ *and*

$$\|\check{\mathbf{\Omega}} - \mathbf{\Omega}^*\|_\infty \leq 4\kappa_{\Gamma^*}\left(1 + \frac{8}{\alpha}\right)\lambda. \tag{29}$$

Furthermore, for any data matrix $\mathbf{X} \in \mathbb{R}^{n\times p}$, *the breakdown point satisfies* $\epsilon_n(\check{\mathbf{\Omega}}, \mathbf{X}) = 50\%$.

Remark 4. *Note that Theorem 5 guarantees that the robust GLasso estimator* $\check{\mathbf{\Omega}}$ *obtained from a semidefinite projection of the robust covariance estimator shares the same level of statistical consistency achieved by the robust GLasso estimator* $\hat{\mathbf{\Omega}}$. *In addition, the precision matrix estimator* $\check{\mathbf{\Omega}}$ *has a breakdown point of 50%. Although other authors also suggest projecting the robust covariance estimator onto the positive semidefinite cone before applying the GLasso [46, 58], they advocate a projection in terms of the Frobenius norm rather than the* ℓ_∞*-norm in the optimization program* (27). *As can be seen in the proof of Theorem 5, minimizing the elementwise* ℓ_∞*-norm is much more natural from the point of view of statistical consistency, since it guarantees that the* ℓ_∞*-error between the precision matrix estimate and the true precision matrix grows by at most a factor of two.*

We now show that although the CLIME is as robust as the GLasso in terms of statistical consistency under the cellwise contamination model, it has much poorer breakdown behavior. Consider the CLIME estimator based on corrupted data:

$$\min \|\mathbf{\Omega}\|_1 \quad \text{s.t.} \quad \|\hat{\mathbf{\Sigma}}(\mathbf{X}^m)\mathbf{\Omega} - I\|_\infty \leq \lambda, \tag{30}$$

where $\hat{\mathbf{\Sigma}}(\mathbf{X}^m)$ is the robust covariance estimator based on a data matrix with at most m arbitrarily corrupted entries per column. Since the CLIME estimator arises as the solution to a constrained linear program, the solution is undefined (infinite) when the problem is infeasible. Indeed, we will show in the following theorem that such a case may arise even by corrupting at most *one* entry in each column of the data matrix.

Theorem 6. *Suppose* $p = 2$. *There exists* $\mathbf{X} \in \mathbb{R}^{n\times p}$ *such that* $\epsilon_n(\hat{\mathbf{\Omega}}, \mathbf{X}) = \frac{1}{n}$, *where* $\hat{\mathbf{\Omega}}$ *denotes the CLIME estimator with regularization parameter* $\lambda = \lambda^*$.

The proof of Theorem 6, supplied in Section 5.5, provides the construction of a data matrix $\mathbf{X} \in \mathbb{R}^{n\times p}$ where the CLIME estimator becomes infeasible after perturbing a single entry in each column. This is in stark contrast to the result in Theorem 5, which establishes that the breakdown point of the robust GLasso estimator is 50%, for *any* data matrix $\mathbf{X}$. We note that in order to avoid the infeasibility problem that arises in the construction of Theorem 6, one could instead consider a version of the CLIME estimator where the regularization

parameter λ is chosen adaptively. However, we note that the estimator thus derived could then have vacuous guarantees from the point of view of statistical consistency, which is the property we are chiefly concerned about in this paper.

The conclusion of Theorem 6 underscores the fact that consistency and breakdown point under cellwise contamination are in some sense orthogonal measures of robustness. As demonstrated in the previous section, the robust CLIME and GLasso both enjoy good rates of statistical consistency when the contamination fraction ϵ is sufficiently small relative to the problem parameters. On the other hand, the results of this section show that the CLIME is extremely non-robust in terms of its breakdown point. Similarly, procedures such as the Gnanadesikan-Kettenring estimator (21) may be shown to be statistically consistent under cellwise contamination (cf. Section 3.2), but as discussed in Oellerer and Croux [46], the breakdown point of the covariance estimator $\hat{\mathbf{\Sigma}}$ is at most 25%, which leads to error propagation in $\hat{\mathbf{\Omega}}$.

Remark 5. *Several additional remarks are in order. As pointed out by a reviewer, alternative breakdown measures have been introduced in the literature specifically for quantify the robustness of covariance and correlation-type estimators [16, 9, 20]. For instance, one might redefine the breakdown point of a covariance (or precision) matrix estimator as the minimum fraction of samples that may be contaminated before changing the sign of individual entries of the matrix. This alternate definition could lead to a more refined notion of breakdown point that would lead to a more favorable comparison of the CLIME and GLasso estimators. However, such a study seems rather complicated, and we leave an analysis of this important question to future work.*

Finally, we note that the notion of breakdown point that we consider in equation (25) is defined with respect to a finite sample, without recourse to probability distributions. Other notions of breakdown point, defined with respect to an ϵ-contaminated distribution, have also been studied in the literature [28]. For some alternative measures of breakdown robustness, the CLIME estimator may have a more controlled breakdown behavior, but we have not explored them here.

5. Proofs

In this section, we provide an outline of the proofs of Corollaries 1 and 2 and Theorems 4, 5, and 6, which are the main technical contributions of our paper. Proofs of all supporting lemmas are contained in the supplement [43].

5.1. Proof of Corollary 1

We first verify the bound (11a). When $i = j$, we have $\bar{\boldsymbol{\rho}}_{ii}^K = \boldsymbol{\rho}_{ii} = 1$, and

$$\begin{aligned}\boldsymbol{r}_{ii}^K &= \frac{2}{n(n-1)} \sum_{k<\ell} \operatorname{sign}^2(X_{ki} - X_{\ell i}) = \frac{2}{n(n-1)} \sum_{k<\ell} (1 - \mathbb{1}(X_{ki} = X_{\ell i})) \\ &= 1 - \frac{2}{n(n-1)} \sum_{k<\ell} \mathbb{1}(X_{ki} = X_{\ell i}).\end{aligned}$$

Hence,

$$\begin{aligned}\left|\sin\left(\frac{\pi}{2}\boldsymbol{r}_{ii}^K\right)-\boldsymbol{\rho}_{ii}\right| &= \left|\sin\left(\frac{\pi}{2}-\frac{\pi}{n(n-1)}\sum_{k<\ell}\mathbb{1}(X_{ki}=X_{\ell i})\right)-1\right| \\ &= \left|\cos\left(\frac{\pi}{n(n-1)}\sum_{k<\ell}\mathbb{1}(X_{ki}=X_{\ell i})\right)-\cos(0)\right| \\ &\le \frac{\pi}{2}q_i,\end{aligned}$$

where

$$q_i=\frac{2}{n(n-1)}\sum_{k<\ell}\mathbb{1}(X_{ki}=X_{\ell i})$$

is a U-statistic, and the last inequality follows from the fact that $\cos(x)$ is 1-Lipschitz. By Hoeffding's inequality for U-statistics, we have

$$P\left(\left|\sin\left(\frac{\pi}{2}\boldsymbol{r}_{ii}^K\right)-1\right|\ge t\right)\le P\left(q_i\ge\frac{2t}{\pi}\right)\le\exp\left(-\frac{4nt^2}{\pi^2}\right). \tag{31}$$

Now suppose $i\neq j$. Since $\boldsymbol{r}_{ij}^K$ is a U-statistic with kernel bounded between -1 and 1, Hoeffding's inequality and the fact that $\sin(x)$ is 1-Lipschitz implies that

$$P\left(\left|\sin\left(\frac{\pi}{2}\boldsymbol{r}_{ij}^K\right)-\sin\left(\frac{\pi}{2}\boldsymbol{\rho}_{ij}^K\right)\right|\ge t\right)\le P\left(|\boldsymbol{r}_{ij}^K-\boldsymbol{\rho}_{ij}^K|\ge\frac{2}{\pi}t\right)\le 2\exp\left(-\frac{nt^2}{\pi^2}\right). \tag{32}$$

Combining inequalities (31) and (32) and taking $t=C\sqrt{\frac{\log p}{n}}$, we conclude that with probability at least $1-2p^{-(C^2/\pi^2-2)}$,

$$\max_{1\le i\le p}\left|\sin\left(\frac{\pi}{2}\boldsymbol{r}_{ii}^K\right)-\boldsymbol{\rho}_{ii}\right|\le C\sqrt{\frac{\log p}{n}},\qquad\text{and} \tag{33a}$$

$$\max_{i\neq j}\left|\sin\left(\frac{\pi}{2}\boldsymbol{r}_{ij}^K\right)-\sin\left(\frac{\pi}{2}\boldsymbol{\rho}_{ij}^K\right)\right|\le C\sqrt{\frac{\log p}{n}}. \tag{33b}$$

Turning to inequality (12a), we have under model (1) that for any pair $i\neq j$,

$$(X_{ki},X_{kj})\overset{\text{i.i.d.}}{\sim}F_{ij}=(1-\gamma_{ij})\Phi_{\boldsymbol{\mu}_{\{i,j\}},\boldsymbol{\Sigma}_{\{i,j\}}}+\gamma_{ij}H_{ij},\qquad\forall 1\le k\le n, \tag{34}$$

where $\Phi_{\boldsymbol{\mu}_{\{i,j\}},\boldsymbol{\Sigma}_{\{i,j\}}}=N(\boldsymbol{\mu}_{\{i,j\}},\boldsymbol{\Sigma}_{\{i,j\}})$ is the marginal distribution of (Y_{ki},Y_{kj}), H_{ij} is a mixture of the distributions of Y_{ki},Y_{kj},Z_{ki}, and Z_{kj}, and $1-\gamma_{ij}=(1-\epsilon_i)(1-\epsilon_j)$.

By Lemma 8, we have $\boldsymbol{\rho}_{ij}^K=\frac{2}{\pi}\sin^{-1}\boldsymbol{\rho}_{ij}+R_{ij}$, where $|R_{ij}|\le 12\gamma_{ij}+17\gamma_{ij}^2$. Setting $R'_{ij}=\frac{\pi}{2}R_{ij}$, we then have

$$\begin{aligned}\left|\sin\left(\frac{\pi}{2}\boldsymbol{\rho}_{ij}^K\right)-\boldsymbol{\rho}_{ij}\right| &= \left|\sin\left(\sin^{-1}(\boldsymbol{\rho}_{ij})+R'_{ij}\right)-\boldsymbol{\rho}_{ij}\right| \\ &= \left|\sin(\sin^{-1}(\boldsymbol{\rho}_{ij}))\cos(R'_{ij})+\cos(\sin^{-1}(\boldsymbol{\rho}_{ij}))\sin(R'_{ij})-\boldsymbol{\rho}_{ij}\right|\end{aligned}$$

$$
\begin{aligned}
&= \left|\boldsymbol{\rho}_{ij}\cos(R'_{ij}) + \sqrt{1-\boldsymbol{\rho}_{ij}^2}\sin(R'_{ij}) - \boldsymbol{\rho}_{ij}\right| \\
&\le \left|\boldsymbol{\rho}_{ij}\left(1-\cos(R'_{ij})\right)\right| + \left|\sqrt{1-\boldsymbol{\rho}_{ij}^2}\sin(R'_{ij})\right| \\
&\le \left[1-\cos(R'_{ij})\right] + \left|\sin(R'_{ij})\right|.
\end{aligned}
$$

Note that $\gamma_{ij} = \epsilon_i + \epsilon_j - \epsilon_i\epsilon_j \le 2\epsilon$, so

$$
|R'_{ij}| \le \frac{\pi}{2}(12\gamma_{ij} + 17\gamma_{ij}^2) \le \frac{\pi}{2}\left(12\cdot 2\epsilon + 17(2\epsilon)^2\right) = 12\pi\epsilon + 34\pi\epsilon^2.
$$

In particular, this bound is less than 1 when $\epsilon \le 0.02$. Then using the fact that $|\sin(x) - x| \le \frac{|x|^3}{3!}$ and $|1-\cos(x)| \le \frac{x^2}{2!}$ for $|x| \le 1$, we conclude that

$$
\begin{aligned}
\max_{1\le i,j\le p}\left|\sin\left(\frac{\pi}{2}\boldsymbol{\rho}_{ij}^K\right) - \boldsymbol{\rho}_{ij}\right| &\le \max_{1\le i,j\le p}\left[|R'_{ij}| + \frac{(R'_{ij})^2}{2} + \frac{|R'_{ij}|^3}{6}\right] \\
&\le 2\max_{1\le i,j\le p}|R'_{ij}| \le 26\pi\epsilon. \qquad (35)
\end{aligned}
$$

Note that we require the bound $\epsilon \le 0.02$ on the level of contamination, but the requirement could be relaxed with a more refined proof technique.

5.2. *Proof of Corollary 2*

We will prove the inequalities with $\bar{\boldsymbol{\rho}}_{ij} = 2\sin\left(\frac{\pi}{6}\left(E(\boldsymbol{r}_{ij}^S) - \Delta\right)\right)$, where Δ is the population-level quantity defined in equation (71), satisfying $|\Delta| \le \frac{12}{n+1}$.

We begin by verifying inequality (11a). When $i = j$, we have $2\sin(\frac{\pi}{6}\boldsymbol{r}_{ii}^S) = 2\sin(\frac{\pi}{6}\boldsymbol{\rho}_{ij}^S) = \boldsymbol{\rho}_{ii} = 1$. We now consider the case when $i \ne j$. By Lemma 10, we have $\boldsymbol{r}_{ij}^S = \frac{n-2}{n+1}U_{ij} + \frac{3}{n+1}\boldsymbol{r}_{ij}^K$, where U_{ij} is a U-statistic with kernel bounded between -3 and 3, and $\boldsymbol{r}_{ij}^K$ is the Kendall's tau correlation. Using the fact that $\sin(x)$ is 1-Lipschitz, we then have

$$
\begin{aligned}
&P\left(2\left|\sin\left(\frac{\pi}{6}\boldsymbol{r}_{ij}^S\right) - \sin\left(\frac{\pi}{6}\left(E(\boldsymbol{r}_{ij}^S) - \Delta\right)\right)\right| \ge t\right) \\
&\quad\le P\left(|\boldsymbol{r}_{ij}^S - E(\boldsymbol{r}_{ij}^S) + \Delta| \ge \frac{3t}{\pi}\right) \\
&\quad= P\left(\left|\frac{n-2}{n+1}(U_{ij} - E(U_{ij})) + \frac{3}{n+1}(\boldsymbol{r}_{ij}^K - \boldsymbol{\rho}_{ij}^K) + \Delta\right| \ge \frac{3t}{\pi}\right) \\
&\quad\le P\left(|U_{ij} - E(U_{ij}))| + \frac{18}{n+1} \ge \frac{3t}{\pi}\right) \\
&\quad\le P\left(|U_{ij} - E(U_{ij})| \ge \frac{3t}{2\pi}\right),
\end{aligned}
$$

where the last inequality follows from the choice $t = C\sqrt{\frac{\log p}{n}}$ and the fact that

$\frac{18}{n+1} \leq \frac{3t}{2\pi}$ when $n \geq \frac{144\pi^2}{C^2 \log p}$. Furthermore, Hoeffding's inequality implies

$$P\left(|U_{ij} - E(U_{ij})| \geq \frac{3t}{2\pi}\right) \leq 2\exp\left(-2\left\lfloor\frac{n}{3}\right\rfloor\left(\frac{3t}{2\pi}\right)^2\frac{1}{6^2}\right) \leq 2\exp\left(-\frac{nt^2}{32\pi^2}\right).$$

Plugging in $t = C\sqrt{\frac{\log p}{n}}$ and using a union bound, we then have

$$\begin{aligned} P\left(\max_{1\leq i,j\leq p} 2\left|\sin\left(\frac{\pi}{6}\boldsymbol{r}_{ij}^S\right) - \sin\left(\frac{\pi}{6}E(\boldsymbol{r}_{ij}^S)\right)\right| \geq C\sqrt{\frac{\log p}{n}}\right) \\ \leq 2p^2\exp\left(-\frac{C^2\log p}{32\pi^2}\right) = 2p^{-\left\{\frac{C^2}{32\pi^2}-2\right\}}. \end{aligned} \tag{36}$$

For inequality (12a), note that under model (1), for any pair $i \neq j$, we have

$$(X_{ki}, X_{kj}) \overset{\text{i.i.d.}}{\sim} F_{ij} = (1-\gamma_{ij})\Phi_{\boldsymbol{\mu}_{\{i,j\}},\boldsymbol{\Sigma}_{\{i,j\}}} + \gamma_{ij}H_{ij}, \qquad \forall 1 \leq k \leq n,$$

where $\Phi_{\boldsymbol{\mu}_{\{i,j\}},\boldsymbol{\Sigma}_{\{i,j\}}} = N(\boldsymbol{\mu}_{\{i,j\}}, \boldsymbol{\Sigma}_{\{i,j\}})$ is the marginal distribution of (Y_{ki}, Y_{kj}), H_{ij} is a mixture of the distributions of Y_{ki}, Y_{kj}, Z_{ki}, and Z_{kj}, and $1-\gamma_{ij} = (1-\epsilon_i)(1-\epsilon_j)$.

By Lemma 9, we have $E(\boldsymbol{r}_{ij}^S) - \Delta = \frac{6}{\pi}\sin^{-1}\left(\frac{\rho_{ij}}{2}\right) + R_{ij}$, where $|R_{ij}| \leq 48\gamma_{ij} + 129\gamma_{ij}^2 + 88\gamma_{ij}^3$. Setting $R'_{ij} = \frac{\pi}{6}R_{ij}$, we then have

$$\begin{aligned} \left|\bar{\boldsymbol{\rho}}_{ij} - \boldsymbol{\rho}_{ij}\right| &= \left|2\sin\left(\sin^{-1}(\boldsymbol{\rho}_{ij}/2) + R'_{ij}\right) - \boldsymbol{\rho}_{ij}\right| \\ &= \left|2\sin(\sin^{-1}(\boldsymbol{\rho}_{ij}/2))\cos(R'_{ij}) + 2\cos(\sin^{-1}(\boldsymbol{\rho}_{ij}/2))\sin(R'_{ij}) - \boldsymbol{\rho}_{ij}\right| \\ &= \left|\boldsymbol{\rho}_{ij}\cos(R'_{ij}) + 2\sqrt{1-\boldsymbol{\rho}_{ij}^2/4}\cdot\sin(R'_{ij}) - \boldsymbol{\rho}_{ij}\right| \\ &\leq \left|\boldsymbol{\rho}_{ij}\left(1-\cos(R'_{ij})\right)\right| + 2\left|\sqrt{1-\boldsymbol{\rho}_{ij}^2/4}\cdot\sin(R'_{ij})\right| \\ &\leq \left[1-\cos(R'_{ij})\right] + 2\left|\sin(R'_{ij})\right|. \end{aligned}$$

Note that $\gamma_{ij} = \epsilon_i + \epsilon_j - \epsilon_i\epsilon_j \leq 2\epsilon$, so

$$\begin{aligned} |R'_{ij}| &\leq \frac{\pi}{6}\left(48\gamma_{ij} + 129\gamma_{ij}^2 + 88\gamma_{ij}^3\right) \\ &\leq \frac{\pi}{6}\left(48\cdot 2\epsilon + 129(2\epsilon)^2 + 88(2\epsilon)^3\right) \\ &\leq 16\pi\epsilon + 86\pi\epsilon^2 + 118\pi\epsilon^3. \end{aligned}$$

In particular, this bound is less than 1 when $\epsilon \leq 0.01$ and $n \geq 15$. Then using the fact that $|\sin(x) - x| \leq \frac{|x|^3}{3!}$ and $|\cos(x) - 1| \leq \frac{x^2}{2!}$ for $|x| \leq 1$, we conclude that

$$\max_{1\leq i,j\leq p}\left|\bar{\boldsymbol{\rho}}_{ij} - \boldsymbol{\rho}_{ij}\right| \leq \max_{1\leq i,j\leq p}\left[2|R'_{ij}| + \frac{(R'_{ij})^2}{2} + \frac{|R'_{ij}|^3}{3}\right]$$

$$\begin{aligned}
&\leq 3 \max_{1\leq i,j\leq p} |R'_{ij}| \\
&\leq 48\pi\epsilon + 258\pi\epsilon^2 + 354\pi\epsilon^3 \\
&\leq 51\pi\epsilon,
\end{aligned}$$

completing the proof.

5.3. *Proof of Theorem 4*

We adopt a proof technique from Chen et al. [17], where we relate the minimax bound for contaminated data to the the minimax rate for uncontaminated data. From Theorem 5 of Ren et al. [49], we have the following result:

Proposition 2. *Suppose* $\frac{d^2 \log p}{n} = o(1)$. *There exist constants* $C', c' > 0$ *such that*

$$\inf_{\hat{\mathbf{\Omega}}} \sup_{\mathbf{\Omega}^* \in \mathcal{U}} P_{\mathbf{\Omega}^*} \left(\|\hat{\mathbf{\Omega}} - \mathbf{\Omega}^*\|_\infty \geq C' \sqrt{\frac{\log p}{n}} \right) \geq c'.$$

Following the notation of Chen et al. [17], we let $\mathcal{M}(0) := C'\sqrt{\frac{\log p}{n}}$ denote the lower bound in the case of uncontaminated data. Clearly, we have

$$\begin{aligned}
&\inf_{\hat{\mathbf{\Omega}}} \sup_{\mathbf{\Omega}^* \in \mathcal{U}, H^*} P_{(\epsilon, \mathbf{\Omega}^*, H^*)} \left(\|\hat{\mathbf{\Omega}} - \mathbf{\Omega}^*\|_\infty \geq C' \sqrt{\frac{\log p}{n}} \right) \\
&\quad \geq \inf_{\hat{\mathbf{\Omega}}} \sup_{\mathbf{\Omega}^* \in \mathcal{U}} P_{(\epsilon, \mathbf{\Omega}^*, N(0, \mathbf{\Omega}^*))} \left(\|\hat{\mathbf{\Omega}} - \mathbf{\Omega}^*\|_\infty \geq C' \sqrt{\frac{\log p}{n}} \right) \\
&\quad = \inf_{\hat{\mathbf{\Omega}}} \sup_{\mathbf{\Omega}^* \in \mathcal{U}} P_{\mathbf{\Omega}^*} \left(\|\hat{\mathbf{\Omega}} - \mathbf{\Omega}^*\|_\infty \geq C' \sqrt{\frac{\log p}{n}} \right) \\
&\quad \geq c'. \qquad (37)
\end{aligned}$$

We now provide a lower bound on the term

$$\inf_{\hat{\mathbf{\Omega}}} \sup_{\mathbf{\Omega}^* \in \mathcal{U}, H^*} P_{(\epsilon, \mathbf{\Omega}^*, H^*)} \left(\|\hat{\mathbf{\Omega}} - \mathbf{\Omega}^*\|_\infty \geq \epsilon \right).$$

Consider $\Theta_1 = I$ and $\Theta_2 = 2\epsilon 11^T + \Theta_1$. Clearly, we have $\|\Theta_1 - \Theta_2\|_\infty = 2\epsilon$. Furthermore, we have the total variation bound

$$\begin{aligned}
TV(N(0, \Theta_1^{-1}), N(0, \Theta_2^{-1}))^2 &\leq \frac{1}{2} D_{KL}(N(0, \Theta_1^{-1})) \| N(0, \Theta_2^{-1})) \\
&\leq \frac{1}{8} \|\Theta_1^{-1} - \Theta_2^{-1}\|_F^2 \\
&= \frac{1}{8} \left(1 - \frac{1}{1+2\epsilon} \right)^2 < \left(\frac{\epsilon}{1-\epsilon} \right)^2.
\end{aligned}$$

Hence, there exists $\epsilon' < \epsilon$ such that $TV(N(0,\Theta_1^{-1}), N(0,\Theta_2^{-1})) = \frac{\epsilon'}{1-\epsilon'}$. We claim that there exist contaminating distributions H_1 and H_2 such that $P_{(\epsilon',\Omega_1,H_1)} = P_{(\epsilon',\Omega_2,H_2)}$. Indeed, note that both P_{Ω_1} and P_{Ω_2} are product distributions:

$$\begin{aligned} dP_{\Omega_1}(x_1,\ldots,x_p) &= p_1(x_1)p_1(x_2)\cdots p_1(x_p), \\ dP_{\Omega_2}(x_1,\ldots,x_p) &= p_{\epsilon'}(x_1)p_1(x_2)\cdots p_1(x_p), \end{aligned}$$

where $p_1(x)$ is a standard normal density and $p_{\epsilon'}(x)$ is the pdf of a $N\left(0, \frac{1}{1+\epsilon}\right)$ random variable. We define

$$\begin{aligned} h_1(x_1,\ldots,x_p) &= q_1(x_1)p_1(x_2)\cdots p_1(x_p), \\ h_2(x_1,\ldots,x_p) &= q_2(x_1)p_1(x_2)\cdots p_1(x_p), \end{aligned}$$

where

$$\begin{aligned} q_1(x_1) &= \frac{(p_2(x_1) - p_1(x_1))\,1\{p_2(x_1) \geq p_1(x_1)\}}{\epsilon'/(1-\epsilon')}, \\ q_2(x_1) &= \frac{(p_1(x_1) - p_2(x_1))\,1\{p_1(x_1) \geq p_2(x_1)\}}{\epsilon'/(1-\epsilon')}. \end{aligned}$$

Note that q_1 and q_2 are both valid pdfs, since

$$\begin{aligned} \int q_1(x_1)dx_1 &= \frac{1-\epsilon'}{\epsilon'} \cdot \int (p_2(x_1) - p_1(x_1))\,1\{p_2(x_1) \geq p_1(x_1)\}\,dx_1 \\ &= \left(\frac{1-\epsilon'}{\epsilon'}\right)\left(1 - \int (p_1(x_1) \wedge p_2(x_1))\,dx_1\right) = \frac{1-\epsilon'}{\epsilon'} \cdot TV(p_1,p_2) \\ &= 1, \end{aligned}$$

and similarly for q_2. We denote the distributions of h_1 and h_2 by H_1 and H_2, respectively. Furthermore, since

$$\begin{aligned} \frac{(1-\epsilon')p_1(x_1) + \epsilon' q_1(x_1)}{1-\epsilon'} &= p_1(x_1) + (p_2(x_1) - p_1(x_1))\,1\{p_2(x_1) \geq p_1(x_1)\} \\ &= p_1(x_1) + (p_2(x_1) - p_1(x_1))\,(1 - 1\{p_1(x_1) \geq p_2(x_1)\}) \\ &= p_2(x_1) + (p_1(x_1) - p_2(x_1))\,1\{p_1(x_1) \geq p_2(x_1)\} \\ &= \frac{(1-\epsilon')p_2(x_1) + \epsilon' q_2(x_1)}{1-\epsilon'}, \end{aligned}$$

we conclude that

$$\begin{aligned} (1-\epsilon')dP_{\Omega_1}(x_1,\ldots,x_p) + \epsilon' dH_1 &= \Big((1-\epsilon')p_1(x_1) + \epsilon' q_1(x_1)\Big)p_1(x_2)\cdots p_1(x_p) \\ &= \Big((1-\epsilon')p_2(x_1) + \epsilon' q_2(x_1)\Big)p_1(x_2)\cdots p_1(x_p) \\ &= (1-\epsilon')dP_{\Omega_2}(x_1,\ldots,x_p) + \epsilon' dH_2, \end{aligned}$$

so $P_{(\epsilon',\Omega_1,H_1)} = P_{\epsilon',\Omega_2,H_2}$, as claimed. Finally, note that since $\epsilon' < \epsilon$, we can construct H_1' and H_2' such that $P_{(\epsilon,\Omega_1,H_1')} = P_{(\epsilon',\Omega_1,H_1)}$ and $P_{(\epsilon,\Omega_2,H_2')} = P_{(\epsilon',\Omega_2,H_2)}$, simply by defining

$$dH_1' = \frac{\epsilon'}{\epsilon} \cdot dH_1 + \frac{\epsilon - \epsilon'}{\epsilon} \cdot dP_{\Omega_1},$$
$$dH_2' = \frac{\epsilon'}{\epsilon} \cdot dH_2 + \frac{\epsilon - \epsilon'}{\epsilon} \cdot dP_{\Omega_2}.$$

Then $P_{(\epsilon,\Omega_1,H_1')} = P_{(\epsilon,\Omega_2,H_2')}$, implying that the cellwise ϵ-contaminated models P_{Ω_1} and P_{Ω_2} have nontrivial intersection. By Le Cam's two-point hypothesis testing method [67, 60], we therefore conclude that

$$\inf_{\hat{\mathbf{\Omega}}} \sup_{\mathbf{\Omega}^* \in \mathcal{U}, H^*} P_{(\epsilon,\mathbf{\Omega}^*,H^*)} \left(\|\hat{\mathbf{\Omega}} - \mathbf{\Omega}^*\|_\infty \geq \epsilon \right) \geq \frac{1}{2}. \tag{38}$$

Combining inequalities (37) and (38), we then have

$$\inf_{\hat{\mathbf{\Omega}}} \sup_{\mathbf{\Omega}^* \in \mathcal{U}, H^*} P_{(\epsilon,\mathbf{\Omega}^*,H^*)} \left(\|\hat{\mathbf{\Omega}} - \mathbf{\Omega}^*\|_\infty \geq C' \sqrt{\frac{\log p}{n}} \vee \epsilon \right) \geq \min \left\{ c', \frac{1}{2} \right\}.$$

5.4. *Proof of Theorem 5*

Note that $\check{\mathbf{\Sigma}}$ is the projection of the robust covariance estimator $\hat{\mathbf{\Sigma}}$ onto the positive semidefinite cone, where the distance is measured in the elementwise ℓ_∞-norm. Furthermore, note that $\|\check{\mathbf{\Sigma}} - \hat{\mathbf{\Sigma}}\|_\infty \leq \|\mathbf{\Sigma}^* - \hat{\mathbf{\Sigma}}\|_\infty$, since $\mathbf{\Sigma}^* \succeq 0$. Hence,

$$\|\check{\mathbf{\Sigma}} - \mathbf{\Sigma}^*\|_\infty \leq \|\check{\mathbf{\Sigma}} - \hat{\mathbf{\Sigma}}\|_\infty + \|\hat{\mathbf{\Sigma}} - \mathbf{\Sigma}^*\|_\infty \leq 2\|\hat{\mathbf{\Sigma}} - \mathbf{\Sigma}^*\|_\infty. \tag{39}$$

Using Lemma 1, we immediately arrive at the bound (29).

Turning to the derivation of the breakdown point, note that by Theorem 1 of Oellerer and Croux [46], we have

$$\epsilon_n(\check{\mathbf{\Omega}}(\mathbf{X}), \mathbf{X}) \geq \epsilon_n^+(\check{\mathbf{\Sigma}}(\mathbf{X}), \mathbf{X}). \tag{40}$$

Consider the estimator $\check{\mathbf{\Sigma}}(\mathbf{X}^m)$, based on corrupted data. We have

$$\|\check{\mathbf{\Sigma}}(\mathbf{X}^m) - \mathbf{\Sigma}^*\|_\infty \leq 2\|\hat{\mathbf{\Sigma}}(\mathbf{X}^m) - \mathbf{\Sigma}^*\|_\infty \leq 2\|\hat{\mathbf{\Sigma}}(\mathbf{X}^m)\|_\infty + 2\|\mathbf{\Sigma}^*\|_\infty, \tag{41}$$

where the first inequality follows from the bound (39), and the second inequality comes from the triangle inequality. Furthermore, note that since $\check{\mathbf{\Sigma}}(\mathbf{X}^m) \succeq 0$ by construction,

$$\begin{aligned} \lambda_1(\check{\mathbf{\Sigma}}(\mathbf{X}^m)) = \left|\left|\left| \check{\mathbf{\Sigma}}(\mathbf{X}^m) \right|\right|\right|_2 &\leq \left|\left|\left| \check{\mathbf{\Sigma}}(\mathbf{X}^m) - \mathbf{\Sigma}^* \right|\right|\right|_2 + \left|\left|\left| \mathbf{\Sigma}^* \right|\right|\right|_2 \\ &\leq p\|\check{\mathbf{\Sigma}}(\mathbf{X}^m) - \mathbf{\Sigma}^*\|_\infty + \left|\left|\left| \mathbf{\Sigma}^* \right|\right|\right|_2, \end{aligned} \tag{42}$$

where we have used the bound $\|\mathbf{A}\|_\infty \leq |||\mathbf{A}|||_2 \leq p\|\mathbf{A}\|_\infty$, for all $\mathbf{A} \in \mathbb{R}^{p \times p}$, in the last inequality. Combining inequalities (41) and (42), we then obtain

$$\lambda_1(\check{\mathbf{\Sigma}}(\mathbf{X}^m)) \leq 2p\|\hat{\mathbf{\Sigma}}(\mathbf{X}^m)\|_\infty + 2p\|\mathbf{\Sigma}^*\|_\infty + \|\mathbf{\Sigma}^*\|_2\,,$$

so

$$\left|\lambda_1(\check{\mathbf{\Sigma}}(\mathbf{X}^m)) - \lambda_1(\check{\mathbf{\Sigma}}(\mathbf{X}))\right| \leq \lambda_1(\check{\mathbf{\Sigma}}(\mathbf{X})) + \left(2p\|\hat{\mathbf{\Sigma}}(\mathbf{X}^m)\|_\infty + 2p\|\mathbf{\Sigma}^*\|_\infty + \|\mathbf{\Sigma}^*\|_2\right). \tag{43}$$

Finally, since the correlation estimators are bounded in magnitude by 1, we have

$$\|\hat{\mathbf{\Sigma}}(\mathbf{X}^m)\|_\infty \leq \max_{1\leq i,j\leq p} \hat{\sigma}_i(\mathbf{X}^m)\hat{\sigma}_j(\mathbf{X}^m), \tag{44}$$

where $\{\hat{\sigma}_i(\mathbf{X}^m)\}_{1\leq i\leq p}$ are the robust scale estimators based on $\mathbf{X}^m$, given by the MAD estimators calculated from the corresponding columns. Furthermore, the breakdown point of the MAD is 50% [36], so the quantity on the right-hand side of inequality (44) is finite when $\frac{m}{n} < 50\%$. Then by inequality (43) and the definition of the explosion breakdown point, we conclude that $\epsilon_n^+(\check{\mathbf{\Sigma}}(\mathbf{X}), \mathbf{X}) \geq 50\%$. By inequality (40), we therefore have $\epsilon_n(\check{\mathbf{\Omega}}(\mathbf{X}), \mathbf{X}) \geq 50\%$, as well.

We now establish that $\epsilon_n(\check{\mathbf{\Omega}}(\mathbf{X}), \mathbf{X}) = 50\%$. Note that if we are allowed to corrupt more than 50% of the entries in each column of the data matrix, the columnwise MAD estimates may be made arbitrarily small (say, smaller than some value a); indeed, we may simply replace more than half of the entries in each column by values in $(0, a)$. Consequently, the overall covariance estimator $\hat{\mathbf{\Sigma}}(\mathbf{X}^m)$ will have all entries bounded in magnitude by $[\Phi^{-1}(0.75)]^{-2}a^2$. We claim that the diagonal elements of $\check{\mathbf{\Sigma}}(\mathbf{X}^m)$ must therefore be bounded in magnitude by $2[\Phi^{-1}(0.75)]^{-2}a^2$. Indeed, note that the matrix $\operatorname{diag}(\hat{\mathbf{\Sigma}}(\mathbf{X}^m))$ is feasible for the projection (27). Hence, we must have

$$\|\hat{\mathbf{\Sigma}}(\mathbf{X}^m) - \check{\mathbf{\Sigma}}(\mathbf{X}^m)\|_\infty \leq \|\hat{\mathbf{\Sigma}}(\mathbf{X}^m) - \operatorname{diag}(\hat{\mathbf{\Sigma}}(\mathbf{X}^m))\|_\infty \leq [\Phi^{-1}(0.75)]^{-2}a^2,$$

implying in particular that

$$\begin{aligned}\|\operatorname{diag}(\check{\mathbf{\Sigma}}(\mathbf{X}^m))\|_\infty &\leq \|\operatorname{diag}(\hat{\mathbf{\Sigma}}(\mathbf{X}^m))\|_\infty + \|\operatorname{diag}(\hat{\mathbf{\Sigma}}(\mathbf{X}^m)) - \operatorname{diag}(\check{\mathbf{\Sigma}}(\mathbf{X}^m))\|_\infty \\ &\leq \frac{2a^2}{[\Phi^{-1}(0.75)]^2},\end{aligned}$$

as claimed. Now note that the first-order optimality condition for the GLasso is given by

$$\check{\mathbf{\Sigma}}(\mathbf{X}^m) - \left(\check{\mathbf{\Omega}}(\mathbf{X}^m)\right)^{-1} + \lambda \cdot \operatorname{sign}\{\check{\mathbf{\Omega}}(\mathbf{X}^m) - \operatorname{diag}(\check{\mathbf{\Omega}}(\mathbf{X}^m))\} = 0,$$

where the sign function is computed entrywise, omitting the diagonal elements of $\check{\mathbf{\Omega}}(\mathbf{X}^m)$. In particular, this implies that the $\operatorname{diag}(\check{\mathbf{\Sigma}}(\mathbf{X}^m)) = \operatorname{diag}\left\{\left(\check{\mathbf{\Omega}}(\mathbf{X}^m)\right)^{-1}\right\}$, so the diagonal elements of $\left(\check{\mathbf{\Omega}}(\mathbf{X}^m)\right)^{-1}$ are also bounded in magnitude by $2[\Phi^{-1}(0.75)]^{-2}a^2$. Hence,

$$\lambda_p\left(\left(\check{\mathbf{\Omega}}(\mathbf{X}^m)\right)^{-1}\right) = \min_{\|\mathbf{v}\|_2=1} \mathbf{v}^T\left(\left(\check{\mathbf{\Omega}}(\mathbf{X}^m)\right)^{-1}\right)\mathbf{v} \leq \min_{1\leq j\leq p} \mathbf{e}_j^T\left(\left(\check{\mathbf{\Omega}}(\mathbf{X}^m)\right)^{-1}\right)\mathbf{e}_j$$

$$\leq \left\| \text{diag} \left\{ \left(\check{\mathbf{\Omega}}(\mathbf{X}^m) \right)^{-1} \right\} \right\|_\infty \leq 2[\Phi^{-1}(0.75)]^{-2} a^2,$$

where the $\mathbf{e}_j$'s are the canonical basis vectors, and we have used the variational representation of eigenvalues of a Hermitian matrix to show that the minimum eigenvalue is bounded by the minimum diagonal entry. This allows us to conclude that

$$\begin{aligned} 1 &= \lambda_p \left(\check{\mathbf{\Omega}}(\mathbf{X}^m) \left(\check{\mathbf{\Omega}}(\mathbf{X}^m) \right)^{-1} \right) \leq \lambda_1 \left(\check{\mathbf{\Omega}}(\mathbf{X}^m) \right) \lambda_p \left(\left(\check{\mathbf{\Omega}}(\mathbf{X}^m) \right)^{-1} \right) \\ &\leq \lambda_1 \left(\check{\mathbf{\Omega}}(\mathbf{X}^m) \right) \cdot \frac{2a^2}{[\Phi^{-1}(0.75)]^2}, \end{aligned}$$

where we have used the inequality $\lambda_p(\mathbf{A}\boldsymbol{B}) \leq \lambda_1(\mathbf{A})\lambda_p(\boldsymbol{B})$, for $\mathbf{A}, \boldsymbol{B} \succeq 0$, in the first inequality [69]. Hence, $\lambda_1 \left(\check{\mathbf{\Omega}}(\mathbf{X}^m) \right) \geq \frac{[\Phi^{-1}(0.75)]^2}{2a^2}$. However, we may choose a to be arbitrarily close to 0, implying that the maximum eigenvalue of $\check{\mathbf{\Omega}}(\mathbf{X}^m)$ may be made arbitrarily large, and the estimator breaks down. This concludes the proof.

5.5. *Proof of Theorem 6*

Clearly, $\epsilon_n(\hat{\mathbf{\Omega}}, \mathbf{X}) \geq \frac{1}{n}$ for any $\mathbf{X}$, by the definition of the breakdown point. To show equality, we now provide a data matrix X and a corrupted data matrix $\mathbf{X}^1$, where $\mathbf{X}^1$ differs from $\mathbf{X}$ in at most one element per column, and the CLIME problem is feasible for $\hat{\mathbf{\Sigma}}(\mathbf{X})$ but infeasible for $\hat{\mathbf{\Sigma}}(\mathbf{X}^1)$.

Let

$$\mathbf{X}^1 = \begin{pmatrix} a_1 & -a_1 \\ a_2 & -a_2 \\ \vdots & \vdots \\ a_n & -a_n \end{pmatrix},$$

where the a_k's are all distinct and a_n has the largest magnitude. Note that the columns of $\mathbf{X}^1$ are perfectly negatively correlated; hence, the correlation matrix (computed from either Kendall's tau or Spearman's rho, for instance) is $\begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}$. Furthermore, we have $\hat{\sigma}_1 = \hat{\sigma}_2 := \hat{\sigma}$, since the data in the two columns are negatives of each other. It follows that $\hat{\mathbf{\Sigma}}(\mathbf{X}^1) = \hat{\sigma}^2 \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}$. Clearly, the problem

$$\beta_1 : \left\| \hat{\mathbf{\Sigma}}(\mathbf{X}^1)\beta_1 - \begin{pmatrix} 1 \\ 0 \end{pmatrix} \right\|_\infty \leq \lambda \tag{45}$$

is infeasible for $\lambda < \frac{1}{2}$. Hence, the CLIME estimator based on $\hat{\mathbf{\Sigma}}(\mathbf{X}^1)$ is infeasible.

On the other hand, we may construct an initial data matrix $\mathbf{X}$ such that the CLIME program based on $\hat{\mathbf{\Sigma}}(\mathbf{X})$ is feasible, simply by altering the last row

of $\mathbf{X}^1$. Suppose we change the last row of $\mathbf{X}^1$ to (a_n, a_n). Then the columns are no longer perfectly negatively correlated, and it is easy to check that the correlation matrix of $\mathbf{X}$ will take the form $\begin{pmatrix} 1 & a \\ a & 1 \end{pmatrix}$, for some $|a| < 1$. Denoting the corresponding estimates of scale as $\hat{\sigma}_1$ and $\hat{\sigma}_2$, we then have

$$\hat{\mathbf{\Sigma}}(\mathbf{X}) = \begin{pmatrix} \hat{\sigma}_1^2 & a\hat{\sigma}_1\hat{\sigma}_2 \\ a\hat{\sigma}_1\hat{\sigma}_2 & \hat{\sigma}_2^2 \end{pmatrix}.$$

Note that $\det\{\hat{\mathbf{\Sigma}}(\mathbf{X})\} = \hat{\sigma}_1^2\hat{\sigma}_2^2(1-a^2) > 0$. It follows that $\hat{\mathbf{\Sigma}}(\mathbf{X})$ is invertible. In particular, the matrix $\left(\hat{\mathbf{\Sigma}}(\mathbf{X})\right)^{-1}$ is always a feasible point for the CLIME program based on $\hat{\mathbf{\Sigma}}(\mathbf{X})$.

Hence, we conclude that the CLIME program breaks down when even one corruption per column is allowed. It follows that $\epsilon_n(\hat{\mathbf{\Omega}}, \mathbf{X}) = \frac{1}{n}$ for the constructed value of $\mathbf{X}$.

6. Simulations

In this section, we perform simulation studies to examine the performance of the two robust covariance matrix estimators introduced in Section 2, and also the robust precision matrix estimators obtained using the GLasso. We will refer to the two type of estimators as `Kendall` and `Spearman`, respectively.

For comparison, we also compute the following robust covariance matrix estimators, which are similarly plugged into the GLasso to obtain robust precision matrix estimators:

Do not consider CLIME here

- `SpearmanU`: The pairwise covariance matrix estimator proposed in Oellerer and Croux [46], where the MAD estimator is combined with Spearman's rho (without transformation):

$$\hat{\mathbf{\Sigma}}_{ij} = \hat{\sigma}_i\hat{\sigma}_j\boldsymbol{r}_{ij}^S, \qquad \text{where} \qquad \hat{\sigma}_i = [\Phi^{-1}(0.75)]^{-1}\hat{d}_i.$$

- `OGK`: The OGK estimator proposed in Maronna and Zamar [45], with scale estimator Q_n.
- `NPD`: The pairwise covariance matrix estimator considered in Tarr et al. [58], where

$$\tilde{\mathbf{\Sigma}}_{ij} = \frac{1}{4}\left(\hat{\sigma}^2_{(i,j),+} - \hat{\sigma}^2_{(i,j),-}\right).$$

  Here, $\hat{\sigma}_{(i,j),+}$ is the Q_n statistic computed from $\{X_{ki} + X_{kj} : 1 \le k \le n\}$ and $\hat{\sigma}_{(i,j),-}$ is the Q_n statistic computed from $\{X_{ki} - X_{kj} : 1 \le k \le n\}$. An NPD projection is applied to $\tilde{\mathbf{\Sigma}}$ to obtain the final positive semidefinite covariance matrix estimator.

Further details for the orthogonalized Gnanedesikan-Kettenring (OGK) and nearest positive definite (NPD) procedures may be found in Maronna and Zamar [45] and Higham [32], respectively. The nonrobust GLasso, which takes

the sample covariance matrix estimator as an input (`SampleCov`), as well as the inverse sample covariance matrix estimator (`InvCov`), applicable in the case $p < n$, are used as points of reference. On the other hand, we also compare the precision matrix estimation performance of the methods above to the one obtained from Tuning-Insensitive Graph Estimation and Regression based on square root Lasso (`Tiger`), a robust regression procedure proposed in Liu and Wang [41] for graphical model estimation.

Take inverse directly?

An implementation of the GLasso that allows the diagonal entries of the precision matrix estimator to be unpenalized is provided in the widely used `glasso` package. In this paper, however, we use the GLasso implementation from the `QUIC` package [34], since it does not require the input covariance matrix to be positive semidefinite, and speeds up substantially over `glasso`. We select the tuning parameter λ in GLasso by cross-validation: We first split the data into K groups, or folds, of nearly equal size. For a given λ and $1 \le k \le K$, we take the k^{th} fold as the test set, and compute the precision matrix estimate $\hat{\mathbf{\Omega}}_\lambda^{(-k)}$ based on the remaining $K-1$ folds. We then compute the negative log-likelihood on the test set, $L^{(k)}(\lambda) = -\log\det\hat{\mathbf{\Omega}}_\lambda^{(-k)} + \operatorname{tr}\left(\hat{\mathbf{\Sigma}}^{(k)}\hat{\mathbf{\Omega}}_\lambda^{(-k)}\right)$, where $\hat{\mathbf{\Sigma}}^{(k)}$ is the robust covariance estimate obtained from the test set. This is done over a logarithmically spaced grid of 15 values between $\lambda_{\max} = \max_{i\neq j}|\hat{\mathbf{\Sigma}}_{ij}|$ and $\lambda_{\min} = 0.01\lambda_{\max}$, where $\hat{\mathbf{\Sigma}}$ is the robust covariance estimate computed from the whole data set. The value of λ that minimizes $\frac{1}{K}\sum_{k=1}^{K} L^{(k)}(\lambda)$ is selected as the final tuning parameter. On the other hand, the `flare` package contains an implementation of `Tiger`. We use the default parameters in the function `sugm()` and we perform model selection using `sugm.select()` with `criterion = cv`.

This is done by R language

N=15

Simulation settings We consider the following sampling schemes, covering different structures of the precision matrix $\mathbf{\Omega}^* \in \mathbb{R}^{p\times p}$:

- Banded: $\mathbf{\Omega}^*_{ij} = 0.6^{|i-j|}$.
- Sparse: $\mathbf{\Omega}^* = \mathbf{B} + \delta\mathbf{I}_p$, where $b_{ii} = 0$ and $b_{ij} = b_{ji}$, with $P(b_{ij} = 0.5) = 0.1$ and $P(b_{ij} = 0) = 0.9$, for $i \neq j$. The parameter δ is chosen such that the condition number of $\mathbf{\Omega}^*$ equals p. The matrix is then standardized to have unit diagonals.
- Dense: $\mathbf{\Omega}^*_{ii} = 1$ and $\mathbf{\Omega}^*_{ij} = 0.5$, for $i \neq j$.
- Diagonal: $\mathbf{\Omega}^* = \mathbf{I}_p$.

experiment 1

For each sampling scheme and dimension $p \in \{120, 400\}$, we generate $B = 100$ samples of size $n = 200$ from the multivariate normal distribution $N(\mathbf{0}, (\mathbf{\Omega}^*)^{-1})$. We then add 5% or 10% of rowwise or cellwise contamination to the data, where the outliers are sampled independently from $N(10, 0.2)$. We also simulate model deviation by generating all observations from either the multivariate t-distribution, $t_3(\mathbf{0}, (\mathbf{\Omega}^*)^{-1})$, or the alternative t-distribution, $t_3^*(\mathbf{0}, (\mathbf{\Omega}^*)^{-1})$, each with three degrees of freedom. Recall that $\mathbf{X} \sim t_\nu(\mathbf{0}, (\mathbf{\Omega}^*)^{-1})$, where $t_\nu(\mathbf{0}, \mathbf{\Omega}^*)^{-1})$ denotes the multivariate t-distribution with ν degrees of freedom, if $\mathbf{X} = \mathbf{Y}/\sqrt{\tau}$, where $\mathbf{Y} \sim N(\mathbf{0}, (\mathbf{\Omega}^*)^{-1})$ and $\tau \sim \Gamma(\nu/2, \nu/2)$. The alternative

experiment 2

t-distribution, denoted by t^*_ν, is proposed in Finegold and Drton [25] as a generalization of the multivariate t-distribution. We say that $\mathbf{X} \sim t^*_\nu(\mathbf{0}, (\mathbf{\Omega}^*)^{-1})$ if $X_i = Y_i/\sqrt{\tau_i}$, for all $1 \leq i \leq p$, where the divisors $\tau_i \sim \Gamma(\nu/2, \nu/2)$ are independent. In this case, the heaviness of the tails are different for different components of $\mathbf{X}$.

Performance measures We assess the performance of the covariance and precision matrix estimators via the deviations $\|\hat{\mathbf{\Sigma}} - \mathbf{\Sigma}^*\|_\infty$ and $\|\hat{\mathbf{\Omega}} - \mathbf{\Omega}^*\|_\infty$, respectively. We also consider the false positive (FP) and false negative (FN) rates:

$$\text{FP} = \frac{|\{(i,j) : \hat{\mathbf{\Omega}}_{ij} \neq 0, \mathbf{\Omega}^*_{ij} = 0\}|}{|\{(i,j) : \mathbf{\Omega}^*_{ij} = 0\}|}, \quad \text{and} \quad \text{FN} = \frac{|\{(i,j) : \hat{\mathbf{\Omega}}_{ij} = 0, \mathbf{\Omega}^*_{ij} \neq 0\}|}{|\{(i,j) : \mathbf{\Omega}^*_{ij} \neq 0\}|}.$$

FP gives the proportion of zero elements in the true precision matrix that are incorrectly estimated to be nonzero, while FN gives the proportion of nonzero elements in the true precision matrix that are incorrectly estimated to be zero. Note that if $\mathbf{\Omega}^*$ has no zero entries, as in the case of the banded and dense structures, the quantity FP is undefined.

Results Not surprisingly, `SampleCov` performs best for clean data. Under rowwise contamination, `OGK` yields the best results in terms of covariance estimation; under cellwise contamination, `Kendall`, `Spearman`, and `SpearmanU` equally share the best performance, while `NPD` and `Tiger` are slightly worse off. `Kendall`, `Spearman`, and `SpearmanU` also perform very well when the data are generated from a multivariate t-distribution or the alternative t-distribution, although these latter cases are not covered by our theory. More detailed comments and tables summarizing the results of the simulations may be found in Appendix A. Empirical results of a similar flavor were obtained in Liu et al. [42], although their paper does not provide theoretical guarantees for the behavior of the estimators under contaminated data.

7. Discussion

We have derived statistical error bounds for high-dimensional robust precision matrix estimators, when data are drawn from a multivariate normal distribution and then observed subject to cellwise contamination. We show that in such settings, the precision matrix estimators that are obtained by plugging in pairwise robust covariance estimators to the GLasso or CLIME routine, as suggested by Oellerer and Croux [46] and Tarr et al. [58], have error bounds that match standard high-dimensional bounds for uncontaminated precision matrix estimation, up to an additive factor involving a constant multiple of the contamination fraction ϵ. Our results for precision matrix estimators are derived via estimation error bounds for robust covariance matrix estimators, which have similar deviation properties.

The results of our paper naturally suggest several venues for future work. In particular, it would be interesting to relate the nonasymptotic statistical error bounds to the behavior of the sensitivity curve of the robust covariance estimator, which is the finite-sample analog of the influence function. We have also left open the question of calculating the breakdown point for the CLIME estimator with respect to more general data matrices, as well as the breakdown behavior of CLIME and GLasso under different notions of breakdown point. Although our results imply the superiority of the GLasso over the CLIME estimator from the perspective of the finite-sample breakdown point, this may only be part of the story.

Lastly, it would be interesting to generalize our study to other classes of distributions. In one direction, it would be possible to study contaminated versions of other distributions besides the multivariate Gaussian, for which the precision matrix encodes information about the underlying graphical model (e.g., Ising models on trees). A harder question to tackle would be the problem of robust graphical model estimation in settings where the structure of the graph is not encoded in the precision matrix alone. Finally, one could consider robust estimation of scatter matrices, when the uncontaminated data are drawn from an elliptical distribution. In that case, the proposed Kendall's tau and Spearman's rho correlation coefficients would still be Fisher consistent upon taking the respective sine transformations, so similar error bounds should hold. As demonstrated in our simulation results, the pairwise covariance estimators based on Kendall's tau and Spearman's rho perform reasonably well when data are generated from either the multivariate t-distribution or the alternative t-distribution. This motivates studying the convergence rates of the same covariance matrix estimators under heavy-tailed or elliptical distributions.

The problem of estimating high-dimensional covariance matrices under various structural assumptions has also been widely studied. Various families of structured covariance matrices have been introduced, including bandable matrices [14], Toeplitz matrices [13], and sparse matrices [8, 15]. The proposed covariance matrix estimators involve regularizing the sample covariance matrix in accordance to structural assumptions. It would be interesting to study robust versions of these structured covariance matrix estimators under a model such as cellwise contamination. Besides graphical models, covariance matrix estimation is also useful for statistical methods such as linear discriminant analysis and principal component analysis. Several high-dimensional procedures have been proposed with proven theoretical guarantees when data are uncontaminated [10, 63], and it would be interesting to study robust adaptations of these procedures, as well.

Acknowledgements

The authors thank the AE and anonymous referees for useful feedback that led to a more complete and coherent manuscript. The authors also thank Xingguo Li for helping debug the TIGER code.

Supplementary Material

Supplement to: High-dimensional robust precision matrix estimation: Cellwise corruption under ϵ-contamination
(doi: 10.1214/18-EJS1427SUPP; .pdf). We provide proofs for the technical lemmas employed in our paper.

References

[1] Agostinelli, C., Leung, A., Yohai, V. J. and Zamar, R. H. (2015). Robust estimation of multivariate location and scatter in the presence of cellwise and casewise contamination. *Test* **24** 441–461. MR3377883

[2] Alqallaf, F. A., Konis, K. P., Martin, R. D. and Zamar, R. H. (2002). Scalable robust covariance and correlation estimates for data mining. In *Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* 14–23. ACM.

[3] Alqallaf, F., van Aelst, S., Yohai, V. J. and Zamar, R. H. (2009). Propagation of outliers in multivariate data. *Ann. Statist.* **37** 311–331. MR2488353 (2010c:62091)

[4] Anderson, T. W. (2003). *An Introduction to Multivariate Statistical Analysis. Wiley Series in Probability and Statistics*. Wiley. MR1990662

[5] Avella-Medina, M., Battey, H. S., Fan, J. and Li, Q. (2016). Robust estimation of high-dimensional covariance and precision matrices. *Biometrika* **103** 1–14.

[6] Banerjee, O., El Ghaoui, L. and d'Aspremont, A. (2008). Model selection through sparse maximum likelihood estimation for multivariate Gaussian or binary data. *J. Mach. Learn. Res.* **9** 485–516. MR2417243

[7] Bickel, P. J. (1964). On some alternative estimates for shift in the p-variate one sample problem. *Ann. Math. Statist.* **35** 1079–1090. MR0165624 (29 ##2904)

[8] Bickel, P. J. and Levina, E. (2008). Covariance regularization by thresholding. *Ann. Statist.* **36** 2577–2604. MR2485008 (2010b:62197)

[9] Boudt, K., Cornelissen, J. and Croux, C. (2012). The Gaussian rank correlation estimator: robustness properties. *Statistics and Computing* **22** 471–483. MR2865030

[10] Cai, T. and Liu, W. (2011). A direct estimation approach to sparse linear discriminant analysis. *J. Amer. Statist. Assoc.* **106** 1566–1577. MR2896857

[11] Cai, T., Liu, W. and Luo, X. (2011). A constrained ℓ_1 minimization approach to sparse precision matrix estimation. *J. Amer. Statist. Assoc.* **106** 594–607. MR2847973 (2012j:62208)

[12] Cai, T. T., Liu, W. and Zhou, H. H. (2016). Estimating sparse precision matrix: Optimal rates of convergence and adaptive estimation. *Ann. Statist.* **44** 455–488. MR3476606

[13] Cai, T. T., Ren, Z. and Zhou, H. H. (2013). Optimal rates of convergence for estimating Toeplitz covariance matrices. *Probab. Theory Related Fields* **156** 101–143. MR3055254

[14] Cai, T. T., Zhang, C.-H. and Zhou, H. H. (2010). Optimal rates of convergence for covariance matrix estimation. *Ann. Statist.* **38** 2118–2144. MR2676885 (2011h:62215)

[15] Cai, T. T. and Zhou, H. H. (2012). Optimal rates of convergence for sparse covariance matrix estimation. *Ann. Statist.* **40** 2389–2420. MR3097607

[16] Capéraà, P. and Guillem, A. I. G. (1997). Taux de résistance des tests de rang d'indépendance. *Canadian Journal of Statistics* **25** 113–124. MR1451675

[17] Chen, M., Gao, C. and Ren, Z. (2015). Robust covariance and scatter matrix estimation under Huber's contamination model. *ArXiv e-prints*.

[18] Chen, M., Gao, C. and Ren, Z. (2016). A general decision theory for Huber's ϵ-contamination model. *Electronic Journal of Statistics* **10** 3752–3774. MR3579675

[19] Croux, C. and Dehon, C. (2010). Influence functions of the Spearman and Kendall correlation measures. *Statistical Methods and Applications* **19** 497–515. MR2728407

[20] Davies, P. L. and Gather, U. (2005). Breakdown and groups. *The Annals of Statistics* **33** 977–1035. MR2195626

[21] Donoho, D. L. (1982). Breakdown properties of multivariate location estimators Technical Report, Harvard University, Boston. URL http://www-stat.stanford.edu/~donoho/Reports/Oldies/BPMLE.pdf.

[22] Donoho, D. and Huber, P. J. (1983). The notion of breakdown point. In *A Festschrift for Erich L. Lehmann. Wadsworth Statist./Probab. Ser.* 157–184. Wadsworth, Belmont, CA. MR689745 (87h:62043)

[23] Fan, J., Liu, H. and Wang, W. (2015). Large covariance estimation through elliptical factor models. *arXiv preprint arXiv:1507.08377*.

[24] Fan, J., Liu, H., Ning, Y. and Zou, H. (2017). High dimensional semiparametric latent graphical model for mixed data. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **79** 405–421. MR3611752

[25] Finegold, M. and Drton, M. (2011). Robust graphical modeling of gene networks using classical and alternative t-distributions. *Ann. Appl. Stat.* **5** 1057–1080. MR2840186 (2012i:62151)

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

[27] Hampel, F. R. (1974). The influence curve and its role in robust estimation. *J. Amer. Statist. Assoc.* **69** 383–393. MR0362657 (50 ##15097)

[28] Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J. and Stahel, W. A. (2011). *Robust Statistics: The Approach Based on Influence Functions. Wiley Series in Probability and Statistics*. Wiley. MR0829458

[29] Han, F. and Liu, H. (2014). Scale-invariant sparse PCA on high-dimensional meta-elliptical data. *J. Amer. Statist. Assoc.* **109** 275–287. MR3180563

[30] Han, F. and Liu, H. (2017). Statistical analysis of latent generalized correlation matrix estimation in transelliptical distribution. *Bernoulli* **23** 23–57. MR3556765

[31] Han, F., Lu, J. and Liu, H. (2015). Robust scatter matrix estimation for high dimensional distributions with heavy tails Technical Report, Princeton University.
[32] Higham, N. J. (2002). Computing the nearest correlation matrix—a problem from finance. *IMA Journal of Numerical Analysis* **22** 329–343. MR1918653
[33] Hoeffding, W. (1948). A class of statistics with asymptotically normal distribution. *Ann. Math. Statistics* **19** 293–325. MR0026294 (10,134g)
[34] Hsieh, C. J., Dhillon, I. S., Ravikumar, P. K. and Sustik, M. A. (2011). Sparse inverse covariance matrix estimation using quadratic approximation. In *Advances in Neural Information Processing Systems 24* (J. Shawe-Taylor, R. S. Zemel, P. L. Bartlett, F. Pereira and K. Q. Weinberger, eds.) 2330–2338. Curran Associates, Inc.
[35] Huber, P. J. (1964). Robust estimation of a location parameter. *Ann. Math. Statist.* **35** 73–101. MR0161415 (28 ##4622)
[36] Huber, P. J. (1981). *Robust Statistics*. John Wiley & Sons, Inc., New York Wiley Series in Probability and Mathematical Statistics. MR606374 (82i:62057)
[37] Kendall, M. G. (1948). *Rank Correlation Methods*. Griffin. MR1527612
[38] Kruskal, W. H. (1958). Ordinal measures of association. *Journal of the American Statistical Association* **53** 814–861. MR0100941
[39] Lauritzen, S. L. (1996). *Graphical Models*. Oxford University Press. MR1419991
[40] Little, R. J. A. and Rubin, D. B. (1986). *Statistical Analysis with Missing Data*. John Wiley & Sons, Inc., New York, NY, USA. MR0890519
[41] Liu, H. and Wang, L. (2017). TIGER: A tuning-insensitive approach for optimally estimating Gaussian graphical models. *Electronic Journal of Statistics* **11** 241–294. MR3606771
[42] Liu, H., Han, F., Yuan, M., Lafferty, J. and Wasserman, L. (2012). High-dimensional semiparametric Gaussian copula graphical models. *Ann. Statist.* **40** 2293–2326. MR3059084
[43] Loh, P.-L. and Tan, X. L. Supplement to: "High-dimensional robust precision matrix estimation: Cellwise corruption under ϵ-contamination." DOI: 10.1214/18-EJS1427SUPP
[44] Maronna, R. A. (1976). Robust M-estimators of multivariate location and scatter. *Ann. Statist.* **4** 51–67. MR0388656 (52 ##9492)
[45] Maronna, R. A. and Zamar, R. H. (2002). Robust estimates of location and dispersion for high-dimensional datasets. *Technometrics* **44** 307–317. MR1939680
[46] Öllerer, V. and Croux, C. (2015). Robust high-dimensional precision matrix estimation. In *Modern Nonparametric, Robust and Multivariate Methods* 325–350. Springer. MR3444335
[47] Puri, M. L. and Sen, P. K. (1971). *Nonparametric Methods in Multivariate Analysis*. John Wiley & Sons, Inc., New York-London-Sydney. MR0298844 (45 ##7893)

[48] Ravikumar, P., Wainwright, M. J., Raskutti, G. and Yu, B. (2011). High-dimensional covariance estimation by minimizing ℓ_1-penalized log-determinant divergence. *Electron. J. Statist.* **5** 935–980. MR2836766

[49] Ren, Z., Sun, T., Zhang, C. H. and Zhou, H. H. (2015). Asymptotic normality and optimalities in estimation of large Gaussian graphical models. *Ann. Statist.* **43** 991–1026. MR3346695

[50] Rousseeuw, P. J. (1984). Least median of squares regression. *J. Amer. Statist. Assoc.* **79** 871–880. MR770281 (86d:62113)

[51] Rousseeuw, P. (1985). Multivariate estimation with high breakdown point. In *Mathematical statistics and applications, Vol. B (Bad Tatzmannsdorf, 1983)* 283–297. Reidel, Dordrecht. MR851060 (88a:62142)

[52] Rousseeuw, P. J. and Croux, C. (1993). Alternatives to the median absolute deviation. *J. Amer. Statist. Assoc.* **88** 1273–1283. MR1245360 (94g:62064)

[53] Serfling, R. and Mazumder, S. (2009). Exponential probability inequality and convergence results for the median absolute deviation and its modifications. *Statist. Probab. Lett.* **79** 1767–1773. MR2566751 (2010j:62103)

[54] Shevlyakov, G. L. and Vilchevski, N. O. (2002). *Robustness in Data Analysis: Criteria and Methods. Modern Probability and Statistics, 6*. VSP.

[55] Smith, S. M., Miller, K. L., Salimi-Khorshidi, G., Webster, M., Beckmann, C. F., Nichols, T. E., Ramsey, J. D. and Woolrich, M. W. (2011). Network modelling methods for FMRI. *NeuroImage* **54** 875–891.

[56] Stahel, W. A. (1981). *Breakdown of Covariance Estimators*. Fachgruppe für Statistik, Eidgenössische Techn. Hochsch.

[57] Swanson, D. C. (2000). *Signal Processing for Intelligent Sensor Systems. Signal Processing and Communications*. CRC Press.

[58] Tarr, G., Müller, S. and Weber, N. C. (2015). Robust estimation of precision matrices under cellwise contamination. *Computational Statistics and Data Analysis*. MR3406222

[59] Troyanskaya, O., Cantor, M., Sherlock, G., Brown, P., Hastie, T., Tibshirani, R., Botstein, D. and Altman, R. B. (2001). Missing value estimation methods for DNA microarrays. *Bioinformatics* **17** 520–525.

[60] Tsybakov, A. B. (2008). *Introduction to Nonparametric Estimation*. Springer Publishing Company, Incorporated. MR2656066

[61] Tukey, J. W. (1962). The future of data analysis. *Ann. Math. Statist.* **33** 1–67. MR0133937 (24 ##A3761)

[62] Van Aelst, S. (2016). Stahel-Donoho estimation for high-dimensional data. *International Journal of Computer Mathematics* **93** 628–639. MR3473777

[63] Vu, V. Q., Cho, J., Lei, J. and Rohe, K. (2013). Fantope projection and selection: A near-optimal convex relaxation of sparse PCA. In *Advances in Neural Information Processing Systems* 2670–2678.

[64] Wegkamp, M. and Zhao, Y. (2016). Adaptive estimation of the copula correlation matrix for semiparametric elliptical copulas. *Bernoulli* **22** 1184–1226. MR3449812

[65] Werhli, A. V., Grzegorczyk, M. and Husmeier, D. (2006). Comparative evaluation of reverse engineering gene regulatory networks with relevance networks, graphical gaussian models and bayesian networks. *Bioinformatics* **22** 2523–2531.

[66] Xue, L. and Zou, H. (2012). Regularized rank-based estimation of high-dimensional nonparanormal graphical models. *Ann. Statist.* **40** 2541–2571. MR3097612

[67] Yu, B. (1997). *Assouad, Fano, and Le Cam* In *Festschrift for Lucien Le Cam: Research Papers in Probability and Statistics* 423–435. Springer New York, New York, NY. MR1462963

[68] Yuan, M. and Lin, Y. (2007). Model selection and estimation in the Gaussian graphical model. *Biometrika* **94** 19–35. MR2367824

[69] Zhang, F. (2011). *Matrix Theory: Basic Results and Techniques. Universitext*. Springer. MR2857760

[70] Zhao, T., Liu, H., Roeder, K., Lafferty, J. and Wasserman, L. (2012). The `huge` package for high-dimensional undirected graph estimation in R. *J. Mach. Learn. Res.* **13** 1059–1062. MR2930633