Discussion of "Estimating structured high-dimensional covariance and precision matrices: Optimal rates and adaptive estimation"*

Krishnakumar Balasubramanian and Ming Yuan[†]

Department of Statistics
University of Wisconsin at Madison
e-mail: krizna@stat.wisc.edu; myuan@stat.wisc.edu

1. Introduction

We would like to first congratulate the authors for this timely survey on high-dimensional covariance matrix estimation. Over the last decade, there has been a flurry of activities in high-dimensional statistics in general and more specifically on covariance estimation. The article provides an organized overview of the key advancements in the area of sparse covariance matrix estimation and sparse inverse covariance matrix estimation. Specifically it enlists the parameter spaces corresponding to the different types of sparse structures commonly assumed in the literature and provides minimax lower bounds and estimators that match these lower bounds for each of the parameter spaces. Furthermore, it includes discussion on two closely related problems, namely sparse principal component analysis and high-dimensional covariance testing. This expository article, without a doubt would serve as a "goto" reference for people who are interested to learn about high dimensional covariance matrix estimation in the future.

In this discussion, we concentrate on a particular assumption made in most of the literature on covariance estimation, namely the sub-Gaussian assumption. It is long recognized by statisticians that data from real-world experiments oftentimes tend to be corrupted with outliers and/or exhibit heavy tails. In such cases, it is not clear that those covariance matrix estimators described in this article remain optimal. In particular, many of these estimators are based upon sample covariance matrices that is known to perform rather poorly if the data is corrupted with outliers, under the classical fixed dimensionality paradigm [1].

To explore this possible pitfall and offer potential solutions, we investigate in this discussion a simple strategy of replacing the sample covariance matrix with a more robust *median covariance matrix*.

2. Median covariance matrix estimator

Let $\{X^{(1)}, \ldots, X^{(n)}\}$ be n random variables taking values in a Banach space $(\mathcal{B}, \|\cdot\|)$, with metric $\|\cdot\|$. Recall that the geometric median (cf. [2]), $\hat{m}(X^{(1)}, X^{(2)}, \ldots, X^{(n)})$, of the n random variables is defined as the point in the space that minimizes the sum of its distance to all observations:

$$\hat{m}(X^{(1)}, X^{(2)}, \ldots, X^{(n)}) = \arg\min_{c \in \mathcal{B}} \sum_{i=1}^{n} \|X^{(i)} - c\|.$$

We note that the minimizer on the right hand side is uniquely defined and therefore the above definition is valid if the space $\mathcal{B}$ is separable, reflexive and strictly convex; see [3] for further discussions about the existence and other properties of $\hat{m}$.

Now consider employing a similar strategy for covariance matrix estimation. More specifically, define

$$\hat{\Sigma}^{\mathrm{med}} = \hat{m}(X^{(1)}X^{(1)^\top}, X^{(2)}X^{(2)^\top}, \ldots, X^{(n)}X^{(n)^\top}),$$

where, for simplicity, we use the standard Frobenious norm $\|A\|_{\mathrm{F}} = \sqrt{\mathrm{tr}(A^\top A)}$ as the metric for calculating $\hat{m}$. Instead of Frobenius norm, one may also consider other metrics, for example those adapted to the manifold of symmetric positive semidefinite matrices, when computing $\hat{m}$.

Once the median covariance matrix is computed, we may impose sparse structures on it to derive estimates suitable for high dimensional problems. To fix ideas, here we focus on the bandable case and consider a banding estimator based on $\hat{\Sigma}^{\mathrm{med}}$:

$$\hat{\Sigma}_{B,k} = (\hat{\Sigma}_{ij}^{\mathrm{med}} I\{|i - j| \le k\}).$$

We now examine the effect of using Σ^{med} as the initial estimator through a numerical experiment.

3. Numerical example

To compare the median covariance matrix based banding estimator against the sample covariance matrix based estimator, we now conduct a set of simulation studies. We generated $n = 50$ random vectors in $\mathbb{R}^{120}$. Of the 50 observations, a random fraction (ρ) are corrupted, and the corrupted observations are sampled from $[-10, 10]^{120}$. The remaining observations are samples from a Gaussian distribution with a certain AR(1) covariance structure similar to [4], that

Comparison between sample covariance matrix and median covariance matrix based banding estimators. Reported here are the estimation error measured by the spectral norm averaged over 50 runs. The numbers in bracket correspond to standard errors

	Sample Cov Based	Median Cov Based
$\rho = 0.10$	3.98 (0.02)	1.76 (0.04)
$\rho = 0.14$	4.76 (0.04)	2.21 (0.05)
$\rho = 0.18$	5.56 (0.04)	2.65 (0.03)
$\rho = 0.20$	6.64 (0.04)	2.99 (0.02)

is $\Sigma_{ij}^* = \eta^{|i-j|}$, with $\eta = 0.7$. To gain insights into the effect of the level of corruption, we consider various values of ρ, namely $\rho \in \{0.10, 0.14, 0.18, 0.20\}$. We compare both the sample covariance matrix based and median covariance matrix based banding estimator in terms of spectral norm. To fix ideas, we fix $k = 5$ for both estimators. The result based on 50 runs is summarized below. The benefit of using the more robust median covariance matrix as the initial estimator is evident from this example.

To conclude, robustness to corrupted and/or heavy-tailed data appears to be an important practical issue when it comes to covariance matrix estimation. Our limited experiment here suggests that it could potentially be gained by simply replacing the sample covariance matrix with a robust covariance matrix estimator. To what extent this is true and what are the other possible strategies to deal with heavy tailed distributions warrant further studies.

References

[1] HUBER, P. (2009). *Robust statistics*, Wiley and Sons. MR0606374

[2] KEMPERMAN, J. (1987). *The median of a finite measure on a Banach space*, Statistical Data Analysis Based on the L1-norm and Related Methods, NorthHolland, Amesterdam. MR0949228

[3] SMALL, C.G. (1990). *A survey of multidimensional medians*, International Statistical Review/Revue Internationale de Statistique.

[4] BICKEL, P. AND LEVINA, E. (2008). *Covariance regularization by thresholding*, The Annals of Statistics. MR2485008