Random matrices and covariance estimation

Kaveh S. Nobari

Lectures in High-Dimensional Statistics

Department of Mathematics and Statistics
Lancaster University

Contents

Motivation

The issue of covariance estimation is intertwined with random matrix theory, since sample covariance is a particular type of random matrix. These slides follow the structure of chapter 6 of Wainwright (2019) to shed light on random matrices in a non-asymptotic setting, with the aim of obtaining explicit deviation inequalities that hold for all sample sizes and matrix dimensions.

In the classical framework of covariance matrix estimation the sample size n tends to infinity while the matrix dimension d is fixed; in this setting the behaviour of sample covariance matrix is characterized by the usual limit theory. In contrast, in high-dimensional settings the data dimension is either comparable to the sample size ($d \asymp n$) or possibly much larger than the sample size $d \gg n$.

Motivation

We begin with the simplest case, namely ensembles of Gaussian random matrices, and we then discuss more general sub-Gaussian ensembles, before moving to milder tail conditions.

First, let us consider rectangular matrices, for instance matrix $A \in \mathbb{R}^{n \times m}$ with $n \geq m$, the ordered singular values are written as follows

$$\sigma_{\max}(A) = \sigma_1(A) \geq \sigma_2(A) \geq \cdots \geq \sigma_m(A) = \sigma_{\min}(A) \geq 0$$

The maximum and minimum singular values are obtained by maximizing the "blow-up factor"

$$\sigma_{\max}(A) = \max_{\forall x} \frac{\|Ax\|_2}{\|x\|_2}, \quad \sigma_{\min}(A) = \min_{\forall x} \frac{\|Ax\|_2}{\|x\|_2}$$

which is obtained when x is the largest and smallest singular vectors respectively - i.e.

$$\sigma_{\max}(A) = \max_{v \in S^{m-1}} \frac{\|Av\|_2}{\|v\|_2}, \quad \sigma_{\min}(A) = \min_{v \in S^{m-1}} \frac{\|Av\|_2}{\|v\|_2}$$

noting that $\|v\|_2 = 1$, since $S^{d-1} := \{v \in \mathbb{R}^d \mid \|v\|_2 = 1\}$ is the Euclidean unit sphere in $\mathbb{R}^d$. We may denote

$$|||A|||_2 = \sigma_{\max}(A)$$

However, covariance matrices are square symmetric matrices, thus we must also focus on symmetric matrices in $\mathbb{R}^d$, denoted $S^{d\times d} := \{Q \in \mathbb{R}^{d\times d} \mid Q = Q'\}$, as well as subset of semi-definite matrices given by

$$S_+^{d\times d} := \{Q \in S^{d\times d} \mid Q \geq 0\}.$$

Any matrix $Q \in S^{d\times d}$ is diagonalizable via unitary transformation, and let us denote the vector of eigenvalues of Q by $\gamma(Q) \in \mathbb{R}^d$ ordered as

$$\gamma_{\max}(Q) = \gamma_1(Q) \geq \gamma_2(Q) \geq \cdots \geq \gamma_d(Q) = \gamma_{\min}(Q)$$

Note the matrix Q is semi-positive definite, which may be expressed as $Q \geq 0$, iff $\gamma_{\min}(Q) \geq 0$.

The Rayleigh-Ritz variational characterization of the minimum and maximum eigenvalues

$$\gamma_{\max}(Q) = \max_{v \in S^{d-1}} v'Qv \quad \text{and} \quad \gamma_{\min}(Q) = \min_{v \in S^{d-1}} v'Qv$$

For symmetric matrix Q, the l_2 norm can be expressed as

$$|||Q|||_2 = \max\{\gamma_{\max}(Q), |\gamma_{\min}(Q)|\} := \max_{v \in S^{d-1}} |v'Qv|$$

Finally, suppose we have a rectangular matrix $A \in \mathbb{R}^{n \times m}$, with $n \geq m$. We know that any rectangular matrix can be expressed using singular value decomposition (SVD hereafter), as follows

$$A = U\Sigma V'$$

wher U is an $n \times n$ unitary matrix, Σ is an $n \times m$ rectangular diagonal matrix with non-negative real numbers on the diagonal up and V is an $n \times n$ unitary matrix. Using SVD, we can express $A'A$ where

$$A'A = V\Sigma' U' U \Sigma V'$$

and since U is an orthogonal matrix, we know that $U'U = I$ where I is the identity matrix.

$$A'A = V(\Sigma'\Sigma)V'$$

Therefore, as the diagonal matrix Σ contains the eigenvalues of matrix A, hence, $\Sigma'\Sigma$ contains the eigenvalues of $A'A$ and it can be thus concluded

$$\gamma_j(A'A) = (\sigma_j(A))^2, \quad j = 1, \cdots, m$$

Let $\{x_1, \cdots, x_n\}$ be a collection of n i.i.d samples from a distribution in $\mathbb{R}^d$ with zero mean and the covariance matrix Σ. A standard estimator of sample covariance matrix is

$$\hat{\Sigma} := \frac{1}{n} \sum_{i=1}^{n} x_i x_i'.$$

Since, each x_i for $i = 1, \cdots, n$ has zero mean, it is guaranteed that

$$\mathbb{E}[x_i x_i'] = \Sigma$$

and the random matrix $\hat{\Sigma}$ is an unbiased estimator of the population covariance Σ. Consequently the error matrix $\hat{\Sigma} - \Sigma$ has mean zero, and goal is to obtain bounds on the error measures in l_2-norm. We are essentially seeking a band of the form

$$\left|\left|\left|\hat{\Sigma} - \Sigma\right|\right|\right| \leq \varepsilon,$$

where as before,

References

Wainwright, M. J. (2019). *High-dimensional statistics: A non-asymptotic viewpoint*, volume 48. Cambridge University Press.