Covariance matrices from sub-Gaussian ensembles

Hao Yan

May 25, 2022

Consider random matrix $X \in \mathbb{R}^{n\times d}$ with i.i.d. rows from a σ-sub-Gaussian, we would like to bound the operator norm of the difference between the sample covariance and the covariance matrix.

To control $\|\widehat{\Sigma}-\Sigma\|_2$ for Gaussian ensembles, the steps are very standard. We first make use of concentration of measures for **Lipschitz functions of Gaussian** random vectors to argue that the eigenvalues of the sample covariance are concentrated around the expectations. The result then follows by applying **Gaussian comparison inequalities** to bound the expectations.

To obtain the bound for sub-Gaussian ensembles, we have to adopt the ϵ-net approach. The following result is what we are going to show.

Theorem 0.1.

$$\mathbb{E}[e^{\lambda\|\widehat{\Sigma}-\Sigma\|_2}] \le e^{c_0\frac{\lambda^2\sigma^4}{n}+4d}, \quad \textit{for all } |\lambda| < \frac{n}{64e^2\sigma^2}.$$

Proof. Let $Q := \widehat{\Sigma}-\Sigma$. We use the variational representation $\|Q\|_2 = \max_{v\in\mathbb{S}^{d-1}} v^\top Qv$. Then all we need to do is to establish some kind of uniform law of large number result.

Step 1. We first **discretize** $\mathbb{S}^{d-1}$. For an ϵ-net $\mathcal{N}_\epsilon$ of the unit sphere, we have

$$|\mathcal{N}_\epsilon| \le \left(1+\frac{2}{\epsilon}\right)^d.$$

For every unit vector v, we can find $v_j \in \mathcal{N}_\epsilon$ such that $\|v_j - v\|_2 \le \epsilon$. Denote $\Delta = v - v_j$. Notice that

$$\begin{aligned} v^\top Qv &= (v_j+\Delta)^\top Q(v_j+\Delta) \\ &\le v_j^\top Qv_j + (2\epsilon+\epsilon^2)\|Q\|_2, \end{aligned}$$

take the maximum at both side, we have

$$(1-2\epsilon-\epsilon^2)\|Q\|_2 \le \max_{v\in\mathcal{N}_\epsilon} |v^\top Qv|.$$

For simplicity, take $\epsilon = 1/5$, we have

$$\|Q\|_2 \le 2\max_{v\in\mathcal{N}_\epsilon} |v^\top Qv|,$$

and $|\mathcal{N}_\epsilon| \le 11^d$.

Step 2. We apply the **symmetrization trick**. First, we have

$$\mathbb{E}[e^{\lambda\|Q\|_2}] \le |\mathcal{N}_\epsilon| \cdot \left(\mathbb{E}[e^{2\lambda v^\top Qv}] + \mathbb{E}[e^{-2\lambda v^\top Qv}]\right).$$

It follows that

$$\begin{aligned} \mathbb{E}[e^{tv^\top Qv}] &= \mathbb{E}[\exp(tv^\top(\widehat{\Sigma}-\Sigma)v)] \\ &= \prod_{i=1}^n \mathbb{E}[e^{\frac{t}{n}\{(v^\top x_i)^2 - v^\top\Sigma v\}}] \\ &= \mathbb{E}[e^{\frac{t}{n}\{(v^\top x_1)^2 - v^\top\Sigma v\}}]^n. \end{aligned}$$

Since $\mathbb{E}[(v^\top x_1)^2] = v^\top \Sigma v$ and $\Phi(t) = \exp(t)$ is a convex nondecreasing function, a standard symmetrization argument implies that

$$\mathbb{E}_{x_1}[e^{\frac{t}{n}\{(v^\top x_1)^2 - v^\top \Sigma v\}}] \leq \mathbb{E}_{x_1,\varepsilon}[e^{\frac{2t}{n}\varepsilon(v^\top x_1)^2}],$$

where ε is the Rademacher random variable.

Step 3. Finally, we do a Taylor expansion. It follows that

$$\begin{aligned}
\mathbb{E}_{x_1,\varepsilon}[e^{\frac{2t}{n}\varepsilon(v^\top x_1)^2}] &= \sum_{k=1}^{\infty} \frac{1}{k!}\left(\frac{2t}{n}\right)^k \mathbb{E}_{x_1,\varepsilon}[\varepsilon^k (v^\top x_1)^{2k}] \\
&= 1 + \sum_{l=1}^{\infty} \frac{1}{(2l)!}\left(\frac{2t}{n}\right)^{2l} \mathbb{E}[(v^\top x_1)^{4l}] \\
&\leq 1 + \sum_{l=1}^{\infty} \frac{1}{(2l)!}\left(\frac{2t}{n}\right)^{2l} \frac{(4l)!}{2^{2l}(2l)!}(\sqrt{8}e\sigma)^{4l} \\
&\leq 1 + \sum_{l=1}^{\infty} \left(\frac{16t}{n}e^2\sigma^2\right)^{2l} \\
&= \frac{1}{1 - \left(\frac{16t}{n}e^2\sigma^2\right)^2}.
\end{aligned}$$

As long as $f(t) := \frac{16t}{n}e^2\sigma^2$ satisfies $f(t) \leq \frac{1}{2}$, we have

$$\frac{1}{1 - f(t)^2} \leq \exp\left(2f(t)^2\right).$$

Putting the pieces together, we have

$$\mathbb{E}[e^{tv^\top Qv}] \leq \exp(2nf(t)^2)$$

valid for all $|t| < \frac{n}{32e^2\sigma^2}$.

Thus, for $|\lambda| < \frac{n}{64e^2\sigma^2}$, we have

$$\mathbb{E}[e^{\lambda\|Q\|_2}] \leq 11^d \cdot 2\exp(2nf(t)^2) = 11^d \cdot 2e^{2048\frac{\lambda^2}{n}e^4\sigma^4} \leq e^{c_0\frac{\lambda^2\sigma^4}{n} + 4d}.$$

$\square$