MATH888: High-dimensional probability and statistics	Fall 2021

Lecture 23 — October 29, 2021

Sebastien Roch, UW-Madison — *Scribe: Yuchen Li*

1 Overview

In the last lecture, we discussed the proof of Theorem 2 from lecture 21. In this lecture, we finish this proof. This lecture is based on Section 8.2 from Wainwright [1].

2 Review of PCA in Spiked Covariance Model

Let $\mathbf{W} \in \mathbb{R}^d$ be an isotropic, subgaussian random vector with mean zero, and let ϵ be an independent real-valued subgaussian random variable with mean zero and variance 1. The **spiked covariance model** is given by the random vector $\mathbf{X}$ with distribution

$$\mathbf{X} \sim \mathbf{W} + \sqrt{\nu}\epsilon\theta^*$$

where $\nu > 0$, $\theta^* \in \mathbb{S}^{d-1}$ are fixed. Since $\mathbf{X}$ is mean-zero and isotropic, the covariance is

$$\mathbf{\Sigma} = I_d + \nu\theta^*\theta^{*\top}$$

where I_d is the $d \times d$ identity matrix. The maximal eigenvalue is $1 + \nu$ with eigenvector θ^*. Now we restate the main theorem we proved partially in last lecture,

Theorem 1 (Cor 8.7 in [1])**.** *Assume $n > d$. Given n iid samples from the spiked covariance model with $(*)$, and assuming that $\sqrt{\frac{\nu+1}{\nu^2}}\sqrt{\frac{d}{n}} \leq C_0$, it holds that if $\hat{\theta}$ is the maximal eigenvector of $\hat{\mathbf{\Sigma}}_n$, then with probability $1 - C_2\exp\{-C_3 d\}$, we have*

$$\left\|\hat{\theta} - \theta^*\right\|_2 \leq C_1\sqrt{\frac{\nu+1}{\nu^2}}\sqrt{\frac{d}{n}}$$

To prove the theorem, we also need the following lemma,

Lemma 2 (Theorem 8.5 in [1])**.** *Consider $\mathbf{P} = \hat{\mathbf{\Sigma}} - \mathbf{\Sigma}$ and $\tilde{\mathbf{P}} = U_2^\top \mathbf{P}\theta^*$ where the columns of $U_2^\top$ forms an orthonormal basis of span$\{\theta^*\}^\perp$. If $\|\mathbf{P}\|_2 < \frac{\nu}{2}$ then*

$$\left\|\hat{\theta} - \theta^*\right\|_2 \leq \frac{2\left\|\tilde{\mathbf{P}}\right\|_2}{\nu - 2\|\mathbf{P}\|_2}$$

3 Proof of Theorem 1

We continue the proof of Theorem 1 by using Lemma 2 in this section. Last time, we decomposed P as

$$\mathbf{P} = \mathbf{P}_1 + \mathbf{P}_2 + \mathbf{P}_3$$

we claimed with probability $1 - C_4 exp\{-C_5 d\}$, we have

$$\|\mathbf{P}_1\|_2 \le \frac{\nu}{8}, \quad \|\mathbf{P}_2\|_2 \le C_6\sqrt{\nu}\sqrt{\frac{d}{n}}, \quad \|\mathbf{P}_3\|_2 \le C_7\sqrt{\frac{d}{n}}$$

thus we have

$$\|\mathbf{P}\|_2 \le \frac{\nu}{4}$$

Now we begin with the following claim,

Claim 3. *Let* $\bar{W} = \frac{1}{n}\sum_{i=1}^n \epsilon_i W^{(i)}$*, with probability* $1 - C_8 \exp\{-C_9 d\}$ *we have*

$$\left\|\bar{W}\right\|_2 \le C_{10}\sqrt{\frac{d}{\nu}}$$

We first review the concept of ϵ net before we prove the Claim 3.

3.1 Review of ε–net

Recall $N \subseteq T$ is ε–net of $K \subseteq T$ if for $\forall x \in K$, $\exists x_0 \in N$ so that $\|x - x_0\| \le \varepsilon$. Also, $N(K, \varepsilon)$ denotes the smallest size of an ε–net of the set K.

Recall the Lemma 4 from lecture 19, we showed the covering number of the unit Euclidean ball B_2^d satisfy the following for any $\varepsilon > 0$:

$$\mathcal{N}\left(B_2^d, \varepsilon\right) \le \left(\frac{2}{\varepsilon} + 1\right)^d. \tag{1}$$

Take

3.2 Proof of Claim 3

Now we prove Claim 3 utilizing ε–net,

Proof. Note

$$\left\|\bar{W}\right\|_2 = \sup_{u \in B_2^d} \langle u, \bar{W}\rangle$$

Let N be the $\frac{1}{2}$–net of B_2^d, $\forall u \in B_2^d$, $\exists z \in N$ such that

$$x = u - z \quad with \quad \|x\|_2 \le \frac{1}{2}$$

also by taking $\varepsilon = \frac{1}{2}$ in inequality 1, we have

$$\mathcal{N}\left(B_2^d, \frac{1}{2}\right) \le 5^d$$

then

$$
\begin{aligned}
\left\|\bar{W}\right\|_2 &= \sup_{u \in B_2^d} \langle u, \bar{W} \rangle \\
&\leq \sup_{z \in N} \langle z, \bar{W} \rangle + \sup_{x \in \frac{1}{2} B_2^d} \langle x, \bar{W} \rangle \\
&= \sup_{z \in N} \langle z, \bar{W} \rangle + \frac{1}{2} \left\|\bar{W}\right\|_2
\end{aligned}
$$

so we have

$$
\left\|\bar{W}\right\|_2 \leq 2 \sup_{z \in N} \langle z, \bar{W} \rangle
$$

Now we let $t = C_{12}\sqrt{\frac{d}{\nu}}$, then the probability

$$
\begin{aligned}
\mathbb{P}\left(2 \sup_{z \in N} \langle z, \bar{W} \rangle \geq t\right) &\leq \sum_{z \in N} \mathbb{P}(2\langle z, \bar{W} \rangle \geq t) \\
&= \sum_{z \in N} \mathbb{P}\left(\frac{2}{n} \sum_{i=1}^{n} \epsilon_i \langle z, \bar{W}^{(i)} \rangle \geq t\right) \\
&\leq 5^d \exp(-C_{11} C_{12}^2 d)
\end{aligned}
$$

note ϵ_i and $\langle z, \bar{W}^{(i)} \rangle$ are sub-Gaussian so the product of them is sub-exponential, we get the last inequality by Bernstein inequality. This finishes the proof. $\square$

Now with the bound of $\left\|\bar{W}\right\|_2$, we could find the bound of $\tilde{P}$ and $\mathbf{P}_2$. This all together could prove Theorem 1 by Lemma 2.

References

[1] Wainwright, M. J. (2019). *High-Dimensional Statistics: A Non-Asymptotic Viewpoint*. CUP.

[2] Roman Vershynin, *High-dimensional probability: An introduction with applications in data science*, Cambridge University Press, 2018.