MATH888: High-dimensional probability and statistics Fall 2021

Lecture 22 — October 27, 2021

Sebastien Roch, UW-Madison *Scribe: Sang Yup Han*

1 Overview

In the last lecture we applications to principal component analysis and previewed the proof of Cor. 8.7 in [1].

In this lecture we finish the proof of Cor. 8.7.

2 Review of Last Lecture

Reviewing last lecture, we talked about the spiked covariance model. As a review of the setup:

Definition 1. *Let $\boldsymbol{W} \in \mathbb{R}^d$ be an isotropic, subgaussian random vector with mean zero and norm ≤ 1, and let ϵ be an independent real-valued subgaussian random variable with mean zero and variance 1.*[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.

With this model in mind, we defined the following matrix:

$$\hat{\mathbf{\Sigma}}_n = \frac{1}{n}\sum_{i=1}^{n} \mathbf{X}^{(i)}\mathbf{X}^{(i)\top}$$

and the corresponding $\hat{\theta} \in \mathbb{R}^d$, the max eigenvalue of $\hat{\mathbf{\Sigma}}_n$. With these definitions in mind, we stated the following theorem from [1]

Theorem 2. *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 $\hat{\theta} - \theta^*{}_2 \leq C_1\sqrt{\frac{\nu+1}{\nu^2}}\sqrt{\frac{d}{n}}$ with probability $1 - C_2\exp\{-C_3 d\}$.*

Recall the perturbation matrix from last lecture, $\mathbf{P} := \hat{\mathbf{\Sigma}}_n - \mathbf{\Sigma}$. Define the following:

$$\tilde{\mathbf{p}} = \mathbf{U_2}^\top \mathbf{P}\theta^*$$

where the columns of $\mathbf{U}_2 \in \mathbb{R}^{d\times d-1}$ form an orthonormal basis of $\theta^{*\top}$

We have the following lemma, (Thm 8.5 in [1]), which we do not prove in lecture, but the proof boils down to linear algebra calculations.

[1] Note [1] uses ξ rather than ϵ

Theorem 3 (Theorem 8.5 in [1])**.** *If* $\|\mathbf{P}\|_2 < \frac{\nu}{2}$ *then*

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

3 Proof of Theorem 2

Now we present the proof of Theorem 2 by applying Theorem 3. To do this, we bound the norm of $\mathbf{P}$.

Proof. The idea of the proof is to divide $\mathbf{P}$ into three contributions.

$$\begin{aligned}\mathbf{P} &= \hat{\mathbf{\Sigma}}_n - \mathbf{\Sigma} \\ &= \nu\left(\frac{1}{n}\sum_{i=1}^{n}\epsilon_i^2 - 1\right)\theta^*\theta^{*\top} + \sqrt{\nu}\left(\bar{\mathbf{w}}\theta^{*\top} + \theta^*\bar{\mathbf{w}}^\top\right) + \left(\frac{1}{n}\sum_{i=1}^{n}\mathbf{w}_i\mathbf{w}_i^\top - I_d\right)\end{aligned}$$

□

where $\bar{\mathbf{w}} = \frac{1}{n}\sum_{i=1}^{n}\epsilon_i\mathbf{w}_i$. Intuitively, the terms break up in the following way. The first term is the main contributions from θ^*, the middle term deals with both the contributions from $\bar{\mathbf{w}}$ and θ^*. The last term deals with the expectation of the values of $\mathbf{w}_i$. Call the first term $\mathbf{P}_1$, the second term $\mathbf{P}_2$, the third term $\mathbf{P}_3$. By the triangle inequality, we have that:

$$\|\mathbf{P}\|_2 < \|\mathbf{P}_1\|_2 + \|\mathbf{P}_2\|_2 + \|\mathbf{P}_3\|_2$$

We go through these individually. Starting with $\mathbf{P}_3$, previously we showed that with probability $1 - 2e^{-u}$ that:

$$\|\mathbf{P}_3\| \leq C(\sqrt{\frac{d+u}{n}} + \frac{d+u}{n})$$

So we take $u = cd$. Use the fact that $\|\mathbf{W}^{(i)}\|_2 \leq 1$ to get that:

$$\|\mathbf{P}_3\| \simeq c'\sqrt{\frac{d}{n}}$$

Turning to $\mathbf{P}_1$, we have that the expectation of ϵ_i^2 is 1 because we took it be mean zero, variance 1 and subgaussian. We can consider each piece of $\mathbf{P}_1$ separately as:

$$\|\mathbf{P}_1\|_2 \leq \nu|\frac{1}{n}\sum_{i=1}^{n}\epsilon_i^2 - 1|\|\theta^*\theta^{*\top}\|_2$$

Note that $\|\theta^*\theta^{*\top}\|_2 = 1$. We also have the following lemma (Lemma 2.7.7 in [2]):

Lemma 4. *If* X, Y *are sub-gaussian, then:*

$$\|XY\|_{\psi_1} \leq \|X\|_{\psi_2}\|Y\|_{\psi_2}$$

In particular, $\|\epsilon_i^2\|_{\psi_1} \leq 1$.

Let $z = |\frac{1}{n}\sum_{i=1}^{n} \epsilon_i^2 - 1|$. By Bernstein's inequality and the fact that the expectation of ϵ^2 is 1, we prove using Cor 2.8.3 in [2] that:

$$\mathbb{P}(z \geq t) \leq 2\exp(-c\min(t^2, t)n)$$

where n is the number of samples. Take $t = c\sqrt{\frac{d}{n}}$. Letting this c be as small as we'd like, then we can get the bound:

$$\|\mathbf{P}_1\|_2 \leq \frac{\nu}{8}$$

Using Lemma 3 above, we can also get that:

$$\|\mathbf{P}_2\|_2 \leq \sqrt{\nu} \cdot 2\|\bar{\mathbf{w}}\theta^{*\top}\|_2 \leq 2\sqrt{\nu}\|\bar{\mathbf{w}}\|_2$$

Now we get the following series of claims.

Claim 5. *With probability* $1 - ce^{-c'd}$*, we have that:*

$$\|\bar{\mathbf{w}}\|_2 \leq c''\sqrt{\frac{d}{n}}$$

Suppose Claim 4 is true, then we have:

Claim 6. *With probability* $1 - ce^{-c'd}$*, we have that:*

$$\|\mathbf{P}_2\|_2 \leq c''\sqrt{\nu}\sqrt{\frac{d}{n}}$$

Suppose both Claim 4 and 5 are true, then we have:

Claim 7. *With probability* $1 - ce^{-c'd}$

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

The key to this claim is the assumption that $\sqrt{\frac{\nu+1}{\nu^2}}\sqrt{\frac{d}{n}} \leq C_0$. By letting C_0 be small enough, then we can get the above claim by combining the remaining inequalities. With this claim, we're one step closer to applying Theorem 2 (Thm 8.5 in [1]) to finish the proof of Cor. 8.7.

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.