High-Dimensional Probability and Statistics

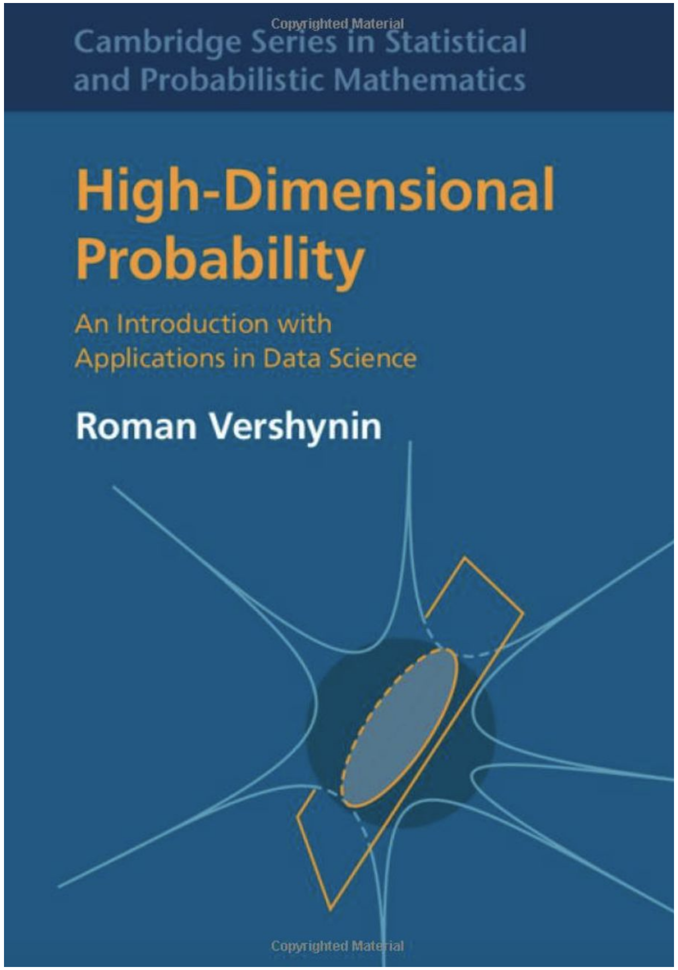
MATH/STAT/ECE 888: Topics in Mathematical Data Science

Sebastien Roch (Math+Stat)
UW-Madison

Fall 2021

Lecture 20 (10/22/21)

Today’s slides based on Vershynin



Matrix Bernstein

Matrix Bernstein inequality

Theorem 5.4.1 (Matrix Bernstein's inequality). *Let $X_1, \ldots, X_N$ be independent, mean zero, $n \times n$ symmetric random matrices, such that $\|X_i\| \le K$ almost surely for all i. Then, for every $t \ge 0$, we have*

$$\mathbb{P}\Big\{\Big\|\sum_{i=1}^{N} X_i\Big\| \ge t\Big\} \le 2n \exp\Big(-\frac{t^2/2}{\sigma^2 + Kt/3}\Big).$$

Here $\sigma^2 = \Big\|\sum_{i=1}^{N} \mathbb{E}\, X_i^2\Big\|$ is the norm of the matrix variance of the sum.

In particular, we can express this bound as the mixture of sub-gaussian and sub-exponential tail, just like in the scalar Bernstein's inequality:

$$\mathbb{P}\Big\{\Big\|\sum_{i=1}^{N} X_i\Big\| \ge t\Big\} \le 2n \exp\Big[-c \cdot \min\Big(\frac{t^2}{\sigma^2}, \frac{t}{K}\Big)\Big].$$

Application: covariance estimation

Like in Section 4.7, we estimate the the second moment matrix $\Sigma = \mathbb{E}\, XX^\mathsf{T}$ by its sample version

$$\Sigma_m = \frac{1}{m} \sum_{i=1}^{m} X_i X_i^\mathsf{T}.$$

Recall that if X has zero mean, then Σ is the covariance matrix of X and Σ_m is the sample covariance matrix of X.

Theorem 5.6.1 (General covariance estimation)**.** *Let X be a random vector in $\mathbb{R}^n$, $n \geq 2$. Assume that for some $K \geq 1$,*

$$\|X\|_2 \leq K \, (\mathbb{E}\, \|X\|_2^2)^{1/2} \quad \textit{almost surely.} \tag{5.16}$$

Then, for every positive integer m, we have

$$\mathbb{E}\, \|\Sigma_m - \Sigma\| \leq C \Big(\sqrt{\frac{K^2 n \log n}{m}} + \frac{K^2 n \log n}{m} \Big) \, \|\Sigma\|.$$

Application: covariance estimation cont'd

Exercise 5.6.4 (Tail bound). ☕☕ Our argument also implies the following high-probability guarantee. Check that for any $u \geq 0$, we have

$$\|\Sigma_m - \Sigma\| \leq C\Big(\sqrt{\frac{K^2 r(\log n + u)}{m}} + \frac{K^2 r(\log n + u)}{m}\Big)\, \|\Sigma\|$$

with probability at least $1 - 2e^{-u}$. Here $r = \mathrm{tr}(\Sigma)/\|\Sigma\| \leq n$ as before.