21-849 High-Dimensional Probability

Lectures by Konstantin Tikhomirov

Notes by Kadin Zhang

Fall 2024

Contents

1 High dimensional probability

1.1 8/26/24 - Subexponential and subgaussian random variables

Definition (Orlicz space): Let $\psi : [0, \infty) \to [0, \infty)$ be convex and increasing. Define the *Orlicz norm* of a random variable X as

$$\|X\|_\psi = \inf\{t > 0 : \mathbf{E}\psi(|X|/t) \le 1\}.$$

The *Orlicz space* L_ψ is the set of random variables on $(\Omega, \Sigma, \mathbf{P})$ with finite Orlicz norm.

Definition (Subgaussian, subexponential): X is *subexponential* if it has finite Orlicz norm with $\psi_1(t) := \exp(t) - 1$. X is *subgaussian* if it has finite Orlicz norm with $\psi_2(t) := \exp(t^2) - 1$.

Theorem (Characterization of subexponential and subgaussian): X is subexponential if and only if

$$\frac{(\mathbf{E}|X|^p)^{1/p}}{p} < \infty.$$

X is subgaussian if and only if

$$\frac{(\mathbf{E}|X|^p)^{1/p}}{\sqrt{p}} < \infty.$$

Theorem: Let $X_1, \ldots, X_n$ independent, mean zero, sub-gaussian. Then $\sum_i X_i$ is sub-gaussian and there is $C > 0$ such that

$$\left\|\sum X_i\right\|_{\psi_2}^2 \le C \sum_i \|X_i\|_{\psi_2}^2.$$

Proof. Combinatorial. □

1.2 8/28/24 - Kchintchine and Hoeffding

Proposition: Let $1 \le p < \infty$. There are c_p, C_p only depending on p such that:

(a) If $\|X\|_{\psi_p} \le 1$, then

$$\mathbf{P}(|X| \ge t) \le \exp(-c_p t^p), \quad t \ge 2.$$

(b) If $\mathbf{P}(|X| \ge t) \le \exp(-t^p)$ for $t \ge 2$, then

$$\|X\|_{\psi_p} \le C_p.$$

Lemma: Let $X_1, \dots, X_n$ independent, mean zero, and $\|X_i\|_{\psi_2} \le 1$. For any real numbers $a_1, \dots, a_n$,

$$\left\| \sum a_i X_i \right\|_{\psi_2}^2 \le C \sum a_i^2.$$

Theorem (Kchintchine's inequality): Let $r_1, \dots, r_n$ be independent Rademacher. Fix $(a_1, \dots, a_n) \in S^{n-1}$. Then $\sum a_i r_i$ is $O(1)$ subgaussian, i.e.

$$\left\| \sum a_i r_i \right\|_{\psi_2} \le C.$$

for some universal constant C.

Proof. Direct from lemma. □

Theorem (Hoeffding's inequality): Let $X_1, \dots, X_n$ independent, mean zero, sub-gaussian. Then for $t > 0$,

$$\mathbf{P}\left(\left| \sum X_i \right| \ge t \right) \le 2 \exp\left(-\frac{ct^2}{\sum_i \|X_i\|_{\psi_2}^2} \right).$$

Proof. By the lemma, $\sum X_i$ is subgaussian with

$$\left\| \sum X_i \right\|_{\psi_2} \le C \sqrt{\sum \|X_i\|_{\psi_2}^2}.$$

So dividing by the RHS gives us a variable with ψ_2 norm at most 1, and we can apply the earlier proposition. □

Theorem (Hoeffding's inequality for bounded RV): Let $X_1, \dots, X_n$ independent taking values in bounded intervals $[a_i, b_i]$. Then for $t > 0$,

$$\mathbf{P}\left(\left| \sum X_i - \sum \mathbf{E} X_i \right| \ge t \right) \le C \exp\left(-\frac{ct^2}{\sum (b_i - a_i)^2} \right).$$

Proof. Note that $X_i - \mathbf{E} X_i$ is mean zero subgaussian with $\|X_i\|_{\psi_2} \le b_i - a_i$. Then we can apply Hoeffding's inequality in its general form. □

1.3 8/30/24 - Example, symmetrization trick

Example: Let $v_1, \dots, v_n \in S^{n-1}$ fixed. Then

$$\mathbf{P}\left(\left\| \sum_i r_i v_i \right\|_2 \ge C \sqrt{n \log n} \right) \le n^{-100}.$$

1.4 9/4/24 - Bernstein's inequality

Theorem (Bernstein's inequality): Let $X_1, \ldots, X_n$ independent, zero mean, sub-exponential. Then

$$\mathbf{P}\left(\left|\sum X_i\right| \geq t\right) \leq 2\exp\left(-C\min\left(\frac{t^2}{\sum_i \|X_i\|_{\psi_1}^2}, \frac{t}{\max_{i\leq n}\|X_i\|_{\psi_1}}\right)\right).$$

Proof. MGF and Markov-Chebyshev. □

Example: Let X random vector zero mean, unit variance, sub-gaussian components bounded by K. By Bernstein inequality,

$$\mathbf{P}\left(\left|\sum X_i^2 - n\right| \geq t\right) \leq 2\exp\left(-C\min\left(\frac{t}{k^2}, \frac{t^2}{k^4 n}\right)\right).$$

Show that

$$\mathbf{P}\left(\left|\|X\|_2 - \sqrt{n}\right| \geq t\right) \leq 2\exp\left(-C_k t^2\right),$$

from which we can conclude that $\|X\|_2 - \sqrt{n}$ is sub-gaussian.

Proof. □

Proposition:

Theorem (Chernoff's inequality): Let X_i independent Bernoulli with parameter p. Then

$$\mathbf{P}\left(\sum_i b_i - p_n \geq tpn\right) \leq \left(\frac{\exp t}{(1+t)^{1+t}}\right)^{pn}.$$

1.5 9/6/24 - Subgaussian RVs, Johnson-Lindenstrauss lemma

Definition (K-subgaussian random vectors): Let X be random vector in $\mathbb{R}^n$, $K < \infty$. X is *K-subgaussian* if

$$\|X\|_{\psi_2} \coloneqq \max_{y\in S^{n-1}} \|\langle X, y\rangle\|_{\psi_2} \leq K.$$

Corollary of Hoeffding's inequality:

Theorem: Let X be a random vector in $\mathbb{R}^n$ with independent, zero mean components, identity covariance, and $\max_{i \le n} \|X_1\|_{\psi_2} \le K$. Then for every $(a_1, \dots, a_n) \in S^{n-1}$,

$$\sum_i a_i X_i$$

is CK-subgaussian.

Theorem (Johnson-Linderstrauss lemma): Consider a dataset of m points in $\mathbb{R}^n$, $\mathcal{C} = \{x_1, \dots, x_m\}$. For every $n \ge 1$, $m \ge 3$, $\varepsilon \in (0, 1/2]$, and $k \ge C\varepsilon^{-2} \log m$, there exists a linear mapping $\phi : \mathbb{R}^n \to \mathbb{R}^k$ where $k \ll n$ such that for all $x, y \in \mathbb{R}^n$,

$$1 - \varepsilon \le \frac{\|\phi(x) - \phi(y)\|_2}{\|x - y\|_2} \le 1 + \varepsilon.$$

1.6 9/9/24 - Sparse JL

1.7 9/11/24 - ε-net arguments

Definition (ε-net): Let (T, ρ) be a metric space and $S \subset T$. A subset $\mathcal{N} \subset T$ is a ε-net for S if for every $x \in S$ there exists $y = y(x) \in \mathcal{N}$ such that $\rho(x, y) \le \varepsilon$. We say $\mathcal{N}$ is a ε-net in S if $\mathcal{N} \subset S$.

Definition (Operator norm): Let A be a $m \times n$ matrix. We define the operator norm as the largest factor by which the linear operator A can stretch a vector:

$$\|A\| := \max_{x \in \mathbb{R}^n \setminus \{0\}} \frac{\|Ax\|_2}{\|x\|_2} = \max_{x \in S^{n-1}} \|Ax\|_2.$$

Note that this is equal to the largest singular value of A.

Proposition (Operator norm on net): Let A be $m \times n$ matrix and $\varepsilon \in [0, 1)$. For an ε-net $\mathcal{N}$ of S^{n-1},

$$\sup_{x \in \mathcal{N}} \|Ax\|_2 \le \|A\| \le \frac{1}{1 - \varepsilon} \sup_{x \in \mathcal{N}} \|Ax\|_2.$$

Proof. Let $x \in S^{n-1}$ for which $\|Ax\|_2 = \|A\|$. Choose $x_0 \in \mathcal{N}$ such that $\|x - x_0\|_2 \le \varepsilon$. Then,

$$\|Ax - Ax_0\|_2 = \|A(x - x_0)\|_2 \le \|A\| \|x - x_0\|_2 \le \varepsilon \|A\|.$$

Triangle inequality gives us

$$\|Ax_0\|_2 \ge \|Ax\|_2 - \|Ax - Ax_0\|_2 \ge (1 - \varepsilon)\|A\|.$$

Dividing by $1 - \varepsilon$,

$$\|A\| \le \frac{1}{1 - \varepsilon} \|Ax_0\|_2 \le \frac{1}{1 - \varepsilon} \sup_{x \in \mathcal{N}} \|Ax\|_2.$$

□

Proposition: If A is symmetric $n \times n$, then $\|A\| = \max_{x \in S^{n-1}} |\langle Ax, x\rangle|$.

Proof. If A is symmetric, it has orthonormal basis $v_1, \dots, v_n$, which are eigenvectors with some eigenvalues $\lambda_1, \dots, \lambda_n$. Then for $x \in S^{n-1}$, write $x = a_1 v_1 + \cdots + a_n v_n$, where $a = (a_1, \dots, a_n)$ is a unit vector.

$$\begin{aligned}\langle Ax, x\rangle &= \langle a_1 A v_1 + \cdots + a_n A v_n, a_1 v_1 + \cdots + a_n v_n\rangle \\ &= a_1^2 \lambda_1 + \cdots + a_n^2 \lambda_n.\end{aligned}$$

If λ_i is maximum, then the optimal choice is $x = v_i$, which yields operator norm λ_i as expected. □

Proposition (Operator norm on net, symmetric case): Let M be $n \times n$ symmetric real matrix and $\mathcal{N}$ be an ε net in S^{n-1}. Then

$$\|M\| \le \frac{1}{1 - 2\varepsilon} \max_{y \in \mathcal{N}} |\langle My, y\rangle|.$$

Proof. Let y_x denote vector in $\mathcal{N}$ with distance at most ε from x by ε net.

$$\begin{aligned}\|M\| &\le \max_{x \in S^{n-1}} |\langle Mx, x\rangle| \\ &= \max_{x \in S^{n-1}} |\langle Mx, y_x\rangle| + \max_{x \in S^{n-1}} |\langle Mx, x - y_x\rangle| \\ &\le \max_{x \in S^{n-1}} |\langle x, My_x\rangle| + \varepsilon \|M\| \\ &\le \max_{x \in S^{n-1}} |\langle y_x, My_x\rangle| + \max_{x \in S^{n-1}} \|\langle x - y_x\rangle, My_x\| + \varepsilon\|M\| \\ &\le \max_{y \in \mathcal{N}} |\langle y, My\rangle| + 2\varepsilon \|M\|.\end{aligned}$$

Note that $\langle x, My_x\rangle = \langle y_x, Mx\rangle$ since M is symmetric. □

1.8 9/13/24 - RIP

Definition (Restricted isometry property): Let A be a $k \times n$ matrix. A satisfies *RIP* with parameters $a, b > 0$ and $s \in \mathbb{N}$ if for every s-sparse vector x in $\mathbb{R}^n$,

$$a\|x\|_2 \le \|Ax\|_2 \le b\|x\|_2.$$

Theorem (RIP implies exact recovery of sparse signals): Let A be $k \times n$ satisfying RIP with parameters a, b, s. Then for every $x \in \mathbb{R}^n$ with $|\{i \le n : x_i \ne 0\}|(1 + b^2/a^2) < s$,

$$x = \arg\min\{\|y\|_1 : Ay = Ax\}.$$

Theorem (RIP for random matrices): Let A be $m \times n$ random matrix with mean zero, unit variance, K-subgaussian entries. Let $n \ge s$, $m \ge C_s \log(en/s)$. Then, with high probability, A is RIP with parameters $0.9\sqrt{m}, 1.1\sqrt{m}, s$. In particular, we can reconstruct any $s/3$-sparse vector.

1.9 9/16/24 - Concentration of singular values

Lemma: S^{n-1} has a ε-net of size at most $(1+2/\varepsilon)^n$.

Theorem: Let A be $N \times n$ matrix with i.i.d. entries of zero mean, unit variance, subgaussian moment bounded by K. Then for any $t > 0$,

$$\mathbf{P}\Big(\sqrt{N} - C\sqrt{n} - Ct \le s_{\min}(A) \le s_{\max}(A) \le \sqrt{N} + C\sqrt{n} + Ct\Big) \ge 1 - 2e^{-ct^2},$$

where C, c only depend on K.

Proof. Let $M := A^\top A$. Let $\mathcal{N}$ be a 1/4-net in S^{n-1} with cardinality at most 9^n. We apply the ε-net argument for operator norm on symmetric matrices on $\frac{1}{N}M - I_n$:

$$\begin{aligned}\left\|\frac{1}{N} - I_n\right\| &\le 2\max_{y\in\mathcal{N}}\left|\left\langle \frac{1}{N}My - I_n y\right\rangle, y\right| \\ &= 2\max_{y\in\mathcal{N}}\left|\frac{1}{N}\|Ay\|_2^2 - 1\right|.\end{aligned}$$

By corollary of Bernstein, for fixed y, since By is a random vector with zero mean, unit variance, and bounded subgaussian moment, $\|By\|_2 - \sqrt{N}$ is subgaussian. So $\|By\|_2^2 - N$ is a sum of zero mean subexponential variables, and we may apply Bernstein's inequality:

$$\begin{aligned}\mathbf{P}(\left|\|By\|_2^2 - N\right| \ge t) &\le 2\exp\left(-c\min\left(\frac{t^2}{N}, t\right)\right) \\ &= 2\exp\left(-\frac{ct^2}{N+t}\right).\end{aligned}$$

Consider the inequality for $t = 2\sqrt{N}(C\sqrt{n}+t) + (C\sqrt{n}+t)^2$, where $C = \sqrt{2\ln 9/c}$. Then we obtain

$$\mathbf{P}\left(\left|\|By\|_2^2 - N\right| \ge 2\sqrt{N}(C\sqrt{n}+t) + (C\sqrt{n}+t)^2\right) \le 2\cdot 9^{-n}\exp(-ct^2/2).$$

Taking union bound over $y \in \mathcal{N}$ and substituting $\left|\frac{1}{N}\|By\|_2^2 - 1\right|$ with $\left\|\frac{1}{N}M - I_n\right\|$,

$$\mathbf{P}\left(\left\|\frac{1}{N}M - I_n\right\| \ge 2(C\sqrt{n}+t)/\sqrt{N} + (C\sqrt{n}+t)^2/N\right) \le 2\exp(-ct^2/2).$$

It suffices to use the deterministic relations

$$\max\big(|N^{-1}s_{\min}^2(A) - 1|, |N^{-1}s_{\max}^2(A) - 1|\big) \le \left\|\frac{1}{N}M - I_n\right\|.$$

$\square$

Corollary: If we let $t = c'\sqrt{n}$ for appropriate choice of c' we have that

$$\mathbf{P}(\sqrt{N} - C\sqrt{n} \le s_{\min}(A) \le s_{\max}(A) \le \sqrt{N} + C\sqrt{n}) \ge 1 - 2e^{-n}.$$

Remark: As $N, n \to \infty$, $s_{\max}(A) = (1+o(1))(\sqrt{N}+\sqrt{n})$ and $s_{\min}(A) = (1+o(1))(\sqrt{N}-\sqrt{n})$. This shows that our bounds are optimal.

1.10 9/18/24 - Covariance estimation

Consider a statistical model $((\mathbb{R}^n)^N, \mathcal{P}, N)$, $N \geq n$, where $\mathcal{P}$ is a family of distributions on $\mathbb{R}^n$ such that for every random vector X from $\mathcal{P}$,

$$\left\| \frac{\langle X, y\rangle}{\sqrt{\mathbf{E}\langle X, y\rangle}} \right\|_{\psi_2} \leq K.$$

We observe sample $\Sigma^{1/2}X_1, \dots, \Sigma^{1/2}X_N$, where $X_1, \dots, X_N$ are i.i.d. mean zero K-subgaussian isotropic vectors and Σ is the true covariance matrix. We then compute the sample covariance matrix as

$$\hat{\Sigma} = \frac{1}{N}\sum_{i=1}^{N}(\Sigma^{1/2}X_i)(\Sigma^{1/2}X_i)^\top.$$

So

$$\hat{\Sigma} - \Sigma = \Sigma^{1/2}\left(\frac{1}{N}\sum_{i=1}^{N} X_i X_i^\top - I_n\right)\Sigma^{1/2}.$$

We can then bound

$$\left\|\hat{\Sigma} - \Sigma\right\| \leq \|\Sigma\| \left\| \frac{1}{N}\sum_{i=1}^{N} X_i X_i^\top - I_n \right\|,$$

thus the ratio $\frac{\|\hat{\Sigma}-\Sigma\|}{\|\Sigma\|}$ is deterministically upper bounded by (letting $M := \frac{1}{N}\sum_{i=1}^{N} X_i X_i^\top$)

$$\max(\lambda_{\max}(M) - 1, 1 - \lambda_{\min}(M)).$$

Let A be the $N \times n$ matrix with rows X_i, $i \leq N$. Then, $M = \frac{1}{N}A^\top A$, so

$$\lambda_{\max}(M) = \frac{1}{N}s_{\max}(A)^2, \quad \lambda_{\min}(M) = \frac{1}{N}s_{\min}(A)^2.$$

A modified form of the ε-net argument for singular values holds for matrices with independent isotropic K-subgaussian rows (rather than i.i.d. mean zero, unit variance, subgaussian moment at most K at every entry). So with high probability in n,

$$s_{\max}(A) \leq \sqrt{N} + C\sqrt{n}, \quad s_{\min}(A) \geq \sqrt{N} - C\sqrt{n}.$$

Note that $N \geq n$, so

$$\lambda_{\max}(M) \leq 1 + C'\sqrt{\frac{n}{N}}, \quad \lambda_{\min}(M) \geq 1 - C'\sqrt{\frac{n}{N}},$$

thus

$$\mathbf{P}\left(\frac{\left\|\hat{\Sigma} - \Sigma\right\|}{\|\Sigma\|} \geq C\sqrt{\frac{n}{N}}\right) \geq 1 - 2e^{-n}.$$

1.11 9/20/24 - Linear algebra facts

Definition (Schatten norm): Let $1 \le p \le \infty$. If A is $m \times n$ matrix with singular values $s_1(A), \dots, s_n(A)$, then define the *Schatten norm* as

$$\|A\|_p := \|(s_1(A), \dots, s_n(A))\|_p.$$

Remark: If $p = \infty$ this is called *spectral norm.* If $p = 2$, then this is called *Frobenius/Hilbert-Schmidt norm*, and

$$\|A\|_2 = \sqrt{\sum_{i,j} a_{ij}^2}.$$

This turns the space of matrices into Euclidean space with dimension equal to the minimum of A, B, or $\operatorname{tr}(AB^\top)$?

Proposition (Circular property of trace): If AB, BA well defined,

$$\operatorname{tr}(AB) = \operatorname{tr}(BA).$$

Corollary: If A, B are PSD, then $\operatorname{tr}(AB) \ge 0$.

Proof. Let $A = M^\top M$, $B = N^\top N$. Then $\operatorname{tr}(AB) = \operatorname{tr}(M^\top M N^\top N) = \operatorname{tr}(MN^\top (MN^\top)^\top) \ge 0$. □

Corollary: Let B PSD, A any symmetric matrix. Then

$$\operatorname{tr}(AB) \le \operatorname{tr}(\|A\| B) = \|A\| \operatorname{tr}(B).$$

Proof.

$$\operatorname{tr}((\|A\| I - A)B) \ge 0.$$

□

Definition (Matrix exponent): Let A be a square matrix. Then

$$e^A = \sum_{i=0} \frac{A^j}{j!}.$$

So

$$spec(e^A) = \exp(spec(A)).$$

Theorem (Golden-Thompson inequality): Let A, B symmetric. Then

$$\operatorname{tr}\exp(A + B) \le \operatorname{tr}(\exp(A)\exp(B)).$$

Theorem (Matrix Kchintchine inequality): Let $A_1, \dots, A_m$ deterministic symmetric $n \times n$ matrices. Let $r_1, \dots, r_m$ independent Rademacher. Then

$$\mathbf{P}\left(\left\|\sum_i r_i A_i\right\| \ge t\right) \le 2n \exp\left(-\frac{t^2}{4\|\sum_i A_i^2\|}\right).$$

Lemma: Let B symmetric random matrix, and $\lambda > 0$ be a parameter. Then

$$\mathbf{P}(\|B\| \geq t) \leq \mathbf{P}(\operatorname{tr}(\exp(\lambda B)) \geq \exp(\lambda t)) + \mathbf{P}(\operatorname{tr}(\exp(-\lambda B)) \geq \exp(\lambda t)).$$

Proof. Condition on realization of B s.t. $\|B\| \geq t$, so either $\lambda_{\max}(B) \geq t$ or $\lambda_{\min}(B) \leq -t$.

If $\lambda_{\max}(B) \geq t$, then $\lambda_{\max}(\exp(\lambda B)) \geq \exp(\lambda t)$, so $\operatorname{tr}(\exp(\lambda B)) \geq \exp(\lambda t)$. □

1.12 9/23/24 - Matrix concentration

Midterm: 2, 3, 4: 2 coffee cups Concentration inequalities / covering problem.

Lemma: Let $C_1, \dots, C_m$ independent symmetric $n \times n$, $a_1, \dots, a_m \in \mathbb{R}$ such that $\exp(a_i I) - \mathbf{E}\exp(\lambda C_i)$ is PSD for all i. Then

$$\mathbf{E}\operatorname{tr}(\exp(\lambda(C_1 + \cdots + C_m))) \leq n\exp(a_1 + \cdots + a_m).$$

Proof. Let $B = \sum_i r_i A_i$.

$$\begin{aligned}
\mathbf{P}(\|B\| \geq t) &\leq \frac{\mathbf{E}\operatorname{tr}(\exp(\lambda B))}{\exp(\lambda t)} \\
&\leq \mathbf{E}\operatorname{tr}(\exp(\lambda C_1)\exp(\lambda(C_2 + \cdots + C_m))) && \text{(G-T)} \\
&= \mathbf{E}\operatorname{tr}(\mathbf{E}(\exp(\lambda C_1))\exp(\lambda(C_2 + \cdots + C_m))) \\
&\leq \mathbf{E}\operatorname{tr}(\exp(a_i I)\exp(\lambda(C_2 + \cdots + C_m))) \\
&\leq \cdots \\
&\leq n\exp(a_1 + \cdots + a_m).
\end{aligned}$$

□

Theorem (Matrix Kchintchine inequality): Let $A_1, \dots, A_m$ deterministic symmetric $n \times n$ matrices. Let $r_1, \dots, r_m$ independent Rademacher. Then

$$\mathbf{P}\left(\left\|\sum_i r_i A_i\right\| \geq t\right) \leq 2n\exp\left(-\frac{t^2}{4\|\sum_i A_i^2\|}\right).$$

Proof.

$$\begin{aligned}
\mathbf{P}\left(\left\|\sum_i r_i A_i\right\| \geq t\right) &\leq 2\mathbf{P}\left(\operatorname{tr}\exp(\lambda\sum_i r_i A_i) \geq \exp(\lambda t)\right) \\
&\leq 2\frac{\mathbf{E}\operatorname{tr}\exp(\lambda\sum_i r_i A_i)}{\exp(\lambda t)}.
\end{aligned}$$

Consider the numerator. By G-T,

$$\mathbf{E}\operatorname{tr}\exp(\lambda\sum_i r_i A_i) \leq \mathbf{E}\operatorname{tr}\exp(\lambda^2(A_1^2 + \cdots + A_m^2))\exp\left(\sum_i(r_i\lambda A_i - \lambda^2 A_i^2)\right).$$

□

Theorem: Let $A_1, \ldots, A_m$ mean zero, independent, $n \times n$ symmetric with $\|\|A_i\|\|_{\psi_1} \leq K$. Then

$$\mathbf{P}(\|A_1 + \cdots + A_m\| \geq t) \leq$$

$$2n \exp\left(- \frac{ct^2}{\|\mathbf{E}(A_1^2 + \cdots + A_m^2)\| + tK \log\left(1 + \frac{mk^2}{\|\mathbf{E}(A_1^2 + \cdots + A_m^2)\|}\right)} \right)$$

Let $\{x_1, \ldots, x_N\}$ where x_i are i.i.d. mean zero vectors with $\mathbf{E} x_i x_i^\top = \Sigma$, and are K-subgaussian. The goal is to show

$$\left\| \Sigma - \frac{1}{N} \sum_{i=1}^N x_i x_i^\top \right\| \leq C_K \left(\sqrt{\frac{k(\Sigma) \log n}{N}} + \frac{k(\Sigma) \log^2(n)}{N} \right) \|\Sigma\|,$$

where $\|\Sigma\|$ is the intrinsic dimension of Σ.

1.13 9/25/24

Definition (Anisotropic random vector): Let $X \sim \Sigma$, $\mathbf{E} X = 0$. Then

(a) $\mathbf{E}\|X\|_2^2 = \mathbf{E} X^\top X = \mathbf{E} \operatorname{tr}(X^\top X)$

1.14 9/30/24 - Non-linear concentration

Theorem (Azuma): Let X be random variable, $\mathcal{F}_n$ a filtration.

$$\mathbf{P}(|X - \mathbf{E}X| \geq t) \leq 2 \exp\left(- \frac{ct^2}{\sum^N \|\mathbf{E}(X \mid \mathcal{F}_n) - \mathbf{E}(X \mid \mathcal{F}_{n-1})\|_\infty^2} \right), \quad t > 0.$$

Proof. From 721. □