Lecture 2. Upper and lower bounds for subgaussian matrices

1. The ε-net method refined

2. Random processes. Multiscale ε-net method: Dudley's inequality

Upper and lower bounds

- Our goal: upper and lower bounds on random matrices.
- In Lecture 1, we proved an upper bound for $N \times n$ subgaussian matrices A:

$$\lambda_{max}(A) = \max_{x \in S^{n-1}} \|Ax\| \leq C(\sqrt{N} + \sqrt{n})$$

  with exponentially large probability.
- How to prove a lower bound for

$$\lambda_{min}(A) = \min_{x \in S^{n-1}} \|Ax\|?$$

- Will try to prove both upper and lower at once: tightly bound $\|Ax\|$ above and below for all $x \in S^{n-1}$.

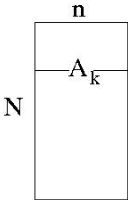
The ε-net method

- We need to tightly bound $\|Ax\|$ above and below for all $x \in S^{n-1}$.
  Discretization: replace the sphere S^{n-1} by a small ε-net $\mathcal{N}$;
  Concentration: for every $x \in \mathcal{N}$, the random variable $\|Ax\|$ is
  close its mean M with high probability (CLT);
  Union bound over all $x \in \mathcal{N} \Rightarrow$
  with high probability, $\|Ax\|$ is close to M for all x.
- Q.E.D.

Subexponential random variables

- What is the distribution of the r.v. $\|Ax\|$ for a fixed $x \in S^{n-1}$?
- Let A_k denote the rows of A. Then

$$\|Ax\|_2^2 = \sum_{k=1}^{N} \langle A_k, x \rangle^2.$$



- A is subgaussian $\Rightarrow$ each $\langle A_k, x \rangle$ is subgaussian.
- But we sum the *squares* $\langle A_k, x \rangle^2$. These are subexponential:

  X is subgaussian $\Leftrightarrow$ X^2 is subexponential.

  X is subexponential iff

$$\mathbb{P}(|X| > t) \leq 2\exp(-Ct) \qquad \text{for every } t > 0.$$

- We have a *sum* of subexponential i.i.d. r.v.'s.
  Central Limit Theorem should be of help:

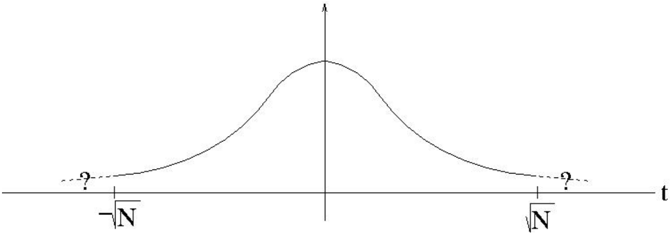
Concentration

Theorem (Bernstein's inequality)

Let $Z_1, \ldots, Z_N$ be independent subexponential centered r.v.'s. Then

$$\mathbb{P}\Big(\frac{1}{\sqrt{N}} \Big| \sum_{k=1}^{N} Z_k \Big| > t \Big) \le \exp(-ct^2) \qquad \text{for } t \le \sqrt{N}.$$

- The subgaussian tail says: CLT is valid in the range $t \le \sqrt{N}$.



- For *subgaussian* random variables, works for all t.
- The range of CLT *propagates* as $N \to \infty$.

Concentration

- Apply CLT to the sum of independent subgaussian random variables
$$\|Ax\|^2 = \sum_{k=1}^{N} \langle A_k, x\rangle^2.$$
- First compute the mean. Since the entries of A have variance 1, we have $\mathbb{E}\langle A_k, x\rangle^2 = 1$.
- Want to bound the deviation from the mean
$$\|Ax\|^2 - N = \sum_{k=1}^{N} \langle A_k, x\rangle^2 - 1,$$
which is a sum of independent subgaussian centered r.v.'s.
- CLT applies:
$$\mathbb{P}\Big(\frac{1}{\sqrt{N}}\big|\|Ax\|^2 - N\big| > t\Big) \le \exp(-ct^2) \qquad \text{for } t \le \sqrt{N}.$$

Concentration

- We proved the concentration bound

$$\mathbb{P}\Big(\frac{1}{\sqrt{N}}\big|\|Ax\|^2 - N\big| > t\Big) \le \exp(-ct^2) \qquad \text{for } t \le \sqrt{N}.$$

- Normalize by dividing by $\sqrt{N}$:

$$\mathbb{P}\Big(\big|\|\bar{A}x\|^2 - 1\big| > s\Big) \le \exp(-cs^2 N) \qquad \text{for } s \le 1.$$

- and can drop the square using the inequality $|a-1| \le |a^2 - 1|$.
- We thus tightly control $\|\bar{A}x\|$ near mean 1 for every fixed vector x.
- Now we need to unfix x, so that our concentration bound holds w.h.p. for all $x \in S^{n-1}$.

Discretization and union bound

- Discretization: approximate the sphere S^{n-1} by an ε-net $\mathcal{N}$ of . Can find with cardinality exponential in n: $|\mathcal{N}| \le (\frac{3}{\varepsilon})^n$.
- Union bound:

$$\mathbb{P}\Big(\exists x \in \mathcal{N}: \; \big|\|\bar{A}x\| - 1\big| > s\Big) \le |\mathcal{N}| \exp(-cs^2 N),$$

- which we can make very small, say $\le \varepsilon^n$, by choosing s appropriately large: $s \sim \sqrt{\frac{n}{N} \log \frac{1}{\varepsilon}} = \sqrt{y \log \frac{1}{\varepsilon}}$.
- Extend from $\mathcal{N}$ to the whole sphere S^{n-1} by approximation:
- Every point $x \in S^{n-1}$ can be ε-approximated by $y \in \mathcal{N}$, thus

$$\big|\|\bar{A}x\| - \|\bar{A}y\|\big| \le \|\bar{A}(x-y)\| \le \varepsilon\|\bar{A}\| \lesssim \varepsilon(1+\sqrt{y}) \le \varepsilon.$$

  (Here we used the upper bound from the last lecture).
- Conclusion: with high probability, for every $x \in S^{n-1}$,

$$\big|\|\bar{A}x\| - 1\big| \le s + \varepsilon \sim \sqrt{y \log \frac{1}{\varepsilon}} + \varepsilon.$$

  For $\varepsilon \le y$, the first term dominates. We have thus proved:

Conclusion:

Theorem (Upper and lower bounds for subgaussian matrices)

Let A be a subgaussian $N \times n$ matrix with aspect ratio $y = n/N$, and let $0 < \varepsilon \le y$. Then, with probability at least $1 - \varepsilon^n$,

$$1 - C\sqrt{y \log \tfrac{1}{\varepsilon}} \le \lambda_{min}(\bar{A}) \le \lambda_{max}(\bar{A}) \le 1 + C\sqrt{y \log \tfrac{1}{\varepsilon}}.$$

- Not yet quite final. Asymptotic theory predicts $1 \pm \sqrt{y}$ w.h.p., while Theorem can only yield $1 \pm \sqrt{y \log \frac{1}{y}}$.
  Will fix this later: prove Theorem with ε of constant order.
- Even in its present form, yields that
  the subgaussian matrices are restricted isometries.
- Indeed, we apply the Theorem w.h.p. for each minor, then take the union bound over all minors.

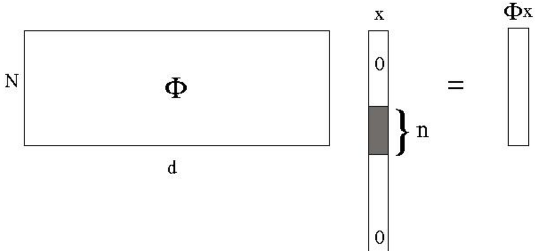
Theorem (Reconstruction from subgaussian measurements)

With exponentially high probability, an $N \times d$ subgaussian matrix Φ is a restricted isometry (for sparsity level n), provided that

$$N \sim n \log \frac{d}{n}.$$

Consequently, by Candes-Tao Restricted Isometry Condition, one can reconstruct any n-sparse vector $x \in \mathbb{R}^d$ from its measurements $b = \Phi x$ using the convex program

$$\min \|x\|_1 \quad \textit{subject to} \quad \Phi x = b.$$



Sharper bounds for subgaussian matrices

- So far, we match the asymptotic theory up to a log factor:
$$1 - C\sqrt{y \log \tfrac{1}{y}} \leq \lambda_{min}(\bar{A}) \leq \lambda_{max}(\bar{A}) \leq 1 + C\sqrt{y \log \tfrac{1}{y}}.$$
- Our goal: remove the log factor.
Would match the asymptotic theory up to a constant C.
- New tool: random processes. Multiscale ε-net method: Dudley's inequality.

From random matrices to random processes

- The desired bounds

$$1 - C\sqrt{y} \leq \lambda_{min}(\bar{A}) \leq \lambda_{max}(\bar{A}) \leq 1 + C\sqrt{y}$$

  simply say that $\|\bar{A}x\|^2$ is concentrated about its mean 1 for all vectors x on the sphere S^{n-1}:

$$\max_{x \in S^{n-1}} \left| \|\bar{A}x\|^2 - 1 \right| \lesssim \sqrt{y}.$$

- For each vector x,

$$X_x := \left| \|\bar{A}x\|^2 - 1 \right|$$

  is a random variable.
  The collection $(X_x)_{x \in T}$, where $T = S^{n-1}$, is a random process.
- Our goal: bound the random process:

$$\max_{x \in T} X_x \leq ? \qquad \text{w.h.p.}$$

General random processes

- Bounding random processes is a big field in probability theory.
- Let $(X_t)_{t \in T}$ be a centered random process on a metric space T. Usually, t is time (thus $T \subset \mathbb{R}$). But not in our case ($T = S^{n-1}$).
- Our goal: bound $\sup_{t \in T} X_t$ w.h.p. in terms of the geometry of T.
- General assumption on the process: controlled "speed". The size of the increments $X_t - X_s$ should be proportional to the "time" – the distance $d(t, s)$.
- An specific form of such assumption:

$$\frac{|X_t - X_s|}{d(t,s)} \quad \text{is subgaussian for every } t, s \in T.$$

  Such processes are called subgaussian random processes. Examples: gaussian processes, e.g. Brownian motion.
- The size of T is measured using the covering numbers $N(T, \varepsilon)$ (the number of ε-balls needed to cover T).

Dudley's Inequality

Theorem (Dudley's Inequality)

For a subgaussian process $(X_t)_{t \in T}$*, one has*

$$\mathbb{E} \sup_{t \in T} X_t \le C \int_0^\infty \sqrt{\log N(T, \varepsilon)}\, d\varepsilon.$$

- LHS probabilistic. RHS geometric.
- Multiscale ε-net method: uses covering numbers for all scales ε.
- ∞ can clearly be replaced by $\mathrm{diam}(T)$. Singularity at 0.
- "With high probability" version: $\frac{\sup_{t \in T} X_t}{\text{RHS}}$ is subgaussian.
- $\sqrt{\log u}$ is simply the inverse of $\exp(u^2)$ (the subgaussian tail).
- Holds for almost any other tail (e.g. subexponential), with corresponding inverse function in RHS.

The random matrix process

- Recall: for upper/lower bounds for subgaussian matrices, we need to bound the maximum of the random process $(X_x)_{x \in T}$ on the unit sphere $T = S^{n-1}$, where

$$X_x := \big| \|\bar{A}x\|^2 - 1 \big|.$$

- To apply Dudley's inequality, we need first to check the "speed" of the process – the tail decay of the increments:

$$I_{x,y} := \frac{X_x - X_y}{\|x - y\|}.$$

- As before, we write $\|\bar{A}x\|^2 = \sum_{k=1}^{N} \langle \bar{A}_k, x \rangle^2$, where $\bar{A}_k$ are the rows of $\bar{A}$. The sum of independent subexponential random variables.
- Use CLT (Bernstein's inequality) ... and get

$$\mathbb{P}(|I_{x,y}| > u) \leq 2\exp(-cN \cdot \min(u, u^2)) \qquad \text{for all } u > 0.$$

Mixture of subgaussian (in the range of CLT) and subexponential.

Applying Dudley's Inequality

- So, we know the "speed" of our random process

$$\mathbb{P}(|I_{x,y}| > u) \le 2\exp(-cN \cdot \min(u, u^2)) \qquad \text{for all } u > 0.$$

- To apply Dudley's inequality, we compute the inverse function of RHS as $\max\left(\frac{\log u}{N}, \sqrt{\frac{\log u}{N}}\right)$; we can bound the max by the sum.
- Then Dudley's inequality gives

$$\mathbb{E}\sup_{x\in T} X_x \lesssim \int_0^{1=\mathrm{diam}(T)} \left(\frac{\log N(T,\varepsilon)}{N} + \sqrt{\frac{\log N(T,\varepsilon)}{N}}\right) d\varepsilon.$$

- Recall: the covering number is exponential in the dimension: $N(T,\varepsilon) \le (\frac{3}{\varepsilon})^n$. Thus $\frac{\log N(T,\varepsilon)}{N} \le \frac{n}{N}\log(\frac{3}{\varepsilon}) = y\log(\frac{3}{\varepsilon})$.
- $\log(\frac{3}{\varepsilon})$ is integrable, as well as its square root. Thus

$$\mathbb{E}\sup_{x\in S^{n-1}} X_x \lesssim y + \sqrt{y} \lesssim \sqrt{y}.$$

- Recalling that $X_x = \left|\|\bar{A}x\|^2 - 1\right|$, we get the desired concentration:

Theorem (Sharp bounds for subgaussian matrices)

Let A be a subgaussian $N \times n$ matrix with aspect ratio $y = n/N$, Then, with high probability,

$$1 - C\sqrt{y} \leq \lambda_{min}(\bar{A}) \leq \lambda_{max}(\bar{A}) \leq 1 + C\sqrt{y}.$$

- High probability $=$ exponential in n.