Matrix deviations, random projections, Dvoretsky-Milman theorem

Recall from Lecture 11:

Thm (Talagrand's comparison inequality): Let $\{X_t\}_{t\in T}$, $\{Y_t\}_{t\in T}$ be separable mean-zero processes, $\{Y_t\}_{t\in T}$ Gaussian w/ canonical metric d, $\{X_t\}_{t\in T}$ subgaussian w.r.t. d. Then

$$\mathbb{E} \sup_{t\in T} X_t \leq C \cdot \mathbb{E} \sup_{t\in T} Y_t$$

We'll derive from this several interesting results in high-dimensional probability, statistics, and geometry.

Chevet's inequality and matrix deviations

Thm (Chevet's inequality): Let $X \in \mathbb{R}^{n\times m}$ have independent, mean-0, σ^2-subgaussian entries. Then for any $S \subset \mathbb{R}^n$, $T \subset \mathbb{R}^m$,

$$\mathbb{E} \sup_{u\in S, v\in T} u^T X v \leq C\sigma\left(w(S)\,\mathrm{rad}(T) + w(T)\,\mathrm{rad}(S)\right)$$

where $w(T) = \mathbb{E}_{g\sim N(0,I)} \sup_{t\in T} g^T t$ is the Gaussian width,

$\mathrm{rad}(T) = \sup_{t\in T} \|t\|_2$ is the radius.

Proof: Set $X_{uv} = u^T X v$. Then $X_{uv} - X_{wz} = \sum_{ij} X_{ij}(u_i v_j - w_i z_j)$

is τ^2-subgaussian for

$$\tau^2 = \sigma^2 \cdot \|uv^T - wz^T\|_F^2 \le \sigma^2 \left(\|(u-w)v^T\|_F + \|w(v-z)^T\|_F\right)^2$$

$$= \sigma^2 \left(\|u-w\|_2 \cdot \|v\|_2 + \|v-z\|_2 \cdot \|w\|_2\right)^2$$

$$\le 2\sigma^2 \left(\|u-w\|_2^2 \cdot \mathrm{rad}(T)^2 + \|v-z\|_2^2 \cdot \mathrm{rad}(S)^2\right)$$

Consider $Y_{uv} = \sqrt{2\sigma^2}\left(g^T u \cdot \mathrm{rad}(T) + h^T v \cdot \mathrm{rad}(S)\right)$

where $g \sim N(0, I_n)$, $h \sim N(0, I_m)$ independent.

$$\Rightarrow \mathbb{E}(Y_{uv} - Y_{wz})^2 = 2\sigma^2 \left(\|u-w\|_2^2 \cdot \mathrm{rad}(T)^2 + \|v-z\|_2^2 \cdot \mathrm{rad}(S)^2\right)$$

By subgaussian comparison thm,

$$\mathbb{E} \sup_{u \in S, v \in T} X_{uv} \le C \cdot \mathbb{E} \sup_{u \in S, v \in T} Y_{uv}$$

$$\le C' \sigma \left(w(S)\,\mathrm{rad}(T) + w(T)\,\mathrm{rad}(S)\right).$$

Example: If $S = B^n$ (unit ball in ℓ_2) then $w(S) = \mathbb{E}\|g\|_2 \le \sqrt{n}$, $\mathrm{rad}(S) = 1$. This recovers $\mathbb{E}\|X\|_{op} \le C(\sqrt{n} + \sqrt{m})$ from Lecture 7.

Example: If $S = [-1,1]^n$ (unit ball in ℓ_∞) then $w(S) = \mathbb{E}\|g\|_1 \asymp n$, $\mathrm{rad}(S) = \sqrt{n}$. So

$$\mathbb{E} \sup_{u \in [-1,1]^n,\, v \in [-1,1]^m} u^T X v \le C\sqrt{mn}\left(\sqrt{n} + \sqrt{m}\right).$$

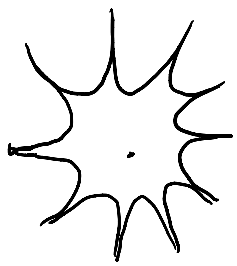
This is the same bound as if $[-1,1]^n$ were replaced by $\sqrt{n} \cdot B^n$.

Example: If $S = \{t \in \mathbb{R}^n : \|t\|_1 \le 1\}$ (unit ball in ℓ_1) then

$w(S) = \mathbb{E}\|g\|_\infty \asymp \sqrt{\log n}$, $\mathrm{rad}(S)=1$. So

$$\mathbb{E}\sup_{u,v:\|u\|_1,\|v\|_1\le 1} u^T X v \le C\left(\sqrt{\log n}+\sqrt{\log m}\right).$$

In high dimensions, ℓ_1-ball looks like:



Gaussian width is much smaller than ℓ_2-ball.

Thm (matrix deviation inequality): Let $X\in\mathbb{R}^{n\times m}$ have independent, mean-0, isotropic σ^2-subgaussian rows (i.e. $\mathbb{E}x_ix_i^T = I$ and u^Tx_i is σ^2-subgaussian for all unit vectors $u\in\mathbb{R}^n$). Then for any $T\subseteq\mathbb{R}^m$,

$$\mathbb{E}\sup_{u\in T}\Big|\|Xu\|_2 - \mathbb{E}\|Xu\|_2\Big| \le C\sigma^2\cdot\tilde{w}(T)$$

where $\tilde{w}(T) = \mathbb{E}_{g\sim N(0,I)}\sup_{t\in T}|g^Tt|$.

Proof: Define $X_u = \|Xu\|_2 - \mathbb{E}\|Xu\|_2$. We claim that

$$\|X_u - X_v\|_{\psi_2} \le C\sigma^2\|u-v\|_2 \quad \forall u,v\in\mathbb{R}^n. \quad (*)$$

Assuming $(*)$, noting $X_0=0$,

$$\mathbb{E}\sup_{u\in T}|X_u| \le \mathbb{E}\sup_{u,v\in T\cup\{0\}}|X_u-X_v| = \mathbb{E}\sup_{u,v\in T\cup\{0\}}\underbrace{X_u-X_v}_{=X_{uv}}$$

where $\|X_{uv} - X_{wz}\|_{\psi_2}^2 \le \left(\|X_u - X_w\|_{\psi_2} + \|X_v - X_z\|_{\psi_2}\right)^2$

$$\le 2C^2\sigma^4\left(\|u-w\|_2^2 + \|v-z\|_2^2\right) = \mathbb{E}\left(Y_{uv} - Y_{wz}\right)^2$$

for $Y_{uv} = \sqrt{2}\cdot C\sigma^2(u^Tg + v^Th)$, $g, h \overset{iid}{\sim} \mathcal{N}(0, I)$.

$$\Rightarrow \mathbb{E} \sup_{u,v \in T \cup \{0\}} X_{uv} \le C \cdot \mathbb{E} \sup_{u,v\in T\cup\{0\}} Y_{uv}$$

$$\le C'\sigma^2\, \mathbb{E} \sup_{t \in T\cup\{0\}} t^Tg \le C'\sigma^2\, \tilde{w}(T).$$

It remains to show $(*)$.

① Suppose first $\|u\|_2 = \|v\|_2 = 1$. Note that

$$\mathbb{P}\left[\frac{\left|\|Xu\|_2 - \|Xv\|_2\right|}{\|u-v\|_2} \ge s\right] = \mathbb{P}\left[\underbrace{\left|\frac{\|Xu\|_2^2 - \|Xv\|_2^2}{\|u-v\|_2}\right|}_{:=Z} \ge s\left(\|Xu\|_2 + \|Xv\|_2\right)\right]$$

$$\le \underbrace{\mathbb{P}\left[|Z| \ge s\sqrt{n}\right]}_{I} + \underbrace{\mathbb{P}\left[\|Xu\|_2 + \|Xv\|_2 \le \sqrt{n}\right]}_{II}.$$

- Xu has mean-zero, variance 1, σ^2-subgaussian entries

  $\Rightarrow \mathbb{P}\left[\left|\|Xu\|_2 - \sqrt{n}\right| \ge t\right] \le 2e^{-\frac{ct^2}{\sigma^4}}$ (Lecture 2)

  Similarly for $\|Xv\|_2$. So $II \le 2e^{-\frac{cn}{\sigma^4}}$.

- $Z = \sum_{i=1}^n Z_i$, $Z_i := \dfrac{(x_i^Tu)^2 - (x_i^Tv)^2}{\|u-v\|_2}$ where

$\mathbb{E} Z_i = 0$, and $Z_i = \dfrac{X_i^T(u-v) \cdot X_i^T(u+v)}{\|u-v\|_2}$, so

$$\|Z_i\|_{\psi_1} \le \frac{1}{\|u-v\|_2} \cdot \|X_i^T(u-v)\|_{\psi_2} \cdot \|X_i^T(u+v)\|_{\psi_2} \le C\sigma^2.$$

$\Rightarrow \mathrm{I} \le 2e^{-c\min\left(\frac{s^2}{\sigma^4}, \frac{s\sqrt{n}}{\sigma^2}\right)}$ by Bernstein inequality.

If $s \le 2\sqrt{n}$ (recalling $\sigma^2 \ge 1$), this shows $\mathrm{I} + \mathrm{II} \le 4e^{-\frac{cs^2}{\sigma^4}}$.

For $s > 2\sqrt{n}$, apply directly

$$\mathbb{P}\left[\frac{\big|\|Xu\|_2 - \|Xv\|_2\big|}{\|u-v\|_2} \ge s\right] \le \mathbb{P}\left[\left\|X \cdot \frac{u-v}{\|u-v\|_2}\right\|_2 \ge s\right]$$

$$\le \mathbb{P}\left[\left\|X \cdot \frac{u-v}{\|u-v\|_2}\right\|_2 - \sqrt{n} \ge \frac{s}{2}\right] \le 2e^{-\frac{cs^2}{\sigma^4}}.$$

Thus $\dfrac{\|Xu\|_2 - \|Xv\|_2}{\|u-v\|_2}$ is $C\sigma^4$-subgaussian, i.e.

$$\|X_u - X_v\|_{\psi_2} \le C \big\| \|Xu\|_2 - \|Xv\|_2 \big\|_{\psi_2} \le C'\sigma^2 \|u-v\|_2.$$

② For general u, v, assume WLOG $\|u\|_2 = 1$ and $\|v\|_2 \ge 1$.

Let $\bar{v} = \frac{v}{\|v\|_2}$. Then

$$\|X_u - X_v\|_{\psi_2} \le \underbrace{\|X_u - X_{\bar{v}}\|_{\psi_2}}_{\le C\sigma^2\|u-\bar{v}\|_2 \text{ by ① above.}} + \|X_{\bar{v}} - X_v\|_{\psi_2}$$

Since $|X_{\bar v} - X_v| = |X_{\bar v} - \|v\|_2 \cdot X_{\bar v}| = \|\bar v - v\|_2 \cdot |X_{\bar v}|$

$$\|X_{\bar v} - X_v\|_{\psi_2} \leq \|\bar v - v\|_2 \cdot \|X_{\bar v}\|_{\psi_2} \leq C\sigma^2 \|\bar v - v\|_2.$$

So
$$\|X_u - X_v\|_{\psi_2} \leq C\sigma^2 (\|u-\bar v\|_2 + \|\bar v - v\|_2)$$
$$\leq C\sigma^2 (\|u - v\|_2 + 2\underbrace{\|\bar v - v\|_2}_{\leq \|u-v\|_2}) \leq C'\sigma^2 \|u-v\|_2.$$

This shows $(*)$. $\blacksquare$

Remark: For $\|u\|_2 = 1$, $\mathbb{P}[\,|\|X u\|_2 - \sqrt{n}| > t] \leq 2e^{-\frac{ct^2}{\sigma^4}}$ so $|\mathbb{E}\|Xu\|_2 - \sqrt{n}| \leq C\sigma^2$. Then

$$\sup_{u\in T} |\mathbb{E}\|Xu\|_2 - \sqrt{n}\cdot\|u\|_2| \leq C\sigma^2 \cdot \sup_{u\in T} \|u\|_2 \leq C\sigma^2 \tilde{w}(T)$$

so we also have

$$\mathbb{E} \sup_{u\in T} |\|Xu\|_2 - \sqrt{n}\|u\|_2| \leq C\sigma^2 \tilde{w}(T).$$

If T is the unit sphere, then $\tilde{w}(T) = w(T) \leq \sqrt{m}$. So the theorem shows

$$s_{\min}(X),\ s_{\max}(X) \in \sqrt{n} \pm O_{\mathbb{P}}(\sqrt{m})$$

in accordance with the tail bound from Lecture 7.

An application with a different choice of T is the following:

Thm (low-rank covariance estimation): Let $X_1, \ldots, X_n \in \mathbb{R}^m$ be iid with $\mathbb{E}X_i = 0$, $\operatorname{Cov} X_i = \Sigma \in \mathbb{R}^{m \times m}$, and $\Sigma^{-1/2} X_i$ is σ^2-subgaussian. Let $\hat{\Sigma} = \frac{1}{n}\sum_{i=1}^{n} X_i X_i^T$ be the sample covariance estimate of Σ. For any $\delta > 0$, there exists $C(\delta) > 0$ s.t.

$$\mathbb{P}\left[\|\hat{\Sigma} - \Sigma\|_{op} > C(\delta)\sigma^4\left(\sqrt{\tfrac{r}{n}} + \tfrac{r}{n}\right)\|\Sigma\|_{op}\right] < \delta$$

where $r = \operatorname{Tr}\Sigma / \|\Sigma\|_{op}$ is the "stable rank" of Σ.

Note: $r \leq \operatorname{rank}(\Sigma) \leq m$ always. Unlike $\operatorname{rank}(\Sigma)$, r is stable with respect to perturbations.

Proof: Let $Z_i = \Sigma^{-1/2} X_i$, $Z = \begin{bmatrix} -z_1^T- \\ \vdots \\ -z_n^T- \end{bmatrix} \in \mathbb{R}^{n \times m}$. Then

$$\|\hat{\Sigma} - \Sigma\|_{op} = \|\Sigma^{1/2}(\tfrac{1}{n} Z^T Z - I)\Sigma^{1/2}\|_{op}$$

$$= \sup_{v \in S^{m-1}} |v^T \Sigma^{1/2}(\tfrac{1}{n} Z^T Z - I)\Sigma^{1/2} v|$$

$$= \sup_{u \in T} \frac{1}{n}\left|\|Zu\|_2^2 - n\|u\|_2^2\right| \qquad T = \Sigma^{1/2} S^{n-1}.$$

With probability $\geq 1-\delta$, by Markov's inequality,

$$\left|\|Zu\|_2 - \sqrt{n}\cdot\|u\|_2\right| \leq C(\delta)\sigma^2 \tilde{w}(T) \quad \forall u \in T$$

$$\Rightarrow \frac{1}{n}\left|\|Zu\|_2^2 - n\|u\|_2^2\right| = \frac{1}{n}\left|\|Zu\|_2 - \sqrt{n}\cdot\|u\|_2\right| \cdot \left|\|Zu\|_2 + \sqrt{n}\|u\|_2\right|$$

$$\leq \frac{1}{n} \cdot C(\delta)\sigma^2 \tilde{w}(T) \cdot \left(C(\delta)\sigma^2 \tilde{w}(T) + 2\sqrt{n} \cdot \|u\|_2\right) \quad \forall u \in T.$$

Here: $\tilde{w}(T) = \mathbb{E} \sup_{u \in T} |g^T u| = \mathbb{E}\|\Sigma^{1/2} g\|_2 \leq \left(\mathbb{E}\|\Sigma^{1/2} g\|_2^2\right)^{1/2} = (\mathrm{Tr}\,\Sigma)^{1/2}$

$\|u\|_2 \leq \|\Sigma^{1/2}\|_{op} = \|\Sigma\|_{op}^{1/2}$

So w/ prob. $\geq 1-\delta$,

$$\|\hat{\Sigma} - \Sigma\|_{op} \leq C'(\delta)\sigma^4 \left(\frac{\mathrm{Tr}\,\Sigma}{n} + \frac{(\mathrm{Tr}\,\Sigma \cdot \|\Sigma\|_{op})^{1/2}}{\sqrt{n}}\right)$$

$$= C'(\delta)\sigma^4 \left(\frac{r}{n} + \sqrt{\frac{r}{n}}\right) \cdot \|\Sigma\|_{op}.$$

$\blacksquare$

Random projections

Let us specialize to $X \in \mathbb{R}^{n \times m}$ w/ iid $N(0,1)$ entries, and consider $P = \frac{1}{\sqrt{n}} X$ as a random projection from $\mathbb{R}^m$ to $\mathbb{R}^n$. For $T \subset \mathbb{R}^m$, what does PT look like?

Cor: For any $\delta > 0$, with probability at least $1-\delta$,

$$\|Pu - Pv\|_2 \in \|u - v\|_2 \pm \frac{C(\delta)}{\sqrt{n}} w(T) \quad \forall u, v \in T.$$

In particular, $\mathrm{diam}(PT) \leq \mathrm{diam}(T) + \frac{C(\delta)}{\sqrt{n}} \cdot w(T)$.

Proof: The preceding theorem + Markov's inequality imply, w/ prob. $\geq 1-\delta$,

$$\sup_{u,v \in T} \left| \|P(u-v)\|_2 - \|u-v\|_2 \right| \leq \frac{C(\delta)}{\sqrt{n}} \tilde{w}(T-T)$$

Note $\tilde{w}(T-T) = \mathbb{E} \sup_{u,v\in T} |g^T(u-v)|$

$$= \mathbb{E} \sup_{u,v\in T} g^T(u-v) = w(T) + w(-T) = 2w(T).$$

This shows the first statement. The second follows from

$$\operatorname{diam}(PT) = \sup_{u,v\in T} \|Pu-Pv\|_2, \quad \operatorname{diam}(T) = \sup_{u,v\in T} \|u-v\|_2. \quad \blacksquare$$

Interpretation: $\operatorname{diam}(PT)$ has a phase transition. When $\operatorname{diam}(T) \gg \frac{1}{\sqrt{n}} w(T)$, i.e. $n \gg \frac{w(T)^2}{\operatorname{diam}(T)^2}$, we have $\operatorname{diam}(PT) \approx \operatorname{diam}(T)$, and P is a near-isometry:

$$\|Pu - Pv\|_2 = \|u-v\|_2 + o(\operatorname{diam} T) \quad \forall u,v\in T.$$

Here $\frac{w(T)^2}{\operatorname{diam}(T)^2}$ is called the <u>stable dimension</u> of T.

If T is finite, then by the maximal inequality

$$w(T) = \mathbb{E} \sup_{t\in T} g^T(t-t_0) \leq \sqrt{2\log|T|}\cdot \operatorname{diam}(T)$$

so the above holds as long as $n \gg \log|T|$, recovering the condition of the Johnson-Lindenstrauss Thm from Lecture 2.

In the complementary regime $\operatorname{diam}(T) \ll \frac{1}{\sqrt{n}} w(T)$, the following shows that PT looks instead like a Euclidean ball in $\mathbb{R}^n$.

Thm (Dvoretsky-Milman): Suppose T containing 0, let $\mathrm{conv}(PT)$ be the convex hull of PT, and B_2^n the unit ball. Then w/ probability at least $1-\delta$,

$$r_- B_2^n \subseteq \mathrm{conv}(PT) \subseteq r_+ B_2^n$$

where $r_\pm = \frac{1}{\sqrt{n}} w(T) \pm C(\delta)\,\mathrm{diam}(T)$.

Proof: The claim $r_- B_2^n \subseteq \mathrm{conv}(PT) \subseteq r_+ B_2^n$ is equivalent to:

$$\text{For all } u \in S^{n-1},\ r_- \le \sup_{x \in PT} x^T u \le r_+.$$

Consider
$$Z := \sup_{u \in S^{n-1}} \left| \sup_{x \in PT} x^T u - \mathbb{E} \sup_{x \in PT} x^T u \right|$$

$$= \frac{1}{\sqrt{n}} \sup_{u \in S^{n-1}} \Big| \underbrace{\sup_{y \in T} y^T X^T u - \mathbb{E} \sup_{y \in T} y^T X^T u}_{:= X_u} \Big|$$

We claim

$$\|X_u - X_v\|_{\psi_2} \le C \cdot \mathrm{diam}(T) \|u-v\|_2 \quad \forall u,v \in S^{n-1} \quad (*)$$

Assuming $(*)$, applying the subgaussian comparison theorem and same argument as in the matrix deviation inequality shows

$$\mathbb{E} \sup_{u \in S^{n-1}} |X_u| \le C' \cdot \mathrm{diam}(T)\, \tilde{w}(S^{n-1}) \le C\sqrt{n} \cdot \mathrm{diam}(T).$$

Then w/ probability $\ge 1-\delta$, $Z = \frac{1}{\sqrt{n}} \sup_{u \in S^{n-1}} |X_u| \le C(\delta) \cdot \mathrm{diam}(T)$.

For every $u \in S^{n-1}$, we have

$$\mathbb{E} \sup_{x \in PT} x^T u = \frac{1}{\sqrt{n}} \mathbb{E} \sup_{y \in T} y^T X^T u = \frac{1}{\sqrt{n}} w(T)$$

so $Z \leq C(\delta) \cdot \mathrm{diam}(T) \Rightarrow r_- \leq \sup_{x \in PT} x^T u \leq r_+$ as desired.

It remains to show $(*)$.

Let $f(x) = \sup_{y \in T} y^T x$. Note $X \mapsto y^T(a + X^T b)$ is $\|y\|_2 \cdot \|b\|_2$-Lipschitz so $X \mapsto f(a + X^T b)$ is $\mathrm{diam}(T) \cdot \|b\|_2$-Lipschitz. Same holds for $X \mapsto f(a - X^T b)$, so by Gaussian concentration,

$$\|f(a + X^T b) - \mathbb{E} f(a + X^T b)\|_{\psi_2} \leq C \,\mathrm{diam}(T) \cdot \|b\|_2$$

$$\|f(a - X^T b) - \mathbb{E} f(a - X^T b)\|_{\psi_2} \leq C \,\mathrm{diam}(T) \cdot \|b\|_2$$

Also $\mathbb{E} f(a + X^T b) = \mathbb{E} f(a - X^T b)$, so by triangle inequality,

$$\|f(a + X^T b) - f(a - X^T b)\|_{\psi_2} \leq C \cdot \mathrm{diam}(T) \cdot \|b\|_2 .$$

Take $b = \frac{u - v}{2}$, $a = X \cdot \frac{u + v}{2}$, noting a is independent of Xb b/c $\left(\frac{u+v}{2}\right)^T \left(\frac{u-v}{2}\right) = 0$. This shows

$$\|f(X^T u) - f(X^T v)\|_{\psi_2} \leq C \,\mathrm{diam}(T) \|u - v\|_2$$

conditional on a, and hence also unconditionally, which is the claim $(*)$. ∎