Liam Hardiman *May 6, 2020*

270C - Homework 2

3.1.4

Let $X = (X_1, \dots, X_n) \in \mathbb{R}^n$ be a random vector with independent, sub-gaussian coordinates X_i that satisfy $E[X_i^2] = 1$. Set $K = \max_i \|X_i\|_{\psi_2}$

(a) Show that

$$\sqrt{n} - CK^2 \leq E\|X\|_2 \leq \sqrt{n} + CK^2.$$

Proof. By Theorem 3.1.1, we have that $\|\|X\|_2 - \sqrt{n}\|_{\psi_2} \leq CK^2$ for some absolute constant C. In particular, we have that for any $t \geq 0$,

$$\Pr\left[\left|\|X\|_2 - \sqrt{n}\right| > t\right] \leq 2\exp(-t^2/(CK^2)^2).$$

This gives

$$\begin{aligned} E\left[\left|\|X\|_2 - \sqrt{n}\right|\right] &= \int_0^\infty \Pr\left[\left|\|X\|_2 - \sqrt{n}\right| > t\right]\, dt \\ &\leq \int_0^\infty 2\exp(-t^2/(CK^2)^2)\, dt \\ &= CK^2\sqrt{\pi}. \end{aligned}$$

Rearranging gives

$$\sqrt{n} - CK^2\sqrt{\pi} \leq E\|X\|_2 \leq \sqrt{n} + CK^2\sqrt{\pi}.$$

□

(b) Can CK^2 be replaced by $o(1)$?

Solution. □

3.1.5

Deduce from the previous exercise that

$$\mathrm{Var}(\|X\|_2) \leq CK^4.$$

Proof. We have that

$$\begin{aligned} \mathrm{Var}[\|X\|_2] &= E\left[(\|X\|_2 - E\|X\|_2)^2\right] \\ &= E\left[(\|X\|_2 - \sqrt{n})^2\right] + 2(\sqrt{n} - E\|X\|_2)E[\|X\|_2 - \sqrt{n}] + (\sqrt{n} - E\|X\|_2)^2. \end{aligned}$$

By the previous exercise, this quantity is less than

$$E\left[(\|X\|_2 - \sqrt{n})^2\right] + 3C^2K^4.$$

Now by theorem 3.1.1, $\|X\|_2 - \sqrt{n}$ is sub-Gaussian with norm K. Consequently, we can bound its second moment, $E[(\|X\|_2 - \sqrt{n})^2] \leq 2K^2$. Putting it all together, we have

$$\mathrm{Var}[\|X\|_2] \leq C^2K^4 + 2K^2.$$

□

3.1.6

Let $X = (X_1, \ldots, X_n) \in \mathbb{R}^n$ be a random vector with independent coordinates X_i that satisfy $EX_i^2 = 1$ and $EX_i^4 \leq K^4$. Show that

$$\mathrm{Var}[\|X\|_2] \leq CK^4.$$

Proof. First we claim that $E(\|X\|_2^2 - n)^2 \leq K^4 n$. This follows from simply expanding $\|X\|_2^4$.

$$\begin{aligned} E(\|X\|_2^2 - n)^2 &= E\left[\|X\|_2^4 - n^2\right] \\ &= \sum_{i=1}^n E[X_i^4] + 2\sum_{i<j} E[X_i^2 X_j^2] - n^2 \\ &\leq nK^4 + n(n-1) - n^2 \\ &\leq K^4 n. \end{aligned}$$

From this we have

$$\begin{aligned} K^4 n &\geq E\left[\left(\|X\|_2^2 - n\right)^2\right] \\ &= E\left[\left(\|X\|_2 - \sqrt{n}\right)^2 \left(\|X\|_2 + \sqrt{n}\right)^2\right] \\ &\geq nE\left[\left(\|X\|_2 - \sqrt{n}\right)^2\right], \end{aligned}$$

so $E(\|X\|_2 - \sqrt{n})^2 \leq K^4$. Since the mean minimizes the mean-square error, i.e.

$$\mathrm{Var}[\|X\|_2] \leq E[(\|X\|_2 - c)^2]$$

for all $c \in \mathbb{R}$, we deduce that $\mathrm{Var}[\|X\|_2] \leq CK^4$. □

3.1.7

Let $X = (X_1, \ldots, X_n) \in \mathbb{R}^n$ be a random vector with independent coordinates X_i with continuous distributions. Assume that the densities of X_i are uniformly bounded by 1. Show that for any $\epsilon > 0$, we have

$$\Pr[\|X\|_2 \le \epsilon\sqrt{n}] \le (C\epsilon)^n.$$

Proof. We square and apply the tried and true "multiply by λ and exponentiate" trick.

$$\Pr[\|X\|_2 \le \epsilon\sqrt{n}] = \Pr\left[-\frac{1}{\epsilon^2}\|X\|_2^2 \ge -n\right]$$

$$\le e^{\lambda n}\prod_{i=1}^{n} E\left[e^{-\lambda X_i^2/\epsilon^2}\right].$$

Let's bound those moment generating functions. If f_i is the density of X_i, then since $\left\|f_i\right\|_{L^\infty} \le 1$ for all i, we have

$$E\left[e^{-\lambda X_i^2/\epsilon^2}\right] = \int_{\mathbb{R}} e^{-\lambda x^2/\epsilon^2} f_i(x)\, dx$$

$$\le \int_{\mathbb{R}} e^{-\lambda X_i^2/\epsilon^2}\, dx$$

$$= \epsilon\sqrt{\pi/\lambda}.$$

Combining this with the preceding paragraph gives

$$\Pr[\|X\|_2 \le \epsilon\sqrt{n}] \le e^{\lambda n}(\epsilon\sqrt{\pi/\lambda})^n = \epsilon^n(e^{\lambda}\sqrt{\pi/\lambda})^n.$$

This holds for any value of $\lambda > 0$, so the result follows by choosing a value of λ. Optimizing gives $\lambda = 1/2$, so

$$\Pr[\|X\|_2 \le \epsilon\sqrt{n}] \le (\epsilon \cdot \sqrt{2\pi e})^n$$

□

3.2.6

Let X and Y be independent, mean zero, isotropic random vectors in $\mathbb{R}^n$. Check that

$$E\|X - Y\|_2^2 = 2n.$$

Proof. I don't have any clever expository things to say here.

$$E\|X - Y\|_2^2 = E[X^tX - X^tY - Y^tX + Y^tY]$$

$$= E[X^tX] - E[X^t]E[Y] - E[Y^t]E[X] + E[Y^tY]$$

$$= 2n.$$

□

3.3.3

Deduce the following properties from the rotation invariance of the normal distribution.

(a) Consider a random vector $g \sim \mathcal{N}(0, I_n)$ and a fixed vector $u \in \mathbb{R}^n$. Then

$$\langle g, u\rangle \sim \mathcal{N}(0, \|u\|_2^2).$$

Proof. Let U be a rotation matrix such that $Uu = \|u\|e_1$. That is, U rotates u so that it lies on the first coordinate axis. We then have

$$\langle g, u\rangle = \langle U^t U g, u\rangle = \langle Ug, \|u\|e_1\rangle.$$

By rotation invariance, $Ug \sim \mathcal{N}(0, I_n)$. Consequently, the above quantity is $\|u\| \cdot g_1 \sim \mathcal{N}(0, \|u\|^2)$ by the definition of the multivariate normal distribution. □

(b) Consider independent random variables $X_i \sim \mathcal{N}(0, \sigma_i^2)$. Then

$$\sum_{i=1}^{n} X_i \sim \mathcal{N}(0, \sigma^2) \quad \text{where} \quad \sigma^2 = \sum_{i=1}^{n} \sigma_i^2.$$

Proof. Consider the vector $u = (\sigma_1, \ldots, \sigma_n)$. If $g \sim \mathcal{N}(0, I_n)$, then by part (a) we have that $\langle g, u\rangle \sim \mathcal{N}(0, \|u\|^2) = \mathcal{N}(0, \sigma^2)$. On the other hand, since each $g_i \sim \mathcal{N}(0, 1)$, we have

$$\langle g, u\rangle = \sum_{i=1}^{n} \sigma_i g_i.$$

The claim follows from the fact that X_i equals $\sigma_i g_i$ in distribution. □

(c) Let G be an $m \times n$ Gaussian random matrix, i.e. the entries of G are independent $\mathcal{N}(0, 1)$ random variables. Let $u \in \mathbb{R}^n$ be a fixed unit vector. Then

$$Gu \sim \mathcal{N}(0, I_m).$$

Proof. The i-th coordinate of Gu is $\langle g_i, u\rangle$, where g_i is the i-th row of G. Now $g_i \sim \mathcal{N}(0, I_n)$, so $\langle g_i, u\rangle \sim \mathcal{N}(0, \|u\|^2) = \mathcal{N}(0, 1)$ by part (a) and all the coordinates of Gu are standard normal random variables.

It remains to show that the covariance matrix of Gu is I_m. Since u is a unit vector we have

$$\mathrm{Cov}(Gu) = E[(Gu)(Gu)^t] = E[G(uu^t)G^t] = E[GG^t].$$

Now since the entries of G are iid standard normal random variables, $E[GG^t] = I_m$ and the result follows. □

3.3.5

Let $X \sim \mathcal{N}(0, I_n)$.

(a) Show that, for any fixed vectors $u, v \in \mathbb{R}^n$, we have

$$E\langle X, u\rangle\langle X, v\rangle = \langle u, v\rangle.$$

Proof. Since $X \sim \mathcal{N}(0, I_n)$, we have $E[XX^t] = I_n$. Consequently, we have

$$E\langle X, u\rangle\langle X, v\rangle = E[u^t X v^t X] = E[u^t X X^t v] = u^t E[XX^t] v = \langle u, v\rangle.$$

□

(b) Given a vector $u \in \mathbb{R}^n$, consider the random variable $X_u = \langle X, u\rangle$. From exercise 3.3.3 we know that $X_u \sim \mathcal{N}(0, \|u\|_2^2)$. Check that

$$\left\|X_u - X_v\right\|_{L^2} = \|u - v\|_2$$

for any fixed vectors $u, v \in \mathbb{R}^n$.

Proof. By part (a) we have $E[X_u X_v] = \langle u, v\rangle$. Thus,

$$\begin{aligned}
\left\|X_u - X_v\right\|_{L^2}^2 &= E[|X_u - X_v|^2] \\
&= E[X_u^2 - 2X_u X_v + X_v^2] \\
&= \|u\|_2^2 - 2\langle u, v\rangle + \|v\|_2^2 \\
&= \|u - v\|_2^2.
\end{aligned}$$

□

3.4.3

(a) Let $X = (X_1, \ldots, X_n) \in \mathbb{R}^n$ be a random vector with sub-Gaussian coordinates X_i. Show that X is a sub-Gaussian random vector.

Proof. We use the moment characterization of sub-Gaussian random variables. Since each X_i is sub-Gaussian, $\left\|X_i\right\|_{L^p} \leq K_i\sqrt{p}$ for constants K_i and all $p \geq 1$. We then have

$$\begin{aligned}
\left\|\langle X, x\rangle\right\|_{L^p} &= \Big\|\sum_{i=1}^n x_i X_i\Big\|_{L^p} \\
&\leq \sum_{i=1}^n |x_i| \cdot \left\|X_i\right\|_{L^p} \\
&\leq \left(\sum_{i=1}^n |x_i| \cdot K_i\right)\sqrt{p},
\end{aligned}$$

so $\langle X, x\rangle$ is sub-Gaussian. $\square$

(b) Find an example of a random vector X with

$$\|X\|_{\psi_2} \gg \max_{i\leq n} \|X_i\|_{\psi_2}.$$

Solution. Let ξ be any real-valued sub-Gaussian random variable and consider the random vector $X = (\sqrt{n}\xi, \dots, \sqrt{n}\xi) \in \mathbb{R}^n$. X is sub-Gaussian by part (a). We clearly have

$$\max_{i\leq n} \|X_i\|_{\psi_2} = \left\|\sqrt{n}\xi\right\|_{\psi_2} = \sqrt{n}\cdot\|\xi\|_{\psi_2}.$$

On the other hand, we have

$$\begin{aligned}\|X\|_{\psi_2} &\geq \left\|\left\langle X, \frac{1}{\sqrt{n}}\mathbf{1}\right\rangle\right\|_{\psi_2} \\ &= \left\|\sum_{i=1}^{n} \sqrt{n}\xi \cdot \frac{1}{\sqrt{n}}\right\|_{\psi_2} \\ &= n\cdot\|\xi\|_{\psi_2},\end{aligned}$$

which is $\omega(\max_{i\leq n}\|X_i\|_{\psi_2})$. $\square$

3.4.4

Let $X \in \mathbb{R}^n$ be a random vector with coordinate distribution. That is, X is uniformly distributed in the set $\{\sqrt{n}e_i : i = 1, \dots n\}$. Show that

$$\|X\|_{\psi_2} \asymp \sqrt{\frac{n}{\log n}}.$$

Proof. $\square$

3.4.5

Let X be an isotropic random vector supported in a finite set $T \subseteq \mathbb{R}^n$. Show that in order for X to be sub-Gaussian with $\|X\|_{\psi_2} = O(1)$, the cardinality of the set must be exponentially large in n:

$$|T| \geq e^{cn}.$$

Proof. $\square$

3.4.10

Let $X = (X_1, \ldots, X_n) \in \mathbb{R}^n$ be random vector with independent, sub-Gaussian coordinates X_i that satisfy $EX_i^2 = 1$. Then

$$\left\| \|X\|_2 - \sqrt{n} \right\|_{\psi_2} \leq CK^2,$$

where $K = \max_i \|X_i\|_{\psi_2}$ and C is an absolute constant. Show that this concentration inequality may not hold for a general isotropic sub-Gaussian random vector X.

Proof. □

3.5.3

Let $A = (a_{ij})$ be a symmetric real $n \times n$ matrix. Suppose that A is either positive semidefinite or has zero diagonal. Assume that, for any numbers $x_i \in \{-1, 1\}$ we have

$$\left| \sum_{i,j} a_{ij} x_i x_j \right| \leq 1. \tag{1}$$

Then, for any Hilbert space H and any vectors $u_i, v_j \in H$ satisfying $\|u_i\| = \|v_j\| = 1$, we have

$$\left| \sum_{i,j} a_{ij} \langle u_i, v_j \rangle \right| \leq 2K,$$

where K is the absolute constant from Grothendieck's inequality.

Proof. Note that (1) can be written as $|\langle Ax, x \rangle| \leq 1$. Observe the following polarization identity.

$$\begin{aligned}
\left\langle A\left(\frac{x+y}{2}\right), \frac{x+y}{2} \right\rangle - \left\langle A\left(\frac{x-y}{2}\right), \frac{x-y}{2} \right\rangle &= \frac{1}{4}[\langle A(x+y), x+y \rangle - \langle A(x-y), x-y \rangle] \\
&= \langle Ax, y \rangle.
\end{aligned}$$

Then, if we set $u = \frac{1}{2}(x+y)$ and $v = \frac{1}{2}(x-y)$, we have.

$$|\langle Ax, y \rangle| \leq |\langle Au, u \rangle| + |\langle Av, v \rangle| \tag{2}$$

Unfortunately, the vectors u and v are in $\{-1, 0, 1\}^n$, not $\{-1, 1\}^n$. If we could show that our hypothesis holds for vectors of this form as well, then the above quantity would be less than 2 and we could apply Grothendieck's inequality. (So far we haven't used the symmetry of A.)

(The idea for this part comes from a stackexchange post) We claim that, under our hypotheses, for any $I \subset [n]$ and any $x \in \{-1, 1\}^n$ we have

$$-1 \leq \sum_i a_{ii} + \sum_{i \neq j \in I} a_{ij} x_i x_j \leq 1.$$

To see this, fix a subset $I \subseteq [n]$ and some vector $y \in \{-1,1\}^n$. Now consider the set of vectors $x \in \{-1,1\}^n$ that agree with y on I. There are $M = 2^{n-|I|}$ such vectors. For any such vector, we have by hypothesis

$$-1 \le \sum_i a_{ii} + \sum_{i \ne j} a_{ij} x_i x_j \le 1.$$

Now let's add all M of these inequalities together. Since each vector is in $\{-1,1\}^n$, the diagonal term $\sum_i a_{ii}$ will appear in each of them. Since each of these vectors agrees on I, we'll get a $\sum_{i \ne j \in I} a_{ij} x_i x_j$ for each of them. Now for every choice of coordinates outside of I, there is an "opposite" choice to make by flipping each coordinate outside of I. These sums cancel with each other, leaving only the diagonal and I terms. This gives

$$-M \le M \sum_i a_{ii} + M \sum_{i \ne j \in I} a_{ij} x_i x_j \le M.$$

Dividing through by M establishes the claim. (I don't think we used the symmetry of A anywhere here.)

Now let's return to our problem. Consider the case where A has zeros on its diagonal and let $u \in \{-1,0,1\}^n$. Let I be the support of u and let $\tilde{u}$ the vector that agrees with u on the support of u and is 1 elsewhere. By our claim we have

$$\langle Au, u \rangle = \sum_{i \ne j} a_{ij} u_i u_j = \sum_{i \ne j \in I} a_{ij} \tilde{u}_i \tilde{u}_j.$$

The last quantity is less than 1 in absolute value by hypothesis, so (2) is less than 2 and we can apply Grothendieck (still haven't used symmetry). Now suppose that A is PSD and let $u \in \{-1,0,1\}^n$. In this case we have $0 \le \langle Au, u \rangle$, so we only need to show that $\langle Au, u \rangle \le 1$. Since A is PSD, its diagonal entries are nonnegative. Letting I and $\tilde{u}$ be as in the zero-diagonal case, we have

$$\langle Au, u \rangle = \sum_{i \in I} a_{ii} + \sum_{i \ne j} a_{ij} u_i u_j \le \sum_i a_{ii} + \sum_{i \ne j \in I} a_{ij} \tilde{u}_i \tilde{u}_j.$$

The last quantity is less than 1 in absolute value, so again we can apply Grothendieck. (I don't think we used symmetry at any point here.)

□

3.5.7

3.6.7

Consider a random vector $g \sim \mathcal{N}(0, I_n)$. Show that for any fixed vectors $u, v \in S^{n-1}$ we have

$$E[\mathrm{sign}\langle g, u \rangle \mathrm{sign}\langle g, v \rangle] = \frac{2}{\pi} \arcsin \langle u, v \rangle.$$

Proof. By rotation invariance, we can assume that g lies in the plane determined by u and v. As shown in Figure 1, $\langle g, u \rangle \langle g, v \rangle$ is positive if and only if the angles between g and u and between g and v are

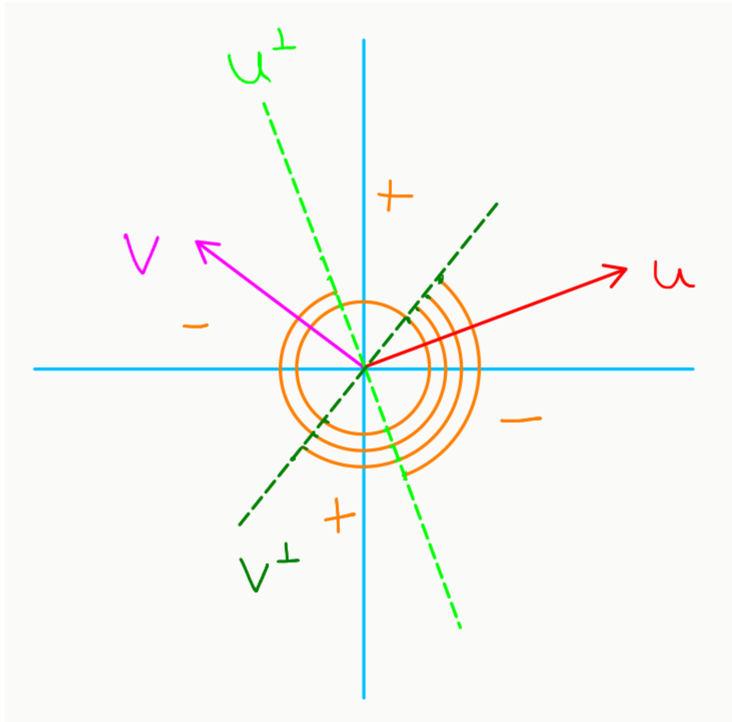

Figure 1: An orange "+" or "-" indicates the sign of $\langle g, u\rangle\langle g, v\rangle$ for g in the indicated region.

both acute or both obtuse. Call the event that g satisfies this condition E and let θ be the angle between u and v. By the rotation invariance of the multivariate normal distribution, the angle $\langle g, u\rangle/\|g\|_2$ is a uniform random variable. We then have

$$\begin{aligned} E[\text{sign}\langle g, u\rangle \text{sign}\langle g, v\rangle] &= \Pr[E] - \Pr[E^C] \\ &= \frac{2\pi - 2\theta}{2\pi} - \frac{2\theta}{2\pi} \\ &= \frac{2}{\pi}\left(\frac{\pi}{2} - \theta\right) \\ &= \frac{2}{\pi}\arcsin\langle u, v\rangle. \end{aligned}$$

□

4.4.3

Let A be an $m \times n$ matrix and $\epsilon \in [0, 1/2)$.

(a) Show that for any ϵ-net $\mathcal{N}$ of the sphere S^{n-1} and any ϵ-net $\mathcal{M}$ of the sphere S^{m-1}, we have

$$\sup_{x\in\mathcal{N}, y\in\mathcal{M}} \langle Ax, y\rangle \le \|A\| \le \frac{1}{1-2\epsilon} \cdot \sup_{x\in\mathcal{N}, y\in\mathcal{M}} \langle Ax, y\rangle. \tag{3}$$

Proof. The lower bound trivially follows from the fact that

$$\|A\| = \max_{x \in S^{n-1}, y \in S^{m-1}} \langle Ax, y \rangle.$$

As for the upper bound, fix vectors $x \in S^{n-1}$ and $y \in S^{m-1}$ that realize the operator norm bound: $\|A\| = \langle Ax, y \rangle$. Let $x_0 \in \mathcal{N}$ and $y_0 \in \mathcal{M}$ be such that $\|x - x_0\|_2 \leq \epsilon$ and $\|y - y_0\|_2 \leq \epsilon$. We have

$$\begin{aligned} |\langle Ax, y \rangle - \langle Ax_0, y_0 \rangle| &= |\langle Ax, y - y_0 \rangle + \langle A(x - x_0), y_0 \rangle| \\ &\leq \|A\| \|x\|_2 \|y - y_0\|_2 + \|A\| \|(x - x_0)\|_2 \|y_0\|_2 \\ &\leq 2\epsilon \|A\|. \end{aligned}$$

By the triangle inequality we then have

$$|\langle Ax_0, y_0 \rangle| \geq \|A\| - 2\epsilon \|A\|,$$

which gives the desired upper bound. □

(b) Moreover, if $m = n$ and A is symmetric, show that

$$\sup_{x \in \mathcal{N}} |\langle Ax, x \rangle| \leq \|A\| \leq \frac{1}{1 - 2\epsilon} \cdot \sup_{x \in \mathcal{N}} |\langle Ax, x \rangle|.$$

Proof. Use the exact same argument from part (a) since the operator norm of a symmetric matrix A is given by

$$\|A\| = \sup_{x \in S^{n-1}} \langle Ax, x \rangle.$$

□

4.6.4

4.7.5