3) $X \sim$ multivariate symmetric Bernoulli.

Obvious X is sub-g.

Since the coordinates are independent, the lemma above gives that

$$\|X\|_{\psi_2} \leq C \| \text{Sym Bernoulli} \| < C \text{ indep of } n.$$

4) Consider the coordinate distribution on $\mathbb{R}^n$: choose one of the coordinate vectors $e_i = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix}$ uniformly at random.

$$Y \sim \text{Unif}\{e_i : i=1,\dots,n\}.$$

$$EY=0, \quad Cov(Y) = \mathbb{E}(YY^T) = \frac{1}{n}\sum_{i=1}^{n} e_i e_i^T = \frac{1}{n} I$$

So if we multiply Y by $\sqrt{n}$ get an isotropic random vector, so let

$$X \sim \text{Unif}(\sqrt{n}\, e_i : i=1,\dots,n).$$

X – bdd $\Rightarrow$ it is sub-gaussian. However it has a very large sub-gaussian norm.

Indeed

$$\|X\|_{\psi_2} \approx \sqrt{\frac{n}{\log n}},$$

so for large n shouldn't really think of the coord distr as sub-gaussian – its sub-g norm is too large.

5) Consider $Y \sim \mathrm{Unif}(S^{n-1})$.

$$\|Y\|_2^2 = Y_1^2 + \dots + Y_n^2 = 1$$

Assuming the coordinates are roughly the same order, we see that Y_i should be of order $\frac{1}{\sqrt{n}}$.

Let's scale by $\sqrt{n}$.

$X \sim \sqrt{n}\, \mathrm{Unif}(S^{n-1})$. $\quad X = (X_1, \dots, X_n)$.

<u>Q:</u> How is X_1 distributed as $n \to \infty$?

We saw that if $g \sim N(0, I_n)$, then $\frac{g}{\|g\|_2} \sim \mathrm{Unif}(S^{n-1})$

So X & $\sqrt{n}\frac{g}{\|g\|_2}$ have the same distr.

So X_1 & $\frac{\sqrt{n}\, g_1}{\|g\|_2}$ have the same distr.

We have $\frac{\|g\|_2^2}{n} = \frac{g_1^2 + \dots + g_n^2}{n} \xrightarrow[\text{by SLLN}]{a.s.} E(g_1^2) = 1$

so $\|g\|_2 \xrightarrow{a.s.} \sqrt{n}$

Thus

$$\frac{\sqrt{n}}{\|g\|_2} g_1 \longrightarrow N(0,1).$$

It follows that if $Y \sim \mathrm{Unif}(S^{n-1})$, then

$$\sqrt{n}\, Y \longrightarrow N(0,1).$$

This is called the projective central limit thm.

Recall that Hoeffding's inequality is the quantitative version of the CLT. The quantitative version of the projective CLT is the following:

Thm: Let X be uniformly distributed on the unit sphere of radius $\sqrt{n}$.

$$X \sim \mathrm{Unif}(\sqrt{n}\, S^{n-1}).$$

Then X is sub-gaussian w/

$$\|X\|_{\psi_2} \le C.$$

Pf: We can represent X i.t.o. a gaussian. Let $g \sim N(0, I_n)$. Then we saw

$$\frac{g}{\|g\|_2} \text{ is } \mathrm{Unif}(S^{n-1}).$$

Let $X = \sqrt{n}\frac{g}{\|g\|_2}$. To show X is sub-gaussian,

NTS $\langle X, t\rangle$ is sub-gaussian $\forall\, t \in S^{n-1}$.

By the rotational invariance of X, it is enough to show the case of $t = (1, 0, \ldots, 0)$, i.e. when $\langle X, t\rangle = X_1$.

(also by the rotational invariance we will have $\|\langle X, t\rangle\|_{\psi_2}$ is indep of $t \in S^{n-1}$).

Let $$p(t) := P(|X_1| \ge t) = P\left(\frac{|g_1|}{\|g\|_2} \ge \frac{t}{\sqrt{n}}\right) = P\left(|g_1| \ge \frac{t\|g\|_2}{\sqrt{n}}\right)$$

We need to bound $p(t)$ from above.

Since $\|X\|_2 = \sqrt{n}$, $X_1 \le \sqrt{n}$ always, so $p(t) = 0$ if $t > \sqrt{n}$.

Thus, only need to bound $p(t)$ when $t < \sqrt{n}$.

Know that $\|g\|_2$ is close to $\sqrt{n}$, so with high probability $\|g\|_2 \geq \frac{\sqrt{n}}{2}$. Let A be the event $\|g\|_2 \geq \sqrt{n}$

More precisely, we know

$$\big\| \|g\|_2 - \sqrt{n} \big\|_{\psi_2} \leq C.$$

Let $Y = \|g\|_2 - \sqrt{n}$.

We have $E(e^{Y^2/\|Y\|_{\psi_2}^2} \leq 2)$ from the defn of $\|\cdot\|_{\psi_2}$.

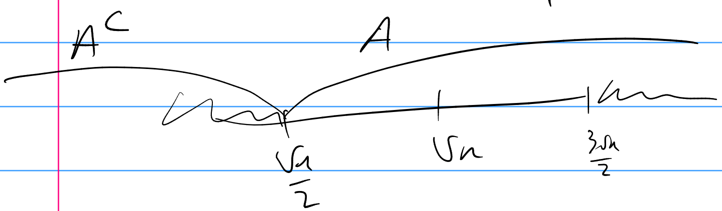
We showed that this is equivalent to

$$P(|Y| \geq t) \leq 2e^{-ct^2/\|Y\|_{\psi_2}^2} \qquad \forall t \geq 0$$

for some absolute const c.

Thus

$$P(A^c) \leq P\left(\big| \|g\|_2 - \sqrt{n} \big| > \frac{\sqrt{n}}{2} \right) \leq 2e^{-cn} \quad \text{for some const } c$$



Now

$$\begin{aligned}
P(t) &\leq P\left(\frac{|g_1|}{\|g\|_2} \geq \frac{t}{\sqrt{n}} \cap A \right) + P(A^c) \\
&\leq P\left(|g_1| \geq \frac{t}{2} \cap A \right) + 2e^{-cn} \\
&\leq P\left(|g_1| \geq \frac{t}{2} \right) + 2e^{-cn} \\
&\leq 2e^{-t^2/8} + 2e^{-cn}
\end{aligned}$$

Since $t < \sqrt{n}$ $\quad \leq 4e^{-c' t^2}$ for some const c'.

Thus X_1 is sub-gaussian w/ a sub-gaussian norm which is independent of n $\triangle$.

Grothendieck's inequality

This summer program started with Caratheodory's Thm, which was a deterministic result, however the proof introduced randomness. Yesterday saw another example in Steve's lecture.
Here is one more such example, this time using higher dim'l gaussians.

Thm (Grothendieck's inequality).
Consider an $m \times n$ matrix (a_{ij}) of real numbers. Suppose that for any $x_i, y_j \in \mathbb{R}$

$$\left|\sum_{ij} a_{ij} x_i y_j\right| \leq \max_i |x_i| \max_j |y_j|$$

(i.e. this multiply the i'th row by x_i, & the j'th column by y_j & sum the entries).
Then for any Hilbert space H and any vectors $u_i, v_j \in H$ we have

$$\left|\sum_{ij} a_{ij} \langle u_i, u_j \rangle\right| \leq K \max_i \|u_i\| \cdot \max_j \|v_j\|.$$

$K \leq 1.783$ is an absolute constant.

Rmk: While the statement holds for a constant $K \leq 1.783$ we will give a proof that gives $K \leq 8$.

Rmk: A Hilbert space is a vector space together with an inner product on it which makes it a complete metric space
(e.g. $\mathbb{R}^n$ w/ the std inner product, or space of square integrable fcns w/ the inner product $\langle f,g\rangle = \int f(x)g(x)\,dx$).

Pf:

1) Given an $m\times n$ matrix A, let $K=K(A)$ be the smallest K which makes the statement true for every Hilbert space H.
Note that $\tilde{K}=\sum_{ij}|a_{ij}|$ works, so the set of K's that work is not empty.
The key point of the thm is that $K(A)$ in fact does not depend on A, n or m.

2) Given $u_i, v_j \in H$ we need to show we can find
$K<2.88$ s.t. $$\left|\sum_{ij} a_{ij}\langle u_i, u_j\rangle\right| \le K \max_i \|u_i\| \max_j \|v_j\|.$$
Once u_i, v_j are selected, the space H does not play any role any more, so we can replace H by its subspace $\tilde{H}$ spanned by all the u_i's & v_j's.
$\tilde{H}$ is of dimension $\le N := m+n$, so it is isometric with a subspace of $\mathbb{R}^N$. Thus, without loss of generality we can assume $H=\mathbb{R}^N$ with the std inner product.

3) We need to bound

$$\left|\sum_{ij} a_{ij} \langle u_i, v_j \rangle\right|.$$

Let's realize $\langle u_i, v_j \rangle$ via random gaussian vectors.

Let $g \sim N(0, I_N)$ & define $U_j = \langle g, u_j \rangle$
$V_j = \langle g, v_j \rangle \ \forall j.$

U_i, V_j are linear combinations of independent mean-zero gaussians, so they are mean-zero gaussians.

Moreover

$$E U_i V_j = E(\overbrace{u_i^T g}^{\langle g, u_i \rangle} \overbrace{g^T v_j}^{\langle g, v_j \rangle}) = u_i^T (E g g^T) v_j = u_i^T v_j = \langle u_i, v_j \rangle$$

so

$$\sum_{ij} a_{ij} \langle u_i, v_j \rangle = E\left(\sum_{ij} a_{ij} U_i V_j\right).$$

This way we could turn the inner product $\langle u_i, v_j \rangle$ into the product $U_i V_j$ (at the cost of adding the expectation) to which we can apply the assumption of the thm:
for a given realization of U_i, V_j we can use the assumption in the thm to write

$$\left|\sum_{ij} a_{ij} U_i V_j\right| \le \max_i |U_i| \max_j |V_j|.$$

The issue here is that U_i, V_j are normal, so they are not bounded, so can be arbitrarily large.

4) Truncate the RVs U_i, V_j – separate into two parts, the 1st is bdd, the 2nd is unlikely (has small probability).

Given R, let

$$U_i^- := U_i \mathbb{1}_{|U_i| \le R\|U_i\|} \quad U_i^+ := U_i \mathbb{1}_{|U_i| > R\|U_i\|}$$

Similarly define

$$V_j^- := V_j \mathbb{1}_{|V_j| \le R\|V_j\|} \quad V_j^+ := V_j \mathbb{1}_{|V_j| > R\|V_j\|}$$

we have

$$U_i = U_i^+ + U_i^-$$

$$V_j = V_j^+ + V_j^-.$$

We have

$$\sum a_{ij} U_i V_j = \underbrace{\sum a_{ij} U_i^- V_j^-}_{S_1} + \underbrace{\sum a_{ij} U_i^+ V_j^-}_{S_2} + \underbrace{\sum a_{ij} U_i^- V_j^+}_{S_3} + \underbrace{\sum_{ij} a_{ij} U_i^+ V_j^+}_{S_4}$$

For S_1 by the hypothesis in the thm

$$|S_1| \le \max_i \|U_i^-\| \max_j \|V_j^-\| \le R^2 \max_i |u_i| \max_j |v_j|$$

$$\text{So} \quad E|S_1| \le R^2 \max_i |u_i| \max_j |v_j|$$

5) For S_2 we write

$$ES_2 = \sum_{ij} a_{ij} E(U_i^+ V_j^-).$$

Consider U_i^-, V_j^+ as elements of the Hilbert space L_2 with the inner product

$$\langle X, Y \rangle_{L_2} = EXY.$$

Our $K = K(A)$ works for any Hilbert space so we have

$$|ES_2| \leq K \max_i \|U_i^+\|_{L_2} \max_j \|V_j^-\|_{L_2}$$

Since $U_i = \langle g, u_i \rangle$, we have $U_i \sim N(0, \|u_i\|^2) \sim \|u_i\| N(0,1)$

Thus

$$\|U_i^+\|_{L_2}^2 = E\, U_i^2 1_{|U_i| > R\|u_i\|} = \|u_i\|^2 E\left(g^2 1_{|g|>R}\right)$$

where $g \sim N(0,1)$.

A simple integration by parts gives

$$\frac{1}{2} E g^2 1_{|g|>R} = E g^2 1_{g>R} = R \cdot \frac{1}{\sqrt{2\pi}} e^{-R^2/2} + P(g>R) \text{ so}$$

$$E g^2 1_{|g|>R} \leq 2\left(R + \frac{1}{R}\right) \frac{1}{\sqrt{2\pi}} e^{-R^2/2} =: C_R$$

We get $\|U_i^+\|_{L_2}^2 \leq \|u_i\|^2 C_R$, $\|V_j^-\|_{L_2} \leq \|V_j\|_{L_2} = \|v_j\|$

Thus $|ES_2| \leq K \cdot C_R \max_i \|u_i\| \max_j \|v_j\|$

Similarly $|ES_3| \leq K \cdot C_R$ —— // ——

& $|ES_4| \leq K\, C_R^2$ —— // ——

So $\left|E \sum_{ij} a_{ij} U_i V_j\right| \leq \left(R^2 + K(2C_R + C_R^2)\right)$ ——

K was the smallest which made ↑ work for all H, so

$$K \leq R^2 + K(2C_R + C_R^2) \text{ so } K \leq \frac{R^2}{1-(2C_R + C_R^2)}.$$

Plug in $R = 2.3$, get $K \leq 8$

$\triangle$