3) $X \sim$ multivariate symmetric Bernoulli. 

Observe $X$ is sub-g.$

Since the coordinates are independent, the lemma above gives that

$$||X||_{L_2} \leq C \cdot \text{Sym Bernoulli} \leq C \cdot \text{max of } X.$$

4) Consider the coordinate distribution on $\mathbb{R}^n$: 

choose one of the coordinate vectors $e_i = (0, \ldots, 1, \ldots, 0)$

uniformly at random.

$$Y \sim \text{Unif} \{ e_i : i \leq n \}.$$

$$EY = 0, \quad \text{Cov}(Y) = E(YY^T) = \frac{1}{n} \sum e_i e_i^T = \frac{1}{n} \Sigma.$$

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

So let

$$X \sim \text{Unif} (\sqrt{n} e_i : i \leq n).$$

$X \sim \text{ball of size } \sqrt{n} \text{ sub-gaussian. However,}$

$X$ has a very large sub-gaussian norm.

Indeed

$$||X||_{L_2} \sim \sqrt{n} \log n,$$

so for large $n$ shouldn't really think of the coord

dists as sub-gaussian - its sub-g norm is too large.
5) Consider $Y \sim \text{Unif}(S^{n-1})$.

$$\|Y\|_2^2 = Y_1^2 + Y_2^2 + \ldots + Y_n^2 = 1$$

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 $\sqrt{n}$.

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

Q: How is $X_1$ distributed as $n \to \infty$?

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

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

So $X_1 \sim \sqrt{n} \frac{g_1}{\|g\|_2}$ have the same dist.

We have

$$\frac{\|g\|_2^2}{n} = n \frac{g_1^2}{\|g\|_2^2} \overset{as}{\to} E(g_1^2) = 1$$

so $\|g\|_2 \overset{as}{\to} \sqrt{n}$

Thus

$$\frac{\sqrt{n}}{\|g\|_2} g_1 \overset{d}{\to} N(0, I).$$

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

$$\sqrt{n} Y \overset{d}{\to} N(0, I).$$

This is called the projection central limit theorem.
Recall that Hoeffding’s Inequality is the quantitative version of the CLT. The qualitative version of the projective CLT is the following:

Thus, let $X$ be uniformly distributed on the unit sphere of radius $\sqrt{n}$.

$$X \sim \text{Unif}(\mathbb{S}^{n-1}).$$

Then $X$ is sub-gaussian if

$$\|X\|_2 \leq C.$$ 

**Proof:** We can represent $X$ initially a Gaussian.

Let $g \sim N(0, I_n)$. Then we saw

$$g \sim \text{Unif}(\mathbb{S}^{n-1}).$$

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

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

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 $\|X_1\|_2$ is indep of $t \in \mathbb{S}^{n-1}$).

Let $p(t) = P(1X_1 > t) = P\left(\frac{|X_1|}{\|X\|_2} > \frac{t}{\sqrt{n}}\right) = P(1\|X\|_2 > \frac{t\|X\|_2}{\sqrt{n}})$.

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

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

Thus, only need to bound $p(t)$ when $t < \sqrt{n}$.
Know that $\|x\|_2$ is close to $S_n$, so with high probability $\|x\|_2 > \frac{S_n}{2}$. Let $A$ be the event $\|x\|_2 > \frac{S_n}{2}$.

More precisely, we know

$$\|\|x\|_2 - S_n\|_2 \leq C.$$ 

Let $Y = \|x\|_2 - S_n$.

We have $E(e^{tY}/\|x\|_2) \leq 2$ from the definition of $\|x\|_2$.

We showed that this is equivalent to

$$P(|Y| > t) \leq 2e^{-c\frac{t^2}{\|x\|_2}} \quad t > 0$$

for some absolute constant $c$.

Thus

$$P(A^c) \leq P(\|\|x\|_2 - S_n\|_2 > \frac{S_n}{2}) \leq 2e^{-cn} \quad \text{for some constant } c.$$

Now

$$P(A) = 1 - P(A^c) \leq 1 - 2e^{-cn} \leq 2e^{-cn} + 2e^{-cn} \leq 2e^{-\frac{c}{2}n} + 2e^{-cn}$$

Since $t < \frac{S_n}{3}$

$$\leq 2e^{-\frac{c}{3}n} + 2e^{-cn} \leq 4e^{-c'n} \quad \text{for some constant } c'$$

Thus $X_i$ is sub-gaussian w.r.t sub-gaussian norm which is independent of $n$. \(\triangle\)
Grothendieck's inequality

This summer program started with Caratheodory's Theorem, which was a deterministic result, however the proof introduced randomness. Yesterday saw another example in Steve's lecture.

Here is one more such example, thus true using higher dimensional concepts.

Theorem (Grothendieck's inequality).
Consider an m x n matrix \((a_{ij})\) of real numbers.
Suppose that for any \(x_i, y_j \in \mathbb{R}\)
\[
\left| \sum_{i,j} a_{ij} x_i y_j \right| \leq \max_i |x_i| \max_j |y_j|
\]
(i.e. 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
\[
\left| \sum_{i,j} a_{ij} \langle u_i, v_j \rangle \right| \leq K \max_i \|u_i\| \cdot \max_j \|v_j\|.
\]

\(K \leq 1.783\) is an absolute constant.

Remark: While the statement holds for a constant \(K \leq 1.783\), we will give a proof that gives \(K = 8\).
A Hilbert space is a vector space together with an inner product on it which makes it a complete metric space. (e.g. \( \mathbb{R}^n \) with the standard inner product, or space of square integrable functions with the inner product \( \langle f, g \rangle = \int f(x)g(x)dx \)).

1) Given a \( m \times n \) matrix \( A \), let \( K = K(A) \) be the smallest \( K \) which makes the statement true for every Hilbert space \( H \).

Note that \( K = \sum |a_{ij}| \) works, so the set of \( K \)'s that work is not empty.

The key point of the theorem is that \( K(A) \) in fact does not depend on \( A \) in \( \mathbb{R} \).

2) Given \( u_i, v_j \in H \) we need to show we can find \( K \geq 0 \) such that \[ \sum_{i,j} a_{ij} \langle u_i, u_j \rangle \leq K \max_{i,j} \max \| u \| \max \| v \|. \]

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 \( \overline{H} \) spanned by all the \( u_i \)'s and \( v_j \)'s.

\( \overline{H} \) is a subspace of \( \mathbb{R}^N \), thus, without loss of generality we can assume \( H = \mathbb{R}^n \) with the standard 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:
\[ g \sim N(0, I_n) \]
and define
\[ U_g = \langle g, u_i \rangle \]
\[ V_g = \langle g, v_j \rangle + v_j. \]

\( U_i, V_j \) are linear combinations of independent mean-zero Gaussians, so they are mean-zero Gaussians.

Moreover,
\[ \langle g, u_i \rangle \quad \langle g, v_j \rangle \]
\[ EU_i V_j = E \left( u_i^T g g^T v_j \right) = u_i^T (E g g^T) v_j = u_i^T v_j = \langle u_i, v_j \rangle. \]

So
\[ \sum_{ij} a_{ij} \langle u_i, v_j \rangle = \sum_{ij} a_{ij} E \langle u_i, v_j \rangle. \]

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 form:

for a given realization of \( U_i, V_j \) we can use the assumption in the form to write
\[ \left| \sum_{ij} a_{ij} U_i V_j \right| \leq \max_i |U_i| \max_j |V_j|. \]

The issue here is that \( U_i, V_j \) are normal, so they are not bounded, so \( \sum_{ij} a_{ij} U_i V_j \) can be arbitrarily large.
4) Truncate the RVs $U_i, V_j$ - separate into two parts, the 1st is small, the 2nd is unlikely (has small probability).

Given $R_i, k_i$

$$U_i^- := U_i \mathbf{1}_{|U_i| \leq R_{i|U_i|}} \quad U_i^+ := U_i \mathbf{1}_{|U_i| > R_{i|U_i|}}$$

Similarly define

$$V_j^- := V_j \mathbf{1}_{|V_j| \leq R_{i|V_j|}} \quad V_j^+ := V_j \mathbf{1}_{|V_j| > R_{i|V_j|}}$$

we have $U_i = U_0^+ + U_i^-$

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

We have

$$\sum a_{ij} U_i V_j = \sum a_{ij} U_i^- V_j^- + \sum a_{ij} U_i^+ V_j^- + \sum a_{ij} U_i^+ V_j^+ + \sum a_{ij} U_i^- V_j^+$$

For $S_1$ by the hypothesis on the sum

$$|S_1| \leq \max_i \|U_i\| \max_j \|V_j\| \leq R^2 \max_i \|U_i\| \max_j \|V_j\|$$

So

$$E|S_1| \leq R^2 \max_i \|U_i\| \max_j \|V_j\|$$

5) For $S_2$ we write

$$E S_2 = \sum_{i,j} a_{ij} E(U_i^+ V_j^-)$$

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 $k_l(k)$ works for any Hilbert space so we have

$$|E_{S_2}| \leq K \max_i \|U_i^+U_i\|_{L^2} \max_j \|V_j\|_{L^2}$$

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

Thus

$$\|U_i\|^2_{L^2} = E U_i^2 \frac{1}{|g|} \rightarrow \text{as } |g| \rightarrow \infty = \|u_i\|^2 E (g^2 \frac{1}{|g|})$$

$$\max_i \|U_i\|_{L^2} \approx N(0, 1).$$

A simple integration by parts gives

$$\frac{1}{2} E g^2 \frac{1}{|g|} \leq R \cdot \frac{1}{500} e^{-R/2} + P(|g| > R)$$

So

$$E g^2 \frac{1}{|g|} \leq 2(R + \frac{1}{R}) \frac{1}{500} e^{-R/2} =: C_R$$

We get

$$\|U_i\|_{L^2} \leq \|u_i\| \cdot C_R, \quad \|V_j\|_{L^2} \leq \|V_j\|_{L^2} = \|V_j\|$$

Thus

$$|E_{S_2}| \leq K \cdot C_R \max \|u_i\| \max \|V_j\|$$

Similarly

$$|E_{S_3}| \leq K \cdot C_R$$

and

$$|E_{S_4}| \leq K \cdot C_R^2$$

So

$$\sum_{i,j} a_{ij} U_i V_j \leq (R^2 + K(2C + C^2))$$

$K$ was the smallest which made the work for all $\lambda$, so

$$K \leq R^2 + K(2C + C^2)$$

So

$$K \leq \frac{R^2}{1 - (2C + C^2)}$$

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