3 Sums of independent random matrices

3.1 Distributions and random vectors in $\mathbb{R}^n$

- A random vector $X$ in $\mathbb{R}^n$ $\Rightarrow$ distribution in $\mathbb{R}^n$
- Expected value $\mu = \mathbb{E}X$
- Covariance matrix $\text{Cov}(X) = \mathbb{E}(X-\mu)(X-\mu)^T = (\text{Cov}(X_i, X_j))_{i,j=1}^n = \mathbb{E}XX^T - \mu\mu^T$.
- Second moment matrix: $\Sigma = \text{Cov}(X) = \mathbb{E}XX^T = (\mathbb{E}X_iX_j)_{i,j}$
  For $\mu = 0$, $\text{Cov}(X) = \Sigma(X)$

Properties of covariance matrices

(a) $\Sigma(X)$ and thus $\text{Cov}(X)$ are symmetric, positive-semidefinite (PSD) $\uparrow$
  (since $\Sigma_j = \mathbb{E}X_jX_j$) (check!)
(b) Hence the eigenvalues $\lambda_i$ of $\text{Cov}(X)$ are real, non-negative;
  SVD: $\text{Cov}(X) = \sum_{i=1}^n \lambda_i u_i u_i^T$
  $\uparrow$ eigenvectors
  Arrange $\lambda_i$ in non-increasing order. $\Rightarrow$ PCA

$\Sigma_2$ $\Sigma_1$ $\lambda_2$ $\lambda_1$ $u_2$ $u_1$ $\uparrow$
Def (Isotropic) A random vector $X$ in $\mathbb{R}^n$ is called isotropic if

$$\Sigma(X) = I.$$ 

(c) If $Z$ is isotropic, $\Sigma = V$ symmetric, PSD matrix $\Rightarrow$

$$X = \Sigma^{1/2} Z \quad \text{has} \quad \Sigma(X) = \Sigma.$$

(d) If $\Sigma(X) = \Sigma$ then

$$Z = \Sigma^{-1/2} X \quad \text{is isotropic,} \quad \Sigma(Z) = I.$$

Remark (c), (d) allow to study many problems for isotropic r. n.'s (by rescaling)

Example $Z \sim N(0, I_n)$ is isotropic

- $\Sigma = V$ symmetric PSD matrix

$$X = \Sigma^{1/2} Z \sim N(0, I),$$

density

$$f_X(x) = \frac{1}{(2\pi)^{n/2} |\Sigma|^{1/2}} \exp \left( -\frac{1}{2} x^T \Sigma^{-1} x \right).$$

(for $n = 1$, this is $N(0, \sigma^2)$)

Prop $X$ is isotropic iff

$$E\langle X, x \rangle^2 = \|x\|^2 \quad \text{for all} \quad x \in \mathbb{R}^n.$$
Lemma: Let \( X \) be an isotropic r.v. in \( \mathbb{R}^n \). Then
\[
E \|X\|^2 = n \quad (\Rightarrow \|X\| = \sqrt{n})
\]

Let \( Y \) be an independent isotropic r.v. in \( \mathbb{R}^n \). Then
\[
E \langle X, Y \rangle^2 = n \quad (\Rightarrow \langle \frac{X}{\|X\|}, \frac{Y}{\|Y\|} \rangle = \frac{1}{\sqrt{n}})
\]

Proof:
1) \( E \|X\|^2 = E \text{tr} XX^T = \text{tr} EXX^T = \text{tr} I = n \)

2) \( E \langle X, Y \rangle^2 = E_{\tilde{Y}} \left[ E_X \langle X, Y \rangle^2 | \tilde{Y} \right] \) (condition on \( Y \))
\[
= E_{\tilde{Y}} \|\tilde{Y}\|^2 \quad \text{(by Prop. p 41)}
\]
\[
= n \quad \text{(by last part 1). Q.E.D.}
\]

HW: Prove that there exist exponentially many pairwise almost orthogonal vectors in \( \mathbb{R}^n \).
Precisely, prove that there exists a set \( \{X_1, X_2, \ldots, X_N\} \subseteq S^n \), \( N \geq \exp(cn) \) such that \( \langle X_i, X_j \rangle < 0.1 \) for all \( i \neq j \).
(The constant 0.1 can be replaced by any other positive constant.)
(Hint: construct \( X_i \) one by one, each time removing a spherical cap centered at \( X_i \).)

1) Gaussian: \( X \sim N(0, I_n) \) is isotropic.
2) More generally, product distribution:
If \( X = (X_1, \ldots, X_n) \), \( X_i \) are independent r.v.'s with
\[
EX_i = 0, \quad \text{Var}(X_i) = 1
\]
then \( X \) is isotropic (since \( \text{Cov}(X_i, X_j) = \delta_{ij} \))

3) Bernoulli: \( X = (X_1, \ldots, X_n) \), \( X_i \) are i.i.d. symmetric Bernoulli \( (\pm 1) \)
\( X \) is isotropic.
(4) Coordinate r.v. \( X \sim \text{Unif}\{ \pm \sqrt{n} e_i \}; i = 1, \ldots, n \). 

(Indeed, \( E(X, x)^2 = \frac{1}{n} \sum_{i=1}^{n} \langle \sqrt{n} e_i, x \rangle^2 = \frac{1}{n} \sum_{i=1}^{n} x_i^2 = \|x\|^2 \)).

Same for \( X \sim \text{Unif}\{ \pm \sqrt{n} e_i \}; i = 1, \ldots, n \).

Remark: Gaussian is the most continuous distribution, coordinate is the most discrete.

(5) Frame:

Def. A frame is a set of vectors \( \{u_i\}_{i=1}^{N} \) in \( \mathbb{R}^n \) which obeys Parseval's identity (exactly or approximately):

\[ \exists A, B > 0 \quad (\text{called the frame bounds}) \quad s.t. \]

\[ A \|x\|^2 \leq \sum_{i=1}^{N} \langle u_i, x \rangle^2 \leq B \|x\|^2 \quad \text{for all } x \in \mathbb{R}^n. \]

If \( A = B \), the set is called a tight frame.

Examples:
(a) Any orthonormal basis in \( \mathbb{R}^n \), with \( A = B = 1 \).
(b) The Mercedes-Benz frame in \( \mathbb{R}^2 \), \( \begin{array}{c} 1 \\ 0 \end{array} \) with \( A = B \) (linearly dependent).
(c) (Geometric representation):

\[ \text{The set } \{P e_i\}_{i=1}^{N} \text{ is a tight frame with } A = B = 1, \]

where \( P \) is any orthogonal proj. in \( \mathbb{R}^N \).

Vice versa, every tight frame can be realized this way (up to isometry).

(d) (Probabilistic interpretation):

Let \( X \sim \text{Unif}\{ x_i : i = 1, \ldots, N \} \). \( X \) is isotropic iff \( \{x_i\}_{i=1}^{N} \) is a tight frame with \( A = B = 1 \).

\[ \left(\text{Indeed, } \|x\|^2 = E(X, X)^2 = \frac{1}{N} \sum_{i=1}^{N} \langle x_i, x \rangle^2 \right). \]
(6) \textbf{Spherical.} \quad X \sim \text{Unif} \left( \frac{S^n}{n} \right)

sphere centered at 0 and with radius \( \frac{1}{n} \)

Then \( X \) is isotropic. Indeed, by rotation invariance, \( \exists \, \lambda \).

\[
\mathbb{E} \langle X, x \rangle^2 = \lambda \|x\|_2^2 \quad \forall x
\]

\( \Rightarrow \) \( X/\lambda \) is isotropic \( \Rightarrow \) by lemma p. 42,

\[
\mathbb{E} \|X/\lambda\|_2^2 = n \quad \Rightarrow \quad \mathbb{E} \|X\|_2^2 = 2n
\]

\( \frac{n}{n} \) since \( x \in \frac{1}{n} S^n \)

\( \Rightarrow \lambda = 1 \) \quad \text{QED.}

(7) \textbf{Uniform on a convex set.} \quad X \sim \text{Unif} (K).

By remarks below def. of isotropic, \( \exists \) affine transformation \( T \).

\( Z \sim \text{Unif}(TK) \) is isotropic. \textbf{Say:} "TK is isotropic."

Isotropic convex sets are "well-conditioned", which is useful

for numerical algorithms (e.g., volume computation)

\textit{SEMINAR THIS WEEK.}

\[\text{Def (Sub-gauss.)} \quad \text{A r.v. } X \text{ in } \mathbb{R}^n \text{ is sub-gaussian if } \forall \text{ one-dimensional}
\]

\[\text{marginals } \langle X, x \rangle, x \in \mathbb{R}^n \text{, are sub-gaussian. The sub-gaussian norm:}
\]

\[
\|X\|_{\psi_2} := \sup_{\|x\|_2 = 1} \|\langle X, x \rangle\|_{\psi_2}
\]

Similar def. for sub-exponential.

\[\text{Not necessarily indep. coordinates.}
\]

\[\text{(marginals vs. coordinates)} \quad \text{let } X = (X_1, \ldots, X_n) \text{ be a r.v.}
\]

(1) \textbf{(Qualitative):} If all \( X_i \) are sub-gaussian r.v.'s

then \( X \) is a sub-gaussian r-vector. \quad \textbf{(Prove)}.

(2) \textbf{(Quantitative).} However, it is possible that

\[
\|X\|_{\psi_2} \gg \max_i \|X_i\|_{\psi_2}
\]

\textbf{(Find an example).}
If the coord's $X_i$ are independent, then there are no counterexamples:

Prop  Let $X=(X_1, \ldots, X_n)$, with $X_i$ independent sub-gauss r.v.'s. Then

$$
\|X\|_{\psi_2} \leq C \max_i \|X_i\|_{\psi_2}.
$$

Proof  This is a direct consequence of rotation invariance.

\[ \forall x=(x_1, \ldots, x_n), \quad \|x\|_{\psi_2} = 1: \]

\[
\|\langle X, x \rangle\|_{\psi_2} = \|\sum_{i=1}^n x_i X_i\|_{\psi_2} \leq C \sum_{i=1}^n \|x_i X_i\|_{\psi_2} \leq C \max_i \|x_i X_i\|_{\psi_2} \leq C \max_i \|x_i\|_{\psi_2} \|X_i\|_{\psi_2} \frac{\sum_{i=1}^n x_i^2}{\|x\|_{\psi_2}^2}.
\]

QED.

Examples:

1. Gaussian, Bernoulli: r.v. vectors are sub-gaussian by Prop. above.

2. Spherical: $X \sim \text{Unif}(\text{on } S^{n-1})$ is sub-gaussian,

\[
\|X\|_{\psi_2} \leq C.
\]

This follows from concentration of measure (Cor. on p. 30).

Indeed, $\forall x, \|x\|_{\psi_2} = 1, \quad f(x) = \langle X, x \rangle$ is a Lipschitz function on $\mathbb{R}^n$ with $\|f\|_{\text{Lip}} = 1 \Rightarrow \|f(x) - Ef(x)\|_{\psi_2} \leq C \|f\|_{\text{Lip}}$. QED.

(note the normalization: $X \in \sqrt{n} S^{n-1}$)

3. Coordinate: $X \sim \text{Unif}\{\pm 1, \ldots, \pm 1\}$: While (qualitatively) $X$ is sub-gaussian (since $\|X\|_{\psi_2} = \sqrt{n}$, $o(n)$), quantitatively $X$ is not a good sub-gaussian vector: $\|X\|_{\psi_2} \leq 1$ (indeed, $\mathbb{P}(\|X\|_{\psi_2} = 1) \approx \exp(-Cn^2)$ is not a sub-gaussian tail).

4. Uniform on a convex set: $X \sim \text{Unif}(K)$, isotropic

- For many sets $K$ (called $\psi_2$-bodies), $X$ is sub-gaussian.

Examples:

(a) For cube $K=[-1,1]^n$, this follows from Prop. above.

(b) For ball $K=\text{Ball}(\ell_2^n)$, this follows from concentration on the ball (Example 8p. 38) in a way similar to the sphere,

see Example 2 above.