High-Dimensional Statistics

Vitaliy Pozdnyakov

March 2020

Solutions to exercises from the book ”High-Dimensional Probability”, Roman Vershynin, 2020
https://www.math.uci.edu/~rvershyn/

Notation:

- cc – coffee cup
- $\sum := \sum_{i=1}^{N}$
- $\prod := \prod_{i=1}^{N}$

Contents

1 Exercise 2.1.4 (1 cc)

Let $g \sim \mathcal{N}(0,1)$. Show that for all $t \geq 1$, we have

$$\mathbb{E}g^2\mathbf{1}\{g > t\} = t\frac{1}{\sqrt{2\pi}}e^{-t^2/2} + \mathbb{P}\{g > t\} \leq \left(t + \frac{1}{t}\right)\frac{1}{\sqrt{2\pi}}e^{-t^2/2}$$

1.1 Solution

$$\mathbb{E}g^2\mathbf{1}\{g > t\} = \int_{-\infty}^{+\infty} x^2\mathbf{1}\{x > t\}\frac{1}{\sqrt{2\pi}}e^{-x^2/2}dx = \frac{1}{\sqrt{2\pi}}\int_t^{+\infty} x^2 e^{-x^2/2}dx$$

Make the substitution $u = x$ and $dv = xe^{-x^2/2}dx$, then $du = dx$ and

$$v = \int dv = \int xe^{-x^2/2}dx = \left[z = -\frac{x^2}{2}; dz = -xdx\right] = -\int e^z dz = -e^z = -e^{-x^2/2}$$

Next, integrate by parts

$$\int_a^b udv = uv\Big|_a^b - \int_a^b vdu$$

$$\frac{1}{\sqrt{2\pi}}\int_t^{+\infty} x^2 e^{-x^2/2}dx = \frac{1}{\sqrt{2\pi}}\left(-xe^{-x^2/2}\Big|_t^{+\infty}\right) + \int_t^{+\infty}\frac{1}{\sqrt{2\pi}}e^{-x^2/2}dx = t\frac{1}{\sqrt{2\pi}}e^{-t^2/2} + \mathbb{P}\{g > t\}$$

Next, apply the theorem "Tails of the normal distribution":

$$\mathbb{P}\{g > t\} \leq \frac{1}{t}\cdot\frac{1}{\sqrt{2\pi}}e^{t^2/2}$$

Thereby,

$$t\frac{1}{\sqrt{2\pi}}e^{-t^2/2} + \mathbb{P}\{g > t\} \leq t\frac{1}{\sqrt{2\pi}}e^{-t^2/2} + \frac{1}{t}\cdot\frac{1}{\sqrt{2\pi}}e^{t^2/2} = \left(t + \frac{1}{t}\right)\frac{1}{\sqrt{2\pi}}e^{-t^2/2}$$

2 Exercise 2.2.3 (1 cc)

(Bounding the hyperbolic cosine)
Show that

$$\cosh(x) \leq \exp(x^2/2) \text{ for all } x \in \mathbb{R}.$$

2.1 Solution

Apply Taylor's expansion for the left side of inequation

$$\cosh(x)' = \sinh(x);\ \cosh(x)'' = \cosh(x);\ \cosh(x)''' = \sinh(x);\ \dots$$

$$\cosh(0) = 1;\ \sinh(0) = 0$$

Then

$$\begin{aligned}\cosh(x) &= \cosh(0) + \frac{\sinh(0)}{1!}x^1 + \frac{\cosh(0)}{2!}x^2 + \frac{\sinh(0)}{3!}x^3 + \dots \\ &= 1 + \frac{x^2}{2} + \frac{x^4}{24} + \frac{x^6}{720} + O(x^8)\end{aligned}$$

For the right side

$$\exp(x^2/2)' = \exp(x^2/2)x;\ \exp(x^2/2)'' = \exp(x^2/2)(x^2+1);\ \exp(x^2/2)''' = \exp(x^2/2)x(x^2+3);\ \dots$$

$$\exp(0^2/2) = 1;\ \exp(0^2/2)0 = 0;\ \exp(0^2/2)(0^2+1) = 1;\ \exp(0^2/2)0(0^2+3) = 0;\ \dots$$

Then

$$\begin{aligned}\exp(x^2/2) &= \exp(0^2/2) + \frac{\exp(0^2/2)0}{1!}x^1 + \frac{\exp(0^2/2)(0^2+1)}{2!}x^2 + \frac{\exp(0^2/2)0(0^2+3)}{3!}x^3 + \dots \\ &= 1 + \frac{x^2}{2} + \frac{x^4}{8} + \frac{x^6}{384} + O(x^8)\end{aligned}$$

Obviously,

$$1 + \frac{x^2}{2} + \frac{x^4}{24} + \frac{x^6}{720} + O(x^8) \leq 1 + \frac{x^2}{2} + \frac{x^4}{8} + \frac{x^6}{384} + O(x^8)$$

Thus,

$$\cosh(x) \leq \exp(x^2/2) \text{ for all } x \in \mathbb{R}.$$

3 Exercise 2.2.7. (2 cc)

Let $X_1, ..., X_N$ be independent random variables. Assume that $X_i \in [m_i, M_i]$ for every i. Prove that for any $t > 0$

$$\mathbb{P}\left\{\sum(X_i - \mathbb{E}X_i) \geq t\right\} \leq \exp\left(-\frac{2t^2}{\sum(M_i - m_i)^2}\right)$$

3.1 Solution

Let $\xi_i = X_i - \mathbb{E}X_i$. Apply Markov's inequality and a property of the exp function

$$\mathbb{P}\left\{\sum \xi_i \geq t\right\} = \mathbb{P}\left\{e^{\lambda \sum \xi_i} \geq e^{\lambda t}\right\} \leq e^{-\lambda t}\mathbb{E}\left[e^{\lambda \sum \xi_i}\right] \text{ for any } \lambda > 0$$

Apply independence assumption

$$\mathbb{E}\left[e^{\lambda \sum \xi_i}\right] = \mathbb{E}\left[\prod e^{\lambda \xi_i}\right] = \prod \mathbb{E}\left[e^{\lambda \xi_i}\right]$$

Apply Hoeffding's Lemma

$$\mathbb{E}\left[e^{\lambda \xi_i}\right] \leq e^{\frac{\lambda^2 (M_i - m_i)^2}{8}}$$

Thus,

$$\mathbb{P}\left\{\sum \xi_i \geq t\right\} \leq e^{-\lambda t} \prod e^{\frac{\lambda^2 (M_i - m_i)^2}{8}} = e^{-\lambda t + \frac{\lambda^2}{8}\sum (M_i - m_i)^2}$$

Let $\lambda = 4t / \sum (M_i - m_i)^2$, then

$$\mathbb{P}\left\{\sum (X_i - \mathbb{E}X_i) \geq t\right\} \leq \exp\left(-\frac{2t^2}{\sum (M_i - m_i)^2}\right)$$

4 Exercise 2.2.8. (2 cc)

(Boosting randomized algorithms)
Imagine we have an algorithm for solving some decision problem (e.g. is a given number p a prime?). Suppose the algorithm makes a decision at random and returns the correct answer with probability $1/2 + \delta$ with some $\delta > 0$, which is just a bit better than a random guess. To improve the performance, we run the algorithm N times and take the majority vote. Show that, for any $\epsilon \in (0, 1)$, the answer is correct with probability at least $1 - \epsilon$, as long as

$$N \leq \frac{1}{2\delta^2} \ln(1/\epsilon)$$

4.1 Solution

Let $X_1, X_2, \ldots X_N$ be a random sequence with probabilities

$$\mathbb{P}\{X_i = 0\} = 1/2 - \delta$$

$$\mathbb{P}\{X_i = 1\} = 1/2 + \delta$$

Then the event {The answer is incorrect} can be written as

$$\begin{aligned}\left\{\sum X_i \geq N/2\right\} &= \left\{\sum (X_i - \mathbb{E}X_i) \geq \sum (1/2 - \mathbb{E}X_i)\right\} \\ &= \left\{\sum (X_i - \mathbb{E}X_i) \geq \sum (1/2 - 1/2 + \delta)\right\} \\ &= \left\{\sum (X_i - \mathbb{E}X_i) \geq \delta N\right\}\end{aligned}$$

Apply Hoeffding inequality

$$\mathbb{P}\left\{\sum (X_i - \mathbb{E}X_i) \geq t\right\} \leq \exp\left(-\frac{2t^2}{\sum (M_i - m_i)^2}\right)$$

Make substitutions $t = \delta N$, $M_i = 1$, $m_i = 0$

$$\mathbb{P}\left\{\sum (X_i - \mathbb{E}X_i) \geq \delta N\right\} \leq \exp(-2\delta^2 N)$$

We can observe that

$$\left\{N \leq \frac{1}{2\delta^2}\ln(1/\epsilon)\right\} \Rightarrow \{\exp(-2\delta^2 N) < \epsilon\}$$

Thereby, $\mathbb{P}\{\text{The answer is incorrect}\} \leq \epsilon$ and $\mathbb{P}\{\text{The answer is correct}\} \geq 1 - \epsilon$ under that constraint.

5 Exercise 2.2.10 (2 cc)

Let $X_1, \ldots, X_N$ be non-negative independent random variables with continuous distributions. Assume that the densities of X_i are uniformly bounded by 1.
(a) Show that the MGF of X_i satisfies $\mathbb{E}e^{-tX_i} \leq 1/t$ for all $t > 0$.
(b) Deduce that, for any $\epsilon > 0$, we have $\mathbb{P}\{\sum X_i \leq \epsilon N\} \leq (e\epsilon)^N$.

5.1 Solution (a)

Let $\phi_i(x)$ be a PDF of X_i

$$\mathbb{E}\exp(-tX_i) = \int_0^\infty e^{-tx}\phi_i(x)dx$$

Since $\phi_i(x) \leq 1$ for all i and x, then

$$\int_0^\infty e^{-tx}\phi_i(x)dx \leq \int_0^\infty e^{-tx}dx = \frac{1}{t} \text{ for all } t > 0$$

5.2 Solution (b)

Rewrite the probability as

$$\mathbb{P}\left\{\sum X_i \leq \epsilon N\right\} = \mathbb{P}\left\{\sum -X_i/\epsilon \geq -N\right\}$$

Apply Markov's inequality and a property of the exp function

$$\mathbb{P}\left\{\sum -X_i/\epsilon \geq -N\right\} = \mathbb{P}\left\{\exp\left(\sum -X_i/\epsilon\right) \geq \exp(-N)\right\} \leq e^N \prod \mathbb{E}e^{-X_i/\epsilon}$$

Apply the result from the section (a) and then

$$\mathbb{P}\left\{\sum X_i \leq \epsilon N\right\} \leq e^N \prod \epsilon = (e\epsilon)^N$$

6 Exercise 2.3.2 (2 cc)

(Chernoff's inequality: lower tails)
Modify the proof of Theorem 2.3.1 to obtain the following bound on the lower tail. For any $t < \mu$, we have

$$\mathbb{P}\{S_N \leq t\} \leq e^{-\mu}\left(\frac{e\mu}{t}\right)^t$$

6.1 Solution

Since $t < \mu$, $t > 0$ and $S_N > 0$ we can rewrite the probability as

$$\mathbb{P}\{S_N \leq t\} = \mathbb{P}\{(t/\mu)^{S_N} \geq (t/\mu)^t\}$$

Apply Markov's inequality

$$\mathbb{P}\{(t/\mu)^{S_N} \geq (t/\mu)^t\} \leq (\mu/t)^t \mathbb{E}(t/\mu)^{S_N}$$

Consider the expectation

$$\mathbb{E}(t/\mu)^{S_N} = \prod \mathbb{E}(t/\mu)^{X_i}$$

$$\mathbb{E}(t/\mu)^{X_i} = (t/\mu)^1 p_i + (t/\mu)^0(1 - p_i) = 1 + [(t/\mu) - 1]p_i \leq \exp\left(\left[\frac{t}{\mu} - 1\right] p_i\right)$$

$$\prod \mathbb{E}(t/\mu)^{X_i} \leq \exp\left(\left[\frac{t}{\mu} - 1\right] \sum p_i\right) = \exp\left(\left[\frac{t}{\mu} - 1\right] \mu\right) = e^t e^{-\mu}$$

Thereby

$$\mathbb{P}\{S_N \leq t\} \leq e^{-\mu} \left(\frac{e\mu}{t}\right)^t$$

7 Exercise 2.3.3 (2 cc)

(Poisson tails)
Let $X \sim Pois(\lambda)$. Show that for any $t > \lambda$, we have

$$\mathbb{P}\{X \geq t\} \leq e^{-\lambda} \left(\frac{e\lambda}{t}\right)^t$$

7.1 Solution

Chernoff's inequality

$$\mathbb{P}\{S_N \geq t\} \leq e^{-\mu} \left(\frac{e\mu}{t}\right)^t$$

Assume that $N \to \infty$, $\max p_i \to 0$ and $\mathbb{E}S_N \to \lambda < \infty$. Then, using Poisson limit theorem we have

$$S_N \to Pois(\lambda)$$

Thereby

$$\mathbb{P}\{X \geq t\} \leq e^{-\lambda} \left(\frac{e\lambda}{t}\right)^t$$

8 Exercise 2.3.5 (3 cc)

(Chernoff's inequality: small deviations)
Show that, in the setting of Theorem 2.3.1, for $\delta \in (0, 1]$ we have

$$\mathbb{P}\{|S_N - \mu| \geq \delta\mu\} \leq 2e^{-c\mu\delta^2}$$

where $c > 0$ is an absolute constant.

8.1 Solution

Consider Chernoff's inequality with $t > \mu$ (upper tails). Let $t = \mu + \delta\mu$, so

$$\mathbb{P}\{S_N - \mu \geq \delta\mu\} \leq e^{-\mu}\left(\frac{e}{1+\delta}\right)^{\mu+\delta\mu}$$

In opposite, if $t < \mu$ and $t = \mu - \delta\mu$ (lower tails) we obtain

$$\mathbb{P}\{S_N - \mu \leq -\delta\mu\} \leq e^{-\mu}\left(\frac{e}{1-\delta}\right)^{\mu-\delta\mu}$$

Apply addition theorem of probability for lower and upper tails

$$\mathbb{P}\left(\{S_N - \mu \geq \delta\mu\} \cup \{S_N - \mu \leq -\delta\mu\}\right) = \mathbb{P}\{|S_N - \mu| \geq \delta\mu\}$$

$$\mathbb{P}\{|S_N - \mu| \geq \delta\mu\} \leq e^{-\mu}\left(\frac{e}{1-\delta}\right)^{\mu-\delta\mu} + e^{-\mu}\left(\frac{e}{1+\delta}\right)^{\mu+\delta\mu} \leq 2e^{-\mu}\left(\frac{e}{1+\delta}\right)^{\mu+\delta\mu}$$

Consider the inequality

$$\begin{aligned} 2e^{-\mu}\left(\frac{e}{1+\delta}\right)^{\mu+\delta\mu} &\leq 2e^{-c\mu\delta^2} \\ \frac{e^{\delta\mu}}{(1+\delta)^{\mu+\delta\mu}} &\leq e^{-c\mu\delta^2} \\ e^{\delta+c\delta^2} &\leq (1+\delta)^{1+\delta} \end{aligned}$$

Take log

$$\begin{aligned} \delta + c\delta^2 &\leq (1+\delta)\ln(1+\delta) \\ c &\leq \frac{(1+\delta)\ln(1+\delta) - \delta}{\delta^2} \end{aligned}$$

Thereby,

$$\mathbb{P}\{|S_N - \mu| \geq \delta\mu\} \leq 2e^{-c\mu\delta^2}$$

for all $c \leq [(1+\delta)\ln(1+\delta) - \delta]/\delta^2$.

9 Exercise 2.3.8 (2 cc)

(Normal approximation to Poisson)
Let $X \sim Pois(\lambda)$. Show that, as $\lambda \to \infty$, we have

$$\frac{X-\lambda}{\sqrt{\lambda}} \to \mathcal{N}(0,1) \text{ in distribution.}$$

9.1 Solution

Rewrite the RV X as a sum $Z_1 + \cdots + Z_N$ where $Z_i \sim Pois(\lambda/N)$ and apply Lindeberg-Levy central limit theorem

$$\frac{X - \mathbb{E}X}{\sqrt{\mathrm{Var}(X)}} = \frac{X - N\mathbb{E}Z_i}{\sqrt{N\,\mathrm{Var}(Z_i)}} = \frac{X-\lambda}{\sqrt{\lambda}} \to \mathcal{N}(0,1) \text{ as } N \to \infty$$

10 Exercise 2.4.2 (1 cc)

(Bounding the degrees of sparse graphs)
Consider a random graph $G \sim G(n, p)$ with expected degrees $d = O(\log n)$. Show that with high probability (say, 0.9), all vertices of G have degrees $O(\log n)$.

10.1 Solution

Since $d_i \geq 0$ and $n > 0$, then for any i we have $d_i = O(\log n) \equiv \exists C_i : d_i \leq C_i \log n$ and we need to prove the following statement

$$\mathbb{P}\{\forall i : d_i \leq C_i \log n\} \geq 0.9$$

where C_i is a constant. We can rewrite probability of the opposite case as

$$\mathbb{P}\{\exists i : d_i \geq C_i \log n\} \leq \sum \mathbb{P}\{d_i \geq C_i \log n\} \leq 0.1$$

Apply Chernoff's inequality for upper bound (since $d \leq C_i \log n$)

$$\sum \mathbb{P}\{d_i \geq C_i \log n\} \leq ne^{-d} \left(\frac{ed}{C_i \log n} \right)^{C_i \log n} \leq 0.1$$

Note that there exists C_d such that $d \leq C_d \log n$ and then

$$ne^{-C_d \log n} \left(\frac{eC_d \log n}{C_i \log n} \right)^{C_i \log n} = n^{C_i - C_d + 1} \left(\frac{C_d}{C_i} \right)^{C_i \log n} \leq 0.1$$

Hence, the statement holds for any C_i such that

$$C_i \geq \frac{-C_d \log n + \log n + \log 10}{\log n} \left[W \left(\frac{-C_d \log n + \log n + \log 10}{C_d e \log n} \right) \right]^{-1}$$

where $W(\cdot)$ is the product log function.

11 Exercise 2.4.3 (2 cc)

(Bounding the degrees of very sparse graphs)
Consider a random graph $G \sim G(n, p)$ with expected degrees $d = O(1)$. Show that with high probability (say, 0.9), all vertices of G have degrees

$$O\left(\frac{\log n}{\log \log n} \right)$$

11.1 Solution

Apply the same approach as in the previous exercise and obtain

$$\sum \mathbb{P}\left\{ d_i \geq C_i \frac{\log n}{\log \log n} \right\} \leq ne^{-C_d} \left(\frac{eC_d \log \log n}{C_i \log n} \right)^{(C_i \log n)/(\log \log n)} \leq 0.1$$

Hence, the statement holds for any C_i such that

$$C_i \geq \frac{(\log n - C_d + \log 10) \log \log n}{\log n} \left[W \left(\frac{\log n - C_d + \log 10}{C_d e} \right) \right]^{-1}$$

where $W(\cdot)$ is the product log function.

12 Exercise 3.1.4 (3 cc)

(Expectation of the norm)
(a) Deduce from Theorem 3.1.1 that $\sqrt{n} - CK^2 \le \mathbb{E}\|X\|_2 \le \sqrt{n} + CK^2$
(b) Can CK^2 be replaced by $o(1)$, a quantity that vanishes as $n \to \infty$?

12.1 Solution (a)

$$
\begin{aligned}
|\mathbb{E}\|X\|_2 - \sqrt{n}| &= \left|\mathbb{E}\left[\|X\|_2 - \sqrt{n}\right]\right| \\
&\le \mathbb{E}\left|\|X\|_2 - \sqrt{n}\right| && \text{(by Jensen's inequality)} \\
&= \left\|\|X\|_2 - \sqrt{n}\right\|_{L^1} && \text{(by the definition of } L^p \text{ norm)} \\
&\le C_1 \left\|\|X\|_2 - \sqrt{n}\right\|_{\psi^2} \sqrt{1} && \text{(using (2.15) with } p = 1) \\
&\le C_2 K^2 && \text{(by Theorem 3.1.1)}
\end{aligned}
$$

12.2 Solution (b)

Let $Y = \frac{1}{n}\sum X_i^2$ then $\mathbb{E}Y = 1$ and $\operatorname{Var} Y = \sigma^2/n$. Consider the inequality

$$
\frac{1 + x - (x-1)^2}{2} \le \sqrt{x} \le \frac{1+x}{2}
$$

$$
\mathbb{E}\frac{1 + Y - (Y-1)^2}{2} \le \mathbb{E}\sqrt{Y} \le \mathbb{E}\frac{1+Y}{2}
$$

$$
1 + \frac{\sigma^2}{2n} \le \mathbb{E}\sqrt{Y} \le 1
$$

Thereby

$$
\lim_{n\to\infty} \frac{\mathbb{E}\|X\|_2}{\sqrt{n}} = 1
$$

13 Exercise 3.2.2 (1 cc)

(Reduction to isotropy)
(a) Let Z be a mean zero, isotropic random vector in $\mathbb{R}^n$. Let $\mu \in \mathbb{R}^n$ be a fixed vector and Σ be a fixed $n \times n$ positive-semidefinite matrix. Check that the random vector $X := \mu + \Sigma^{1/2} Z$ has mean μ and covariance matrix $\operatorname{cov}(X) = \Sigma$.
(b) Let X be a random vector with mean μ and invertible covariance matrix $\Sigma = \operatorname{cov}(X)$. Check that the random vector $Z := \Sigma^{-1/2}(X - \mu)$ is an isotropic, mean zero random vector.

13.1 Solution (a)

$$
\mathbb{E}X = \mathbb{E}\left[\mu + \Sigma^{1/2} Z\right] = \mathbb{E}\mu + \Sigma^{1/2}\mathbb{E}Z = \mu
$$

$$
\begin{aligned}
\operatorname{cov}(X) &= \mathbb{E}\left[(X-\mu)(X-\mu)^T\right] \\
&= \mathbb{E}\left[\left(\Sigma^{1/2}Z\right)\left(\Sigma^{1/2}Z\right)^T\right] \\
&= \mathbb{E}\left[\Sigma^{1/2}ZZ^T\left(\Sigma^{1/2}\right)^T\right] \\
&= \mathbb{E}\left[\Sigma^{1/2}\left(\Sigma^{1/2}\right)^T\right] \\
&= \mathbb{E}\left[\Sigma^{1/2}\Sigma^{1/2}\right] = \Sigma
\end{aligned}
$$

13.2 Solution (b)

$$
\mathbb{E}Z = \mathbb{E}\left[\Sigma^{-1/2}(X-\mu)\right] = \Sigma^{-1/2}\left[\mathbb{E}X-\mu\right] = 0
$$

$$
\begin{aligned}
\operatorname{cov}(Z) = \mathbb{E}ZZ^T &= \mathbb{E}\left[\Sigma^{-1/2}(X-\mu)(X-\mu)^T\left(\Sigma^{-1/2}\right)^T\right] \\
&= \Sigma^{-1/2}\mathbb{E}\left[(X-\mu)(X-\mu)^T\right]\left(\Sigma^{-1/2}\right)^T \\
&= \Sigma^{-1/2}\operatorname{cov}(X)\left(\Sigma^{-1/2}\right)^T \\
&= \Sigma^{-1/2}\Sigma\left(\Sigma^{-1/2}\right)^T \\
&= \Sigma^{-1/2}\Sigma^{-1/2}\Sigma = I_n
\end{aligned}
$$

14 Exercise 3.2.6 (1 cc)

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

$$
\mathbb{E}\|X-Y\|_2^2 = 2n
$$

14.1 Solution

$$
\mathbb{E}\|X-Y\|_2^2 = \mathbb{E}\left[\sum(X_i-Y_i)^2\right] = \sum(\mathbb{E}X_i^2 - \mathbb{E}X_iY_i + \mathbb{E}Y_i^2)
$$

$\mathbb{E}X_i^2 = \mathbb{E}Y_i^2 = 1$ by isotropy, $\mathbb{E}X_iY_i = \mathbb{E}X_i\mathbb{E}Y_i$ by independence and $\mathbb{E}X_i = \mathbb{E}Y_i = 0$ by task description. Thereby,

$$
\mathbb{E}\|X-Y\|_2^2 = \sum(1-0+1) = 2n
$$

15 Exercise 3.3.3 (2 cc)

(Rotation invariance)
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 $\langle g, u\rangle \sim \mathcal{N}(0, \|u\|_2^2)$
(b) Consider independent random variables $X_i \sim \mathcal{N}(0, \sigma_i^2)$. Then $\sum X_i \sim \mathcal{N}(0, \sigma^2)$ where $\sigma^2 = \sum \sigma_i^2$
(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 $Gu \sim \mathcal{N}(0, I_m)$.

15.1 Solution (a)

Let Q be an orthogonal matrix such that $Q(u/\|u\|_2) = [1, 0, \dots, 0]^T$. Next, use the property of the orthogonal matrix

$$\langle u, g\rangle = \langle Qg, Qu\rangle = \|u\|_2 \left\langle Qg, [1, 0, \dots, 0]^T \right\rangle$$

By the rotation invariance we obtain

$$Qg \sim \mathcal{N}(0, I_n) \Rightarrow \left\langle Qg, [1, 0, \dots, 0]^T \right\rangle \sim \mathcal{N}(0, 1)$$

and finally use the property of Gaussian distribution

$$\|u\|_2 \mathcal{N}(0, 1) = \mathcal{N}(0, \|u\|_2^2)$$

15.2 Solution (b)

Consider the PDF $f_Z(z)$ where $Z = X + Y$ with independent $X \sim \mathcal{N}(0, \sigma_X^2)$ and $Y \sim \mathcal{N}(0, \sigma_Y^2)$

$$\begin{aligned} f_Z(z) &= \int_{-\infty}^{+\infty} f(z - x) f(x) dx \\ &= \int_{-\infty}^{+\infty} \frac{1}{\sigma_Y \sqrt{2\pi}} \exp\left[-\frac{(z - x)^2}{2\sigma_Y^2}\right] \frac{1}{\sigma_X \sqrt{2\pi}} \exp\left[-\frac{x^2}{2\sigma_X^2}\right] dx \\ &= [\dots] = \frac{1}{\sqrt{\sigma_X^2 + \sigma_Y^2}\sqrt{2\pi}} \exp\left[-\frac{z^2}{2(\sigma_X^2 + \sigma_Y^2)}\right] \end{aligned}$$

Thereby $Z \sim \mathcal{N}(0, \sigma_X^2 + \sigma_Y^2)$ and consequently $\sum X_i \sim \mathcal{N}(0, \sigma^2)$ where $\sigma^2 = \sum \sigma_i^2$

15.3 Solution (c)

Denote g_{ij} as an entry of the matrix G and v_i as an entry of the vector Gu

$$\begin{pmatrix} v_1 \\ v_2 \\ \dots \\ v_m \end{pmatrix} := Gu = \begin{pmatrix} g_{11} & g_{12} & \dots & g_{1n} \\ g_{21} & g_{22} & \dots & g_{2n} \\ \dots & \dots & \dots & \dots \\ g_{m1} & g_{m2} & \dots & g_{mn} \end{pmatrix} \begin{pmatrix} u_1 \\ u_2 \\ \dots \\ u_n \end{pmatrix}$$

Then $v_i = g_{i1}u_1 + g_{i2}u_2 + \dots + g_{in}u_n \sim \mathcal{N}(0, \|u\|_2^2)$ by solution (a) and (b). Since u is a unit vector we obtain that $v_i \sim \mathcal{N}(0, 1)$. Entries g_{ij} are independent, then entries v_i are independent too and consequently $\forall i \neq j : \mathrm{cov}(v_i, v_j) = 0$. Thereby $Gu \sim \mathcal{N}(0, I_n)$.

16 Exercise 3.3.6 (1 cc)

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 $u, v \in \mathbb{R}^n$ be unit orthogonal vectors. Prove that Gu and Gv are independent $\mathcal{N}(0, I_m)$ random vectors.

16.1 Solution

Let Q be an orthogonal transformation matrix such that $Qu = [1, 0, \dots, 0]^T$ and $Qv = [0, 1, 0, \dots, 0]^T$ and then

$$
\begin{aligned}
Gu &= GuQQ^T && \text{(since } Q \text{ is orthogonal)} \\
&= G[1, 0, \ldots, 0]^T Q^T \\
Gv &= GvQQ^T = G[0, 1, 0, \ldots, 0]^T Q^T
\end{aligned}
$$

We see that vectors Gu and Gv can be represent as sums of entries from different columns of G. If the entries of G are independent then Gu and Gv are independent too. Check the expectations and covariance matrix

$$
\mathbb{E}Gu = u\mathbb{E}G = u\mathbf{0} = \mathbf{0}
$$

$$
\mathbb{E}Gu(Gu)^T = \mathbb{E}Guu^T G^T = \mathbb{E}GG^T = I_n
$$

17 Exercise 3.3.9 (2 cc)

Show that $\{u_i\}_{i=1}^N$ is a tight frame in $\mathbb{R}^n$ with bound A if and only if $\sum u_i u_i^T = AI_n$.

17.1 Solution

Check that the left and right side of the equation $\sum u_i u_i^T = AI_n$ are hold the definition of a tight frame:

$$
A\|x\|_2^2 = \sum \langle u_i, x\rangle^2 \text{ for any } x \in \mathbb{R}^n
$$

First, the left side

$$
x^T AI_n x = A\|x\|_2^2
$$

Next, the right side

$$
x^T \left(\sum u_i u_i^T\right) x = \sum x^T u_i u_i^T x = \sum \left(u_i^T x\right)^T u_i^T x = \sum \langle u_i, x\rangle^2
$$

18 Exercise 4.1.1 (1 cc)

Suppose A is an invertible matrix with singular value decomposition $A = \sum s_i u_i v_i^T$. Check that $A^{-1} = \sum (1/s_i) v_i u_i^T$

18.1 Solution

Consider the matrix form of SVD

$$
A = U\Sigma V^T, A^{-1} = U\Sigma^{-1}V^T
$$

Use the fact that U and V are unitary matrix and check that matrix A is invertible indeed

$$
AA^{-1} = V\Sigma U^T U\Sigma^{-1}V^T = VI_nV^T = I_n
$$

19 Exercise 4.1.2 (2 cc)

Prove the following bound on the singular values s_i of any matrix A: $s_i \leq (1/\sqrt{i})\|A\|_F$

19.1 Solution

Use the definition of the Frobenius norm

$$
\begin{aligned}
\frac{1}{i}\|A\|_F^2 = \frac{1}{i}\sum_{j=1}^{n} s_j^2 = & \ \frac{1}{i}(s_1^2 + \cdots + s_i^2 + \cdots + s_n^2) && (\forall i : s_i \geq s_{i+1}) \\
\geq & \ \frac{1}{i}(s_i^2 + \cdots + s_i^2 + \cdots + s_n^2) && \text{(substite the first } i \text{ terms by } s_i) \\
= & \ \frac{1}{i}(is_i^2 + \cdots + s_n^2) \\
= & \ s_i^2 + \frac{1}{i}(s_{i+1}^2 + \ldots s_n^2) \\
\geq & \ s_i^2
\end{aligned}
$$

Thereby

$$
s_i^2 \leq \frac{1}{i}\|A\|_F^2 \Rightarrow s_i \leq \frac{1}{\sqrt{i}}\|A\|_F
$$

20 Exercise 4.1.3 (2 cc)

(Best rank k approximation)
Let A_k be the best rank k approximation of a matrix A. Express $\|A - A_k\|^2$ and $\|A - A_k\|_F^2$ in terms of the singular values s_i of A.

20.1 Solution

Use the Eckart-Young-Mirsky theorem and obtain

$$
\|A - A_k\|^2 = \|V[\Sigma(A) - \Sigma(A_k)]U^T\|^2 = \|\Sigma(A) - \Sigma(A_k)\|_\infty = s_{k+1}^2
$$

Use the property of the Frobenius norm $\|A\|_s = \|s\|_2$

$$
\|A - A_k\|_F^2 = \sum_{i=k+1}^{n} s_i^2
$$

21 Exercise 4.1.8 (1 cc)

(Isometries and projections from unitary matrices)
Canonical examples of isometries and projections can be constructed from a fixed unitary matrix U. Check that any sub-matrix of U obtained by selecting a subset of columns is an isometry, and any sub-matrix obtained by selecting a subset of rows is a projection.

21.1 Solution

Denote $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n$ as column-vectors in the matrix U. Let V be a matrix constructed from a fixed set of columns of U with indexes $\mathbf{i} = (i_1, i_2, \ldots, i_n)$

$$
V := (\mathbf{u}_{i_1}, \mathbf{u}_{i_2}, \ldots, \mathbf{u}_{i_n})
$$

Check an isometry by the property (c) from Exercise 4.1.4: $\|Ax\|_2 = \|x\|_2$ for all $x \in \mathbb{R}^n$

$$
\begin{aligned}
\|Vx\|_2^2 = \langle Vx, Vx\rangle &= \langle V^T V x, x\rangle \\
&= \langle I_n x, x\rangle \qquad \text{(since U is unitary)} \\
&= \|x\|_2^2
\end{aligned}
$$

Consequenly V is an isometry. V^T is a subset of rows of U and note that V is an isometry if and only if V^T is a projection.

22 Exercise 4.2.9 (3 cc)

(Allowing the centers to be outside K)
In our definition of the covering numbers of K, we required that the centers x_i of the balls $B(x_i, \varepsilon)$ that form a covering lie in K. Relaxing this condition, define the exterior covering number $\mathcal{N}^{ext}(K, d, \varepsilon)$ similarly but without requiring that $x_i \in K$. Prove that

$$
\mathcal{N}^{ext}(K, d, \varepsilon) \leq \mathcal{N}(K, d, \varepsilon) \leq N^{ext}(K, d, \varepsilon/2)
$$

22.1 Solution

The lower bound follow from the fact that $K \subset T$ where (T, d) is a metric space. Obviously an infimum by the subset can not be less than infimum by the whole set.
Next, let $x_1, \ldots, x_{\mathcal{N}}$ be centers of balls of the form $B(x_i, \varepsilon/2)$. By the definition of the covering, for each x_i there exists at least one $k_i \in K$ such that $k_i \in B(x_i, \varepsilon/2)$. Apply the trianle inequality

$$
B(x_i, \varepsilon/2) \subseteq B(k_i, \varepsilon)
$$

and consequently $B(k_i, \varepsilon), \ldots, B(k_{\mathcal{N}^{ext}}, \varepsilon)$ is an classical (not exterior) covering of size N^{ext}. Thereby

$$
\mathcal{N}(K, d, \varepsilon) \leq N^{ext}(K, d, \varepsilon/2)
$$

23 Exercise 4.4.2 (1 cc)

Let $x \in \mathbb{R}^n$ and $\mathcal{N}$ be an ε-net of the sphere S^{n-1}. Show that

$$
\sup_{y \in \mathcal{N}} \langle x, y\rangle \leq \|x\|_2 \leq \frac{1}{1-\varepsilon} \sup_{y \in \mathcal{N}} \langle x, y\rangle
$$

23.1 Solution

First, consider the lower bound. By the definition of the dot product we have

$$
\langle x, y\rangle = \|x\|_2 \|y\|_2 \cos\theta
$$

Use the fact that $y \in \mathcal{N} \subset S^{n-1}$, so $\|y\|_2 = 1$. Also $\cos\theta \leq 1$, consequently

$$
\sup_{y \in \mathcal{N}} \langle x, y\rangle \leq \sup_{y \in S^{n-1}} \langle x, y\rangle = \|x\|_2
$$

Next, use the similar approach as in Lemma 4.4.1 for the upper bound. Fix a vector $y \in S^{n-1}$ for which

$$
\langle x, y\rangle = \|x\|_2
$$

and choose $y_0 \in \mathcal{N}$ that approximates y so that

$$\|y - y_0\|_2 \le \varepsilon$$

By the definition of the dot product, this implies

$$\langle x, y - y_0 \rangle = \|x\|_2 \|y - y_0\|_2 \cos\theta \le \|x\|_2 \varepsilon \cos\theta \le \|x\|_2 \varepsilon$$

Using a distributive law, we find that

$$\langle x, y_0 \rangle = \langle x, y \rangle - \langle x, y - y_0 \rangle \ge \|x\|_2 - \|x\|_2 \varepsilon = (1 - \varepsilon)\|x\|_2$$

Dividing both sides of this inequality by $1 - \varepsilon$, we complete the proof.

24 Exercise 4.4.3 (2 cc)

(Quadratic form on a net)
Let A be an $m \times n$ matrix and $\varepsilon \in [0, 1/2)$.
(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

$$\sup_{x \in \mathcal{N}, y \in \mathcal{M}} \langle Ax, y \rangle \le \|A\| \le \frac{1}{1 - 2\varepsilon} \sup_{x \in \mathcal{N}, y \in \mathcal{M}} \langle Ax, y \rangle$$

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

$$\sup_{x \in \mathcal{N}} |\langle Ax, x \rangle| \le \|A\| \le \frac{1}{1 - 2\varepsilon} \sup_{x \in \mathcal{N}} |\langle Ax, x \rangle|$$

24.1 Solution (a)

First, consider the lower bound. Using the same explanations as in Exercise 4.4.2 and the lower bound from Lemma 4.4.1 we obtain

$$\|A\| \ge \sup_{x \in \mathcal{N}} \|Ax\|_2 \ge \sup_{x \in \mathcal{N}, y \in \mathcal{M}} \langle Ax, y \rangle$$

Next, again apply Lemma 4.4.1 and Exercise 4.4.2 for the upper bound

$$\begin{aligned}
\|A\| \le & \frac{1}{1 - \varepsilon} \sup_{x \in \mathcal{N}} \|Ax\|_2 \\
\le & \frac{1}{1 - \varepsilon} \sup_{x \in \mathcal{N}} \left(\frac{1}{1 - \varepsilon} \sup_{y \in \mathcal{M}} \langle Ax, y \rangle \right) = \quad \frac{1}{1 - 2\varepsilon + \varepsilon^2} \sup_{x \in \mathcal{N}, y \in \mathcal{M}} \langle Ax, y \rangle \\
\le & \frac{1}{1 - 2\varepsilon} \sup_{x \in \mathcal{N}, y \in \mathcal{M}} \langle Ax, y \rangle && \text{(since } \varepsilon \in [0, 1/2))
\end{aligned}$$

24.2 Solution (b)

First, use Lemma 4.4.1 and the Cauchy–Bunyakovsky–Schwarz inequality for the lower bound

$$\begin{aligned}
\|A\| &\ge \sup_{x \in \mathcal{N}} \|Ax\|_2 = \sup_{x \in \mathcal{N}} \|Ax\|_2 \|x\|_2 && \text{(by the fact that } \|x\|_2 = 1) \\
&\ge \sup_{x \in \mathcal{N}} |\langle Ax, x \rangle| && \text{(by Cauchy–Bunyakovsky–Schwarz inequality)}
\end{aligned}$$

Next, use the similar approach as in Lemma 4.4.1. Fix a vector $x \in S^{n-1}$ for which

$$\|Ax\|_2 = \|Ax\|_2 \|x\|_2 = |\langle Ax, x \rangle|$$

and choose $x_0 \in \mathcal{N}$ that approximates x so that

$$\|x - x_0\|_2 \leq \varepsilon$$

By the Cauchy–Bunyakovsky–Schwarz inequality, this implies

$$|\langle Ax, x - x_0\rangle| \leq \|Ax\|_2 \|x - x_0\|_2 \leq \|Ax\|_2 \varepsilon$$

Using the trianle inequality of absolute value, we find that

$$|\langle Ax, x_0\rangle| \geq |\langle Ax, x\rangle| - |\langle Ax, x - x_0\rangle| \geq \|Ax\|_2 - \|Ax\|_2\varepsilon = (1-\varepsilon)\|Ax\|_2$$

Dividing both sides of this inequality by $1-\varepsilon$, we prove that

$$\|Ax\|_2 \leq \frac{1}{1-\varepsilon} \sup_{x\in\mathcal{N}} |\langle Ax, x\rangle|$$

Finally, apply the upper bound of Lemma 4.4.1 and complete the proof

$$\begin{aligned}
\|A\| \leq & \frac{1}{1-\varepsilon} \sup_{x\in\mathcal{N}} \|Ax\|_2 \\
\leq & \frac{1}{1-\varepsilon} \sup_{x\in\mathcal{N}} \left(\frac{1}{1-\varepsilon} \sup_{x\in\mathcal{N}} |\langle Ax, x\rangle|\right) = \quad \frac{1}{1-2\varepsilon+\varepsilon^2} \sup_{x\in\mathcal{N}} |\langle Ax, x\rangle| \\
\leq & \frac{1}{1-2\varepsilon} \sup_{x\in\mathcal{N}} |\langle Ax, x\rangle| \qquad\qquad (\text{since } \varepsilon \in [0, 1/2))
\end{aligned}$$

25 Exercise 4.4.4 (3 cc)

(Deviation of the norm on a net)
Let A be an $m \times n$ matrix, $\mu \in \mathbb{R}$ and $\varepsilon \in [0, 1/2)$. Show that for any ε-net $\mathcal{N}$ of the sphere S^{n-1}, we have

$$\sup_{x\in S^{n-1}} |\|Ax\|_2 - \mu| \leq \frac{C}{1-2\varepsilon} \sup_{x\in\mathcal{N}} |\|Ax\|_2 - \mu|$$

25.1 Solution

Assume without loss of generality that $\mu = 1$. Rewrite $\|Ax\|_2^2 - 1$ in the quadratic form

$$\|Ax\|_2^2 - 1 = \langle Rx, x\rangle$$

where $R = A^T A - I_n$. Next, use the definition of the operator norm and the result from Exercise 4.4.3 (b)

$$\sup_{x\in S^{n-1}} |\langle Rx, x\rangle| = \|R\| \leq \frac{1}{1-2\varepsilon} \sup_{x\in\mathcal{N}} |\langle Rx, x\rangle|$$

Applying the elementary inequality

$$\max(|z-1|, |z-1|^2) \leq |z^2 - 1|, z > 0$$

for $z = \|Ax\|_2$, we conclude that

$$|\langle Rx, x\rangle| \leq |\|Ax\|_2 - 1|$$

And thereby

$$\sup_{x \in S^{n-1}} |\|Ax\|_2 - 1| \leq \frac{1}{1-2\varepsilon} \sup_{x \in \mathcal{N}} |\|Ax\|_2 - 1|$$

Reject the assumption $\mu = 1$ and complete the proof.