Exercise 1 (Exact Volume of ℓ_p Balls)
Let $\Gamma(a) = \int_0^\infty t^{a-1}e^{-t}dt$, $a > 0$ denote the Gamma function. You will prove that

$$\mathrm{vol}_n(B_p^n) = (2\Gamma(1/p)/p)^n/\Gamma(n/p+1).$$

1. Show that $\int_{\mathbb{R}^n} e^{-\|x\|_p^p}dx = (2\Gamma(1/p)/p)^n$.
   (Hint: Split the integral along each coordinate and apply a change of variables.)

2. Show that $\int_{\mathbb{R}^n} e^{-\|x\|_p^p}dx = \mathrm{vol}_n(B_p^n)\Gamma(n/p+1)$.
   (Hint: Use the identity $e^{-\|x\|_p^p} = \int_{\|x\|_p}^\infty pt^{p-1}e^{-t^p}dt$)

Solution:

1. *By splitting up the integral along coordinates, we get that*

$$\int_{\mathbb{R}^n} e^{-\|x\|_p^p}dx = \int_{\mathbb{R}^n}\prod_{i=1}^n e^{-|x_i|^p}dx \underbrace{=}_{Fubini} \prod_{i=1}^n \int_{\mathbb{R}} e^{-|x_i|^p}dx_i = (\int_{\mathbb{R}} e^{-|t|^p}dt)^n = (2\int_0^\infty e^{t^p}dt)^n. \quad (1)$$

Applying the change of variable $t \to t^{1/p}$, we see that

$$\int_0^\infty e^{t^p}dt = \frac{1}{p}\int_0^\infty t^{1/p-1}e^t dt = \frac{1}{p}\Gamma(1/p). \quad (2)$$

Combining (1), (2), *get the desired expression*

$$\int_{\mathbb{R}^n} e^{-\|x\|_p^p} = (2\Gamma(1/p)/p)^n.$$

2. *Using the identity, we see that*

$$\begin{aligned}
\int_{\mathbb{R}^n} e^{-\|x\|_p^p}dx &= \int_{\mathbb{R}^n}\int_{\|x\|_p}^\infty pt^{p-1}e^{-t^p}dtdx \\
&= \int_{\mathbb{R}^n}\int_0^\infty 1[\|x\|_p \le t]pt^{p-1}e^{-t^p}dtdx \\
&= \int_0^\infty pt^{p-1}e^{-t^p}\int_{\mathbb{R}^n} 1[\|x\|_p \le t]dxdt \quad (\textit{ by Fubini }) \\
&= \int_0^\infty pt^{p-1}e^{-t^p}\mathrm{vol}_n(tB_p^n)dt \\
&= \int_0^\infty pt^{p+n-1}e^{-t^p}\mathrm{vol}_n(B_p^n)dt. \quad (3)
\end{aligned}$$

Applying the change of variable $t \to t^{1/p}$ to the last line, we get

$$\int_0^\infty pt^{p+n-1}e^{-t^p}dt = \int_0^\infty t^{n/p}e^{-t}dt = \Gamma(n/p+1). \quad (4)$$

Combining (3), (4) *yields the result.*

Exercise 2 (Duality between ℓ_2 norm minimization and ellipsoid maximation)
Let $A \in \mathbb{R}^{m\times n}$ be a non-singular matrix and $b \in \mathbb{R}^n$.

1. Show that

$$\min\{\|x\|_2 : x \in \mathbb{R}^m, A^\mathsf{T} x = b\} \geq \max\{y^\mathsf{T} b : y \in \mathbb{R}^n, \|Ay\|_2 \leq 1\}.$$

(Hint: Use Cauchy-Schwarz.)

2. Show that the value of both systems is equal to $\sqrt{b^\mathsf{T}(A^\mathsf{T} A)^{-1} b}$ and hence both have same value. (Hint: Use the fact that $\|Ay\|_2 \leq 1$ defines an ellipsoid.)

3. Let $A_1, \dots, A_k$ be matrices where $A_i \in \mathbb{R}^{m_i \times n}$, $i \in [k]$, let $\lambda_1, \dots, \lambda_k > 0$ and $b \in \mathbb{R}^n$. Assume that $\sum_{i=1}^k \lambda_i A_i^\mathsf{T} A_i \succ 0$ (positive definite). Use the above to show that the systems

$$1.\ \min\{\sqrt{\sum_{i=1}^k \lambda_i \|x_i\|^2} : x_i \in \mathbb{R}^{m_i}, i \in [k], \sum_{i=1}^k \lambda_i A_i^\mathsf{T} x_i = b\}$$

$$2.\ \max\{b^\mathsf{T} y : y \in \mathbb{R}^n, \sum_{i=1}^k \lambda_i \|A_i y\|_2^2 \leq 1\}$$

are strong duals to each other. That is, show that any solution to (1) has value at least that of (2), and the optimal solutions have same value.

Solution:

1. *Let $x \in \mathbb{R}^m$ satisfy $A^\mathsf{T} x = b$ and $\|Ay\|_2 \leq 1$. Then*

$$\|x\|_2 \geq \|Ay\|_2 \|x\|_2 \underbrace{\geq}_{\text{Cauchy-Schwarz}} (Ay)^\mathsf{T} x = y^\mathsf{T} b.$$

2. *Since A is non-singular, note that $A^\mathsf{T} A$ is invertible. Furthermore, $((A^\mathsf{T} A)^{-1})^\mathsf{T} = ((A^\mathsf{T} A)^\mathsf{T})^{-1} = (A^\mathsf{T} A)^{-1}$, so the inverse is symmetric. Let $y = (A^\mathsf{T} A)^{-1} b / \sqrt{b^\mathsf{T}(A^\mathsf{T} A)^{-1} b}$. From here, we get that*

$$y^\mathsf{T} b = \sqrt{b^\mathsf{T}(A^\mathsf{T} A)^{-1} b},$$
$$\|Ay\|_2^2 = y^\mathsf{T}(A^\mathsf{T} A) y = b^\mathsf{T}(A^\mathsf{T} A)^{-\mathsf{T}}(A^\mathsf{T} A)(A^\mathsf{T} A)^{-1} b / (b^\mathsf{T}(A^\mathsf{T} A)^{-1} b) = 1.$$

Therefore the value of the right hand side program is at least $\sqrt{b^\mathsf{T}(A^\mathsf{T} A)^{-1} b}$.

Let $x = A(A^\mathsf{T} A)^{-1} b$. From here, we get that

$$A^\mathsf{T} x = (A^\mathsf{T} A)(A^\mathsf{T} A)^{-1} b = b,$$
$$\|x\|_2^2 = x^\mathsf{T} x = b^\mathsf{T}(A^\mathsf{T} A)^{-\mathsf{T}}(A^\mathsf{T} A)(A^\mathsf{T} A)^{-1} b = b^\mathsf{T}(A^\mathsf{T} A)^{-1} b.$$

Thus the value of the left hand side program is at most $\sqrt{b^\mathsf{T}(A^\mathsf{T} A)^{-1} b}$. By Part 1, both programs therefore has the same value, as needed.

3. *Let $A \in \mathbb{R}^{m \times n}$, where $m = \sum_{i=1}^k m_i$, denote the block diagonal matrix*

$$A = \begin{pmatrix} \sqrt{\lambda_1} A_1 \\ \sqrt{\lambda_2} A_2 \\ \dots \\ \sqrt{\lambda_k} A_k \end{pmatrix}.$$

A direct computation reveals that $\|Ay\|_2^2 = \sum_{i=1}^k \lambda_i \|A_i y\|_2^2$. By Parts 1 and 2, we therefore have that

$$\max\{y^\mathsf{T} b : \sum_{i=1}^k \lambda_i \|A_i y\|_2^2 \le 1\} = \min\{\|x\|_2 : x \in \mathbb{R}^m, A^\mathsf{T} x = b\}.$$

We now work on the second hand program to get it in the desired form. Let us decompose $x = (x_1, \dots, x_k)$, where $x_i \in \mathbb{R}^{m_i}$. From here, we see that

$$b = A^\mathsf{T} x = \sum_{i=1}^k \sqrt{\lambda_i} A_i^\mathsf{T} x_i = \sum_{i=1}^k \lambda_i A_i^\mathsf{T} (x_i / \sqrt{\lambda_i}),$$
$$\|x\|^2 = \sum_{i=1}^k \|x_i\|_2^2 = \sum_{i=1}^m \lambda_i \|x_i / \sqrt{\lambda_i}\|_2^2.$$

Therefore, letting $z_i = x_i / \sqrt{\lambda_i}$, $\forall i \in [k]$, we see that

$$\min\{\|x\|_2 : x \in \mathbb{R}^m, A^\mathsf{T} x = b\} = \min\{\sqrt{\sum_{i=1}^k \lambda_i \|z_i\|_2^2} : z_i \in \mathbb{R}^{m_i}, i \in [k], \sum_{i=1}^k \lambda_i A_i^\mathsf{T} z_i = b\},$$

as needed.

Exercise 3 (1/n-Concavity of Determinant) Let $A, B \succ 0$ be $n \times n$ positive definite matrices.

1. Show that $\det(A)^{1/n} = \min\{\mathrm{tr}(AX)/n : X \succ 0, \det(X) = 1\}$, where equality on the right hand side is uniquely attained at $X = A^{-1} \det(A)^{1/n}$. (Hint: Recall that the trace of a matrix is the sum of the eigenvalues while the determinant is the product. Compare the two via the arithmetic mean - geometric mean (AM-GM) inequality.)

2. Conclude that for $\lambda \in (0, 1)$, $\det(\lambda A)^{1/n} + (1 - \lambda) \det(B)^{1/n} \le \det(\lambda A + (1 - \lambda) B)^{1/n}$ with equality iff $A \in \mathbb{R}_+ B$.

Solution:

1. *Let $\lambda_1, \dots, \lambda_n > 0$ denote the eigen values of A. These are all positive since $A \succ 0$.*

   Let $X = A^{-1} \det(A)^{1/n}$. Recall that $\det(A) = \prod_{i=1}^n \lambda_i > 0$ and that the eigenvalues of A^{-1} are $1/\lambda_1, \dots, 1/\lambda_n > 0$. Since $A^{-\mathsf{T}} = (A^\mathsf{T})^{-1} = A^{-1}$ is symmetric, X is positive definite. Second, $\det(X) = \det(A^{-1})(\det(A)^{1/n})^n = \det(A)^{-1} \det(A) = 1$. Thus, the value of the program is at most $\mathrm{tr}(XA)/n = \det(A)^{1/n} \mathrm{tr}(I_n)/n = \det(A)^{1/n}$.

   Now take $X \succ 0$ such that $\det(X) = 1$. Letting $X^{1/2}$ denote the unique positive definite square root of X, we see that AX is similar to $X^{1/2}(AX)X^{-1/2} = X^{1/2} A X^{1/2}$. Since the latter is positive *definite, AX has positive eigen values $\gamma_1, \ldots, \gamma_n > 0$ and is diagonalizable. Furthermore, since $\det(AX) = \det(A)\det(X) = \det(A)$, we see that $\prod_{i=1}^n \gamma_i = \det(A)$. From here, we have that*

$$\mathrm{tr}(AX)/n = \sum_{i=1}^n \gamma_i/n \geq \prod_{i=1}^n \gamma_i^{1/n} = \det(A)^{1/n}$$

where the first inequality is a consequence of the AM-GM inequality. By the equality conditions for AM-GM, the above inequality holds at equality iff $\gamma_1 = \gamma_2 = \cdots = \gamma_n = \det(A)^{1/n}$. Since AX is diagonalizable, equality thus implies that $AX = \det(A)^{1/n} I_n$, where I_n is the $n \times n$ identity. That is, equality holds iff $X = A^{-1}\det(A)^{1/n}$ as needed.

2. *From Part 1, we see that*

$$\begin{aligned}\det(\lambda A + (1-\lambda)B)^{1/n} &= \min_{\det(X)=1, X \succ 0} \mathrm{tr}((\lambda A + (1-\lambda)B)X)/n \\ &= \min_{\det(X)=1, X \succ 0} \lambda \mathrm{tr}(AX)/n + (1-\lambda)\mathrm{tr}(BX)/n \\ &\geq \lambda \min_{\det(X)=1, X \succ 0} \mathrm{tr}(AX)/n + (1-\lambda) \min_{\det(X)=1, X \succ 0} \mathrm{tr}(BX)/n \\ &= \lambda \det(A)^{1/n} + (1-\lambda)\det(B)^{1/n}.\end{aligned}$$

By part 1, since $\lambda \in (0,1)$, the above holds at equality iff

$$(\lambda A + (1-\lambda)B)^{-1} \det(\lambda A + (1-\lambda)B)^{1/n} = A^{-1}\det(A)^{1/n} = B^{-1}\det(B)^{1/n}.$$

The above implies that $A = B\det(A)^{1/n}/\det(B)^{1/n} \Rightarrow A \in \mathbb{R}_+ B$. Now if $A \in \mathbb{R}_+ B$, since $A \succ 0$ this implies that $A = \gamma B$, with $\gamma > 0$. From here, it is direct to check the that equality conditions are satisfied, thus proving the statement.

Exercise 4 (Approximation by a Simplex)
Let $K \subset \mathbb{R}^n$ be a convex body. Let $\Delta = \mathrm{conv}(p_1, \ldots, p_{n+1})$ denote a maximum volume simplex contained in K and let $c = \sum_{i=1}^{n+1} p_i/(n+1)$ denote its center.

1. Show that such a simplex exists and give an example body K where it is not unique.
2. For $i \in [n+1]$, let η_i denote a unit normal to the hyperplane $H_i = \mathrm{aff.hull}(p_j : j \in [n+1] \setminus \{i\})$, pointing in the direction of p_i, and let $s_i \in \mathbb{R}^n$ satisfy $H_i = \{x \in \mathbb{R}^n : \langle \eta_i, x\rangle = s_i\}$ (note that by assumption $\langle \eta_i, p_i \rangle > s_i$). Prove that

$$K \subseteq \{x \in \mathbb{R}^n : |\langle \eta_i, x\rangle - s_i| \leq \langle \eta_i, p_i\rangle - s_i\}.$$

   (Hint: Show that if not one can replace p_i to make a simplex of larger volume.)
3. Conclude that $\Delta - c \subseteq K - c \subseteq (n+2)(\Delta - c)$.

Solution:

1. *If $K = B_2^n$, then if Δ is a maximum volume simplex then so is $R\Delta$ where R is any orthogonal transformation since $RB_2^n = B_2^n$ and R is measure preserving. Since Δ is clearly a strict subset of B_2^n, the maximizer cannot be unique.*

2. *Take $x \in K$. By the base times height formula in $\mathbb{R}^n$, we have that*

$$\begin{aligned}\mathrm{vol}_n(p_1,\dots,p_{i-1},x,p_i,\dots,p_n) &= \mathrm{vol}_{n-1}(p_1,\dots,p_{i-1},p_{i+1},\dots,p_n)\cdot \mathrm{dist}(H_i,x)/n \\ &= \mathrm{vol}_{n-1}(p_1,\dots,p_{i-1},p_{i+1},\dots,p_n)\cdot |\langle \eta_i, x\rangle - s_i|/n,\end{aligned}$$

where $\mathrm{dist}(H_i,x)$ is the Euclidean distance between the hyperplane H_i and x. By the definition of Δ, plugging in p_i for x above maximizes volume. Given the above formula, this implies that

$$\langle \eta_i, p_i\rangle - s_i = |\langle \eta_i, p_i\rangle - s_i| \geq |\langle \eta_i, x\rangle - s_i| \ \forall x \in K,$$

as needed.

3. *After shifting K and Δ by $-c$, we may assume that $0 = c = \sum_{i=1}^{n+1} p_i/(n+1)$. We now wish to show that $\Delta \subseteq K \subseteq (n+2)\Delta$. Since the first inclusion is by assumption, we may focus on proving $K \subseteq (n+2)\Delta$. To begin, we first claim that Δ can be expressed in inequality form as follows:*

$$\Delta = \{x \in \mathbb{R}^n : \langle \eta_i, x\rangle \geq s_i, \forall i \in [n+1]\}, \tag{5}$$

where s_i is as above (after shifting Δ). To see this, let

$$A = \begin{pmatrix} p_1 & p_2 & \dots & p_{n+1} \\ 1 & 1 & \dots & 1 \end{pmatrix}.$$

Note that $\mathrm{conv}(p_1,\dots,p_{n+1})$ has non-zero volume iff $p_1,\dots,p_n$ are affinely independent and hence iff A is invertible. From here, by definition

$$x \in \mathrm{conv}(p_1,\dots,p_{n+1}) \Leftrightarrow A^{-1}\begin{pmatrix} x \\ 1\end{pmatrix} \geq 0.$$

To prove the claim it suffices to show that the i^{th} row of A^{-1} is a positive multiple of $(\eta_i, -s_i)$. Note that this is true iff $\langle \eta_i, p_j\rangle - s_i > 0$ for $i = j$ and $= 0$ otherwise. This last property follows by construction and hence the claim holds.

Given the inequality description, to prove that $K \subseteq (n+2)\Delta$ it suffices to show that

$$\langle \eta_i, x\rangle \geq (n+2)s_i, \ \forall x \in K, \ i \in [n+1].$$

Take $x \in K$ and $i \in [n+1]$. Since $0 = c \in \Delta$ by assumption, we have that $s_i \leq 0$ in (5). If $\langle \eta_i, x\rangle \geq s_i$, the desired inequality is trivial since $s_i \leq 0$. Thus, we may assume that $\langle \eta_i, x\rangle \leq s_i$. In this situation, by part 2 we have that

$$\langle \eta_i, p_i\rangle - s_i \geq |\langle \eta_i, x\rangle - s_i| = s_i - \langle \eta_i, x\rangle. \tag{6}$$

From here, we see that

$$\begin{aligned} 0 = \langle \eta_i, c\rangle &= \sum_{j\neq i}\langle \eta_i, p_j\rangle/(n+1) + \langle \eta_i, p_i\rangle/(n+1) = s_i n/(n+1) + \langle \eta, p_i\rangle/(n+1) \\ &\Rightarrow \langle \eta, p_i\rangle = -s_i n. \end{aligned} \tag{7}$$

Combining (6), (7), we have that

$$\langle \eta_i, x\rangle - s_i \geq s_i - \langle \eta_i, p_i\rangle = s_i(n+1) \Rightarrow \langle \eta_i, x\rangle \geq s_i(n+2),$$

as needed.