Convex Optimization - HW5

IGOR LUCINDO CARDOSO

December 7, 2025

Disclaimer

This work was assisted by Gemini in the following ways: improvement of text content, clarification of mathematical definitions, and rapid generation of graphs and tables.

Course Evaluation: I confirm that I have submitted the course evaluation.

Problem 1

1.1 - First, let's simplify the term inside the summation using the definition $\sigma(t) = 1/(1 + e^{-t})$.

$$\log(\sigma(t)) = \log\left(\frac{1}{1 + e^{-t}}\right) = \log(1) - \log(1 + e^{-t}) = -\log(1 + e^{-t}).$$

Maximizing the objective function is equivalent to minimizing its negative. We formulate the negative-likelihood minimization problem as follows:

$$\min_{w \in \mathbb{R}^n, b \in \mathbb{R}} \quad \sum_{i \in [d]} \log(1 + e^{-y_i(\langle x_i, w \rangle + b)})$$

$$\text{s.t.} \quad \|w\|_2 \leq R.$$

1.2 - First, we define the exponential cone $K_{\exp}$ and the second-order cone $\mathcal{C}$ as:

$$K_{\exp} = \{(x, y, z) \mid y > 0, z \geq y e^{x/y}\} \cup \{(x, 0, z) \mid x \leq 0, z \geq 0\},$$
$$\mathcal{C} = \{(x, t) \in \mathbb{R}^n \times \mathbb{R} \mid \|x\|_2 \leq t\}.$$

To rewrite the problem in conic form, we introduce auxiliary variables u_i for each term in the log-likelihood sum. The optimization problem is equivalent to minimizing $\sum u_i$ subject to the constraints:

$$u_i \geq \log(1 + e^{-y_i(\langle x_i, w \rangle + b)}) \iff 1 \geq e^{-u_i} + e^{-u_i - y_i(\langle x_i, w \rangle + b)}.$$

We represent this sum of exponentials using auxiliary variables $r_{i,1}, r_{i,2}$, satisfying $r_{i,1} + r_{i,2} \leq 1$. The individual exponential terms are handled by $K_{\exp}$ constraints. The full conic formulation is:

$$\min_{w,b,u,r} \quad \sum_{i \in [d]} u_i$$

$$\begin{aligned}
\text{s.t.} \quad & r_{i,1} + r_{i,2} \leq 1, & \forall i \in [d] \\
& (-u_i, 1, r_{i,1}) \in K_{\exp}, & \forall i \in [d] \\
& (-u_i - y_i(\langle x_i, w \rangle + b), 1, r_{i,2}) \in K_{\exp}, & \forall i \in [d] \\
& (w, R) \in \mathcal{C}.
\end{aligned}$$

1.3 - Knowing that the dual cone of $K_{\exp}$ is given by

$$K_{\exp}^* = \{(u, v, w) \mid u < 0, -u e^{v/u} \leq w\} \cup \{(0, v, w) \mid v \geq 0, w \geq 0\}.$$

With dual variables $\lambda_i \geq 0$ for the linear inequality, $(a_{i,1}, \beta_{i,1}, \gamma_{i,1}) \in K_{\exp}^*$, $(a_{i,2}, \beta_{i,2}, \gamma_{i,2}) \in K_{\exp}^*$, and $(\mu, \tau) \in \mathcal{C}$ (where $\mathcal{C}$ is self-dual), we construct the Lagrangian:

$$\mathcal{L} = \sum_{i \in [d]} u_i + \sum_{i \in [d]} \lambda_i (r_{i,1} + r_{i,2} - 1) - \sum_{i \in [d]} \langle (a_{i,1}, \beta_{i,1}, \gamma_{i,1}), (-u_i, 1, r_{i,1}) \rangle$$
$$- \sum_{i \in [d]} \langle (a_{i,2}, \beta_{i,2}, \gamma_{i,2}), (-u_i - y_i(\langle x_i, w \rangle + b), 1, r_{i,2}) \rangle - \langle (\mu, \tau), (w, R) \rangle$$

Grouping terms by primal variables (w, b, u, r), the Lagrangian simplifies to:

$$\mathcal{L} = \sum_{i \in [d]} u_i (1 + a_{i,1} + a_{i,2}) + \sum_{i \in [d]} r_{i,1} (\lambda_i - \gamma_{i,1}) + \sum_{i \in [d]} r_{i,2} (\lambda_i - \gamma_{i,2})$$
$$+ \left\langle w, \sum_{i \in [d]} a_{i,2} y_i x_i - \mu \right\rangle + b \left(\sum_{i \in [d]} a_{i,2} y_i \right) - \sum_{i \in [d]} (\lambda_i + \beta_{i,1} + \beta_{i,2}) - \tau R$$

Minimizing $\mathcal{L}$ with respect to the primal variables yields the optimality conditions: $a_{i,1} + a_{i,2} = -1$, $\gamma_{i,1} = \gamma_{i,2} = \lambda_i$, $\sum a_{i,2} y_i = 0$, and $\mu = \sum a_{i,2} y_i x_i$. Therefore the dual problem is given by

$$\max \min_{w,b,u,r} \mathcal{L}.$$

Which is simplified to

$$\max \ - \sum_{i \in [d]} (\lambda_i + \beta_{i,1} + \beta_{i,2}) - \tau R$$
$$\text{s.t.} \ \sum_{i \in [d]} a_{i,2} y_i = 0$$
$$(-1 - a_{i,2}, \beta_{i,1}, \lambda_i) \in K_{\exp}^* \qquad\qquad \forall i \in [d]$$
$$(a_{i,2}, \beta_{i,2}, \lambda_i) \in K_{\exp}^* \qquad\qquad \forall i \in [d]$$
$$(\sum_{i \in [d]} a_{i,2} y_i x_i, \tau) \in \mathcal{C}.$$

1.4 - We implemented the primal and dual formulations in Python using the `MOSEK` solver. The experiment was conducted with $d = 50$ randomly generated data points in $\mathbb{R}^5$ and a radius constraint of $R = 5$. The obtained numerical results are:

Problem	Optimal Value
Primal Formulation	17.887166028
Dual Formulation	17.887166020
Duality Gap	$\mathbf{7.39 \times 10^{-9}}$

Both the primal and dual solvers converged to an `Optimal` status. The duality gap is negligible ($\approx 10^{-9}$), effectively zero within numerical precision. This confirms that strong duality holds for this problem and validates the correctness of our derived dual formulation.

Problem 2

2.1 - We reformulate the problem into a conic form using the exponential cone K_{exp}. First, we substitute the variables x, y, z with their logarithmic counterparts $\tilde{x} = \log x$, $\tilde{y} = \log y$, and $\tilde{z} = \log z$. The objective function $xy + y^2 z$ is replaced by minimizing $a + b$, subject to the epigraph constraints:

$$a \geq e^{\tilde{x}+\tilde{y}} \quad \Longleftrightarrow \quad (\tilde{x} + \tilde{y}, 1, a) \in K_{\text{exp}}$$
$$b \geq e^{2\tilde{y}+\tilde{z}} \quad \Longleftrightarrow \quad (2\tilde{y} + \tilde{z}, 1, b) \in K_{\text{exp}}.$$

Next, we handle the constraints by introducing auxiliary variables u_1, u_2 for the first constraint and v_1, v_2 for the second constraint, such that $u_1 + u_2 \leq 1$ and $v_1 + v_2 \leq 1$. The terms map to the exponential cone as follows:

$$u_1 \geq 0.1\sqrt{x} = e^{\log(0.1)+0.5\tilde{x}} \quad \Longleftrightarrow \quad (\log(0.1) + 0.5\tilde{x}, 1, u_1) \in K_{\text{exp}}$$
$$u_2 \geq 2y^{-1} = e^{\log(2)-\tilde{y}} \quad \Longleftrightarrow \quad (\log(2) - \tilde{y}, 1, u_2) \in K_{\text{exp}}$$
$$v_1 \geq z^{-1} = e^{-\tilde{z}} \quad \Longleftrightarrow \quad (-\tilde{z}, 1, v_1) \in K_{\text{exp}}$$
$$v_2 \geq yx^{-2} = e^{\tilde{y}-2\tilde{x}} \quad \Longleftrightarrow \quad (\tilde{y} - 2\tilde{x}, 1, v_2) \in K_{\text{exp}}.$$

Collecting these constraints, we obtain the full conic reformulation:

$$
\begin{aligned}
\min \ & a + b \\
\text{s.t. } & u_1 + u_2 \leq 1 \\
& v_1 + v_2 \leq 1 \\
& (\tilde{x} + \tilde{y}, 1, a) \in K_{\text{exp}} \\
& (2\tilde{y} + \tilde{z}, 1, b) \in K_{\text{exp}} \\
& (\log(0.1) + 0.5\tilde{x}, 1, u_1) \in K_{\text{exp}} \\
& (\log(2) - \tilde{y}, 1, u_2) \in K_{\text{exp}} \\
& (-\tilde{z}, 1, v_1) \in K_{\text{exp}} \\
& (\tilde{y} - 2\tilde{x}, 1, v_2) \in K_{\text{exp}}.
\end{aligned}
$$

2.2 - Let K_{exp}^* be the dual cone of K_{exp}. We introduce dual variables $\lambda_1, \lambda_2 \geq 0$ for the linear inequalities, and $(x_k^*, y_k^*, z_k^*) \in K_{\text{exp}}^*$ for $k = 1, \ldots, 6$ corresponding to the six exponential cone constraints derived in 2.1.
We construct the Lagrangian:

$$
\begin{aligned}
\mathcal{L} = & (a + b) + \lambda_1(u_1 + u_2 - 1) + \lambda_2(v_1 + v_2 - 1) \\
& - \langle (x_1^*, y_1^*, z_1^*), (\tilde{x} + \tilde{y}, 1, a) \rangle - \langle (x_2^*, y_2^*, z_2^*), (2\tilde{y} + \tilde{z}, 1, b) \rangle \\
& - \langle (x_3^*, y_3^*, z_3^*), (\log(0.1) + 0.5\tilde{x}, 1, u_1) \rangle - \langle (x_4^*, y_4^*, z_4^*), (\log(2) - \tilde{y}, 1, u_2) \rangle \\
& - \langle (x_5^*, y_5^*, z_5^*), (-\tilde{z}, 1, v_1) \rangle - \langle (x_6^*, y_6^*, z_6^*), (\tilde{y} - 2\tilde{x}, 1, v_2) \rangle
\end{aligned}
$$

Grouping terms by primal variables $(a, b, u, v, \tilde{x}, \tilde{y}, \tilde{z})$, the Lagrangian simplifies to:

$$
\begin{aligned}
\mathcal{L} = & a(1 - z_1^*) + b(1 - z_2^*) + u_1(\lambda_1 - z_3^*) + u_2(\lambda_1 - z_4^*) + v_1(\lambda_2 - z_5^*) + v_2(\lambda_2 - z_6^*) \\
& + \tilde{x}(-x_1^* - 0.5x_3^* + 2x_6^*) + \tilde{y}(-x_1^* - 2x_2^* + x_4^* - x_6^*) + \tilde{z}(-x_2^* + x_5^*) \\
& - (\lambda_1 + \lambda_2) - \sum_{k=1}^{6} y_k^* - x_3^* \log(0.1) - x_4^* \log(2)
\end{aligned}
$$

Minimizing $\mathcal{L}$ with respect to the primal variables yields the optimality conditions:

$$
\begin{aligned}
& z_1^* = 1, \quad z_2^* = 1, \\
& z_3^* = z_4^* = \lambda_1, \quad z_5^* = z_6^* = \lambda_2, \\
& 2x_6^* = x_1^* + 0.5x_3^*, \quad x_4^* = x_1^* + 2x_2^* + x_6^*, \quad x_5^* = x_2^*.
\end{aligned}
$$

3

Therefore the dual problem is given by $\max \min \mathcal{L}$, which is simplified to:

$$\max \quad -(\lambda_1 + \lambda_2) - \sum_{k=1}^{6} y_k^* - x_3^* \log(0.1) - x_4^* \log(2)$$

$$\text{s.t.} \quad x_1^* + 0.5x_3^* - 2x_6^* = 0$$
$$x_1^* + 2x_2^* - x_4^* + x_6^* = 0$$
$$x_2^* - x_5^* = 0$$
$$(x_1^*, y_1^*, 1) \in K_{\text{exp}}^*, \quad (x_2^*, y_2^*, 1) \in K_{\text{exp}}^*$$
$$(x_3^*, y_3^*, \lambda_1) \in K_{\text{exp}}^*, \quad (x_4^*, y_4^*, \lambda_1) \in K_{\text{exp}}^*$$
$$(x_5^*, y_5^*, \lambda_2) \in K_{\text{exp}}^*, \quad (x_6^*, y_6^*, \lambda_2) \in K_{\text{exp}}^*.$$

2.3 - Yes, strong duality holds for this problem.
Explanation: The problem is a convex optimization problem (specifically, a Geometric Program in convex form). For such problems, strong duality is guaranteed if **Slater's condition** is satisfied. Slater's condition requires the existence of a strictly feasible point—a point $(x, y, z) > 0$ that satisfies the inequality constraints with strict inequality.
Recalling the primal constraints:

$$0.1\sqrt{x} + \frac{2}{y} < 1 \quad \text{and} \quad \frac{1}{z} + \frac{y}{x^2} < 1.$$

We can demonstrate strict feasibility by constructing a specific point, for example, $(x, y, z) = (10, 3, 2)$:

- **Constraint 1:**
$$0.1\sqrt{10} + \frac{2}{3} \approx 0.316 + 0.667 = 0.983 < 1.$$

- **Constraint 2:**
$$\frac{1}{2} + \frac{3}{10^2} = 0.5 + 0.03 = 0.53 < 1.$$

Since strictly feasible points exist, Slater's condition is satisfied. Therefore, strong duality holds, and the duality gap is zero.

Problem 3

3.1

Proof. To prove that $\text{tr} : \mathbb{R}^{n \times n} \to \mathbb{R}$ is a linear operator, we show that it satisfies both additivity and homogeneity properties. Using the definition of matrix addition and scalar multiplication, the (i, i)-th entry of the matrix $(\alpha A + \beta B)$ is given by $\alpha A_{ii} + \beta B_{ii}$. Therefore:

$$\text{tr}(\alpha A + \beta B) = \sum_{i \in [n]} (\alpha A + \beta B)_{ii} = \sum_{i \in [n]} (\alpha A_{ii} + \beta B_{ii}).$$

By the linearity of finite summation, we can distribute the sum:

$$\sum_{i \in [n]} \alpha A_{ii} + \sum_{i \in [n]} \beta B_{ii} = \alpha \sum_{i \in [n]} A_{ii} + \beta \sum_{i \in [n]} B_{ii} = \alpha \text{tr}(A) + \beta \text{tr}(B).$$

Since both additivity and homogeneity properties are satisfied, tr is a linear operator. $\square$

3.2

(a)

Proof. To prove that the covariance matrix Σ is Positive Semidefinite (PSD), we must show that $x^\top \Sigma x \geq 0$ for any vector $x \in \mathbb{R}^n$.
Using the linearity of the expectation operator, we substitute the definition of Σ:

$$x^\top \Sigma x = x^\top \mathbb{E}[(\xi - \mu)(\xi - \mu)^\top] x = \mathbb{E}[x^\top (\xi - \mu)(\xi - \mu)^\top x].$$

As the term $x^\top (\xi - \mu)$ is a scalar, we get:

$$x^\top \Sigma x = \mathbb{E}[(x^\top (\xi - \mu))^2]. \tag{1}$$

Since $(x^\top (\xi - \mu))^2$ is a squared real number, it is always non-negative. By the monotonicity of expectation, the expected value of a non-negative random variable is non-negative:

$$\mathbb{E}[(x^\top (\xi - \mu))^2] \geq 0.$$

Thus, $x^\top \Sigma x \geq 0$ for all $x \in \mathbb{R}^n$, which implies that Σ is PSD. $\square$

(b)

Proof. First, we calculate the expectation of Z using the linearity of the expectation operator:

$$\mathbb{E}[x^\top \xi] = x^\top \mathbb{E}[\xi] = x^\top \mu.$$

Now, by the definition of variance for a scalar random variable, we have:

$$\text{Var}(\langle \xi, x \rangle) = \mathbb{E}[(x^\top \xi - x^\top \mu)^2] = \mathbb{E}[(x^\top (\xi - \mu))^2].$$

From equation (1), we conclude:

$$\text{Var}(\langle \xi, x \rangle) = x^\top \Sigma x.$$

$\square$

(c)

Proof. We compute the trace of the covariance matrix Σ. Using the linearity of the trace operator and the linearity of expectation, we can commute the two operators:

$$\mathrm{tr}(\Sigma) = \mathrm{tr}\left(\mathbb{E}[(\xi - \mu)(\xi - \mu)^\top]\right) = \mathbb{E}\left[\mathrm{tr}\left((\xi - \mu)(\xi - \mu)^\top\right)\right].$$

Recall that for any vector v, the matrix $vv^\top$ has diagonal entries $(vv^\top)_{ii} = v_i^2$. Therefore, the trace (sum of diagonal entries) is the squared Euclidean norm:

$$\mathrm{tr}\left((\xi - \mu)(\xi - \mu)^\top\right) = \sum_{i \in [n]} (\xi_i - \mu_i)^2.$$

Substituting this back into the expectation:

$$\mathrm{tr}(\Sigma) = \mathbb{E}[\sum_{i \in [n]} (\xi_i - \mu_i)^2].$$

By the linearity of expectation, we can distribute the expectation across the sum:

$$\mathrm{tr}(\Sigma) = \sum_{i \in [n]} \mathbb{E}\left[(\xi_i - \mu_i)^2\right].$$

Since $\mu_i = \mathbb{E}[\xi_i]$, the term $\mathbb{E}[(\xi_i - \mu_i)^2]$ is exactly the definition of the variance of the i-th component, $\mathrm{Var}(\xi_i)$. Thus:

$$\mathrm{tr}(\Sigma) = \sum_{i \in [n]} \mathrm{Var}(\xi_i).$$

$\square$

Problem 4

4.1 — Since there is no distributional ambiguity, we assume the true distribution is known to be the joint distribution $\mathbb{P}$ where components are independent Uniform$(10, 15)$. We formulate the standard stochastic programming problem, which minimizes the expected cost under the reference distribution $\mathbb{P}$:

$$\min_{x \in \mathbb{R}_+^n} \quad \mathbb{E}_\mathbb{P}\left[\langle h, (x - \xi)_+ \rangle + \langle p, (\xi - x)_+ \rangle\right].$$

4.2 — Let $\beta = 0.05$. The Conditional Value at Risk (CVaR) at level β is equivalent to the worst-case expectation over an ambiguity set of distributions defined by a likelihood ratio constraint.
The minimax formulation is:

$$\min_{x \in \mathbb{R}_+^n} \sup_{\mathbb{Q}} \mathbb{E}_\mathbb{Q}[\langle h, (x - \xi)_+ \rangle + \langle p, (\xi - x)_+ \rangle]$$

$$\text{s.t. } \langle 1, \mathbb{Q} \rangle = 1$$

$$\mathbb{Q} \leq \frac{1}{\beta}\mathbb{P}$$

$$\mathbb{Q} \geq 0.$$

To derive the single-level minimization problem, we compute the dual of the inner maximization problem below.

$$\sup_{\mathbb{Q}} \mathbb{E}_\mathbb{Q}[\langle h, (x - \xi)_+ \rangle + \langle p, (\xi - x)_+ \rangle]$$

$$\text{s.t. } \langle 1, \mathbb{Q} \rangle = 1$$

$$\mathbb{Q} \leq \frac{1}{\beta}\mathbb{P}$$

$$\mathbb{Q} \geq 0.$$

We introduce the dual variables $v, w \geq 0$, where v, w are bounded continuous functions that vanishes at infinite, $\alpha \in \mathbb{R}$. Hence the dual problem is given as follows.

$$\inf_{\mathbb{Q}} \alpha + \frac{1}{\beta}\mathbb{E}_\mathbb{P}[w]$$

$$\text{s.t. } \langle h, (x - \xi)_+ \rangle + \langle p, (\xi - x)_+ \rangle + v - \alpha - w = 0$$

$$v \geq 0$$

$$w \geq 0$$

$$\alpha \in \mathbb{R}.$$

Which simplifies to

$$\min \alpha + \frac{1}{\beta}\mathbb{E}_\mathbb{P}[(\langle h, (x - \xi)_+ \rangle + \langle p, (\xi - x)_+ \rangle - \alpha)_+]$$

$$\text{s.t. } \alpha \in \mathbb{R}.$$

Substituting this back, we obtain the associated single-level minimization problem:

$$\min \alpha + \frac{1}{\beta}\mathbb{E}_\mathbb{P}[(\langle h, (x - \xi)_+ \rangle + \langle p, (\xi - x)_+ \rangle - \alpha)_+]$$

$$\text{s.t. } \alpha \in \mathbb{R}$$

$$x \in \mathbb{R}_+^n.$$

4.3 - Let μ and Σ denote the mean and covariance of ξ, respectively. With this consideration, we set up the moment-based DRO formulation as follows.

$$\min_{x\in\mathbb{R}^n_+} \sup_{\mathbb{Q}} \mathbb{E}_{\mathbb{Q}}[\langle h, (x-\xi)_+\rangle + \langle p, (\xi-x)_+\rangle]$$

$$\text{s.t. } \langle 1, \mathbb{Q}\rangle = 1$$

$$\mathbb{Q} \geq 0$$

$$\|\Sigma^{-\frac{1}{2}}(\mathbb{E}_{\mathbb{Q}}[\xi] - \mu)\|_2 \leq \sqrt{\lambda_1}$$

$$\lambda_2\Sigma - \mathbb{E}_{\mathbb{Q}}[(\xi-\mu)(\xi-\mu)^\top] \succeq 0.$$

Once again, to derive the single-level minimization problem, we compute the dual of the inner maximization, which is given as by the formulation below.

$$\inf \alpha - \langle \Sigma^{-\frac{1}{2}}\mu, q\rangle + \sqrt{\lambda_1}\|q\|_2 + \lambda_2\langle\Sigma, Q\rangle$$

$$\text{s.t. } \alpha - \langle\Sigma^{-\frac{1}{2}}\xi, q\rangle + \langle(\xi-\mu)(\xi-\mu)^\top, Q\rangle \geq \langle h, (x-\xi)_+\rangle + \langle p, (\xi-x)_+\rangle \qquad \forall \xi \in \Xi$$

$$\alpha \in \mathbb{R}$$

$$Q \succeq 0$$

$$q \in \mathbb{R}^n.$$

Substituting this back, we obtain the associated single-level minimization problem:

$$\min \alpha - \langle \Sigma^{-\frac{1}{2}}\mu, q\rangle + \sqrt{\lambda_1}\|q\|_2 + \lambda_2\langle\Sigma, Q\rangle$$

$$\text{s.t. } \alpha - \langle\Sigma^{-\frac{1}{2}}\xi, q\rangle + \langle(\xi-\mu)(\xi-\mu)^\top, Q\rangle \geq \langle h, (x-\xi)_+\rangle + \langle p, (\xi-x)_+\rangle \qquad \forall \xi \in \Xi$$

$$\alpha \in \mathbb{R}$$

$$Q \succeq 0$$

$$q \in \mathbb{R}^n$$

$$x \in \mathbb{R}^n_+.$$

4.4 — We implemented both the Stochastic Programming (SP) and the Moment-Based Distributionally Robust Optimization (DRO) formulations in Python using the `MOSEK` solver with $n = 10$. The obtained numerical results are:

Problem	Optimal Decision (x^*)	Optimal Value (Cost)
Stochastic Programming (SP)	13.0000	155.0000
Distributionally Robust (DRO)	12.6190	251.4881

The Stochastic Programming solution ($x^* = 13.00$) yields a lower cost because it optimizes for the known Uniform distribution. The DRO solution ($x^* \approx 12.62$) results in a significantly higher cost (251.49) because it guards against the worst-case distribution satisfying the moment constraints. The difference of approximately 96.49 represents the Price of Robustness.