UGA Master1 Applied Maths
Jérôme Malick – Spring 2026

# OR Complementary – a selection of exercises

**Exercise 1 – Support Functions.** Let $C$ be a subset of $\mathbb{R}^n$; recall that the support function of $C$ is

$$\sigma_C(x) = \sup_{y \in C} x^\top y \qquad \text{for } x \in \mathbb{R}^n.$$

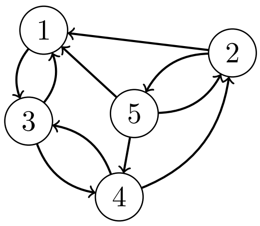
**a)** Show that $\sigma_C \colon \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is convex.

**b)** Calculate the support function for the following subsets of $\mathbb{R}^n$:

- $C$ the Euclidean ball of radius 1 (draw a picture in $\mathbb{R}^2$);
- $C = (\mathbb{R}^+)^n$ the positive orthant;
- $C = [a, b]$ the segment joining two points $a$ and $b$ in $\mathbb{R}^n$.

**c)** Show: $\sigma_C = \sigma_{\mathrm{conv}\, C}$. (In words: a support function doesn't distinguish between $C$ and its convex hull).

**Exercise 2 – (Google) PageRank.** The problem of ranking webpages is of the utmost importance for search engines. To this end, a popular approach is to represent webpages as a graph where the nodes are the pages themselves and the edges are the links between them (if page $i$ contains a links pointing toward page $j$, there is a directed edge from node $i$ to node $j$ in the graph). Then, a page/node has a high score if there are many links pointing toward it, especially coming from highly ranked pages. To fix ideas, consider the graph below of $N = 5$ pages.




of incidence matrix $A = \begin{bmatrix} 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 \end{bmatrix}$

We could choose the number of incoming links, as a score: node 1 would be ranked first with 3, nodes $2, 3, 4$ second with 2, 5 last with 1. The drawback of this scoring is that $2, 3, 4$ have the same score but are different in nature, as 3 is pointed by the most important page. To correct this phenomenon, the Google founders proposed an (implicit) scoring, similar to the following.

The score $x_i$ of page $i$ is equal to the sum over the pages $j$ pointing toward $i$ of the scores $(x_j)$ divided by their number of outgoing links $n_j$, that is,

$$x_i \;=\; (1-\alpha) \sum_{j \in \mathcal{P}_i} \frac{x_j}{n_j} \;+\; \frac{\alpha}{N} \sum_{j=1}^{N} x_j \qquad (1)$$

where $\alpha$ is a "damping" parameter in $(0, 1)$ and $\mathcal{P}_i$ is the set of nodes pointing toward $i$.

**a)** Let $x \in \mathbb{R}^N$ be the vector of the pages scores. Write the score equation (1) as a linear equation $x = Rx$ with $R$ defined from the incidence matrix $A$.

**b)** Show that $R^\top e = e$, i.e. $R$ is column-stochastic (that is, its elements are non negative and its columns sum to ones).

**c)** Deduce first from **b** that 1 is an eigenvalue of $R$. Deduce also from **b** that $|||R||| = 1$ for a matrix norm, and then that the spectral radius is $\rho(R) = 1$.

**d)** Conclude with Perron-Frobenuis: the vector of score $x$, satisfying $\sum_i x_i = 1$, exists and is unique.

**e)** For the graph above, compute the score vector and show that 3 is the most important page.

**Exercise 3 – Pure Nash.** What are the pure Nash equilibria of the two following games?

| | | Player 2 A | B | C |
|---|---|---|---|---|
| Player 1 | a | (3,1) | (2,3) | (10,8) |
| | b | (4,5) | (3,0) | (6,4) |
| | c | (2,2) | (5,4) | (8,3) |
| | d | (7,6) | (4,5) | (5,4) |

| | | Player 2 A | B | C |
|---|---|---|---|---|
| Player 1 | a | (3,1) | (2,3) | (10,2) |
| | b | (4,5) | (3,0) | (6,4) |
| | c | (2,2) | (5,4) | (12,3) |
| | d | (5,6) | (4,5) | (9,7) |

**Exercise 4 – Small parametric game.** Consider this game depending on the parameter $x \in \mathbb{R}$:

| | | Player 2 A | B |
|---|---|---|---|
| Player 1 | A | (0.5,0.5) | $(x, 1-x)$ |
| | B | $(1-x, x)$ | (0.5, 0.5) |

**a)** What are the pure Nash equilibrium of this game, depending on $x$ ?

**b)** Given $(q, 1-q)$ a mixed strategy for Player 2, what is the expected payoff for Player 1 if he plays A? Same question if Player 1 plays B.

**c)** Following the notation of the course, let a mixed Nash equilibrium $\big((p^*, 1-p^*), (q^*, 1-q^*)\big)$ (not a pure one, so $p^* \notin \{0, 1\}$). Show that we have: $0.5q^\star + (1-q^\star)x - (1-x)q^\star - 0.5(1-q^\star) = 0$. Explain briefly why this makes sense and why this property is called "indifference".

**d)** What are the mixed Nash equilibrium of this game, depending on $x$ ?

**Exercise 5 – Mixed Nash.** Same as the previous exercise. Give the pure and mixed Nash equilibria for the following game, depending on the parameter $x \in \mathbb{R}$,

| | | Player 2 A | B |
|---|---|---|---|
| Player 1 | A | (0.5,0.5) | (0,1) |
| | B | (1,0) | $(\frac{1-x}{2}, \frac{1-x}{2})$ |

**Exercise 6 – Linear vs. non-linear duality.** Consider the optimization problem (in $\mathbb{R}$)

$$\begin{cases} \max \quad \varphi(x) = x \\ \quad x \leqslant 0, \; x \in \{-2, 1\}. \end{cases}$$

**a)** Write the dual problem associated to relaxing the contraint $x \leqslant 0$. Show that the duality gap is 2.

**b)** Solve the convexified problem (with $x \in [-2, 1]$). Show that the convexified optimal value is equal to the optimal dual value.

**c)** Redo the two above questions with $\varphi(x) = -x^2$. Do we get the same final equality?

**Exercise 7 – Pricing for a mixed-integer problem.** We consider the optimization problem in $\mathbb{R}^2$

$$F(d) := \begin{cases} \min \quad 5p_1 + 10p_2 \\ \quad p_1 + p_2 \geqslant d \\ \quad p \in \{0, 3\} \times [0, 1] \end{cases} \qquad (P_d)$$

**a)** Find the optimal solution $p(d)$, depending on $d \in [0, 4]$. Draw the graph of $F$.

**b)** Write the optimization problem as a max and introducing the Lagrangian

$$L_0(p; u) := -5p_1 - 10p_2 - u(-p_1 - p_2)\,,$$

to dualize $(P_0)$. Compute the optimal solution $p^u$ of maximizing the Lagrangian, depending on $u \geqslant 0$. Draw the graph of the associated dual function $\theta_0(u)$.

**c)** Form the dual of $(P_d)$, and express the dual function $\theta_d$ with the help of $\theta_0$. What is the minimum of $\theta_d$ for $d = 2$?

**d)** Observe graphically that the dual optimal solution is the slope of the convex enveloppe of $F$.

**Exercise 8 – Dualize other contraintes.** With course notation, we consider

$$\begin{cases} \max \quad \varphi(x) \\ \quad x \in X \\ \quad c(x) \in B \end{cases}$$

where $B$ is a subset of $\mathbb{R}^n$. We assume that we have an oracle solving $\theta(u) := \max_{x\in X} \varphi(x) - u^\top c(x)$.

**a)** Adding a slack variable, write the dual problem.

**b)** Apply the result to $B = \{0\}$, $B = \mathbb{R}^n_+$ and $B$ the $\ell_2$-ball of radius $\varepsilon$.

**Exercise 9 – Augmented Lagrangian relaxation.** We start this exercice with studying the following simple optimization problem in $\mathbb{R}^2$

$$\begin{cases} \max \quad -x_1 - 2x_2 \\ \quad x_1 + x_2 = 3 \\ \quad x_1 \in [0,2],\ x_2 \in \{0,2\}. \end{cases} \tag{P}$$

**a)** By observing that (P) reduces to the trivial problem

$$\begin{cases} \max \quad -x_1 - 4 \\ \quad x_1 = 1 \\ \quad x_1 \in [0,2], \end{cases}$$

give the optimal solution and the optimal value of (P).

**b)** What is the optimal solution and the optimal value of the convexified problem ? (where the constraint $x_2 \in \{0,2\}$ is replaced by $x_2 \in [0,2]$).

**c)** Write the Lagrangian and the dual function $\theta$ associated to the dualization in (P) of the constraint $x_1 + x_2 - 3 = 0$.

**d)** Draw the graph of $\theta$. Give the dual optimal solution, the dual optimal value, and the duality gap.

Let's now turn to the general framework of the course

$$\begin{cases} \max \quad \varphi(x) \\ \quad c(x) = 0,\ x \in X. \end{cases}$$

For a parameter $\rho > 0$, we define the augmented Lagrangian function by

$$L^\rho(x;u) := \varphi(x) - u^\top c(x) - \rho\|c(x)\|^2$$

and the associated augmented dual function by

$$\theta^\rho(u) := \max_{x\in X} L^\rho(x;u).$$

**d)** Show that $\theta^\rho \colon \mathbb{R}^n \to \mathbb{R}\cup\{+\infty\}$ is convex. Show that for any dual variable $u$ and any primal feasible variable $x \in X$ such that $c(x) = 0$, we have $\theta^\rho(u) \geqslant \varphi(x)$.

**e)** Fix $\bar{u}$ and $x(\bar{u}) \in X$ such that $\theta^\rho(\bar{u}) = L^\rho(x(\bar{u});\bar{u})$. Prove that, if $c(x(\bar{u})) = 0$, then $\bar{u}$ minimizes $\theta^\rho$, $x(\bar{u})$ is a primal optimal solution, and that there is no duality gap.

Augmented Lagrangians have the following nice property. Contrary to *standard* Lagrangian duality, *augmented* Lagrangian duality always zeroes the duality gap and recovers primal solutions (when $\rho$ is large enough). The aim of this exercise is to prove this property for (P) and $\rho = 3$.

**f)** Write the augmented Lagrangian and the augmented dual function $\theta^3$ (that is, $\theta^\rho$ for $\rho = 3$) associated to the dualization of $x_1 + x_2 - 3 = 0$ in problem (P). Show that $\theta^3$ can be cast as

$$\theta^3(u) = \max\{\theta_0^3(u), \theta_2^3(u)\}$$

with two concave functions that we denote by $\theta_0^3$ and $\theta_2^3$ (no need to develop them explicitly).

**g)** Show that $\theta^3(-1) = -5$.

**h)** Conclude that $\bar{u} = -1$ minimizes $\theta^3$ and that there is no duality gap.

**i)** Thus solving the augmented Lagrangian dual allows us to solve the primal problem! But there is no free lunch: what is the big disadvantage of augmented Lagrangian (versus the usual Lagrangian)?

**Exercise 10 – Max-cut.** Consider a undirect graph whose nodes are numbered from 1 to $n$ and edges have weights $w_{ij} \in \mathbb{R}$. We are interested in the max-cut problem (separating nodes into two groups such that the sum of the weights of the cut edges is maximum).

**a)** For each node, we associate: $x_i = 1$ if we put $i$ in the first group and $x_i = -1$ in the second group. Model the problem as a quadratic problem under the constraints ${x_i}^2 = 1$.

**b)** Apply the Lagrangian duality mechanism to write the dual problem. [Hint: you will need to introduce a constraint of the type $X \in \mathcal{S}_n^+$].

**c)** Observe that the dual problem is indeed convex. Show that the problem is non-degenerate, *i.e.* there exists $u \in \mathbb{R}^n$ such that $W/4 + \mathrm{Diag}(u)$ is positive definite

**d)** Using the result of **Exercise 8** write the dual of the dual problem. How does this bi-dual relate with the max-cut problem ?

**Exercise 11 – $\ell_\infty$-fitting as an LP.** Assume we have $m$ observations $(a_i, b_i) \in \mathbb{R}^n \times \mathbb{R}$, stored as a vector $b \in \mathbb{R}^m$ and a matrix $A \in \mathbb{R}^{n\times m}$ (with the ${a_i}^\top$'s as lines). We would like to compute $x \in \mathbb{R}^n$ such that $Ax - b$ is as small as possible for the $\ell_\infty$-norm; that is, to solve

$$\min_{x\in\mathbb{R}^n} \|Ax - b\|_\infty$$

where $\|u\|_\infty = \max\{|u_i|\,,\ i = 1, \ldots, m\}$ for $u \in \mathbb{R}^m$.

**a)** Show that this convex optimization problem can be cast as a linear optimization problem.

**b)** Explicit vectors and matrices $(c, G, h)$ to write this linear problem as the following form

$$\begin{cases} \min_u \ c^\top u \\ \quad Gu \leqslant h \end{cases}$$

so that we could solve the problem by using an off-the-shelf LP solver.

**c)** Assume moreover that the entries of $A$ and $b$ are all positive. Consider now the same problem but in logarithmic scale and with $x \geqslant 0$

$$\min_{x\in(\mathbb{R}_+)^n} \max_{i=1,\ldots,n} |\log({a_i}^\top x) - \log(b_i)|.$$

This problem can no longer be written as a linear problem, but as a conic optimisation problem. [Hint: positive semidefinite $2 \times 2$-matrices come into play...].

**Exercise 12 – Dantzig Selector.** We consider a regression model $y = A\theta + \xi$ where the noise is Gaussian $\xi \sim \mathcal{N}(0, \sigma I_m)$. The observations are $A \in \mathbb{R}^{m\times n}$ and $y \in \mathbb{R}^m$; $\theta \in \mathbb{R}^n$ is the unknown parameter we wish to estimate. In the over-parameterized case (i.e., when the size $n$ of $\theta$ is large compared to $m$, the size of $y$), the "Dantzig selector" consists in solving the optimization problem

$$\min_{\theta\in\mathbb{R}^n} \|\theta\|_1, \quad \text{subject to } \|A^\top(A\theta - y)\|_\infty \leqslant \kappa\sigma$$

where $\kappa > 0$ is a hyperparameter.

**a)** Let $e(\theta) = 1/2\|A\theta - y\|_2^2$ be the quadratic error of the model. Observe that $\nabla e(\theta) = A^\top(A\theta - y)$.

**b)** By introducing additional variables, reformulate this problem as a linear problem.

**c)** Construct the vectors and matrices $(c, G, h)$ to write this linear problem in canonical form (to be able to solve it later using an available solver)

$$\min_x \; c^\top x \quad \text{subject to } Gx \leqslant h$$

**Exercise 13 – Proof of Von Neumann in the case of matrix games.** Let $e \in \mathbb{R}^n$ be the vector of all ones $e = [1, \dots, 1]^\top$ and $\Delta = \{x \in (\mathbb{R}_+)^n,\ e^\top x = 1\}$ the simplex in $\mathbb{R}^n$. We consider a zero-sum matrix game with two players (P1 and P2) and a payoff matrix $A \in \mathbb{R}^{n \times n}$. Each player makes a choice between $n$ actions, randomly and independently, following their own mixed strategies ($x$ for P1 and $y$ for P2). The goal of P1 is to have the expected payoff $g(x, y) = x^\top Ay$ as large as possible while the goal of P2 is to have it as low as possible ($g_1 = g$ and $g_2 = -g$).

**a)** Recall what is a the mixed strategy. What is the interest of considering mixed strategies rather than pure strategies ? Recall what is the payoff matrix in the case of rock-paper-scissor.

**b)** Show that the min-max problem can be written as the following linear problem

$$\max_{x \in \Delta} \min_{y \in \Delta} x^\top Ay \quad \Longleftrightarrow \quad \begin{cases} \max_{t,x} \quad t \\ \quad x \geqslant 0,\ e^\top x = 1 \\ \quad A^\top x \geqslant te \end{cases}$$

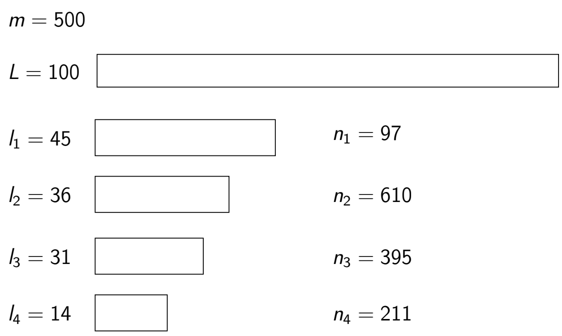
**c)** Apply Lagrangian duality to the above linear problem by dualizing two contraints: the constraint $e^\top x - 1 = 0$ with a first dual variable $\tau \in \mathbb{R}$, as well as the constraint $te - A^\top x \leqslant 0$ with a second dual variable $u \in (\mathbb{R}_+)^n$. [Keep the constraint $x \geqslant 0$; no need to dualize it.]

**d)** Show that the optimal values of the two following optimization problems are the same:

$$\begin{cases} \max_{t,x} \quad t \\ \quad x \geqslant 0,\ e^\top x = 1 \\ \quad A^\top x \geqslant te \end{cases} = \begin{cases} \min_{\tau,u} \quad \tau \\ \quad u \geqslant 0,\ e^\top u = 1 \\ \quad Au \leqslant \tau e \end{cases}$$

**e)** Show that this gives a proof of the Von Neumann theorem in the framework of this exercice.

**Exercise 14 – Lagrangian decomposition for cutting-stock.** The problem consists in minimizing the number of stock pieces of width $L$, used to meet demands $n_1, \dots, n_I$, for items $i = 1, \dots, I$, to be cut at their width $l_1, \dots, l_I$. We assume that every $l_j$ is smaller than $L$ and that there are enough stock pieces, say $m$, available for a feasible cutting. We denote by $n \in \mathbb{R}^I$ (respectively $l \in \mathbb{R}^I$) the vector of entries $n_i$ (resp. $l_i$) for all $i$. In the example drawn here: we have $m = 500$ pieces of width $L = 100$ where to cut $I = 4$ types of items; the demand consists in different numbers of items $n_i$ with different lengths $l_i \leqslant 100$ for the $I = 4$ types of items.



A possible formulation for the cutting-stock problem is the following integer linear problem:

$$(\mathrm{P})\quad \begin{cases} \min_{y,z} & \sum_{k=1}^{m} y^k \\ \sum_{k=1}^{m} z_i^k \geqslant n_i & \text{for all } i=1,\dots,I \\ \sum_{i=1}^{I} z_i^k\, l_i \leqslant L\, y_k & \text{for all } k=1,\dots,m \\ y^k \in \{0,1\},\ z_i^k \in \mathbb{N} & \text{for all } i=1,\dots,I,\ k=1,\dots,m \end{cases}$$

**a)** Explain the modelling as (P) : what is the role of the variables ? and the meaning of the objective and the constraints ?

**b)** Let us dualize the $I$ demand-covering constraints $\sum_{k=1}^{m} z_i^k \geqslant n_i$. Re-write the above problem as a max with the course's notation: introduce $\varphi$, $c$ and $X$.

**c)** For a dual variable $u \in (\mathbb{R}_+)^I$, define the Lagrangian function and show that it is decomposable with respect to $k$.

**d)** Observe then that the associated dual function, denoted by $\theta$, can be written as the juxtaposition of $m$ identical max problems, that is,

$$\theta(u) = -n^\top u + \sum_{k=1}^{m} v(u) = -n^\top u + m\, v(u)$$

where $v(u)$ is the optimal solution of a max problem to be specified.

**e)** Show moreover that $v(u)$ can be explicitly written as:

$$v(u) = \begin{cases} 0 & \text{if } u^\top z(u) \leqslant 1 \\ u^\top z(u) - 1 & \text{otherwise} \end{cases}$$

where $z(u)$ is the optimal solution of the following integer knapsack problem, parameterized by $u$

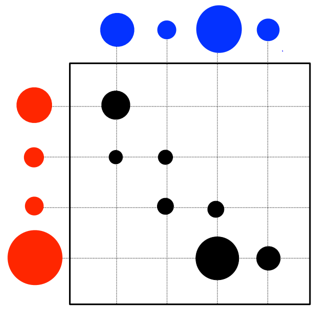
$$\begin{cases} \min & u^\top z \\ & l^\top z \leqslant L,\ z \in \mathbb{N}^I. \end{cases}$$

**f)** Discuss the complexity and the practical difficulty of computing $\theta(u)$, compared to solving (P). Explain what would be an "oracle" for $\theta$ providing a linearization of the (convex) function $\theta$.

**g)** What does the dual optimal value correspond to, in the problem (P) ? How does it compare with the continuous relaxation consisting in relaxing all the integrity constraints of (P) (i.e. $y^k \in [0,1]$, $z_i^k \in [0, M]$ with $M$ an upper bound).

**Exercise 15 – Optimal Transport.** Let $a \in \mathbb{R}_+^n$ and $b \in \mathbb{R}_+^m$ be two positive vectors such that $\sum_{i=1}^{n} a_i = 1$ $\sum_{j=1}^{m} b_j = 1$ (thus representing discrete probability densities). In the figure, the discrete We want to perform optimal transport from $a$ to $b$: we need to find a matrix $P = (P_{ij}) \in \mathbb{R}_+^{n\times m}$ that represents how each $a_i$ is distributed towards the $b_j$ given associated costs $C_{ij} \geqslant 0$. This problem is formulated as

$$W(a,b) = \begin{cases} \min & \sum_{i=1}^{n} \sum_{j=1}^{m} C_{ij} P_{ij} \\ \sum_{j=1}^{m}; P_{ij} = a_i, & \text{for all } i=1,\dots,n \\ \sum_{i=1}^{n}; P_{ij} = b_j, & \text{for all } j=1,\dots,m \\ P_{ij} \geqslant 0 & \text{for all } i=1,\dots,n \text{ and } j=1,\dots,m \end{cases}$$



distribution $a \in \mathbb{R}^4$ is in red and $b \in \mathbb{R}^4$ in blue. We have $n = m = 4$ with $a_4 \geqslant a_1 \geqslant a_2 \geqslant a_4$ and $b_3 \geqslant b_1 \geqslant b_4 \geqslant b_2$. The black dots represent the non-zero coefficients of $P$.

**a)** Consider now the dualization of all constraints on rows ($a_i - \sum_{j=1}^{m} P_{ij} = 0$ for all $i$) and columns ($b_j - \sum_{i=1}^{n} P_{ij} = 0$ for all $j$). Put the problem in the form given in the course, introduce the associated Lagrangian, and define the dual function. We will denote the dual variables $\lambda^a = (\lambda_i^a)i = 1,\dots,n \in \mathbb{R}^n$ and $\lambda^b = (\lambda^b j)j = 1,\dots,m \in \mathbb{R}^m$.

**b)** Show that there is no duality gap. Deduce that

$$W(a,b) = \left\{ \begin{array}{ll} \max & a^\top \lambda^a + b^\top \lambda^b \\ & \lambda_i^a + \lambda_j^b \leqslant C_{ij}, \quad \text{for all } i = 1, \dots, n \text{ and } j = 1, \dots, m \end{array} \right.$$

**Exercise 16 – Optimal Transport and Wasserstein Distance.** With the notation of the previous exercice, consider the case where $n = m$ and $C$ defines a distance on $\{1, \dots, n\}$, that is: $C_{i,j} = C_{j,i} \geqslant 0$ for all $i, j$; $C_{i,j} = 0$ if and only if $i = j$; $C_{ij} \leqslant C_{ik} + C_{kj}$ for all $i, j, k$ (triangle inequality). Denote $\Sigma_n = \{a \in \mathbb{R}_+^n : \sum_{i=1}^n a_i = 1\}$ the simplex of $\mathbb{R}^n$.

**a)** Observe that $W$ is positive and symetric on $\Sigma_n$. Show also that $W(a, b) = 0 \iff a = b$.

**b)** Fix $a, b, c \in \Sigma_n$; take $P$ and $Q$ optimal transport plans for $W(a, b)$ and $W(b, c)$ respectively. If $b_i > 0$ for all $i$, show that the matrix $S = P \operatorname{diag}(1/b_1, \dots, 1/b_n) Q$ satisfies $Se = a$ and $S^\top e = b$ where $e = (1, \dots, 1)^\top$ is the vector of all ones.

**c)** Deduce that we have $W(a, c) \leqslant W(a, b) + W(b, c)$, for all $a, b, c \in \Sigma_n$.

**d)** Conclude that $W$ is a distance on $\Sigma_n$; it is called the Wasserstein distance.

**Exercise 17 – Entropy-regularized Optimal Transport.** Let's come back to the optimal transport problem, to which we will add entropic regularization:

$$H(P) = \sum_{i=1}^{n} \sum_{j=1}^{m} P_{ij}(\log(P_{ij}) - 1) \qquad \text{(where log is the natural logarithm).}$$

We therefore consider the problem, with $\varepsilon > 0$,

$$\text{(P)} \qquad \min_{P \in \mathcal{U}(a,b)} \quad \sum_{i=1}^{n} \sum_{j=1}^{m} C_{ij} P_{ij} + \varepsilon\, H(P)$$

where $\mathcal{U}(a, b)$ is the set of transport plans from $a \in \mathbb{R}_+^n$ to $b \in \mathbb{R}_+^m$ (with $\sum_{i=1}^n a_i = 1$ et $\sum_{j=1}^m b_j = 1$).

**a)** Let $\varphi \colon \mathbb{R}_+ \to \mathbb{R}$ be a function defined and continuous on $\mathbb{R}_+$

$$\varphi(t) = \left\{ \begin{array}{ll} t \log(t) & \text{if } t > 0 \\ 0 & \text{if } t = 0. \end{array} \right.$$

Show that $\varphi$ is strictly convex on $\mathbb{R}_+^*$.

**b)** Show that $\varphi$ is in fact strictly convex on all of $\mathbb{R}_+$. [Hint: we can observe that, for $0 < \alpha < 1$ and $t > 0$, we have $\log(\alpha t) < \log(t)$.]

**c)** Deduce that the function $H \colon \mathbb{R}_+^{n \times m} \to \mathbb{R}$ is continuous and strictly convex. Show that there exists a unique solution to (P). Let's denote it $P_\varepsilon$.

**d)** By introducing the matrix $K = (K_{ij}) \in \mathbb{R}^{n \times m}$ defined by $K_{ij} = \exp(-C_{ij}/\varepsilon)$ for all $(i, j)$, rearrange the objective to show[1] that [Hint: use the fact that the sum of $P_{ij}$ is constant.]

$$P_\varepsilon = \underset{P \in \mathcal{U}(a,b)}{\operatorname{argmin}} \quad \sum_{i=1}^{n} \sum_{j=1}^{m} P_{ij} \log(P_{ij}/K_{ij}).$$

Let's now return to the initial problem (P) and study a dual approach to compute $P_\varepsilon$. We will dualize all equality constraints of $\mathcal{U}(a, b)$, but not the positivity constraints. The function $\varphi_\alpha \colon \mathbb{R}_+ \to \mathbb{R}$ defined for $\alpha \in \mathbb{R}$ by

$$\varphi_\alpha(t) = \left\{ \begin{array}{ll} \varepsilon\, t \log(t) + \alpha\, t & \text{if } t > 0 \\ 0 & \text{if } t = 0 \end{array} \right.$$

will appear in the developments.

[1] Cultural note: this means that $P_\varepsilon$ can be interpreted as the projection, in the sense of Kullback-Leibler divergence, of $K$ (called Gibbs kernel) onto $\mathcal{U}(a, b)$.

**e)** Reformulate (P) in the form given in class (changing the sign to have a max). Define the Lagrangian and the dual function $\theta$. We will denote the dual variables $\lambda^a \in \mathbb{R}^n$ and $\lambda^b \in \mathbb{R}^m$.

**f)** Show that

$$\theta(\lambda^a, \lambda^b) = -a^\top \lambda^a - b^\top \lambda^b \sum_{i=1}^{n} \sum_{j=1}^{m} \min_{P_{ij} \geqslant 0} \varphi_{\alpha_{ij}}(P_{ij})$$

for some $\alpha_{ij} \in \mathbb{R}$ that you will specify.

**g)** Calculate the minimum on $\mathbb{R}_+$ of the function $\varphi_\alpha$.

**h)** Deduce that

$$\theta(\lambda^a, \lambda^b) = -a^\top \lambda^a - b^\top \lambda^b \varepsilon \sum_{i=1}^{n} \sum_{j=1}^{m} \exp\left((-C_{ij} + \lambda_i^a + \lambda_j^b)/\varepsilon\right)$$

Compare with the dual of the non-regularized problem ($\varepsilon = 0$) seen in class. Interpret the impact of regularization on the dual.

**i)** Deduce that $\theta$ is differentiable and give the expressions for $\frac{\partial}{\partial \lambda_i^a}\theta(\lambda^a, \lambda^b)$ for all $i$, as well as $\frac{\partial}{\partial \lambda_j^b}\theta(\lambda^a, \lambda^b)$ for all $j$.

**j)** Show that the unique solution optimizing the Lagrangian, for fixed $(\lambda^a, \lambda^b)$, is

$$(P_{\lambda^a,\lambda^b})ij = \exp\left((-Cij + \lambda_i^a + \lambda_j^b)/\varepsilon\right) \qquad \text{for all } (i,j).$$

Rewrite the partial derivatives of $\theta$ at $(\lambda^a, \lambda^b)$ in terms of $P_{\lambda^a,\lambda^b}$.

**k)** Write the dual problem. What do you propose for solving it numerically?

**l)** Assuming we have the dual solutions $(\bar{\lambda}^a, \bar{\lambda}^b)$; show that $P_{\bar{\lambda}^a,\bar{\lambda}^b}$ is feasible. Deduce that there is no duality gap and that $P_\varepsilon = P_{\bar{\lambda}^a,\bar{\lambda}^b}$.

**m)** Deduce the classical expression of $P_\varepsilon$, with the matrix $K$ from question **e**:

$$P_\varepsilon = \operatorname{diag}(\exp(\bar{\lambda}^a/\varepsilon))K \operatorname{diag}(\exp(\bar{\lambda}^b/\varepsilon)).$$

Notation: for a vector $\lambda$, we denote by $\operatorname{diag}(\exp(\lambda))$ the diagonal matrix with coefficients $\exp(\lambda_i)$ on the diagonal.