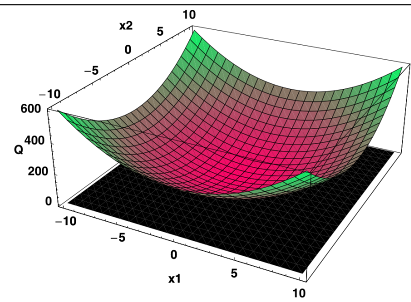
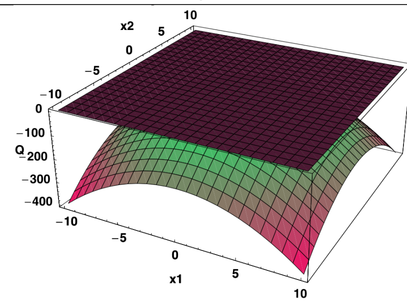
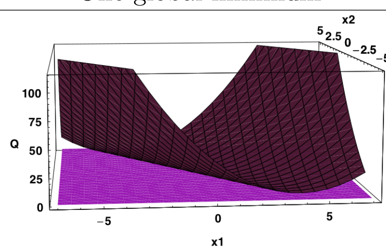
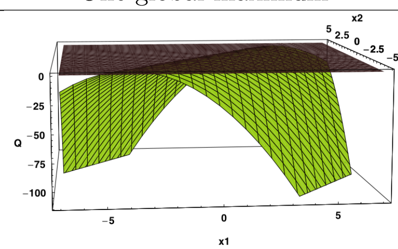
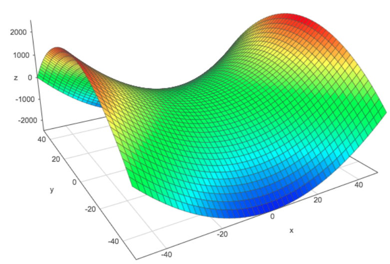
TD 2 - Optimization Correction

November 2023

Reminder for a symmetric real matrix $A \in \mathbb{R}^{n \times n}$.

	Positive	Negative
Definite	 $\forall x \in \mathbb{R}^n \setminus \{0\}, x^\top Ax > 0$ One global minimum	 $\forall x \in \mathbb{R}^n \setminus \{0\}, x^\top Ax < 0$ One global maximum
Semi-definite	 $\forall x \in \mathbb{R}^n, x^\top Ax \geq 0$ Multiple local minima	 $\forall x \in \mathbb{R}^n, x^\top Ax \leq 0$ Multiple local maxima
Indefinite	 $\exists x \in \mathbb{R}^n, x^\top Ax \geq 0$ & $\exists y \in \mathbb{R}^n, y^\top Ay \leq 0$ No maximum nor minimum (saddle point or minimax point)	

Exercise I Consider a linear problem in standard form

$$\min_{x \in \mathbb{R}^n} c^\top x$$

$$\text{such that } \begin{cases} Ax = b \\ x \succeq 0 \end{cases}$$

where $c \in \mathbb{R}^n$, $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$ are fixed. Find the lower bound for the problem.

Solution The problem P can be written as

$$\begin{aligned} &\underset{x\in\mathbb{R}^n}{min} \quad f(x) = c^\top x \\ &\text{such that } \begin{cases} h(x) \leqslant 0 & \text{with } h(x) = -x \\ g(x) = 0 & \text{with } g(x) = Ax - b \end{cases} \end{aligned}$$

Compute the Lagrangian :

$$\begin{aligned} L(x,\lambda,\mu) &= f(x) + \mu^\top g(x) + \lambda^\top h(x) \qquad\qquad \text{with } \mu \in \mathbb{R}^m \text{and } \lambda \in \mathbb{R}^n \\ &= c^\top x + \mu^\top (Ax - b) - \lambda^\top x \\ &= -\mu^\top b + (c^\top + \mu^\top A - \lambda^\top)x \\ &= -\mu^\top b + (c + A^\top \mu - \lambda)^\top x \end{aligned}$$

The dual problem is then :

$$\begin{aligned} l(\lambda,\mu) &= \underset{x}{min}\Big(L(x,\mu,\lambda)\Big) \\ &= -b^\top \mu + \underset{x}{min}\Big((c + A^\top \mu - \lambda)^\top x\Big) \\ &= \begin{cases} -b^\top \mu, & \text{if } c + A^\top \mu - \lambda = 0 \\ -\infty, & \text{otherwise} \end{cases} \end{aligned}$$

The best lower bound for the problem is $-b^\top \mu$ when $A^\top \mu - \lambda = 0$

Exercise II We consider a non-convex problem

$$\begin{aligned} &\min_{x\in\mathbb{R}^n} x^\top Ax \\ &\text{such that } x_i^2 = 1, \forall i = 1, ..., n \end{aligned}$$

where $A = A^T$, is given, but we do not assume that it is positive definite. Find the lower bound for the problem.

Solution The function to be minimised : $f(x) = x^\top Ax$

The constraint : $\forall i = 1, ..., n, x_i^2 = 1 \Leftrightarrow \begin{bmatrix} x_1^2 \\ x_2^2 \\ \vdots \\ x_n^2 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}$ which can be expressed as $x^2 = \overrightarrow{1}$

Compute the Lagrangian

$$\begin{aligned} L(x,\mu) &= x^\top Ax + \mu^\top (x^2 - \overrightarrow{1}) \\ &= (x^\top (A + diag(\mu))x - \mu^\top \overrightarrow{1} \end{aligned}$$

as $\mu^\top x^2 = \begin{bmatrix} \mu_1 & \mu_2 & ... & \mu_n \end{bmatrix} \begin{bmatrix} x_1^2 \\ x_2^2 \\ \vdots \\ x_n^2 \end{bmatrix} = \sum_{i=1}^n \mu_i x_i^2$ and

$$x\top diag(\mu)x = \begin{bmatrix} x_1 & x_2 & ... & x_n \end{bmatrix} \begin{bmatrix} \mu_1 & 0 & ... & 0 \\ 0 & \mu_2 & ... & 0 \\ \vdots & & \ddots & \vdots \\ 0 & ... & 0 & \mu_n \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} \mu_1 x_1 & \mu_2 x_2 & ... & \mu_n x_n \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \sum_{i=1}^n \mu_i x_i^2$$

The dual function is then

$$\begin{aligned} l(\mu) &= \min_x \; x^\top (A + diag(\mu))x - \mu^\top \overrightarrow{1} \\ &= \begin{cases} -\mu^\top \overrightarrow{1}, & \text{if } A + diag(\mu) \succeq 0 \\ -\infty, & \text{otherwise} \end{cases} \end{aligned}$$

The best lower bound for the problem is $-\mu^\top \overrightarrow{1}$ when $A + diag(\mu)$ is positive semi-definite.

Exercise III Show that calculating the projection of a point $x_0 \in \mathbb{R}^n$ on the hyperplane with equation $x^\top a = b$ where $a \in \mathbb{R}^n$ and $b \in \mathbb{R}$ are fixed, can be formulated as a constrained optimization problem. Give its solution.

Solution The hyperplane equation can be expressed as :

$$\mathcal{H} = \left\{ x \in \mathbb{R}^n : a^\top x - b = 0 \right\}$$

The distance between the point x_0 and its projection x^* on the hyperplan $\mathcal{H}$ is $\|x - x_0\|^2$

The constrained optimization problem can be expressed as :

$$\min_x \; f(x) = \frac{1}{2}\|x - x_0\|^2$$
$$\text{such that } g(x) = a^\top x - b = 0$$

(Note that the 1/2 factor in $f(x)$ is added for convenience when computing the gradient of the Lagrangian.)

The Lagragian is then :

$$L(x, \lambda) = \frac{1}{2}\|x - x_0\|^2 + \lambda(a^\top x - b) \qquad \text{with } \lambda \in \mathbb{R}$$

We solve :

$$\begin{cases} \nabla_x L(x, \lambda) = 0 \\ g(x) = 0 \end{cases}$$

$$\Leftrightarrow \begin{cases} x - x_0 + \lambda a = 0 \\ a^\top x - b = 0 \end{cases}$$

$$\Leftrightarrow \begin{cases} x = x_0 - \lambda a \\ a^\top (x_0 - \lambda a) - b = 0 \Rightarrow \lambda = \frac{a^\top x_0 - b}{\|a\|^2} \end{cases}$$

Then the optimal solution is

$$x^* = x_0 - \frac{a^\top x_0 - b}{\|a\|^2}.a$$

Exercise IV Find the dimensions (radius r and height h) of a cylindrical can for maximizing its volume such that its surface is given by S.

Solution A cylindrical can with radius r and height h has

- volum $V = \Pi.r^2.h$
- surface $S = 2\Pi rh + 2\Pi r^2$

The optimization problem can then be written as :

$$\max_{r,h} \; V = \Pi r^2 h$$
$$\text{such that } S = 2\Pi rh + 2\Pi r^2$$

a) Search for an optimal point

We compute the Lagrangian

$$L(r,h,\lambda) = \Pi r^2 h + \lambda(2\Pi rh + 2\Pi r^2 - S) \qquad \text{with } \lambda \in \mathbb{R}$$

We solve

$$P : \begin{cases} \nabla_r L = 2\Pi rh + 2\Pi\lambda h + 4\Pi\lambda r = 0 \\ \nabla_h L = \Pi r^2 + 2\Pi\lambda r = 0 \\ 2\Pi rh + 2\Pi r^2 = S \end{cases}$$

$$\Leftrightarrow \begin{cases} rh + \lambda h + 2\lambda r = 0 \\ r^2 + 2\lambda r = 0 \quad \Leftrightarrow \quad \lambda = -r/2 \\ 2\Pi rh + 2\Pi r^2 = S \end{cases}$$

$$\Leftrightarrow \begin{cases} rh - rh/2 - r^2 = 0 \quad \Leftrightarrow \quad h = 2r \\ \lambda = -r/2 \\ rh + r^2 = \frac{S}{2\Pi} \end{cases} \qquad \Leftrightarrow \begin{cases} \lambda = \frac{-1}{2}\sqrt{\frac{S}{6\Pi}} \\ h = 2\sqrt{\frac{S}{6\Pi}} \\ r = \sqrt{\frac{S}{6\Pi}} \end{cases}$$

We found an optimal point (r^*, h^*) and now we check that it is the global maximum.

b) Check that the optimal point is the global maximum

Without any constraint, we would have checked that $\forall x, \quad x^\top \cdot \nabla^2 f(x^*)x < 0$ (second derivative test).
With a constraint, we need to check that the Hessian matrix of the Lagrangian is negative definite at the tangent space of the optimal point, i.e. $\forall x \in \mathcal{T}(g)_{x^*}, \quad x^\top \cdot \nabla^2 L(x^*) \cdot x < 0$, with $\mathcal{T}(g)_{x^*}$ the tangent hyperplane to the constraint $g(x)$ hypersurface at x^*.
We compute the Hessian matrix of the Lagrangian :

$$\begin{aligned} \nabla^2 L(r,h) &= \begin{bmatrix} \frac{\partial^2 L}{\partial^2 r} & \frac{\partial^2 L}{\partial h \partial r} \\ \frac{\partial^2 L}{\partial r \partial h} & \frac{\partial^2 L}{\partial^2 h} \end{bmatrix} \\ &= \Pi \begin{bmatrix} 2h + 4\lambda & 2r + 2\lambda \\ 2r + 2\lambda & 0 \end{bmatrix} \\ \nabla^2 L(r^*, h^*) &= \Pi \begin{bmatrix} 2r^* & r^* \\ r^* & 0 \end{bmatrix} \qquad \text{(with } h^* = 2r^* \text{ and } \lambda = -r^*/2 \text{ at the optimal point)} \end{aligned}$$

We have the constraint $g(r,h) = 2\Pi rh + 2\Pi r^2 - S$.

$$\nabla g(r,h) = \begin{bmatrix} \frac{\partial g}{\partial r} \\ \frac{\partial g}{\partial h} \end{bmatrix} = 2\Pi \begin{bmatrix} h + 2r \\ r \end{bmatrix}$$

At the optimal point

$$\nabla g(r^*, h^*) = 2\Pi \begin{bmatrix} h^* + 2r^* \\ r^* \end{bmatrix} = 2\Pi \begin{bmatrix} 4r^* \\ r^* \end{bmatrix} = 2\Pi r^* \begin{bmatrix} 4 \\ 1 \end{bmatrix}$$

The tangent space $\mathcal{T}(g)_x$ to the constraint $g(x)$, is normal to the constraint gradient $\nabla g(x)$ at the optimal point x^*.
It can then be defined as :

$$\begin{aligned} \mathcal{T}(g)_{x^*} &= \left\{ z : \begin{bmatrix} 4 & 1 \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} = 0 \right\} \\ &= \left\{ z : 4z_1 + z_2 = 0 \right\} \\ &= \left\{ z : (z_1, -4z_1) \right\} \end{aligned}$$

Then we have, $\forall z \in \mathcal{T}(g)_{x^*}$,

$$\begin{aligned}
z^\top \cdot \nabla^2 L(x^*) \cdot z &= \Pi \begin{bmatrix} z_1 & -4z_1 \end{bmatrix} \begin{bmatrix} 2r^* & r^* \\ r^* & 0 \end{bmatrix} \begin{bmatrix} z_1 \\ -4z_1 \end{bmatrix} \\
&= \Pi z_1^2 \begin{bmatrix} 1 & -4 \end{bmatrix} \begin{bmatrix} 2r^* & r^* \\ r^* & 0 \end{bmatrix} \begin{bmatrix} 1 \\ -4 \end{bmatrix} \\
&= \Pi z_1^2 \begin{bmatrix} -2r^* & r^* \end{bmatrix} \begin{bmatrix} 1 \\ -4 \end{bmatrix} \\
&= -6\Pi r^* z_1^2 < 0
\end{aligned}$$

So the Hessian matrix of the Lagrangian is negative definite and the optimal point is indeed the maximum.

Exercise V We consider the entropy maximization problem

$$\min_{x \in \mathbb{R}^n} \sum_{i=1}^{n} x_i \log(x_i)$$

$$\text{such that } \begin{cases} Ax \preceq b \\ \sum_{i=1}^n x_i = \vec{1}^\top x = 1 \\ x_i \geqslant 0, & \forall i = 1, ..., n \end{cases}$$

where $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$ are fixed.

1. Define the primal problem for the entropy maximization problem
2. Solve the entropy maximization problem

Solution 1) Compute the Lagrangian

$$L(x, \lambda, \mu) = \sum_i x_i \log(x_i) + \lambda^\top (Ax - b) + \mu(\vec{1}^\top x - 1)$$

$$\begin{aligned}
l(\lambda, \mu) &= \min_x L(x, \lambda, \mu) \\
&= -\lambda^\top b - \mu + \min_x \left(\sum_i x_i \log(x_i) + (\vec{1}\mu + A^\top \lambda)^\top x \right) \\
&= -\lambda^\top b - \mu - e^{-\mu - 1} \sum_{i=1}^n e^{-a_i^\top \lambda}
\end{aligned}$$

The dual problem is then

$$\begin{cases} \max l(\lambda, \mu) & = -b^\top \lambda - \mu - e^{-\mu-1} \sum_{i=1}^n e^{-a_i^\top \lambda} \\ \lambda \succeq 0 \end{cases}$$

We can simplify the dual problem by maximizing over μ.

$$\Rightarrow \mu^* = \log \sum_{i=1}^n e^{-a_i^\top \lambda} - 1$$

The dual problem becomes then

$$\begin{cases} \max l(\lambda, \mu) & = \max \ -\lambda^\top b - \log \left(\sum_{i=1}^n e^{-a_i^\top \lambda} \right) \\ \lambda \succeq 0 \end{cases}$$

2) We assume that the weak form of Slater's condition holds, i.e.

$$\exists x \text{ such that } \begin{cases} Ax \preceq b \\ \vec{1}^\top x = 1 \end{cases}$$

Then the strong duality holds and an optimal solution (λ^*, μ^*) exists.

$$\Rightarrow L(x, \lambda^*, \mu^*) = \sum_i x_i \log(x_i) + \lambda^{*\top}(Ax - b) + \mu^*(\overrightarrow{1}^\top x - 1)$$

which is strictly convex.

$$\Rightarrow x_i^* = \frac{1}{exp(a_i^\top \lambda^* + \mu^* + 1)} \quad \forall i = 1, .., n$$

If x^* is primal feasible then it is optimal. Otherwise, the problem does not have optimal.