Optimization in Machine Learning
(ACM 40990)

Dr Lennon Ó Náraigh

Constrained Optimization #2

1. Consider the OP

$$\min(x+y) \qquad \text{subject to: } \begin{cases} c_1(\boldsymbol{x}) \geq 0, \\ c_2(\boldsymbol{x}) \geq 0, \end{cases}$$

where $c_1(\boldsymbol{x}) = 1 - x^2 - (y-1)^2$ and $c_2 = -y$. Show that the LICQ does not hold at $\boldsymbol{x}_* = (0,0)^T$.

We have $\nabla c_1 = -2x\mathbf{i} - 2(y-1)\mathbf{j}$. At $\boldsymbol{x}_* = (0,0)^T$ we have:

$$\nabla c_1 = 2\mathbf{j}. \tag{1a}$$

Also,

$$\nabla c_2 = -\mathbf{j}. \tag{1b}$$

The vectors in Equations (1a) and (1b) are not linearly independent, hence the LICQ does not hold at $\boldsymbol{x}_* = (0,0)^T$.

2. Consider the feasible set:

$$\Omega = \{\boldsymbol{x} \in \mathbb{R}^2 | y \geq 0,\, y \leq x^2\}.$$

   (a) For $\boldsymbol{x}_* = (0,0)^T$, write down $T_\Omega(\boldsymbol{x}_*)$ and $\mathcal{F}_\Omega(\boldsymbol{x}_*)$.

   (b) Is the LICQ satisfied at $\boldsymbol{x}_*$?

   (c) If the objective function is $f(\boldsymbol{x}) = -y$, verify that the KKT conditions are satisfied at $\boldsymbol{x}_*$.

   (d) Find a feasible sequence $\{\boldsymbol{z}_k\}_{k=0}^{\infty}$ approaching $\boldsymbol{x}_*$ with $f(\boldsymbol{z}_k) < f(\boldsymbol{x}_*)$, for all k.

We have $c_1(\boldsymbol{x}) = y$ and $c_2(\boldsymbol{x}) = x^2 - y$. Both constraints are active at $\boldsymbol{x}_* = (0,0)^T$.

(a) We therefore have $\nabla c_1 = \mathbf{j}$ and $\nabla c_2 = 2x\mathbf{i} - \mathbf{j}$. At $\boldsymbol{x}_*$, we have $\nabla c_2 = -\mathbf{j}$. Hence,

$$\mathcal{F}_\Omega(\boldsymbol{x}_*) = \left\{ \boldsymbol{d} \in \mathbb{R}^2 | \begin{array}{c} \nabla c_1 \cdot \boldsymbol{x} \geq 0, \\ \nabla c_2 \cdot \boldsymbol{x} \geq 0 \end{array}, \text{at } \boldsymbol{x}_* \right\}.$$

Hence, $d_2 \geq 0$ and $d_2 \leq 0$, hence $d_2 = 0$ and thus,

$$\mathcal{F}_\Omega(\boldsymbol{x}_*) = \{(d_1, 0) | d_1 \in \mathbb{R}\}.$$

For the tangent cone, we consider the regularized constraint $c_{2,\epsilon} = x^2 - y + \epsilon$, where $\epsilon > 0$ is a small positive parameter. Hence, on the boundary $c_{2\,\epsilon} = 0$ we have $y = x^2 + \epsilon$. As $x \to 0 x_* = 0$, we linearize the constraint $c_{2\epsilon}$: the linearized form of the constraint $c_{2,\epsilon} = 0$ is simply $y = \epsilon$. Feasible sequences then have the form

$$\boldsymbol{z}_k = \boldsymbol{x}_* + t_k \boldsymbol{d} + \boldsymbol{\delta}_k t_k,$$

where $\boldsymbol{\delta}_k$ is an error term with $\|\boldsymbol{\delta}_k\| \to 0$ as $k \to \infty$. By inspection of the Figure 1, $\boldsymbol{d} = (d_1, d_2)$, where d_1 is arbitrary and $0 \leq d_2 \leq \epsilon$. We take $\epsilon \downarrow 0$ to get:

$$T_\Omega(\boldsymbol{x}_*) = \{(d_1, 0) | d_1 \in \mathbb{R}\}.$$

(b) By direct calculation, we have:

$$\begin{aligned} \nabla c_1(\boldsymbol{x}_*) &= \mathbf{j}, \\ \nabla c_2(\boldsymbol{x}_*) &= -\mathbf{j}, \end{aligned}$$

these are not linearly independent, so the LICQ does not hold.

Remark: From class notes, we know that:

$$\text{LICQ} \implies T_\Omega(\boldsymbol{x}_*) = \mathcal{F}_\Omega(\boldsymbol{x}_*).$$

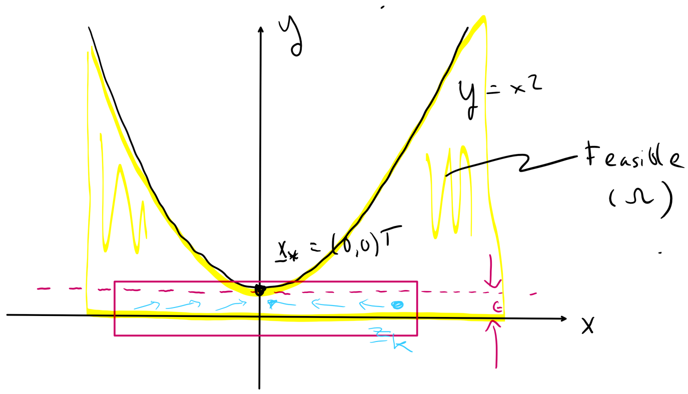

Figure 1: Construction of the tangent cone at $\boldsymbol{x}_* = (0,0)^T$

The contrapositive statement is:

$$\text{LICQ does not hold} \Longleftarrow T_\Omega(\boldsymbol{x}_*) \neq \mathcal{F}_\Omega(\boldsymbol{x}_*).$$

These are the only two statements we can be sure about *a priori*. So, just because the LICQ does not hold, that does not tell us anything about $T_\Omega(\boldsymbol{x}_*)$ and $\mathcal{F}_\Omega(\boldsymbol{x}_*)$.

(c) We have $f(\boldsymbol{x}) = -y$, so

$$\mathcal{L} = -y - \lambda_1 y - \lambda_2(x^2 - y).$$

The KKT conditions here are:

$$\begin{cases} \nabla_x \mathcal{L}(\boldsymbol{x}_*, \lambda_1^*, \lambda_2^*) = 0, \\ \text{No Equality Constraints} \\ c_1(\boldsymbol{x}_*) \geq 0,\ c_2(\boldsymbol{x}_*) \geq 0 \\ \lambda_1^* \geq 0, \lambda_2^* \geq 0, \\ \lambda_1^* c_1(\boldsymbol{x}_*) = 0,\ \lambda_2^* c_2(\boldsymbol{x}_*) = 0. \end{cases}$$

We have $\nabla_x \mathcal{L} = 0$, hence,

$$0 = \frac{\partial \mathcal{L}}{\partial x} = -2\lambda_2 x, \qquad 0 = \frac{\partial \mathcal{L}}{\partial y} = (-\lambda_1 - 1) + \lambda_2 \tag{2}$$

KKT2 is satisfied automatically. Both constraints are active, so KKT3 is satisfied, and so is KKT5. We therefore solve for λ_1^* and λ_2^* in Equation (2) to verify KKT4.

From Equation (2) we have $\lambda_2 x = 0$ and $x = x_* = 0$, hence λ_2 is undetermined. From the same equation, we have $\lambda_1 + 1 = \lambda_2$. As the LICQ is not satisfied, the Lagrange multipliers are not necessarily unique. So the valid (non-unique) Lagrange multipliers satisfying KKT 1-5 are:

$$(\lambda_1^*, \lambda_2^*) = \{(\lambda_1, \lambda_2) | \lambda_1 \geq 0, \lambda_2 = \lambda_1 + 1\}.$$

(d) By inspection, we consider a curve

$$\vec{(}\alpha) = (\alpha, \alpha^2)^T,$$

which is on the boundary of Ω satisfying $c_2(\boldsymbol{x}) = 0$ and $c_1(\boldsymbol{x} \geq 0)$. We introduce:

$$\widetilde{f}(\alpha) = f(\boldsymbol{x}(\alpha)) = y(\alpha) = -\alpha^2.$$

We have $f(\alpha) < 0$ for all $\alpha \neq 0$. A feasible sequence $\boldsymbol{z}_k$ approaching $\boldsymbol{x}_*$ with $f(\boldsymbol{z}_k) > f(\boldsymbol{x}_*)$ is therefore:

$$\boldsymbol{z}_k = \boldsymbol{x}(\alpha_k), \qquad \alpha_k = \pm 1/k, \qquad k \in \{1, 2, \cdots\}.$$

See Figure 2.

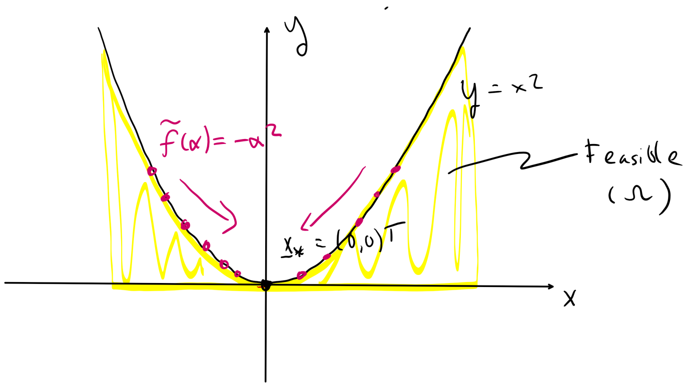

Figure 2: Construction of feasible sequences $\boldsymbol{z}_k$ such that $\boldsymbol{z}_k \to \boldsymbol{x}_* = (0,0)^T$ as $k \to \infty$

3. Consider the half-space defined by:

$$H_\alpha = \{\boldsymbol{x} \in \mathbb{R}^n | \langle \boldsymbol{a}, \boldsymbol{x} \rangle + \alpha \geq 0\},$$

where $\boldsymbol{a} \in \mathbb{R}^n$ is a constant non-zero vector and $\alpha \in \mathbb{R}$ is a constant scalar. Formulate and solve the OP for finding the point $\boldsymbol{x} \in H_\alpha$ with the smallest Euclidean norm.

The OP to minimize is:

$$\min f(\boldsymbol{x}), \qquad f(\boldsymbol{x}) = \tfrac{1}{2}\sum_{i=1}^{n} x_i^2,$$

subject to $c_1(\boldsymbol{x}) geq 0$, where

$$c_1(\boldsymbol{x}) = \sum_{i=1}^{n} a_i x_i + \alpha.$$

As such, we introduce the Lagrangian

$$\mathcal{L} = \tfrac{1}{2}\sum_{i=1}^{n} x_i^2 - \lambda \left(\sum_{i=1}^{n} a_i x_i + \alpha \right).$$

We have $\partial\mathcal{L}/\partial x_i = x_i - \lambda a_i$. We therefore have:

- KKT1: $x_i - \lambda a_i = 0$.
- KKT2: No equality constraints.
- KKT3: $\sum_i x_i a_i + \alpha \geq 0$.
- KKT4: $\lambda \geq 0$.
- KKT5: $\lambda \left(\sum_i x_i a_i + \alpha \right) = 0$.

Thus,

- KKT1 gives $x_i = \lambda a_i$.
- KKT3 gives $\sum_i a_i x_i + \alpha \geq 0$, hence:

$$\lambda \sum_i a_i^2 + \alpha \geq 0.$$

- KKT5 therefore becomes:

$$\lambda \left(\lambda \sum_i a_i^2 + \alpha \right) \geq 0.$$

So a solution is:

$$\lambda = \begin{cases} -\alpha / \sum a_i^2, & \alpha < 0 \text{ (Active constraint)}, \\ 0, & \alpha > 0 \text{ (Inactive constraint)}. \end{cases}$$

When $\lambda = 0$ we have $\boldsymbol{x} = 0$ (Case 1). When $\lambda \neq 0$ we have:

$$\boldsymbol{x} = \frac{\alpha \boldsymbol{a}}{\sum_i a_i^2} \text{ (Case 2).}$$

This makes geometric sense: Case 1 is illustrated in Figure 3. and here $\boldsymbol{x}_*$ is the shortest distance between the line (plane) $\langle \boldsymbol{x}, \boldsymbol{a} \rangle + \alpha = 0$ and the origin. In contrast, Case 2 is illustrated in Figure 4. Now, the origin is in the feasible set, so the feasible vector of shortest distance is the zero vector.

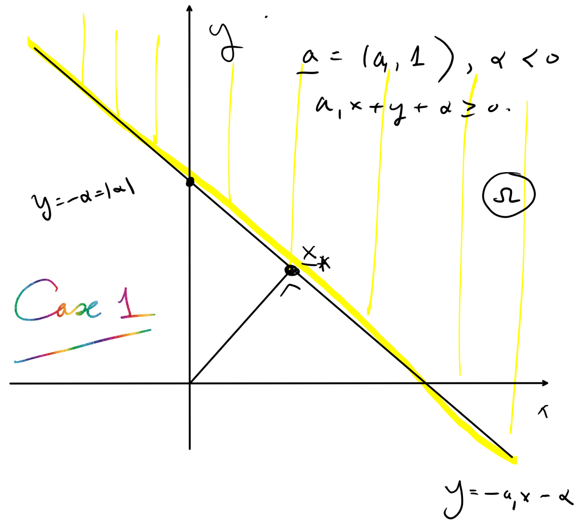

Figure 3: Simple illustration of Case 1 in 2D for the constraint equation $a_1x+y+\alpha=0$

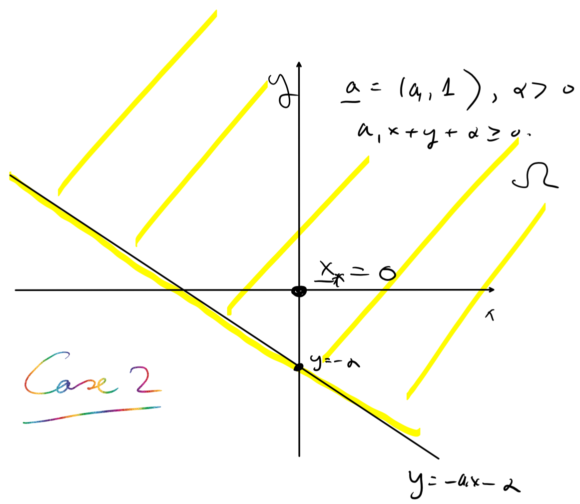

Figure 4: Simple illustration of Case 2 in 2D for the constraint equation $a_1x+y+\alpha=0$

4. Consider the following modification of the example in class notes. Her, t is a parameter that is fixed prior to solving the problem:

$$\min_{\boldsymbol{x}\in\mathbb{R}^2} f(\boldsymbol{x}),$$

where

$$f(\boldsymbol{x}) = \left(x - \tfrac{3}{2}\right)^2 + (y-t)^4,$$

subject to:

$$\begin{bmatrix} 1-x-y \\ 1-x+y \\ 1+x-y \\ 1+x+y \end{bmatrix} \geq 0.$$

(a) For what values of t does the point $\boldsymbol{x}_* = (1,0)^T$ satisfy the KKT conditions?

(b) Show that when $t=1$, only the first constraint is active at the solution and find the solution.

We have:

$$\mathcal{L} = \left(x - \tfrac{3}{2}\right)^2 + (y-t)^4 - \lambda_1(1-x-y) - \lambda_2(1-x+y) - \lambda_3(1+x-y) - \lambda_4(1+x+y).$$

Thus,

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial x} &= 2(x-3/2) + \lambda_1 + \lambda_2 - \lambda_3 - \lambda_4, \\ \frac{\partial \mathcal{L}}{\partial y} &= 4(y-t)^3 + \lambda_1 - \lambda_2 + \lambda_3 - \lambda_4. \end{aligned}$$

We solve $\nabla\mathcal{L}(\boldsymbol{x}_*) = 0$, where $\boldsymbol{x}_* = (0,0)^T$. KKT1 then becomes:

$$\text{KKT1}: \begin{cases} \lambda_1 + \lambda_2 - \lambda_3 - \lambda_4, & = 1, \\ \lambda_1 - \lambda_2 + \lambda_3 - \lambda_4, & = -4(-t)^3. \end{cases}$$

KKT5 gives: $\lambda_i c_i = 0$, hence:

$$\lambda_1 \times 0 = 0, \qquad \lambda_2 \times 0 = 0, \qquad \lambda_3 = 0, \qquad \lambda_4 = 0.$$

So only c_1 and c_2 are active. Hence:

$$\begin{aligned} \lambda_1 + \lambda_2 &= 1, \\ \lambda_1 - \lambda_2 &= -4(-t)^3, \end{aligned}$$

hence $2\lambda_1 = 1 - 4(-t)^3$. We require $\lambda_1 \geq 0$.

- If $t \geq 0$ we are fine, as then $2\lambda_1 = 1 - (-1)^3 t^3 \geq 0$.

- If $t \leq 0$, we have $-t = |t|$, and we require $2\lambda 1 = 1 - 4|t|^3 \geq 0$, hence $|t| \leq 1/4^{1/3}$,

So overall we require:

$$t \geq -\tfrac{1}{4^{1/3}}.$$

Furthermore,

$$\lambda_2 = 1 - \lambda_1,$$

hence

$$\begin{aligned} \lambda_2 &= 1 - \lambda_1, \\ &= 1 - \left[\tfrac{1}{2} - \tfrac{1}{2}2(-t)^3\right], \\ &= \tfrac{1}{2} + 2(-t)^3. \end{aligned}$$

We also require $\lambda_2 \geq 0$, which by the same reasoning as before gives $t \leq 1/4^{1/3}$ so overall, we require:

$$-\tfrac{1}{4^{1/3}} \leq t \leq \tfrac{1}{4^{1/3}}.$$

For part (ii) we first use an **elementary method**. Using geometric reasoning, we guess that the solution is $x_* \in L_1$, where L_1 is the line $y = 1 - x$. We have (with $t = 1$):

$$\begin{aligned} \widetilde{f}(x) &= f(x, y = 1 - x), \\ &= \left(x - \tfrac{3}{3}\right)^2 + x^4, \\ &= x^4 + x^2 - 3x + \tfrac{9}{4}. \end{aligned}$$

A plot of $\widetilde{f}(x)$ reveals a minimum x_* less than one (Figure 5). Using ordinary calculus, the minimum must satisfy $\widetilde{f}'(x) = 0$, hence

$$2x^3 + x - \tfrac{3}{2} = 0. \tag{3}$$

Using a numerical method (e.g. Wolfram Alpha), we obtain a the minimum x_*:

$$x_* \approx 0.728.$$

We now show compute the minimum using the **KKT conditions**. We solve $\nabla_x \mathcal{L} = 0$. We have:

$$\begin{aligned} 2(x - 3/2) + \lambda_1 + \lambda_2 - \cancel{\lambda_3} - \cancel{\lambda_4} &= 0, \\ 4(y - 1)^3 + \lambda_1 - \lambda_2 + \cancel{\lambda_3} - \cancel{\lambda_4} &= 0. \end{aligned}$$

Introduce $X = x - 3/2$ and $Y = y - 1$. So we have:

$$\begin{aligned} \lambda_1 + \lambda_2 &= -2X, \\ \lambda_1 - \lambda_2 &= -4Y^3. \end{aligned}$$

Hence,

$$\lambda_1 = -X - 2Y^3, \qquad \lambda_2 = -X + 2Y^3.$$

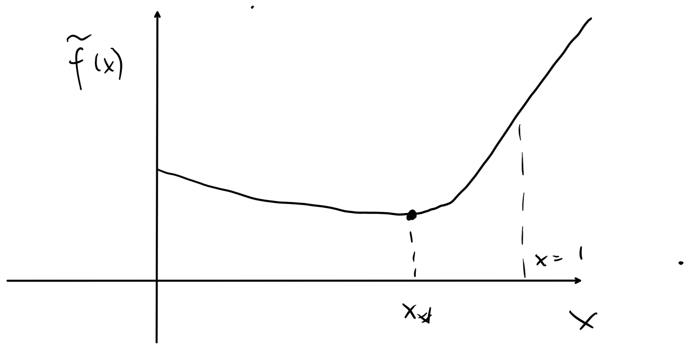

Figure 5: Plot of $\widetilde{f}(x)$ on the interval $[0, 1]$

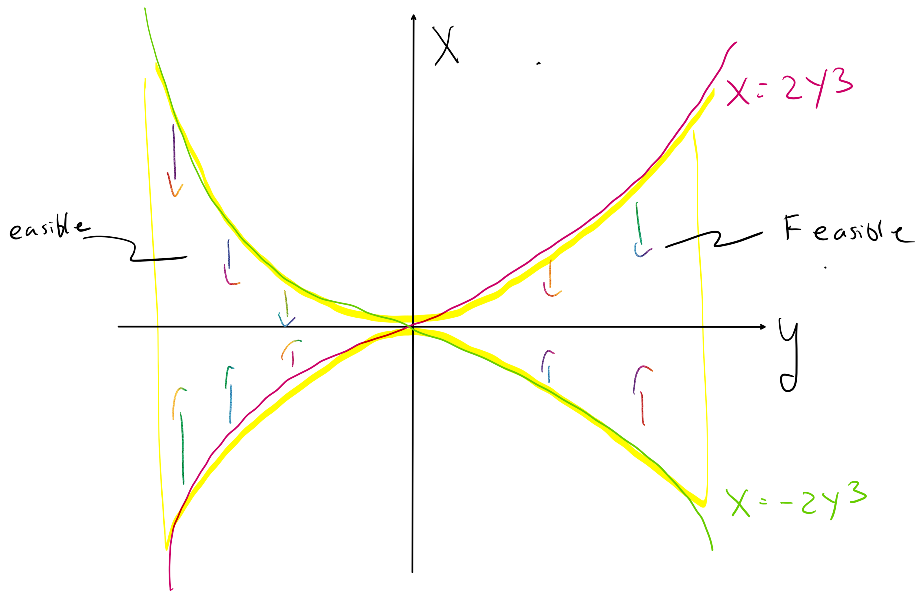

Figure 6: Allowed region where X and Y satisfy $X \leq -2Y^3$ and $X \leq 2Y^3$.

We require $\lambda_1 \geq 0$ and $\lambda_2 \geq 0$, hence

$$X \leq -2Y^3, \qquad X \leq 2Y^3.$$

The allowed region is shown in Figure 6 Overall therefore, $X \leq -2|Y|^3$.

We now look at the complementarity conditions, starting with CC1:

$$\begin{aligned}
\lambda(1 - x - y) &= 0, \\
\lambda_1 \left[-x - (y - 1)\right] &= 0, \\
\lambda_1 \left[-(x - 3/2) - (y - 1) - 3/2\right] &= 0, \\
\lambda_1 \left[-X - Y - 3/2\right] &= 0.
\end{aligned}$$

But $\lambda_1 = -X - 2Y^3$, hence:

$$(x + 2Y^3)\,(X_Y + 3/2) = 0, \qquad \text{CC1}.$$

Similarly, we look at CC2:

$$\begin{aligned}
\lambda_1(1-x+y) &= 0,\\
\lambda_2\left[-(x-1-1/2)-1/2+(y-1)+1\right] &= 0,\\
\lambda_2\left[-X+Y+1/2\right] &= 0.
\end{aligned}$$

But $\lambda_2=\lambda_2=-X+2Y^3$, hence:

$$\left(2Y^3-X\right)(X-Y-1/2)=0, \qquad \text{CC2.}$$

We look at a particular solution for CC1:

$$X+Y+3/2=0,$$

hence $y=1-x$. Then, to make CC2 hold, we require:

$$2Y^3=X,$$

hence:

$$\begin{aligned}
2Y^3 &= X,\\
2(y-1)^3 &= x-3/2,\\
2(1-x-1)^3 &= x-3/2,\\
2(-x)^3 &= x_3/2,\\
2x^3+x-3/2=0.
\end{aligned}$$

This is exactly Equation (3), so the minimum is at

$$(x_*, 1-x_*), \qquad x_*\approx 0.728,$$

which is the same answer we got using the elementary method.

5. Solve the OP in Question 4 (part (ii)) numerically, using Matlab or Python. Compare your answer with the answer obtained previously.