MATH60005/70005: Optimisation (Autumn 23-24)

Week 9: Problem Session

Dr Dante Kalise, Dr Estefanía Loayza-Romero and Sara Bicego
Department of Mathematics
Imperial College London, United Kingdom
{dkaliseb,kloayzar,s.bicego21}@imperial.ac.uk

1. Solve the problem

$$\begin{aligned} \min \quad & x_1^2 + 2x_2^2 + 4x_1x_2 \\ \text{s.t.} \quad & \mathbf{x} \in \Delta_2 \,. \end{aligned}$$

2. **Orthogonal regression**. Suppose we have $\mathbf{a}_1, \dots, \mathbf{a}_m \in \mathbb{R}^n$. For a given $\mathbf{0} \neq \mathbf{x} \in \mathbb{R}^n$ and $y \in \mathbb{R}$, we define the hyperplane:

$$H_{\mathbf{x},y} := \left\{\mathbf{a} \in \mathbb{R}^n : \mathbf{x}^\top \mathbf{a} = y\right\}$$

In the orthogonal regression problem, we seek to find a nonzero vector $\mathbf{x} \in \mathbb{R}^n$ and $y \in \mathbb{R}$ such that the sum of squared Euclidean distances between the points $\mathbf{a}_1, \dots, \mathbf{a}_m$ to $H_{\mathbf{x},y}$ is minimal:

$$\min_{\mathbf{x},y} \left\{ \sum_{i=1}^m d\left(\mathbf{a}_i, H_{\mathbf{x},y}\right)^2 : \mathbf{0} \neq \mathbf{x} \in \mathbb{R}^n, y \in \mathbb{R} \right\} .$$

Let $\mathbf{A}$ be the matrix

$$\mathbf{A} = \begin{bmatrix} \mathbf{a}_1^\top \\ \mathbf{a}_2^\top \\ \vdots \\ \mathbf{a}_m^\top \end{bmatrix}$$

Show the optimal solution of the orthogonal regression problem is given by $\mathbf{x}$ which is an eigenvector of the matrix $\mathbf{A}^\top(\mathbb{I}_m - \frac{1}{m}\mathbf{e}\mathbf{e}^\top)\mathbf{A}$ associated with the minimum eigenvalue and $y = \frac{1}{m}\sum_{i=1}^m \mathbf{a}_i^\top \mathbf{x}$.

3. Consider the problem

$$\begin{aligned} \min \quad & x_1^2 - x_2 \\ \text{s.t.} \quad & x_2 = 0 \,, \end{aligned}$$

and its equivalent formulation

$$\begin{array}{ll} \min & x_1^2 - x_2 \\ \text{s.t.} & x_2^2 \leq 0\,. \end{array}$$

Determine KKT conditions for both problems, are they equivalent and solvable?

Solutions

1. Since $\mathbf{x} \in \Delta_2$, we our problem reads

$$\min_{\mathbf{x}} \left\{ f(\mathbf{x}) := x_1^2 + 2x_2^2 + 4x_1x_2 \right\} \quad \text{s.t.} \quad \begin{cases} x_1 + x_2 = 1 & (x_1 + x_2 - 1 = 0) \\ x_1 \geq 0 & (-x1 \leq 0) \\ x_2 \geq 0 & (-x2 \leq 0) \end{cases}$$

By KKT condition for this Linearly Constrained Problem, if $\mathbf{x}^*$ is a local minimizer of $f(\mathbf{x})$ over Δ_2, then there exist $\lambda_1, \lambda_2 \geq 0$ and $\mu \in \mathbb{R}$ such that

$$\begin{cases} \nabla_x \mathcal{L} = 0 \\ \lambda_i(-x_i) = 0, \quad i = 1, 2 \\ x_1 + x_2 = 1 \end{cases} \qquad \text{for the Lagrangian} \quad \mathcal{L}(\mathbf{x}, \lambda, \mu) := f(\mathbf{x}) + \sum_{i=1}^{2} \lambda_i(-x_i) + \mu(x_1 + x_2 - 1)\,.$$

Our objective function is quadratic, as

$$f(\mathbf{x}) = \mathbf{x}^\top A\mathbf{x} = \mathbf{x}^\top \begin{bmatrix} 1 & 2 \\ 2 & 2 \end{bmatrix} \mathbf{x}\,,$$

with $Tr(A) = 3$ and $Det(A) = -2$. This implies that f is not convex, hence the KKT condition are only necessary.

The associated KKT system is $\begin{cases} 2x_1 + 4x_2 - \lambda_1 + \mu = 0 \\ 4x_2 + 4x_1 - \lambda_2 + \mu = 0 \\ \lambda_1 x_1 = 0 \\ \lambda_2 x_2 = 0 \\ x_1 + x_2 = 1 \end{cases}$ which we address by considering the following 4 cases.

- **Case $\lambda_1 = \lambda_2 = 0$:** the KKT system becomes

$$\begin{cases} 2x_1 + 4x_2 + \mu = 0 & (1) \\ 4x_2 + 4_x 1 + \mu = 0 & (2) \\ x_1 + x_2 = 1 & (3) \end{cases}$$

  and by considering $(2) - (1)$ we obtain $2x_1 = 0 \implies x_1 = 0$, and so $x_2 = 1$, $\mu = -4$. Thus, $(0, 1)$ is a KKT point.

- **Case $\lambda_1, \lambda_2 > 0$:** we need $x_1 = x_2 = 0$ which is unfeasible (it violates the last condition).
- **Case $\lambda_1 > 0, \lambda_2 = 0$:** we have $x_1 = 0$ which for feasibility implies $x_2 = 1$, then

$$\begin{cases} 4 + \mu - \lambda_1 = 0 \\ 4 + \mu = 0 \end{cases} \implies \begin{cases} \mu = -4 \\ \lambda_1 = 0 \end{cases} \qquad \text{and so } (0,1) \text{ solves the system.}$$

- **Case $\lambda_1 = 0, \lambda_2 > 0$:** we obtain the KKT point $(1, 0)$.

For optimality, we need to compare $f(0,1)$ and $f(1,0)$:

$$\begin{aligned} f(0,1) &= 0 + 2 + 0 = 2 \\ f(1,0) &= 1 + 0 + 0 = 1 \end{aligned} \qquad f(0,1) < f(1,0) \implies (0,1) \text{ is a local minimum.}$$

2. First, we need to find an explicit expression for $d\left(\mathbf{a}_i, H_{\mathbf{x},y}\right) = \|\mathbf{a}_i - \mathbb{P}_{H_{\mathbf{x},y}}(\mathbf{a}_i)\|$, where $\mathbb{P}_{H_{\mathbf{x},y}}(\mathbf{a}_i)$ is the orthogonal projection of $\mathbf{a}_i$ onto the hyperplane $H_{\mathbf{x},y}$. During the lectures, we have computed the orthogonal projection onto affine spaces $\mathbf{C} = \{\mathbf{v} \in \mathbb{R}^n \colon \mathbf{A}\mathbf{v} = \mathbf{b}\}$ given by

$$\mathbb{P}_{\mathbf{C}}(\mathbf{z}) = \mathbf{z} - \mathbf{A}^\top(\mathbf{A}\mathbf{A}^\top)^{-1}(\mathbf{A}\mathbf{z} - \mathbf{b})\,.$$

Since a hyperplane is a particular case of an affine space with $A = x^\top$ and $\mathbf{b} = y$, we obtain

$$\begin{aligned} \mathbb{P}_{H_{\mathbf{x},y}}(\mathbf{a}) &= \mathbf{a} - \mathbf{x}(\mathbf{x}^\top\mathbf{x})^{-1}(\mathbf{x}^\top\mathbf{a} - y)\,, \\ &= \mathbf{a} - \frac{(\mathbf{x}^\top\mathbf{a} - y)}{\|\mathbf{x}\|^2}\,\mathbf{x}\,. \end{aligned}$$

Altogether, implies

$$\begin{aligned} d\left(\mathbf{a}_i, H_{\mathbf{x},y}\right) &= \left\| \mathbf{a}_i - \mathbf{a}_i + \frac{(\mathbf{x}^\top\mathbf{a}_i - y)}{\|\mathbf{x}\|^2}\mathbf{x} \right\|\,, \\ &= \frac{|\mathbf{x}^\top\mathbf{a}_i - y|}{\|\mathbf{x}\|}\,, \end{aligned}$$

and the orthogonal regression problem is given by

$$\min \left\{ \sum_{i=1}^m \frac{(\mathbf{x}^\top\mathbf{a} - y)^2}{\|\mathbf{x}\|^2} : \mathbf{x} \neq \mathbf{0}\; y \in \mathbb{R} \right\}\,.$$

If we fix $\mathbf{x}$, then the optimiser with respect to y is given by

$$y = \frac{1}{m}\sum_{i=1}^m \mathbf{a}_i^\top\mathbf{x} = \frac{1}{m}\mathbf{e}^\top\mathbf{A}\mathbf{x}\,,$$

where $\mathbf{e} = [1, 1, \ldots, 1]^\top$.

Using the above expression,we obtain

$$
\begin{aligned}
\sum_{i=1}^{m}(\mathbf{a}_i^\top \mathbf{x} - y) &= \sum_{i=1}^{m}\left(\mathbf{a}_i^\top \mathbf{x} - \frac{1}{m}\mathbf{e}^\top \mathbf{A}\mathbf{x}\right)^2 , & (1)\\
&= \sum_{i=1}^{m}(\mathbf{a}_i^\top \mathbf{x})^2 - \frac{2}{m}\sum_{i=1}^{m}(\mathbf{e}^\top \mathbf{A}\mathbf{x})(\mathbf{a}_i^\top \mathbf{x}) + \frac{1}{m}(\mathbf{e}^\top \mathbf{A}\mathbf{x})^2 , & (2)\\
&= \|\mathbf{A}\mathbf{x}\|^2 - \frac{1}{m}(\mathbf{e}^\top \mathbf{A}\mathbf{x})^2 , & (3)\\
&= \mathbf{x}\mathbf{A}^\top \left(\mathbb{I}_m - \frac{1}{m}\mathbf{e}\mathbf{e}^\top\right)\mathbf{A}\mathbf{x} , & (4)
\end{aligned}
$$

and therefore the problem can be reformulated as

$$
\min_{\mathbf{x}} \left\{ \frac{\mathbf{x}^\top \left[\mathbf{A}^\top \left(\mathbb{I}_m - \frac{1}{m}\mathbf{e}\mathbf{e}^\top\right)\mathbf{A}\right]\mathbf{x}}{\|\mathbf{x}\|^2} : \mathbf{x} \neq \mathbf{0} \right\} ,
$$

whose optimal solution corresponds to the eigenvector associated with the smallest eigenvalue of the matrix $\mathbf{A}^\top \left[\mathbb{I}_m - \frac{1}{m}\mathbf{e}\mathbf{e}^\top\right]\mathbf{A}$.

3. We start with the solution via KKT of

$$
\min x_1^2 - x_2 \quad \text{s.t.} \quad x_2 = 0
$$

which is a minimization of a convex cost with convex constraints, hence KKT conditions are both necessary and sufficient:

$$
\mathcal{L}(\mathbf{x}, \mu) = x_1^2 - x_2 + \mu x_2 , \qquad \nabla_x \mathcal{L} = 0 \implies \begin{cases} 2x_1 &= 0 \\ \mu - 1 &= 0 \end{cases}
$$

from which we can conclude that $(0, 0)$ is the only KKT point (and minimizer).

In the alternative formulation

$$
\min x_1^2 - x_2 \quad \text{s.t.} \quad x_2^2 \leq 0 ,
$$

we have again convex cost and convex constraint, but the Slater's condition is not satisfied, as there is no $x_2 \in \mathbb{R}$ such that $x_2^2 < 0$. Then, KKT condition are only sufficient. For the associated Lagrangian

$$
\mathcal{L}(\mathbf{x}, \lambda) = x_1^2 - x_2 + \lambda x_2^2 ,
$$

we have

$$
\nabla_x \mathcal{L} = 0 \iff \begin{cases} 2x_1 = 0 \\ 2\lambda x_2 - 1 = 0 \\ \lambda x_2^2 = 0 \end{cases}
$$

but since the last two equations cannot be satisfied simultaneously, the KKT system has no feasible solution, even though the problem has a feasible optimal solution at $x_1 = x_2 = 0$.

Constrained Least Squares (extra exercise)

An application of this framework can be found in reformulating the (RLS) problem

$$\min_{\mathbf{x}} \|A\mathbf{x} - b\|^2 + \lambda \|\mathbf{x}\|^2$$

as a Constrained Least Squares (CLS) problem

$$\min_{\mathbf{x}} \|A\mathbf{x} - b\|^2 \qquad \text{s.t.} \qquad \|\mathbf{x}\|^2 \leq \alpha, \qquad \alpha > 0$$

which has convex cost and nonlinear convex constraint, and the Slater's condition is satisfied with$\hat{\mathbf{x}} = 0$, since $\|\hat{\mathbf{x}}\|^2 = 0 < \alpha$. Thus, KKT conditions are necessary and sufficient.

The associated Lagrangian reads

$$\mathcal{L}(\mathbf{x}, \lambda) = \|A\mathbf{x} - b\|^2 + \lambda(\|\mathbf{x}\|^2 - \alpha)\,, \qquad \lambda \in \mathbb{R}_+,$$

and so, a KKT point $\mathbf{x}^*$ satisfies

$$\nabla_x \mathcal{L}(\mathbf{x}^*) = 0 \quad \Longleftrightarrow \quad \begin{cases} 2A^\top(A\mathbf{x}^* - b) + 2\lambda \mathbf{x}^* = 0 \\ \lambda(\|\mathbf{x}^*\|^2 - \alpha) = 0 \\ \|\mathbf{x}^*\|^2 \leq \alpha \quad \Longrightarrow \ \lambda \geq 0 \end{cases}$$

We then distinguish the two cases

- **Case $\lambda = 0$:** using the first expression we have

  $$\mathbf{x}^* = (\mathbf{A}^\top \mathbf{A})^{-1}\mathbf{A}^\top \mathbf{b} = \mathbf{x}_{LS}$$

  where $\mathbf{x}_{LS}$ is the solution of the ordinary Least Squares problem. If $\|\mathbf{x}_{LS}\|^2 \leq \alpha$, then $\mathbf{x}_{LS}$ is the solution of the (CLS) problem.

- **Case $\lambda > 0$:** if we have to consider this is because $\|\mathbf{x}_{LS}\|^2 > \alpha$. Furthermore, $\lambda > 0$ implies – due to the second equation in the KKT system – that $\|\mathbf{x}^*_\lambda\|^2 = \alpha$. Then $\mathbf{x}^*_\lambda = (A^\top A + \lambda \mathbb{I})^{-1} A^\top b$ and we want to find λ such that

  $$\|\mathbf{x}^*_\lambda\|^2 = \alpha = \|(A^\top A + \lambda \mathbb{I})^{-1} A^\top b\|^2\,.$$

  By defining the function $F : \mathbb{R} \longrightarrow \mathbb{R}$ such that

  $$F(\lambda) = \|(A^\top A + \lambda \mathbb{I})^{-1} A^\top b\|^2 - \alpha\,.$$

  The problem reduces to find the zeros of F on $[0, +\infty[$. Let us start noticing that $F(0) = \|\mathbf{x}_{LS}\|^2 - \alpha > 0$, and that F is strictly decreasing with $\lim_{\lambda \to \infty} F(\lambda) = -\alpha < 0$. Thus, there exists a unique solution λ^* such that $F(\lambda^*) = 0$.

To conclude, the solution of the CLS problem reads

$$\mathbf{x}^*_{CLS} = \begin{cases} \mathbf{x}_{LS} & \text{if } \|\mathbf{x}_{LS}\|^2 \leq \alpha\,, \\ (A^\top A + \lambda^* \mathbb{I})^{-1} A^\top b & \text{otherwise} \end{cases}$$

where λ^* satisfies $F(\lambda^*) = 0$.