

MATH60005/70005: Optimisation (Autumn 22-23)

Week 9: Problem Session

Dr Dante Kalise
Department of Mathematics
Imperial College London, United Kingdom
dkaliseb@imperial.ac.uk

Sara Bicego (GTA)
Department of Mathematics
Imperial College London, United Kingdom
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:

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

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:

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

Solve this problem using KKT conditions.

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

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

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

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

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

$$\begin{cases} 2x_1 + 4x_2 - \lambda_1 + \mu = 0 \\ 4x_2 + 4x_1 - \lambda_2 + \mu = 0 \\ \lambda_1 x_1 = 0 \\ \lambda_2 x_2 = 0 \\ x_1 + x_2 = 1 \end{cases} \quad \text{which we address by}$$

considering the following 4 cases.

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

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

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

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

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

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

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

2. A Linear Least Squares problem, given data $\{z_i = (x_i, y_i)\}_{i=1}^m$, reads

$$\min_{z_i} \sum_{i=1}^m \|y_i - (\mathbf{a}^\top x_i + \mathbf{b})\|^2,$$

which is related to the projection of the points z_i along the y -axis. If we are interested in the least orthogonal projection problem instead, the formulation will be

$$\min_{z_i} \sum_{i=1}^m d(z_i, H_{x,y})^2 \quad \text{for} \quad H_{x,y} := \{\mathbf{z} \in \mathbb{R}^n : \mathbf{x}^\top \mathbf{z} = y\}$$

where $\mathbf{x} \in \mathbb{R}^n \setminus 0$ and $y \in \mathbb{R}$. As seen in the course notes, this can be written as

$$\min_{(\mathbf{x}, y)} \sum_{i=1}^m \frac{(\mathbf{x}^\top z_i - y)^2}{\|\mathbf{x}\|^2}.$$

If we fix \mathbf{x} , then the optimisation with respect to y holds as

$$y = \frac{1}{m} \sum_{i=1}^m z_i^\top \mathbf{x},$$

then, we can reformulate the problem as

$$\min_{(\mathbf{x}, y)} \left\{ \sum_{i=1}^m \frac{(\mathbf{x}^\top z_i - \frac{1}{m} \sum_{i=1}^m z_i^\top \mathbf{x})^2}{\|\mathbf{x}\|^2} = \frac{\mathbf{x}^\top \mathbf{z}^\top (\mathbb{I}_m - \frac{1}{m} \mathbf{e} \mathbf{e}^\top) \mathbf{z} \mathbf{x}}{\|\mathbf{x}\|^2} \right\}$$

where $\mathbf{x} \neq 0$ and $\mathbf{z} = (z_1, \dots, z_m)^\top$. From here, it can be observed that the optimal solution corresponds to the smallest eigenvalue of $\mathbf{z}^\top (\mathbb{I}_m - \frac{1}{m} \mathbf{e} \mathbf{e}^\top) \mathbf{z}$



3. We start with the solution via KKT of

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

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

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

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

In the alternative formulation

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

we have again convex cost and convex constraint, but no Slater condition, 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

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

but since the last two equation 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

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

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

in as a Constrained Least Squares (CLS) problem

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

which has convex cost and nonlinear convex constraint, and Slater condition $\mathbf{x}^* = 0$, $\|\mathbf{x}^*\|^2 < \alpha$. Thus, KKT conditions are necessary and sufficient.

The associated Lagrangian reads

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

and so

$$\nabla_{\mathbf{x}} \mathcal{L} \iff \begin{cases} 2\mathbf{A}^\top(\mathbf{Ax} - \mathbf{b}) + 2\lambda\mathbf{x} = 0 \\ \lambda(\|\mathbf{x}\|^2 - \alpha) = 0 \\ \|\mathbf{x}\|^2 \leq \alpha \implies \lambda \geq 0 \end{cases}$$

We then distinguish the two cases



- **Case $\lambda = 0$:** the optimizer is the LSS solution $\mathbf{x}^* = \mathbf{x}_{LSS} = (A^\top A)^{-1} A^\top b$, provided that $\mathbf{x}_{LSS} \leq \alpha$.
- **Case $\lambda > 0$:** 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

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

This is a nonlinear problem for λ , which can be reformulated as finding the zeros of

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

Since $F(0) = \|\mathbf{x}_{LSS}\|^2 - \alpha 0$ and at the same time $\lim_{\lambda \rightarrow \infty} = -\alpha < 0$, we have existence of a solution λ^* such that $F(\lambda^*) = 0$.

To conclude, the solution of the CLS problem reads

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

where λ satisfies $F(\lambda) = 0$.

