José Antonio Álvarez Ocete

Take home exam - part II
Optimization

March 28, 2022

Problem 1
We have worked out the elementary vision of Lagrange multipliers, assuming that, from $g(x,y) = 0$, we can find a function $y = h(x)$ such that $g(x, h(x)) = 0$.

But sometimes, what we get is that there is an h such that $g(h(y), y) = 0$. Rewrite the Lagrange multiplier analysis in the lecture slides under this assumption.

First of all, we will change the question's notation to match the one in the lectures notes. For $f, g : \mathbb{R}^2 \to \mathbb{R}$, consider the following minimization problem:

$$\begin{aligned} &\min f(x,y) \\ &\text{s.t. } h(x,y) = 0. \end{aligned}$$

Assuming the **implicit function theorem** holds, we can find a function $x = \phi(y)$ such that $h(\phi(y), y) = 0$ and, thus, we can write:

$$f(x,y) = f(\phi(y), y) = \psi(y).$$

At a minimum y^* with $x^* = \phi(y^*)$ we thus have

$$0 = \psi'(y^*) = \frac{\partial f}{\partial x}(x^*, y^*)\phi'(y^*) + \frac{\partial f}{\partial y}(x^*, y^*). \tag{1}$$

But since $h(\phi(y), y) = 0$, we also have

$$0 = \frac{\partial h}{\partial x}(x^*, y^*)\phi'(y^*) + \frac{\partial h}{\partial y}(x^*, y^*) \implies \phi'(y^*) = -\frac{\frac{\partial h}{\partial y}(x^*, y^*)}{\frac{\partial h}{\partial x}(x^*, y^*)} \tag{2}$$

Putting together 1 and 2 we arrive at

$$0 = \frac{\partial f}{\partial y}(x^*, y^*)\frac{\partial h}{\partial x}(x^*, y^*) - \frac{\partial f}{\partial x}(x^*, y^*)\frac{\partial h}{\partial y}(x^*, y^*).$$

That is, at (x^*, y^*), $\nabla f \perp \left(\frac{\partial h}{\partial x}, -\frac{\partial h}{\partial y}\right)$ and, since $\left(\frac{\partial h}{\partial x}, -\frac{\partial h}{\partial y}\right) \perp \nabla h$, we have $\nabla f \parallel \nabla h$, i.e. $\nabla f(x^*, y^*) = -\mu^* \nabla h(x^*, y^*)$ for some $\mu^* \neq 0$.

Thus, for the **Lagrangian**

$$\mathcal{L}(x, y; \mu) = f(x,y) + \mu h(x,y),$$

we have that at a minimum (x^*, y^*) there is a $\mu^* \neq 0$ such that:

$$\nabla \mathcal{L}(x^*, y^*; \mu^*) = \nabla f(x^*, y^*) + \mu^* \nabla h(x^*, y^*) = 0.$$

Problem 2

We want to solve the following constrained restriction problem:

$$\begin{aligned} \min \quad & x^2 + 2xy + 2y^2 - 3x + y \\ \text{s.t} \quad & x + y = 1 \\ & x, y \geq 0. \end{aligned}$$

Argue first that f is convex and then:

1. Write its Lagrangian with α, β the multipliers of the inequality constraints.
2. Write the KKT conditions.
3. Use them to solve the problem. For this consider separately the $(\alpha = \beta = 0)$, $(\alpha > 0, \beta = 0)$, $(\alpha = 0, \beta > 0)$, $(\alpha > 0, \beta > 0)$ cases.

Firstly we will show that f is a convex function. A twice differenciable function is convex if and only if its Hessian matrix is positive semidefinite. In our case, we will show that the Hessian matrix is actually positive definite by showing that is symmetric and that all its eigenvalues are positive:

$$\nabla f(x, y) = (2x + 2y + 4y - 3, 2x + 4y + 1)$$
$$\text{Hess } f(x, y) = \begin{pmatrix} 2 & 2 \\ 2 & 4 \end{pmatrix}$$

Indeed, Hess f is symmetric. Let us compute its eigenvalues:

$$|\text{Hess } f - \lambda I| = 0 \iff \begin{vmatrix} 2-\lambda & 2 \\ 2 & 4-\lambda \end{vmatrix} = \lambda^2 - 6\lambda + 4 = 0 \iff \lambda = 3 \pm \sqrt{5} \geq 0$$

which shows that Hess f is positive definite, and thus, convex.

Next, let us solve the given minimization problem. We start by computing the Lagrangian. We have to change the sign of the inequalities $x, y \geq 0$ to $-x, -y \leq 0$ to this end:

$$\mathcal{L}(x, y; \alpha, \beta, \gamma) = x^2 + 2xy + 2y^2 - 3x + \gamma(1 - x - y) + y - \alpha x - \beta y$$

In order to compute the KKT conditions for this problem we compute the partial derivatives of the Lagrangian:

$$\frac{\partial \mathcal{L}}{\partial x}(x, y; \alpha, \beta, \gamma) = 2x + 2y - 3 - \gamma - \alpha$$
$$\frac{\partial \mathcal{L}}{\partial y}(x, y; \alpha, \beta, \gamma) = 2x + 4y + 1 - \gamma - \beta$$

Hence, the KKT conditions are:

$$\begin{cases} 2x + 2y - 3 - \gamma - \alpha & = 0 \\ 2x + 4y + 1 - \gamma - \beta & = 0 \\ \alpha x & = 0 \\ \beta y & = 0 \end{cases}$$

where $\alpha, \beta \geq 0$. Finally, let us study the different cases for (α, β) to find the KKT points:

- Case $\alpha, \beta > 0$: Using the last two equations we obtain $x = y = 0$, but our initial equality restriction states that $x + y = 1$, impossible. There are no KKT points in this case.
- Case $\alpha = \beta = 0$: By adding the first two KKT conditions we obtain:

  $$-2y - 4 - \alpha + \beta = 0. \tag{3}$$

  Using this equation we obtain $y = -2 < 0$, which invalidates the initial restriction $y > 0$. Thus, there are no KKT points in this case.
- Case $\alpha > 0, \beta = 0$: Using the fourth KKT condition we obtain $x = 0$ and since $x + y = 1$, $y = 1$. However, we still have to verify that the KKT conditions and the rest of restrictions are fulfilled. The first KKT condition shows that $\alpha = -6 < 0$, and thus the computed point is not a valid KKT point. Thus, there are no KKT points in this case.
- Case $\alpha = 0, \beta > 0$: Using the third KKT condition we obtain $y = 0$ and since $x + y = 1$, $x = 1$. The first KKT conditions shows that $\gamma = -1$ (which is valid), and the second that $\beta = 4$. We have obtained a set of values $(x = 1, y = 0, \alpha = 0, \beta = 4, \gamma = -1)$ that satisfy all the KKT conditions and problem restrictions. Thus, $(x = 1, y = 0)$ is a KKT point.

Since $(x = 1, y = 0)$ is the only KKT point, it must be our minimum. Finally, the minimum value of $f(x, y)$ where (x, y) satisfy our restrions is $f(1, 0) = -2$.

Problem 3
Let $f : S \subset \mathbb{R}^d \to \mathbb{R}$ be a convex function on the convex set S and we extend it to an $\tilde{f} : \mathbb{R}^d \to \mathbb{R}$ as:

$$\tilde{f}(x) = \begin{cases} f(x) & \text{if } x \in S \\ +\infty & \text{if } x \notin S \end{cases}$$

Show that $\tilde{f}$ is a convex function on $\mathbb{R}^d$. Assume that $a + \infty = \infty$ and $a \cdot \infty = \infty$ for $a > 0$.

Since $\tilde{f}$ is not differentiable, we will prove its convexity using the defintion:

$$f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda) f(y) \tag{4}$$

Let us distinguish cases:

- If $x, y \in S$, equation 4 holds due to the convexity of f.
- If $x, y \notin S$, equation 4 holds:

$$\begin{aligned} f(\lambda x + (1-\lambda)y) &\le \lambda f(x) + (1-\lambda) f(y) \\ &= \lambda \infty + (1-\lambda)\infty \\ &= \infty \end{aligned}$$

- Suppose $x \in S, y \notin S$ (the opposite case is analogous). Then:

$$\begin{aligned} f(\lambda x + (1-\lambda)y) &\le \lambda f(x) + (1-\lambda) f(y) \\ &= \lambda \infty + (1-\lambda) f(y) \\ &= \infty. \end{aligned}$$

  Hence, equation 4 holds.

Problem 4
Prove Jensen's inequality: if f is convex on $\mathbb{R}^d$ and $\sum_1^k \lambda_i = 1$, with $0 \le \lambda_i \le 1$, we have for any $x_1, \dots, x_k \in \mathrm{R}^n$

$$f\left(\sum_1^k \lambda_i x_i\right) \le \sum_1^k \lambda_i f(x_i)$$

Hint: just write $\sum_1^k \lambda_i x_i = \lambda_1 x_1 + (1-\lambda_1) u$ for an appropriate u and apply repeatedly the definition of a convex function. Start with $k = 3$ and carry on.

We will prove this inequality by induction:

- Base case 1 ($k = 1$): trivial when we realize that $\lambda_1 = 1$.
- Base case 2 ($k = 2$): we need to prove that

$$f(\lambda_1 x_1 + \lambda_2 x_2) \le \lambda_1 f(x_1) + \lambda_2 f(x_2),$$

  but this is true since f is convex.

- Recursive case: Let $k \in \mathbb{N}, k > 2$ and suppose that

$$f\left(\sum_1^k \lambda_i x_i\right) \le \sum_1^k \lambda_i f(x_i).$$

  We will show that

$$f\left(\sum_1^{k+1} \lambda_i x_i\right) \le \sum_1^{k+1} \lambda_i f(x_i). \tag{5}$$

If $\lambda_{k+1} = 1$, and since $\sum_{i=0}^{k+1} \lambda_i = 1$, then $\lambda_1 = \ldots = \lambda_k = 0$. This is the first base case, so the inequality 5 holds.

Suppose that $\lambda_{k+1} < 1$, we may rewrite our linear combination as:

$$\sum_{i=0}^{k+1} \lambda_i x_i = \lambda_{k+1} x_{k+1} + (1 - \lambda_{k+1}) \left(\sum_{i=0}^{k} \frac{\lambda_i}{1 - \lambda_{k+1}} x_i \right).$$

Then:

$$\begin{aligned}
f\left(\sum_{i=0}^{k+1} \lambda_i x_i\right) &= f\left(\lambda_{k+1} x_{k+1} + (1 - \lambda_{k+1}) \left(\sum_{i=0}^{k} \frac{\lambda_i}{1 - \lambda_{k+1}} x_i \right)\right) \\
(1) \quad &\leq \lambda_{k+1} f(x_{k+1}) + (1 - \lambda_{k+1}) f\left(\sum_{i=0}^{k} \frac{\lambda_i}{1 - \lambda_{k+1}} x_i \right) \\
(2) \quad &\leq \lambda_{k+1} f(x_{k+1}) + (1 - \lambda_{k+1}) \sum_{i=0}^{k} \frac{\lambda_i}{1 - \lambda_{k+1}} f(x_i) \\
&= \lambda_{k+1} f(x_{k+1}) + \sum_{i=0}^{k} \lambda_i f(x_i) \\
&= \sum_{i=0}^{k+1} \lambda_i f(x_i)
\end{aligned}$$

where in (1) we used that f is convex and in (2) the induction hypothesis. This proves inequality 5 and the whole theorem.

Problem 5
Prove that the following function is convex

$$f(x) = \begin{cases} x^2 - 1 & |x| > 1 \\ 0 & |x| \leq 1 \end{cases}$$

and compute its proximal. Which are the fixed points of this proximal?

Let us prove that f is convex by using the definition: f is convex if and only if for all $x, y \in \mathbb{R}$ and $\lambda \in [0, 1]$ the following equality holds:

$$f(\lambda x + (1 - \lambda) y) \leq \lambda f(x) + (1 - \lambda) f(y) \tag{6}$$

First, note that $f(x) \geq 0$ for all $x \in \mathbb{R}$. We will call $z \equiv \lambda x + (1 - \lambda) y$, and distinguish cases:

- Case 1: If $|z| \leq 1$, $f(z) = 0$, and equation 6 holds:

$$0 = f(z) \leq \underbrace{\lambda}_{\geq 0} \underbrace{f(x)}_{\geq 0} + \underbrace{(1 - \lambda)}_{\geq 0} \underbrace{f(y)}_{\geq 0}$$

- Case 2: If $|z| > 1$, at least either x or y have an absolute value greater than 1. Let us distinguish cases:
  - If both fulfill the condition, $|x|, |y| > 1$ (and since $|z|$ is also greater than 1), we are just checking if the mapping $x \mapsto x^2 - 1$ is convex, which is trivial. Thus, equation 6 holds.
  - Our last case is the most complex one and requires us to use the expression of f explicitely. Suppose that $|x| > 1$ and $|y| \leq 1$ (the opposite case is analogous). Then:

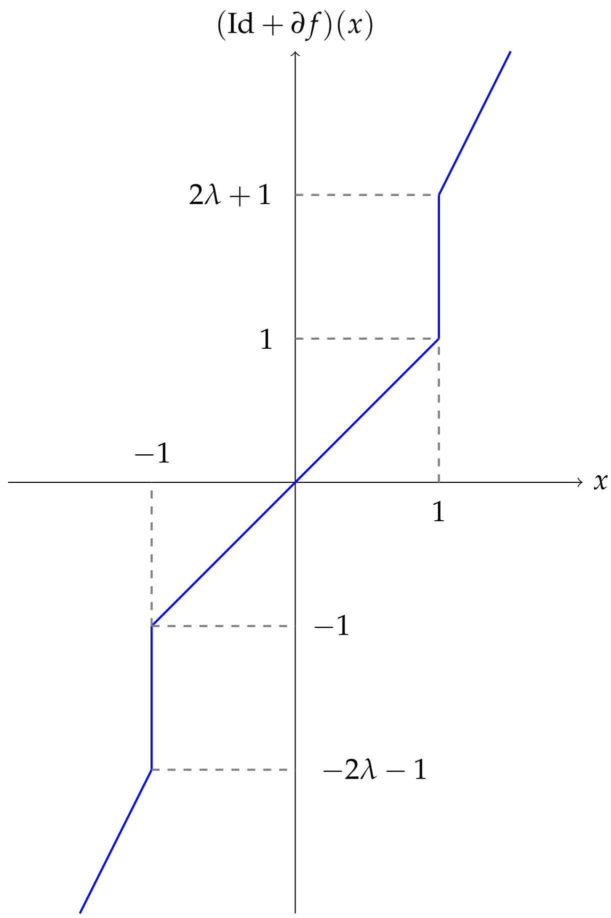
$$
\begin{aligned}
& f(z) \leq \lambda f(x) + (1-\lambda) f(y) = \lambda f(x) & \\
\iff \quad & (\lambda x + (1-\lambda) y)^2 - 1 \leq \lambda \left(x^2 - 1\right) & \\
\iff \quad & \lambda x^2 - \lambda + 1 - (\lambda x + (1-\lambda) y)^2 & \geq 0 \\
\iff \quad & \lambda x^2 + (1-\lambda) - \lambda^2 x^2 - (1-\lambda)^2 y^2 - 2\lambda(1-\lambda) xy & \geq 0 \\
\iff \quad & (1-\lambda) + x^2 \lambda (1-\lambda) - (1-\lambda)^2 y^2 - 2\lambda(1-\lambda) xy & \geq 0 \\
(1) \iff \quad & 1 + \lambda x^2 - (1-\lambda) y^2 - 2\lambda xy & \geq 0 \\
\iff \quad & \lambda(x^2 + y^2 + 2xy) + 1 - y^2 & \geq 0 \\
\iff \quad & \underbrace{\lambda(x-y)^2}_{\geq 0} + \underbrace{1 - y^2}_{\geq 0} & \geq 0
\end{aligned}
$$

    where in (1) we divided by $(1-\lambda)$ and used that $(1-\lambda) > 0$. In the last inequality we know that $1 - y^2 \geq 0$ because $|y| \leq 1$.

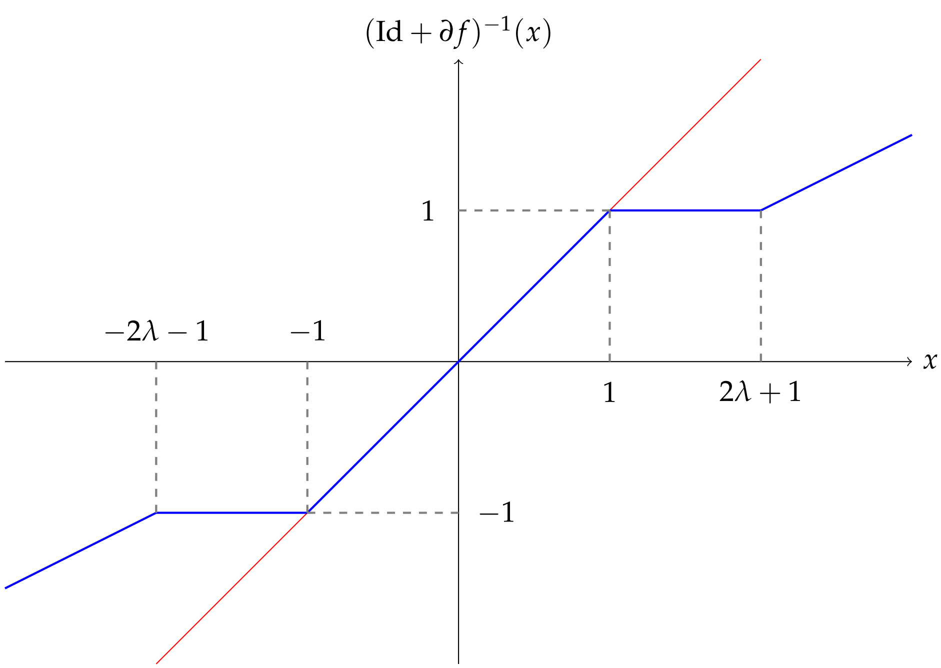
That proves that f is convex. Let us compute the proximal. First, we obtain the subgradient and $(\mathrm{Id} + \lambda \partial f)(x)$:

$$
\partial f(x) = \begin{cases} 2x & x < -1 \\ [-2, 0] & x = -1 \\ 0 & x \in [-1, 1] \\ [0, 2] & x = 1 \\ 2x & x > 1 \end{cases} \quad \lambda \partial f(x) = \begin{cases} 2\lambda x & x < -1 \\ [-2\lambda, 0] & x = -1 \\ 0 & x \in [-1, 1] \\ [0, 2\lambda] & x = 1 \\ 2\lambda x & x > 1 \end{cases} \quad (\mathrm{Id} + \lambda \partial f)(x) = \begin{cases} x(2\lambda + 1) & x < -1 \\ [-2\lambda - 1, -1] & x = -1 \\ x & x \in [-1, 1] \\ [1, 2\lambda + 1] & x = 1 \\ x(2\lambda + 1) & x > 1 \end{cases}
$$

In order to obtain the proximal $\mathrm{prox}_f(x) = (\mathrm{Id} + \partial f)^{-1}(x)$, we plot the previously obtained function:



And then compute its inverse by rotating 90 degrees clockwise around the origin and flipping around the vertical axis:



From the figure we may obtain the analytic expression of the proximal:

$$\operatorname{prox}_f(x) \begin{cases} \frac{x}{2\lambda+1} & x \leq -\epsilon - \lambda \\ -1 & -2\lambda - 1 \leq x \leq -1 \\ x & |x| < 1 \\ 1 & 1 \leq x \leq 2\lambda + 1 \\ \frac{x}{2\lambda+1} & x \geq 2\lambda + 1 \end{cases}$$

Finally, the fixed points are those that match the identity function (the red line in the previous figure). Those are $[-1, 1]$. As we already know, these are the minimums of our original function f.