IB Optimisation: Lecture 3

Mike Tehranchi

University of Cambridge

29 April 2020

We now consider the general constrained optimisation problem

$$\text{minimise } f(x) \text{ subject to } g(x) = b, \quad x \in X.$$

No convexity assumptions are made now.

Introduce a new function $L : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ defined by

$$L(x, \lambda) = f(x) + \lambda^\top (b - g(x))$$

This function is called the *Lagrangian* of the problem.
For a vector $\lambda = (\lambda_1, \ldots, \lambda_m)^\top$, the component λ_i is called the *Lagrange multiplier* for the i-th functional constraint.

Theorem (The Lagrangian sufficiency theorem)

Let x^* be feasible for the problem. Suppose there exists a $\lambda^* \in \mathbb{R}^m$ such that

$$L(x^*, \lambda^*) \le L(x, \lambda^*) \text{ for all } x \in X.$$

Then x^* is optimal.

Proof: For any feasible x and any λ we have

$$L(x, \lambda) = f(x) + \lambda^\top (b - g(x)) = f(x)$$

since $g(x) = b$. Hence if x^* is feasible, then

$$\begin{aligned} f(x^*) &= L(x^*, \lambda^*) \\ &\leq L(x, \lambda^*) \text{ for all } x \in X \text{ by assumption} \\ &= f(x) \text{ for all feasible } x. \end{aligned}$$

□

Example. Consider

$$\text{minimise } x_1^2 + 3x_2^2 \text{ subject to } 4x_1 + x_2 = 7$$

Claim: $(x_1^*, x_2^*) = (\frac{12}{7}, \frac{1}{7})$ is optimal.

Consider the Lagrangian

$$L(x_1, x_2, \lambda) = x_1^2 + 3x_2^2 + \lambda(7 - 4x_1 - x_2)$$

Note that

$$L(x_1, x_2, \tfrac{6}{7}) = (x_1 - \tfrac{12}{7})^2 + 3(x_2 - \tfrac{1}{7})^2 + 3$$

so

$$L(x_1, x_2, \tfrac{6}{7}) \geq L(\tfrac{12}{7}, \tfrac{1}{7}; \tfrac{6}{7})$$

for *all* (x_1, x_2). We're done by the Lagrangian sufficiency theorem.

First interpretation of the Lagrange multiplier λ^*: a certificate of optimality.

For a general optimisation problem, is it always possible find numbers $(\lambda_1, \ldots, \lambda_m)$ to serve as a certificate of optimality?

If x^* and λ^* exist as in the Lagrangian sufficiency theorem, we have

$$\inf_{x \in X} L(x, \lambda^*) = f(x^*) > -\infty$$

Step (1). Identify the set of feasible Lagrange multipliers

$$\Lambda = \{\lambda \in \mathbb{R}^m : \inf_{x \in X} L(x, \lambda) > -\infty\}.$$

Step (2). For each $\lambda \in \Lambda$ find the optimal solution to the unconstrained problem to

$$\text{minimise } L(x, \lambda) \text{ subject to } x \in X.$$

Let $x(\lambda)$ be the minimiser.

Step (3). Find a $\lambda^* \in \Lambda$ such that $x^* = x(\lambda^*)$ is feasible for the original problem, that is, $g(x^*) = b$.

In general, it might not be possible to do steps (1) through (3). But, if it is possible, the resulting x^* is optimal by the Lagrangian sufficiency theorem. (By step (2) we have that $L(x^*, \lambda^*) \leq L(x, \lambda^*)$ for all $x \in X$, and by step (3) we have that x^* is feasible.)

Example. (Maximum likelihood estimator of the multinomial distribution)
Given constants $n_1, \ldots, n_k > 0$, consider the problem to

$$\text{maximise } \sum_{i=1}^{k} n_i \log p_i \text{ subject to } \sum_{i=1}^{k} p_i = 1, \; p_i > 0 \text{ for all } i.$$

The Lagrangian is

$$\begin{aligned} L(p,\lambda) &= \sum_{i=1}^k n_i \log p_i + \lambda \left(1 - \sum_{i=1}^k p_i\right) \\ &= \lambda + \sum_{i=1}^k (n_i \log p_i - \lambda p_i) \end{aligned}$$

Step (1). Note that if $\lambda \leq 0$ then $n_i \log p_i - \lambda p_i \to \infty$ as $p_i \to \infty$. Hence

$$\Lambda = \{\lambda \in \mathbb{R} : \sup_{p>0} L(p, \lambda) < \infty\} = \{\lambda : \lambda > 0\}$$

Step (2). We solve

$$\frac{\partial L}{\partial p_i} = \frac{n_i}{p_i} - \lambda = 0 \Rightarrow p_i(\lambda) = \frac{n_i}{\lambda}.$$

Since

$$D^2 L = \begin{pmatrix} -\frac{n_1}{p_1^2} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & -\frac{n_k}{p_k^2} \end{pmatrix}$$

is non-positive definite for all p, we have found the maximum.

Step (3) The constraint $\sum_{i=1}^{k} p_i = \sum_{i=1}^{k} \frac{n_i}{\lambda} = 1$ yields $\lambda = \sum_{i=1}^{k} n_i$. By the Lagrangian sufficiency theorem,

$$p_i^* = \frac{n_i}{\sum_{j=1}^{k} n_j}$$

is optimal.

Notation. If $x, y \in \mathbb{R}^n$ then we write $x \geq y$ if $x_i \geq y_i$ for all $1 \leq i \leq n$

A major focus of this course are problems with inequality constraints. Consider

$$P: \text{ minimise } f(x) \text{ subject to } g(x) \le b, \quad x \in X.$$

This problem can be put into equality form by introducing *slack variables*:

$$P': \text{ minimise } f(x) \text{ subject to } g(x) + z = b, \quad x \in X, z \ge 0.$$

Notice that $\binom{x}{z} \in \mathbb{R}^{n+m}$ is feasible for problem P' if and only if $x \in \mathbb{R}^n$ is feasible for problem P and $z = b - g(x)$.

The Lagrangian is

$$L(x, z, \lambda) = f(x) + \lambda^\top(b - g(x) - z) = f(x) + \lambda^\top(b - g(x)) - \lambda^\top z$$

We now apply the Lagrangian method.

Step (1). Note that if $\lambda_i > 0$ for some i then

$$-\lambda^\top z = -\lambda_1 z_1 - \ldots - \lambda_i z_i - \ldots - \lambda_m z_m \to -\infty \text{ as } z_i \to \infty.$$

Hence

$$\inf_{x \in X, z \geq 0} L(x, z, \lambda) > -\infty \text{ only if } \lambda \leq 0$$

That is, the inequality constraint $g(x) \leq b$ for the variable x introduces a *sign constraint* $\lambda \leq 0$ for the Lagrange multiplier λ. In particular, we have

$$\Lambda = \{\lambda \in \mathbb{R}^m : \lambda \leq 0, \inf_{x \in X} [f(x) + \lambda^\top (b - g(x))] > -\infty\}$$

Step (2). Note that for $\lambda \le 0$ we have $\inf_{z \ge 0}(-\lambda^\top z) = 0$. That is to say, for each $\lambda \in \Lambda$, the optimal $z = z(\lambda)$ satisfies the *complementary slackness* condition $\lambda^\top z = 0$.

- If i-th Lagrange multiplier λ_i is non-zero, then $z_i = 0$ so the i-th functional constraint is *tight*, that is, holds with equality.
- If the i-th functional constraint is not tight so that $z_i > 0$ then i-th Lagrange multiplier λ_i is zero.

To find the $x = x(\lambda)$ we solve the unconstrained problem to

$$\text{minimise } f(x) + \lambda^\top (b - g(x))$$

as usual.

Step (3). As usual, pick $\lambda^* \in \Lambda$ so that $x^* = x(\lambda^*)$ and $z^* = z(\lambda^*)$ are feasible, i.e. $g(x^*) \leq b$.

Consider

$$P: \text{minimise} \quad x_1 - 3x_2 \quad \text{subject to} \quad \begin{array}{rcl} x_1^2 + x_2^2 & \leq & 4 \\ x_1 + x_2 & \leq & 2 \end{array}$$

Introducing slack variables, the problem is

$$P': \text{minimise} \quad x_1 - 3x_2 \quad \text{subject to} \quad \begin{array}{l} x_1^2 + x_2^2 + z_1 \qquad\quad = \; 4 \\ x_1 + x_2 \qquad\qquad + z_2 = \; 2 \\ z_1, z_2 \geq 0 \end{array}$$

The Lagrangian is

$$L = x_1 - 3x_2 + \lambda_1(4 - x_1^2 - x_2^2 - z_1) + \lambda_2(2 - x_1 - x_2 - z_2)$$

By the sign constraint, we consider Lagrange multipliers $\lambda_1, \lambda_2 \leq 0$. Note that

$$D^2 L = \begin{pmatrix} -2\lambda_1 & 0 \\ 0 & -2\lambda_1 \end{pmatrix}$$

so the Hessian is non-negative definite. Hence to find the minimum we need only solve $\frac{\partial L}{\partial x_1} = 0 = \frac{\partial L}{\partial x_2}$ yielding

$$1 - 2\lambda_1 x_1 - \lambda_2 = 0$$
$$-3 - 2\lambda_1 x_2 - \lambda_2 = 0$$

We now have to analyse cases.

- Case $\lambda_1 = 0$. This yields $\lambda_2 = 1$ and $\lambda_2 = -3$, a contradiction.
- Case $\lambda_1 < 0$, $\lambda_2 < 0$. Note that by complementary slackness

$z_1 = z_2 = 0$ so both functional constraints are tight. Hence we have four equations and four unknowns:

$$\begin{aligned} 1 - 2\lambda_1 x_1 - \lambda_2 &= 0 \\ -3 - 2\lambda_1 x_2 - \lambda_2 &= 0 \\ x_1^2 + x_2^2 &= 4 \\ x_1 + x_2 &= 2. \end{aligned}$$

Solving the bottom two equations yields the two solutions $(x_1, x_2) = (2, 0)$ and $(0, 2)$. Plugging these into the first equations yields $(x_1, x_2, \lambda_1, \lambda_2) = (2, 0, 1, -3)$ and $(0, 2, -1, 1)$. Unfortunately, neither solution works since the sign constraint $\lambda \leq 0$ is violated for both.

• Case $\lambda_1 < 0$, $\lambda_2 = 0$. Now by complementary slackness $z_1 = 0$ so the first functional constraints is tight. Hence we have three equations and three unknowns:

$$\begin{aligned} 1 - 2\lambda_1 x_1 &= 0 \\ -3 - 2\lambda_1 x_2 &= 0 \\ x_1^2 + x_2^2 &= 4 \end{aligned}$$

From the first equations we get $x_1 = \frac{1}{2\lambda_1}$, $x_2 = -\frac{3}{2\lambda_1}$ and from the third equation $\lambda_1 = \pm\frac{\sqrt{10}}{4}$. But $\lambda_1 < 0$, so the solution

$$(x_1, x_2) = \left(-\sqrt{\frac{2}{5}}, 3\sqrt{\frac{2}{5}}\right)$$

is optimal by the Lagrangian sufficiency theorem.