Lagrange multipliers

Sylvain Rubenthaler

28 octobre 2019

Université Nice SOPHIA ANTIPOLIS

Gradients

When $f:\mathbb{R}^n \to \mathbb{R}$ is sufficiently differentiable. We can write a first order development, using many equivalent notations for the first order differential of f:

$$\begin{aligned} f(x+u) &= f(x) + df(x)(u) + o(\|u\|) \\ &= f(x) + f'(x).u + o(\|u\|). \end{aligned}$$

We have

$$f'(x).u = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(x) \times u_i .$$

So, if we define the gradient of f by

$$\nabla f(x) = \begin{pmatrix} \frac{\partial f}{\partial x_1}(x) \\ \vdots \\ \frac{\partial f}{\partial x_n}(x) \end{pmatrix},$$

Gradients

we then get

$$f'(x).u = \langle \nabla f(x), u \rangle$$

($\langle \ldots, \ldots \rangle$ is the scalar product).

⚠This is not a real proof.

Maximization of f

Suppose, we have $f : \mathbb{R}^n \to \mathbb{R}$ and $g_1, \ldots, g_p : \mathbb{R}^n \to \mathbb{R}$. We set $M = \{x : g_i(x) = c_i, 1 \leq i \leq p\}$ (for some constants c_i). We are interested in

$$\operatorname*{argmax}_{x \in M} f(x),$$

which is the same as finding

$$\begin{cases} \max f(x) \\ \text{under } g_i(x) = c_i, 1 \leq i \leq p. \end{cases} \tag{1}$$

Necessary condition

Suppose we have found a maximum in x_0. Then, for any "small" move u such that $g_i'(x_0).u = 0$ (that is, a move that stays in M), we have

$$f(x_0 + u) = f(x_0) + f'(x_0).u + o(\|u\|).$$

So $f'(x_0).u = 0$. This means that

$$\nabla f(x_0) \in \mathrm{Span}(\nabla g_1(x_0), \ldots, \nabla g_p(x_0)).$$

Lagrangian function

We set

$$L(x, \lambda_1, \ldots, \lambda_p) = f(x) - \sum_{i=1}^{p} \lambda_i (g_i(x) - c_i)$$

$(L : \mathbb{R}^n \times \mathbb{R}^p \to \mathbb{R})$. In x_0, there exist $\lambda_1^{(0)}$, ..., $\lambda_p^{(0)}$ such that

$$\nabla f(x_0) = \sum_{i=1}^{p} \lambda_i^{(0)} \nabla g_i(x_0).$$

Lagrangian function

Let us compute

$$\nabla L(x,\lambda_1,\ldots,\lambda_p)=\begin{pmatrix} \frac{\partial f}{\partial x_1}(x)-\sum_{i=1}^p \lambda_i \frac{\partial g_i}{\partial x_1}(x) \\ \vdots \\ \frac{\partial f}{\partial x_n}(x)-\sum_{i=1}^p \lambda_i \frac{\partial g_i}{\partial x_n}(x) \\ -(g_1(x)-c_1) \\ \vdots \\ -(g_p(x)-c_p) \end{pmatrix}.$$

We observe that

$$\nabla L(x_0,\lambda_1^{(0)},\ldots,\lambda_p^{(0)})=0.$$

Working our way back

The conclusion is that, when trying to find the solution of (1), a good candidate is x_0 such that there exists $(\lambda_i^{(0)})_{1 \le i \le p}$ with $\nabla L(x_0, \lambda_1^{(0)}, \dots, \lambda_p^{(0)}) = 0$.

The coefficients $\lambda_i^{(0)}$ are called "Lagrange multipliers".

Maximization of f

Suppose, we have $f : \mathbb{R}^n \to \mathbb{R}$ and $g_1, \ldots, g_p : \mathbb{R}^n \to \mathbb{R}$. We are interested in

$$\begin{cases} \max f(x) \\ \text{under } g_i(x) \le c_i\,,\, 1 \le i \le p\,. \end{cases} \tag{2}$$

for some $c_1, \ldots, c_p$.

Necessary condition

Suppose we have found a maximum x_0. We then divide the indexes into

- for $1 \le i \le k$, $g_i(x_0) = c_i$ (we say the constraint is binding)
- for $k+1 \le i \le p$, $g_i(x_0) < c_i$ (we say the constraint is not binding)

for some k in $\{0, 1, \ldots p\}$.

Necessary condition

For v such that $g_i'(x_0).v = 0$ $(1 \leq i \leq k)$, we have

$$f'(x_0).v = 0$$

(because x_0 is a maximum of f restrained to $\{x : g_i(x) = c_i, 1 \leq i \leq k\}$). So there exist $\lambda_1^{(0)}, \ldots, \lambda_k^{(0)}$ such that

$$\nabla f(x_0) = \sum_{i=1}^{k} \lambda_i^{(0)} \nabla g_i(x_0).$$

We set

$$\lambda_{k+1}^{(0)} = \cdots = \lambda_p^{(0)} = 0.$$

Necessary condition

Let us now take $h > 0$ ("small"). We have, for all i in $\{1, 2, \ldots, k\}$,

$$f(x_0 - h\nabla g_i(x_0)) = f(x_0) - h\langle \nabla f(x_0), \nabla g_i(x_0)\rangle + o(h),$$

and we should have $f(x_0 - h\nabla g_i(x_0)) \leq f(x_0)$. So

$$\langle \nabla f(x_0), \nabla g_i(x_0)\rangle \geq 0.$$

In the case where the $\nabla g_i(x_0)$ are orthogonal, the above implies $\lambda_i^{(0)} \geq 0$.

Lagrangian function

We define

$$L(x,\lambda_1,\dots,\lambda_p) = f(x) - \sum_{i=1}^{p} \lambda_i (g_i(x) - c_i)$$

$(L : \mathbb{R}^n \times \mathbb{R}^p \to \mathbb{R})$. The gradient of L is (the same as before)

$$\nabla L(x,\lambda_1,\dots,\lambda_p) = \begin{pmatrix} \frac{\partial f}{\partial x_1}(x) - \sum_{i=1}^{p} \lambda_i \frac{\partial g_i}{\partial x_1}(x) \\ \vdots \\ \frac{\partial f}{\partial x_n}(x) - \sum_{i=1}^{p} \lambda_i \frac{\partial g_i}{\partial x_n}(x) \\ -(g_1(x) - c_1) \\ \vdots \\ -(g_p(x) - c_p) \end{pmatrix}.$$

Lagrangian function

We get that

$$\begin{cases} \frac{\partial L}{\partial x_1}(x_0) = 0, \ldots, \frac{\partial L}{\partial x_n}(x_0) = 0\,, \\ \lambda_i^{(0)}(g_i(x_i) - c_i) = 0\,, 1 \leq i \leq p\,, \\ g_i(x_0) \leq c_i\,, 1 \leq i \leq p\,, \\ \langle \nabla f(x_0), \nabla g_i(x_0) \rangle \geq 0\,, 1 \leq i \leq k\,. \end{cases} \tag{3}$$

Working our way back

The conclusion is that, when trying to find the solution of (2), a good candidate is x_0 such that there exists $(\lambda_i^{(0)})_{1 \leq i \leq p}$ such that Equation (3) is statisfied.