Lagrange multipliers and KKT conditions

Vincenzo Bonifaci

May 5, 2016

The method of *Lagrange multipliers* allows one to write necessary conditions for any smooth optimization problem with equality constraints. It can be further generalized to inequality constraints, yielding the *Karush-Kuhn-Tucker* (KKT) conditions. Lagrange multipliers and KKT conditions have many uses in applied mathematics; often, the multiplier variables also have interesting interpretations.

1 Method of Lagrange multipliers

Consider a smooth[1] real-valued function $F(x, y)$ defined over $\mathbb{R}^{n\times m}$ ($x \in \mathbb{R}^n$, $y \in \mathbb{R}^m$). We study the constrained minimization problem:

$$\begin{aligned} &\min F(x,y) && && \text{(P)}\\ &\text{s.t. } h_k(x,y) = 0 && k = 1,\dots,l \\ &\quad x_i \geq 0 && i = 1,\dots,n \\ &\quad y_j \gtrless 0 && j = 1,\dots,m, \end{aligned}$$

where each h_k is an linear (affine) constraint.

For a vector $\lambda \in \mathbb{R}^l$, let

$$L(x, y, \lambda) \stackrel{\text{def}}{=} F(x, y) + \lambda^\top h(x, y) = F(x, y) + \sum_{k=1}^{l} \lambda_k h_k(x, y).$$

The theory of Lagrange multipliers asserts that if (x, y) is an optimal solution to (P), then there exists a set of real values $\lambda_1, \dots, \lambda_l$ with the following properties:

1. For each $j = 1, \dots, m$, $\frac{\partial}{\partial y_j} L(x, y, \lambda) = 0$;
2. For each $i = 1, \dots, n$, if $x_i > 0$, then $\frac{\partial}{\partial x_i} L(x, y, \lambda) = 0$;
3. For each $i = 1, \dots, n$, if $x_i = 0$, then $\frac{\partial}{\partial x_i} L(x, y, \lambda) \geq 0$.

[1] "Smooth", here, means differentiable and with a continuous derivative.

Example 1.1. Consider $\min\{x^2 + y^2 : x + y \geq 1\}$. We transform the inequality constraint into an equality constraint by adding a variable $z \geq 0$ and rewriting the feasible region as $\{(x, y, z) \in \mathbb{R}^3 : x + y = 1 + z,\ z \geq 0\}$. The Lagrange function is

$$L(x, y, z, \lambda) = x^2 + y^2 + \lambda(x + y - z - 1).$$

Condition 1 gives $2x + \lambda = 0$ and $2y + \lambda = 0$. Thus, $2x + \lambda = 2y + \lambda$, i.e., $x = y$ is necessary for optimality of (x, y). Condition 2 gives: if $z > 0$, then $-\lambda = 0$. Condition 3 gives: if $z = 0$, then $-\lambda \geq 0$. But $\lambda = 0$ is impossible, since it would imply $x = y = 0$ which is not feasible. Thus $z = 0$ and $x + y = 1 + z = 1$, from which we obtain $x = y = 1/2$.

2 Karush-Kuhn-Tucker conditions

The method of Lagrange multipliers is a special case of a more general result due to Karush, Kuhn, and Tucker (KKT). In this case we consider the constrained optimization problem

$$\begin{aligned} &\min F(x) && && \text{(P)} \\ &\text{s.t. } h_k(x) = 0 && k = 1, \dots, l \\ &\quad g_j(x) \leq 0 && j = 1, \dots, m. \end{aligned}$$

where $x \in \mathbb{R}^n$, $F : \mathbb{R}^n \to \mathbb{R}$ is a smooth function, and each h_k, g_j is also a smooth function from $\mathbb{R}^n$ to $\mathbb{R}$.

The KKT conditions apply whenever the constraints are linear (affine) or, alternatively, whenever there exists a point $\bar{x}$ such that $h_k(\bar{x}) = 0$ for each k and $g_j(\bar{x}) < 0$ for each j. If this is true, assume that x is an optimal point; then Karush, Kuhn and Tucker proved that there must exist real values μ_j $(j = 1, \dots, m)$, λ_k $(k = 1, \dots, l)$ such that:

$$\nabla F(x) + \mu^\top \nabla g(x) + \lambda^\top \nabla h(x) = 0,$$

or in other words, for each $i = 1, \dots, n$,

$$\frac{\partial}{\partial x_i} F(x) + \sum_{j=1}^{m} \mu_j \frac{\partial}{\partial x_i} g_j(x) + \sum_{k=1}^{l} \lambda_k \frac{\partial}{\partial x_i} h_k(x) = 0.$$