Lagrange Multipliers

by Alec Johnson

November 2, 2008

The basic idea of Lagrange Multipliers is that a differentiable function f cannot have a local maximum at a point $\mathbf{r}_0$ when restricted to a differentiable surface or curve (given by the level sets $g_i(\mathbf{r}) = C_i$) if the gradient of f has a component parallel to the surface. In other words, f can only have a maximum at a point $\mathbf{r}_0$ where f and g are smooth if ∇f is perpendicular to the surface at p_0. But this means that the gradient of f must be in the span of the gradients of the constraint equations.

Proposition 1 (Lagrange Multiplers)**.** *Let $f : \mathbb{R}^n \mapsto \mathbb{R}$ and let $\{g_i(\mathbf{r}) = C_i\}_{i=1}^m$ be m constraint equations, where $g_i : \mathbb{R}^n \mapsto \mathbb{R}$. Let $\mathbf{r}_0$ be a local extremum of f on the manifold (surface or curve) defined by the constraint equations, and suppose that f and g_i are differentiable at $\mathbf{r}_0$.*

Then there exist $\lambda_i, i = 1, \ldots, m$ such that at $\mathbf{r}_0$

$$\nabla f = \sum_{i=1}^{m} \lambda_i \nabla g_i.$$

Justification of 1. Let $\mathbf{r}(t)$ represent an arbitrary differentiable path in the contraint manifold, i.e. $g_i(\mathbf{r}(t)) = C_i$ for all i, and suppose that $\mathbf{r}(0) = \mathbf{r}_0$ and $\mathbf{r}'(0) \neq 0$. Then the function $h(t) := f(\mathbf{r}(t))$ has a local extremum at $t = 0$, so $0 = \frac{d}{dt}h(0) = \frac{d}{dt}f(\mathbf{r}(t))|_{t=0} = \mathbf{r}'(0) \cdot \nabla f$. Similarly, differentiating the constraint equations gives $0 = \frac{d}{dt}g_i(\mathbf{r}(t))|_{t=0} = \mathbf{r}'(0) \cdot \nabla g_i$. So we have shown that if $\mathbf{u} \cdot \nabla g_i = 0$ then $\mathbf{u} \cdot \nabla f = 0$, i.e., if $\mathbf{u}$ is perpendicular to ∇g_i for all i then it must be perpendicular to ∇f. This means that ∇f must be a linear combination of the ∇g_i. (To see this, write ∇f as the sum of a vector $(\nabla f)_\parallel$ in the span of ∇g_i and a vector $(\nabla f)_\perp$ in the orthogonal complement of the g_i. Taking $\mathbf{r}_0 = (\nabla f)_\perp$ shows that $(\nabla f)_\perp$ must be zero.) □

Proposition 2 (Necessary/Sufficient Condition for Local Extremum)**.** *Suppose that in a neighborhood of $\mathbf{r}_0$ $f : \mathbb{R}^n \mapsto \mathbb{R}$ and $g_i : \mathbb{R}^n \mapsto \mathbb{R}$ are smooth functions and $\nabla g(\mathbf{r}_0) \neq 0$. A sufficient [respectively necessary] condition for f to have a local minimum (respectively local maximum) at $\mathbf{r}_0$ subject to the constraints $g_i(\mathbf{r}) = g_i(\mathbf{r}_0)$ is: there exist constants λ_i such that for the function $f_L(\mathbf{r}) = f(\mathbf{r}) - \sum_i \lambda_i g_i$, $\nabla f_L(\mathbf{r}_0) = 0$ and $\nabla\nabla f_L(\mathbf{r}_0)$ is [semi]positive (respectively [semi]negative) definite when restricted to the orthogonal complement of the constraint gradients g_i.*

Justification of 2. We show the case where positive definite implies local minimum. Assume the hypotheses. It is enough to show that $\frac{d^2}{dt^2}f(\mathbf{r}(t))|_{t=0} > 0$ if $\mathbf{r}'(0) \neq 0$. But $\frac{d^2}{dt^2}f(\mathbf{r}(t)) = \frac{d}{dt}(\mathbf{r}'(t) \cdot \nabla f(\mathbf{r}(t))) = \mathbf{r}''(t) \cdot \nabla f + \mathbf{r}' \cdot (\nabla\nabla f) \cdot \mathbf{r}'$. But $\nabla f(\mathbf{r}_0) = \sum_i \lambda_i \nabla g_i(\mathbf{r}_0)$, and twice differentiating $g_i(\mathbf{r}(t)) = 0$ shows that $\mathbf{r}'' \cdot \nabla g_i = -\mathbf{r}' \cdot (\nabla\nabla g_i) \cdot \mathbf{r}'$, so $\frac{d^2}{dt^2}f(\mathbf{r}(t))|_{t=0} = \mathbf{r}' \cdot \nabla\nabla(f - \sum_i \lambda_i g_i) \cdot \mathbf{r}'$, which is positive by the assumption that $f_L := f - \sum_i \lambda_i g_i$ is positive definite for vectors $\mathbf{r}'(0)$ in the orthogonal complement of the constraint gradients (i.e. the tangent space of the constraint manifold). □

Another way to view Lagrange multipliers is that they are a means of penalizing "infeasabilities", i.e. perturbations in disallowed directions. Lagrange multipliers solve constrained optimization problems by adding a term whose derivative is always zero precisely for allowed perturbations. In particular, suppose $h(\mathbf{r}, \lambda) := f(\mathbf{r}) + \lambda g(\mathbf{r})$ has an extremum at $\mathbf{r}_0, \lambda_0$. Then at this point $0 = d_t h(\mathbf{r}(t), \lambda(t)) = r' \cdot \nabla(f + \lambda g) + \lambda' g$ $(\forall \mathbf{r}', \lambda')$. Note that this is equivalent to $\nabla(f + \lambda g) = 0$ and $g = 0$. But this implies that $\mathbf{r}' \cdot \nabla f = 0$ if $\mathbf{r}' \cdot \nabla g = 0$, i.e., $d_t f(\mathbf{r}) = 0$ if $d_t g(\mathbf{r}) = 0$, which is what it means for f to have a stationary point at $\mathbf{r}_0$ subject to the constraint $g = 0$. (I don't see how to reverse this proof to show the existence of a lambda without resorting to an argument such as what I used here for Proposition 1).