Lagrange Multipliers

May 13, 2020

Abstract

We consider a special case of Lagrange Multipliers for constrained optimization. The class quickly sketched the "geometric" intuition for Lagrange multipliers, but let's consider a short algebraic deriviation.

We consider a function with linear constraints:

$$f(x) \text{ subject to } Ax = b$$

Here $f : \mathbb{R}^d \to \mathbb{R}$, $x \in \mathbb{R}^d$, $A \in \mathbb{R}^{n \times d}$, and $b \in \mathbb{R}^n$. The technique we consider is to turn the problem from a constrained problem into an unconstrained problem using the Lagrangian,

$$L(x, \mu) = f(x) + \mu^T (Ax - b) \text{ in which } \mu \in \mathbb{R}^n$$

We'll show that the critical points of the constrained function f are critical points of $L(x, \mu)$.

Finding the Space of Solutions Assume the constraints are satisfiable, then let x_0 be such that $Ax_0 = b$. Let $\textbf{rank}(A) = r$ then let $\{u_1, \dots, u_k\}$ be an orthonormal basis for the null space of A in which $k = n - r$. We write this basis as a matrix:

$$U = [u_1, \dots, u_k] \in \mathbb{R}^{d \times k}$$

Since U is a basis, any solution for $f(x)$ can be written as $x = x_0 + Uy$. This captures all the *free parameters* of the solution. Thus, we consider the function:

$$g(y) = f(x_0 + Uy) \text{ in which } g : \mathbb{R}^k \to \mathbb{R}$$

The critical points of g are critical points of f. Notice that g is unconstrained, so we can use standard calculus to find its critical points.

$$\nabla_y g(y) = 0 \text{ equivalently } \nabla f(x_0 + Uy)U = 0.$$

To make sure the types are clear: $\nabla_y g(y) \in \mathbb{R}^k$, $\nabla f(z) \in \mathbb{R}^d$ and $U \in \mathbb{R}^{n \times k}$. In both cases, 0 is the 0 vector in $\mathbb{R}^k$.

The above condition says that z is a critical point for g then $\nabla f(x)$ must be *orthogonal* to U. However, U is the null space of A and the rowspace is orthogonal to it. In particular, observe that for any element of the rowspace $z = A^T \mu \in \mathbb{R}^d$ then for any $u = Uy$ is orthogonal, since:

$$z^T u = \mu^T A u = \mu^T A u = \mu^T 0 = 0$$

That is, $Au = 0$.

Thus, we can rewrite this orthogonality condition as there is some $\mu \in \mathbb{R}^n$ (depending on x) such that

$$\nabla f(x) + \mu^T A = 0.$$

The Clever Lagrangian We now observe that the critical points of the Lagrangian are (by differentiating and setting to 0)

$$\nabla_x L(x, \mu) = \nabla f(x) + \mu^T A = 0 \text{ and } \nabla_\mu L(x, \mu) = Ax - b$$

The first condition is exactly the condition that x be a critical point and the second condition says the that the constraint be satisfiied. Thus, if x is a critical point there exists some μ as above, and (x, μ) is a critical point for L.

Generalizing to Nonlinear Equality Constraints This is a much more general technique. If you want to handle non-linear equality constraints, then the main idea is the same and you need a little extra machinery: the implicit function theorem. However, the key idea is taht you find the space of solutions and you optimize. In that case,

$$f(z) \text{ s.t. } g(x) = c \text{ leads to } L(x, \mu) = f(x) + \mu^T (g(x) - c)$$

The gradient condition here is $\nabla f(x) = \mu^T \nabla g(x)$, which we can view as saying, *"at a critical point, the gradient of the surface must be parallel to the gradient of the function."* This connects us back to the picture that we drew during lecture.

Extra Resources If you find resources you like, post them on Piazza!