Lagrange Multipliers

May 16, 2020

Abstract

We consider a special case of Lagrange Multipliers for constrained optimization. The class quickly sketched the "geometric" intuition for Lagrange multipliers, and this note considers a short algebraic derivation.

In order to minimize or maximize a function with linear constraints, we consider finding the critical points (which may be local maxima, local minima, or saddle points) of

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

Here $f : \mathbb{R}^d \to \mathbb{R}$ is a convex (or concave) function, $x \in \mathbb{R}^d$, $A \in \mathbb{R}^{n\times d}$, and $b \in \mathbb{R}^n$. To find the critical points, we cannot just set the derivative of the objective equal to 0.[1] 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 $\mathsf{rank}(A) = r$, then let $\{u_1, \ldots, u_k\}$ be an orthonormal basis for the null space of A in which $k = d-r$. Note if $k = 0$, then x_0 is uniquely defined. So we consider $k > 0$. We write this basis as a matrix:

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

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.

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

[1] See the example at the end of this document.

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

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

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

Since we can decompose $\mathbb{R}^d$ as a direct sum of $\mathsf{null}(A)$ and the rowspace of A, we know that any vector orthogonal to U must be in the rowspace. We can rewrite this orthogonality condition as follows: there is some $\mu \in \mathbb{R}^n$ (depending on x) such that

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

for a certain x such that $Ax = A(x_0 + Uy) = Ax_0 = b$.

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

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

The first condition is exactly the condition that x be a critical point in the way we derived it above, and the second condition says that the constraint be satisfied. 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 Lagrange multipliers are a much more general technique. If you want to handle non-linear equality constraints, then you will need a little extra machinery: the implicit function theorem. However, the key idea is that you find the space of solutions and you optimize. In that case, finding the critical points of

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

The gradient condition here is $\nabla f(x) + J^T\mu = 0$, where J is the Jacobian matrix of g. For the case where we have a single constraint, the gradient condition reduces to $\nabla f(x) = -\mu_1 \nabla g_1(x)$, 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.

Example: Need for constrained optimization We give a simple example to show that you cannot just set the derivatives to 0. Consider $f(x_1, x_2) = x_1$ and $g(x_1, x_2) = x_1^2 + x_2^2$ and so:

$$\max_x f(x) \text{ subject to } g(x) = 1.$$

This is just a linear functional over the circle, and it is compact, so the function must achieve a maximum value. Intuitively, we can see that $(1, 0)$ is the maximum possible value (and hence a critical point). Here, we have:

$$\nabla f(x) = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \text{ and } \nabla g(x) = 2 \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

Notice that $\nabla f(x)$ is not zero anywhere on the circle–it's constant! For $x \in \{(1, 0), (-1, 0)\}$, $\nabla f(x) = \lambda \nabla g(x)$ (take $\lambda \in \{1/2, -1/2\}$, respectively). On the other hand, for any other point on the circle $x_2 \neq 0$, and so the gradient of f and g are *not* parallel. Thus, such points are not critical points.

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