Lagrange Multipliers

Suhas Arehalli

December 13, 2015

1 Preliminaries

Definition 1.1 (Manifold)**.** A smooth k-manifold M in $\mathbb{R}^n$ can be defined as a set of points

$$M = \{\mathbf{x} \in \mathbb{R}^n | F(\mathbf{x}) = \mathbf{0}\}$$

For some function $F : \mathbb{R}^n \to \mathbb{R}^{n-k}$ such that $[\mathbf{D}F(\mathbf{x})]$ has maximal rank for all $\mathbf{x}$. This is equivalent to the conventional definition in terms of local charts and such.

Definition 1.2 (Tangent Space)**.** The tangent space to a manifold at point $\mathbf{a}$ is defined as the set of vectors that locally describe movement on the manifold. Since staying on the manifold requires that it's defining constaint function F remains 0, we can formally define the tangent space at $\mathbf{a}$ as

$$T_{\mathbf{a}}M = \{\mathbf{x} \in \mathbb{R}^n | [\mathbf{D}F(\mathbf{a})]\mathbf{x} = \mathbf{0}\}$$
$$T_{\mathbf{a}}M = ker[\mathbf{D}F(\mathbf{a})]$$

2 Background

Given a function $f : \mathbb{R}^n \to \mathbb{R}$, we'd like to find a vector $\mathbf{x}$ on a smooth k-manifold M embedded in $\mathbb{R}^n$. That maximizes or minimizes the function f. This is called *Constrained Optimization*, and can be solved using the method of Lagrange Multipliers.

Due to the definition of a manifold we gave earlier, we can phrase this problem in a somewhat more accessable light. Suppose we have a system of p *constraint equations* of n variables represented by the equation $F(x) = 0$ for $F : \mathbb{R}^n \to \mathbb{R}^p$. This defines a set of points in $\mathbb{R}^n$ that satisfy these equations, and if we find that the derivative of F is onto, then we know that these points form a smooth manifold. Now given these constraints, it seems natural in many practical scenarios that we'd want to find extremum that satisfy these constaints. For instance, we'd like to know how to minimize cost or maximize profit or utility given some set of constraints. Given some cost or utility function f, Lagrange multipliers give us a way to answer that question.

3 Statement

Theorem 3.1 (Method of Lagrange Multipliers)**.** *Suppose M is a smooth k-manifold described by the function $F : \mathbb{R}^n \to \mathbb{R}^{n-k}$ such that $F(\mathbf{x}) = 0$ and $\mathbf{a}$ is an extremum of some function $f : \mathbb{R}^n \to \mathbb{R}$ on M. Then*

$$[\mathbf{D}f(\mathbf{a})] = \sum_{i}^{n-k} \lambda_i \, [\mathbf{D}F_i(\mathbf{a})]$$

Or in terms of gradients,

$$\nabla f(\mathbf{a}) = \sum_{i}^{n-k} \lambda_i \nabla F_i(\mathbf{a})$$

$$\nabla f(\mathbf{a}) \in \mathit{span}\left(\{\nabla F_1(\mathbf{a}), \ldots, \nabla F_{n-k}(\mathbf{a})\}\right)$$

4 Intuition

First let's discuss *why* this should be true, and then discuss a strategy for proving it.

Naively, we begin to optimize functions by using the first derivative test - a function is at a local extremum if movement in any direction will no longer increase/descrease the value of the function. More precisely, if the directional derivative is 0 for all vectors ($[\mathbf{D}f(\mathbf{a})]\mathbf{v} = \mathbf{0}, \forall \mathbf{v} \in \mathbb{R}^n$), then we're at a potential extremum.

Note, however, that in order to maximize f on the manifold, we cannot simply seek out a maximum for f. There's no guarantee that any maximum of f exists in $\mathbb{R}^n$, let alone lies on M. Thus we must begin working from the manifold, rather than from the function we optimize.

The main benefit of working on the manifold is that we have few directions we can move, and thus it is easier to be a local extremum. Consider that the gradient ∇f is considered the direction of greatest increase, and it's negation is the direction of greatest decrease. If we want to be at an extremum, these must be orthogonal to the manifold - no movement on the manifold can change the value of f. However, the consider each of the individual constraint functions, $F_1, ..., F_{n-k}$. Each of these has a gradient function ∇F_i which is orthogonal to the manifold - In fact, these span the orthogonal complement of the tangent space of the manifold at a point. Thus, at an extrema, ∇f should be in the span of $\{\nabla F_1, ..., \nabla F_{n-k}\}$, which is exactly what the method of Lagrange Multipliers says.

5 Proof

At a local maximum of f on M, $\mathbf{a}$, we must be able to move in every direction on the manifold without increasing the value of the f (otherwise, a point ε in that direction would give a greater value of f than $\mathbf{a}$, contradicting the fact that $\mathbf{a}$ is a local maximum. The same argument goes for minima - we must be able to move in every direction without decreasing f. Since derivatives are continuous on smooth manifolds, we must have

$$\forall \mathbf{v} \in T_{\mathbf{a}}M = \ker[\mathbf{D}F(\mathbf{a})]$$

that

$$\begin{aligned}[\mathbf{D}f(\mathbf{a})]\mathbf{v} &= 0 \\ \mathbf{v} &\in \ker[\mathbf{D}f(\mathbf{a})]\end{aligned}$$

Which implies that

$$\ker[\mathbf{D}F(\mathbf{a})] \subset \ker[\mathbf{D}f(\mathbf{a})]$$

This is often one stopping point for this result, but deriving the original equation requires further manipulation. First note that since $[\mathbf{D}F(\mathbf{a})]\mathbf{v} = \mathbf{0}$ (since $\mathbf{v} \in \ker[\mathbf{D}F(\mathbf{a})]$), then $\forall i \in \{1, \ldots, n-k\}$

$$[\mathbf{D}F_i(\mathbf{a})]\mathbf{v} = 0$$

plus, from the above

$$[\mathbf{D}f(\mathbf{a})]\mathbf{v} = 0$$

For all $\mathbf{v} \in T_{\mathbf{a}}M$. We can consider these in terms of gradients:

$$\begin{aligned}\nabla F_1 \cdot \mathbf{v} &= 0 \\ &\cdots \\ \nabla F_{n-k} \cdot \mathbf{v} &= 0 \\ \nabla f \cdot \mathbf{v} &= 0\end{aligned}$$

Now it is clear to see that $\{\nabla f, \nabla F_1, ..., \nabla F_{n-k}\}$ lie in the orthogonal complement to the Tangent Space $T_{\mathbf{a}}M$ (they're all orthogonal to every vector in the tangent space!).

Now, since they both live in $\mathbb{R}^n$, the orthogonal complement should have dimension $n - dim(T_{\mathbf{a}}M) = n - k$, since the sum of dimension of any space and it's orthogonal complement must be the dimension of the space it's embedded in. Now we have $n - k + 1$ vectors in a space with dimension $n - k$. Thus they cannot be linearly independent, and one must be a linear combination of the others. thus we can write

$$\nabla f = \sum_{i}^{n-k} \lambda_i \nabla F_i$$

Or, back in matrix notation,

$$[\mathbf{D}f(\mathbf{a})] = \sum_{i}^{n-k} [\mathbf{D}F_i(\mathbf{a})]$$

Concluding our proof.