DRP HW 2 - Lagrange Multipliers

Owen Huang

July 3, 2024

Notes

This homework proves the Lagrange multiplier theorem in its full generality, using some tools from analysis which we will just blackbox. I will include some problems at the end to practice with actual functions.

1 Analysis Prelim and Warmup

You may use any results here without proof (though you should make sure to state when you are invoking a theorem).

Theorem 1.1 (Mean Value Theorem)**.** *Let $f : [a, b] \to \mathbb{R}$ be continuous, and differentiable in (a, b). There exists a $c \in (a, b)$ for which*

$$f'(c) = \frac{f(b) - f(a)}{b - a}$$

Lemma 1.2. *Suppose $f : \mathbb{R} \to \mathbb{R}$ is continuous, and $f(x) > c$ for some x. Then there exists $\epsilon > 0$ so that $f(y) > c$ for every $y \in (x - \epsilon, x + \epsilon)$*

Warmup Problem

We are going to prove the following theorem, which you must have learned in calculus class, but are not quite sure why it is correct!

Theorem 1.3. *Suppose that $f : [a, b] \to \mathbb{R}$ is continuously differentiable. If $x^* \in (a, b)$ is a local extrema, then $f'(x^*) = 0$.*

Proof. **a**

We are going to prove this statement by contradiction. Suppose that x^* is a local extrema, but $f'(x^*) \neq 0$. Without loss of generality, suppose that $f'(x^*) > 0$. We will show that x^* cannot be a local extrema by producing an y close to x^* so that $f(y)$ is either larger or smaller than $f(x^*)$, depending on if your extrema is a maximum or a minimum. Use the above lemma to find an interval around x^* so that the derivative is positive throughout.

b

Apply the mean value theorem to your chosen interval. Make sure you choose your endpoints carefully to obtain the result you want. [hint: one endpoint should be x^* itself!].

c

Now, f' is positive in the entire interval, so how can we use this to derive a contradiction? $\square$

In fact, the above result is true for multiple dimensions - that if x maximizes $f : \mathbb{R}^n \to \mathbb{R}$, then $\nabla f(x) = 0$. We will use this the proof od the main theorem of the next section. But first, we need to discuss the implicit function theorem.

Theorem 1.4 (Implicit Function Theorem)**.** *Let $f : \mathbb{R}^{n+m} \to \mathbb{R}^m$ be continuously differentiable, and suppose that $f(\vec{x}, \vec{y}) = 0$ at some $\vec{x} \in \mathbb{R}^n$ and $\vec{y} \in \mathbb{R}^m$. Suppose further that the Jacobian of f, when restricted to the final m columns is invertible (which we shall denote by $J_m(f)$). Then there exists a ball $B \subset \mathbb{R}^n$ of finite radius containing $\vec{x}$ and a unique continuously differentiable function $g : B \to \mathbb{R}^m$ satisfying $g(\vec{x}) = \vec{y}$ and $f(\vec{x}, g(\vec{x})) = 0$ for all $\vec{x} \in B$. Moreover, the Jacobian of g, evaluated at a point $\vec{x} \in \mathbb{R}^n$ is given by*

$$J(g)[\vec{x}] = -J_m(f)^{-1}[\vec{x}]J(f)[\vec{x}]$$

Okay. Looks messy I know. But let me take a second to explain the notation at the bottom. Recall that for a function $f : \mathbb{R}^m \to \mathbb{R}^n$, we have that $f(x_1, \ldots, x_m) = (y_1, \ldots y_n)$. Because the output is just n real numbers, all independent of each other, we can just look at each of them individually. Explicitly, for any y_i, we can consider it as a real-valued function $y_i = f_i(x_1, \ldots x_m)$. But now, we are back in the real-valued case so we may take the partial derivatives $\partial f_i / \partial x_j$. The **Jacobian** of f, $J(f)$ is the $n \times m$ matrix containing all of the possible partial derivatives:

$$J(f) = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_m} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_n}{\partial x_1} & \cdots & \frac{\partial f_n}{\partial x_m} \end{bmatrix}$$

Now, as written, each component in the matrix is still a function - so to make this truly a matrix, we'd have to evaluate it at a point. That is, for some $\vec{x} \in \mathbb{R}^m$, need to define

$$J(f)[\vec{x}] = \begin{bmatrix} \frac{\partial f_1}{\partial x_1}(\vec{x}) & \cdots & \frac{\partial f_1}{\partial x_m}(\vec{x}) \\ \vdots & \ddots & \vdots \\ \frac{\partial f_n}{\partial x_1}(\vec{x}) & \cdots & \frac{\partial f_n}{\partial x_m}(\vec{x}) \end{bmatrix}$$

Now, in the statement of the theorem, f is a function from $\mathbb{R}^{n+m}$ to $\mathbb{R}^m$, so its Jacobian has dimensions $m \times (n + m)$. Explicitly,

$$J(f) = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} & \frac{\partial f_1}{\partial x_{n+1}} & \cdots & \frac{\partial f_1}{\partial x_{n+m}} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1} & \cdots & \frac{\partial f_m}{\partial x_n} & \frac{\partial f_m}{\partial x_{n+1}} & \cdots & \frac{\partial f_m}{\partial x_{n+m}} \end{bmatrix}$$

Where the last m columns form $J_m(f)$ and they correspond to the partial derivatives involving the last m entries of the input.

2 The Lagrange Multiplier Theorem

We are now ready to state and prove the *Lagrange Multiplier Theorem*, and important result that shows up often in machine learning and constrained optimization. Intuitively, it provides necessary conditions for "optimal points" of a function. Of course, these conditions are not also sufficient, but in most cases they will be (ie., many functions we optimize are convex, and it turns out that is enough to make this condition an "if and only if"). Here f will be our *objective function*, which we are trying to optimize, and g is a *constraint function* which somehow captures the limits to which we can optimize. The function g is important to place boundedness/compactness conditions on our problem, because if we didn't, then many objective functions could possibly shoot off to infinity!

Theorem 2.1 (Lagrange Multipliers). *Let $f : \mathbb{R}^n \to \mathbb{R}$ and $g : \mathbb{R}^n \to \mathbb{R}^d$ continuously differentiable, with $d < n$. For any $\vec{c}$ in the range of g, denote $M_{\vec{c}} = \{\vec{x} \in \mathbb{R}^n \mid g(\vec{x}) = \vec{c}\} \subseteq \mathbb{R}^n$ to be the level sets of g. Let $\vec{x}^*$ be a local extremum of $f(\vec{x})$ subject to the constraint that $\vec{x} \in M_{\vec{c}}$, and suppose that $J(g)[\vec{x}^*]$ has rank n. Then there is a unique vector $\vec{\lambda} \in \mathbb{R}^d$ so that $\nabla f[\vec{x}^*] = \lambda J(g)[\vec{x}^*]$.*

Before we prove this theorem,

a

Rewrite the statement of the theorem to the familiar case $f, g : \mathbb{R}^3 \to \mathbb{R}$, and use the gradient operator ∇ instead of the Jacobian [why are these the same when the function is real-valued?]

Proof. **b**

Let x^* be a constrained local extrema of f. By rearranging coordinates, we can assume without loss of generality that the last d columns of $J(g)[\vec{x}^*]$ are all linearly independent. Apply the implicit function theorem to find a ball $B \subset \mathbb{R}^n$ containing $\vec{x}^*$, a subset $V \subset \mathbb{R}^{n-d}$ and a continuously differentiable function $h : B \to \mathbb{R}^d$ so that

$$M_{\vec{c}} \cap B = \{(\vec{x}, h(\vec{x})) \mid \vec{x} \in V\}$$

c

Let us now write vectors $\vec{x} \in \mathbb{R}^n$ as $(\vec{x}_1, \vec{x}_2)$ where the second component is in $\mathbb{R}^d$ and the first is in $\mathbb{R}^{n-d}$. Consider the function $F : \mathbb{R}^n \to \mathbb{R}$ given by $F(\vec{x}_1, \vec{x}_2) = f(\vec{x}_1, h(\vec{x}_1))$. Now, if $\vec{x}^*$ is a constrained extrema of f, then it is an unconstrained extrema of F! By the warmup problem to this homework, we have that $\nabla F(\vec{x}^*) = 0$. Denote ∇F_{n-d} to be the first $n - d$-coordinates of this vector and ∇F_d to be the final d coordinates. However, as F is a composition of two functions (f and h), we can use the chain rule to obtain an explicit expression for it. Use the multivariable chain rule (ie., $\nabla F = J(f)J(h)$ to show that

$$\nabla F_{n-d}(\vec{x}^*) + \nabla F_d(\vec{x}^*)J(h)[(\vec{x}^*)] = 0$$

d

Now we have this pesky $J(h)$ term in the equation which we would like to get rid of. To do so, we need to use the constraint that $f(\vec{x}^*, h(\vec{x}^*)) = \vec{c}$. Remember that we are trying to optimize on $M_{\vec{c}}$ so the function $F(\vec{x}, h(\vec{x}))$ is constant when $\vec{x} \in V$. Use implicit differentiation to show that

$$J(h)[\vec{x}^*] = -J(g)_d$$

□