Columbia MA Math Camp

Optimization

Vinayak Iyer [a]

July 22, 2020

[a] Material adapted from notes by David Thompson and Xingye Wu

Table of Contents

Some Definitions

Let $\mathcal{D} \subseteq \mathbb{R}^n$ and $f : \mathcal{D} \to \mathbb{R}$.

- A point c is a **maximum point** or **global maximum** of f if $f(c) \geq f(x)$ for all $x \in \mathcal{D}$
- A point c is a **local maximum** of f if there exists an $\epsilon > 0$ such that $f(c) \geq f(x)$ for all $x \in B_\epsilon(c)$
- If f is differentiable, a point such that $f'(c) = 0$ is a **critical point** of f

Maximum and minimum points are also called **extreme points, extremum** and **optimal points**

First-Order Conditions

Proposition 1.1
Let $\mathcal{D} \subseteq \mathbb{R}^n$ be an open set, and $f : \mathcal{D} \to \mathbb{R}$ a differentiable function. If f has a local extreme point *at x, then $f'(x) = 0$*

Notes : This could mean that x is either a maximum or minimum or neither

Proposition 1.2
Let $\mathcal{D}$ be an open set of $\mathbb{R}^n$ and $f : \mathcal{D} \to \mathbb{R}$ a C^2 function.

- *If f has a* local maximum (minimum) *at x, the Hessian of f at x is* negative (positive) semi-definite
- *If $f'(x) = 0$ and $H(x)$ is* negative (positive) definite, *then x is a* strict local maximum (minimum)

(Proof of second result): Since $H(x)$ is negative definite, there exists $\epsilon > 0$ such that $H(\zeta)$ is negative definite for all $\zeta \in B_\epsilon(x)$. Using the exact form of Taylor's theorem, for any $y \in B_\epsilon(x)$:

$$\begin{aligned} f(y) &= f(x) + f'(x)(y-x) + \frac{1}{2}(y-x)^T H(\zeta)(y-x) \\ &< f(x) \end{aligned}$$

Summarizing

- Critical points are a **necessary (not sufficient)** condition for extrema
- The second derivative can give us **local sufficient conditions**

To this point we've only discussed maximization over an open set

- Maxima need not exist on an open set (e.g. $f(x) = x$ on $(0, 1)$)
- If you're maximizing over a closed set S, you can decompose it as $S = int(S) \cup Boundary(S)$. Need to check the boundary

General recipe to maximize a function :

- Find all critical points. If there are many, SOC can help filter
- Find the critical point with the largest value
- Check if the function takes on a higher value along the boundary

Sufficient Conditions for Global Extrema

For convex functions, optimization is dramatically simpler, as evidenced by the following proposition :

Proposition 1.3
Let $\mathcal{D}$ be a convex open set of $\mathbb{R}^n$ and $f : \mathcal{D} \to \mathbb{R}$ be a convex function. Then :

- *The set of minimizers of f is* convex
- *If f is* strictly convex, *it has* at most one minimizer
- *Any* local minimum *of f is a* global minimum
- *If f is* differentiable, *then x is a* global minimum *of f iff $f'(x) = 0$*

The same results hold for concave functions, replacing "minimizers" with "maximizers"

Note :

- Note that we already proved the second result for quasiconvex functions.

Table of Contents

Constrained Optimization Problems

- In economics, it's more common to maximize an objective function subject to some constraints
- For example, in consumer theory, you will see problems of this form:

$$\max_{c_1, c_2} \log c_1 + \alpha \log c_2 \ \text{ s.t. } \ p_1 c_1 + p_2 c_2 = M$$

- One approach to these problems is to put the constraints in the objective. For instance, in the above example

$$c_1 = \frac{M - p_2 c_2}{p_1}$$

So we could do the unconstrained maximization problem:

$$\max_{c_2} \log \left(\frac{M - p_2 c_2}{p_1} \right) + \alpha \log c_2$$

[Do More Examples]

A more formal statement

Theorem 2.1
Let $f : \mathbb{R}^n \to \mathbb{R}$ *and* $g : \mathbb{R}^n \to \mathbb{R}^k$ *be* C^1 *functions, and consider the program:*

$$\max_{x \in \mathbb{R}^n} f(x) \text{ s.t. } g(x) = 0$$

If x^* *is a local maximum and* x^* *satisfies the constraint qualification ;* $rank(g'(x^*)) = k$, *then there exist* k Lagrange multipliers $\lambda = (\lambda_1, ..., \lambda_k)^T \in \mathbb{R}^k$ *such that the first-order condition holds:*

$$f'(x^*) + \lambda^T g'(x^*) = 0$$

It is common to talk about the **Lagrangian** of a system:

$$\mathcal{L}(x, \lambda) = f(x) + \lambda^T g(x)$$

The first-order conditions wrt x and λ give us the critical point of the Lagrangian.

A sketch of Lagrange's Theorem in two variables

- Write $x = (x_1, x_2)$. Let $x^* = (x_1^*, x_2^*)$ be a local maximum of f subject to g
- By the IFT, we can write $x_2 = h(x_1)$, with $h'(x_1) = -\frac{g_1(x_1,x_2)}{g_2(x_1,x_2)}$
- We now do unconstrained optimization of $f(x_1, h(x_1))$. The FOC is

$$f_1(x^*) + f_2(x^*)h'(x_1^*) = 0$$

- Define $\lambda = -\frac{f_2(x^*)}{g_2(x^*)}$. Then

$$\begin{aligned} f_1 + \lambda g_1(x^*) &= 0 \\ f_2 + \lambda g_2(x^*) &= 0 \end{aligned}$$

The general case is similar, just with more cumbersome matrix notation.

Comments on Lagrange's Theorem

- The Lagrange condition is a necessary condition.
- As with unconstrained optimization, there are second order conditions that let you check whether a critical point of the Lagrangian is a local maximum or minimum (see FMEA Section 3.4 for details)
- In order to get sufficient conditions for global maxima along the constraint, we need additional structure
- If the constraint qualification fails, the theorem says nothing. So you need to check points where the CQ fails separately

An example of a sufficient condition

Proposition 2.1
Let $f : \mathbb{R}^n \to \mathbb{R}$ be strictly quasiconcave and consider the program

$$\max_x f(x) \text{ s.t. } Ax = b \tag{1}$$

where A is an $m \times n$ matrix with $m < n$. If (x^*, λ^*) is a critical point of the Lagrangian and $f'(x^*) \neq 0$, then x^* solves (1).

Proof.
The FOC of the Lagrangian implies $f'(x^*) + \lambda^T A = 0$. Suppose there were an $\hat{x}$ such that $A\hat{x} = b$ and $f(\hat{x}) > f(x^*)$. Since f is strictly quasiconcave:

$$\begin{aligned} 0 &< f'(x^*)(\hat{x} - x^*) \\ &= -\lambda^T A(\hat{x} - x^*) \\ &= 0 \end{aligned}$$

a contradiction. □

Note : $f(x)$ being strictly quasiconcave and the constraint being linear ensures that the Lagrangian is strictly quasiconcave as well (which we have shown before implies a unique maximizer)

Interpretation of the multipliers

Define the **"value function"** V as follows

$$V(b) = \max_{x} f(x) \text{ s.t. } g(x) = b$$

Form the Lagrangian :

$$\mathcal{L}(x, \lambda, b) = f(x) + \lambda^T (b - g(x))$$

Write the solution of this problem as $x^*(b), \lambda^*(b)$. Then

$$V(b) = \mathcal{L}(x^*(b), \lambda^*(b), b)$$

Interpretation of the multipliers (cont.)

Using the chain rule, we have :

$$V'(b) = \frac{\partial \mathcal{L}}{\partial x}\frac{dx^*}{db} + \frac{\partial \mathcal{L}}{\partial \lambda}\frac{d\lambda^*}{db} + \frac{\partial \mathcal{L}}{\partial b}$$

However, we know $\frac{\partial \mathcal{L}}{\partial x}$ and $\frac{\partial \mathcal{L}}{\partial \lambda}$ are 0 at x^*, λ^*, so we have

$$V'(b) = \lambda^T$$

Interpretation: Lagrange multipliers measure the marginal value of loosening a constraint by 1 unit