Static Optimization

Divya Bhagia

Boston College

August 22, 2019

Overview

Overview

1. Unconstrained Optimization
2. Constrained Maximization: Equality Constraints
3. Constrained Maximization: Inequality Constraints
4. Envelope Theorem

Unconstrained Maximization

Everywhere in today's lecture we will consider a function $f : \mathbb{R}^N \to \mathbb{R}$.

Definition

The vector $\bar{x} \in \mathbb{R}^N$ is a **local maximizer** of f if there is an open neighborhood of $\bar{x}$, $A \subset \mathbb{R}^N$, such that $f(\bar{x}) \geq f(x)$ for every $x \in A$. If $f(\bar{x}) \geq f(x)$ for every $x \in \mathbb{R}^N$, then we say that $\bar{x}$ is a **global maximizer** of f (or simply a maximizer).

The concepts of **local minimizer** and **global minimizer** are defined analogously.

Unconstrained Maximization: FOC (Necessary)

Theorem

Suppose that f is differentiable and that $\bar{x} \in \mathbb{R}^N$ is a local maximizer or minimizer of f. Then $\bar{x}$ is a critical point of f defined as

$$\nabla f(\bar{x}) = 0$$

Proof: Denote by $e^n \in \mathbb{R}^N$ the vector having it's nth entry equal to 1 and all other entries equal to 0. Suppose that $\bar{x}$ is a local maximizer or minimizer of f but $df(\bar{x})/dx_n = a > 0$ for some n (analogous argument if $a < 0$). By definition of partial differential

$$\frac{df(\bar{x})}{dx_n} = \lim_{\varepsilon \to 0} \frac{f(\bar{x} + \varepsilon e^n) - f(\bar{x})}{\varepsilon} = \lim_{\varepsilon \to 0} \frac{f(\bar{x} - \varepsilon e^n) - f(\bar{x})}{-\varepsilon} > 0$$

This implies that for some arbitrarily small $\varepsilon > 0$,

$$f(\bar{x} - \varepsilon e^n) < f(\bar{x}) < f(\bar{x} + \varepsilon e^n)$$

So $\bar{x}$ cannot be a local maximizer or minimizer of f. □

Unconstrained Maximization: FOC (Not Sufficient)

By the previous theorem every local maximizer or minimizer is a critical point but the converse is not true. Example:

$$f(x_1, x_2) = x_1^2 - x_2^2$$

$$\nabla f(x) = \begin{bmatrix} \frac{\partial f(x_1, x_2)}{\partial x_1} \\ \frac{\partial f(x_1, x_2)}{\partial x_2} \end{bmatrix} = \begin{bmatrix} 2x_1 \\ 2x_2 \end{bmatrix}$$

At the origin $\nabla f(0, 0) = 0$, but it is neither a local maximizer or minimizer of f.

Unconstrained Maximization: SOC

Theorem

Suppose that the function $f : \mathbb{R}^N \to \mathbb{R}$ is twice continuously differentiable and $\bar{x}$ is a critical point i.e. $\nabla f(\bar{x}) = 0$.

1. If $\bar{x} \in \mathbb{R}^N$ is a local maximizer, then the (symmetric) $N \times N$ matrix $D^2 f(\bar{x})$ is negative semidefinite. (*Necessary*)
2. If $D^2 f(\bar{x})$ is negative definite, then $\bar{x}$ is a local maximizer. (*Sufficient*)

Replacing negative by positive, the same is true for local minimizers.

Proof: For an arbitrary direction of displacement $z \in \mathbb{R}^N$ and scaler ε, mean-value expansion of the function $f(\bar{x} + \varepsilon z)$

$$f(\bar{x} + \varepsilon z) - f(\bar{x}) = \varepsilon \nabla f(\bar{x})^T z + \frac{1}{2}\varepsilon^2 z^T D^2 f(\bar{x} + \alpha \varepsilon z) z$$

for some $\alpha \in (0, 1)$. Since $\nabla f(\bar{x}) = 0$ and $(1/\varepsilon^2)[f(\bar{x} + \varepsilon z) - f(\bar{x})] \leq 0$, which implies that if ε is small enough then $z^T D^2 f(\bar{x}) z \leq 0$. In addition, if $z^T D^2 f(\bar{x}) z < 0$ for any $z \neq 0$ then $f(\bar{x} + \varepsilon z) < f(\bar{x})$ for small ε. □

Unconstrained Maximization: SOC

In the case when $D^2 f(\bar{x})$ is negative semidefinite but not negative definite, we cannot assert that $\bar{x}$ is a local maximizer. Consider for example:

$$f(x) = x^3$$

Then since

$$\frac{\partial^2 f(0)}{dx} = 0$$

$D^2 f(0)$ is negative semidefinite but 0 is neither a local maximizer nor a local minimizer of this function.

Unconstrained Maximization

Theorem

Any critical point $\bar{x}$ of a concave function f is a global maximizer of f. And analogously, any critical point of a covex function f is a global minimizer of f.

Proof: Recall that in the lecture on *Convexity* we proved that for concave functions we have,

$$f(x) \leq f(\bar{x}) + \nabla f(\bar{x})^T (x - \bar{x})$$

for every x in the domain. Since $\nabla f(\bar{x}) = 0$, $\bar{x}$ is a global maximizer. $\square$

Overview

Constrained Maximization: Equality Constraints

We will study the problem

$$\underset{x \in \mathbb{R}^N}{Max} \quad f(x) \quad s.t.$$
$$g_1(x) = b_1$$
$$\vdots$$
$$g_M(x) = b_m$$

where functions $f, g_1, \ldots, g_M$ are defined on $\mathbb{R}^N$ and $N \geq M$.

Constrained Maximization: Equality Constraints

Definition

The set of all $x \in \mathbb{R}^N$ satisfying the constraints of the problem above is called the **constraint set** and is denoted by

$$C = \left\{ x \in \mathbb{R}^N : g_m(x) = b_m \quad \forall m = 1, \ldots, M \right\}$$

Definition

The feasible point $\bar{x} \in C$ is a **local constrained maximizer** if there exists an open neighborhood of $\bar{x}$, say $A \subset \mathbb{R}^N$, such that $f(\bar{x}) \geq f(x)$ for all $x \in A \cap C$. The point $\bar{x}$ is a **global constrained maximizer** if $f(\bar{x}) \geq f(x)$ for all $x \in C$.

The concepts of **local constrained minimizer** and **global constrained minimizer** are defined analogously.

Langrange Method: FOC

Theorem

Suppose that the objective and constraint functions of the problem are differentiable and that $\bar{x} \in C$ is a local constrained maximizer. Assume also that the $M \times N$ matrix

$$\begin{bmatrix} \frac{dg_1(\bar{x})}{dx_1} & \dots & \frac{dg_1(\bar{x})}{dx_N} \\ & \ddots & \\ \frac{dg_M(\bar{x})}{dx_1} & \dots & \frac{dg_M(\bar{x})}{dx_N} \end{bmatrix}$$

has rank M (*constraint qualification*). Then there are numbers $\lambda_m \in \mathbb{R}$, one for each constraint, such that

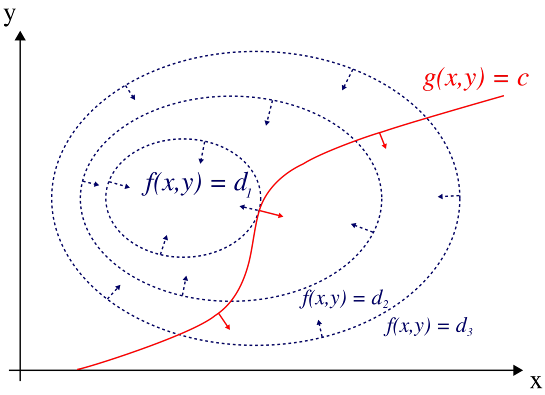
$$\frac{df(\bar{x})}{dx_n} = \sum_{m=1}^{M} \lambda_m \frac{dg_m(\bar{x})}{dx_n} \quad \text{for every} \quad n = 1, \dots, N \tag{1}$$

The numbers λ_m are referred to as **Lagrange multipliers**.

Langrange Method: Intuition

Let's consider the following problem in two dimensions

$$\max_{\{x,y\}\in\mathbb{R}^2} f(x,y) \quad \text{s.t.} \quad g(x,y) = c$$



Source: Wikipedia

Langrange Method: Proof

Proof

We will be using the implicit function theorem to prove this result. Let us divide vector $(x_1, ..., x_N)$ into two vectors $\mathbf{x_1} \in \mathbb{R}^{N-M}$ and $\mathbf{x_2} \in \mathbb{R}^M$. Let function $g : \mathbb{R}^{N-M} \times \mathbb{R}^M \to \mathbb{R}^M$ be defined as

$$g(x) = [g_1(x), .., g_m(x)]^T = [b_1, ..., b_M]^T = b$$

By implicit function theorem, in the neighborhood of $(\bar{\mathbf{x}}_1, \bar{\mathbf{x}}_2)$

$$g(\mathbf{x}_1, h(\mathbf{x}_1)) = 0$$

where h is C^1 and hence differentiating yields

$$\underbrace{\nabla_1 g(\mathbf{x}_1, h(\mathbf{x}_1))}_{M \times N-M} = -\underbrace{\nabla_2 g(\mathbf{x}_1, h(\mathbf{x}_1))}_{M \times M} \underbrace{\nabla h(\mathbf{x}_1)}_{M \times N-M}$$

▸ Implicit Function Theorem

Langrange Method: Proof

Proof (cont.)

Now the following two statements are equivalent

1. $(\bar{\mathbf{x}}_1, \bar{\mathbf{x}}_2)$ is a local max of $f(\mathbf{x}_1, \mathbf{x}_2)$ s.t. $g(\mathbf{x}_1, \mathbf{x}_2) = b$
2. $\bar{\mathbf{x}}_1$ is local max of $F(\mathbf{x}_1) = f(\mathbf{x}_1, h(\mathbf{x}_1))$

$$\begin{aligned}\nabla F(\bar{\mathbf{x}}_1) &= \nabla_1 f(\bar{\mathbf{x}}_1, h(\bar{\mathbf{x}}_1)) + \nabla_2 f(\bar{\mathbf{x}}_1, h(\bar{\mathbf{x}}_1)) \nabla h(\bar{\mathbf{x}}_1) \\ &= \nabla_1 f(\bar{\mathbf{x}}_1, h(\bar{\mathbf{x}}_1)) - \nabla_2 f(\bar{\mathbf{x}}_1, h(\bar{\mathbf{x}}_1))[\nabla_2 g(\bar{\mathbf{x}}_1, h(\bar{\mathbf{x}}_1))]^{-1} \nabla_1 g(\bar{\mathbf{x}}_1, h(\bar{\mathbf{x}}_1)) \\ &= 0\end{aligned}$$

Setting $\lambda^*_{1\times M} = \nabla_2 f(\bar{\mathbf{x}}_1, h(\bar{\mathbf{x}}_1))[\nabla_2 g(\bar{\mathbf{x}}_1, h(\bar{\mathbf{x}}_1))]^{-1}$. This implies that,

$$\nabla_1 f(\bar{\mathbf{x}}_1, h(\bar{\mathbf{x}}_1)) - \lambda^* \nabla_1 g(\bar{\mathbf{x}}_1, h(\bar{\mathbf{x}}_1)) = 0$$

$$\nabla_2 f(\bar{\mathbf{x}}_1, h(\bar{\mathbf{x}}_1)) - \lambda^* \nabla_2 g(\bar{\mathbf{x}}_1, h(\bar{\mathbf{x}}_1)) = 0$$

□

Langrange Method

Given variables $x = (x_1, \ldots, x_N)$ and $\lambda = (\lambda_1, \ldots, \lambda_N)$ we can define the **Lagrangian function**

$$L(x, \lambda) = f(x) - \sum_m \lambda_m g_m(x)$$

Then the conditions given by (1) are the unconstrained first-order conditions of maximization of the above function with respect to the variables $x = (x_1, \ldots, x_N)$. Similarly $g(x) = 0$ are first-order conditions of $L(.,.)$ with respect to $\lambda = (\lambda_1, \ldots, \lambda_N)$. Such that

$$\frac{dL(\bar{x}, \lambda)}{dx_n} = 0 \quad \text{for} \quad n = 1, \ldots, N \quad \text{and}$$
$$\frac{dL(\bar{x}, \lambda)}{d\lambda_m} = 0 \quad \text{for} \quad m = 1, \ldots, M$$

Constrained Maximization: Second Order Conditions

Following is the Bordered Hessian Matrix in context of a constrained optimization problem:

$$B = \begin{bmatrix} 0 & \dots & 0 & \frac{\partial g_1}{\partial x_1} & \dots & \frac{\partial g_1}{\partial x_N} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & \dots & 0 & \frac{\partial g_M}{\partial x_1} & \dots & \frac{\partial g_M}{\partial x_N} \\ \frac{\partial g_1}{\partial x_1} & \dots & \frac{\partial g_M}{\partial x_1} & \frac{\partial^2 L}{\partial^2 x_1} & \dots & \frac{\partial^2 L}{\partial x_N \partial x_1} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \frac{\partial g_1}{\partial x_N} & \dots & \frac{\partial g_M}{\partial x_N} & \frac{\partial^2 L}{\partial x_1 \partial x_N} & \dots & \frac{\partial^2 L}{\partial^2 x_N} \end{bmatrix}$$

Given that conditions of Lagrange Theorem are satisfied. If the last $N - M$ leading principal minors of B evaluated at $(\bar{x}_1, ..., \bar{x}_N, \lambda_1, ..., \lambda_M)$

1. Alternate in sign where the last minor $B_{N+M} = B$ has the sign $(-1)^N$, then $\bar{x}$ is a local max
2. Have the same sign as $(-1)^M$, then $\bar{x}$ is a local min

Constrained Maximization: Inequality Constraints

The problem now is

$$\underset{x\in\mathbb{R}^N}{Max} \quad f(x) \quad s.t.$$

$$g_1(x) = b_1$$

$$\vdots$$

$$g_M(x) = b_m$$

$$h_1(x) \leq c_1$$

$$\vdots$$

$$h_K(x) \leq c_k$$

where every function is defined on $\mathbb{R}^N$ and $N \geq M + K$.

Kuhn-Tucker Theorem

Theorem

Suppose that $\bar{x} \in C$ is a local maximizer of the problem with inequality constraints. Also assume that the constraint qualification is satisfied. Then there are multipliers $\lambda_m \in \mathbb{R}$, one for each equality constraint, and $\lambda_k \in \mathbb{R}_+$, one for each inequality constraint, such that

1. For every $n = 1, \ldots, N$

$$\frac{df(\bar{x})}{dx_n} = \sum_{m=1}^{M} \lambda_m \frac{dg_m(\bar{x})}{dx_n} + \sum_{k=1}^{K} \lambda_k \frac{dh_k(\bar{x})}{dx_n}$$

2. For every $k = 1, \ldots, K$,

$$\lambda_k (h_k(\bar{x}) - c_k) = 0$$

Envelope Theorem

Write the maximization problem as

$$\underset{x \in \mathbb{R}^N}{Max} \quad f(x; q) \quad s.t.$$

$$g_1(x; q) = b_1$$

$$\vdots$$

$$g_M(x; q) = b_m$$

Theorem

Consider the value function $v(q)$ for the above problem. Assume that it is differentiable at $\bar{q} \in \mathbb{R}^S$ and that $(\lambda_1, \ldots, \lambda_M)$ are the values of the Lagrange multipliers associated with the maximizer solution $x(\bar{q})$ at $\bar{q}$. Then

$$\frac{dv(\bar{q})}{dq_s} = \frac{df(x(\bar{q}); \bar{q})}{dq_s} - \sum_{m=1}^{M} \lambda_m \frac{dg_m(x(\bar{q}); \bar{q})}{dq_s} \quad \text{for} \quad s = 1, \ldots, S$$

Appendix

Implicit Function Theorem

Theorem

Let $g : X_1 \times X_2 \to \mathbb{R}^M$ be C^1 where $X_1 \times X_2 \subset \mathbb{R}^N \times \mathbb{R}^M$. If

$$g(\bar{\mathbf{x}}_1, \bar{\mathbf{x}}_2) = 0 \qquad \text{and} \qquad \nabla_2 g(\bar{\mathbf{x}}_1, \bar{\mathbf{x}}_2) \neq 0$$

then there exists r_1 and r_2 such that for every $\mathbf{x}_1 \in \mathbb{R}^N$ with $|\mathbf{x}_1 - \bar{\mathbf{x}}_1| < r_1$, there exists unique $\mathbf{x}_2 \in \mathbb{R}^M$ with $|\mathbf{x}_2 - \bar{\mathbf{x}}_2| < r_2$ such that

$$g(\mathbf{x}_1, \mathbf{x}_2) = 0$$

In addition, define $h : B_{r_1}(\bar{\mathbf{x}}_1) \to B_{r_2}(\bar{\mathbf{x}}_2)$ then

$$g(\mathbf{x}_1, h(\mathbf{x}_2)) = 0$$

and h is C^1 on $B_{r_1}(\bar{\mathbf{x}}_1)$.