Optimization Theory

Lectures 4-6

Unconstrained Maximization

Problem: Maximize a function $f:\mathbb{R}^n \to \mathbb{R}$ within a set **A** $\subseteq \mathbb{R}^n$.

Typically, **A** is $\mathbb{R}^n$, or the non-negative orthant $\{x \in \mathbb{R}^n | x \geq 0\}$

Existence of a maximum:

Theorem. If **A** is compact (i.e., closed and bounded) and f is continuous, then a maximum exists.

Proof: For each $x \in$ **A**, the set $\{z \in \mathbf{A} | f(z) \geq f(x)\}$ is a closed subset of a compact set, hence compact, and the intersection of any finite number of these sets is non-empty. Therefore, by the finite intersection property, they have a non-empty intersection, which is a maximand. □

Uniqueness of a maximum:

Def: A function f is *strictly concave* on **A** (i.e., $x, y \in$ **A**, $x \neq y$, $0 < \theta < 1$ implies $f(\theta x+(l-\theta)y) > \theta f(x) + (l-\theta)f(y)$).

Theorem. If **A** is convex and f is strictly concave and a maximum exists, then it is unique.

Proof: If $x \neq y$ are both maxima, then $(x+y)/2 \in$ **A** and by strict concavity, $f((x+y)/2)$ gives a higher value, a contradiction. □

A vector $y \in \mathbb{R}^n$ points into **A** (from $x^0 \in$ **A**) if $x^0 + \theta y \in$ **A** for all sufficiently small positive scalars θ.

Assume that f is twice continuously differentiable, and let f_x denote its vector of first derivatives and f_{xx} denote its array of second derivatives. Assume that x^o achieves a maximum of f on **A**. Then, a Taylor's expansion gives

$$f(x^o) \geq f(x^o+\theta y) = f(x^o) + \theta f_x(x^o)\cdot y + (\theta^2/2)y' f_{xx}(x^o)y + R(\theta^2)$$

for all y that point into **A** and small scalars $\theta > 0$, where $R(\varepsilon)$ is a remainder satisfying $\lim_{\varepsilon\to 0} R(\varepsilon)/\varepsilon = 0$.

First-Order Condition (FOC): $f_x(x^o)\cdot y \le 0$ for all y that point into **A** (implies $f_x(x^o)\cdot y = 0$ when both y and -y point into A, and $f_x(x^o) = 0$ when x^o is interior to **A** so that all y point into **A**). When A is the non-negative orthant, the FOC is $\partial f(x^o)/\partial x_i \le 0$, and $\partial f(x^o)/\partial x_i = 0$ if $x_i > 0$ for $i = 1,\ldots,n$.

Proof: In Taylor's expansion, take θ sufficiently small so quadratic term is negligible. □

Second-Order Condition (SOC): $y' f_{xx}(x^o)y \leq 0$ for all y pointing into A with $f_x(x^o)\cdot y = 0$ (implies $y' f_{xx}(x^o)y \leq 0$ for all y when x^o is interior to **A**).

The FOC and SOC are necessary at a maximum. If FOC holds, and a strict form of the SOC holds,

$y' f_{xx}(x^o)y < 0$ for all $y \neq 0$ pointing into A with $f_x(x^o)\cdot y = 0$,

then x^o is a unique maximum within some neighborhood of x^o.

Proof: In Taylor's expansion, take θ sufficiently small so remainder term is negligible. □

Inequality-Constrained Maximization

Suppose g:**A**×**D** → $\mathbb{R}$ and h:**A**×**D** → $\mathbb{R}^m$ are continuous functions on a convex set **A**, and define **B**(**y**) = {**x**∈**A**| h(**x**,**y**) ≥ **0**} for each **y** ∈ **D**. Typically, **A** is the non-negative orthant of $\mathbb{R}^n$. Maximization in **x** of g(**x**,**y**) subject to the constraint h(**x**,**y**) ≥ 0 is called a *nonlinear mathematical programming problem*.

Define $r(\mathbf{y}) = \max_{\mathbf{x} \in \mathbf{B}(\mathbf{y})} g(\mathbf{x},\mathbf{y})$ and $f(\mathbf{y}) = \operatorname{argmax}_{\mathbf{x} \in \mathbf{B}(\mathbf{y})} g(\mathbf{x},\mathbf{y})$. If $\mathbf{B}(\mathbf{y})$ is bounded, then it is compact, guaranteeing that $r(\mathbf{y})$ and $f(\mathbf{y})$ exist.

Define a *Lagrangian* $L(\mathbf{x},\mathbf{p},\mathbf{y}) = g(\mathbf{x},\mathbf{y}) + \mathbf{p}\cdot h(\mathbf{x},\mathbf{y})$. A vector $(\mathbf{x}^0,\mathbf{p}^0,\mathbf{y})$ with $\mathbf{x}^0 \in \mathbf{A}$ and $\mathbf{p}^0 \geq \mathbf{0}$ is a (global) *Lagrangian Critical Point* (LCP) at $\mathbf{y} \in \mathbf{D}$ if

$$L(\mathbf{x},\mathbf{p}^0,\mathbf{y}) \leq L(\mathbf{x}^0,\mathbf{p}^0,\mathbf{y}) \leq L(\mathbf{x}^0,\mathbf{p},\mathbf{y})$$

for all $\mathbf{x} \in \mathbf{A}$ and $\mathbf{p} \geq \mathbf{0}$.

Note that a LCP is a *saddle-point* of the Lagrangian, which is <u>maximized</u> in $\mathbf{x}$ at $\mathbf{x}^0$ given $\mathbf{p}^0$, and <u>minimized</u> in $\mathbf{p}$ at $\mathbf{p}^0$ given $\mathbf{x}^0$. The variables in the vector $\mathbf{p}$ are called *Lagrangian multipliers* or *shadow prices*.

Example: Suppose $\mathbf{x}$ is a vector of policy variables available to a firm, $g(\mathbf{x})$ is the firm's profit, and excess inventory of inputs is $h(\mathbf{x},\mathbf{y}) = \mathbf{y} - q(\mathbf{x})$, where $q(\mathbf{x})$ specifies the vector of input requirements for $\mathbf{x}$. The firm must operate under the constraint that excess inventory is non-negative.

The Lagrangian $L(\mathbf{x},\mathbf{p},\mathbf{y}) = g(\mathbf{x}) + \mathbf{p}\cdot(\mathbf{y}-q(\mathbf{x}))$ can then be interpreted as the overall profit from operating the firm and selling off excess inventory at prices $\mathbf{p}$. In this interpretation, a LCP determines the firm's implicit reservation prices for the inputs, the opportunity cost of selling inventory rather than using it in production.

The problem of minimizing in **p** a function t(**p**,**y**) subject to v(**p**,**y**) $\geq$ **0** is the same as that of maximizing -t(**p**,**y**) subject to this constraint, with the associated Lagrangian -t(**p**,**y**) + **x**·v(**p**,**y**) with shadow prices **x**. Defining L(**x**,**p**,**y**) = t(**p**,**y**) - **x**·v(**p**,**y**), a LCP for this constrained minimization problem is then ($\mathbf{x}^0$,$\mathbf{p}^0$,**y**) such that L(**x**,$\mathbf{p}^0$,**y**) $\leq$ L($\mathbf{x}^0$,$\mathbf{p}^0$,**y**) $\leq$ L($\mathbf{x}^0$,**p**,**y**) for all **x** $\geq$ **0** and **p** $\geq$ **0**. Note that this is the same string of inequalities that defined a LCP for a constrained maximization problem. The definition of L(**x**,**p**,**y**) is in general different in the two cases, but we next consider a problem where they coincide.

2.16.2. An important specialization of the nonlinear programming problem is *linear programming*, where one seeks to maximize g(**x**,**y**) = **b**·**x** subject to the constraints h(**x**,**y**) = **y** - **Sx** ≥ **0** and **x** ≥ **0**, with **S** an m×n matrix. Associated with this problem, called the *primal* problem, is a second linear programming problem, called its *dual*, where one seeks to minimize **y**·**p** subject to **S**′**p** - **b** ≥ **0** and **p** ≥ **0**.

The Lagrangian for the primal problem is L($\mathbf{x}$,$\mathbf{p}$,$\mathbf{y}$) = g($\mathbf{x}$,$\mathbf{y}$) + $\mathbf{p}$·h($\mathbf{x}$,$\mathbf{y}$) = $\mathbf{b}\cdot\mathbf{x} + \mathbf{p}\cdot\mathbf{y} - \mathbf{p}'\mathbf{S}\mathbf{x}$. The Lagrangian for the dual problem is L($\mathbf{x}$,$\mathbf{p}$,$\mathbf{y}$) = $\mathbf{y}\cdot\mathbf{p} - \mathbf{x}\cdot(\mathbf{S}'\mathbf{p} - \mathbf{b})$ = $\mathbf{b}\cdot\mathbf{x} + \mathbf{p}\cdot\mathbf{y} - \mathbf{p}'\mathbf{S}\mathbf{x}$, which is exactly the same as the expression for the primal Lagrangian. Therefore, if the primal problem has a LCP, then the dual problem has exactly the same LCP.

2.16.3. **Theorem**. If ($\mathbf{x}^0$,$\mathbf{p}^0$,$\mathbf{y}$) is a LCP, then $\mathbf{x}^0$ is a maximand in $\mathbf{x}$ of g($\mathbf{x}$,$\mathbf{y}$) subject to h($\mathbf{x}$,$\mathbf{y}$) $\geq$ $\mathbf{0}$ for each $\mathbf{y} \in \mathbf{D}$.

Proof: The inequality L($\mathbf{x}^0$,$\mathbf{p}^0$,$\mathbf{y}$) $\leq$ L($\mathbf{x}^0$,$\mathbf{p}$,$\mathbf{y}$) gives ($\mathbf{p}^0$ - $\mathbf{p}$)·h($\mathbf{x}^0$,$\mathbf{y}$) $\leq$ 0 for all $\mathbf{p} \geq \mathbf{0}$. Then $\mathbf{p} = \mathbf{0}$ implies $\mathbf{p}^0$·h($\mathbf{x}^0$,$\mathbf{y}$) $\leq$ 0, while taking $\mathbf{p}$ to be $\mathbf{p}^0$ plus various unit vectors implies h($\mathbf{x}^0$,$\mathbf{y}$) $\geq$ $\mathbf{0}$, and hence $\mathbf{p}^0$·h($\mathbf{x}^0$,$\mathbf{y}$) = 0. These are called the <u>complementary slackness</u> conditions, and state that if a constraint is not binding, then its Lagrangian multiplier is zero. The inequality L($\mathbf{x}$,$\mathbf{p}^0$,$\mathbf{y}$) $\leq$ L($\mathbf{x}^0$,$\mathbf{p}^0$,$\mathbf{y}$) then gives g($\mathbf{x}$,$\mathbf{y}$) + $\mathbf{p}$·h($\mathbf{x}$,$\mathbf{y}$) $\leq$ g($\mathbf{x}^0$,$\mathbf{y}$). Then, $\mathbf{x}^0$ satisfies the constraints. Any other $\mathbf{x}$ that also satisfies the constraints has $\mathbf{p}$·h($\mathbf{x}$,$\mathbf{y}$) $\geq$ 0, so that g($\mathbf{x}$,$\mathbf{y}$) $\leq$ g($\mathbf{x}^0$,$\mathbf{y}$). Then $\mathbf{x}^0$ solves the constrained maximization problem. □

2.16.4. **Theorem**. Suppose $g:\mathbf{A}\times\mathbf{D} \to \mathbb{R}$ and $h:\mathbf{A}\times\mathbf{D} \to \mathbb{R}^m$ are continuous concave functions on a convex set **A**. Suppose $\mathbf{x}^0$ maximizes $g(\mathbf{x},\mathbf{y})$ subject to the constraint $h(\mathbf{x},\mathbf{y}) \geq 0$. Suppose the *constraint qualification* that for each vector **p** satisfying $\mathbf{p} \geq \mathbf{0}$ and $\mathbf{p} \neq \mathbf{0}$, there exists $\mathbf{x} \in \mathbf{A}$ such that $\mathbf{p}\cdot h(\mathbf{x},\mathbf{y}) > 0$. [A sufficient condition for the constraint qualification is that there exist $\mathbf{x} \in \mathbf{A}$ at which $h(\mathbf{x},\mathbf{y})$ is strictly positive.] Then, there exists $\mathbf{p}^0 \geq \mathbf{0}$ such that $(\mathbf{x}^0,\mathbf{p}^0,\mathbf{y})$ is a LCP.

Proof: Define the sets $\mathbf{C}_1 = \{(\lambda,\mathbf{z})\in\mathbb{R}\times\mathbb{R}^m \mid \lambda\le g(\mathbf{x},\mathbf{y})$ and $\mathbf{z} \le h(\mathbf{x},\mathbf{y})$ for some $\mathbf{x} \in \mathbf{A}\}$ and $\mathbf{C}_2 = \{(\lambda,\mathbf{z})\in\mathbb{R}\times\mathbb{R}^m \mid \lambda>g(\mathbf{x}^0,\mathbf{y})$ and $\mathbf{z} > \mathbf{0}\}$. Show as an exercise that $\mathbf{C}_1$ and $\mathbf{C}_2$ are convex and disjoint. Since $\mathbf{C}_2$ is open, they are strictly separated by a hyperplane with normal $(\mu,\mathbf{p})$; i.e., $\mu\lambda'+\mathbf{p}\mathbf{z}' > \mu\lambda''+\mathbf{p}\mathbf{z}''$ for all $(\lambda',\mathbf{z}') \in \mathbf{C}_2$ and $(\lambda'',\mathbf{z}'') \in \mathbf{C}_1$, This inequality and the definition of $\mathbf{C}_2$ imply that $(\mu,\mathbf{p}) \ge \mathbf{0}$. If $\mu = 0$, then $\mathbf{p} \ne \mathbf{0}$ and taking $\mathbf{z}' \to \mathbf{0}$ implies $0 \ge \mathbf{p}\cdot h(\mathbf{x},\mathbf{y})$ for all $\mathbf{x} \in \mathbf{A}$, violating the constraint qualification. Therefore, we must have $\mu > 0$. Define $\mathbf{p}^0 = \mathbf{p}/\mu$. Then, the separating inequality implies $g(\mathbf{x}^0,\mathbf{y}) \ge g(\mathbf{x},\mathbf{y}) + \mathbf{p}^0\cdot h(\mathbf{x},\mathbf{y})$ for all $\mathbf{x} \in \mathbf{A}$. Since $\mathbf{p}^0 \ge \mathbf{0}$ and $h(\mathbf{x}^0,\mathbf{y}) \ge \mathbf{0}$, taking $\mathbf{x} = \mathbf{x}^0$ implies $\mathbf{p}^0\cdot h(\mathbf{x}^0,\mathbf{y}) = 0$. Therefore , $L(\mathbf{x},\mathbf{p},\mathbf{y}) = g(\mathbf{x},\mathbf{y}) + \mathbf{p}\cdot h(\mathbf{x},\mathbf{y})$ satisfies $L(\mathbf{x},\mathbf{p}^0,\mathbf{y}) \le L(\mathbf{x}^0,\mathbf{p}^0,\mathbf{y})$. Also, $\mathbf{p} \ge \mathbf{0}$ implies $\mathbf{p}\cdot h(\mathbf{x}^0,\mathbf{y}) \ge 0$, so that $L(\mathbf{x}^0,\mathbf{p}^0,\mathbf{y}) \le L(\mathbf{x}^0,\mathbf{p},\mathbf{y})$. Therefore, $(\mathbf{x}^0,\mathbf{p}^0,\mathbf{y})$ is a LCP. □

2.16.5. **Theorem**. Suppose maximizing $g(\mathbf{x})$ subject to the constraint $q(\mathbf{x}) \le \mathbf{y}$ has LCP $(\mathbf{x}^0,\mathbf{p}^0,\mathbf{y})$ and $(\mathbf{x}^0+\Delta\mathbf{x},\mathbf{p}^0+\Delta\mathbf{p},\mathbf{y}+\Delta\mathbf{y})$. Then $(\mathbf{p}^0+\Delta\mathbf{p})\cdot\Delta\mathbf{y} \le g(\mathbf{x}^0+\Delta\mathbf{x}) - g(\mathbf{x}^0) \le \mathbf{p}^0\cdot\Delta\mathbf{y}$.

Proof: The inequalities $L(\mathbf{x}^0+\Delta\mathbf{x},\mathbf{p}^0,\mathbf{y}) \le L(\mathbf{x}^0,\mathbf{p}^0,\mathbf{y}) \le L(\mathbf{x}^0,\mathbf{p}^0+\Delta\mathbf{p},\mathbf{y})$ and $L(\mathbf{x}^0,\mathbf{p}^0+\Delta\mathbf{p},\mathbf{y}+\Delta\mathbf{y}) \le L(\mathbf{x}^0+\Delta\mathbf{x},\mathbf{p}^0+\Delta\mathbf{p},\mathbf{y}+\Delta\mathbf{y}) \le L((\mathbf{x}^0+\Delta\mathbf{x},\mathbf{p}^0,\mathbf{y}+\Delta\mathbf{y})$ imply $L(\mathbf{x}^0,\mathbf{p}^0+\Delta\mathbf{p},\mathbf{y}+\Delta\mathbf{y}) - L(\mathbf{x}^0,\mathbf{p}^0+\Delta\mathbf{p},\mathbf{y}) \le L(\mathbf{x}^0+\Delta\mathbf{x},\mathbf{p}^0+\Delta\mathbf{p},\mathbf{y}+\Delta\mathbf{y}) - L(\mathbf{x}^0,\mathbf{p}^0,\mathbf{y}) \le L((\mathbf{x}^0+\Delta\mathbf{x},\mathbf{p}^0,\mathbf{y}+\Delta\mathbf{y}) - L(\mathbf{x}^0+\Delta\mathbf{x},\mathbf{p}^0,\mathbf{y})$. Then, cancellation of terms gives the result. □

This result justifies the interpretation of a Lagrangian multiplier as the rate of increase in the constrained optimal objective function that results when a constraint is relaxed, and hence as the shadow or implicit price of the constrained quantity.

Classical Programming Problem

Consider the problem of maximizing $f:\mathbb{R}^n \rightarrow \mathbb{R}$ subject to equality constraints

$h^j(x) - c_j = 0$ for $j = 1, \dots, m$ (in vector notation, $h(\mathbf{x}) - \mathbf{c} = \mathbf{0}$)

where $m < n$.

Lagrangian: $L(x,p) = f(x) - p\cdot[h(x) - c]$

The vector (x^o,p^o) is said to be a local (interior) Lagrangian Critical Point if

(11) $L_x(x^o,p^o) = 0$, $L_p(x^o,p^o) = h(x^o) - c = 0$, and

$z'L_{xx}(x^o,p^o)z \leq 0$ if z satisfies $h_x(x^o)z = 0$,

where p^o is unconstrained in sign.

Theorem. If (x^o,p^o) is a local LCP and $z'L_{xx}(x^op^o)z < 0$ if $z \neq 0$ satisfies $h_x(x^o)z = 0$, then x^o is a unique local maximum of f subject to $h(x) = 0$.

Proof: Note that $0 = L_p(x^o, p^o) = h(x^o) - c$ implies that x^o is feasible, and that $L(x^o,p^o) = f(x^o)$. A Taylor's expansion of $L(x^o+\theta z,p^o)$ yields

$$L(x^o+\theta z,p^o) = L(x^o,p^o) + \theta L_x(x^o,p^o)\cdot z + (\theta^2/2)z'L_{xx}(x^o,p^o)z + R(\theta^2),$$

the $R(\theta^3)$ term is a residual.

Theorem. If (x^o, p^o) is a local LCP and $z'L_{xx}(x^o p^o)z < 0$ for all $z \neq 0$ satisfying $h_x(x^o)z = 0$, then x^o is a unique local maximum of f subject to $h(x) = 0$.

Proof: Note that $0 = L_p(x^o, p^o) = h(x^o) - c$ implies that x^o is feasible, and that $L(x^o, p^o) = f(x^o)$. Taylor's expansions yield

$$f(x^o+\theta z) = f(x^o) + \theta f_x(x^o)\cdot z + (\theta^2/2)z'f_{xx}(x^o)z + R(\theta^2),$$

$$h(x^o+\theta z) = c + \theta h_x(x^o)\cdot z + (\theta^2/2)z'h_{xx}(x^o)z + R(\theta^2),$$

the $R(\theta^3)$ terms are residuals.

Using $L_x(x^o, p^o) = 0$,

$$L(x^o+\theta z, p^o) = L(x^o,p^o) + \theta L_x(x^o,p^o)\cdot z + (\theta^2/2)z' L_{xx}(x^o,p^o)z + R(\theta^2)$$

$$= f(x^o) + (\theta^2/2)z' L_{xx}(x^o, p^o, q^o)z + R(\theta^2)$$

A point $x^o + \theta z$ satisfying the constraints, with θ small, must satisfy $h_x(x^o)z = 0$. Then, the negative semidefiniteness of L_{xx} subject to these constraints implies

$$f(x^o + \theta z) \leq f(x^o) + (\theta^2/2)z' L_{xx}(x^o,p^o)z + R(\theta^2) \leq f(x^o) + R(\theta^2).$$

If L_{xx} is negative definite subject to these constraints then the SOC is sufficient for x^o to be a local maximum. □

Theorem. If x^o maximizes f(x) subject to h(x) - c = 0, and the constraint qualification holds that the m×n array $B = h_x(x^o)$ is of full rank m, then there exist Lagrange multipliers p^0 such that (x^0,p^0) is a local LCP.

Proof: The hypothesis of the theorem implies that

$$\begin{aligned} f(x^o) &\geq f(x^o+\theta z) \\ &= f(x^o) + \theta f_x(x^o)\cdot z + (\theta^2/2)z' f_{xx}(x^o)z + R(\theta^2) \end{aligned}$$

for all z such that

$$\begin{aligned} c &= h(x^o+\theta z) \\ &= c + \theta h_x(x^o)\cdot z + (\theta^2/2)z' h_{xx}(x^o)z + R(\theta^2). \end{aligned}$$

Taking θ small, these conditions imply

$x^o)\cdot z = 0$ and $z'f_{xx}(x^o)z$ for any z satisfying $h_x(x^o)\cdot z = 0$

ecall that $B = h_x(x^o)$ is m×n of rank m, and define the dempotent) n×n matrix $M = I - B'(BB')^{-1}B$. Since BM 0, each column of M is a vector z meeting the ondition $h_x(x^o)\cdot z = 0$, implying that $f_x(x^o)\cdot z = 0$, or

$$0 = Mf_x(x^o) = f_x(x^o) - h_x(x^o)'p^0,$$

ıere

$$p^0 = (BB')^{-1}Bf_x(x^o).$$

Define $L(x,p) = f(x) - p \cdot [h(x) - c]$. Then

(21) $$L_x(x^o,p^o) = f_x(x^o) - h_x(x^o)' p^o$$
$$= [I - B'(BB')^{-1}B]f_x(x^o) = 0.$$

The construction guarantees that $L_p(x^o,p^o) = 0$. Finally, Taylor's expansion of the Lagrangian establishes that $z'L_{xx}(x^o,p^o)z \leq 0$ for all z satisfying $h_x(x^o) \cdot z = 0$. Therefore, the constrained maximum corresponds to a local LCP. □