AAE 635 Lecture 3: Unconstrained Optimization

3.1 REVIEW

[Results from the one-input model, find input demand curve, slope of the input demand curve…]

3.2 MOTIVATION: N-INPUT MODEL

Consider the case of a firm producing one output $y \in R$ using n inputs $\mathbf{x} = (x_1, x_2, \ldots, x_n) \in R^n$. Let $\mathbf{F} \subset R^{n+1}$ be the feasible set representing the firm technology, where $(y, \mathbf{x}) \in \mathbf{F}$ means that the production decision $(y, \mathbf{x})$ is technologically feasible. As before, it will be useful to put some structure on the representation of the production technology. Define the production function as

$$g(\mathbf{x}) = \underset{y}{Max}\{y : (y, \mathbf{x}) \in \mathbf{F}\}.$$

Then, express technological feasibility as

$$\mathbf{F} = \{(y, \mathbf{x}) : y \le g(\mathbf{x}), \mathbf{x} \ge 0, y \ge 0\}.$$

For a competitive firm in a market economy, let $p \in R$ be the output price, and $\mathbf{w} = (w_1, w_2, \ldots, w_n) \in R^n$ denote the n input prices for $\mathbf{x}$. Then, the firm profit is $\pi = p\, y - \mathbf{w} \cdot \mathbf{x}'$, where py is firm revenue, and $\mathbf{w} \cdot \mathbf{x}' = \Sigma_i\, w_i \cdot x_i$ is firm cost of production. Then, profit maximization is denoted by

$$\underset{y,\mathbf{x}}{Max}\{p \cdot y - \mathbf{w} \cdot \mathbf{x}' : (y, \mathbf{x}) \in \mathbf{F}\}$$

or

$$\underset{y,\mathbf{x}}{Max}\{p \cdot y - \mathbf{w} \cdot \mathbf{x}' : y \le g(\mathbf{x}), \mathbf{x} \ge 0, y \ge 0\}$$

or, if p > 0,

$$\underset{\mathbf{x}}{Max}\{p \cdot g(\mathbf{x}) - \mathbf{w} \cdot \mathbf{x}' : \mathbf{x} \ge 0\}.$$

This is an unconstrained maximization problem representing economic rationality for a competitive firm. To analyze this problem in general, we can rely on some mathematical tools. Such tools are reviewed here.

3.2.1 SOME MATHEMATICAL TOOLS

Consider a function $f(\mathbf{x})$, where $\mathbf{x} = (x_1, x_2, \ldots, x_n) \in R^n$. Below, it will be convenient to treat $\mathbf{x}$ as a $(n\times1)$ vector $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$, to avoid the notation complication.

The function $f(\mathbf{x})$ is *continuous* at $\mathbf{x}$ if $[\lim_{\mathbf{y}\to\mathbf{x}} f(\mathbf{y})] = f(\mathbf{x})$. Here " $\lim_{\mathbf{y}\to\mathbf{x}} f(\mathbf{y})$" means the limit reached by $f(\mathbf{y})$ for all sequences of points $\mathbf{y} \ne \mathbf{x}$ becoming arbitrarily close to $\mathbf{x}$.

If $f(\mathbf{x})$ is differentiable, then its *derivative* is $f'(\mathbf{x}) = f_{\mathbf{x}}(\mathbf{x}) = \frac{\partial f}{\partial \mathbf{x}} = [\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, ..., \frac{\partial f}{\partial x_n}]$, where $\frac{\partial f}{\partial x_i} = \lim_{\varepsilon \to 0}[\frac{f(\mathbf{y}) - f(\mathbf{x})}{\varepsilon} : \mathbf{y} = (x_1, ..., x_{i-1}, x_i + \varepsilon, x_{i+1}, ..., x_n)]$ is the partial derivative of $f(\mathbf{x})$ with respect to x_i, $i = 1, \ldots, n$.

A differentiable function is always continuous (but a continuous function is not always differentiable).

A function $f(\mathbf{x})$ that is differentiable and where $f'(\mathbf{x})$ is continuous is said to *continuously differentiable*.

If $f(\mathbf{x})$ is twice differentiable, then its second derivative is $f''(\mathbf{x}) = f_{\mathbf{xx}}(\mathbf{x}) = \frac{\partial^2 f}{\partial \mathbf{x}^2} =$

$$\begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix}$$

, which is a $n \times n$ matrix, where $\frac{\partial^2 f}{\partial x_i^2}$ is the second partial derivative of f with respect to x_i (defined as the partial derivative of $\frac{\partial f}{\partial x_i}$ with respect to x_i), and $\frac{\partial^2 f}{\partial x_i \partial x_j}$ is the partial derivative of $\frac{\partial f}{\partial x_i}$ with respect to x_j.

A function $f(\mathbf{x})$ that is twice differentiable and where $f''(\mathbf{x})$ is continuous is said to be *twice continuously differentiable*.

<u>Young theorem</u>: If $f(\mathbf{x})$ is *twice continuously differentiable*, then $\frac{\partial^2 f}{\partial x_i \partial x_j} = \frac{\partial^2 f}{\partial x_j \partial x_i}$ for all $i \neq j$. In this case, the second cross-derivatives are invariant to the order of differentiation, and the ($n \times n$) matrix $f''(\mathbf{x})$ is *symmetric*.

The ($n \times n$) matrix $\mathbf{A}$ is *negative (positive) semi-definite* if $\mathbf{v}^T \cdot \mathbf{A} \cdot \mathbf{v} \leq (\geq)\ 0$ for all ($n \times 1$) vectors $\mathbf{v}$. (Here, $\mathbf{v}^T$ is a ($1 \times n$) vector defined as the transpose of the vector $\mathbf{v}$, and $\mathbf{v}^T \cdot \mathbf{A} \cdot \mathbf{v}$ is called a *quadratic form*).

The ($n \times n$) matrix $\mathbf{A}$ is *negative (positive) definite* if $\mathbf{v}^T \cdot \mathbf{A} \cdot \mathbf{v} < (>)\ 0$ for all ($n \times 1$) vectors $\mathbf{v} \neq 0$.

<u>Concavity:</u> The function $f(\mathbf{x})$ is *concave* if, for any two points $\mathbf{x}^1$ and $\mathbf{x}^2$, it satisfies

$$f(\theta \cdot \mathbf{x}^1 + (1-\theta) \cdot \mathbf{x}^2) \geq \theta \cdot f(\mathbf{x}^1) + (1-\theta) \cdot f(\mathbf{x}^2),$$

for any θ, $0 \leq \theta \leq 1$.

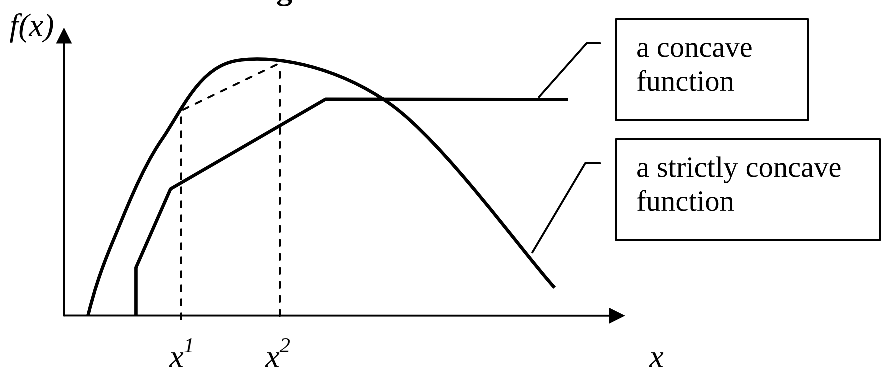
The function $f(\mathbf{x})$ is *strictly concave* if, for any two points $\mathbf{x}^1$ and $\mathbf{x}^2$, it satisfies

$$f(\theta \cdot \mathbf{x}^1 + (1-\theta) \cdot \mathbf{x}^2) > \theta \cdot f(\mathbf{x}^1) + (1-\theta) \cdot f(\mathbf{x}^2),$$

for any θ, $0 < \theta < 1$.

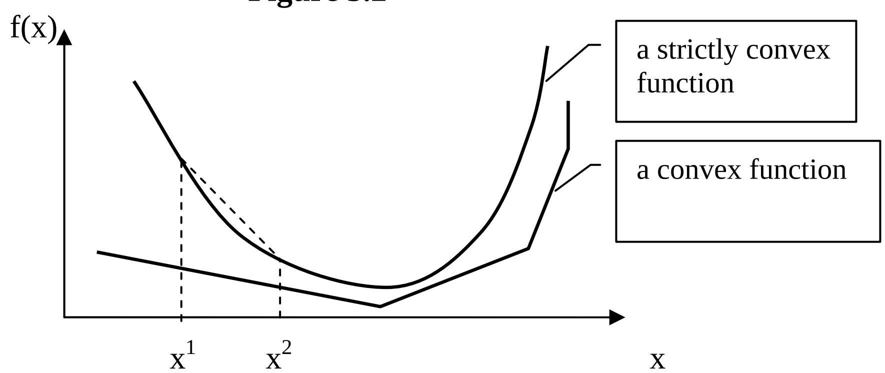
For a concave function, any line joining two points ($f(\mathbf{x}^1)$, $\mathbf{x}^1$) and ($f(\mathbf{x}^2)$, $\mathbf{x}^2$) cannot be above the function. And for a strictly concave function, any line joining two distinct points ($f(\mathbf{x}^1)$, $\mathbf{x}^1$) and ($f(\mathbf{x}^2)$, $\mathbf{x}^2$) must be below the function. See Figure 3.1. (Note: a concave or strictly concave function is not necessarily differentiable).

Figure 3.1



For a convex function, any line joining two points ($f(\mathbf{x}^1)$, $\mathbf{x}^1$) and ($f(\mathbf{x}^2)$, $\mathbf{x}^2$) cannot be below the function. And for a strictly convex function, any line joining two distinct points ($f(\mathbf{x}^1)$, $\mathbf{x}^1$) and ($f(\mathbf{x}^2)$, $\mathbf{x}^2$) must be above the function. See Figure 3.2. (Again, note that a convex or strictly convex function is not necessarily differentiable).

Figure 3.2



Some important results:

- A linear function, $f(x) = a + b\,x$, is both concave and convex.
- Assume that $f(\mathbf{x})$ is continuously differentiable.
  - The function $f(\mathbf{x})$ is concave *if and only if*
    $$f(\mathbf{x}^2) - f(\mathbf{x}^1) \le f'(\mathbf{x}^1)\cdot(\mathbf{x}^2 - \mathbf{x}^1)$$
    for any two points $\mathbf{x}^1$ and $\mathbf{x}^2$. (Recall that $f'(\mathbf{x}^1)$ is a $(1\times n)$ vector, $\mathbf{x}$ is a $(n\times 1)$ vector, and $f'(\mathbf{x}^1)\cdot(\mathbf{x}^2 - \mathbf{x}^1) = \Sigma_i\,[\frac{\partial f}{\partial x_i}\cdot(x_i^2 - x_i^1)]$).
  - The function $f(\mathbf{x})$ is strictly concave *if and only if*
    $$f(\mathbf{x}^2) - f(\mathbf{x}^1) < f'(\mathbf{x}^1)\cdot(\mathbf{x}^2 - \mathbf{x}^1)$$
    for any two distinct points $\mathbf{x}^1$ and $\mathbf{x}^2$.
- Assume that $f(\mathbf{x})$ is continuously differentiable.
  - The function $f(\mathbf{x})$ is concave *if and only if*

$[f'(\mathbf{x}^2) - f'(\mathbf{x}^1)]\,(\mathbf{x}^2 - \mathbf{x}^1) \leq 0$,
for any two points $\mathbf{x}^1$ and $\mathbf{x}^2$.

- The function $f(\mathbf{x})$ is strictly concave *if and only if*
  $[f'(\mathbf{x}^2) - f'(\mathbf{x}^1)]\,(\mathbf{x}^2 - \mathbf{x}^1) < 0$,
  for any two distinct points $\mathbf{x}^1$ and $\mathbf{x}^2$.

This is an important an intuitive result: it means that (strict) concavity is equivalent to situations where the marginal values $f'(\mathbf{x})$ are (strictly) declining in $\mathbf{x}$.

- Assume that $f(\mathbf{x})$ is twice continuously differentiable.
  - The function $f(\mathbf{x})$ is concave *if and only if* $f''(\mathbf{x})$ is a negative semi-definite matrix for all $\mathbf{x}$.
  - If the function $f''(\mathbf{x})$ is negative definite for all $\mathbf{x}$, then $f(\mathbf{x})$ is strictly concave.

Note: Similar results apply to convex functions (just replace $f(\mathbf{x})$ by $-f(\mathbf{x})$ in the above results…).

Taylor series

Define the *neighborhood* of $\mathbf{x}^*$ as the set of points $\{\mathbf{x}: \|\mathbf{x} - \mathbf{x}^*\| < \varepsilon$, for some $\varepsilon > 0\}$ where $\|\mathbf{x} - \mathbf{x}^*\| = (\Sigma_i\,(x_i - x_i^*)^2)^{1/2}$ is the Euclidian distance between $\mathbf{x}$ and $\mathbf{x}^*$. In the case where ε is "small", this defines a "close neighborhood".

Taylor series: Let $f(\mathbf{x})$ be a twice continuously differentiable function. Then, in a close neighborhood of $\mathbf{x}^*$, $f(\mathbf{x})$ can be approximated well by the first-order Taylor series expansion

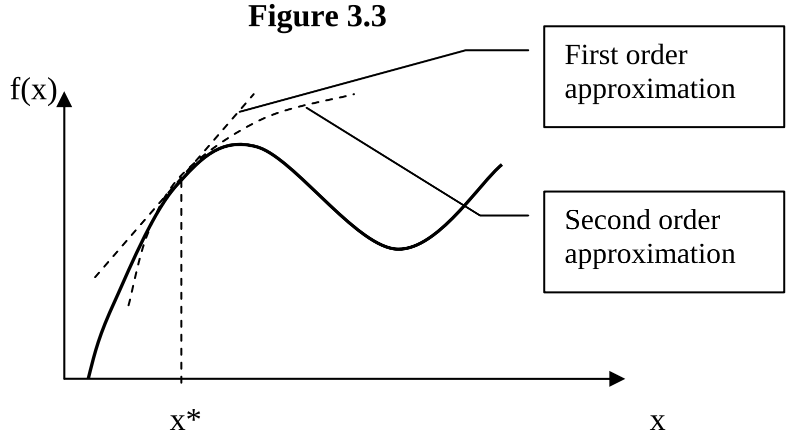
$$f(\mathbf{x}) \cong f(\mathbf{x}^*) + f'(\mathbf{x}^*)\,(\mathbf{x} - \mathbf{x}^*),$$

or by the second-order Taylor series expansion

$$f(\mathbf{x}) \cong f(\mathbf{x}^*) + f'(\mathbf{x}^*)\,(\mathbf{x} - \mathbf{x}^*) + .5\,(\mathbf{x} - \mathbf{x}^*)^T f''(\mathbf{x}^*)\,(\mathbf{x} - \mathbf{x}^*).$$

This is illustrated in Figure 3.3, showing that a differentiable function can be approximated well *locally* by a polynomial function (e.g., linear or quadratic).

Figure 3.3



Implicit function theorem

Consider the system of n equations involving $(n + m)$ variables $(\mathbf{x}, \mathbf{y})$

$$h_i(\mathbf{x}, \mathbf{y}) = 0,\ i = 1, \ldots, n,$$

where $\mathbf{x} = (x_1, \ldots, x_n) \in \mathbf{R}^n$ and $\mathbf{y} = (y_1, \ldots, y_m) \in \mathbf{R}^m$. Using matrix notation, this can be written as
$h(\mathbf{x}, \mathbf{y}) = 0$,

where $h(\mathbf{x}, \mathbf{y}) = \begin{bmatrix} h_1(\mathbf{x},\mathbf{y}) \\ h_2(\mathbf{x},\mathbf{y}) \\ \vdots \\ h_n(\mathbf{x},\mathbf{y}) \end{bmatrix}$ is a $(n\times1)$ vector of n functions.

We want to solve this system of n equations for the n unknown $\mathbf{x}$, taking the values of the m variables y as given. Denote this solution by $\mathbf{x^*(y)}$, which must satisfy $h(\mathbf{x^*(y)}, \mathbf{y}) = 0$.

Define the $(n\times n)$ *Jacobian matrix* $h_{\mathbf{x}}(\mathbf{x}, \mathbf{y}) = \begin{bmatrix} \frac{\partial h_1}{\partial x_1} & \frac{\partial h_1}{\partial x_2} & \cdots & \frac{\partial h_1}{\partial x_n} \\ \frac{\partial h_2}{\partial x_1} & \frac{\partial h_2}{\partial x_2} & \cdots & \frac{\partial h_2}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial h_n}{\partial x_1} & \frac{\partial h_n}{\partial x_2} & \cdots & \frac{\partial h_n}{\partial x_n} \end{bmatrix}$ as the matrix of partial derivatives of $h(\mathbf{x}, \cdot)$ with respect to $\mathbf{x}$. And define the $(n\times m)$ matrix $h_{\mathbf{y}}(\mathbf{x}, \mathbf{y}) =$ $\begin{bmatrix} \frac{\partial h_1}{\partial y_1} & \frac{\partial h_1}{\partial y_2} & \cdots & \frac{\partial h_1}{\partial y_m} \\ \frac{\partial h_2}{\partial y_1} & \frac{\partial h_2}{\partial y_2} & \cdots & \frac{\partial h_2}{\partial y_m} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial h_n}{\partial y_1} & \frac{\partial h_n}{\partial y_2} & \cdots & \frac{\partial h_n}{\partial y_m} \end{bmatrix}$ as the matrix of partial derivatives of $h(\cdot, \mathbf{y})$ with respect to $\mathbf{y}$.

<u>Theorem</u>: Assume that the functions $h(\mathbf{x}, \mathbf{y})$ are continuously differentiable in $(\mathbf{x}, \mathbf{y})$, and that the determinant of the Jacobian matrix is non-zero, $det(h_{\mathbf{x}}) \neq 0$, at $(\mathbf{x}^*(\mathbf{y}), \mathbf{y})$. Then,

- $\mathbf{x^*(y)}$ is a differentiable function
- and its derivatives are given by the $(n\times m)$ matrix

$$\frac{\partial \mathbf{x}^*}{\partial \mathbf{y}} = \mathbf{x_y}^* \equiv \begin{bmatrix} \frac{\partial x_1^*}{\partial y_1} & \frac{\partial x_1^*}{\partial y_2} & \cdots & \frac{\partial x_1^*}{\partial y_m} \\ \frac{\partial x_2^*}{\partial y_1} & \frac{\partial x_2^*}{\partial y_2} & \cdots & \frac{\partial x_2^*}{\partial y_m} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial x_n^*}{\partial y_1} & \frac{\partial x_n^*}{\partial y_2} & \cdots & \frac{\partial x_n^*}{\partial y_m} \end{bmatrix} = -[h_{\mathbf{x}}]^{-1}\, h_{\mathbf{y}}, \text{ evaluated at } (\mathbf{x^*(y)}, \mathbf{y}).$$

<u>Proof</u>: Taking a first order Taylor series expansion of $h(\mathbf{x}, \mathbf{y}) = 0$ evaluated at point $(\mathbf{x^*(y^*)}, \mathbf{y^*})$ gives

$$h_{\mathbf{x}}(\mathbf{x^*(y^*)}, \mathbf{y^*})\, [\mathbf{x} - \mathbf{x^*}] + h_{\mathbf{y}}(\mathbf{x^*(y^*)}, \mathbf{y^*})\, [\mathbf{y} - \mathbf{y^*}] \cong 0.$$

If $det(h_{\mathbf{x}}) \neq 0$, then $[h_{\mathbf{x}}]^{-1}$ exists, implying that

$$[\mathbf{x} - \mathbf{x^*}] \cong -[\, h_{\mathbf{x}}(\mathbf{x^*}, \mathbf{y^*})]^{-1} \cdot h_{\mathbf{y}}(\mathbf{x^*}, \mathbf{y^*}) \cdot [\mathbf{y} - \mathbf{y^*}].$$

This means that, in a close neighborhood of (**x*(y*), y***), the function **x*(y)** can be approximated well by a linear function of **y**. It follows that **x*(y)** is differentiable and that $\frac{\partial \mathbf{x}^*}{\partial \mathbf{y}} = \mathbf{x_y}^* = -[h_\mathbf{x}]^{-1} h_\mathbf{y}$.

Homogeneous functions

A function $f(\mathbf{x})$ is *homogeneous of degree r* if, for all **x**, and $t > 0$,

$$f(t\,\mathbf{x}) = t^r f(\mathbf{x}).$$

Homogeneous functions have two important properties:

- If $f(\mathbf{x})$ is homogeneous of degree r and continuously differentiable, then $f'(\mathbf{x})$ is homogeneous of degree (r-1).

- (***Euler* theorem**) If $f(\mathbf{x})$ is homogeneous of degree r and continuously differentiable, then
  $$f'(\mathbf{x})\,\mathbf{x} = r\,f(\mathbf{x})$$
  or, equivalently,
  $$\Sigma_i \left(\frac{\partial f}{\partial x_i}\right) x_i = r\,f(\mathbf{x}).$$

3.3 UNCONSTRAINED OPTIMIZATION

Consider the optimization problem

$$\underset{x}{Max} \ \{ f(\mathbf{x}) : \mathbf{x} \geq 0, \mathbf{x} \in \mathbf{R}^n \}.$$

A solution (also called a *global solution*) to the maximization problem is a point **x*** satisfying $f(\mathbf{x}^*) \geq f(\mathbf{x})$ for all $\mathbf{x} \geq 0$.

A *local solution* to $f(\mathbf{x})$ is a point x* satisfying $f(\mathbf{x}) \geq f(\mathbf{x}^*)$ for all **x** in some neighborhood of **x***.

An *interior point* $\mathbf{x} = (x_1, \ldots, x_n)^T$ is a point such that $x_i > 0$, $i = 1, \ldots, n$.

Consider the first and second derivatives of $f(\mathbf{x})$:

$f'(\mathbf{x}) = \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \ldots, \frac{\partial f}{\partial x_n}\right)$ = ($1 \times n$) vector of first derivatives

and

$$f''(\mathbf{x}) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix} = (n \times n) \text{ matrix of second derivatives.}$$

Assume that $f(\mathbf{x})$ is twice continuous differentiable, and that $\mathbf{x}^* > 0$ is an interior solution to maximization problem. The key results are presented next.

<u>Note</u>: Note that $\mathbf{x}^*$ maximizes $f(\mathbf{x})$ if and only if $\mathbf{x}^*$ also minimizes - $f(\mathbf{x})$. Thus all the results presented below can be adapted to minimization problems (after replacing $f(\mathbf{x})$ by - $f(\mathbf{x})$ in the analysis).

3.3.1 First order necessary condition (FONC)

$f'(\mathbf{x}^*) = 0$ is a *necessary condition* for $\mathbf{x}^*$ to be an interior maximum of $f(\mathbf{x})$.

<u>Proof</u>: Let $\mathbf{x}^*$ be a global maximum of $f(\mathbf{x})$. A first order Taylor series approximation of $f(\mathbf{x})$ evaluated at point $\mathbf{x}^*$ gives

$$f(\mathbf{x}) \cong f(\mathbf{x}^*) + f'(\mathbf{x}^*)\,(\mathbf{x} - \mathbf{x}^*).$$

By definition of a maximum, $f(\mathbf{x}^*) \geq f(\mathbf{x})$ for all $\mathbf{x}$. In a close neighborhood of $\mathbf{x}^*$, it follows that

$$f(\mathbf{x}^*) \geq f(\mathbf{x}^*) + f'(\mathbf{x}^*)\,(\mathbf{x} - \mathbf{x}^*),$$

or

$$0 \geq f'(\mathbf{x}^*)\,(\mathbf{x} - \mathbf{x}^*) \text{ for all feasible } \mathbf{x}.$$

When $\mathbf{x}^*$ is an interior point, this is possible only if $f'(\mathbf{x}^*) = 0$ (otherwise, we can always choose $x_i - x_i^* > 0$ when $\frac{\partial f}{\partial x_i} < 0$ or $x_i - x_i^* < 0$ when $\frac{\partial f}{\partial x_i} > 0$, which would contradict the above inequality).

<u>Note</u>:

$$f'(\mathbf{x}^*) = 0,$$

or,

$$\left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \ldots, \frac{\partial f}{\partial x_n}\right) = 0$$

or

$$\frac{\partial f(\mathbf{x}^*)}{\partial x_i} = 0,\ i = 1, \ldots, n,$$

are called the *first order necessary conditions* (FONC). They constitute a system of n equations in n unknown, $\mathbf{x}^* = (x_1^*, \ldots, x_n^*)$.

3.3.2 Second order necessary condition (SONC)

$f'(\mathbf{x}^*) = 0$ and $f''(\mathbf{x}^*)$ = a negative semi-definite matrix are *necessary conditions* for $\mathbf{x}^*$ to be an interior maximum of $f(\mathbf{x})$.

<u>Proof</u>: Let $\mathbf{x}^*$ be a global maximum of $f(\mathbf{x})$. A second order Taylor series approximation of $f(\mathbf{x})$ evaluated at point $\mathbf{x}^*$ gives

$$f(\mathbf{x}) \cong f(\mathbf{x}^*) + f'(\mathbf{x}^*)\,(\mathbf{x} - \mathbf{x}^*) + .5\,(\mathbf{x} - \mathbf{x}^*)^T f''(\mathbf{x}^*)\,(\mathbf{x} - \mathbf{x}^*).$$

By definition of a maximum, $f(\mathbf{x}^*) \geq f(\mathbf{x})$ for all $\mathbf{x}$. In a close neighborhood of $\mathbf{x}^*$, it follows that

$$f(\mathbf{x}^*) \geq f(\mathbf{x}^*) + f'(\mathbf{x}^*)\,(\mathbf{x} - \mathbf{x}^*) + .5\,(\mathbf{x} - \mathbf{x}^*)^T f''(\mathbf{x}^*)\,(\mathbf{x} - \mathbf{x}^*),$$

or

$$0 \geq f'(\mathbf{x}^*)\,(\mathbf{x} - \mathbf{x}^*) + .5\,(\mathbf{x} - \mathbf{x}^*)^T f''(\mathbf{x}^*)\,(\mathbf{x} - \mathbf{x}^*), \text{ for all feasible } \mathbf{x}.$$

When $f'(\mathbf{x}^*) = 0$, this implies that $\mathbf{v}^T f''(\mathbf{x}^*)\, \mathbf{v} \leq 0$, with $\mathbf{v} = k\,(\mathbf{x} - \mathbf{x}^*)$, k being a positive constant. Since $\mathbf{v}$ can take any value when $\mathbf{x}^*$ is an interior point, it follows that $f''(\mathbf{x}^*)$ is a negative semi-definite matrix.

$f''(\mathbf{x}^*)$ being a negative semi-definite matrix is called the *second order necessary conditions* (SONC) for maximizing $f(\mathbf{x})$.

Note: Similarly, for a *minimization* problem, $f'(\mathbf{x}^*) = 0$ and $f''(\mathbf{x}^*)$ being a positive semi-definite matrix are necessary conditions for $\mathbf{x}^*$ to be an interior solution to the minimization of $f(\mathbf{x})$. Then, for minimization, the *second order necessary conditions* (SONCmin) are: $f''(\mathbf{x}^*)$ is a positive semi-definite matrix.

3.3.3 **Second order sufficient condition** (SOSC)

$f'(\mathbf{x}^*) = 0$ and $f''(\mathbf{x}^*)$ = a negative definite matrix are *sufficient conditions* for $\mathbf{x}^*$ to be a local interior solution to the maximization problem.

Proof: A second order Taylor series approximation of f(x) evaluated at point x* gives

$$f(\mathbf{x}) \cong f(\mathbf{x}^*) + f'(\mathbf{x}^*)\,(\mathbf{x} - \mathbf{x}^*) + .5\,(\mathbf{x} - \mathbf{x}^*)^T f''(\mathbf{x}^*)\,(\mathbf{x} - \mathbf{x}^*).$$

When $f''(\mathbf{x}^*)$ is negative definite, $(\mathbf{x} - \mathbf{x}^*)^T f''(\mathbf{x}^*)\,(\mathbf{x} - \mathbf{x}^*) < 0$. Thus, $f'(\mathbf{x}^*) = 0$ and $f''(\mathbf{x}^*)$ being negative definite, implies that $f(\mathbf{x}) - f(\mathbf{x}^*) < 0$ for all $\mathbf{x}$ in a close neighborhood of $\mathbf{x}^*$. Thus, $\mathbf{x}^*$ is a local maximum.

$f''(\mathbf{x}^*)$ being a negative definite matrix is called the *second order sufficient conditions* (SOSC).

Note: Similarly, for a *minimization* problem, $f'(\mathbf{x}^*) = 0$ and $f''(\mathbf{x}^*)$ being a positive definite matrix are sufficient conditions for x* to be a local interior solution to the minimization of $f(\mathbf{x})$. Then, for minimization, the *second order sufficient conditions* (SOSCmin) are: $f''(\mathbf{x}^*)$ = a positive definite matrix.

3.3.4 **Optimization under concavity**

If $f(\mathbf{x})$ is *concave*, then $f'(\mathbf{x}^*) = 0$ is a necessary and sufficient condition for $\mathbf{x}^*$ to be a *global interior solution* to the maximization of $f(\mathbf{x})$.

Proof: The proof for necessity was obtained earlier. To prove sufficiency, note that a continuously differentiable and concave function $f(\mathbf{x})$ satisfies

$$f(\mathbf{x}) - f(\mathbf{x}^*) \leq f'(\mathbf{x}^*)\,(\mathbf{x} - \mathbf{x}^*),$$

for any feasible $\mathbf{x}$ and $\mathbf{x}^*$. Then, $f'(\mathbf{x}^*) = 0$ implies that $f(\mathbf{x}) - f(\mathbf{x}^*) \leq 0$ for all feasible $\mathbf{x}$, i.e., that $\mathbf{x}^*$ is a global maximum of $f(\mathbf{x})$.

This is an intuitive result for two reasons.

- First, concavity of $f(\mathbf{x})$ is equivalent to diminishing marginal values for $f(\mathbf{x})$. This implies that, if we find a local interior maximum $\mathbf{x}^*$ satisfying $f'(\mathbf{x}^*) = 0$, $\mathbf{x}$ values lower than $\mathbf{x}^*$ must be associated with $f'(\mathbf{x}) \geq 0$, implying $f(\mathbf{x}) \leq f(\mathbf{x}^*)$. And $\mathbf{x}$ values higher than $\mathbf{x}^*$

must be associated with $f'(\mathbf{x}) \le 0$, implying $f(\mathbf{x}) \le f(\mathbf{x}^*)$. Thus, under diminishing marginal values, **x*** must be a global maximum.

- Second, concavity of $f(\mathbf{x})$ is equivalent to $f''(\mathbf{x})$ = negative semi-definite for all **x**. This implies the SONC: $f''(\mathbf{x}^*)$ = negative semi-definite. (However, SONC does not imply concavity since SONC holds only at **x*** and not for all **x**. On this issue, note that SOSC ($f''(\mathbf{x}^*)$ = negative definite) implies strict concavity *locally*; however it does not imply strict concavity globally because it holds only at **x*** and not for all **x**).

Note: Similarly, for a *minimization* problem, if $f(\mathbf{x})$ is *convex*, then $f'(\mathbf{x}^*) = 0$ is a necessary and sufficient condition for **x*** to be a *global interior solution* to the minimization of $f(\mathbf{x})$.

Similar intuitive interpretations apply.

- First, convexity of $f(\mathbf{x})$ is equivalent to increasing marginal values for $f(\mathbf{x})$. This implies that, if we find a local interior minimum **x*** satisfying $f'(\mathbf{x}^*) = 0$, x values lower than **x*** must be associated with $f'(\mathbf{x}) \le 0$, implying $f(\mathbf{x}) \ge f(\mathbf{x}^*)$. And **x** values higher than **x*** must be associated with $f'(\mathbf{x}) \ge 0$, implying $f(\mathbf{x}) \ge f(\mathbf{x}^*)$. Thus, under diminishing marginal values, **x*** must be a global minimum.
- Second, convexity of $f(\mathbf{x})$ is equivalent to $f''(\mathbf{x})$ = positive semi-definite for all **x**. This implies the SONCmin: $f''(\mathbf{x}^*)$ = positive semi-definite. (However, SONCmin does not imply convexity since SONCmin holds only at **x*** and not for all **x**. On this issue, note that SOSCmin ($f''(\mathbf{x}^*)$ = positive definite) implies strict convexity *locally*; however it does not imply strict convexity globally because it holds only at **x*** and not for all **x**).

3.3.5 Summary

These results suggest the following strategy to identify an interior solution to the *maximization* of $f(\mathbf{x})$:

- Find **x*** satisfying FONC: $f'(\mathbf{x}^*) = 0$. (This involves solving a system of n equations in n unknown).
- If SONC are not satisfied at **x***, then **x*** cannot be a solution to the maximization problem.
- If SOSC are satisfied at **x***, then **x*** is a *local solution* to the maximization problem. (Note that this does not imply that it is global solution…)
- If $f(\mathbf{x})$ is concave, then **x*** is a *global solution* to the maximization problem.

Checking the SOSC: $f''(\mathbf{x}^*)$ = negative definite.

- either find (using a computer) the characteristic values (or eigenvalues) of $f''(\mathbf{x}^*)$ and check that they are all negative; or
- check that the following determinants. Consider the $(n\times n)$ matrix $\mathbf{H} = f''(\mathbf{x}^*)$ (called the *Hessian* of f). Define the $(k\times k)$ matrix $\mathbf{C}_k$ from **H** by deleting from **H** the first $(n-k)$ rows and the first $(n-k)$ columns, $k = 1, \ldots, n$. Then,

  check that $(-1)^k \det(C_k) > 0$, $k = 1, \ldots, n$,

  i.e., $det(\mathbf{C}_k)$ *alternate in sign*: < 0 for $k = 1$, > 0 for $k = 2$, < 0 for $k = 3$, etc.
  In the one variable case ($n = 1$), this gives: $f''(\mathbf{x}^*) < 0$.

In the two variable case (n = 2), this gives: $f_{22} < 0$ for k = 1, and $det\begin{bmatrix} f_{11} & f_{12} \\ f_{12} & f_{22} \end{bmatrix} = f_{11}\, f_{22} - f_{12}^2 > 0$ for k = 2, where $f_{ij} = \frac{\partial^2 f}{\partial x_i \partial x_j}$ evaluated at **x***. (Note that $f_{22} < 0$ and $f_{11}\, f_{22} > f_{12}^2 \geq 0$ imply that $f_{11} < 0$).
And so on…

Similarly, you can use the following strategy to identify an interior solution to the *minimization* of f(**x**):

- Find **x*** satisfying FONC, f '(**x***) = 0. (This involves solving a system of n equations in n unknown).
- If SONCmin are not satisfied at **x***, then **x*** cannot be a solution to the minimization problem.
- If SOSCmin are satisfied at **x***, then **x*** is a *local solution* to the minimization problem. (Note that this does not imply that it is global solution…)
- If f(**x**) is convex, then **x*** is a *global solution* to the minimization problem.

<u>Checking the SOSCmin</u>: f "(**x***) = positive definite.

- either find (using a computer) the eigenvalues of f "(**x***) and check that they are all positive; or
- check that the following determinants. Using the $\mathbf{C}_k$ matrices defined above,

  check that $det(C_k) > 0$, $k = 1, \ldots, n$,

  i.e., $det(\mathbf{C}_k)$ are *positive* for all k.
  In the one variable case (n = 1), this gives: f "(**x***) > 0.
  In the two variable case (n = 2), this gives: $f_{22} > 0$ for k = 1, and $\det\begin{bmatrix} f_{11} & f_{12} \\ f_{12} & f_{22} \end{bmatrix} = f_{11}\, f_{22} - f_{12}^2 > 0$ for k = 2, where $f_{ij} = \frac{\partial^2 f}{\partial x_i \partial x_j}$ evaluated at **x***. (Note that $f_{22} > 0$ and $f_{11}\, f_{22} > f_{12}^2 \geq 0$ imply that $f_{11} > 0$).