# Mathematics for Economists

Lin Zhu
Tsinghua University
School of Economics and Management

Fall 2013

# Chapter 1

# Introduction

'Economics has been defined as the study of making the best use of scarce resources', or put in other words, economic activities can be viewed as reallocation of resources to achieve the best allocation, from the agents' perspective. In microeconomics, economic agents are often confronted by problems of this sort:

- How does a consumer select a consumption bundle to maximize his/her utility, subject to his/her budget constraint?
- How does a competitive firm make decision on production plan to maximize its profit, subject to its resource constraint?
- How does a firm manager select inputs to minimize costs, subject to its output target?

These problems can be categorized as static optimization. In contrast, in macroeconomics setup, agents focus more on dynamic optimization problems. In this chapter, we use several examples to illustrate the problem setup, in a simplied manner, that we will be dealing with in this course.

## 1.1 Basic Examples of Optimization in Economics

**Example 1.1 (Unconstrained Optimization): Profit maximization by a perfectly competitive firm** (i.e. all prices are treated as given)

$$\pi(p,w) = \max_x \{pF(x) - w'x\}, \tag{1.1}$$

where $w = (w_1, \ldots, w_n)'$ is vector of factor prices, $x = (x_1, \ldots, x_n)'$ is vector of factor inputs, $F(x)$ is production function and $p$ is price of the output good. Then $\pi(p,w)$ is the maximum profit function, also called **value function** of problem (1.1), and $x^*(p,w) := \arg\max\{pF(x) - w'x\}$ is vector of factor demand functions.
**Two inputs**: (capital-labor)

$$\pi(p,r,w) = \max_{K,L} \{pF(K,L) - rK - wL\}. \tag{1.1a}$$

**Example 1.2 (Equality constrained optimization): Cost minimization by a perfectly competitive firm**

$$\begin{aligned} C(w,y) &= \min_x w'x \qquad (1.2) \\ \text{s.t. } F(x) &= y, \end{aligned}$$

$C(w,y)$ is the cost function, which is the value function of problem (1.2). A similar problem is the expenditure minimization problem faced by a consumer:

$$\begin{aligned} E(p,u) &= \min_x p'x \qquad (1.2a) \\ \text{s.t. } U(x) &= u, \end{aligned}$$

here $p$ denotes price vector of consumption goods and $U(x)$ is the utility function. $E(p,u)$ is called expenditure function and the resulting optimal solution $X^*(p,u)$ is termed as Hicksian demand function or compensated demand function.

Interpretation: the minimum cost to achieve a certain level of output target for the firm; the minimum expenditure to reach a certain level of happiness for the consumer.

**Example 1.3 (Inequality constrained optimization): Utility maximization by a representative consumer**

$$\begin{aligned} V(p,I) &= \max_c U(c) \\ \text{s.t. } p'c &\leq I,\ c \geq 0, \end{aligned} \tag{1.3}$$

where $c=(c_1,\ldots,c_n)'$ is vector of consumption bundle, $p=(p_1,\ldots,p_n)'$ is vector of prices (taken by consumer as given) and $I$ is the income. $V(p,I)$ is called **indirect utility function,** which is the value function of problem (1.3), and $C_j^*(p,I)$, $j=1,\ldots,n$ denote optimal solutions, called Mashallian demand functions.

**Example 1.4: General equilibrium in a static economy** which includes both consumers and firms
*Consumers*, indexed by $i=1,\ldots,n$, solve the problem:

$$\begin{aligned} & \max_{c^i,l^i} \left\{U^i\left(c^i\right)-v\left(l^i\right)\right\} \\ \text{s.t.}\quad & p'c^i \leq I^i \\ & 0 \leq l^i \leq 1 \\ & I^i = wl^i + \sum_{j=1}^{m} \alpha_{ij}\pi_j, \end{aligned} \tag{1.4}$$

where $l^i$ is the supply of labor, $w$ is the wage rate, $\alpha_{ij}$ is the share of consumer $i$ in the profit of firm $j$. It holds that $\sum_{i=1}^m \alpha_{ij}=1$ for all firms $j=1,\ldots,m$. Consumers take $p$, $w$ and $\pi_j$ as given, choose their consumption $c^i$ and labor supply $l^i$.
*Firms*, indexed by $j=1,\ldots,m$, produce distinct goods, each of which has a unique production function $f^j$. Each firm solves the problem

$$\pi_j = \max_{L_j}\left\{p_j f^j\left(L_j\right) - wL_j\right\}.$$

Firms take $p_j$, $w$ as given, and choose labor inputs $L^j$ to maximize their profits.
*Competitive Equilibrium*: the vector of prices $p$ and the wage $w$ are set such that all markets clear, i.e. the labor market satisfies:

$$\sum_{i=1}^{n} l^i = \sum_{j=1}^{m} L_j$$

and the goods markets satisfies:

$$\sum_{i=1}^{n} c_j^i = y_j,$$

where $c_j^i$ is the $j$-th coordinate of the optimal solution $c^i$.

Note: We could define wage rates specific to firms, but in equilibrium the wage rates will be equal across firms, why?

**Example 1.5: Overlapping generation models**
Let there be a single good (produced and consumed). Let each individual live for two periods: young age (when individual works, consumes and saves for retirement) and old age (when individual retires and consumes savings carried from young age).

An individual who is young in the time period $t$ solves the following problem:

$$\begin{aligned} & \max U\left(c_t^y, c_t^o, l_t\right) \\ \text{s.t.} \quad & c_t^y + s_t \leq (1-\tau) w_t l_t \\ & c_t^o = (1+r_t) s_t + T_{t+1};\ 0 \leq l_t \leq 1 \end{aligned} \tag{1.5}$$

where $c_t^y, c_t^o$ are quantities of a single consumption/capital good consumed when young and old, respectively; $l_t$ is labor supply when young; $s_t$ is savings when young; $\tau$ is tax rate collected by government; $w_t$ is wage rate and $r_t$ is interest rate in period $t$; $T_{t+1}$ is lum-sum transfer to the old (social security) at time $t+1$ by government.

Let there always be $N$ identical young agents, which implies that the number of old agents is also always equal to $N$. Let the production in the economy be perfectly competitive with constant returns to scale production function $F(K, L)$. One can define competitive equilibrium for this overlapping generations economy as a sequence of variables $\{c_t^y, c_t^o, l_t, s_t, w_t, r_t, T_t\}$ such that:

(i) for any t, $\{c_t^y, c_t^o, l_t, s_t\}$ solves (1.5) given $w_t, r_t, T_{t+1}$;

(ii) factor markets clear, i.e. $K_{t+1} = N s_t$, $L_t = N l_t$ for all $t$;

(iii) factor markets are perfectly competitive, hence factor returns are equal to their respective marginal products:

$$\begin{aligned} 1 + r_t &= F_K\left(K_{t+1}, L_{t+1}\right) \\ w_t &= F_L\left(K_t, L_t\right); \end{aligned}$$

(iv) $T_{t+1} = \tau w_{t+1} l_{t+1}$, that is, tax revenues from the working young are transferred to the retired old; this constitute the so called "pay as you go" social security system.

**Example 1.6 (dynamic programming): Optimal intertemporal choice (Ramsey-Cass-Koopmans Model)**
Let there be one good produced and consumed in each time period, which can also be used as capital investment.

$$\begin{aligned} V(y_0) &= \max \sum_{t=0}^{T} \beta^t U(c_t) \\ \text{s.t. } y_t &= c_t + s_t,\ t = 0, \ldots, T, \\ y_{t+1} &= f(s_t),\ t = 0, \ldots, T-1, \\ c_t &\geq 0,\ s_t \geq 0,\ t = 0, \ldots, T, \end{aligned} \tag{1.6}$$

where $c_t$ is consumption at time $t$, $s_t$ is saving at time $t$, $y_t$ is quantity produced at time $t$, $f(\cdot)$ is the production function, and $0 < \beta < 1$ is the discount factor (time preference). Here the time horizon can be either finite or infinite, and $y_0 > 0$ is the initial endowment.
*Special case A*: $f(x) = (1+r)x$ which is the "consumption-saving" problem;
*Special case B*: $f(x) = x$ so that $y_{t+1} = s_t$, which is the so-called "cake-eating" problem.

## 1.2 Mathematical Preliminaries (Review)

### 1.2.1 Linear Vector Space

Key concepts: Linear independence, norm, hyperplane, matrices, convex sets, rank of a matrix in relation to linear independence

**Definition 1.1**: $L$ is a *linear vector space* (over the field of real numbers $\mathbb{R}$) if and only if (abbreviated as iff hereafter) it has the following properties:

(i) $x, y \in L \Rightarrow \alpha x + \beta y \in L$ for all $\alpha, \beta \in \mathbb{R}$;
(ii) $L$ contains an element zero "0" such that: $x + 0 = x$ and $x \cdot 0 = 0$ for all $x \in L$.
Note: $L'$ is a linear subspace of $L$ iff $L' \subset L$ and $L'$ is a linear vector space itself.

**Definition 1.2**: $H \subset L$ is a *hyperplane* if whenever $x, y \in H$ then then the straight line through $x$ and $y$ belongs to $H$, i.e. $x, y \in H \Rightarrow \forall \alpha \in \mathbb{R}$, it follows that $\alpha x + (1-\alpha)y \in H$.

**Exercise**: Prove that the following statements are true:
(1) If $a \in H$, where $H$ is a hyperplane in a linear vector space $L$, then $L' = \{x - a : x \in H\}$ is a linear subspace of $L$.
(2) The set $\{x \in L : f(x) = c\}$ is a hyperplane, where $f(\cdot)$ is a linear functional. (*Linear functional* $f$: $f(ax + by) = af(x) + bf(y)$).

**Definition 1.3**: *Linear dependence and independence*:
The vectors $x^1, x^2, \ldots, x^m \in L$ are *linearly dependent* if there are real numbers $\alpha_1, \ldots, \alpha_m \in \mathbb{R}$, at least one of which is nonzero, such that $\sum_{i=1}^m \alpha_i x^i = 0$;
The vectors $x^1, x^2, \ldots, x^m \in L$ are *linearly independent* if they are not linearly dependent, or equivalently: $\sum_{i=1}^m \alpha_i x^i = 0 \Rightarrow \alpha_i = 0$ for all $i = 1, \ldots, m$.

**Definition 1.4**: *Norm* in a vector space:
Given a vector space $V$ over the real field $\mathbb{R}$, a *norm*, denoted as $\|\cdot\|$, on $V$ is a function mapping from $V$ to $\mathbb{R}$ with the following properties: for all $a \in \mathbb{R}$ and all $u, v \in V$,
(i) *Positive homogeneity*: $\|av\| = |a| \|v\|$;
(ii) *Triangle inequality*: $\|u + v\| \leq \|u\| + \|v\|$;
(iii) $\|v\| = 0$ implies that $v$ is the zero vector in $V$.
*Seminorm* is a *norm* without property (iii).
*Euclidean norm* in $n$-dimensional Euclidean space $\mathbb{R}^n$ is defined as: for $x = (x_1, \ldots, x_n)' \in \mathbb{R}^n$, $\|x\| := \sqrt{x_1^2 + \cdots + x_n^2}$.

**Definition 1.5**: *Basis* in a vector space:
A *basis* $B$ of a vector space $V$ (over the real field $\mathbb{R}$) is a set of vectors such that
(i) for every finite subset $B_0 \subset B$, the vectors in $B_0$ are linearly independent, i.e. suppose $B_0 = \{v_1, \ldots, v_n\}$, for all $a_1, \ldots, a_n \in \mathbb{R}$, if $a_1 v_1 + \cdots + a_n v_n = 0$, then $a_1 = \cdots = a_n = 0$;
(ii) for every $v \in V$, there exist $a_1, \ldots, a_m \in \mathbb{R}$ and $v_1, \ldots, v_m \in B$ such that $v = a_1 v_1 + \cdots + a_m v_m$.
If $B$ contains finite number of elements, $V$ is said to be *finite-dimensional* vector space, otherwise, $V$ is *infinite-dimensional*.
*Digression*: for a *finite-dimensional* vector space $V$, all norms are equivalent in the sense that: if $\|\cdot\|_a$ and $\|\cdot\|_b$ are two norms defined on $V$, then $\|v\|_a = 0 \Leftrightarrow \|v\|_b = 0$.

**Definition 1.6**: *Dimension*:
*Dimension of a linear space* $L$ is the number of vectors in the basis;
*Dimension of a hyperplane*: We learned that each hyperplane can be shifted to become a linear subspace, then the dimension of the hyperplane is defined as the dimension of the shifted linear subspace.

Exercise: Prove the hyperplane $H := \left\{(x_1, \ldots, x_n)' \in \mathbb{R}^n : p_1 x_1 + \cdots + p_n x_n = I\right\}$ (budget constraint equation) is an $(n-1)$-dimensional hyperplane.

*Inner product* "$\cdot$" in Euclidean space is simply defined as: for $x = (x_1, \ldots, x_n)', y = (y_1, \ldots, y_n)' \in \mathbb{R}^n$, $x \cdot y := x'y = x_1 y_1 + \cdots + x_n y_n$.
**Definition 1.6:** *Orthonormal basis* for Euclidean Space $\mathbb{R}^n$:
$B = \left\{x^1, \ldots, x^n\right\}$ is said to be a set of orthonormal basis if $B$ is a basis for $\mathbb{R}^n$ and $x^i \cdot x^j = 1$ if $i = j$, and $x^i \cdot x^j = 0$ if $i \neq j$, for all $i, j = 1, \ldots, n$.

### 1.2.2 Basic Topological Concepts

*Key concepts*: Convergence of a sequcence, continuity of a function, convex sets, seperating hyperplane, seperating hyperplane theorem.

**Definition 1.7**: *Convergence of a sequence*:
A sequence of elements, $\{x^n\} \subset \mathbb{R}^n$ is said to *converge* to a point $x \in \mathbb{R}^n$, if for any $\delta > 0$, there is a number $N(\delta)$, such that for all $n > N(\delta)$, $\|x^n - x\| < \delta$.

**Definition 1.8**: *Continuity* of a function
A function $f : \mathbb{R}^n \to \mathbb{R}^m$ is said to *continuous* at a point $x$ if for *all* sequences $\{x^n\}$ converging to $x$, the corresponding sequence $\{f(x^n)\}$ converges to the point $f(x)$.
*Upper semi-continuity*: $\forall x_n \to x$, it holds that $\lim_{n\to\infty} f(x_n) \leq f(x)$. *Lower semi-continuity* is similarly defined.

Note: we generally impose continuity in the optimization problem, but there are cases in which continuity is too strong, and hence a less strong assumption is imposed, for example, upper semi-continuity.

**Definition 1.9**: *Convex* set
A set $C$ is *convex* if for all $x, y \in C$ and for all $\alpha \in [0,1]$, $\alpha x + (1-\alpha) y \in C$.

**Definition 1.10**: *Seperating hyperplane* (recall the hyperlane can be equivalently defined by a linear functional)
A hyperplane $H = \{x : f(x) = c\}$ seperates two sets $C_1$ and $C_2$, if for all $x \in C_1$, $f(x) \leq c$ and for all $x \in C_2$, $f(x) \geq c$. (or reverse both inequalities)

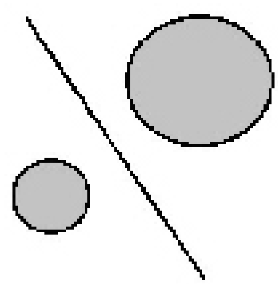


Fig 1.1 A seperating hyperplane

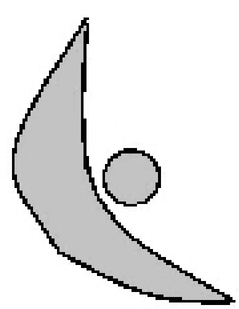


Fig 1.2 No seperating hyperplane exists

**Theorem 1.1 (Seperating Hyperplane Theorem)**: Suppose $C_1$ and $C_2$ are non-empty, convex sets in $\mathbb{R}^n$ such that $C_1 \cap C_2$ has no interior points and at least one set has nonempty interior, then there exists a hyperplane $\{x : a \cdot x = c\}$ with the defining vector $a \in \mathbb{R}^n$ that seperates $C_1$ and $C_2$.
Remark: The theorem allows for two sets to intersect on the boundary.

### 1.2.3 Convex and Quasi-Convex Functions

*Key concepts*: convex/concave function, quasi-convex/concave, Jensen's inequality

**Definition 1.11**: *Convex function*
A real-valued function $f$ defined on a convex set $C$, i.e. $f : C \mapsto \mathbb{R}$, is a *convex function* if the following inequality holds:

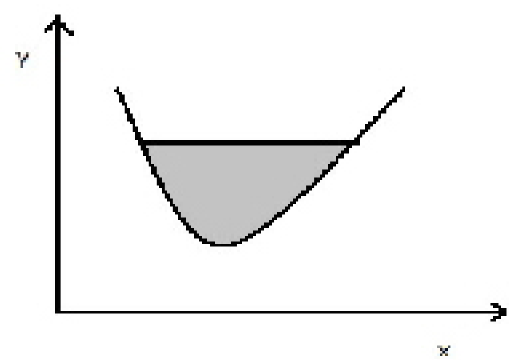
$$f(\alpha x + (1-\alpha) y) \leq \alpha f(x) + (1-\alpha) f(y) \tag{1.7}$$

for all $x, y \in C$ and $\alpha \in [0,1]$.
*Note*: *Concave* function is similarly defined by reversing the above inequality.
A function is *strictly convex* if the above inequality holds strictly for $\alpha \in (0,1)$.

Digression:
(1) For continuous function $f$, convexity can be equivalently defined as

$$f\left(\frac{x+y}{2}\right) \leq \frac{1}{2} f(x) + \frac{1}{2} f(y).$$

(2) Another equivalent definition is given by graph: $f$ is convex on $C \subset \mathbb{R}^n$ if its supergraph, i.e. the set $\overline{G}_f := \{(x,y) \in C \times \mathbb{R} : y \geq f(x)\}$ is convex. (See the graph below.)




**Exercise**: Von Neumann-Morgenstern Expected Utility

$$EU = \int U(x)\, p(x)\, dx, \text{ where } \int p(x)\, dx = 1,$$

or if the probability distribution is discrete,

$$EU = \sum_{i=1}^{m} U(x_i)\, p_i, \text{ where } \sum_{i=1}^{m} p_i = 1.$$

Show that if $U(x)$ is concave, then von Neumann-Morgenstern expected utility defines risk averse preferences, i.e. $E[U(X)] \leq U(EX)$.

*Jensen's Inequality for convex functions*

The function $f$ defined on a convex set $C$ is a convex function iff for any $x^1, \dots, x^m$ in $C$ the following inequality holds:

$$f\left(\sum_{j=1}^{m} \alpha_j x^j\right) \leq \sum_{j=1}^{m} \alpha_j f\left(x^j\right) \tag{1.8}$$

for all $\alpha_j$ such that $\sum_{j=1}^{m} \alpha_j = 1$.
Note: the definition of convex function is a special of this inequality with $m = 2$.
Another convenient form of Jensen's inequality:

$$\left(\sum_{j=1}^{m} \gamma_j\right) f\left(\frac{\sum_{j=1}^{m} \gamma_j x^j}{\sum_{j=1}^{m} \gamma_j}\right) \leq \sum_{j=1}^{m} \gamma_j f\left(x^j\right) \tag{1.9}$$

for any $\gamma_j \geq 0$, $j = 1, \dots, m$ (not all $\gamma_j$ are zeros simutaneously).

**Exercise**: Apply (1.9) to the function $f(x) = x^2$ to derive the Cachy-Schwartz inequality:

$$\sum_{j=1}^{m} \alpha_j y_j \leq \left(\sum_{j=1}^{m} \alpha_j^2\right)^{1/2} \left(\sum_{j=1}^{m} y_j^2\right)^{1/2},$$

for all $\alpha = (\alpha_1, \dots, \alpha_m)'$ and $y = (y_1, \dots, y_m)' \in \mathbb{R}^m$.

**Definition 1.12**: *Quasi-Convexity/Concavity*
(1) A function $f$ defined on a convex set $C \subset \mathbb{R}^n$ is *quasi-convex* if for all $c \in \mathbb{R}$,

$$A_c^- := \{x \in C : f(x) \leq c\}$$

is a convex set.
(2) Similarly, $f$ is a *quasi-concave* function if for all $c \in \mathbb{R}$,

$$A_c^+ := \{x \in C : f(x) \geq c\}$$

is a convex set.

Note the differences between the definitions of quasi-convex function and convex function using graph.

Equivalent definitions:
(1) A function $f$ defined on a convex set $C$ is *quasi-convex* if

$$f(\alpha x + (1-\alpha) y) \leq \max\{f(x), f(y)\}$$

for all $\alpha \in [0,1]$ and $x, y \in C$;
(2) A function $f$ defined on a convex set $C$ is *quasi-concave* if

$$f(\alpha x + (1-\alpha) y) \geq \min\{f(x), f(y)\}$$

for all $\alpha \in [0,1]$ and $x, y \in C$.

**Exercise**: Prove the equivalence between the above two definitions.

### 1.2.4 Gradients, Jacobian Matrix and Hessian Matrix

*Key concepts*: derivative, gradient, Jacobian matrix, Hessian matrix, Taylor rule

Differentiability is an important concept in optimization, and indeed, in many applications, the functional forms we specify pocess this property, for example utility functions. For $f : \mathbb{R} \mapsto \mathbb{R}$, the formal definition of *derivative* of $f$ at $x$ is

$$f'(x) = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h};$$

and for $f : \mathbb{R}^n \mapsto \mathbb{R}$, the *partial derivative* at $x = (x_1, \ldots x_n)'$ w.r.t. $x_i$ is

$$\frac{\partial f(x)}{\partial x_i} = \lim_{h \to 0} \frac{f(x_i + h, x_{-i}) - f(x)}{h}$$

where $x_{-i}$ denotes all other coordinates except $x_i$.

**Definition 1.13**: *Gradient* and *Jacobian Matrix*
(1) Suppose $f$ is a function of multivariate variables, $f : \mathbb{R}^n \mapsto \mathbb{R}$, the gradient of $f$ at $x$, denoted as $\nabla f(x)$, is the $n$-dimensional (column) vector consisting of all partial derivatives:

$$\nabla f(x) = \left( \frac{\partial f(x)}{\partial x_1}, \ldots, \frac{\partial f(x)}{\partial x_n} \right)';$$

(2) Suppose $f$ is a vector function, $f : \mathbb{R}^n \mapsto \mathbb{R}^m$, the Jacobian matrix of $f$ at $x$, denoted as $J_f(x)$, is the $m \times n$ matrix as follows:

$$J_f(x) = \begin{bmatrix} \frac{\partial f_1(x)}{\partial x_1} & \cdots & \frac{\partial f_1(x)}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m(x)}{\partial x_1} & \cdots & \frac{\partial f_m(x)}{\partial x_n} \end{bmatrix}$$

where $f(x) = (f_1(x), \ldots, f_m(x))'$.

First-order Approximation:
For a small vector $v = (v_1, \ldots, v_n)' \in \mathbb{R}^n$ and $f : \mathbb{R}^n \mapsto \mathbb{R}$, the first-order approximation of $f(x^0 + v)$ is given by

$$f(x^0 + v) \approx f(x^0) + \nabla f(x^0) \cdot v;$$

or if $f : \mathbb{R}^n \mapsto \mathbb{R}^m$, we have

$$f(x^0 + v) \approx f(x^0) + J_f(x^0)\, v.$$

Indeed, it is easy to show that

$$\frac{f(x^0 + v) - f(x^0) - \nabla f(x^0) \cdot v}{\|v\|} \to 0$$

as $\|v\| \to 0$, hence the above approximation can be more precisely written as

$$f(x^0 + v) = f(x^0) + \nabla f(x^0) \cdot v + o(\|v\|)$$

where $o(\|v\|)$ denotes a term of smaller order than $\|v\|$, i.e. $o(\|v\|) / \|v\| \to 0$ as $\|v\| \to 0$.

Note: the gradient vector $\nabla f(x^0)$ at $x^0$ points in the direction of the steepest increase of the function $f$, since $f(x) - f(x^0) \approx \nabla f(x^0) \cdot (x - x^0)$ when $x$ is close to $x^0$, then the inner product will have largest value when the direction $(x - x^0)$ (with fixed length $\|x - x^0\|$) is the same as the that of gradient $\nabla f(x^0)$. (why?)

To characterize the convexity/concavity using gradient (first-order derivative), we first consider a differentiable function of single variable, i.e. $f : \mathbb{R} \mapsto \mathbb{R}$. Take a look at the graph below:

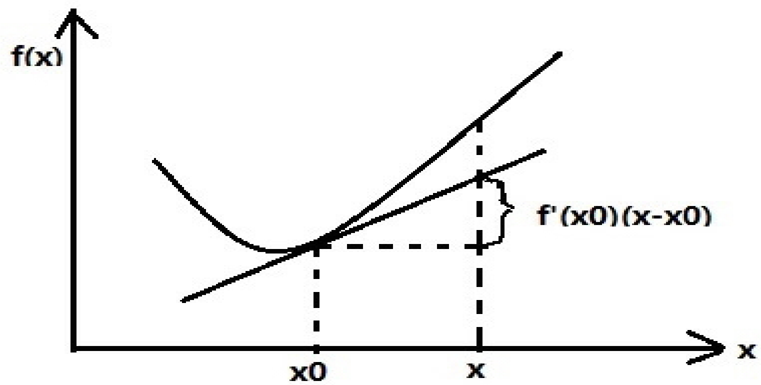



Fig 1.4 Total variation property of a convex function

We know that for a differentiable function $f(x)$ to be convex on $[a,b]$ if and only if $f'(x)$ is non-decreasing on $[a,b]$. Then using this fact, we can prove the following result.

<u>Total variation inequality for convex function</u>: Let $f$ be differentiable on $[a,b]$, then $f$ is convex function *iff* the inequality below is true for any $x$ and $x^0$ in $(a,b)$

$$f(x) - f\left(x^0\right) \geq f'\left(x^0\right)\left(x - x^0\right).$$

This characterization can be carried over to multivariate case.

**Theorem 1.2** (Total variation inequality for convex function): Let $f$ be differentiable in $C \in \mathbb{R}^n$, $f$ is convex *iff* the following inequality is true

$$f(x) - f\left(x^0\right) \geq \nabla f\left(x^0\right) \cdot \left(x - x^0\right)$$

for all $x^0, x \in C$.

**Proof**: For any fixed $x^0, x \in C$, define $F(t) = f\left(x^0 + t\left(x - x^0\right)\right)$, then $F(0) = f\left(x^0\right)$ and $F(1) = f(x)$. Since $f(x)$ is convex, so if $F$ (prove as exercise). Notice that $F'(t) = \nabla f\left(x^0 + t\left(x - x^0\right)\right) \cdot \left(x - x^0\right)$ by chain rule, hence by mean value theorem,

$$\begin{aligned} f(x) - f\left(x^0\right) &= F(1) - F(0) \\ &= F'(t') \end{aligned}$$

for some $t' \in [0,1]$. Since $F(t)$ is convex, the derivative $F'(t)$ is increasing in $t$, implying

$$f(x) - f\left(x^0\right) \geq F'(0) = \nabla f\left(x^0\right) \cdot \left(x - x^0\right).$$

<u>Note</u>: If $\left\|x - x^0\right\| = 1$, then $\frac{\partial F}{\partial t}|_{t=0} = \nabla f\left(x^0\right) \cdot \left(x - x^0\right)$ is the <u>*directional derivative*</u> of $f$ at $x^0$ in the direction $\left(x - x^0\right)$

Suppose $f : \mathbb{R}^n \mapsto \mathbb{R}$ is twice differentiable, the <u>*Hessian matrix*</u> of $f$ is

$$H_f(x) = \begin{bmatrix} \frac{\partial^2 f(x)}{\partial x_1^2} & \frac{\partial^2 f(x)}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f(x)}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f(x)}{\partial x_2 \partial x_1} & \frac{\partial^2 f(x)}{\partial x_2^2} & \cdots & \frac{\partial^2 f(x)}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f(x)}{\partial x_n \partial x_1} & \frac{\partial^2 f(x)}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f(x)}{\partial x_n^2} \end{bmatrix};$$

the <u>second order approximation</u> is

$$f(x) = f\left(x^0\right) + \nabla f\left(x^0\right) \cdot \left(x - x^0\right) + \left(x - x^0\right)' H_f\left(x^0\right)\left(x - x^0\right) + o\left(\left\|x - x^0\right\|^2\right).$$

Note: The Hessian matrix is symmetric, i.e. $\frac{\partial^2 f(x)}{\partial x_i \partial x_j} = \frac{\partial^2 f(x)}{\partial x_j \partial x_i}$, whenever the second order derivatives are continuous.

## 1.2.4 First and Second Order Conditions for Unconstrained Optimization

### 1.2.4.1 Functions of a Single Variable.

**Theorem 1.3** (Fermat): *First order necessary conditions (FONC) of local extremum*

Let $f : [a,b] \mapsto \mathbb{R}$ attains a local extremum at an *interior* point $x^0$. If $f$ is differentiable at $x^0$, then $f'\left(x^0\right) = 0$.

Proof: (use first order approximation and proof by contradiction.)

Interpretation of Fermat Theorem:

If $x^0$ is an extremum point of a function $f$, then at least one of the following applies:

(i) $f'\left(x^0\right) = 0$;

(ii) $x^0$ is on the boundary of the domain of $f$;

(iii) $f$ is not differentiable at $x^0$.

But neither of (i)-(iii) guarantees that $x^0$ is an extremum point. Thus the union of (i), (ii), (iii) constitutes a necessary condition of extremum.

Example: inflection points, e.g. $f(x) = x^3$.

**Corollary 1.1:** Let $f$ be a convex (concave) function defined on an interval $[a,b]$ and $f'\left(x^0\right) = 0$, where $x^0 \in (a,b)$. Then $x^0$ is a *global minimum (maximum)* of $f$ on $[a,b]$.

Proof: Apply Theorem 1.2 (Total variation theorem for convex function) to obtain $f(x) - f\left(x^0\right) \geq f'\left(x^0\right)\left(x - x^0\right) = 0.$ ■

Note: If $f$ is a convex and differentiable, then the condition $f\left(x^0\right) = 0$ at an interior point $x^0$ is both necessary and sufficient for $x^0$ to be the global minimum of $f$.

**Theorem 1.4** *Second order sufficient conditions (SOSC) and second order necessary conditions (SONC) of extremum*

(i) (*SONC*) Let $x^0 \in (a,b)$ be a *local maximum* of $f$, which is twice differentiable in a neighborhood of $x^0$, then $f''\left(x^0\right) \leq 0$.

(ii) (*SOSC*) Let $x^0 \in (a,b)$, if $f'\left(x^0\right) = 0$ and $f''\left(x^0\right) < 0$, then $x^0$ is a **strict** *local maximum*.

Proof: (use the second order approximation.)

### 1.2.4.2 Functions of Several Variables

**Theorem 1.5** (FONC): Consider $f : C \mapsto \mathbb{R}$, where $C \subset \mathbb{R}^n$, that attains a local extremum at $x^0 \in int(C)$. Let $f$ be differentiable in a neighborhood of $x^0$, then $\nabla f\left(x^0\right) = 0$.

Proof: Let $F(t) = f\left(x^0 + t\left(x' - x^0\right)\right)$ for an arbitray $x'$ in the neighborhood of $x^0$, then apply Theorem 1.3.

Similarly as univariate case, for convex function of several variables, we have

**Corollary 1.2:** Let $f : C \mapsto \mathbb{R}$ be convex (concave) and differentiable in $C$. Let $\nabla f\left(x^0\right) = 0$ for some interior $x^0 \in C$, then $x^0$ is the global minimum (maximum) of $f$ in $C$.

To obtain second order conditions for convex/concave optimization, we first review the following concepts.

**Definition 1.14**: A symmetric matrix $A$ is:

*negative definite* if $x'Ax < 0$, for any $x \neq 0$;

*negative semi-definite* if $x'Ax \leq 0$, for any $x \neq 0$;

*positive definite* if $x'Ax > 0$, for any $x \neq 0$;

*positive semi-definite* if $x'Ax \geq 0$, for any $x \neq 0$.

**Theorem 1.6 (SONC)**: Let $f$ attains local minimum (maximum) at $x^0$ and be twice continuously differentiable in a neighborhood of $x^0$. Then $\nabla f\left(x^0\right) = 0$ (FONC) and its Hessian matrix $H_f\left(x^0\right)$ is positive (negative) semi-definite.

**Theorem 1.7 (SOSC)**: Let $\nabla f\left(x^0\right) = 0$ and let $H_f\left(x^0\right)$ be positive (negative) definite, then $x^0$ is a **strict** *local minimum* (maximum) of $f$.

The following theorems gives simple conditions for verifying a matrix being positive/negative (semi-) definite. We first introduce several concepts in matrix algebra.
**Definition 1.14**: *Principal Minor* and *Lead Principal Minor*
Let $A$ be an $n \times n$ matrix. A $k \times k$ submatrix of $A$ formed by deleting $(n-k)$ columns, say columns $i_1, i_2, \ldots, i_{n-k}$ and the same rows $i_1, i_2, \ldots, i_{n-k}$ from $A$ is called a $k^{th}$ *order principle submatrix* of $A$. The determinant of a $k \times k$ principal submatrix is called a $k^{th}$ *order principal minor* of $A$. The $k^{th}$ *order leading principal minor* is the determinant of the submatrix formed by first $k$ rows and first $k$ columns.

**Theorem 1.8 (***Silvester condition***):** Let $A$ be a $(n \times n)$ symmetric matrix with *leading principal minors* $\Delta_i$ of order $i = 1, \ldots, n$ respectively. Then
(i) $A$ is positive definite iff $\Delta_i > 0$ for all $i = 1, \ldots, n$;
$A$ is positive semi-definite iff $\Delta_i \geq 0$ for all $i = 1, \ldots, n$ and the same conditions apply to all *principal minors* (not just leading ones) of respective orders;
(ii) $A$ is negative definite iff $(-1)^i \Delta_i > 0$ for $i = 1, \ldots, n$;
$A$ is negative semi-definite iff $(-1)^{-i} \Delta_i \geq 0$ for all $i = 1, \ldots, n$ and the same conditions apply to all principal minors (not just leading ones) of respective orders;

**Theorem 1.9 (**Diagonal Dominance Condition**):**
Let $A$ be a symmetric matrix with strictly dominant diagonal, i.e.

$$|a_{ii}| > \sum_{j=1, j \neq i}^{n} |a_{ij}| \text{ for } i = 1, \ldots, n$$

Then:
(i) $A$ is positive definite if $a_{ii} > 0$, for all $i = 1, \ldots, n$;
(ii) $A$ is negative definite if $a_{ii} < 0$, for all $i = 1, \ldots, n$.
Note: Diagonal dominance is sufficient but not necessary for definiteness.

An alternative characterization of definiteness is via Eigenvalues. Recall eigenvalues $\lambda$ solve

$$\det(A - \lambda I) = 0,$$

and we know that if $A$ is symmetric, all roots of above equation are real numbers. (Hessian matrix is symmetric for almost all of our applications.)
**Theorem 1.10 (**Eigenvalue Condition**):** If A is symmetric, then:
(i) $A$ is positive definite iff all its eigenvalues are $> 0$;
$A$ is positive semi-definite iff all its eigenvalues are $\geq 0$;
(ii) $A$ is negative definite iff all its eigenvalues are $< 0$;
$A$ is negative semi-definite iff all its eigenvalues are $\leq 0$;

The following results characterize SOCs for convexity/concavity.
**Theorem 1.11** (*Second order necessary and sufficient condition of convexity(concavity)*): Let function $f$ be twice differentiable in an open convex set $C$, then $f$ is convex (concave) **iff** the Henssian matrix $H_f(x)$ is positive semi-definitie (respectively, negative semi-definite) **at all** $x \in C$.
Proof: "$\Rightarrow$" *Necessity*: Let $f$ be convex, then for any $x^0 \in C$ and $v \in \mathbb{R}^n$, there exists $t_0 > 0$ such that $x' = x^0 + t_0 v \in C$. Let $F(t) = f\left(tx' + (1-t)x^0\right)$, $t \in [0,1]$, then convexity of $f$ implies convexity of $F$, which further implies $F''(0) \geq 0$. Notice that

$$\begin{aligned} F''(0) &= \left(x' - x^0\right)' H_f\left(x^0\right)\left(x' - x^0\right) \\ &= t_0^2 v' H_f\left(x^0\right) v, \end{aligned}$$

hence $v' H_f\left(x^0\right) v \geq 0$ for any $v \in \mathbb{R}^n$.
"$\Leftarrow$" Sufficiency: For $f$ to be convex, it suffices to prove $F(t) = f(tx + (1-t)y)$ is convex for any $x, y \in C$. (do the rest algrebra as exercise.)

Note: The necessary and sufficient condition in the above theorem can be similarly carried over to the **strictly** convex (concave) case, for example, $f$ is strictly convex iff the Hessian matrix $H_f(x)$ is positive definite at all $x \in C$.

To give calculus conditions for quasi-convexity/concavity, we first introduce the following definition.
**Definition 1.15**: Suppose $f : \mathbb{R}^n \mapsto \mathbb{R}$ is a twice differentiable. The Bordered Hessian of $f$, denoted as $\overline{H}_f$, is the $(n+1) \times (n+1)$ matrix given by

$$\overline{H}_f(x) = \begin{bmatrix} 0 & \frac{\partial f}{\partial x_1} & \cdots & \frac{\partial f}{\partial x_n} \\ \frac{\partial f}{\partial x_1} & \frac{\partial^2 f}{\partial x_1^2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f}{\partial x_n} & \frac{\partial^2 f}{\partial x_n \partial x_1} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix}.$$

The following theorem characterize necessary conditions and sufficient conditions for quasi-convex/concave functions:

**Theorem 1.12 (Second Order Conditions for Quasi-Convexity/Concavity)**
Let $f : C \mapsto \mathbb{R}$ be a twice differentiable function defined on an open and convex set $C \subset \mathbb{R}^n$, and let $\overline{\Delta}_r$ be the leading principal minors of the Bordered Hessian matrix $\overline{H}_f$, then
(1) A necessary condition for the quasi-concavity of $f$ is that

$$(-1)^{r+1} \overline{\Delta}_r \geq 0, \ r = 2, 3, \ldots, n+1 \text{ and all } x \in C;$$

the necessary condition for quasi-convexity is $\overline{\Delta}_r \leq 0$ for all $r = 2, 3, \ldots, n+1$ and all $x \in C$;
(2) A sufficient condition for $f$ to be quasi-concave is that

$$(-1)^{r+1} \overline{\Delta}_r > 0, \ r = 2, 3, \ldots, n+1 \text{ and all } x \in C;$$

If $\overline{\Delta}_r < 0$ for all $r = 2, 3, \ldots, n+1$ and all $x \in C$, then $f$ is quasi-convex.

Exercise: Proof that $f(x) = x^3 + x$ is quasi-concave but not concave;

Here are some differences between concave and quasi-concave functions:
(1) A critical point (with zero first-order derivative) of a quasi-concave function need not be a local maximum, e.g. $f(x) = x^3$.
(2) The sum of quasi-concave functions needs not to be quasi-concave, while sum of concave functions is concave. For example, $f_1(x) = x^3$ and $f_2(x) = -x$ are both quasi-concave, but the sum $f(x) = x^3 - x$ is neither quasi-concave or quasi-convex.
(3) Any monotonic transformation of a concave function is quasi-concave.

### 1.2.5 Example

Profit maximization by a price-taking firm with Cobb-Douglas production function
The maximization problem is:

$$\max_{K,L} \left\{\Pi(K, L) = pF(K, L) - rK - wL\right\},$$

where $F(K, L) = AK^\alpha L^\beta$ and the coefficient $\alpha$ and $\beta$ are non-negative, $A$ is the technology parameter, $p$ is the price of output product. The FONCs of extremum are given by the system of equations $\nabla \Pi(K, L) = 0$ :

$$\frac{\partial \Pi}{\partial K} = \alpha p A K^{\alpha-1} L^\beta - r = 0 \text{ and } \frac{\partial \Pi}{\partial L} = \beta p A K^\alpha L^{\beta-1} - w = 0.$$

Therefore the FONCs can be written as

$$\begin{aligned} F_K &= \alpha A K^{\alpha-1} L^{\beta} = \frac{r}{p}, \\ F_L &= \beta A K^{\alpha} L^{\beta-1} = \frac{w}{p}, \end{aligned}$$

which gives the optimal capital labor ratio $k^* = \frac{K^*}{L^*} = \frac{\alpha w}{\beta r}$, therefore, $K^* = \frac{\alpha w}{\beta r} L^*$.

Substituting this in either of the equations above and assuming $\alpha + \beta \neq 1$, we obtain the unique solution of the system:

$$L^* = \left[\frac{\alpha^{\alpha}\beta^{1-\alpha}}{r^{\alpha}w^{1-\alpha}}\right]^{\frac{1}{1-\alpha-\beta}} \quad ; \quad K^* = \frac{\alpha w}{\beta r}\left[\frac{\alpha^{\alpha}\beta^{1-\alpha}}{r^{\alpha}w^{1-\alpha}}\right]^{\frac{1}{1-\alpha-\beta}}.$$

Consider the Hessian matrix of $\Pi(K, L)$ and analyze whether the above solution satisfies the second order necessary and sufficient conditions for maximum:

$$H_{\Pi}(K, L) = \begin{bmatrix} \Pi_{KK} & \Pi_{KL} \\ \Pi_{LK} & \Pi_{LL} \end{bmatrix} = p \begin{bmatrix} F_{KK} & F_{KL} \\ F_{LK} & F_{LL} \end{bmatrix},$$

where $F_{KK} = \alpha(\alpha - 1) A K^{\alpha-2} L^{\beta}$, $F_{KL} = \alpha\beta A K^{\alpha-1} L^{\beta-1}$, $F_{KK} = \beta(\beta - 1) A K^{\alpha} L^{\beta-2}$. Then

$$H_{\Pi}(K, L) = pAK^{\alpha-2}L^{\beta-2} \begin{bmatrix} \alpha(\alpha-1)L^2 & \alpha\beta KL \\ \alpha\beta KL & \beta(\beta-1)K^2 \end{bmatrix},$$

hence $\Delta_1 = \alpha(\alpha - 1)L^2$, $\Delta_1' = \beta(\beta - 1)K^2$, $\Delta_2 = (1 - \alpha - \beta)\alpha\beta K^2 L^2$.

**Exercise**:

(1) Use the above formulae to determine conditions on coefficient $\alpha$ and $\beta$ that ensure sufficient conditions of maximum;

(2) Use the above formulae to determine conditions on coefficient $\alpha$ and $\beta$ that ensure (i) concavity of the profit function; (ii) strict concavity of the profit function;

(3) Notice in the above analysis we require $\alpha + \beta \neq 1$, please analyze the case $\alpha + \beta = 1$.

### 1.2.6 Homogeneous and Homothetic Functions

**Definition 1.16**: $f : \mathbb{R}^n_+ \mapsto \mathbb{R}_+$ is homogeneous of degree $r$ if $f(\lambda x) = \lambda^r f(x)$ for all $\lambda \geq 0$, $x \geq 0$.

**Theorem 1.13 (**Euler**)**: Let $f$ be homogeneous of degree $r$ and differentiable. Then

$$\nabla f(x) \cdot x = r f(x).$$

Proof: By definition $f(\lambda x) = \lambda^r f(x)$. Differentiate its left- and right-hand sides with respect to $\lambda$ :

$$\frac{\partial f(\lambda x)}{\partial \lambda} = \nabla f(\lambda x) \cdot x, \qquad \frac{\partial \lambda^r f(x)}{\partial \lambda} = r\lambda^{r-1} f(x),$$

let $\lambda = 1$, then $\nabla f(x) \cdot x = r f(x)$.

**Corollary 1.3**: Let $f : \mathbb{R}^n_+ \mapsto \mathbb{R}_+$ be homogeneous of degree $r$ and differentiable. Then $f_i := \frac{\partial f(x)}{\partial x_i}$ is homogeous of degree $(r - 1)$ for all $i = 1, \ldots, n$.
Proof: Differentiate both sides of $f(\lambda x) = \lambda^r f(x)$ w.r.t. $x_i$.

**Corollary 1.4**: Let $f : \mathbb{R}^n_+ \mapsto \mathbb{R}_+$ be homogeneous of degree 1 and twice differentiable. Then its Hessian matrix $H_f(x)$ is singular (i.e. $\det H_f(x) = 0$) for all $x$.

Proof: According to Corollary 1.3, $f_i(x)$ is homogeneous of degree 0 for each $i = 1, \ldots, n$. Apply Theorem 1.12 seperately to each function $f_i(x)$ :

$$\nabla f_i(x) \cdot x = 0,$$

implying $H_f(x)\,x = 0$. Hence $H_f(x)$ is singular.

**Corollary 1.5**: Let $f : \mathbb{R}^n_+ \mapsto \mathbb{R}_+$ be homogeneous of degree $r \geq 0$ and differentiable. Then

$$\frac{f_i(\lambda x)}{f_j(\lambda x)} = \frac{f_i(x)}{f_j(x)}, \text{ for } i, j = 1, \ldots, n$$

i.e., the marginal rates of substitution are *scale-independent.*

**Example**: Application of Euler's theorem
Consider the profit maximization example of section 1.2.5:

$$\max_{K,L} \{\Pi(K,L) = pF(K,L) - rK - wL\}.$$

Recall that solution exists if $\alpha + \beta \leq 1$, find the average total cost at optimum:

$$ATC = \frac{Total\ Cost}{Output} = \frac{rK^* + wL^*}{F(K^*, L^*)}.$$

By FONCs, $F_K = r/p$, $F_L = w/p$. Then, since $F(\lambda K, \lambda L) = \lambda^{\alpha+\beta} F(K,L)$, applying Euler's Theorem we can write

$$ATC = \frac{(F_K K^* + F_L L^*)\,p}{F(K^*, L^*)} = \frac{(\alpha+\beta)\,F(K^*, L^*)\,p}{F(K^*, L^*)} = (\alpha+\beta)\,p.$$

**Definition 1.17**: A function $f : \mathbb{R}^n_+ \mapsto \mathbb{R}_+$ is homothetic if $f(x) = h(g(x))$ where $h : \mathbb{R}_+ \mapsto \mathbb{R}_+$ is strictly increasing and $g : \mathbb{R}^n_+ \mapsto \mathbb{R}_+$ is homogeneous of degree $k$.

Application: An important application of homogeneous/homothetic functions is that homogeneous/homothetic utility functions rule out "income effects" on demand, i.e. for constant prices, consumers demand goods in the same proportion as income changes.