Simon Fraser University, Department of Economics
Econ 798, Introduction to Mathematical Economics
Prof. B. Antoine
(based on the notes of Prof. A. Karaivanov)

Part 2: Elements of Analysis

1 Real analysis

1.1 Functions

- **Definition (function):** Consider two sets A and B and suppose that with each $x \in A$ we associate an element $y \in B$ which we denote by $f(x)$. Then f is said to be a *function* from A to B and is denoted $f : A \to B$.

- The elements $f(x)$ are called the *values* of f.

- The set of all possible values of f is called the *range* of f.

- The set of $x \in A$ for which f is defined is called the *domain* of f.

  – Exercise: find the ranges and domains of the following functions:
  (i) $\sqrt{x}$; (ii) $\frac{1}{x^2+2x-3}$; $\log(x)$.

- **Types of functions:** Let A and B be two sets and let f be a function from A to B.

  – if $\forall y \in B$, there exists $x \in A$ such that $f(x) = y$, we say that f maps A **onto** B.
  (f is said to be *surjective*).

  – if $\forall x_1, x_2 \in A$ with $x_1 \neq x_2$, we have that $f(x_1) \neq f(x_2)$, we say that f is **one-to-one** function of A to B.
  (f is said to be *injective*).

  – examples:
  (i) $f(x) = 2x + 5$ is both one-to-one and onto;
  (ii) $f(x) = 1 - (x - 1)^2$ with $A = [0, 2]$ and $B = [0, 1]$ is onto but not one-to-one;
  (iii) $f(x) = sin(x)$ is neither onto nor one-to-one for $B = \mathbb{R}$.

- **Inverse:** if $E \subseteq B$, then $f^{-1}(E)$ denotes the set of all $x \in A$ such that $f(x) \in E$. We call f^{-1} the *inverse* of f. If f^{-1} is a function (the set of x s.t. $f(x) = y$ for all $y \in E$ is a singleton) then it is called the inverse function of f.

– Example: the inverse of $f(x) = 2x + 5$ is $f^{-1}(y) = (y - 5)/2$.

– Note: to find the inverse function (when it exists!), simply solve the equation $y = f(x)$ to find x as a function of y, and call the result $f^{-1}(y)$.

1.2 Sequences

- **Definition (convergent sequence):** a sequence $\{p_n\}$ in a metric space X is said to be *convergent* if $\exists p \in X$ such that $\forall \epsilon > 0$, $\exists N \in \mathbb{N}$ such that $n \geq N$ implies $d(p_n, p) < \epsilon$. We also say that p is the limit of $\{p_n\}$ and write:

$$p_n \to p \qquad \text{or} \qquad \lim_{n\to\infty} p_n = p$$

- if $\{p_n\}$ does not converge it is said to *diverge*. Notice that the definition of convergence depends on the set X. E.g., the sequence $1/n$ converges in $\mathbb{R}$ (to 0), but not in $\mathbb{R}_+$.

- **Examples:**

(i) show that $\{a_n\}$ where $a_n = (-1)^n$ does not converge.

(ii) Show that $\{p_n\}$ with $p_n = 1/n$ converges to $p = 0$.

- **Definition (bounded sequence):** a sequence $\{p_n\}$ in the metric space X is bounded if all its elements are bounded (that is, $\exists m \in X$ and a positive real number M s.t. $d(p_n, m) < M$ for all n).

- **Theorem:** Let $\{p_n\}$ be a sequence in a metric space X

  (a) $\{p_n\}$ converges to $p \in X$ if and only if every neighborhood of p contains p_n for all but finitely many n.

  (b) if $\{p_n\}$ converges then it is bounded.

  – proof: Rudin, pp. 48-49.

- **Application (alternative definition for closed set):** a set S is *closed* if for every sequence $\{x_n\}$ such that (i) $x_n \in S$ for all n and (ii) $x_n \to x$, we have $x \in S$.
(i.e. S is closed iff every convergent sequence in S has its limit point in S)

- **Theorem:** Suppose $\{x_n\}$, $\{y_n\}$ are sequences in $\mathbb{R}^k$, $\{\beta_n\}$ is a sequence in $\mathbb{R}$ and $x_n \to x$, $y_n \to y$, $\beta_n \to \beta$. Then:

  (a) $\lim_{n\to\infty}(x_n + y_n) = x + y$

  (b) $\lim_{n\to\infty}(x_n y_n) = xy$

  (c) $\lim_{n\to\infty}(\beta_n x_n) = \beta x$

- proof: Rudin, p. 51

- **Definition (subsequence):** Given a sequence $\{p_n\}$, consider a sequence $\{n_k\}$ of positive integers for which $n_1 < n_2 < \cdots$. The sequence $\{p_{n_k}\}$ is called a *subsequence* of $\{p_n\}$.

- **Theorem:** (proof, Rudin p52)
  - a sequence $\{p_n\}$ converges to p iff every subsequence of $\{p_n\}$ converges to p.
  - if $\{p_n\}$ is a sequence in a compact metric space X, then there exists a subsequence of $\{p_n\}$ that converges to a point in X.
  - (Bolzano-Weierstrass Theorem) every bounded sequence in $\mathbb{R}^k$ contains a convergent subsequence.

1.3 Limit, continuity and monotonicity

- **Definition (limit of a function):** Let X and Y be metric spaces, $E \subseteq X$, and $f : E \to Y$. We write $f(x) \to q$ as $x \to p$ or $\lim_{x\to p} f(x) = q$ if $\exists q \in Y$ such that $\forall \epsilon > 0$, there exists $\delta > 0$ such that $d(f(x), q) < \epsilon$ for all $x \in E$ for which $d(x, p) < \delta$.

- **Properties:** Let $\lim_{x\to p} f(x) = A$ and $\lim_{x\to p} g(x) = B$, then:
  - $\lim_{x\to p} f(x) + g(x) = A + B$
  - $\lim_{x\to p} f(x)g(x) = AB$
  - $\lim_{x\to p} \frac{f(x)}{g(x)} = \frac{A}{B}$ if $B \neq 0$.

- **Definition (continuous function):** Suppose X and Y are metric spaces, $E \subseteq X$, $p \in E$ and $f : E \to Y$. The function f is said to be *continuous* at p if $\forall \epsilon > 0$, $\exists \delta > 0$ such that $d(f(x), f(p)) < \epsilon$ for all $x \in E$ for which $d(x, p) < \delta$.
  - in other words, for a continuous function f, the limit of $f(x)$ as $x \to p$ **exists** and **equals** $f(p)$ (not something else). Draw example.

- **Definition (continuity):** The function $f : A \to B$ is continuous if $\lim_{n\to\infty} f(x_n) = f(x)$ for each convergent sequence $\{x_n\}$ in A with limit $x \in A$.
  - in simple words: the limit must exist and must equal the same value, $f(x)$ for all such sequences.
  - if a function is not continuous at a point we say it is *discontinuous* at that point.

- **Result 1:** If $f(x)$ is continuous at p and $g(x)$ is continuous at $f(p)$, then the function $h(x) \equiv g(f(x))$ is continuous at p. (Rudin, p.86).

- **Result 2:** Let f and g be continuous functions on the metric space X. Then $(f+g)$ and fg are continuous functions on X; and f/g is continuous on X whenever $g \neq 0$.

- **Definition (monotonicity):** Let f be a real function on (a,b). Then f is said to be monotonically increasing on (a,b) if $a < x < y < b$ implies $f(x) \leq f(y)$.
  f is said to be monotonically decreasing if $a < x < y < b$ implies $f(x) \geq f(y)$.

2 Calculus

2.1 Differentiation

We start with univariate functions, $f : \mathbb{R} \to \mathbb{R}$, then move on to multivariate functions, $f : \mathbb{R}^n \to \mathbb{R}$, and finally to vector-valued functions, $f : \mathbb{R}^n \to \mathbb{R}^k$.

2.1.1 Univariate functions

(a) **Derivative of a function:**

Let f be a function defined on $[a,b] \subseteq \mathbb{R}$. For any $x \in [a,b]$, form the ratio:

$$\phi(t) = \frac{f(t) - f(x)}{t - x}, \quad a < t < b, \ t \neq x$$

and define

$$f'(x) \equiv \lim_{t \to x} \phi(t)$$

If this limit **exists**, it is called the (first) derivative of the function f.

- basically, we associate with f another function, denoted f'; its domain is the set of points on which the above limit exists.

- **Definition (differentiable function)**

  (i) If f' is defined at some point x, we say that f is differentiable at x.

  (ii) If f' is defined at all $x \in E$, we say that f is differentiable on E.

- **visual interpretation:** the derivative, $f'(x)$ can be interpreted as the slope of f at point x (use the definition and draw a picture).

(b) **Differentiability and continuity**:

Remember the definition of a continuous function (here applied to $\mathbb{R}$)

- **Definition (continuity)** (using $d(x, y) = |x - y|$)
  The function $f : \mathbb{R} \to \mathbb{R}$ is continuous at x if $\forall \epsilon > 0$, $\exists \delta > 0$ such that $|f(y) - f(x)| < \epsilon$, whenever $|y - x| < \delta$.

  The following important theorem reveals the connection between differentiability and continuity.

- **Theorem 1:** Consider a function f and a point x in its domain. Differentiability of f at x is a sufficient but not necessary condition for continuity at x.

  - Thus, differentiability implies continuity but not the other way around. (counterexamples: kinks)

- **Definition (continuously differentiable function):** If f' is a continuous function on $[a, b]$ then f is called continuously differentiable, denoted C^1.

- **Theorem 2 (rules of differentiation)**

  Suppose f and g are defined on $[a, b]$ and are both differentiable at some $x \in [a, b]$. Then, the functions $(f + g)$, fg, and f/g are also differentiable at x and:

  (a) $(f + g)'(x) = f'(x) + g'(x)$
  (b) $(fg)'(x) = f'(x)g(x) + f(x)g'(x)$
  (c) $(f/g)'(x) = \frac{g(x)f'(x) - f(x)g'(x)}{g^2(x)}$, $g(x) \neq 0$

- **Theorem 3 (the "chain rule")**

  Suppose f is continuous on $[a, b]$, its derivative f' exists at some $x \in [a, b]$, the function g is defined on an interval I containing the range of f, and g is differentiable at $f(x)$. If $h(t) \equiv g(f(t))$, $(a \leq t \leq b)$, then h is differentiable at x and $h'(x) = g'(f(x))f'(x)$.

- **Derivative of inverse function**

  Let $f(x)$ be a one-to-one function (e.g., strictly monotonic). Then, if $f^{-1}(y)$ is the inverse of $y = f(x)$ and we assume f^{-1} is differentiable, we have that:

$$(f^{-1}(y))' = \frac{1}{f'(x)}$$

2.1.2 Applications of derivatives

- **Definition (local maximum)**
  Let f be a real function defined on the interval X. We say that f has a local maximum at $x \in X$ if there exists $\delta > 0$ such that $f(y) \leq f(x)$ for all $y \in X$ with $|y - x| < \delta$.

- **Theorem 6 (first-order condition for a maximum)**
  Let f be defined on the interval $[a, b]$. If f has a local maximum at $x \in (a, b)$ and if $f'(x)$ exists, then $f'(x) = 0$.

  The above theorem provides a necessary condition for a function to have a local maximum at a given (interior) point: its derivative at the point must be equal to zero. Note: the derivative being zero is not a sufficient condition for a maximum! (it can correspond to a minimum, to, or to an inflection point).

- **Theorem 7 (Derivative signs)**
  Suppose f is differentiable in (a, b). Then:

  (a) If $f'(x) \geq 0$ for all $x \in (a, b)$, then f is monotonically increasing.

  (b) If $f'(x) = 0$ for all $x \in (a, b)$, then f is constant.

  (c) If $f'(x) \leq 0$ for all $x \in (a, b)$, then f is monotonically decreasing.

  Obviously, the above theorem is very useful when doing comparative statics.
  The next theorem is useful when we have a function of the form $f(x) = m(x)/n(x)$ and we are interested in $\lim_{x \to x_0} f(x) = \lim_{x \to x_0} m(x) / \lim_{x \to x_0} n(x)$ and this limit tends to $0/0$, or $\pm\infty/\pm\infty$.

- **Theorem 9 ("L'Hopital's rule")**
  Suppose $m(x)$ and $n(x)$ are real, differentiable functions in (a, b) and $n'(x) \neq 0$ for all $x \in (a, b)$ where $-\infty \leq a \leq b \leq \infty$. Suppose also that $m'(x)/n'(x) \to C$ as $x \to c$.
  If $m(x) \to 0$ and $n(x) \to 0$ as $x \to a$ or $m(x)$ and $n(x) \to \pm\infty$ as $x \to c$, then $m(x)/n(x) \to C$ as $x \to c$.

- **Example:** use L'Hopital's rule to find the limits:
  (i) $\lim_{x \to 0} \frac{e^x - 1}{x}$; (ii) $\lim_{x \to 0} \frac{e^x - 1 - x}{x^2}$.

- **Definition (Higher-order derivatives):** If f has a derivative f' on an interval on $\mathbb{R}$ and if f' is itself differentiable, we denote the derivative of f' by f'' and call it the second derivative of f. Proceeding in the same way, $f^{(n)}$ is called the n-th order derivative of f, where $f^{(n)} = (f^{(n-1)})'$. If f'' is continuous, we say that f is twice continuously differentiable (denoted C^2).

- **Differential:**
  Take a small number Δx. Then we have the following approximation:

  $$\frac{f(x_0 + \Delta x) - f(x_0)}{\Delta x} = \frac{\Delta y}{\Delta x} \sim f'(x_0)$$

  or, $f(x_0 + \Delta x) - f(x_0) \sim f'(x_0)\Delta x$.

- **Definition:** The differential of f at x_0 is $dy \equiv f'(x_0)\Delta x$ (or often written as $dy = f'(x)dx$).

  The meaning of the differential is the change in the functional value resulting from a small change in the argument.

- **Taylor series approximation**
  We showed above that $f(x_0 + \Delta x) \sim f(x_0) + f'(x_0)\Delta x$. This is called the first-order Taylor approximation of f around x_0. This approximation idea can be generalized by the following theorem.

- **Theorem 10 (Taylor)** Suppose f is a real function on $[a, b]$, n is a positive integer, the functions $f^{(k)}$, $k = 0, 1, \cdots, n-1$ are continuous on $[a, b]$, and $f^{(n)}(t)$ exists for all $t \in (a, b)$.
  Let α, x be distinct points in $[a, b]$ and define

$$P(x) = \sum_{k=0}^{n-1} \frac{f^{(k)}(\alpha)}{k!}(x - \alpha)^k$$

  Then, there exists $\beta \in (\alpha, x)$ such that $f(x) = P(x) + \frac{f^{(n)}(\beta)}{n!}(x - \alpha)^n$.

  Note: above $f^{(0)}(\alpha) = f(\alpha)$ and $k!$ denotes "k factorial" defined as $0! = 1$ and $k! = k(k-1)!$ for any $k \in \mathbb{N}$.

- **Remarks:**

- $P(x)$ is called the *Taylor series approximation* (Taylor expansion) of $f(x)$ around α.

- if the approximation is around $\alpha = 0$, the resulting formula is called the Maclaurin expansion of $f(x)$.

- for $n = 1$, the above gives the so-called Mean Value Theorem, since $P(x) = f(\alpha)$.

- **Theorem (the "mean value theorem"):** If f is a real continuous function on $[a, b]$ and differentiable on (a, b), then there exists a point $x \in (a, b)$ at which:

$$f(b) - f(a) = (b - a)f'(x)$$

- geometric interpretation: there exists a point x between a and b at which the slope of the function equals the slope of the straight line through the points with coordinates $(a, f(a))$ and $(b, f(b))$.

- **Exercises:**

– take $f(x) = (1+x)^3 = 1 + 3x + 3x^2 + x^3$. Then $f(0) = 1$; $f'(0) = 3$; $f''(0) = 6$; $f'''(0) = 6$. Use Taylor's formula with $\alpha = 0$ to show: $f(x) \sim 1 + 3x + 1/2!6x^2$ (2nd order Taylor expansion of $(1+x)^3$ around 0) and $f(x) \sim 1 + 3x$ (1st order Taylor expansion of $(1+x)^3$ around 0).

– For $f(x) = e^x$, we have $f^{(k)}(x) = e^x$ so $f^{(k)}(0) = 1$ for all k. Thus,

$$e^x \sim \sum_{k=0}^{n-1} \frac{1}{k!} x^k$$

(use Taylor's formula at $x_0 = 0$ and $e^0 = 1$). Plug in $x = 1$ to obtain a formula for $e = 1 + 1 + 1/2 + 1/6 + \cdots$.

2.2 Multivariate functions

Now, we look at multivariate functions $f : \mathbb{R}^n \to \mathbb{R}$, that is, such that $x = (x_1, x_2, \cdots, x_n) \in \mathbb{R}^n$ and $y = f(x) \in \mathbb{R}$ (real-valued functions of n arguments).

- **Level curves:** Assume we have the function $f(x_1, x_2) : \mathbb{R}^2 \to \mathbb{R}$. We can fix the value of $f(x_1, x_2)$ to some c and plot a graph in (x_1, x_2) space. This is called a level curve of f. Formally, the level curve of f for value c is the set $\{x \in \mathbb{R}^n : f(x) = c\}$. It has many applications in economic theory, e.g., isoquants or indifference curves.

  – simplex: The level curve defined as $c = x_1 + x_2 + \cdots + x_n$ for a given $c \in \mathbb{R}$ is called a simplex in $\mathbb{R}^n$.

2.2.1 Partial and total differentiation

- **Partial derivatives**

  Suppose we want to see how f changes if **only one** of its arguments changes by a small amount, *holding all other arguments constant.* As before, form the ratio:

$$\frac{f(x_1^0, x_2^0, \cdots, x_i^0 + \Delta x_i, \cdots, x_n^0) - f(x_1^0, x_2^0, \cdots, x_i^0, \cdots, x_n^0)}{\Delta x_i}$$

  If the limit of this ratio exists as $\Delta x_i \to 0$, we call,

$$\frac{\partial f(x^0)}{\partial x_i} \equiv \lim_{\Delta x_i \to 0} \frac{f(x_1^0, x_2^0, \cdots, x_i^0 + \Delta x_i, \cdots, x_n^0) - f(x_1^0, x_2^0, \cdots, x_i^0, \cdots, x_n^0)}{\Delta x_i}$$

  the *partial derivative* of f at x^0 with respect to x_i. It is read "dee" f, "dee" x_i.

- **Partial differential:**

  Similarly to the univariate case, we call $df_{x_i}(x^0) \equiv \frac{\partial f(x^0)}{\partial x_i} \Delta x_i$, the partial differential of f at x^0 with respect to x_i. Its interpretation is analogous to before.

- **Total differentiation:**
  Consider $y = f(x_1, \cdots, x_n)$ and suppose we want to know how f changes as we vary all its arguments $x_1, \cdots, x_n$ by small amounts, Δx_i $(i = 1, \cdots, n)$. We have:
  $$\Delta y = f(x_1 + \Delta x_1, \cdots, x_n + \Delta x_n) - f(x_1, \cdots, x_n) = \sum_{i=1}^{n} \frac{\Delta y}{\Delta x_i} \Delta x_i$$
  Taking limits as $\Delta x_i \to 0$, we obtain that the change in the functional value is:
  $$df(x) = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i} dx_i$$
  which is called the total differential of f, i.e., the *total change* in its value resulting from small changes in *all arguments*.

- **Gradient of f:**
  We can stack the n partial derivatives of f in the $n \times 1$ vector (column vector of size n):
  $$\nabla f \equiv \frac{\partial f}{\partial x} \equiv \left[\frac{\partial f}{\partial x_1} \; \frac{\partial f}{\partial x_2} \; \cdots \; \frac{\partial f}{\partial x_n} \right]^T$$
  - Note: here we use superscript T to denote transpose of a vector or matrix so that we do not confuse it with the derivative notation. The convention is that x is taken as the column vector of size n (and x' or x^T is the row vector).
  - The elements of the gradient indicate in which direction f moves as we change its respective arguments. The gradient is useful for finding maxima (minima). If some of its elements are positive (negative), then we know that we must move in that direction (increase/decrease that x_i) to reach maximum (minimum). At an (interior) extremum (or critical point) of f it must therefore be true that $\nabla f = 0_{n,1}$.

- **Theorem 12**
  Let $f : \mathbb{R}^n \to \mathbb{R}$ be continuously differentiable. At any point x^0 in the domain of f where $\nabla f(x^0) \neq 0$, the gradient evaluated at x^0 points at the direction in which f increases most rapidly.
  - Note that we can express the total differential of f at x^0 in terms of the gradient:
  $$df(x^0) = \left(\nabla f(x^0)\right)^T dx \qquad \text{where } dx = [dx_1, \cdots, dx_n]^T$$

2.2.2 Higher-order derivatives of multivariate functions

Let f be C^2 (twice continuously differentiable). Then, we can compute
$$\frac{\partial}{\partial x_i} \frac{\partial f}{\partial x_j} \equiv \frac{\partial^2 f}{\partial x_i \partial x_j}$$
the second (cross-) partial derivatives of f for any $i, j = 1, \cdots, n$.

- **Theorem 13 (Young)**
  If f is C^2, we have:
  $$\frac{\partial}{\partial x_i}\frac{\partial f}{\partial x_j} = \frac{\partial}{\partial x_j}\frac{\partial f}{\partial x_i} \qquad i, j = 1, \cdots, n$$
  that is, the order of differentiation does not matter.

  – Young's theorem is useful for computing covariance matrices.

- **Definition (the "Hessian matrix"):**
  The matrix of all second partial derivatives of f:
  $$H(.)_{n\times n} \equiv \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \cdots & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix}$$
  is called the Hessian. It is a symmetric matrix (Why?).

- **Taylor series expansions in $\mathbb{R}^n$:**
  As in the univariate case, we can expand f around a given point $x^0 \in \mathbb{R}^n$. The following are the most frequently used results:

- First-order Taylor series expansion:

$$f(x) \equiv f(x^0) + (\nabla f(x^0))^T (x - x^0)$$

- Second-order Taylor series expansion:

$$f(x) \equiv f(x^0) + (\nabla f(x^0))^T (x - x^0) + \frac{1}{2}(x - x^0)^T H(x^0)(x - x^0)$$

2.2.3 Implicit Functions

Sometimes it is not possible to have a closed form solution for $y = f(x)$. For example, suppose all we know is that y and x satisfy:

$$y^2 x_1 + 2y x_1 x_2 + \sqrt{y} x_2 + 3 = 0$$

We cannot solve for y as an explicit function of $x = (x_1, x_2)$.

- in general, an implicit function is given in the form:

  $$F(y, x_1, \cdots, x_n) = 0$$

  which, totally differentiating both sides, implies

  $$\frac{\partial F}{\partial y}dy + \frac{\partial F}{\partial x_1}dx_1 + \cdots + \frac{\partial F}{\partial x_n}dx_n = 0 \quad (*)$$

- setting all but one (dx_i) of the dx's to zero, we obtain the so-called *Implicit Function Rule*:

$$\frac{\partial y}{\partial x_i} = -\frac{\frac{\partial F}{\partial x_i}}{\frac{\partial F}{\partial y}}$$

which allows us to find the partial derivatives of y and hence study its properties, even without knowing the function itself. In order for the implicit function $y = f(x)$ to exist and the above to hold, however, we need some conditions, as given below.

- **Theorem 14 (the "Implicit function theorem")**
If $F(y, x_1, \cdots, x_n) = 0$ and if
(i) F has continuous partial derivatives $\frac{\partial F}{\partial y}$, $\frac{\partial F}{\partial x_i}$, $i = 1, \cdots, n$.
(ii) $\frac{\partial F(y^0, x^0)}{\partial y} \neq 0$ at some (y^0, x^0) for which $F(y^0, x^0) = 0$ [note: x^0 is vector here!]
Then there exists $N_\epsilon(x^0)$ such that $y = f(x)$ for all $x \in N_\epsilon(x^0)$ and f is C^1.

- **Application: level curves**
Consider the level curve for $f(x) = b$ written as implicit function: $F(b, x_1, \cdots, x_n) = 0$. From (*), we have:

$$(\nabla F)^T dx = 0$$

that is, to stay on the level curve, the inner product of the gradient of F with any vector of changes in x must be zero.

2.3 Vector-valued functions

Let $x \in \mathbb{R}^n$ and consider the functions $f_1(x), f_2(x), \cdots, f_m(x)$ each from $\mathbb{R}^n$ to $\mathbb{R}$. This can be written more compactly as: $f : \mathbb{R}^n \to \mathbb{R}^m$, where $x \in \mathbb{R}^n$ and $f(x) \in \mathbb{R}^m$, or,

$$f(x) = [f_1(x_1, \cdots, x_n), \cdots, f_m(x_1, \cdots, x_n)]^T$$

Such function f is called a vector-valued function.

2.3.1 The Jacobian matrix

If $f_1(.), f_2(.), \cdots, f_m(.)$ are differentiable, we can write the total differential of the vector-valued function $f : \mathbb{R}^n \to \mathbb{R}^m$ as:

$$\begin{bmatrix} df_1 \\ \cdots \\ df_m \end{bmatrix} = \begin{bmatrix} (\nabla f_1)^T dx \\ \cdots \\ (\nabla f_m)^T dx \end{bmatrix}$$

or, in matrix notation,

$$df_{m\times 1} = J_{m\times n} dx_{n\times 1}$$

- **Definition (Jacobian matrix):** The m-by-n matrix J consisting of the transposed gradients of $f_1, \cdots, f_m$ is called the Jacobian of f.

- The first-order Taylor series expansion for a vector valued function f takes the form:

$$f(x) \sim f(x^0) + J(x^0)(x - x^0)$$

2.3.2 Derivatives of matrix functions (FOR REFERENCE ONLY)

Let a be $(n, 1)$ vector (column-vector or simply vector) and A be a (m, n) matrix. Let also x be a vector of size n. We then have the following rules for differentiation of matrix (vector) functions:

- **Rule 1:**
$$\frac{\partial(a^T x)}{\partial x} = a$$

- **Rule 2:**
$$\frac{\partial(Ax)}{\partial x} = A^T$$

- **Rule 3:** Let A be a square matrix. Then,
$$\frac{\partial(x'Ax)}{\partial x} = (A + A^T)x$$
Special case with A symmetric:
$$\frac{\partial(x'Ax)}{\partial x} = 2Ax$$

- **Rule 4:** Let A be a square matrix. Then,
$$\frac{\partial(x'Ax)}{\partial A} = xx^T$$

In the above rules you take the partial derivatives of the respective function with respect to each element of x or A and then stack those into a vector or matrix of the same size as x or A respectively.

2.4 Homogenous functions

We now introduce some particular types of functions, which are often used in economics because of certain "nice" properties they have.

- **Definition (homogeneous function):** Let $f : \mathbb{R}^n \to \mathbb{R}$. We say that f is homogeneous of degree k if $\forall t \in \mathbb{R}_+$, we have $f(tx) = t^k f(x)$.

- **examples:** a linear function $y = ax$ is homogenous of degree 1 while a quadratic form, $y = x^T Ax$ is homogenous of degree 2.

- **Properties:** Consider the function $f : \mathbb{R}^n \to \mathbb{R}$ with f continuously differentiable and homogeneous of degree k. Then f has the following properties:

  1. The partial derivatives of f, $\partial f/\partial x_i$ are homogeneous of degree $(k-1)$.
     Hint: differentiate both sides of $f(tx) = t^k f(x)$ w.r.t. x_i, to get $t\partial f(tx)/\partial x_i = t^k \partial f(x)/\partial x_i$.
  2. Euler's theorem:
     $$\frac{\partial f(x)}{\partial x_1}x_1 + \frac{\partial f(x)}{\partial x_2}x_2 + \cdots + \frac{\partial f(x)}{\partial x_n} = kf(x)$$
     Hint: differentiate both sides of $f(tx) = t^k f(x)$ w.r.t. to t and evaluate at $t = 1$.
  3. The tangent hyperplanes to the level sets of f have constant slope along any ray from the origin.
     Example: indifference curves for Cobb-Douglas preferences.

2.4.1 Homothetic functions

Homothetic functions are also very often used in economics: they represent a generalization of homogeneous functions.

- **Definition (Homothetic function):** A function $h : \mathbb{R}^n \to \mathbb{R}$ is homothetic if it is a positive monotone transformation (PMT) of a homogeneous function, that is, we can write $h(x) = g(f(x))$ where g is monotonically increasing (as defined earlier) and f is homogeneous.

  homogeneity $\Rightarrow$ homotheticity but the opposite is not true in general.

- **Examples:** $h(x) = e^{ax}$ or $h(x) = ln(ax)$

- Homothetic functions frequently used in economics:

1. The Cobb-Douglas function: $f(x_1, x_2) = Ax_1^\alpha x_2^\beta$ for $A, \alpha, \beta > 0$. It has the following properties:

   (i) f is homogeneous of degree $(\alpha + \beta)$.

   (ii) ("returns to scale"):

   * If $\alpha+\beta > 1$, we say that f exhibits increasing returns to scale (since $f(\lambda x_1, \lambda x_2) > \lambda f(x_1, x_2)$, i.e. if we double the inputs, output more than doubles!).

* If $\alpha + \beta = 1$, f exhibits constant returns to scale.
* If $\alpha + \beta < 1$, f exhibits decreasing returns to scale.

(iii) Notice that $\alpha = \partial f/\partial x_1 \times x_1/f$ and $\beta = \partial f/\partial x_2 \times x_2/f$. This can be interpreted in the following way. Imagine that f is a production function with x_1 being capital, K and x_2 being labor, L. Then, since in equilibrium input prices must equal their marginal products, i.e. $r = \partial f/\partial x_1$ and $w = \partial f/\partial x_2$, with r is the rental rate on capital and w is the wage rate. Thus $\alpha = rK/f$ is the share of capital in output and $\beta = wL/f$ is the labor's share.

(iv) The function $\ln(f) = \ln A + \alpha \ln x_1 + \beta \ln x_2$ is homothetic as it is a PMT of $f(x)$.

2. The CES function: $f(x_1, x_2) = A(\delta x_1^\rho + (1-\delta)x_2^\rho)^{1/\rho}$ for $A > 0$, $\rho \leq 1$. It has the following properties:

   (i) f is linearly homogeneous (degree 1).

   (ii) the Cobb-Douglas function is a special case of the CES function when $\rho \to 0$ (use L'Hopital's rule to show it!).

2.5 Concavity and convexity of a function

- **Definition (concave function):** Let $f : \mathbb{R}^n \to \mathbb{R}$. We say that f is concave if $\forall x, y$ and $\forall \lambda \in [0, 1]$, we have:

$$f(\lambda x + (1-\lambda)y) \geq \lambda f(x) + (1-\lambda) f(y)$$

- A function f is **convex** if $-f$ is concave, i.e., if the sign in the above inequality is $\leq$.

The next few theorems describe important properties of concave functions.

- **Theorem 15**
  Let $f_i : \mathbb{R}^n \to \mathbb{R}$, $i = 1, \cdots, k$ be concave functions. Then, with $\alpha_i \in \mathbb{R}_+$, $i = 1, \cdots, n$,

$$f(x) = \sum_{i=1}^{k} \alpha_i f_i(x)$$

  is also concave.

- **Theorem 16**
  Let $f : \mathbb{R}^n \to \mathbb{R}$. Then the following statements are equivalent:

  (a) f is concave.

  (b) $f(\lambda_1 x^1 + \cdots + \lambda_n x^n) \geq \lambda_1 f(x^1) + \cdots + \lambda_n f(x^n)$, $\forall x^i$ and all $\lambda_i \in [0, 1]$ with $\sum_i \lambda_i = 1$ (this generalizes the definition).

(c) The set $\{(x, \alpha) \in \mathbb{R}^{n+1} \ : \ \alpha \in \mathbb{R}, f(x) \geq \alpha\}$ is convex.

(d) $\forall x, y$ we have $f(y) - f(x) \leq (\nabla f(x))^T (y - x)$.

(e) (when defined) the Hessian $H(x)$ of $f(x)$ is negative semi-definite for all x.

The theorem gives different ways of proving that a function is concave.

- **Example (Insurance)** Suppose we have a utility maximizing consumer with utility function over wealth u satisfying $u' > 0$ and $u'' < 0$. Let there be two possible outcomes for the agent's wealth x: a bad one with wealth x_b and a good one $x_g > x_b$, which occur with probabilities p_b and $p_g(= 1 - p_b)$, respectively. Assume that the consumer is an expected utility maximizer, that is, she maximizes the function:

$$u(x) = p_b u(x_b) + p_g u(x_g)$$

From the strict concavity of u we know that:

$$u(p_b x_b + p_g x_g) > p_b u(x_b) + p_g u(x_g)$$

i.e., $u(E(x)) > E(u(x))$ (with $E(x)$ the expected value of r.v. x) thus the consumer strictly prefers to obtain the expected outcome, $p_b x_b + p_g x_g$ for sure than being subjected to the risk. This implies that she would like to give up some consumption in order to be insured against the risk, i.e. she would prefer some

$$\tilde{x} < p_b x_b + p_g x_g$$

offered by an insurance company to the risky outcome without insurance.

2.6 Quasi-concavity and Quasi-convexity

- **Definition (quasi-concave function):** Let $f : \mathbb{R}^n \to \mathbb{R}$. We say that f is quasi-concave if $\forall x, y$ and all $\lambda \in [0, 1]$, we have:

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

- **Definition (quasi-convex function):** A function f is quasi-convex if $-f$ is quasi-concave.

- **Theorem 18**
Consider $f : \mathbb{R}^n \to \mathbb{R}$. The following statements are equivalent:

(a) $\{x : f(x) \geq a\}, \forall a \in \mathbb{R}$ is a convex set (this set is called upper contour set)

(b) f is quasi-concave.

- Notes:

* Observe the difference between part (a) of Theorem 18 and part (c) of Theorem 16. It is clear that concavity implies quasi-concavity but not vice-versa. For example take $f(x) = x^2$, with $U = \{x \in \mathbb{R} : x \geq 0\}$. Clearly f is quasi-concave, since for any $x > y$, $(\lambda x + (1 - \lambda)y)^2 > y^2$, but it is convex, not concave.

* In general, it is easy to show that any monotonically increasing or decreasing univariate function ($n = 1$ above) is quasi-concave.