# Multivariate Calculus

Ichiro Obara

UCLA

September 13, 2019

# Function of Many Variables

We introduce functions that map a point/vector in $\Re^n$ to another point in $\Re^m$, and generalize the notions of the derivative etc. to such functions. For a function $f : \Re^n \to \Re^m$, we denote the $i$th component of function by $f_i(\mathbf{x})$ for $i = 1, ..., m$.

Most function used in Economics have more than one variable in them. The first such function you will see is a utility function $u : \Re^n \to \Re$, which represents an individual's preference over the consumption of $n$ goods.

Here are some other functions:

- **Production Function:** $F(K, L)$, where $K$ is capital and $L$ is labor.
- **Demand Function:** $x(p, w)$, $p \in \Re^n$ is a price vector and $w$ is wealth.
- **Cost/Expenditure Function:** $e(p, u)$ is the minimum amount of expenditure to achieve utility $u$ when the price vector is $p$.

# Special Functions

The following special functions are especially important.

- **Linear Function**: $f(\mathbf{x}) = \sum_{i=1}^{n} a_i x_i$ is a linear function, which maps a point in $\Re^n$ to $\Re$.
- **Quadratic Function**: The function of the form $f(\mathbf{x}) = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} x_i x_j$ is called **quadratic function**. We can represent this using $n \times n$ matrix $A$ as $f(\mathbf{x}) = \mathbf{x}^\top A\mathbf{x}$, where $a_{ij}$ is the $ij$ element of $A$. This is a generalization of quadratic function with single variable: $ax^2$.

# Sequence, Convergence and Limit

We extend some definitions regarding sequences and convergence to $\Re^n$ and review some basic facts.

- A **sequence** of vectors $\mathbf{x}_1, \mathbf{x}_2, \ldots \in \Re^n$ is denoted by $\{\mathbf{x}_n\}_n$.
- A sequence $\{\mathbf{x}_n\}_n$ is **bounded** if there exists a number $K \in \Re$ such that $\|\mathbf{x}_n\| \leq K$ for every $n$.
- A sequence $\{\mathbf{x}_n\}_n$ **converges to** $\mathbf{x}^* \in \Re^n$ if, for any $\epsilon > 0$, there exists an integer $N$ such that $\|\mathbf{x}_n - \mathbf{x}^*\| < \epsilon$ for every $n \geq N$. We write this as $\lim_{n\to\infty} \mathbf{x}_n = \mathbf{x}^*$ or $\mathbf{x}_n \to \mathbf{x}^*$ and call $\mathbf{x}^*$ the **limit** of $\{\mathbf{x}_n\}_n$.

## Some Useful Properties

- A sequence of vectors $\mathbf{x}_n$ converges to $\mathbf{x}^*$ if and only if $x_{i,n}$ converges to $x_i^*$ for every $i = 1, ..., n$.
- A convergent sequence has only one limit.
- A convergent sequence is bounded.
- If $\|\mathbf{x}_n\| \leq K$ for every $n$ for a convergent sequence, then $\|\mathbf{x}^*\| \leq K$.
- Every subsequence of a convergent sequence has the same limit.
- Every bounded sequence has a convergent subsequence (Weierstrass Theorem).

# Open Set and Closed Set

In $\Re^n$, we use open sets and closed sets instead of open/closed intervals.

- A set $X$ in $\Re^n$ is **open** if, for any $\mathbf{x} \in X$, there exists $\epsilon > 0$ such that $B_\epsilon(\mathbf{x}) \subset X$ ($B_\epsilon(\mathbf{x}) = \{\mathbf{x}' \in \Re^n | \, \|\mathbf{x}' - \mathbf{x}\| < \epsilon\}$ is $\epsilon$-**ball** around $\mathbf{x}$).
- A set $X$ in $\Re^n$ is **closed** if every convergent sequence $\{\mathbf{x}_n\}_n$ in $X$ has its limit in $X$.
- A set in $\Re^n$ is closed if and only if its complement is open.

# Compact Set

A set $X \in \Re^n$ is **compact** if it is closed and bounded.

We will use compactness to establish the existence of maximum and minimum of continuous functions on compact sets.

# Exercises

1. Is $\Re^n$ an open set? Is it a closed set?
2. Let $\{\mathbf{x}_n\}_n$ be a convergent sequence in $\Re^n$ with the limit $\mathbf{x}^*$. Is $\{\mathbf{x}_n\}_n$ a compact set? How about if $\mathbf{x}^*$ is added to $\{\mathbf{x}_n\}_n$?

# Continuous Function

Given all the generalized definitions about convergence, the definition of continuity is the same.

## Continuous Function

Function $f : X \to \Re^m$ is continuous at $\mathbf{x} \in X \subset \Re^n$ if, for any sequence $\{\mathbf{x}_n\}_n$ in $X$ that converges to $\mathbf{x} \in \Re^n$, $f(\mathbf{x}_n)$ converges to $f(\mathbf{x}) \in \Re^m$. $f$ is continuous if it is continuous at every $x$ in its domain $X$.

# Monotonic Function

Let's also define monotonicity in $\Re^n$.

## Monotonic Function

Function $f : X \to \Re^m$ is **increasing (strictly increasing)** if $f(x') \geq (>) f(x)$ for any $x' \geq (>) x$, **decreasing (strictly decreasing)** if $f(x') \leq (<) f(x)$ for any $x' \geq (>) x$.

# Partial Derivative

Let $f$ be a function from $\Re^n$ to $\Re$. Fix $\mathbf{x}$ except for $x_j$ and regard it as a function of $x_j$, which we denote by $f(x_j, \mathbf{x}_{-j})$. Since it is a function of single variable, we can define the derivative with respect to $x_j$ given $\mathbf{x}_{-j}$ as usual.

This derivative, which is the limit of $\frac{f(x_j+\Delta x_j, \mathbf{x}_{-j})-f(x_j, \mathbf{x}_{-j})}{\Delta x_j}$ as $\Delta x_j \to 0$, is called the **partial derivative** of $f$ with respect to $x_j$ at $\mathbf{x} \in \Re^n$ and denoted by $D_j f$. We denote the row vector of partial derivatives $(D_1 f, ..., D_n f)$ by $Df$.

For function from $\Re^n$ to $\Re^m$, let $D_j f_i$ be the partial derivative of the $i$th component function $f_i$ with respect to $x_j$. $DF$ is the $m \times n$ matrix of partial derivatives, where the $i$th row is given by the row vector $Df_i$.

# Higher Derivatives

We say that $f : \Re^n \to \Re^m$ is a $\mathcal{C}_1$ function if $D_j f_i$ exists and continuous for every $i, j$.

$f$ is a $\mathcal{C}_2$ function if $D_j f_i$ is $\mathcal{C}_1$ function for every $i, j$. More generally, $f$ is a $\mathcal{C}_k$ function if $D_j f_i$ is $\mathcal{C}_{k-1}$ function for every $i, j$

The important special case is $\mathcal{C}_2$ function with $m = 1$. For such functions, there are $n \times n$ second order partial derivatives. Let $D^2 f$ be the $n \times n$ matrix whose $ij$ element is $D_j D_i f$. For this class of functions, $D_i D_j f = D_j D_i f$ holds for any $i, j$ (Young's Theorem). Hence $D^2 f$ is a symmetric matrix.

**Examples**

- **Linear Function**: $f(\mathbf{x}) = \sum_{i=1}^{n} a_i x_i$.
  - $Df = (a_1, ..., a_n)$ and $D^2 f$ is a $n \times n$ matrix with all elements $= 0$.
- **Quadratic Function**: $f(\mathbf{x}) = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} x_i x_j = \mathbf{x}^\top A\mathbf{x}$.
  - $Df(\mathbf{x}) = \sum_{j=1}^{n} a_{ij} x_j$ and $D^2 f(\mathbf{x}) = A$.

# Total Derivative

Remember that the derivative of a single variable function is the slope of its linear approximation. Here we generalize this idea to define **total derivative** for functions from $\Re^n$ to $\Re^m$.

First consider a function $f$ with $m = 1$ as a special case. We use a linear function to approximate $f$, so the linear approximation around point $\mathbf{x}^0 \in \Re^n$ would look like $f(\mathbf{x}^0) + \sum_{j=1}^{n} a_j(x_j - x_j^0) = A(\mathbf{x} - \mathbf{x}^0)$. We say that a row vector $A = (a_1, ..., a_n)$ is the **total derivative** of $f$ at $\mathbf{x}^0$ if $\frac{|f(\mathbf{x}^0+\Delta\mathbf{x})-f(\mathbf{x}^0)-A\Delta\mathbf{x}|}{\|\mathbf{x}\|}$ converges to 0 as $\Delta\mathbf{x} \to 0$.

# Total Derivative

Now define the total derivative for $f$ with $m \geq 1$.

## Differentiable Function and Derivative

A function $f : \Re^n \to \Re^m$ is **differentiable** at $\mathbf{x}^0 \in \Re^n$ if there exists a $m \times n$ matrix $A$ such that $\frac{|f(\mathbf{x}^0 + \Delta\mathbf{x}) - f(\mathbf{x}^0) - A\Delta\mathbf{x}|}{\|\Delta\mathbf{x}\|}$ converges to 0 as as $\Delta\mathbf{x} \in \Re^n$ converges to 0. $A$ is the **derivative** of $f$ at $\mathbf{x}$. $f$ is differentiable if it is differentiable at every $\mathbf{x}$.

**Remark on Total Derivative and Partial Derivative:**

- If the total derivative of $f$ exists, then all the partial derivatives exist. However, even if all the partial derivatives exist, the total derivative may not exist.
- If all the partial derivatives exist and continuous (i.e. $f$ is a $\mathcal{C}_1$ function), then the total derivative of $f$ exists and is actually given by $Df$.
- If $f : \Re^n \to \Re$ is a $\mathcal{C}_2$ function, then there exist the second total derivative and it is given by $n \times n$ matrix $D^2 f$.

# Some Useful Rules

**Chain Rule:** For two functions $f : \Re^n \to \Re^m$ and $g : \Re^m \to \Re^k$, let $h : \Re^n \to \Re^k$ by $h(\mathbf{x}) = g\left(f(\mathbf{x})\right)$ be their **composite function**. If $f$ and $g$ are $\mathcal{C}_1$ functions, $h$ is a $\mathcal{C}_1$ function as well and its derivative at $x$ is given by the $k \times n$ matrix $Dg(f(\mathbf{x}))Df(\mathbf{x})$.

**Product Rule:** Let $f : \Re^n \to \Re^m$ and $g : \Re^m \to \Re^k$ be $\mathcal{C}_1$ functions. Then $h = f \cdot g = f^\top g$ is a $\mathcal{C}_1$ function from $\Re^n$ to $\Re^1$ (regard the value of $f$ and $g$ as column vectors). Its derivative is given by $Dh = g^\top Df + f^\top Dg$.

# First/Second Order Approximation

Here we provide a version of Taylor's theorem that evaluates the first/second order approximation for functions with many variables.

Take any $\mathcal{C}_1$ single variable function $f$ and pick some point $x_0$ and $x > x_0$. Since $f(x) = f(x_0) + f'(t)(x - x_0)$ for some $t \in (x_0, x)$ by the mean value theorem, the difference between $f(x)$ and the first order Taylor polynomial is $(f'(t) - f'(x_0))\,(x - x_0)$. This term converges to 0 faster than $x - x_0$ as $f'(t) \to f'(x_0)$ as $x \to x_0$. This means $f(x_0 + \Delta x) = f(x_0) + f'(x_0)\Delta x + o(\Delta x)$

Similarly, let $f$ be a $\mathcal{C}_2$ single variable function. Then the difference between $f(x)$ and the second order Taylor polynomial is $\frac{f''(t) - f''(x_0)}{2}(x - x_0)^2$. So we have $f(x_0 + \Delta x) = f(x_0) + f'(x_0)\Delta x + \frac{f''(x_0)}{2}\left(\Delta x\right)^2 + o(\left(\Delta x\right)^2)$.

These approximation results generalize to functions of many variables. Let $f$ be a function from $\Re^n$ to $\Re^m$.

- If $f$ is a $\mathcal{C}_1$ function, then $f(\mathbf{x}^0 + \Delta\mathbf{x}) = f(\mathbf{x}^0) + Df(\mathbf{x}^0)\Delta\mathbf{x} + o(\|\Delta\mathbf{x}\|)$ at any $\mathbf{x}^0 \in \Re^n$.
- If $f$ is a $\mathcal{C}_2$ function, then $f(\mathbf{x}^0 + \Delta\mathbf{x}) = f(\mathbf{x}^0) + Df(\mathbf{x}^0)\Delta\mathbf{x} + \frac{(\Delta\mathbf{x})^\top D^2 f(\mathbf{x}^0)\Delta\mathbf{x}}{2} + o(\|\Delta\mathbf{x}\|^2)$ at any $\mathbf{x}^0 \in \Re^n$.

where $o(\mathbf{h}) \to 0$ as $\mathbf{h} \to \mathbf{0}$ in $\Re^n$.

# Exercises

1. Find the partial derivative of the following functions with respect to $x$.
   - $f(x,y) = \left(x^{\frac{1}{\rho}} + y^{\frac{1}{\rho}}\right)^{\rho}$ $(\rho \neq 0)$.
   - $f(x,y) = x^{\alpha} g(x,y)^{(1-\alpha)}$.
2. Let $f : \Re^n \to \Re^n$ be a $C_2$ function.
   - Write down the first order approximation of $f$ at $\mathbf{x}^0$.
   - We can use the above approximation to do a high dimensional version of Newton's method. Show why the updating formula is given by $\mathbf{x}_{t+1} = \mathbf{x}_t - \left(D_2 f(\mathbf{x}_t)\right)^{-1} Df(x_t)$.
3. Let $f(\mathbf{x}) = \mathbf{x}^\top A\mathbf{x}$ be a quadratic function. Suppose that $Df(\mathbf{x}^*) = \mathbf{0}$ at $\mathbf{x}^*$ and $A$ is negative definite. Explain why $\mathbf{x}^*$ achieves a maximum

# Implicit Function Theorem

Consider a circle in $\Re^2$ defined by $f(x, a) = x^2 + a^2 = 1$ and the point $(1/\sqrt{2}, 1/\sqrt{2})$ on the circle. $x$ is a function of $a$ in the neighborhood of $(1/\sqrt{2}, 1/\sqrt{2})$. Denote this function by $x(a)$. What is $Dx(a)$ at $a = 1/\sqrt{2}$?

One way to obtain this value is to derive $x(a)$ explicitly and compute the derivative. But this is not always an easy thing to do.

- **Implicit function theorem** allows us to compute this value without deriving $x(a)$ explicitly.
- For this example, just differentiate $f(x(a), a) = 0$ with respect to $a$. Then you get $x'(a) = -\frac{D_a f(x,a)}{D_x f(x,a)}$ by the chain rule. Hence

$$Dx(a)|_{a=1/\sqrt{2}} = -\frac{\sqrt{2}}{\sqrt{2}} = -1.$$

- This would not work if $D_x f(x, a) = 0$ (at $(a, x) = (1, 0)$ for example).

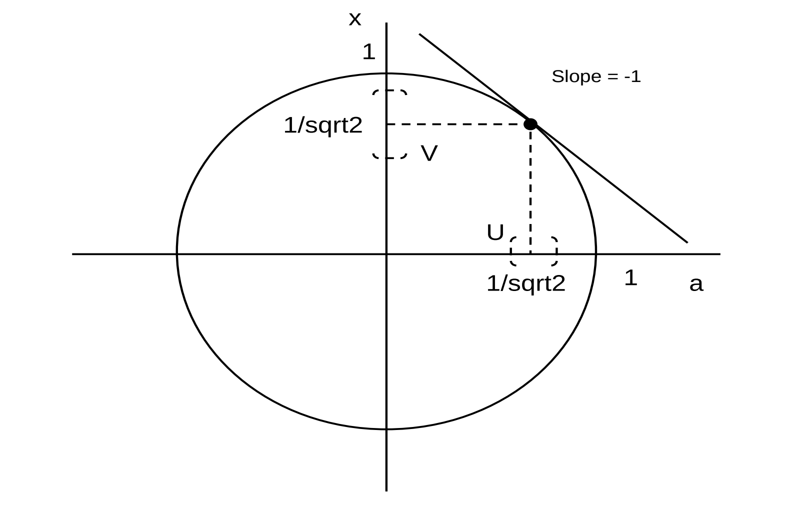
# Implicit Function Theorem

- This result generalizes. Let $F : \Re^n \times \Re^m \to \Re^n$ be a $\mathcal{C}^1$ function.

## Implicit Function Theorem

If $Rank\ D_x F(\mathbf{x}', \mathbf{a}') = n$ at $(\mathbf{x}', \mathbf{a}') \in \Re^n \times \Re^m$, then there exist open neighborhoods $V \subset \Re^n$ of $\mathbf{x}'$, $U \subset \Re^m$ of $\mathbf{a}'$, and a $C^1$ function $f : U \to V$ such that

1. $x = f(a)$ if and only if $F(x, a) = 0$ on $V \times U$, and
2. $Df(\mathbf{a}') = -D_x F(\mathbf{x}', \mathbf{a}')^{-1} D_a F(\mathbf{x}', \mathbf{a}')$.


x
1
Slope = -1
1/sqrt2
V
U
1/sqrt2
1
a

# Exercises

1. Consider a consumer with a utility function $u(x, y)$ for two goods. This consumer's **indifference curve** is implicitly defined by $u(x, y) = \underline{u}$, where $\underline{u}$ is the level of utility.
   - Apply the implicit function theorem to find the **marginal rate of substitution** between $x$ and $y$ ($\frac{dy}{dx}$).
   - Find the linear approximation of the indifference curve at $(x_0, y_0)$ and show that it can be expressed by the hyperplane $D_x u(x_0, y_0)(x - x_0) + D_y u(x_0, y_0)(y_0) = 0$.
   - Find the marginal rate of substitution at $(x_0, y_0)$ for $u(x, y) = \left(x^{\frac{1}{2}} + y^{\frac{1}{2}}\right)^2$. What is the **elasticity of substitution**: $\frac{x}{y}\frac{dy}{dx}$?