Mathematical Economics (ECON 471)
Lecture 3
Calculus of Several Variables & Implicit Functions

Teng Wah Leo

1 Calculus of Several Variables

1.1 Functions Mapping between Euclidean Spaces

Where as in univariate calculus, the function we deal with are such that $f : \mathbf{R}^1 \longrightarrow \mathbf{R}^1$. When dealing with calculus of several variables, we now deal with $f : \mathbf{R}^m \longrightarrow \mathbf{R}^n$, where m and n can be the same or otherwise. What are some instances of this in economics? All you need to do is to consider the production function that used in its barest form labour and capital, and maps it into to range of output values/quantities of an object produced. Yet another example is to consider the vector of goods that is absorbed/consumed by an individual agent that is transformed into felicity via the utility function. Examples such as this abound in economics, and consequently our ground work in Euclidean space.

As in the single variable calculus case, where differentiability is dependent on continuity of the function within the domain. Likewise, this is the case in multivariate calculus. For a function that maps a vector of variables in $\mathbf{R}^m$ onto $\mathbf{R}^n$, let $\mathbf{x}_0$ be a vector in the domain, and let the image under the function f of $\mathbf{x}_0$ be $\mathbf{y} = f(\mathbf{x}_0)$. Then for any sequence of values $\{x_i\}_{i=1}^{\infty}$ in $\mathbf{R}^m$ that converges to $\mathbf{x}_0$, if the sequence of images $\{f(x_i)\}_{i=1}^{\infty}$ likewise converges to $f(\mathbf{x}_0)$ in $\mathbf{R}^n$, we say that the function f is continuous function. Put another way, what this says for the function f to be continuous, it will has to be able to map each of its n components (where each component is denoted as f_i, $i \in \{1, 2, \dots, n\}$) must be able to map from $\mathbf{R}^m \longrightarrow \mathbf{R}^1$. Further, if you have two functions f and g that are independently continuous, then likewise the following operations $f + g$, $f - g$ and f/g still retain their underlying functions continuity.

You may be asking what is a sequence. As a minor deviation, in words, a sequence of real numbers is merely an assignment of real numbers to each natural number (or positive integers). A sequence is typically written as $\{x_1, x_2, \ldots, x_n, \ldots\} = \{x_i\}_{i=1}^{\infty}$, where x_1 is a real number assigned to the natural number 1, ... The three types of sequences are as follows,

1. sequences whose values gets closer and closer relative to their adjacent values as the sequence lengthens,
2. sequences whose values gets farther and farther apart to adjacent values as the sequence lengthens,
3. sequences that exhibits neither of the above, vacillating between those patterns.

For a sequence $\{x_i\}_{i=1}^{\infty}$, we call a value l a limit of this sequence if for a small number ϵ, there is a positive integer N such that for all $i \geq N$, x_i is in the interval generated by ϵ about l. More precisely, $|x_i - l| < \epsilon$. We conclude then that l is the limit of the sequence, and write it as follows,

$$\lim x_i = l \qquad \text{or} \quad \lim_{i\to\infty} x_i = l \qquad \text{or} \quad x_i \to l$$

Note further that each sequence can have only one limit. There of course other ideas. We will bring those definitions on board as and when it becomes a necessity to ease their understanding and context.

2 Implicit Functions, Partial & Total Derivatives

2.1 Partial Derivatives

When dealing with functions in several variables, what do we do to examine the marginal effects the independent variables have on the dependent variable. This now leads us to **partial derivatives**. Technically, the act of differentiating a function remains the same, but as the name suggest, their are some conceptual differences. Consider a function f in two variables x and y, $z = f(x, y)$. When we differentiate the function f with respect to the variable x, we are still finding the effect that x has on z, however we have to hold the variable y constant. That is we treat the second variable as if it were a parameter.

Geometrically, this means we are examining the rate of change of z with respect to x along a values of y. A partial derivative is written as,

$$\begin{aligned}
\text{Given:} \qquad\qquad y &= f(x_1, x_2, \ldots, x_{i-1}, x_i, x_{i+1}, \ldots, x_n) \\
\frac{\partial}{\partial x_i} f(x_1, x_2, \ldots, x_{i-1}, x_i, x_{i+1}, \ldots, x_n) &= f_{x_i}(x_1, x_2, \ldots, x_{i-1}, x_i, x_{i+1}, \ldots, x_n)
\end{aligned}$$

Notice that in taking the derivative, we are no longer using d but the expression ∂ which is usually read as **partial**.

To get a better handle of the process, let's go through some examples.

Example 1 *Let an individual's utility function be U, and it is a function of two goods x_1 and x_2. Let U be a Cobb-Douglas function,*

$$U = Ax_1^{\alpha} x_2^{1-\alpha}$$

Then the marginal utility from the consumption of an additional unit of good 1 is just the partial derivative of the utility function with respect to x_1,

$$\begin{aligned}
\frac{\partial U(x_1, x_2)}{\partial x_1} &= A\alpha x_1^{\alpha-1} x_2^{1-\alpha} \\
&= A\left(\frac{x_2}{x_1}\right)^{1-\alpha}
\end{aligned}$$

Example 2 *An additively separable function is a function made up of several functions which are added one to another. For instance, let x_i, $i \in \{1, 2, \ldots, n\}$ be n differing variables, and let $f(.)$ be a additively separable function of the following form,*

$$f(x) = \sum_{i=1}^{n} h_i(x_i)$$

In truth, the number of sub-functions need not be restricted to being just n, and likewise, each of the sub-functions can be a function of a vector of variables, and not just a corresponding variable of the same sequence. Then the partial derivatives in the above is just,

$$\frac{\partial f(x)}{\partial x_i} = h_i'(x_i) = \frac{\partial h_i(x_i)}{\partial x_i}$$

Example 3 *A multiplicatively separable function is one why the sub-functions are multiplied to each other, such as in the following,*

$$f(x) = \prod_{i=1}^{n} h_i(x_i)$$

The note regarding generality in example 2 holds here as well. Also notice that by take logs, you would often be able to change a multiplicatively separable function to a additive one. The first derivative of $f(x)$ is,

$$\frac{\partial f(x)}{\partial x_i} = h_i'(x_i) \prod_{j=1, j\neq i}^{n} h_j(x_j) = \frac{\partial h_i(x_i)}{\partial x_i} \prod_{j=1, j\neq i}^{n} h_j(x_j)$$

You should ask yourself what would the derivative be if all the sub-functions are functions of x_i. Ask yourself the same question for the case of the additively separable function as well. *Finally, note that for the n variables, there will be n partial derivatives in total.*

2.2 Total Derivatives

The next question to ask is whether if there are instances where we with to find the rate of change of a function with all of the variables, perhaps to examine the entire surface of the function. Just as the first derivative of a univariate is just the tangent, we might be interested in the **tangent plane** to the function. In this case of a multi-variable function, we would be finding the **total derivative**. Let F be a twice continuously differentiable function, in the variables x_i where $i \in \{1, 2, \ldots, n\}$. Then it's total derivative at $\mathbf{x}^*$is,

$$dF(\mathbf{x}) \quad = \quad \frac{\partial F(\mathbf{x}^*)}{\partial x_1} dx_1 + \frac{\partial F(\mathbf{x}^*)}{\partial x_2} dx_2 + \cdots + \frac{\partial F(\mathbf{x}^*)}{\partial x_n} dx_n$$

It is worthwhile to reiterate here that derivatives are to be precise approximations only. We can rewrite the partial derivatives in vector form,

$$\nabla F_{x^*} = \begin{bmatrix} \frac{\partial F(\mathbf{x}^*)}{\partial x_1} \\ \frac{\partial F(\mathbf{x}^*)}{\partial x_2} \\ \vdots \\ \frac{\partial F(\mathbf{x}^*)}{\partial x_n} \end{bmatrix}$$

which in turn is known as the **Jacobian** or the **Jacobian Derivative** of F at x^* or the **gradient vector**.

2.3 Implicit Functions & Implicit Differentiation

Although the analysis in partial functions and partial derivatives of the previous section is couched in terms of holding the other variables constant, it is common to have variables "correlated" with each other. Since we have held the other variables constant, we do not need to consider those effects, but it does not negate their existence. We will discuss in this section, this idea of implicit functions and their derivatives.

Example 4 *Consider the utility function of an individual in terms of two goods, 1 and 2. The quantities consumed of each are x_1 and x_2 respectively, and let the utility function describing the felicity of this individual be $U(.,.)$. Obviously, if we are interested in the marginal utility of the consumption of good 1 on the individual's felicity, all we would need to do is to find the partial derivative, $U_{x_1}(x_1, x_2) = \frac{\partial U(x_1,x_2)}{\partial x_1}$. However, we are just as interested in the marginal rate of substitution (MRS) between the consumption of the two goods since, as you might recall, the equilibrium choice of the individual is in equating the marginal rate of substitution of the individual with the price ratio of the two goods. But what is the MRS but the slope of the indifference curve in a two dimensional diagram. We know that as we traverse along a given utility curve, the level of felicity is unaltered. However, we have to give up some quantities of one good for the other, in other words, the quantity of one is dependent on the other.*

$$\begin{aligned}\frac{\partial U(x_1,x_2)}{\partial x_1} + \frac{\partial U(x_1,x_2)}{\partial x_2}\frac{dx_2}{dx_1} &= 0 \\ \Rightarrow \frac{dx_2}{dx_1} &= \frac{U_{x_1}}{U_{x_2}}\end{aligned}$$

Notice that what we have done here is to use the Chain rule.

It is important to note that in performing the differentiation of the functions, that the variables must be continuous in the other. In other words if there are two variables which are dependent on each other, say x_1 and x_2, then x_1 should be a continuous function of x_2 and vice versa if we wish to perform the analysis the other way around. Let's define the **Chain Rule** formally here. First let $f : \mathbf{R}^n \to \mathbf{R}^1$ be a twice continuously differentiable function for all it's variables x_i where $i \in \{1, 2, \ldots, n\}$. Further, let x_i, $\forall i$ be functions of t and is also twice continuously differentiable. Then,

$$\begin{aligned}F(t) &= f(x_1(t), x_2(t), \ldots, x_n(t)) \\ \Rightarrow \frac{dF(t)}{dt} &= \frac{\partial f(\mathbf{x}(t))}{\partial x_1 t}\frac{dx_1(t)}{dt} + \frac{\partial f(\mathbf{x}(t))}{\partial x_2 t}\frac{dx_2(t)}{dt} + \cdots + \frac{\partial f(\mathbf{x}(t))}{\partial x_n t}\frac{dx_n(t)}{dt} \\ &= \frac{\partial f(\mathbf{x}(t))}{\partial x_1 t}x_1'(t) + \frac{\partial f(\mathbf{x}(t))}{\partial x_2 t}x_2'(t) + \cdots + \frac{\partial f(\mathbf{x}(t))}{\partial x_n t}x_n'(t)\end{aligned}$$

2.4 Direction & Gradients

You would recall in our previous discussion that we can always write the parameterized equation for a line in one parameter as,

$$\mathbf{x} \;=\; \mathbf{x}^* + t\mathbf{v}$$

where $\mathbf{v}$ is just the direction vector. Let F be the function, and $\mathbf{x}$ be the variable, so that the function evaluated along the line is,

$$h(t) = F(\mathbf{x}^* + t\mathbf{v}) \;=\; F(x_1^* + tv_1, x_2^* + tv_2, \ldots, x_n^* + tv_n)$$

Therefore we can use Chain Rule to examine the **derivative of $\mathbf{x}^*$ in the direction of $\mathbf{v}$** evaluated at $t = 0$,

$$h'(t=0) \;=\; \frac{\partial F(\mathbf{x}^*)}{\partial x_1}v_1 + \frac{\partial F(\mathbf{x}^*)}{\partial x_2}v_2 + \cdots + \frac{\partial F(\mathbf{x}^*)}{\partial x_n}v_n$$

which can be written in matrix form as,

$$\begin{bmatrix} \frac{\partial F(\mathbf{x}^*)}{\partial x_1} & \ldots & \frac{\partial F(\mathbf{x}^*)}{\partial x_n} \end{bmatrix} . \begin{bmatrix} v_1 \\ \vdots \\ v_2 \end{bmatrix} = \nabla F(\mathbf{x}^*)\mathbf{v}$$

Then the derivative desired could then be manipulated through varying the direction vector $\mathbf{v}$.

This idea works likewise when $F = [f_1, f_2, \ldots, f_m] : \mathbf{R}^n \rightarrow \mathbf{R}^m$, where each f_i is a function of a vector $\mathbf{x}$ in $\mathbf{R}^n$. In which case, the Jacobian derivative vector now becomes a $m \times n$ matrix. To see this, let $\mathbf{x} \equiv \mathbf{x}(\mathbf{t})$ as before with the difference being you have a different t_i $i \in \{1, \ldots, q\}$ for each of the m functions. Then

$$\nabla h(\mathbf{t}) = \begin{bmatrix} \frac{\partial f_1(\mathbf{x}^*)}{\partial x_1} & \ldots & \frac{\partial f_n(\mathbf{x}^*)}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m(\mathbf{x}^*)}{\partial x_1} & \ldots & \frac{\partial f_m(\mathbf{x}^*)}{\partial x_n} \end{bmatrix} . \begin{bmatrix} \frac{\partial x_1}{\partial t_1} & \ldots & \frac{\partial x_1}{\partial t_q} \\ \vdots & \ddots & \vdots \\ \frac{\partial x_n}{\partial t_1} & \ldots & \frac{\partial x_n}{\partial t_q} \end{bmatrix}$$

which in turn can be written more succinctly as,

$$\nabla h(\mathbf{t}) = \nabla F(\mathbf{x}(\mathbf{t})).\nabla \mathbf{x}(\mathbf{t})$$

You should read your text for more details on the interpretation.

2.5 Higher Order Derivatives & Definiteness

Just as we have higher order differentials in the univariate function case, so to we have in the multivariate case. We will discuss the interpretation here regarding concavity/convexity as well. Before we can find any higher order derivatives, likewise here, the functions has to be at least twice differentiable. The key difference in the multivariate case is that we not only have to find the "rate of change in the rate of change", we would also have to find the "rate of change due to the the change in another variable". We call these the **cross partial derivatives**. In other words, for a function f which is a function of x_i $i \in \{1, \ldots, n\}$, the cross partial derivative is $\frac{\partial^2 f}{\partial x_i \partial x_j}$, where $i \neq j$. For a function in n variables, there are n^2 or $n \times n$ cross partials with the typical (i, j) entry as above. We can form this cross partials into a $n \times n$ matrix of second order derivatives call the **Hessian Matrix**.

$$\nabla^2 f(\mathbf{x}) \equiv \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_2 \partial x_1} & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_1} \\ \frac{\partial^2 f}{\partial x_1 \partial x_2} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_1 \partial x_n} & \frac{\partial^2 f}{\partial x_2 \partial x_n} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix}$$

This leads us to **Young's Theorem**. Although *a priori* there is no reason to believe that cross partial $\frac{\partial^2 f}{\partial x_i \partial x_j}$ and $\frac{\partial^2 f}{\partial x_j \partial x_i}$ are the same, it turns out that usually/generally they are which is what Young's Theorem says.

So what's the big deal with knowing all this! We will get there soon enough. Consider a commonly observed function, the quadratic function,

$$\sum_{i \leq j} a_{ij} x_i x_j$$

or more generally in matrix form,

$$\mathbf{x}^T.\mathbf{A}.\mathbf{x}$$

How do we know when the function is concave or convex, a primary concern in economics, or any optimization process. The key point here is that for $\mathbf{A}_{n \times n}$, a symmetric matrix, it is

1. **positive definite** if $\mathbf{x}^T.\mathbf{A}.\mathbf{x} > 0 \ \forall x \neq \mathbf{0}$, $\mathbf{x} \in \mathbf{R}^n$,

2. **positive semidefinite** if $\mathbf{x}^T.\mathbf{A}.\mathbf{x} \geq 0 \ \forall x \neq \mathbf{0}$, $\mathbf{x} \in \mathbf{R}^n$,

3. **negative definite** if $\mathbf{x}^T.\mathbf{A}.\mathbf{x} < 0 \; \forall x \neq \mathbf{0}$, $\mathbf{x} \in \mathbf{R}^n$,

4. **negative semidefinite** if $\mathbf{x}^T.\mathbf{A}.\mathbf{x} \leq 0 \; \forall x \neq \mathbf{0}$, $\mathbf{x} \in \mathbf{R}^n$, and

5. **indefinite** if $\mathbf{x}^T.\mathbf{A}.\mathbf{x} > 0$ for some $\mathbf{x} \in \mathbf{R}^n$, and < 0 for others in $\mathbf{R}^n$.

What are the implications then to our concern with optimization problems. Recall in the univariate case, a function is concave if it's second order derivative is less than or equal to zero (≤ 0). On the other hand, it is negative is the second order derivative is greater than or equal to zero. Here in the context of a multivariate function, the function is concave on some region if the second order derivative is negative semidefinite $\forall x$ in the region. Similarly, it is convex if it is positive semidefinite. But what do those terms mean?

Before we can link these ideas to optimization, we need additional terminology and definitions. Let $\mathbf{A}$ be $n \times n$ matrix. A k^{th} order **principal submatrix** of $\mathbf{A}$ is a $k \times k$ ($k \leq n$) submatrix generated from $\mathbf{A}$ with $n - k$ of the same sequenced columns and rows deleted from $\mathbf{A}$. For example, consider 3×3 matrix,

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$$

Then it would have one third order principal submatrix, which would be the matrix $\mathbf{A}$ itself. It would have three second order principal submatrix,

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \qquad \begin{bmatrix} a_{11} & a_{13} \\ a_{31} & a_{33} \end{bmatrix} \qquad \begin{bmatrix} a_{22} & a_{23} \\ a_{23} & a_{33} \end{bmatrix}$$

and three first order principal submatrix a_{11}, a_{22} and a_{33}. A k^{th} order principal minor is the determinant of these principal submatrices. In turn, the k^{th} order **leading principal submatrix** of $\mathbf{A}$ such as the previous one discussed, is the submatrix with the last $n - k$ rows and columns deleted, and is denoted as A_k. Its determinant is then referred to as the k^{th} order **leading principal minor** denoted as $|A_k|$. Then we can determine whether a function is "concave" or "convex" by the following,

1. Matrix $\mathbf{A}$ is **positive definite** iff all its n leading principal minors are strictly positive.

2. Matrix **A** is **negative definite** iff its n leading principal minors alternate in sign as follows,

$$|A_1| < 0 \quad |A_2| > 0 \quad |A_3| < 0 \ldots$$

   That is the k^{th} order leading principal minor has the sign of $(-1)^k$.

3. When the signs does not accord with the above, the matrix is **indefinite**

Note that just as we have strict concavity versus concavity, we have the same here in terms of positive and negative semidefinite matrices, which corresponds with convexity and concavity respectively.