14.102, Math for Economists
Fall 2004
Lecture Notes, 10/5/2004

These notes are primarily based on those written by Andrei Bremzen for 14.102 in 2002/3, and by Marek Pycia for the MIT Math Camp in 2003/4. I have made only minor changes to the order of presentation, and added a few short examples. The usual disclaimer applies; questions and comments are welcome.

Nathan Barczi
nab@mit.edu

8 Differentiability and Derivatives

8.1 The Derivative

Weierstraß theorem guarantees that an optimum exists (under specified restrictions on the set and the function). However, it gives absolutely no clue how to find an optimum. Our next step on the path towards learning to optimize (which is essentially our main objective in this course) is studying necessary and/or sufficient conditions for optima. As usually, more definitive results require more structure, and in this case it is the notion of differentiability.

Informally, a function is called "smooth" if it is continuous and its graph has no kinks. Here is a formal definition:

Definition 146 *Function $f : X \subset \mathbb{R}^n \to \mathbb{R}^m$ is called differentiable at point $x_0 \in X$ if it can be decomposed as follows:*

$$f(x) = f(x_0) + D_{x_o} f \cdot (x - x_0) + \alpha(x - x_0), \text{ where}$$

- *$\alpha(x-x_0)$ is small compared to $x-x_0$, i.e., $\frac{\|\alpha(x-x_0)\|}{\|x-x_0\|} \to 0$ as $\|x - x_0\| \to 0$.*
- *$D_{x_0} f$ is a linear function from $\mathbb{R}^n$ to $\mathbb{R}^m$.*

A function is said to be differentiable on X if it is differentiable at any point $x_0 \in X$.

If $n = m = 1$, then $D_{x_0} f$ is a number, called the *derivative* of f at x_0, denoted by $f'(x_0)$ or $\frac{df}{dx}(x_0)$;

If $m = 1$ and $n > 1$, then $D_{x_0} f$ is a row $1 \times n$ vector called the *gradient* of f at x_0, denoted $\nabla f(x_0)$;

Finally, if $n > 1$ and $m > 1$, then $D_{x_0} f$ is an $m \times n$ matrix called the *Jakobi matrix,* sometimes denoted $J_f(x_0)$.

Generally, $D_{x_0}f$ is a linear function that can help approximate f around x_0. It does not have to exist though.

Notice that this definition explicitly generalizes the usual definition for functions mapping the real line to the real line to more dimensions:

Definition 147 *The derivative of f at x is*

$$f'(x) \equiv \frac{df}{dx}(x) = \lim_{h\to 0} \frac{f(x+h)-f(x)}{h}.$$

When the the limit exists the function is called differentiable at x. We use the term differentiable function to denote functions differentiable at every point in the domain.

The deriviate has a nice geometric interpretation: it is the slope of the tangent. Another way to write it is:

$$f(x+h) = f(x) + hf'(x) + h\epsilon(h) = f(x) + hf'(x) + o(h) \qquad (14)$$

$o(h)$ is a quantity that can be written $o(h) = h\epsilon(h)$ where $\epsilon(h) \to 0$ as $h \to 0$. The notation $o(h)$ is often used (e.g. in 14.381) and denotes a quantity negligible relative to small h i.e. such that $\lim_{h\to 0} \frac{o(h)}{h} = 0$.

Instead of h one can also write dx. Using the dx notation one often drops $o(dx)$:

$$f(x+dx) = f(x) + f'(x)dx + o(dx) = f(x) + f'(x)dx$$

Lemma 148 *If a function is differentiable at a point, it is continuous at that point.*

Example 149 *$f(x) = |x|$ is differentiable everywhere except at $x_0 = 0$.*

Example 150 *Usual functions are differentiable on their domains: ln, exp, sin, cos, polynomials, radicals.*

Any function of n variables $f(x_1, ..., x_n)$ can also be viewed as a function of one variable x_1 which depends on parameters $x_2, ..., x_n$. Similarly it can be viewed as a function of x_2 only, that depends on parameters $x_1, x_3, ..., x_n$. and so on. If f is differentiable as a function of x_1 at x_0, its derivative (called *partial derivative of f with respect to x_1*) is denoted $\frac{\partial f}{\partial x_1}(x_0)$; similarly we can define $\frac{\partial f}{\partial x_2}, ..., \frac{\partial f}{\partial x_n}$. The following lemma is straightforward but important conceptually.

Lemma 151 *If $f : X \subset \mathbb{R}^n$ is differentiable as a function of n variables at point x_0, then it is also differentiable as a function of any of its variable at x_0 when the other variables are viewed as parameters, and $\nabla f(x_0) = \left(\frac{\partial f}{\partial x_1}(x_0), ..., \frac{\partial f}{\partial x_n}(x_0)\right)$.*

The converse generally is not true:

Example 152 *Let $f(x, y) = sign(xy)$. Then all (i.e. both) partial derivatives of f at $(0, 0)$ exist, but f is not differentiable at $(0, 0)$ (and even not continuous).*

Definition 153 *A function $f : X \subset \mathbb{R}^n \to \mathbb{R}$ is said to belong to class $C^1[X]$ if it is differentiable on X and its gradient (or simply derivative for $n = 1$) is continuous on X.*

Proposition 154 *Differentiable functions f,g are continuous and:*

- $(f+g)' = f' + g'$, $(fg)' = f'g + g'f$, $(\frac{f}{g})' = \frac{f'g - g'f}{g^2}$
- $(f \circ g)'(x) = f'(g(x))g'(x)$
- $(\ln)'(x) = 1/x, (\exp)' = \exp, (x^n)' = nx^{n-1}, (sin)\prime = cos, (cos)\prime = -sin$
- $(f^\alpha)' = \alpha f' f^{\alpha - 1}$

Proof of the formula on differential of the product. Let $f(x+h) = f(x) + hf'(x) + o(h)$ and $g(x+h) = g(x) + hg'(x) + o(h)$. Note that we use same $o(h)$ even though the two o's are really different. Then

$$\begin{aligned} f(x+h)g(x+h) &= [f(x) + hf'(x) + o(h)][g(x) + hg'(x) + o(h)] \\ &= f(x)g(x) + h\left(f'(x)\, g(x) + f(x)\, g'(x)\right) + o(h) \end{aligned}$$

and after moving $f(x)g(x)$ to the LHS, dividing by h, and taking limit as $h \to 0$ we obtain the formula from the proposition.

8.2 Mean Value Theorem

Lemma 155 *(Fermat's or "hills are flat at the top" theorem) Let f be differentiable on (a,b) and let $c \in (a,b)$ be a* max *for f: $f(c) \geq f(x)$ for any $a < x < b$. Then $f'(c) = 0$*

Proof. $f(c+h) \leq f(c)$ so $(f(c+h) - f(c))/h \leq 0$ for $h > 0$ so right limit is less or equal to zero, so $f'(c) \leq 0$. For $h < 0$, get $f'(c) \geq 0$. QED. ■

Of course same thing true for a minimum. Careful: need differentiability assumption and open set to avoid corner solutions.

This theorem is the basic tool used to find maxima and minima. If a function is differentible, then to find its maximum we need to check

- values at the boundary of the domain
- internal points x satysfying so called "first order condition"

$$f'(x) = 0$$

Theorem 156 *Mean Value Theorem. Let f be continuous on $[a,b]$ and be differentiable on (a,b). Then there is $c \in (a,b)$ such that $f(b) - f(a) = f'(c)(b - a)$.*

Hint for the proof: consider $f(t) - \frac{f(b)-f(a)}{b-a}(t-a)$ and apply Fermat's Theorem.

Proposition 157 *Let f be continuous on $[a,b]$, differentiable on (a,b) and $f'(x) > 0$ for all $x \in (a,b)$. Then f is strictly increasing. Similarly $f'(x) < 0$ implies f strictly is decreasing.*

For proof use Mean Value Theorem. Converse not exactly true. Example $x \to x^3$.

Proposition. Let f be continuous on $[a,b]$ and be differentiable on (a,b).

1. $f'(x) \geq 0$ for all x iff f is weakly increasing.
2. $f'(x) \leq 0$ for all x iff f is weakly decreasing.
3. $f'(x) = 0$ for all x iff f is constant.

8.3 High Order Derivatives and Taylor Expansions.

The derivative of a function $f(t)$ of one variable is itself a function $f'(t)$ of one variable. Therefore we can easily define the *second derivative* as the derivative of the derivative (this does not have to exist, of course). Similarly we define the third derivative, the fourth derivative and so on. We say that $f(x) \in C^k(X)$ if the k-th derivative of f exists at each point of X and is continuous on X (that immediately implies that all derivatives of lower order exist and are continuous as well.

Example 158 *Usual functions are infinitely many times differentiable on their domains: ln, exp, sin, cos, polynomials, radicals.*

The very definition of differentiability tells us how a differentiable function behaves around point x_0 : it is equal to $f(x_0)$ plus a linear term $f'(x_0) \cdot (x - x_0)$ plus some α which is "of the higher order than linear", i.e., goes to zero "faster" than $x - x_0$ itself (in the precise sense defined above). As always,imposing additional structure (in our case, assuming existence of higher order derivatives at x_0) enables us to come to more specific conclusions (in our case to further decompose the black box α into simpler pieces.

Theorem 159 *(Taylor Decomposition in $\mathbb{R}^1$): Suppose that $f : X \subset \mathbb{R} \to \mathbb{R}$ is a C^k function and x_0 is a point in X. Then $f(x)$ can be decomposed as follows:*

$$f(x) = f(x_0) + f'(x_0) \cdot (x - x_0) + \frac{f''(x_0)}{2!} \cdot (x - x_0)^2 + ... + \frac{f^{(k)}(x_0)}{k!} \cdot (x - x_0)^k + o(x - x_0)^k, \text{ where } \frac{o(x-x_0)^k}{(x-x_k)^k} \to 0 \text{ as } (x - x_0) \to 0.$$

Essentially this means that a smooth enough function locally can be very well approximated by a polynomial whose coefficients are related to derivatives of the function.

Having seen this k-term Taylor decomposition for a C^k function one may be tempted to conclude that a C^∞ function can be similarly decomposed into an infinite series $\sum a_k$, where $a_k = \frac{f^{(k)}(x_0)}{k!}(x - x_0)^k$. This conclusion is generally *wrong:* such infinite series does not have to converge at all (making the infinite sum meaningless) and, moreover, even if it does converge, it may converge to something *other than* $f(x)$, as the following example illustrates.

Example 160 *Let* $f(x) = \begin{cases} e^{-\frac{1}{x^2}}, & x \neq 0 \\ 0, & x = 0 \end{cases}$*. Then* $f(x)$ *is infinitely differentiable at all points (including zero) and all its derivatives at* $x_0 = 0$ *are equal to zero, thus making the above mentioned infinite sum equal to zero at any point* x*, whereas* $f(x) > 0$ *for all* $x \neq 0$*.*

Things get more complicated when we consider functions of multiple variables. For a function $f(x) : \mathbb{R}^n \to \mathbb{R}$, its gradient $\nabla f(x)$ is a function $\mathbb{R}^n \to \mathbb{R}^n$, so its derivative will be an $n \times n$ matrix (called the *Hessian* of f), which will be a matrix function of the point at which it is evaluated. If we attempted to differentiate that, we would end up with objects of structure that is unknown to us (something like a 3D matrix), so we restrict ourselves to second order derivatives.

Lemma 161 *If* $f(x_1, ..., x_n)$ *is a* C^2 *function around point* $x_0 \in \mathbb{R}^n$*, then* $\frac{\partial^2 f}{\partial x_i \partial x_j}(x_0) = \frac{\partial^2 f}{\partial x_j \partial x_i}(x_0)$*. Thus the Hessian* $H_f(x_0)$*, which is defined as an* $n \times n$ *matrix of crosspartials of* $f(x)$ *(i.e.,* $h_{ij} = \frac{\partial^2 f}{\partial x_i \partial x_j}$*), is a symmetric matrix.*

Finally we state the Taylor decomposition theorem for $\mathbb{R}^n$. For the reasons discussed above, we only go up to the second term.

Theorem 162 *(Taylor decomposition in* $\mathbb{R}^n$*): Suppose that* $F : X \subset \mathbb{R}^n \to \mathbb{R}$ *is a* C^2 *function and* x_0 *is a point in* X*. Then* $F(x)$ *can be decomposed as follows:*

$$F(x) = F(x_0) + \nabla F(x_0) \cdot (x - x_0) + (x - x_0)' H_F(x_0)(x - x_0) + \beta(x - x_0), \text{ where } \frac{\|\beta(x-x_0)\|}{\|x-x_0\|^2} \to 0 \text{ as } (x - x_0) \to 0.$$

8.4 Implicit Function Theorem

Definition 163 *A* ***level curve*** *for a function* $f : U \subset R^n \to R$ *is a curve in* R^n *defined by* $f(x) = C$

Let $F(x, y) = x^2 + y^2$ (which is a C^1 function) and suppose we are interested in the level curve $F(x, y) = c$. Pick a point (x^*, y^*) on this curve and ask yourself if y is expressible as a function of x along the curve $F(x, y) = c$ around this point. The answer is "almost always yes": if $y^* > 0$, then for x close enough to x^* we have $y(x) = \sqrt{c^2 - x^2}$; if $y^* < 0$ then (again, for x close enough to x^*) we have $y(x) = -\sqrt{c^2 - x^2}$. Note that in either case $y(x)$ is a C^1 function and its derivative $y'(x^*)$ can be easily found by differentiating through the identity $F(x, y(x)) = c$ at point x^* : by the chain rule we have $\frac{\partial F}{\partial x}(x^*, y^*) + \frac{\partial F}{\partial y}(x^*, y^*) \cdot y'(x^*) = 0$, so $y'(x^*) = -\frac{\frac{\partial F}{\partial x}(x^*, y^*)}{\frac{\partial F}{\partial y}(x^*, y^*)}$. Note that for the last formula to work, we need $\frac{\partial F}{\partial y}(x^*, y^*) \neq 0$ and the two points where

$\frac{\partial F}{\partial y}(x^*, y^*) = 0$ are exactly points $(1, 0)$ and $(-1, 0)$ where y can not be locally represented as a well defined function of x.

The following theorem captures the intuition for the example above:

Theorem 164 *(Implicit Function Theorem): Let $F(x_1, ..., x_n)$ be a C^1 function around the point $(x_1^*, ..., x_n^*, y^*)$ such that $\frac{\partial F}{\partial y}(x_1^*, ..., x_n^*, y^*) \neq 0$. Denote $c = F(x_1^*, ..., x_n^*, y^*)$. Then there exists a C^1 function $y = y(x_1, ..., x_n)$ defined around $(x_1^*, ..., x_n^*)$ such that:*

- $F((x_1, ..., x_n, y(x_1, ..., x_n)) = c$
- $y^* = y(x_1^*, ..., x_n^*)$
- $\frac{\partial y}{\partial x_i}(x_1^*, ..., x_n^*) = -\frac{\frac{\partial F}{\partial x_i}(x_1^*, ..., x_n^*, y^*)}{\frac{\partial F}{\partial y}(x_1^*, ..., x_n^*, y^*)}$.

What this theorem tells you is basically that if you have one equation you can, provided some regularity conditions, solve for one unknown as a function of the other unknowns. In economic language we say that one variable (in our case y) is determined *endogenously* as a function of the other (*exogenous*) variables.

It is important to note that, typically, the regularity condition (that the relevant partial is nonzero) will be satisfied for more than one variable. Consequently, this theorem may be applicable in more than one way, i.e., you can choose which variable to express in terms of others. Hence, which variables are exogenous and which are endogenous is determined from the economic story behind your equation, math per se does not help you to figure it out.

Finally, a similar theorem applies for more than one (say, k) equations. In this case, provided you have similar regularity conditions (that have to do with the rank of relevant Jakobi matrix), you can (locally) solve for k variables (to be therefore endogenous) in terms of the other $n - k$ (which are, therefore, considered exogenous).

Example 165 *Consider the following ISLM model developed by Bernanke and Blinder to study the lending channel of monetary policy:*

$$L(\rho, i, y) = \lambda(\rho, i) D(1 - \tau) \quad (15)$$

$$D(i, y) = m(i) R \quad (16)$$

$$y = Y(i, \rho) \quad (17)$$

Equation (15) gives the market clearing condition in the loan market (the left-hand side is demand, the right-hand side supply). D is the deposits held by banks, τ is the fraction of deposits required to be held in reserves by banks to back up their deposits, $\lambda()$ is the fraction of non-required reserves that are supplied as loans, i is the interest rate on bo nds, and ρ is the interest rate on bank lonas. Note that $L_\rho < 0$, $L_y > 0$, and $\lambda_\rho > 0$. Equation (16) is the money market clearance condition (LM curve), where $D()$ is the demand for deposits, m is the money multiplier, and R is the reserves of the banking system. Equation (17) closes the model, and is the goods market clearing condition.

Using Equations (15) and (16), we can find an implicit function for ρ *in terms of* $i, y,$ *and* R*:*

$$F(i, y, R, \rho) = L(i, y, \rho) - \lambda(\rho, i)(1-\tau)m(i)R = 0$$

Moreover, we can use this implicit function theorem to show that $\frac{\partial \rho}{\partial y} > 0$ *and* $\frac{\partial \rho}{\partial R} < 0:$

$$\begin{aligned} \frac{\partial \rho}{\partial y} &= -\frac{F_y}{F_\rho} = -\frac{L_y}{L_\rho - \lambda_\rho(1-\tau)m(i)R} > 0 \\ \frac{\partial \rho}{\partial R} &= -\frac{F_R}{F_\rho} = -\frac{-\lambda(\rho, i)(1-\tau)m(i)}{L_\rho - \lambda_\rho(1-\tau)m(i)R} < 0 \end{aligned}$$

(Intuitively, a higher income level, y*, increases the demand for loans, and all else being equal it must be that the interest rate rises in order to raise supply to meet this demand; a higher amount of reserves* R *increases the amount of money and deposits in the economy, driving up the supply of loans, so the interest rate must fall to restore equilibrium.)*