ARE211, Fall 2009

CALCULUS1: TUE, OCT 20, 2009 PRINTED: AUGUST 27, 2009 (LEC# 16)

Contents

4. Univariate and Multivariate Differentiation

4.1. Univariate calculus: the derivative

Assume you all know how to calculate the derivative of a single variable function, i.e., given f, calculate $\frac{df(\cdot)}{dx}$, denoted also $f'(\cdot)$. Important to know the difference between $f'(\cdot)$, which is a function and $f'(x)$, which is a number, the function evaluated at a point.

I'll try to be careful to use this notation from now on: $g(\cdot)$ is a RULE, represents a function.

Since $f'(\cdot)$ is a function, like any other, it may have a derivative; if it does, call it $f''(\cdot)$. Do example $f(x) = x^2$.

Lots of standard kinds of functions you have to be able to differentiate in your sleep. Equivalent of being able to spell. Brainless activity.

4.2. **The fundamental idea: linear approximations to nonlinear functions**

The most important thing to keep in mind about calculus is:
(Calculus Mantra) When you do calculus, you are, ALWAYS, approximating a small change in a differentiable function by evaluating THE (unique) appropriate LINEAR function at the magnitude of the change. That is, for sufficiently small dx, $f(x + dx) - f(x) \approx f'(x)dx$.

$df^x(\cdot) = f'(x)(\cdot)$ is the linear function that approximates the difference between $f(x+\cdot)$ and $f(x)$.

Draw picture of the comparison: concave example.

- graph of $f(\cdot)$ with points x^* and $x^* + dx^*$.
- look at difference in the value of the function at the two points.
- look at the graph of the linear function $df = f'(x^*)dx$:
- note that the difference between $(f(x^* + dx^*) - f(x^*))$ and $f'(x^*)dx^*$ will be arbitrarily small provided that dx^* is sufficiently small.
- Explain the enormous practical importance of this.

When dx is bigger, don't do so well at approximation. Could do better if you threw in the second derivative, i.e.,

$$f(x + dx) - f(x) \quad \approx \text{even better} \quad f'(x)dx + \frac{1}{2}f''(x)dx^2.$$

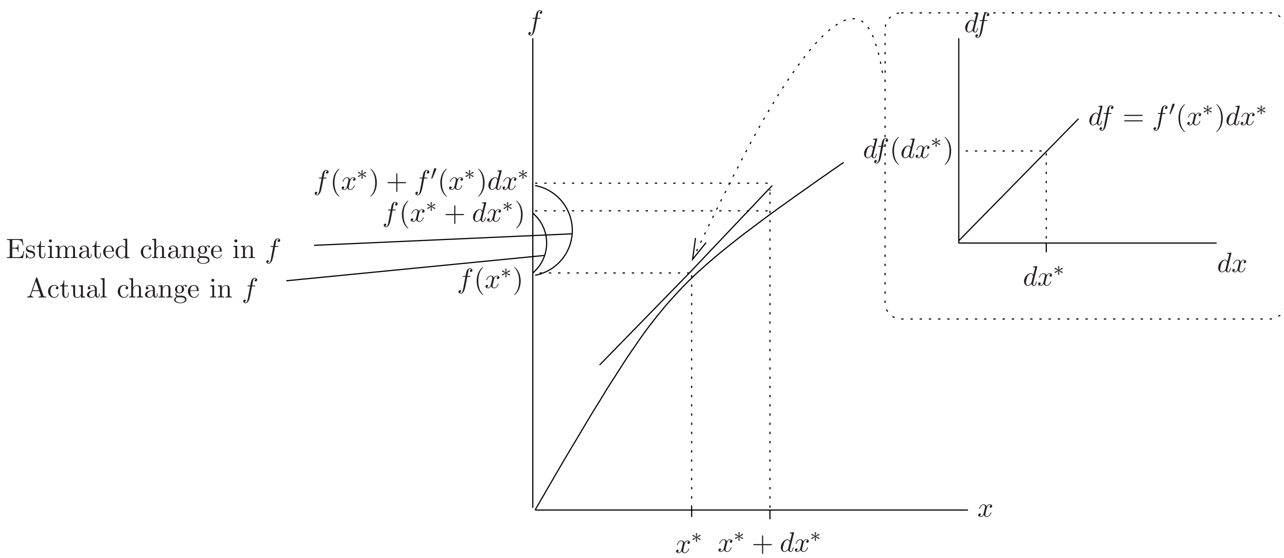

FIGURE 1. Linear Approximations of the change in a nonlinear function

Note that in the above figure, $f''(x^*) < 0$. Note also that the linear approximation gives an overestimate of the change, so that adding the second term helps.

This is the germ of a crucial theorem called Taylor's theorem, that we'll see more of later.

4.3. **Univariate Calculus: the differential**

You've probably seen the expression $dy = f'(x)dx$, known as the *differential*, before. What sort of object is this? What's the relationship between the differential and the derivative? Answer is ridiculously simple. Recall from last time about linear functions from $\mathbb{R}^1$ to $\mathbb{R}^1$ and scalars; every such function is uniquely defined by a single scalar. E.g., I talked about the *function* p, when I really meant the rule $f(\cdot)$ defined by $f(x) = px$. Well... The differential is to the derivative as the linear function is to the scalar/vector that defines it.

The relationship between the differential and the derivative is the same as the relationship between any linear function and the scalar/vector/matrix that defines it.

Defn: the differential of f at x is the linear function defined by the scalar $f'(x)$. In other words, the differential of f at x is the unique linear function whose coefficient (slope) is equal to the derivative of f at x.

Restate: The most important thing to keep in mind about calculus is: when you do calculus on f at x, you are, ALWAYS, approximating a small change in f, starting from x, by evaluating *the differential of f at x* at the magnitude of the change. This is the idea that underlies *all* of comparative statics.

That is, $dy = f'(x)dx$ is the value of the approximation to the change in f when you shift from x to $x + dx$.

Economic example: have $\pi(p)$, i.e., profits are a function of prices (assuming optimal input choices for prices p.) If p change is small enough, $d\pi \approx \pi'(p)dp$.

It is the ability to make computations like this—we call them comparative statics—that distinguish economists from art historians and sociologists. Example:

- Let $f(x) = x^3 + 2x^2$
- set $x^* = 1$, $dx^* = 0.1$
- then $f(x^*) = 3$, $f'(x^*) = 3x^{*2} + 4x^* = 7$, $f(x^* + dx^*) = 1.1^3 + 2 * 1.21 = 3.751$.
- The true difference $f(x^* + dx^*) - f(x^*)$ is .751.
- the estimated difference is $f'(x^*)dx^* = 0.70$, not bad.

4.4. **Multivariate calculus: functions from $\mathbb{R}^n$ to $\mathbb{R}$**

4.4.1. *Partial Derivative.* . Cookbook approach: If $f : \mathbb{R}^n \to \mathbb{R}$, then $f_i(\cdot) \equiv \frac{\partial f(\cdot)}{\partial x_i}$ is the (usual) derivative of the single variable function you get by treating all other variables as constant.

Example:

$$
\begin{aligned}
f(\mathbf{x}) &= x_1^2 + 2x_1x_2 + x_2^2 \\
&= x_1^2 + 2x_1\alpha + \beta
\end{aligned}
$$

$$
\begin{aligned}
f_1(\mathbf{x}) &= 2x_1 + 2\alpha + 0 \\
&= 2x_1 + 2x_2
\end{aligned}
$$

A more graphical view of partial derivatives. Take a cross-section of the graph of the function along the i'th axis, and look at the slope of the single-dimensional function you obtain in this way. This slope is the i'th partial derivative.

Similarly, you could take a diagonal cross-section, and get a different one-dimensional slope. Generally, you get what is called a *directional derivative.* Partial derivatives are just special kinds of directional derivatives: in particular, they tell you the slopes you obtain when you take cross-sections along the various axes.

4.4.2. *The Gradient.* Given a function $f : \mathbb{R}^n \to \mathbb{R}$, the gradient of f is just the function from $\mathbb{R}^n$ to $\mathbb{R}^n$ that maps each $\mathbf{x} \in \mathbb{R}^n$ to the vector of partial derivatives of f at $\mathbf{x}$. That is,

$$\bigtriangledown f(\cdot) = \begin{bmatrix} f_1(\cdot) \\ \vdots \\ f_i(\cdot) \\ \vdots \\ f_n(\cdot) \end{bmatrix}$$

Emphasize that $\bigtriangledown f(\cdot)$ is the exact analog for $f : \mathbb{R}^n \to \mathbb{R}$. as $f'(\cdot)$ is for $f : \mathbb{R} \to \mathbb{R}$. In fact the words "slope of f," "gradient of f" and "derivative of f" are synonomous.

Emphasize the important distinction between the "gradient of f" which is a *function*, written $\bigtriangledown f$ or $\bigtriangledown f(\cdot)$ and the "gradient of f at $\mathbf{x}$," which is a vector.

4.4.3. *Crosspartial Derivative.* A cross-partial derivative is just an entry in the matrix which is the derivative of the derivative. That is, take a function $f : \mathbb{R}^n \to \mathbb{R}$. The i'th partial derivative of this function $f_i(\cdot)$ is a function, like any other, and if the function is differentiable, it has derivatives:

$f_{ij}(\cdot) \equiv \frac{\partial^2 f}{\partial x_i \partial x_j}(\cdot)$ is the j'th partial derivative of the function $f_i(\cdot) \equiv \frac{\partial f}{\partial x_i}(\cdot)$.

Thus, the Hessian of f just consists of the matrix of crosspartials of f.

4.4.4. *The differential for functions from $\mathbb{R}^n$ to $\mathbb{R}$.* Recall from earlier: the most important thing to keep in mind about calculus is that when you do calculus on f at x, you are, ALWAYS, approximating a small change in f, starting from x, by evaluating *the differential of f at x* at the magnitude of the change. We'll look at two numerical examples, then look at the picture.

Numerical Example 1: Use the example above: $f(x,y) = x^2 + y^2$; $\bigtriangledown f(\cdot) = (2x, 2y)$; Consider $(dx, dy) = (0.1, 0.1)$, set $(x^0, y^0) = (100, 100)$ and approximate $f(x^0 + dx, y^0 + dy) - f(x^0, y^0)$ by $\bigtriangledown f(x^0, y^0) * (0.1, 0.1) = (200, 200) * (0.1, 0.1) = 40$. Note that this approximation is virtually the same as the one you get when you use the directional derivative! Now evaluate at f at both $(x^0, y^0) = (100, 100)$ and $(x^0 + dx, y^0 + dy) = (100.1, 100.1)$, and compare: $f(x^0, y^0) = 20,000$ $f(x^0 + dx, y^0 + dy) = 20040.02$ Actual difference is 40.02.

Numerical Example 2: $y = f(\ell, k)$; $\bigtriangledown f(\cdot) = (f_\ell(\cdot), f_k(\cdot))$. Fix $(\bar{\ell}, \bar{k})$. Evaluate the difference $f(\bar{\ell}, \bar{k}) - f(\ell, k)$ by the value of the differential, evaluated at $(\bar{\ell}, \bar{k}) - (\ell, k)$:

$$f(\bar{\ell}, \bar{k}) - f(\ell, k) = dy \quad = \quad f_\ell(\bar{\ell}, \bar{k})d\ell + f_k(\bar{\ell}, \bar{k})dk$$

Now for the picture:

Exactly the same thing for functions from $\mathbb{R}^n$ to $\mathbb{R}$. Evaluate the difference $f(\mathbf{x}+\mathbf{dx}) - f(\mathbf{x})$, where here $\mathbf{dx}$ and $\mathbf{x}$'s are both vectors, by the value of the differential, evaluated at the difference, $\mathbf{dx}$.

What's the differential, in terms that we've used before? Recall that it is a linear function from $\mathbb{R}^n$ to $\mathbb{R}$; so it is characterized by a vector of coefficients. What is the vector? Answer: the gradient, evaluated at x.

Thus: $df = \bigtriangledown f(\mathbf{x}) \cdot \mathbf{dx}$. See Fig. 2 below. As the figure indicates, the differential here plays exactly the same role as it played for functions from $\mathbb{R}$ to $\mathbb{R}$. In Fig. 2, the tangent plane lies below the function, at least a neighborhood of $\mathbf{x}$. Then, as we move from $\mathbf{x}$ to $\mathbf{x}+\mathbf{dx}$, the value of the function declines; but the height of the tangent plane declines even further. Applying the calculus mantra, we approximate the change in f, i.e., $f(\mathbf{x} + \mathbf{dx}) - f(\mathbf{x})$, by translating the tangent plane so that $\mathbf{x}$ becomes the new origin, then approximate the change in f when you move from $\mathbf{x}$ to $\mathbf{x} + \mathbf{dx}$ by evaluating the height of this translated plane at the point $\mathbf{dx}$. Notice from the top panel that the absolute value of $df^{\mathbf{x}}(\mathbf{dx})$ over-estimates the absolute magnitude of the change in the function.

Sometimes economists refer to the differential as the *total differential*. It is important to recognize that the total derivative of a function and the total differential are completely different things. The

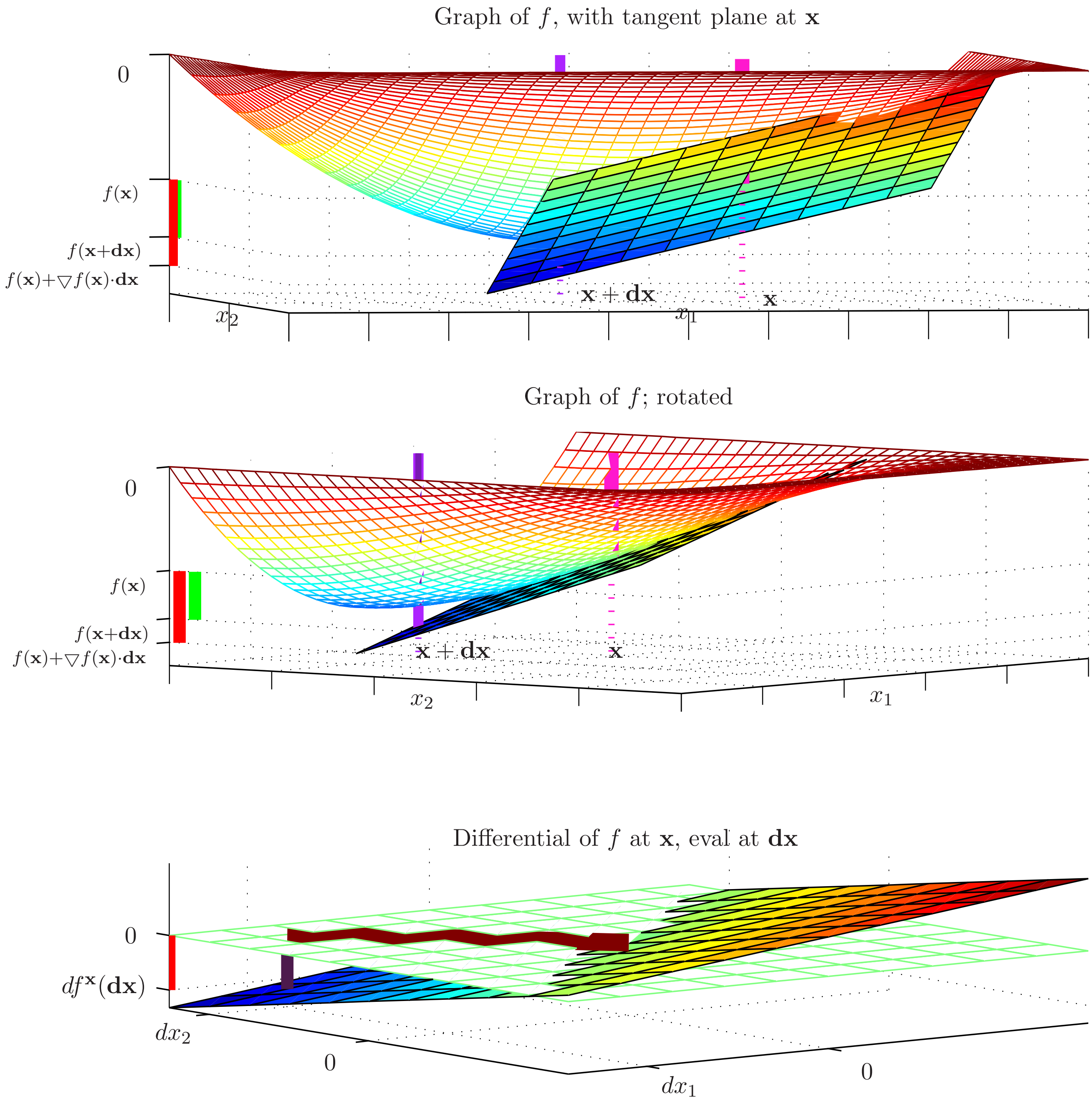

FIGURE 2. Linear Approximation of the change in a nonlinear multivariate function

total derivative is related to (but not the same as) a directional derivative (i.e., you take the cross section of the function in a certain direction). The total differential is the name of a linear function.