Approximation, Taylor Polynomials, and Derivatives

Derivatives for functions $f : \mathbb{R}^n \to \mathbb{R}$ will be central to much of Econ 501A, 501B, and 520 — and also to most of what you'll do as professional economists. The derivative of a function f is simply a linearization, or linear (or affine) approximation of f. For *real* functions, $f : \mathbb{R} \to \mathbb{R}$, this is pretty straightforward, and it's something you already know. So we'll start there, and then generalize to functions $f : \mathbb{R}^n \to \mathbb{R}$.

Suppose, then, that we want to approximate the values of $f(x) = x^2$. This is as simple as it gets: all we have to do is multiply x times x and we get $f(x)$ *exactly*, not merely an approximation. But this example will actually be instructive, as we'll see.

Here's a second example: We wish to evaluate, or approximate, the values of $f(x) = e^x$ — let's say we want to approximate e^x at $x = 1$. So we're actually approximating the value of e. This one is not as obvious as $f(x) = x^2$.

Let's first use the simple example of $f(x) = x^2$ to develop our ideas and some useful notation. Suppose we want to approximate the value of $f(x)$ for values of x near $\overline{x} = 1$, as in Figure 1, because we know that $f(\overline{x}) = 1$. Let's use $\overline{y}$ to denote $f(\overline{x})$ — *i.e.*, $\overline{y} = f(\overline{x})$. For any $x \in \mathbb{R}$, let's write

$$\begin{aligned} \Delta x &= x - \overline{x}, \qquad \textit{i.e.},\ x = \overline{x} + \Delta x; \\ \Delta y &= y - \overline{y} \ = \ f(x) - f(\overline{x}) \ = \ f(\overline{x} + \Delta x) - f(\overline{x}) \ = \ F(\Delta x); \end{aligned}$$

we're defining F to be $F(\Delta x) := f(\overline{x} + \Delta x) - f(\overline{x})$, so that $\Delta y = F(\Delta x)$. Notice that Δy is the *exact* change in y that takes place, given by $\Delta y = F(\Delta x)$, as in Figure 2, and that the graph of F is the same as the graph of f but with the coordinates shifted.

We want to find a function, say $G(\Delta x)$, that gives a *best approximation* of $\Delta y = F(\Delta x)$ — we want G to be a best approximation of the *exact* function $F(\Delta x)$. Equivalently, we want a function $g(x) = f(\overline{x}) + G(\Delta x)$ that approximates $f(x)$.

What we want is a *simple* function G that will be a good approximation of F. So let's say we want to find the best *linear* function $G(\Delta x) = a\Delta x$ to approximate $F(\Delta x)$. In other words, we want to know what the coefficient a should be in order to make the function $G(\Delta x) = a\Delta x$ the best linear approximation of the nonlinear function $F(\Delta x)$. We even say that this best linear approximating function is the *linearization* of F near $\overline{x}$. (Note that $G(0) = 0$: at $\Delta x = 0$, G coincides with F.)

Intuition about the diagram in Figure 3 suggests that the best linear approximation to F, near $\overline{x} = 1$, is the tangent to the graph of F (which is also the graph of f) at $\overline{x} = 1$. If that's the case, then the best coefficient a is the slope of the tangent — a should be the derivative $f'(\overline{x})$, the slope of the tangent to the graph of f at $\overline{x}$. Moreover, if a is any other number, such as $\widetilde{a}$ in Figure 4, then the approximation, at least near $\overline{x}$, will not be as good.

We make this intuition precise by saying that a best approximation $G(\Delta x)$ is one that satisfies the equation

$$\lim_{\Delta x \to 0} \frac{1}{\Delta x} \big[F(\Delta x) - G(\Delta x) \big] = 0, \tag{1}$$

i.e., as Δx grows small the "error" of the approximation, $F(\Delta x) - G(\Delta x)$ grows small a lot faster.

The equation (1) can also be written as

$$\lim_{\Delta x \to 0} \frac{\big[f(\overline{x} + \Delta x) - f(\overline{x}) \big] - a\Delta x}{\Delta x} = 0, \tag{2}$$

or as

$$\lim_{\Delta x \to 0} \frac{\big[f(\overline{x} + \Delta x) - f(\overline{x}) \big]}{\Delta x} = a. \tag{3}$$

We have to tie up one loose end here: the left-hand side of the equation (3), and also of (1) and (2), is the limit of a function of Δx. We need to know the limit exists, and that it's unique, in order to say that the coefficient a that we're looking for is this limit. In general, of course, the limit might or might not exist. So we say that a best linear approximation to the function F is the function $G(\Delta x) = a\Delta x$, where a is given by (3), *if the limit exists.* And if it does, we know it's unique, because we know that's a property of the limit of a function.

This motivates the definition of the derivative of a real function:

Definition: Let $f : \mathbb{R} \to \mathbb{R}$ and let $\overline{x} \in \mathbb{R}$. The **derivative** of f at $\overline{x}$, denoted $f'(\overline{x})$, is the number $a \in \mathbb{R}$ for which the function $G(\Delta x) = a\Delta x$ is a **best linear approximation** (**BLA**) of $F(\Delta x) := f(\overline{x} + \Delta x) - f(\overline{x})$ — *i.e.*,

$$f'(\overline{x}) := \lim_{\Delta x \to 0} \frac{\big[f(\overline{x} + \Delta x) - f(\overline{x}) \big]}{\Delta x} \tag{4}$$

if this limit exists, in which case we say that f is **differentiable** at $\overline{x}$.

So far, we've just been reviewing things you already know. Before we move ahead, let's go back to our example of $f(x) = x^2$ and see how all this works for that function.

Example 1: Let $f : \mathbb{R} \to \mathbb{R}$ be the function defined by $f(x) = x^2$.

Example 2: Let $f : \mathbb{R} \to \mathbb{R}$ be the function defined by $f(x) = e^x$.

I haven't typed up the examples yet.

There are two directions in which we need to generalize what we've done so far: (i) we need to study 2nd-order (quadratic) and higher-order approximations, and (ii) we need to do the same things for functions whose domain is $\mathbb{R}^n$ as we've done for functions with domain $\mathbb{R}$. We'll do the higher-order approximations first.

Taylor Polynomials

In the example in the previous section we suggested that in addition to a best linear approximation we could also define a best quadratic approximation to a function f, or to the function $F(\Delta x) = f(\overline{x}+\Delta x) - f(\overline{x})$. Here we're actually going to go farther and define a best approximation of order n, the best n-th degree polynomial approximation of F, for any $n \in \mathbb{N}$. As we did in the linear-approximation case, where $n = 1$, we start with the fact that $F(0) = 0$ — *i.e.*, $f(\overline{x} + \Delta x) = f(\overline{x})$ when $\Delta x = 0$. We're looking for the best n-th degree polynomial to approximate F, so we're looking for the best function

$$G_n(\Delta x) = a_1\Delta x + a_2(\Delta x)^2 + a_3(\Delta x)^3 + \cdots + a_n(\Delta x)^n. \tag{5}$$

Let's use the notation $G_n^{(k)}$, $F^{(k)}$, and $f^{(k)}$ to denote the k-th derivatives of the functions G_n, F, and f; and note that $F^{(k)}(\Delta x) = f^{(k)}(\overline{x} + \Delta x)$ — in particular, $F^{(k)}(0) = f^{(k)}(\overline{x})$.

By analogy with the case $n = 1$ we'll guess that for every n the best approximation of order n satisfies the condition that

$$G_n^{(n)}(\Delta x) = F^{(n)}(\Delta x) \ \text{ at } \Delta x = 0, \tag{6}$$

— *i.e.*, not only does the value of G equal the value of F at $\Delta x = 0$, but the first derivative (the slope) of the linear approximation $G_1(\cdot)$ has the same value as F' at $\Delta x = 0$ (*i.e.*, at $\overline{x}$); the second derivative (the curvature) of $G_2(\cdot)$ has the same value as $F''(\cdot)$ at $\Delta x = 0$; and so on, with $G_n^{(n)}(0) = F^{(n)}(0)$ for every n.

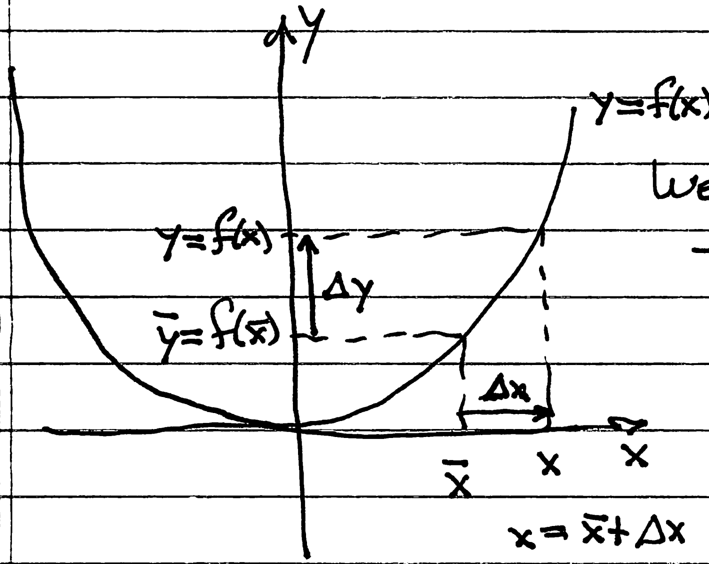
Combining (5) and (6) for each n, we have

$$a_1 = f'(\overline{x}),\ a_2 = \frac{1}{2}f''(\overline{x}),\ a_3 = \frac{1}{3}(\frac{1}{2})f'''(\overline{x}),\ \cdots,\ a_k = \frac{1}{k!}f^{(k)}(\overline{x}),\cdots,\ a_n = \frac{1}{n!}f^{(n)}(\overline{x}). \tag{7}$$

Exercise: Verify that (7) is correct — *i.e.*, that $a_k = \frac{1}{k!}f^{(k)}(\overline{x})$ for each $k = 1,\ldots,n$ in the function $G_n(\cdot)$ in (5).

The function G_n in Equation (5), with the coefficients as in (7), is called the homogeneous n-th degree **Taylor polynomial** of f, which approximates the increment Δy. The non-homogeneous form of the Taylor polynomial, which approximates the value of f at $x = \overline{x} + \Delta x$, is

$$\begin{aligned} P_n(x) &= f(\overline{x}) + G_n(\Delta x) \\ &= f(\overline{x}) + f'(\overline{x})\Delta x + \frac{1}{2}f''(\overline{x})(\Delta x)^2 + \frac{1}{6}f'''(\overline{x})(\Delta x)^3 + \cdots + \frac{1}{n!}f^{(n)}(\overline{x})(\Delta x)^n. \end{aligned}$$



WE WANT TO APPROXIMATE f(x)
– i.e., APPROXIMATE Δy IF WE KNOW f(x̄).

FIGURE 1

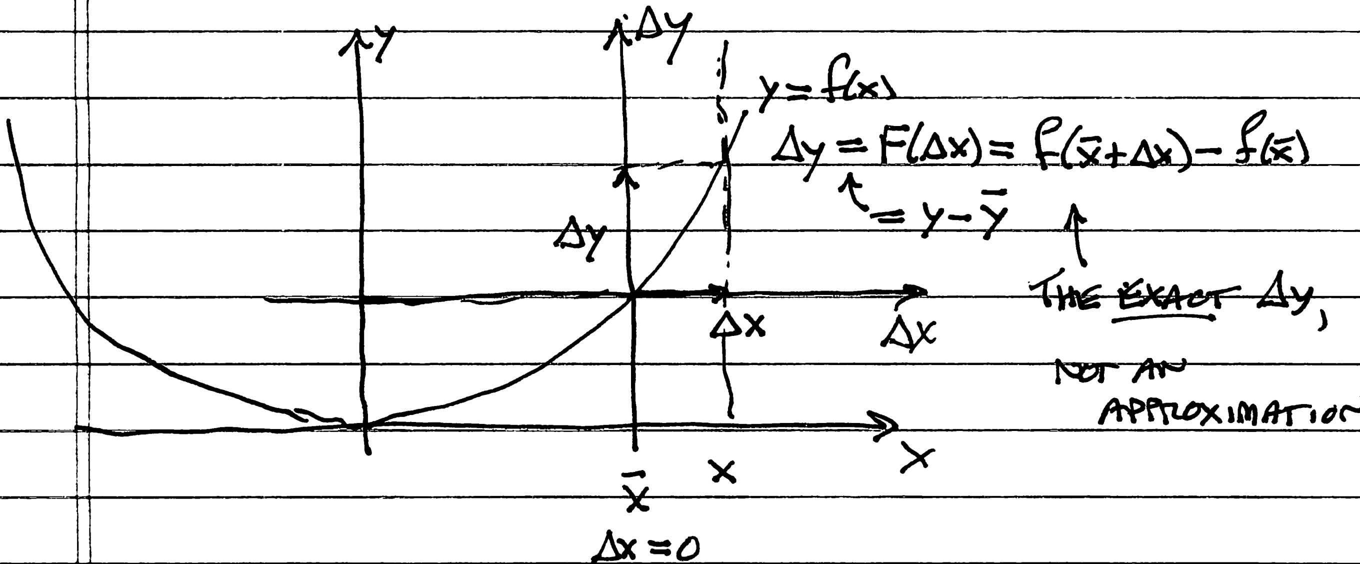


FIGURE 2

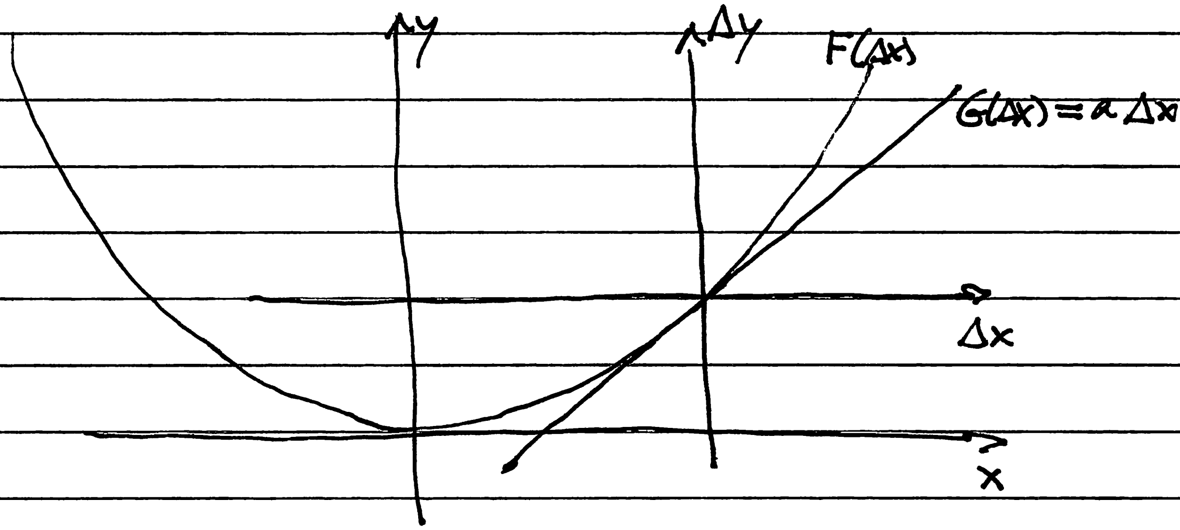


Figure 3

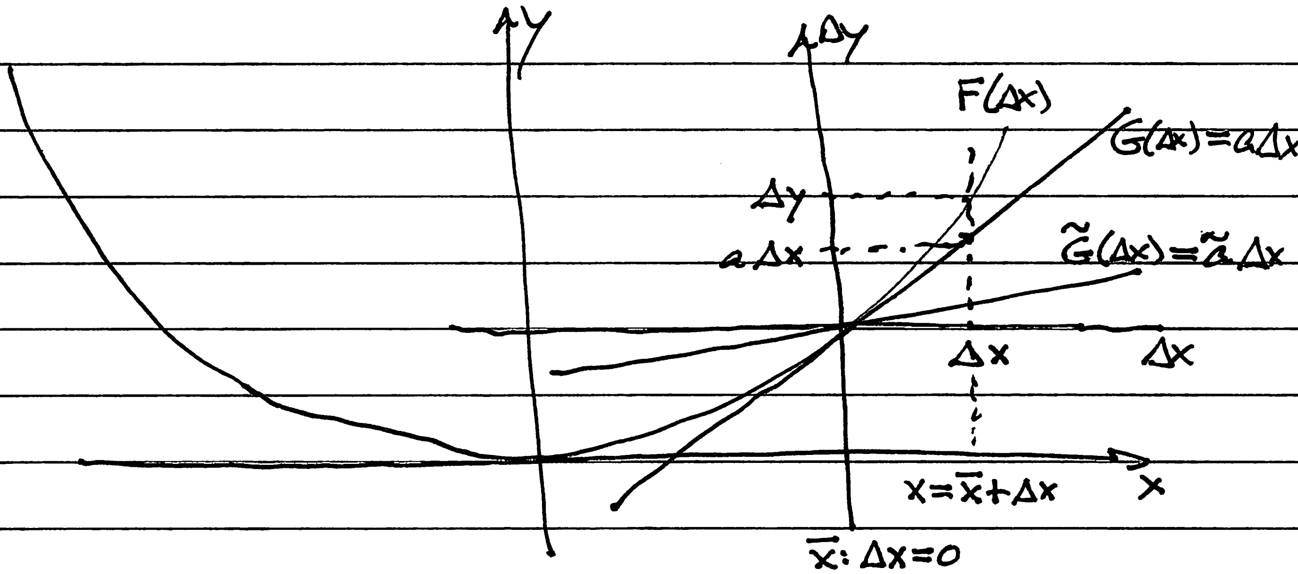


Figure 4