

FAME Week Three Class Notes – Derivatives II

Andie Creel

4th November, 2025

1 Maximum and minimum

If you're out hiking, how do you know if you're at the top of a hill? You will walk on flats at the top. We often want to know the peaks and valleys of slopes. When are we at the top? When are we at the bottom?

In applied setting we frequently want to know when we can maximize an objective or minimize an objective? How much time does a critter spend vigilant to maximize its chance of survival? How much money should a business invest in climate mitigation strategies to maximize their profit?

These questions are interesting on their own. These are the questions that motivate *why* we need to learn calculus, because these questions can be answered *using* calculus. Calculus isn't important on its own, it's important because it can help us answer important questions.

1.1 How do we know if we're going to a peak or a valley?

We can use calculus to find a peak or find a valley. Consider a mountain range that can be modeled with the equation

$$Y = F(x) = ax - bx^2$$
$$a > 0 \text{ and } b > 0$$

where is a peak/valley of this mountain range? At what level of x is $F(x)$ maximized or minimized? We can see when the derivative of the function equals zero, then solve for x . This value of x tells us where the function is "flat".

$$\frac{dF(x)}{dx} = a - 2bx = 0 \implies$$
$$x = \frac{a}{2b}$$

How do we know if $x = a/(2b)$ this is a maximum or minimum (peak or valley)? We can use the **second derivative**. If the second derivative is negative, it's a maximum. If the second derivative is positive, it's a minimum. (Use Eli's tricks of smiley faces and frowny faces to remember this. If the second derivative is negative, the function is a frowny. If the second derivative is positive, the function is a smiley).

A function is concave if the second derivative is negative. A function is convex if the second derivative is positive.

$$\frac{d^2F(x)}{dx^2} = -2b.$$

Assuming $b > 0$ implies $-2b < 0$ therefore we have a local maximum because the second derivative is negative.

Physicists do a good job keeping of functions, first derivatives and second derivatives. It's (respectively) location, speed, acceleration, jerk ("if someone's being really mean to you, you can say you're being a really big third derivative" - Eli).

2 Deriving the equation for taking the mean/average

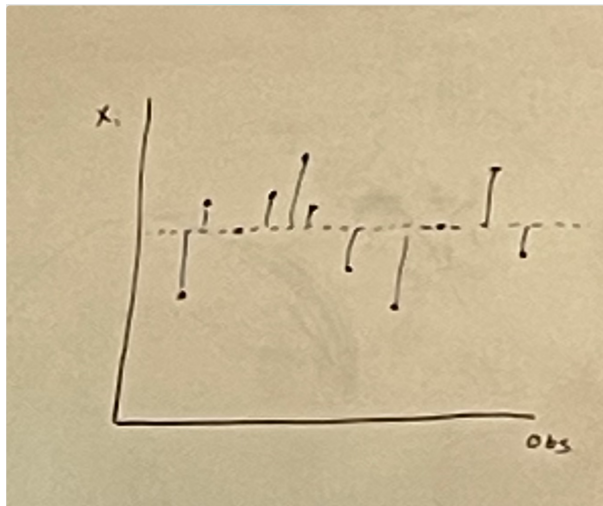


Figure 1: Dotted line is mean

How do we find an average? We want to find a constant, \bar{x} , that minimizes the total distance of all x_i values from \bar{x} , squared. Why are we squaring it? We don't care if x_i is greater than \bar{x} or less than \bar{x} . We just want to find a \bar{x} where all x_i are not that far away from \bar{x} .

Figure 1 shows all the x_i points as the dots and the dashed line as \bar{x} . Notice that the y-axis is x , and the x-axis is just showing us different observations, which we denote at i .

We can define the mean as the constant \bar{x} that minimizes the differences between itself and all other x_i observations,

$$\bar{x} = \mathbf{argmin} \sum_i^N (x_i - \bar{x})^2$$

Great, we have an equation defining the mean. However, this is not very easy to compute (how do you use an argmin?? IDK!). It's defined clearly, but we can't calculate it easily.

2.1 Deriving eqn for mean

Objective: Find the \bar{x} that minimizes the distance from itself and all other x_i , where the distance between two points x, y is measured as $(x - y)^2$.

To achieve this objective, we first need a function that measures the total distance of all x_i and \bar{x} . That function is written as:

$$F(\bar{x}) = \sum_i^N (x_i - \bar{x})^2 \quad (1)$$

$$= (x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + (x_3 - \bar{x})^2 + \dots + (x_N - \bar{x})^2 \quad (2)$$

If we're wanting to *minimize* (getting back to the argmin) the distance from x_i to \bar{x} , we can take a derivative of Equation 2 and set it equal to zero:

$$\frac{dF(\bar{x})}{d\bar{x}} = -2(x_1 - \bar{x}) - 2(x_2 - \bar{x}) - 2(x_3 - \bar{x}) - \dots - 2(x_N - \bar{x}) = 0$$

We can simplify by dividing both sides by -2 , then group the \bar{x} into one term,

$$x_1 + x_2 + x_3 + \dots x_n - N\bar{x} = 0 \quad (3)$$

$$x_1 + x_2 + x_3 + \dots x_n = N\bar{x} \quad (4)$$

simplify the left-hand side of 4 by changing it back to summation notation

$$\sum_i^N x_i = N\bar{x} \implies \quad (5)$$

$$\bar{x} = \frac{1}{N} \sum_i^N x_i \quad (6)$$

This is how the **arithmetic mean** was derived. When someone wanted to find an \bar{x} that was not that different from all the other x_i , they solved this problem!

To confirm that \bar{x} *minimizes* the distances between all x_i and itself, you can take a second derivative and confirm that the second derivative is positive, which means the function is convex and we've found the minimum of that convex function.

2.2 Problem set hint

The hardest problem on the problem set is a challenge problem asking us to derive the process for linear regression.

Linear regression finds a conditional mean *aka* a mean conditioned on x , which is just a line.

In the problem set, we will want to minimize the distance between y_i and $F(x_i)$,

$$\text{distance} = (y_i - F(x_i))^2. \quad (7)$$

What's $F(x)$ going to be? The previously mentioned conditional mean, *i.e.*, a line. In the problem set, we parameterize a line that minimizes the distance between the data points and the line. Solving this problem solves for the same algorithms that computers use when they fit linear regressions.

3 Mean Value Theorem

Definition of Mean Value Theorem: For a smooth and continuous function, there is at least one point on a given interval where the derivative is equal to the average rate of change over the interval.

Intuition: The theorem says that if you draw the secant line (the straight line) connecting the point $(a, f(a))$ (where a is on the x-axis and $f(x)$ is on the y-axis) and the point $(b, f(b))$, then somewhere in the interval (a, b) there will be a point c where the tangent line (the line that touches the curve at exactly one point and has the same slope as the curve at that point) is parallel to the secant line. In other words, at some point in the interval, the function's derivative matches the average rate of change over the interval,

$$f'(c) = \frac{f(b) - f(a)}{b - a}. \quad (8)$$

4 Taylor Series: A Polynomial Approximation

We know that straight lines are good approximators because a line is a conditional mean, and means are good at minimizing error.

Series are good approximators of sequences of numbers, because series are functions.

A **Taylor Series** is an extremely useful series for approximating the relationship of sequence of numbers (data is a sequence of numbers). Taylor series underpins a lot of the math we do today, particularly for any relationship that isn't linear.

$$f(x) = f(a) + \frac{f'(a)}{1!}(x-a)^1 + \frac{f''(a)}{2!}(x-a)^2 + \frac{f'''(a)}{3!}(x-a)^3 + \dots + \frac{f^k(a)}{k!}(x-a)^k + R \quad (9)$$

where R is the residual, or higher order terms.

4.1 0th order Taylor Series

Consider a 0th order taylor series: it would be a **mean** aka just a flat line equal to $f(a)$.

If you're only looking at the means of a dataset, that means you're doing a 0th order taylor approximation.

4.2 First order Taylor Series

Consider a first order taylor series:

$$f(x) = f(a) + \frac{f'(a)}{1!}(x-a)^1 \quad (10)$$

$$= A + B(x-a) \quad (11)$$

$$= A - aB + Bx \quad (12)$$

$$= m + Bx \quad (13)$$

It's a **line** with a slope. A first order Taylor series is a linear approximation. Linear regression is a first order Taylor approximation.

5 Mean Value Theorem and Taylor Series

Remember the Taylor series. A **Taylor Series** can approximate *any continuous function* $f(x)$. The modeler decides if they want to approximate $f(x)$ with a 0th order taylor series (mean), a 1st order taylor series (line), a 2nd order taylor series (parabola), a 3rd order taylor series (at this point, you better have a LOT of data!!). We can approximate $f(x)$ with our taylor series,

$$f(x) \approx f(a) + \frac{f'(a)}{1!}(x-a)^1 + \frac{f''(a)}{2!}(x-a)^2 + \frac{f'''(a)}{3!}(x-a)^3 + \dots + \frac{f^k(a)}{k!}(x-a)^k + R \quad (14)$$

where R is the residual.

Remember that what the **Mean Value Theorem** tells us: If you have two points, the slope between those two points will be parallel to the tangent line of the original function. So, the Taylor series is a large application of the mean value theorem, where a Taylor series approximates for the slope at a point many many times.

5.1 Discussion of models

Any Taylor Series approximation is a model. If someone says "I don't do models, let's only do means" they actually *are* suggesting a model. A mean is a model, it's a zero order Taylor Series approximation!

Higher order Taylor Series will approximate an underlying function better than a lower order. *However*, we do not always have enough data to fit a higher order Taylor series because it would require more estimating parameters and your data set may not have enough power (aka enough data) to do so.

5.2 Exercise: What is $\sqrt{10}$?

What's is $\sqrt{10}$? The function we are considering is

$$f(x) = \sqrt{x} = x^{\frac{1}{2}}. \quad (15)$$

We're interested when $x = 10$. Let's do a **0th order Taylor series approximation**:

$$f(x) = f(a) \quad (16)$$

what should we choose a to be? Let's choose $a = 9$ because 9 is close to 10.

$$f(10) \approx f(a) \quad (17)$$

$$\approx f(9) \quad (18)$$

$$\approx \sqrt{9} \quad (19)$$

$$\approx 3 \quad (20)$$

So the 0th order approximation of $f(10) \approx f(9) = 3$. It's close!

Now, let's do a **first order taylor series approximation**

$$f(x) \approx f(a) + \frac{f'(a)}{1!}(x - a)^1. \quad (21)$$

We need to find $f'(a)$,

$$f(a) = a^{1/2} \quad (22)$$

$$f'(a) = \frac{1}{2}a^{-1/2} \quad (23)$$

We can plug 23 back into 21 to get

$$f(x) \approx f(a) + \frac{1}{2} \frac{a^{-1/2}}{1} (x - a)^1 \quad (24)$$

Now, let's find $f(10)$ where $x = 10$ by using the taylor expansion around $f(9)$ where $a = 9$,

$$f(10) \approx f(9) + \frac{f'(9)}{1!}(10 - 9)^1 \quad (25)$$

$$\approx \sqrt{9} + \frac{1}{2}(9)^{-1/2} \times 1 \quad (26)$$

$$\approx 3 + \frac{1}{2} \frac{1}{(9)^{1/2}} \quad (27)$$

$$\approx 3\frac{1}{6} \quad (28)$$

So our first order approximation of $f(10) \approx 3\frac{1}{6}$. This is even closer than our original approximation, which was 3!

We can then do a **second order taylor series approximation** and get even closer.

$$f(x) \approx f(a) + \frac{f'(a)}{1!}(x - a)^1 + \frac{f''(a)}{2!}(x - a)^2 \quad (29)$$

we already found what $f'(a)$ is. What's $f''(a)$?

$$f(a) = a^{1/2} \quad (30)$$

$$f'(a) = \frac{1}{2}a^{-1/2} \text{ (power rule)} \quad (31)$$

$$f''(a) = -\frac{1}{4}a^{-3/2} \text{ (power rule again)} \quad (32)$$

We can plug in 32 into 29 (along with what we found in our first order approximation) to get

$$f(x) \approx f(a) + \frac{1}{2} \frac{a^{-1/2}}{1} (x - a)^1 + \frac{1}{2} \left(-\frac{1}{4}\right) a^{-3/2} (x - a)^2 \quad (33)$$

we know from our first order that the first two terms simplify to $3\frac{1}{6}$. Lets simplify this expression

$$f(10) \approx 3 + \frac{1}{6} - \frac{1}{8} \frac{1}{9^{3/2}} (10 - 9)^2 \quad (34)$$

$$= 3 + \frac{1}{6} - \frac{1}{8} \frac{1}{27} \quad (35)$$

$$= 3.16203703704 \quad (36)$$

So what is $f(10) = \sqrt{10}$? $\sqrt{10} = 3.16227766017$. So we can see that our second order approximation got pretty close!

5.3 Example: Logistic growth

Consider the logistic growth function:

$$G(N) = rN \left(1 - \frac{N}{K}\right) \quad (37)$$

What would we get if we "taylor expanded" this function? Let's work with the per capital growth rate

$$\frac{G(N)}{N} = r \left(1 - \frac{N}{K}\right) \quad (38)$$

and see if we can use a Taylor series that would approximate this right hand side equation

$$\frac{G(N)}{N} \approx G(a) + G'(a)(N - a) \quad (39)$$

$$\approx G(a) + G'(a)N - G'(a)a \quad (40)$$

$$\approx G(a) - G'(a)a + G'(a)N \quad (41)$$

Let $r = -G'(a)a$.

$$\approx r + G(a) + G'(a)N \quad (42)$$

Let $G(a) = 0$ and rename $G'(a) = \frac{-r}{K}$

$$\approx r + \frac{-r}{K}N \quad (43)$$

$$\approx r(1 - \frac{N}{K}) \quad (44)$$

And now we've shown that 44 is equivalent to 38. Which means we've shown that we can derive the logistic growth equation using a Taylor series approximation.

6 Partial derivatives

Consider a function that maps from $R^2 \rightarrow R^1$,

$$Z = F(x, y). \quad (45)$$

Consider a situation where we can change x with policy, but we can't change y and we want to maximize outcome Z . In this case, we would take the derivative of $F(x, y)$ w.r.t x (treat y like a constant, assume it isn't changing).

We can take a **partial derivative** which notation that are all equivalent:

$$\frac{\partial F(x, y)}{\partial x} = F_x(x, y) = F_1(x, y) \quad (46)$$

This is how we find a **relative max/min**. Conditioned on y staying constant, we can find the relative max/min (relative to that level of y).

6.1 Absolute max/min

If you want to find the absolute maximum or minimum you need to take *all* partial derivatives and set them all equal to zero and solve that system of equations. If your function mapped from $R^N \rightarrow R^1$ then you would take N partial derivatives and solve the system of N equations.

7 Total Derivatives

Total derivative in a function is equal to the sum of the partials multiplied by the change.

Consider the function $F = f(x, y)$. We can get our partial derivatives. Recall that when we take a partial derivative wrt x , we're holding y constant.

$$\frac{\partial f}{\partial x} = f_x \quad (47)$$

$$\frac{\partial f}{\partial y} = f_y \quad (48)$$

Now consider how we would calculate the change in F if both x and y changed. We'd need to get a full derivative,

$$df = f_x dx + f_y dy \quad (49)$$

$$= \frac{\partial f}{\partial x} dx + \frac{\partial f}{\partial y} dy \quad (50)$$

$$\Delta f = \frac{\partial f}{\partial x} \Delta x + \frac{\partial f}{\partial y} \Delta y \quad (51)$$

These are all equivalent ways of writing the full derivative.

7.1 Example of maximizing a multi-variate function

Consider the functions $F(x, y) = -x^2 - y^2$. Let's say that we want to maximize this function, and find the *global* maximum. To do so, we need to take the partial derivatives and set them equal to zero, then solve for x, y .

$$\frac{\partial F}{\partial x} = -2x = 0 \implies x = 0 \quad (52)$$

$$\frac{\partial F}{\partial y} = -2y = 0 \implies y = 0 \quad (53)$$

So the maximum point of $F(x, y) = -x^2 - y^2$ is $(0, 0)$, where $x = 0$ and $y = 0$.

7.2 Implicit Function Theorem (envelope theorem in econ)

Now consider if you want to stay on one level set. For example, if you want to stay on a contour line or on a utility level. That would mean we want the value of $f(x, y)$ to stay the same, $df = \Delta f = 0$, even when we change both the x and y . We can set the full derivative to 0 in order to achieve this.

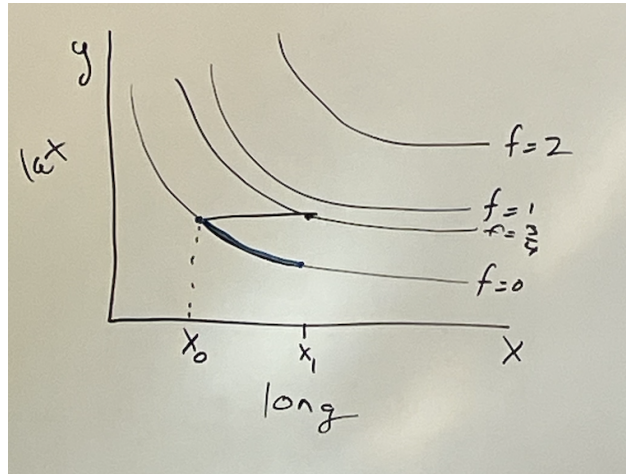


Figure 2: The $f=0$ represents an altitude

$$df = \frac{\partial f}{\partial x} dx + \frac{\partial f}{\partial y} dy = 0 \implies \quad (54)$$

$$dy \frac{\partial f}{\partial y} = -dx \frac{\partial f}{\partial x} \implies \quad (55)$$

$$\frac{dy}{dx} = \frac{-\frac{\partial f}{\partial x}}{\frac{\partial f}{\partial y}} \quad (56)$$

Equation 56 is the implicit function theorem. What equation 56 tells us is "if I want to stay at the original value of $f(x, y)$, and I change x , how do I need to change y in order to stay at the original value of f ". You could also think of this as you're hiking and you're at a certain altitude. The same questions would be "if I change my latitude, how do I need to change my longitude in order to stay at the same altitude?"

The envelope theorem is a special case of the implicit function theorem when first derivatives are set to zero. This leads to math simplifying because we can say that, at the optimum, a bunch of derivatives will equal zero.