Lecture 7
Taylor Series and the Hessian Matrix

Saurabh

June 16, 2025

Contents

1 The Building Blocks: Why Approximation Works

Before we can build a skyscraper, we need to understand the bricks. In our case, before building the powerful Taylor series, we need to understand two fundamental theorems in calculus: Rolle's Theorem and the Mean Value Theorem (MVT). They provide the logical foundation for why the Taylor series isn't just a random formula, but something that is guaranteed to work.

1.1 Rolle's Theorem: The "Flat Spot" Guarantee

The Mountain Trail Analogy

!

Imagine you're hiking on a smooth, continuous mountain trail. You start at an elevation of 1000m and, after some time, you end your hike back at the same elevation of 1000m.

Rolle's Theorem simply states that at some point during your hike, you must have been on a perfectly flat piece of ground. Your path could have gone up and then down, or down and then up, but to get back to the starting elevation, there must be at least one point where the slope was exactly zero. This single guaranteed "flat spot" is the crucial first piece of our puzzle.

Rolle's Theorem

Let f be a function that satisfies three conditions:

1. f is continuous on the closed interval $[a, b]$. (The trail has no sudden gaps.)
2. f is differentiable on the open interval (a, b). (The trail is smooth, with no sharp corners.)
3. $f(a) = f(b)$. (You start and end at the same height.)

Then, there exists at least one number c in (a, b) such that $f'(c) = 0$.

This theorem is the seed from which the Taylor series grows. The proof of the full Taylor's theorem cleverly constructs a special function that meets the conditions for Rolle's Theorem, guaranteeing a certain property that helps us build our approximation step-by-step.

1.2 The Mean Value Theorem (MVT): Average vs. Instantaneous

The Mean Value Theorem is a tilted version of Rolle's Theorem. Instead of starting and ending at the same height, we now allow them to be different.

The Highway Trip Analogy

!

Imagine you drive 200 km in 2 hours. Your average speed was 100 km/h. The Mean Value Theorem guarantees that at some specific moment during your trip, your speedometer must have read exactly 100 km/h. You might have gone faster at times and slower at others, but at some point, your instantaneous speed must have equaled your average speed. The MVT connects the overall, average behavior to a specific, instantaneous moment.

Mean Value Theorem (MVT)

Let f be a function that satisfies two conditions:

1. f is continuous on the closed interval $[a, b]$.
2. f is differentiable on the open interval (a, b).

Then, there exists a number c in (a, b) such that:

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

The left side is the **instantaneous rate of change** at some point c. The right side is the **average rate of change** over the whole interval.

Connecting MVT to Taylor Series: This is where the magic begins. Let's rearrange the MVT formula:

$$f(b) = f(a) + f'(c)(b - a)$$

This looks suspiciously like the first-order Taylor expansion! It says we can find the value of a function at point b by taking its value at point a and adding a "correction term". This term is the distance $(b - a)$ multiplied by the slope at some *unknown* point c in between. The Taylor series is the result of asking: "What if we made this more precise by using the slope *at point* a, and then added more correction terms for curvature using the second, third, and higher derivatives?"

2 The Magic of Approximation: Taylor Series in 1D

2.1 Intuition: Building a "Function Clone"

Imagine you have a very complex, curvy function. Its formula might be a nightmare to work with directly. A Taylor series allows us to create a simpler "clone" of the function that is almost identical, at least around a specific point of interest. This clone is a polynomial, which is built from simple additions and multiplications, making it much easier to handle.

The Master Forger Analogy

!

A master forger wants to replicate a complicated signature on a document. They can't see the whole signature at once, but they can analyze it deeply at its starting point.

- **Zeroth-Order Approx. (The Position):** They put a dot where the signature starts. This matches the position, $f(a)$. It's a poor copy, but it's in the right place.
- **First-Order Approx. (The Slope):** They draw a short, straight line that matches the starting angle (slope) of the signature, $f'(a)$. This linear copy is much better for a tiny distance.
- **Second-Order Approx. (The Curvature):** They adjust their line to have the same initial bend or curve as the signature, using $f''(a)$. This creates a parabola that's a much more convincing copy, hugging the real signature for longer.
- **Higher-Order Approximations:** The forger gets even more precise, matching how the curvature itself is changing ($f'''(a)$), and so on. Each new derivative adds another layer of sophistication to the polynomial forgery.

The Taylor series is the ultimate forger, using all the derivative information at a single point to build a perfect polynomial clone of the function in that neighborhood.

Taylor's Theorem (with Remainder)

If a function f is differentiable n times in an interval containing a point a, then for any x in that interval, we can write:

$$f(x) = \underbrace{\sum_{k=0}^{n-1} \frac{f^{(k)}(a)}{k!}(x-a)^k}_{\text{Taylor Polynomial } P_{n-1}(x)} + \underbrace{R_n(x)}_{\text{Remainder}}$$

The remainder term, derived from repeatedly applying Rolle's Theorem, tells us the "error" of our approximation:

$$R_n(x) = \frac{f^{(n)}(c)}{n!}(x-a)^n \quad \text{for some number } c \text{ between } a \text{ and } x.$$

2.2 Examples: Taylor Series in Action

Example 1: Approximating $f(x)$

The Goal: Create a polynomial approximation for $f(x) = e^x$ centered at $a = 0$.
The Derivatives: The function $f(x) = e^x$ is special because its derivative is always itself.

$$\begin{aligned} f(x) &= e^x & \implies \quad & f(0) = 1 \\ f'(x) &= e^x & \implies \quad & f'(0) = 1 \\ f''(x) &= e^x & \implies \quad & f''(0) = 1 \end{aligned}$$

All derivatives at $a = 0$ are 1.
The Construction: Plugging these into the Taylor series formula (also called a Maclaurin series when $a = 0$):

$$\begin{aligned} e^x &\approx \frac{f(0)}{0!}x^0 + \frac{f'(0)}{1!}x^1 + \frac{f''(0)}{2!}x^2 + \frac{f'''(0)}{3!}x^3 + \dots \\ e^x &\approx \frac{1}{1} + \frac{1}{1}x + \frac{1}{2}x^2 + \frac{1}{6}x^3 + \dots \\ e^x &= 1 + x + \frac{x^2}{2} + \frac{x^3}{6} + \dots \end{aligned}$$

This polynomial is a fantastic approximation of e^x for x close to 0.

Example 2: Approximating $f(x)$

The Goal: Create a polynomial approximation for $f(x) = \cos(x)$ centered at $a = 0$.
The Derivatives: Let's find the derivatives and evaluate them at $a = 0$:

$$\begin{aligned} f(x) &= \cos(x) & \implies \quad & f(0) = 1 \\ f'(x) &= -\sin(x) & \implies \quad & f'(0) = 0 \\ f''(x) &= -\cos(x) & \implies \quad & f''(0) = -1 \\ f'''(x) &= \sin(x) & \implies \quad & f'''(0) = 0 \\ f^{(4)}(x) &= \cos(x) & \implies \quad & f^{(4)}(0) = 1 \end{aligned}$$

The pattern of values at $a = 0$ is $1, 0, -1, 0, 1, \dots$
The Construction: The Taylor series only includes the non-zero terms:

$$\begin{aligned} \cos(x) &\approx \frac{f(0)}{0!}x^0 + \frac{f''(0)}{2!}x^2 + \frac{f^{(4)}(0)}{4!}x^4 + \dots \\ \cos(x) &= 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \dots \end{aligned}$$

This shows that for small angles x, $\cos(x) \approx 1 - x^2/2$, a very useful approximation in physics and engineering.

3 Expanding to a Hilly Landscape: Taylor Series in 2D

What if our function doesn't describe a 1D road, but a 2D surface, like a hilly landscape defined by $z = f(x, y)$? How do we approximate the height of the landscape around a point (a, b)?

The core idea is the same, but now we use partial derivatives. The first-order approximation uses the partial derivatives $f_x(a, b)$ and $f_y(a, b)$ to define a **tangent plane** to the surface at (a, b). For a better fit, the second-order approximation uses the second partial derivatives (f_{xx}, f_{yy}, f_{xy}) to capture the **curvature** of the surface—is it shaped like a bowl, a dome, or a saddle?

Let's look at the second-order Taylor expansion around a critical point (a, b), where the tangent plane is flat ($f_x(a, b) = 0$ and $f_y(a, b) = 0$). Let $x = a + h$ and $y = b + k$.

$$f(a+h, b+k) \approx f(a, b) + \underbrace{hf_x(a, b) + kf_y(a, b)}_{\text{First-order term} = 0 \text{ at a critical point}} + \frac{1}{2!}\left(h^2 f_{xx} + 2hk f_{xy} + k^2 f_{yy}\right)$$

This simplifies to a crucial relationship:

$$f(a+h, b+k) - f(a, b) \approx \frac{1}{2}\left(h^2 f_{xx}(a, b) + 2hk f_{xy}(a, b) + k^2 f_{yy}(a, b)\right)$$

The left side is the change in height as we move away from the critical point. The right side is a **quadratic form** whose sign determines the nature of the critical point. This quadratic form is the key to understanding extrema, and it leads us directly to the Hessian matrix.

4 The Judge of Extrema: The Hessian Matrix

The quadratic form from the 2D Taylor expansion looks complicated. The Hessian matrix is a tool to organize these second partial derivatives and analyze them systematically.

The Hessian Matrix

For a two-variable function $f(x, y)$, the Hessian matrix is the matrix of its second partial derivatives:

$$H(x, y) = \nabla^2 f(x, y) = \begin{pmatrix} f_{xx} & f_{xy} \\ f_{yx} & f_{yy} \end{pmatrix}$$

Since for most well-behaved functions $f_{xy} = f_{yx}$ (Clairaut's Theorem), the Hessian is a symmetric matrix.

The question "Is (a, b) a local maximum or minimum?" boils down to the question: "Is the surface curving up in all directions, down in all directions, or both?" The Hessian matrix holds the answer.

4.1 The Second Derivative Test

Let (a, b) be a critical point where $\nabla f(a, b) = \mathbf{0}$. We evaluate the Hessian at this point, $H(a, b)$, and compute its determinant, which we'll call D.

$$D = \det(H) = f_{xx} f_{yy} - f_{xy}^2$$

The Bowl

At a critical point, you are on flat ground. The Hessian tells you what kind of landscape is immediately around you. Think of the sign of the Taylor approximation's quadratic term.

!

- **Local Minimum:** You are at the bottom of a bowl. Any step you take leads uphill. The surface is "concave up" in all directions. This corresponds to the quadratic term being positive for all non-zero (h, k), which happens when $D > 0$ and $f_{xx} > 0$.
- **Local Maximum:** You are at the top of a hill or dome. Any step you take leads downhill. The surface is "concave down" in all directions. This corresponds to the quadratic term being negative, which happens when $D > 0$ and $f_{xx} < 0$.
- **Saddle Point:** You are on a mountain pass or a Pringles chip. In some directions you go up, and in others you go down. The quadratic term can be positive or negative depending on the direction. This happens when $D < 0$.
- **Inconclusive:** The test fails. The landscape might be perfectly flat, or have a more complex shape like a trough, that our test can't classify. This happens when $D = 0$.

The Second Derivative Test for Extrema

Let (a, b) be a critical point of $f(x, y)$. Let $D = f_{xx}(a, b)f_{yy}(a, b) - [f_{xy}(a, b)]^2$.

1. If $D > 0$ and $f_{xx}(a, b) > 0$, then $f(a, b)$ is a **local minimum**.
2. If $D > 0$ and $f_{xx}(a, b) < 0$, then $f(a, b)$ is a **local maximum**.
3. If $D < 0$, then $f(a, b)$ is a **saddle point**.
4. If $D = 0$, the test is inconclusive.

4.2 Hessian Problems and Solutions

Problem 1: Finding Local Extrema

Problem: Find the local extreme values of the function $f(x, y) = xy - x^2 - y^2 - 2x - 2y + 4$.

Step 1: Find Critical Points. We set the first partial derivatives to zero.

$$f_x = y - 2x - 2 = 0$$
$$f_y = x - 2y - 2 = 0$$

From the first equation, $y = 2x + 2$. Substituting this into the second gives: $x - 2(2x + 2) - 2 = 0 \implies x - 4x - 4 - 2 = 0 \implies -3x = 6 \implies x = -2$. Then $y = 2(-2) + 2 = -2$. The only critical point is $(-2, -2)$.

Step 2: Compute Second Partial Derivatives.

$$f_{xx} = -2$$
$$f_{yy} = -2$$
$$f_{xy} = 1$$

These are constant, so they have these values at $(-2, -2)$.

Step 3: Apply the Second Derivative Test. We calculate the determinant of the Hessian, D:

$$D = f_{xx} f_{yy} - f_{xy}^2 = (-2)(-2) - (1)^2 = 4 - 1 = 3$$

Conclusion: Since $D = 3 > 0$ and $f_{xx} = -2 < 0$, the function has a **local maximum** at $(-2, -2)$. The maximum value is $f(-2, -2) = 8$.

Problem 2: Finding Local Extrema

Problem: Find the local extreme values of $f(x, y) = x^3 + y^3 - 3xy + 1$.
Step 1: Find Critical Points.

$$f_x = 3x^2 - 3y = 0 \implies y = x^2$$
$$f_y = 3y^2 - 3x = 0 \implies x = y^2$$

Substitute $y = x^2$ into the second equation: $x = (x^2)^2 = x^4$. $x^4 - x = 0 \implies x(x^3 - 1) = 0$. This gives $x = 0$ or $x = 1$.

- If $x = 0$, then $y = 0^2 = 0$. Critical Point: $(0, 0)$.
- If $x = 1$, then $y = 1^2 = 1$. Critical Point: $(1, 1)$.

Step 2: Compute Second Partial Derivatives. $f_{xx} = 6x$, $f_{yy} = 6y$, $f_{xy} = -3$.
Step 3: Apply the test at each critical point.

Analysis of (0,0): $f_{xx}(0, 0) = 0$, $f_{yy}(0, 0) = 0$, $f_{xy}(0, 0) = -3$. $D = (0)(0) - (-3)^2 = -9$. Since $D < 0$, the point **(0,0) is a saddle point**.

Analysis of (1,1): $f_{xx}(1, 1) = 6$, $f_{yy}(1, 1) = 6$, $f_{xy}(1, 1) = -3$. $D = (6)(6) - (-3)^2 = 36 - 9 = 27$. Since $D > 0$ and $f_{xx} > 0$, the point **(1,1) is a local minimum**.

Problem 3: Finding Absolute Extrema

Problem: Find the absolute max/min values of $f(x, y) = x^2 + y^2 - x - y$ in the region $D = \{(x, y) : x^2 \le y \le 1\}$.

Step 1: Find Interior Critical Points. $f_x = 2x - 1 = 0 \implies x = 1/2$. $f_y = 2y - 1 = 0 \implies y = 1/2$. The critical point is $(1/2, 1/2)$. Check if it's in D: $(1/2)^2 \le 1/2 \le 1$ is $1/4 \le 1/2 \le 1$, which is true. So $(1/2, 1/2)$ is an interior candidate. $f(1/2, 1/2) = -1/2$.

Step 2: Check the Boundary of the Region.

Boundary Part 1: The Parabola $y = x^2$ for $-1 \le x \le 1$. Substitute $y = x^2$ into f: $g(x) = f(x, x^2) = x^2 + (x^2)^2 - x - x^2 = x^4 - x$. Find critical points for $g(x)$: $g'(x) = 4x^3 - 1 = 0 \implies x = (1/4)^{1/3} \approx 0.63$. This gives a candidate point $(0.63, 0.63^2) \approx (0.63, 0.40)$. $f(0.63, 0.40) \approx -0.47$. We must also check the endpoints of the parabola's domain: $x = -1$ and $x = 1$.

- At $x = -1$, $y = (-1)^2 = 1$. Point is $(-1, 1)$. $f(-1, 1) = 1 + 1 - (-1) - 1 = 2$.
- At $x = 1$, $y = 1^2 = 1$. Point is $(1, 1)$. $f(1, 1) = 1 + 1 - 1 - 1 = 0$.

Boundary Part 2: The Line $y = 1$ for $-1 \le x \le 1$. Substitute $y = 1$ into f: $h(x) = f(x, 1) = x^2 + 1 - x - 1 = x^2 - x$. Find critical points for $h(x)$: $h'(x) = 2x - 1 = 0 \implies x = 1/2$. This gives a candidate point $(1/2, 1)$. $f(1/2, 1) = (1/4) + 1 - 1/2 - 1 = -1/4$. (The endpoints $x = -1, x = 1$ have already been checked).

Step 3: Compare All Candidate Values. Our list of candidate values is: $\{-1/2, -0.47, 2, 0, -1/4\}$. The largest is 2. The smallest is -1/2.

Conclusion: The **absolute maximum is 2**, which occurs at the boundary point $(-1, 1)$. The **absolute minimum is -1/2**, which occurs at the interior point $(1/2, 1/2)$.