LECTURE 8

Derivatives, Taylor's Theorem, and Analyticity

MATH S4061 | Introduction to Modern Analysis I

Tuesday, 23 June 2026

And what are these fluxions? The velocities of evanescent increments.

George Berkeley, The Analyst (1734)

Contents of this lecture

Overview. This lecture begins the differential part of the course. The derivative is first defined by the difference quotient, then immediately reinterpreted as the best first-order linear approximation. That viewpoint makes the usual rules of differentiation bookkeeping in powers of the increment, and it also explains the multivariable definition. The lecture then turns the derivative into a global tool through Fermat's interior-extremum theorem, Rolle's theorem, the Mean Value Theorem, monotonicity, and Darboux's theorem. It closes with Taylor's theorem and the distinction between smooth and analytic functions, first over $\mathbb{R}$ and then over $\mathbb{C}$. The companion text is Rudin, Chapter 5, with the closest matches marked by the locators R 5.1–R 5.16.

8.1 Derivatives as First-Order Approximations

Definition 8.1.1

Derivative at a point

Let $f : (a,b) \to \mathbb{R}$ and let $x_0 \in (a,b)$. The function f is *differentiable at* x_0 (R 5.1) if the limit

$$\lim_{x\to x_0} \frac{f(x)-f(x_0)}{x-x_0} = \lim_{h\to 0} \frac{f(x_0+h)-f(x_0)}{h}$$

exists. This limit is the *derivative* of f at x_0, denoted $f'(x_0)$.[1]

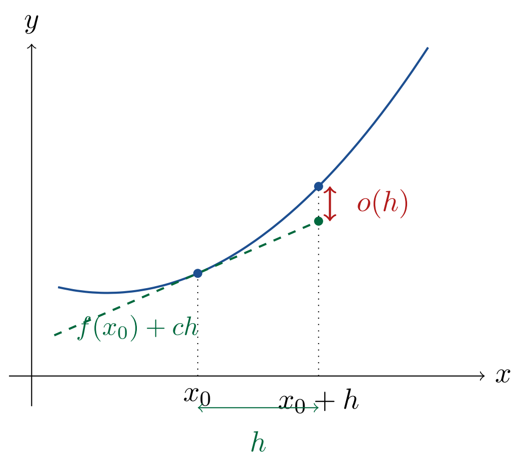

Figure 8.1.2. BEST LINEAR APPROXIMATION. Near x_0, differentiability says the graph of f is shadowed by an affine line whose error is smaller than the displacement.

Proposition 8.1.3

Differentiability as a first-order expansion

For $f : (a,b) \to \mathbb{R}$ and $x_0 \in (a,b)$, the following are equivalent:

$$f \text{ is differentiable at } x_0 \qquad \Longleftrightarrow \qquad f(x_0+h) = f(x_0) + ch + o(h)$$

for some $c \in \mathbb{R}$, where $o(h)$ means an error $E(h)$ satisfying

$$\frac{E(h)}{h} \longrightarrow 0 \qquad (h \to 0).$$

In that case $c = f'(x_0)$. This is Rudin's derivative definition R 5.1 written in the increment variable.

[1] The difference quotient is how you *compute* a derivative; the picture behind it is what to keep. Zoom in on the graph at x_0 and it looks more and more like a straight line, and $f'(x_0)$ is that line's slope—the limit of the chord slopes $(f(x) - f(x_0))/(x - x_0)$ as the far point slides in. 8.1.3 says this outright: to be differentiable is to be matched by *some* straight line with an error smaller than the step h. It is that reading, not the quotient, that survives into $\mathbb{R}^n$ (8.1.4), where one cannot divide by a vector.

Proof 8.1.3

Proof by Rearranging the Difference Quotient.

(1) Suppose f is differentiable at x_0. Define

$$E(h) = f(x_0 + h) - f(x_0) - f'(x_0)h.$$

Then the desired expansion holds with $c = f'(x_0)$, and for $h \neq 0$,

$$\frac{E(h)}{h} = \frac{f(x_0 + h) - f(x_0)}{h} - f'(x_0) \longrightarrow 0.$$

(2) Conversely, if $f(x_0 + h) = f(x_0) + ch + E(h)$ and $E(h)/h \to 0$, then

$$\frac{f(x_0 + h) - f(x_0)}{h} = c + \frac{E(h)}{h} \longrightarrow c.$$

Hence the difference quotient converges and $f'(x_0) = c$.

∎

[Direct] increment form of R 5.1.

Definition 8.1.4

Differentiability from $\mathbb{R}^n$ to $\mathbb{R}^m$

Let $F : \mathbb{R}^n \to \mathbb{R}^m$ and let $x_0 \in \mathbb{R}^n$. The function F is *differentiable at* x_0 if there exists a linear map $A : \mathbb{R}^n \to \mathbb{R}^m$ such that

$$F(x_0 + h) = F(x_0) + A(h) + E(h), \qquad \frac{\|E(h)\|}{\|h\|} \longrightarrow 0 \quad (h \to 0).$$

The linear map A is unique; in the standard bases it is represented by the Jacobian matrix. Compare Rudin's one-parameter vector-valued derivative in R 5.16; the present definition replaces one real increment by a vector increment.

Remark 8.1.5

Why the linear map replaces division

In one real variable, the quotient

$$\frac{f(x_0 + h) - f(x_0)}{h}$$

is available because nonzero real increments have inverses. A vector increment $h \in \mathbb{R}^n$ has no comparable division operation, so the higher-dimensional definition subtracts a candidate linear map and asks that the remaining vector error be smaller than $\|h\|$. Complex differentiability is a special two-dimensional exception: identifying $\mathbb{R}^2$ with $\mathbb{C}$ restores division by a nonzero increment, but imposes a much stronger condition than ordinary real differentiability. Compare the vector-valued one-parameter derivative in R 5.16. More basically still, the difference quotient $[f(x) - f(x_0)]/(x - x_0)$ already presumes that the domain and target of f are one and the same complete field—for instance $\mathbb{R}$ or $\mathbb{C}$—so that subtraction and division are both available.

8.2 Rules for Derivatives

Theorem 8.2.1

Differentiability implies continuity

If f is differentiable at x_0, then f is continuous at x_0 (R 5.2).

Proof 8.2.1

Proof from the First-Order Expansion.

(1) By 8.1.3,

$$f(x_0 + h) = f(x_0) + f'(x_0)h + E(h), \qquad \frac{E(h)}{h} \to 0.$$

(2) Since $E(h) = h(E(h)/h)$, we have $E(h) \to 0$. Therefore

$$f(x_0 + h) - f(x_0) = f'(x_0)h + E(h) \longrightarrow 0,$$

which is continuity at x_0.

■

[Direct] cf. R 5.2.

Corollary 8.2.2

Discontinuity obstructs differentiability

If f is not continuous at x_0, then f is not differentiable at x_0. This is the contrapositive of Rudin's differentiability-implies-continuity theorem R 5.2.

Proof 8.2.2

Proof by Contrapositive.

(1) The contrapositive of 8.2.1 says precisely that differentiability at x_0 forces continuity at x_0.

(2) Therefore a failure of continuity at x_0 rules out differentiability there.

■

[A9]

Proposition 8.2.3

Algebra rules for derivatives

Suppose f and g are differentiable at x_0. Then

$$(f \pm g)'(x_0) = f'(x_0) \pm g'(x_0), \qquad (cf)'(x_0) = c\, f'(x_0),$$

$$(fg)'(x_0) = f'(x_0)g(x_0) + f(x_0)g'(x_0).$$

If $g(x_0) \neq 0$, then f/g is differentiable at x_0 and

$$\left(\frac{f}{g}\right)'(x_0) = \frac{f'(x_0)g(x_0) - f(x_0)g'(x_0)}{g(x_0)^2}.$$

These are Rudin's algebra rules for derivatives (R 5.3).

Proof 8.2.3

Proof by First-Order Bookkeeping.

(1) Write

$$f(x_0+h) = f(x_0) + f'(x_0)h + E_f(h), \qquad g(x_0+h) = g(x_0) + g'(x_0)h + E_g(h),$$

with $E_f(h)/h \to 0$ and $E_g(h)/h \to 0$.

(2) Adding the expansions gives

$$(f+g)(x_0+h) = (f+g)(x_0) + \big(f'(x_0)+g'(x_0)\big)h + \big(E_f(h)+E_g(h)\big),$$

and the last error is $o(h)$ because

$$\frac{E_f(h)+E_g(h)}{h} = \frac{E_f(h)}{h} + \frac{E_g(h)}{h} \longrightarrow 0.$$

The difference and scalar-multiple rules are the same calculation.

(3) Multiplying the expansions and collecting the constant and first-order terms gives

$$\begin{aligned}(fg)(x_0+h) = f(x_0)g(x_0) &+ f'(x_0)g(x_0)h + f(x_0)g'(x_0)h \\ &+ E_f(h)\big(g(x_0)+g'(x_0)h+E_g(h)\big) \\ &+ E_g(h)\big(f(x_0)+f'(x_0)h\big) + f'(x_0)g'(x_0)h^2.\end{aligned}$$

After division by h, the last line tends to 0 because the error ratios tend to 0 and the bracketed factors remain bounded. Hence the coefficient of h is the derivative of fg.

(4) If $g(x_0) \neq 0$, then by 8.2.1 the values $g(x_0+h)$ stay nonzero for all sufficiently small h. Thus

$$\begin{aligned}\left(\frac{1}{g}\right)'(x_0) &= \lim_{h\to 0} \frac{1}{h}\left(\frac{1}{g(x_0+h)} - \frac{1}{g(x_0)}\right) \\ &= -\lim_{h\to 0} \frac{g(x_0+h)-g(x_0)}{h} \cdot \frac{1}{g(x_0+h)g(x_0)} = -\frac{g'(x_0)}{g(x_0)^2}.\end{aligned}$$

Applying the product rule to $f \cdot (1/g)$ gives the quotient rule.

■

[Direct] cf. R 5.3.

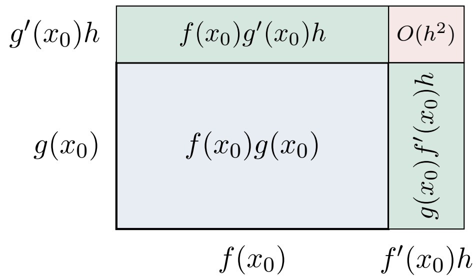

Figure 8.2.4. THE PRODUCT RULE. The two first-order strips contribute to the derivative of the area; the corner is second-order.

Proposition 8.2.5

Chain rule

Suppose f is differentiable at x_0 and g is differentiable at $f(x_0)$. Then $g \circ f$ is differentiable at x_0 and

$$(g \circ f)'(x_0) = g'(f(x_0))f'(x_0).$$

This is the chain rule (R 5.5).

Proof 8.2.5

Proof by Composing First-Order Errors.

(1) Let

$$k(h) = f(x_0 + h) - f(x_0).$$

Differentiability of f gives $k(h) = f'(x_0)h + E_1(h)$, with $E_1(h)/h \to 0$; in particular $k(h) \to 0$.

(2) Differentiability of g at $f(x_0)$ means there is an error $E_2(u)$ such that

$$g(f(x_0) + u) = g(f(x_0)) + g'(f(x_0))u + E_2(u), \qquad \frac{E_2(u)}{u} \to 0.$$

To avoid dividing by a possibly zero displacement, define

$$H(u) = \begin{cases} E_2(u)/u, & u \neq 0, \\ 0, & u = 0. \end{cases}$$

Then $H(u) \to 0$ as $u \to 0$ and $E_2(u) = H(u)u$ for every u.

(3) Substituting $u = k(h)$ gives

$$\begin{aligned} g(f(x_0 + h)) &= g(f(x_0)) + g'(f(x_0))k(h) + H(k(h))k(h) \\ &= g(f(x_0)) + g'(f(x_0))f'(x_0)h + g'(f(x_0))E_1(h) + H(k(h))k(h). \end{aligned}$$

(4) The remainder divided by h tends to 0:

$$g'(f(x_0))\frac{E_1(h)}{h} + H(k(h))\frac{k(h)}{h} \longrightarrow 0 + 0 \cdot f'(x_0) = 0.$$

Therefore the coefficient of h is the derivative of $g \circ f$ at x_0.

■

[Direct] cf. R 5.5.

Proposition 8.2.6

Power rule

For every integer $n \geq 0$,

$$\frac{d}{dx}x^n = nx^{n-1},$$

where the case $n = 0$ is read separately as the derivative of the constant function 1, namely 0. R 5.4

Proof 8.2.6

Proof by Induction Using the Product Rule.

(1) For $n = 0$, the function $x^0 = 1$ is constant, so its derivative is 0.

(2) For $n = 1$, the difference quotient of x is 1.

(3) Assume $(x^n)' = nx^{n-1}$. Since $x^{n+1} = x \cdot x^n$, the product rule gives

$$(x^{n+1})' = (x)'x^n + x(x^n)' = x^n + nx^n = (n+1)x^n.$$

The formula follows by induction.

■

[A10]

8.3 Oscillation Examples

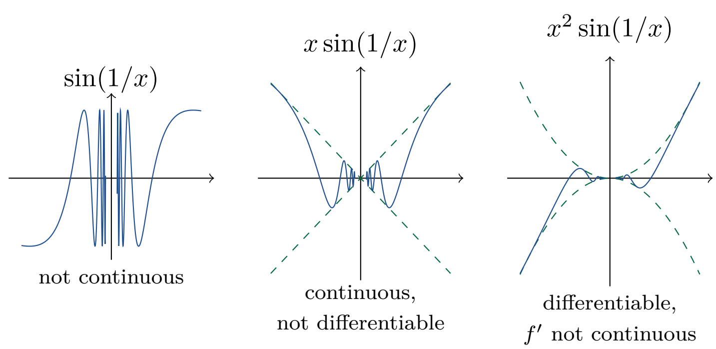

Figure 8.3.1. THREE OSCILLATORY FUNCTIONS. Multiplying by powers of x progressively tames the oscillation of $\sin(1/x)$ near 0.

Example 8.3.2

$\sin(1/x)$ *is not continuous at* 0

Define $f(0) = 0$ and $f(x) = \sin(1/x)$ for $x \neq 0$. Along sequences approaching 0, the values can approach different limits; for example one may choose points where $\sin(1/x) = 1$ and others where $\sin(1/x) = -1$. Thus f is not continuous at 0, and by 8.2.2 it is not differentiable there (cf. R 5.6).

Example 8.3.3

$x\sin(1/x)$ *is continuous but not differentiable at* 0

Let $f(0) = 0$ and $f(x) = x\sin(1/x)$ for $x \neq 0$. Since

$$|x\sin(1/x)| \leq |x|,$$

the squeeze theorem gives $f(x) \to 0 = f(0)$. But the difference quotient at 0 is

$$\frac{f(h) - f(0)}{h} = \sin(1/h),$$

which has no limit. Hence f is continuous at 0 but not differentiable there (R 5.6).

Example 8.3.4

$x^2\sin(1/x)$ *is differentiable but not* C^1 *at* 0

Let $f(0) = 0$ and $f(x) = x^2\sin(1/x)$ for $x \neq 0$. Then

$$\frac{f(h) - f(0)}{h} = h\sin(1/h) \longrightarrow 0,$$

so $f'(0) = 0$. For $x \neq 0$, the product and chain rules give

$$f'(x) = 2x\sin(1/x) - \cos(1/x).$$

The first term tends to 0, but $\cos(1/x)$ has no limit as $x \to 0$. Thus f is differentiable at 0, yet f' is not continuous at 0 (R 5.6).

Remark 8.3.5

The hierarchy already separates

The three examples show that differentiability is stronger than continuity, and continuous differentiability is stronger than differentiability. Each extra factor of x narrows the envelope around the oscillation and buys one more degree of regularity at the origin (cf. R 5.6).

8.4 Extrema and the Mean Value Theorem

Definition 8.4.1

Local extrema

Let $f : E \to \mathbb{R}$ and let $p \in E$. The function f has a *local maximum* at p if there exists $\delta > 0$ such that

$$q \in E, \quad |q - p| < \delta \quad \Longrightarrow \quad f(q) \le f(p).$$

A *local minimum* is defined by reversing the inequality (R 5.7).

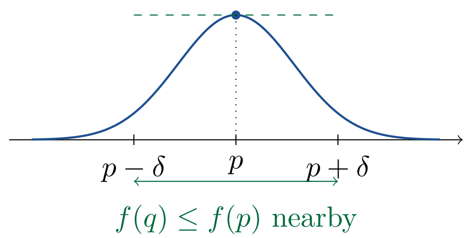

Figure 8.4.2. A LOCAL MAXIMUM. On a sufficiently small window around p, the value $f(p)$ dominates nearby values of the function.

Theorem 8.4.3

Interior extremum theorem

Let $f : (a, b) \to \mathbb{R}$ be differentiable. If $x_0 \in (a, b)$ is a local maximum or a local minimum, then

$$f'(x_0) = 0.$$

This is Fermat's interior-extremum theorem (R 5.8).

Proof 8.4.3

Proof by One-Sided Difference Quotients.

(1) Assume x_0 is a local maximum. For x sufficiently close to x_0,

$$f(x) - f(x_0) \le 0.$$

(2) If $x < x_0$, then $x - x_0 < 0$, so

$$\frac{f(x) - f(x_0)}{x - x_0} \ge 0.$$

Passing to the left-hand limit gives $f'(x_0) \ge 0$.

(3) If $x > x_0$, then $x - x_0 > 0$, so

$$\frac{f(x) - f(x_0)}{x - x_0} \le 0.$$

Passing to the right-hand limit gives $f'(x_0) \le 0$.

(4) The two one-sided limits are equal because $f'(x_0)$ exists. Therefore

$$0 \le f'(x_0) \le 0,$$

and $f'(x_0) = 0$. The minimum case is identical with the inequalities reversed. ■

[A12] cf. R 5.8.

Remark 8.4.4

Why the hypotheses matter

The point must be interior because the proof uses both sides. At an endpoint only one one-sided inequality is available. The converse also fails: $f(x) = x^3$ has $f'(0) = 0$, but 0 is not a local extremum (cf. R 5.8).

Proposition 8.4.5

Rolle's theorem

Let $f : [a, b] \to \mathbb{R}$ be continuous on $[a, b]$ and differentiable on (a, b). If $f(a) = f(b)$, then there exists $c \in (a, b)$ such that

$$f'(c) = 0.$$

This is the Rolle special case behind Rudin's generalized mean-value theorem (R 5.9).

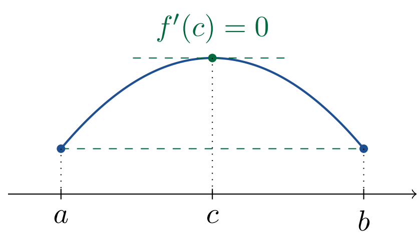

Figure 8.4.6. ROLLE'S THEOREM. Equal endpoint heights force a horizontal tangent at some interior point, unless the function is constant.

Proof 8.4.5

Proof by the Extreme-Value Theorem and Interior Extrema.

(1) Since $[a, b]$ is compact and f is continuous, f attains a maximum and a minimum on $[a, b]$ by the extreme-value theorem from Lecture 7.

(2) If either extreme is attained at an interior point $c \in (a, b)$, then 8.4.3 gives $f'(c) = 0$.

(3) If no extreme is attained in the interior, then both the maximum and minimum occur at endpoints. Since $f(a) = f(b)$, the maximum and minimum values coincide. Hence f is constant on $[a, b]$, and any $c \in (a, b)$ satisfies $f'(c) = 0$. ■

[A6+A12] Rolle step behind R 5.9.

Corollary 8.4.7

Mean Value Theorem

Let $f : [a, b] \to \mathbb{R}$ be continuous on $[a, b]$ and differentiable on (a, b). Then there exists $c \in (a, b)$ such that

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

This is the Mean Value Theorem (R 5.10).[2]

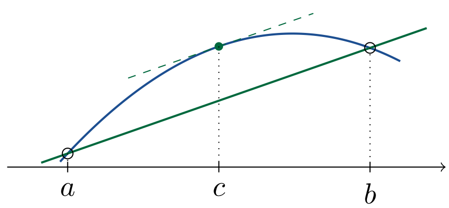

Figure 8.4.8. MEAN VALUE THEOREM. Some interior tangent is parallel to the chord joining the endpoints.

Proof 8.4.7

Proof by Tilting the Chord Flat.

(1) Let L be the secant line

$$L(x) = f(a) + \frac{f(b) - f(a)}{b - a}(x - a),$$

and define $h = f - L$.

(2) The function h is continuous on $[a, b]$ and differentiable on (a, b). Also

$$h(a) = 0, \qquad h(b) = 0.$$

(3) By Rolle's theorem, there exists $c \in (a, b)$ with $h'(c) = 0$. Since

$$h'(x) = f'(x) - \frac{f(b) - f(a)}{b - a},$$

the conclusion follows.

■

[Constr] cf. R 5.10.

[2]This theorem looks modest and carries most of the chapter. Read it as a bridge from *local* to *global*: it turns information about f' at single points into control over how much f can change across a whole interval. Nearly every later result passes through it—the monotonicity test 8.4.9 is the first (the *sign* of f' alone decides whether f rises or falls), and "$f' \equiv 0 \Rightarrow f$ constant" is the same move. The figure above is the whole content: the average slope over $[a, b]$, the chord, is genuinely attained as a tangent slope at some interior c.

Corollary 8.4.9

Monotonicity test

Let f be differentiable on (a, b).

1. If $f'(x) \geq 0$ for all $x \in (a, b)$, then f is nondecreasing.
2. If $f'(x) = 0$ for all $x \in (a, b)$, then f is constant.
3. If $f'(x) \leq 0$ for all $x \in (a, b)$, then f is nonincreasing.

These are the standard monotonicity consequences of the Mean Value Theorem (R 5.11).

Proof 8.4.9

Proof by Applying the Mean Value Theorem on Subintervals.

(1) Fix $x_1 < x_2$ in (a, b). By 8.4.7, there exists $\xi \in (x_1, x_2)$ such that

$$f(x_2) - f(x_1) = f'(\xi)(x_2 - x_1).$$

(2) Since $x_2 - x_1 > 0$, the sign of $f(x_2) - f(x_1)$ is the sign of $f'(\xi)$. The three conclusions follow immediately from the three sign hypotheses.

■

[Direct] cf. R 5.11.

8.5 Darboux's Theorem

Proposition 8.5.1

Darboux's theorem

Let f be differentiable on an interval containing $a < b$. If λ lies strictly between $f'(a)$ and $f'(b)$, then there exists $c \in (a, b)$ such that

$$f'(c) = \lambda.$$

Thus derivatives have the intermediate value property, even though derivatives need not be continuous (R 5.12).

Proof 8.5.1

Proof by Tilting and Taking an Interior Extremum.

(1) Assume $f'(a) < \lambda < f'(b)$; the reverse case is obtained by changing signs. Define

$$g(x) = f(x) - \lambda x.$$

Then $g'(a) < 0$ and $g'(b) > 0$.

(2) Since g is continuous on $[a, b]$, it attains a minimum at some point $c \in [a, b]$.

(3) The condition $g'(a) < 0$ means that just to the right of a the values of g are smaller than $g(a)$, so the minimum is not attained at a. Similarly, $g'(b) > 0$ implies that just to the left of b the values of g are smaller than $g(b)$, so the minimum is not attained at b.

(4) Hence the minimum is attained at an interior point $c \in (a, b)$. By 8.4.3, $g'(c) = 0$, and therefore

$$f'(c) - \lambda = 0.$$

■

[A12+Constr] cf. R 5.12.

Remark 8.5.2

What Darboux rules out

Darboux's theorem forbids derivatives from having jump discontinuities, because a jump would skip the intermediate values. The theorem does not say that derivatives are continuous: 8.3.4 gives a derivative with an oscillatory discontinuity at 0 (cf. R 5.12).

8.6 Taylor Polynomials and Taylor's Theorem

Definition 8.6.1

Taylor polynomial

Let $n \geq 1$, and suppose f has derivatives through order $n-1$ at α. The *Taylor polynomial of degree at most $n-1$ centered at α* is

$$P_{n-1}(x) = \sum_{k=0}^{n-1} \frac{f^{(k)}(\alpha)}{k!}(x-\alpha)^k, \qquad f^{(0)} = f.$$

Rudin introduces higher derivatives in R 5.14 and this polynomial in R 5.15.

Proposition 8.6.2

Matching and uniqueness

The Taylor polynomial P_{n-1} is the unique polynomial of degree at most $n-1$ satisfying

$$P_{n-1}^{(k)}(\alpha) = f^{(k)}(\alpha), \qquad k = 0, 1, \ldots, n-1.$$

This is the matching property built into Rudin's Taylor polynomial in R 5.15.

Proof 8.6.2

Proof in the Shifted Polynomial Basis.

(1) Write a general degree-at-most $n-1$ polynomial in the shifted basis:

$$Q(x) = \sum_{k=0}^{n-1} c_k (x-\alpha)^k.$$

(2) Differentiating k times and evaluating at α leaves only the k-th shifted monomial:

$$Q^{(k)}(\alpha) = k!\, c_k.$$

(3) The matching conditions therefore force

$$c_k = \frac{f^{(k)}(\alpha)}{k!},$$

exactly the coefficients of P_{n-1}. Hence the matching polynomial exists and is unique.

■

[Direct]

Theorem 8.6.3

Taylor's theorem with Lagrange remainder

Let $n \geq 1$. Suppose $f : [a, b] \to \mathbb{R}$ has $f^{(n-1)}$ continuous on $[a, b]$ and $f^{(n)}$ exists on (a, b). If $\alpha, \beta \in [a, b]$ are distinct and P_{n-1} is the Taylor polynomial centered at α, then there is a point ξ strictly between α and β such that

$$f(\beta) = P_{n-1}(\beta) + \frac{f^{(n)}(\xi)}{n!}(\beta - \alpha)^n.$$

This is Rudin's Taylor theorem with Lagrange remainder (R 5.15).[3]

Proof 8.6.3

Proof by Iterated Rolle's Theorem.

(1) Assume $\alpha < \beta$; the other order is identical on the interval between them. Choose M by

$$M = \frac{f(\beta) - P_{n-1}(\beta)}{(\beta - \alpha)^n},$$

so that $f(\beta) = P_{n-1}(\beta) + M(\beta - \alpha)^n$.

(2) Define

$$g(x) = f(x) - P_{n-1}(x) - M(x - \alpha)^n.$$

By the matching property in 8.6.2,

$$g(\alpha) = g'(\alpha) = \cdots = g^{(n-1)}(\alpha) = 0,$$

[3] This extends the lecture's opening idea. In 8.1.3 a derivative let us stand in for f near α by a *line*—a degree-one polynomial—with error $o(h)$; here f is stood in for by a degree-$(n-1)$ *polynomial*, the one matching f and its first $n-1$ derivatives at α (8.6.2), and the remainder names exactly what that polynomial misses. More matched derivatives, a closer fit near α. From here the remainder is the whole question: whether it shrinks as the degree grows is what decides whether the polynomials rebuild f (8.7.1).

and by the choice of M, $g(\beta) = 0$.

(3) Apply Rolle's theorem repeatedly. First, $g(\alpha) = g(\beta) = 0$ gives a point $c_1 \in (\alpha, \beta)$ with $g'(c_1) = 0$. Then $g'(\alpha) = g'(c_1) = 0$ gives a point $c_2 \in (\alpha, c_1)$ with $g''(c_2) = 0$. Continuing n times gives a point $\xi \in (\alpha, \beta)$ with

$$g^{(n)}(\xi) = 0.$$

(4) Since $P_{n-1}^{(n)} = 0$ and $\frac{d^n}{dx^n} M(x-\alpha)^n = n!M$,

$$0 = g^{(n)}(\xi) = f^{(n)}(\xi) - n!M.$$

Thus $M = f^{(n)}(\xi)/n!$, and substituting this into the definition of M gives the stated formula.

■

[A6+Constr] cf. R 5.15.

Remark 8.6.4

The qualitative little-o form

The Lagrange formula gives the quantitative error. If, in addition, $f^{(n)}$ is locally bounded near α–for instance if $f^{(n)}$ is continuous near α–then

$$f(t) - P_{n-1}(t) = O\big((t-\alpha)^n\big),$$

and hence

$$\frac{f(t) - P_{n-1}(t)}{(t-\alpha)^{n-1}} \longrightarrow 0.$$

The boundedness condition is doing real work here; the existence of $f^{(n)}$ alone does not by itself provide local boundedness. This is the qualitative consequence of the Lagrange remainder in R 5.15.

8.7 Real and Complex Analyticity

Definition 8.7.1

Real analyticity at a point

A smooth real function f is *real analytic at* α if its Taylor series at α converges to f on some neighbourhood of α:

$$f(t) = \sum_{k=0}^{\infty} \frac{f^{(k)}(\alpha)}{k!}(t-\alpha)^k$$

for all t sufficiently close to α. This goes beyond Rudin's finite Taylor theorem R 5.15: one must also control the limiting behaviour of the remainders as the degree tends to infinity.

Counterexample 8.7.2

A smooth function need not be analytic

Define

$$f(x) = \begin{cases} e^{-1/x^2}, & x \neq 0, \\ 0, & x = 0. \end{cases}$$

This function is C^∞, and all derivatives at 0 vanish.[4] Its Taylor series at 0 is therefore identically 0, but $f(x) > 0$ for $x \neq 0$. Hence smoothness does not imply real analyticity. This is beyond Rudin's finite Taylor theorem R 5.15.

FAILS: the claim that knowing all derivatives at one point determines a smooth real function.

Theorem 8.7.3

Weierstrass Function

There exists a continuous function $W : \mathbb{R} \to \mathbb{R}$ that is differentiable at no point. This is an advanced separation result beyond Rudin Chapter 5.

Proof 8.7.3

Quoted Advanced Result.

↪ This theorem is stated as part of the regularity landscape; its proof is beyond the lecture.

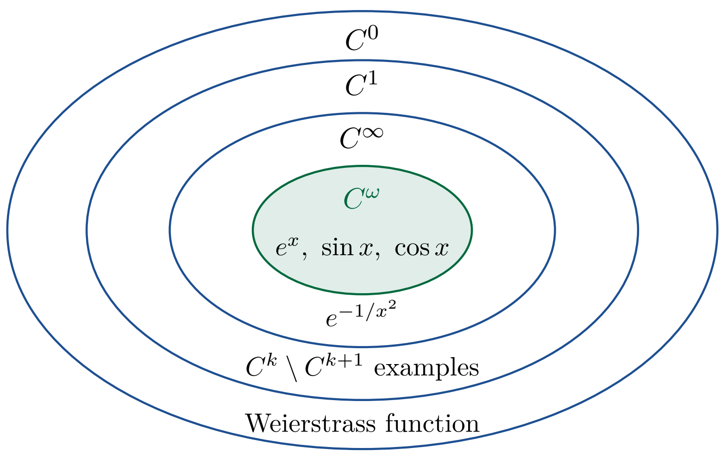

Figure 8.7.4. THE REAL REGULARITY TOWER. For real functions the inclusions between regularity classes are strict; each rung contains functions not lying on the next rung.

Remark 8.7.5

Derivative growth and Taylor convergence

Taylor convergence is controlled by the growth of the coefficients $f^{(r)}(\alpha)/r!$. If these coefficients grow too quickly, the formal Taylor series can have radius of convergence 0. A useful sufficient condition in

[4] It is tempting to assume that a function and *all* its derivatives at one point must determine it. Over $\mathbb{R}$ they do not, and this is the example to remember. Every derivative of f at 0 is zero, so its Taylor series at 0 is the zero series—which rebuilds the zero function, not f. The series converges perfectly well; it just converges to the *wrong* function. So for a smooth real function the Taylor series is a *hope*, not a promise; "analytic" (8.7.1) is the word for the functions that keep it, and most smooth ones do not.

the other direction is uniform control of derivatives on a neighbourhood: if the Lagrange remainders satisfy a bound forcing

$$\left|\frac{f^{(n)}(\xi)}{n!}(t-\alpha)^n\right| \longrightarrow 0,$$

then the Taylor polynomials converge back to f near α. Compare the finite Lagrange-remainder control in R 5.15.

Theorem 8.7.6

Holomorphic functions are analytic

Let $U \subseteq \mathbb{C}$ be open. If $f : U \to \mathbb{C}$ is complex differentiable at every point of U, then f is complex analytic: near each point of U, it is represented by a convergent power series. This complex-analysis theorem is beyond Rudin Chapter 5.[5]

Proof 8.7.6

Quoted Theorem from Complex Analysis.

↪ The result is recorded to contrast real and complex differentiability; its proof belongs to complex analysis.

Remark 8.7.7

The real tragedy and the complex collapse

Over $\mathbb{R}$, the classes $C^1, C^2, \dots, C^\infty, C^\omega$ do not collapse. Over $\mathbb{C}$, the hypothesis is not ordinary real differentiability of a map $\mathbb{R}^2 \to \mathbb{R}^2$; it is complex differentiability on an open set, a much more rigid condition. Under that holomorphic hypothesis, one derivative already forces local power-series representation. This is the contrast with Rudin's finite Taylor theorem R 5.15 and beyond the real-variable scope of Chapter 5.

8.8 The Cauchy Mean Value Theorem and L'Hôpital's Rule

Theorem 8.8.1

Cauchy Mean Value Theorem

Let $f, g : [a,b] \to \mathbb{R}$ be continuous on $[a,b]$ and differentiable on (a,b). Then there exists $t \in (a,b)$ such that

$$[f(b)-f(a)]g'(t) = [g(b)-g(a)]f'(t).$$

If in addition g' is nonzero on (a,b) and $g(b) \neq g(a)$, this is the symmetric ratio form

$$\frac{f(b)-f(a)}{g(b)-g(a)} = \frac{f'(t)}{g'(t)}.$$

[5] Why is $\mathbb{C}$ so much more rigid than $\mathbb{R}$? Complex differentiability demands far more than it appears to. The quotient $(f(z)-f(z_0))/(z-z_0)$ must approach the *same* limit no matter which direction z comes in from in the plane—a heavy constraint with no counterpart on the line, where a point has only a left and a right. Imposed at every point of U, that one requirement already forces a convergent power series, and the strict real tower $C^1 \supsetneq C^2 \supsetneq \cdots \supsetneq C^\omega$ of 8.7.4 collapses to a single floor.

Taking $g(x) = x$ recovers the ordinary Mean Value Theorem 8.4.7; this is the generalized mean value theorem behind Rudin (R 5.9).[6]

Proof 8.8.1

Proof by Rolle Applied to a Combined Function.

(1) Define $h : [a, b] \to \mathbb{R}$ by

$$h(x) = \big[f(b) - f(a)\big]g(x) - \big[g(b) - g(a)\big]f(x).$$

As a linear combination of f and g, the function h is continuous on $[a, b]$ and differentiable on (a, b).

(2) Evaluate at the endpoints. Both reduce to the same value:

$$\begin{aligned} h(a) &= \big[f(b) - f(a)\big]g(a) - \big[g(b) - g(a)\big]f(a) = f(b)g(a) - g(b)f(a), \\ h(b) &= \big[f(b) - f(a)\big]g(b) - \big[g(b) - g(a)\big]f(b) = f(b)g(a) - g(b)f(a). \end{aligned}$$

Hence $h(a) = h(b)$.

(3) By Rolle's theorem 8.4.5, there exists $t \in (a, b)$ with $h'(t) = 0$. Since

$$h'(x) = \big[f(b) - f(a)\big]g'(x) - \big[g(b) - g(a)\big]f'(x),$$

this gives $\big[f(b) - f(a)\big]g'(t) = \big[g(b) - g(a)\big]f'(t)$, which is the claim. When g' is nonzero on (a, b), the Mean Value Theorem applied to g forces $g(b) \neq g(a)$, and dividing by $\big[g(b) - g(a)\big]g'(t) \neq 0$ produces the ratio form.

■

[A6+A12] Rolle step 8.4.5; cf. R 5.9.

Theorem 8.8.2

L'Hôpital's Rule, the 0/0 case

Let $-\infty \leq a < b \leq +\infty$, and suppose f and g are differentiable on (a, b) with $g'(x) \neq 0$ for all $x \in (a, b)$. Assume

$$f(x) \to 0 \quad \text{and} \quad g(x) \to 0 \qquad (x \to a^+),$$

and that, for some $A \in \mathbb{R}$,

$$\frac{f'(x)}{g'(x)} \longrightarrow A \qquad (x \to a^+).$$

[6] The ordinary Mean Value Theorem compares f against the clock x; the Cauchy version compares f against *any* second travelling quantity g, so it reads slopes "per unit g" rather than "per unit x." That is exactly the leverage L'Hôpital's rule 8.8.2 needs to compare two functions that are both vanishing at the same point: one cannot use the clock, because the destination is 0/0, so one races f directly against g.

Then

$$\frac{f(x)}{g(x)} \longrightarrow A \qquad (x \to a^+).$$

This is L'Hôpital's rule for the indeterminate form 0/0 (R 5.13).[7]

Proof 8.8.2

Proof by the Cauchy Mean Value Theorem and Pinching.

(1) We treat the case $A \neq \pm\infty$. Fix any real number $r > A$. Since $f'(x)/g'(x) \to A$ as $x \to a^+$, there exists $c \in (a, b)$ such that

$$\frac{f'(x)}{g'(x)} < r \qquad \text{for all } x \in (a, c).$$

(2) Let $x, y \in (a, c)$ with $x < y$. The Cauchy Mean Value Theorem 8.8.1, applied to f and g on $[x, y]$, supplies a point $t \in (x, y) \subseteq (a, c)$ with

$$\frac{f(x) - f(y)}{g(x) - g(y)} = \frac{f'(t)}{g'(t)} < r.$$

(The denominator $g(x) - g(y)$ is nonzero because g' does not vanish on (a, b), by the Mean Value Theorem.)

(3) Hold $y \in (a, c)$ fixed and let $x \to a^+$. Then $f(x) \to 0$ and $g(x) \to 0$, so the left-hand side tends to $\dfrac{-f(y)}{-g(y)} = \dfrac{f(y)}{g(y)}$, and the strict inequality passes to the limit as

$$\frac{f(y)}{g(y)} \leq r \qquad \text{for all } y \in (a, c).$$

(4) The number $r > A$ was arbitrary, so $\limsup_{y \to a^+} f(y)/g(y) \leq A$. A symmetric argument with any $q < A$ produces some $d \in (a, b)$ with

$$\frac{f(y)}{g(y)} \geq q \qquad \text{for all } y \in (a, d),$$

whence $\liminf_{y \to a^+} f(y)/g(y) \geq A$. The two bounds squeeze the ratio, giving $f(y)/g(y) \to A$ as $y \to a^+$.

■

[Constr] uses 8.8.1; cf. R 5.13.

8.9 Differentiation of Vector-Valued Functions

Definition 8.9.1

[7] The rule is not "differentiate top and bottom and hope." What licenses it is the Cauchy Mean Value Theorem 8.8.1: on any short interval the genuine ratio f/g equals a ratio of *derivatives* f'/g' at some interior point, and as the interval shrinks toward a that interior point is trapped near a too, where f'/g' is already close to A. The pinching by $r > A$ and $q < A$ below is just the ε–free way of saying "close to A from both sides."

Differentiable vector-valued function

Let $f : [a, b] \to \mathbb{R}^k$, written in components $f = (f_1, \dots, f_k)$ with each $f_i : [a, b] \to \mathbb{R}$. The function f is *differentiable at* x if the limit

$$f'(x) := \lim_{h \to 0} \frac{f(x+h) - f(x)}{h} \in \mathbb{R}^k$$

exists. Here the numerator $f(x+h) - f(x)$ is a vector and the increment h is a scalar, so the quotient is a vector divided by a number. This is Rudin's one-parameter vector-valued derivative (R 5.16).[8]

Proposition 8.9.2

Differentiation is componentwise

A function $f = (f_1, \dots, f_k) : [a, b] \to \mathbb{R}^k$ is differentiable at x if and only if each component f_i is differentiable at x, in which case

$$f'(x) = (f_1'(x), \dots, f_k'(x)).$$

In particular, a complex-valued function $f = f_1 + i f_2 : [a, b] \to \mathbb{C}$ is differentiable exactly when f_1 and f_2 are, with $f'(x) = f_1'(x) + i f_2'(x)$. This is Rudin's componentwise differentiation (R 5.16).

Proof 8.9.2

Proof by Reading the Quotient Componentwise.

(1) For $h \neq 0$ the difference quotient is the vector whose i-th coordinate is the scalar difference quotient of f_i:

$$\frac{f(x+h) - f(x)}{h} = \left(\frac{f_1(x+h) - f_1(x)}{h}, \dots, \frac{f_k(x+h) - f_k(x)}{h} \right).$$

(2) A sequence of vectors in $\mathbb{R}^k$ converges if and only if each coordinate sequence converges, and the limit is taken coordinatewise. Hence the vector limit defining $f'(x)$ exists precisely when every $f_i'(x)$ exists, and then $f'(x) = (f_1'(x), \dots, f_k'(x))$.

(3) Identifying $\mathbb{C}$ with $\mathbb{R}^2$ through $f_1 + i f_2 \leftrightarrow (f_1, f_2)$ specialises the statement to the complex case, giving $f' = f_1' + i f_2'$.

■

[Direct] cf. R 5.16.

Remark 8.9.3

Which rules survive, and which break

[8] This is a genuinely different object from the multivariable derivative 8.1.4, even though both involve $\mathbb{R}^k$. Here the *input* is still one real number, so one may legitimately divide the vector displacement by the scalar step h; the derivative is itself a vector (a velocity). In 8.1.4 the input is a vector, division is impossible, and the derivative is a linear map. Same target, opposite side.

The sum, scalar-multiple, and product rules carry over to vector-valued functions. In particular, if $f, g : [a, b] \to \mathbb{R}^k$ are differentiable then the scalar-valued dot product $f \cdot g$ is differentiable with

$$(f \cdot g)' = f' \cdot g + f \cdot g'.$$

Two things *fail*, however. There is no quotient rule, simply because $\mathbb{R}^k$ has no division for $k \geq 2$. More importantly, the equality form of the Mean Value Theorem breaks: there need be no single c with $f(b) - f(a) = f'(c)(b - a)$. The standard witness is $f(x) = (\cos x, \sin x)$ on $[0, 2\pi]$, for which $f(2\pi) - f(0) = (0, 0)$, yet

$$\|f'(x)\| = \|(-\sin x, \cos x)\| = 1 \neq 0 \quad \text{for every } x,$$

so no interior point can make $f'(c)\,(2\pi - 0)$ vanish (cf. R 5.18).

Proposition 8.9.4

Mean Value Inequality for vector-valued functions

Let $f : [a, b] \to \mathbb{R}^k$ be continuous on $[a, b]$ and differentiable on (a, b). Then there exists $x \in (a, b)$ such that

$$\|f(b) - f(a)\| \leq (b - a)\,\|f'(x)\|.$$

This inequality replaces the (failed) equality form of the Mean Value Theorem in the vector-valued setting (R 5.19).

Proof 8.9.4

Proof by Projecting Onto the Displacement and Cauchy–Schwarz.

(1) Set $z = f(b) - f(a) \in \mathbb{R}^k$ and define the real-valued function $\varphi : [a, b] \to \mathbb{R}$ by $\varphi(t) = z \cdot f(t)$, the projection of f onto the fixed displacement vector z. Then φ is continuous on $[a, b]$ and differentiable on (a, b) with $\varphi'(t) = z \cdot f'(t)$.

(2) By the scalar Mean Value Theorem 8.4.7, there exists $x \in (a, b)$ with

$$\varphi(b) - \varphi(a) = \varphi'(x)(b - a) = z \cdot f'(x)\,(b - a).$$

The left-hand side is $z \cdot f(b) - z \cdot f(a) = z \cdot \big(f(b) - f(a)\big) = z \cdot z = \|z\|^2$.

(3) By the Cauchy–Schwarz inequality, $z \cdot f'(x) \leq \|z\|\,\|f'(x)\|$, so

$$\|z\|^2 = z \cdot f'(x)\,(b - a) \leq \|z\|\,\|f'(x)\|\,(b - a).$$

(4) If $z \neq 0$, divide by $\|z\| > 0$ to obtain $\|z\| \leq (b - a)\|f'(x)\|$, which is the claim. If $z = 0$ the left-hand side $\|f(b) - f(a)\|$ is zero and the inequality holds trivially for any $x \in (a, b)$.

■

[Constr] uses 8.4.7; cf. R 5.19.

Remark 8.9.5

The physical reading

If $f(t)$ is the position of a particle at time t, then $f'(t)$ is its velocity vector and $\|f'(t)\|$ its speed. The inequality 8.9.4 then says that the straight-line distance between the endpoints, $\|f(b) - f(a)\|$, is at most the maximum speed times the elapsed time—distance travelled cannot exceed speed multiplied by duration, exactly as expected (cf. R 5.19).

Appendix

Supplementary material—additional proofs, context, and other treatments; not part of the lecture proper.

The first variation of the energy functional. This worked example was presented in lecture as an extended illustration of the abstract derivative 8.1.3; it is enrichment, and is *not* examinable. It carries the best-linear-approximation definition into a setting where the "point" is an entire curve, so the increment is itself a curve and there is nothing to divide by—exactly the situation 8.1.5 anticipated. The first variation yields only a *necessary* condition (a critical curve); a separate Cauchy–Schwarz bound (8.A.9) then shows that curve is the *global* minimiser.

Setup. Fix $p, q \in \mathbb{R}^k$ and let

$$\mathcal{C} = \{\, \gamma : [0,1] \to \mathbb{R}^k \mid \gamma(0) = p,\ \gamma(1) = q,\ \gamma \in C^1 \,\}$$

be the C^1 curves joining them: $\gamma(t)$ is the point reached at time t, the endpoint conditions pin p and q, and $\gamma'(t) \in \mathbb{R}^k$ is the velocity. A single "point" of $\mathcal{C}$ is a whole curve, so $\mathcal{C}$ is infinite-dimensional; it is an *affine*, not a vector, space (adding two curves through p, q lands at $2p, 2q$), which is why the perturbation directions form the separate set $\mathcal{V}$ of 8.A.2.

Definition 8.A.1

The energy functional

The *energy* of a curve $\gamma \in \mathcal{C}$ is

$$E(\gamma) = \int_0^1 \|\gamma'(t)\|^2 \, dt, \qquad \|\gamma'(t)\|^2 = \gamma'(t) \cdot \gamma'(t) = \sum_{j=1}^{k} \big(\gamma_j'(t)\big)^2,$$

the integral of the squared speed, so $E : \mathcal{C} \to \mathbb{R}$ sends a curve to a single real number. We ask which $\gamma \in \mathcal{C}$ has least energy. No naive difference quotient $\big(E(\gamma + h) - E(\gamma)\big)/h$ is available—"h" would be a curve, and one cannot divide by a curve—so we use the abstract derivative 8.1.3, which asks only for the best linear approximation, never for a quotient.

Definition 8.A.2

Admissible variations

To probe E near γ we nudge it in a direction drawn from

$$\mathcal{V} = \{\, \eta : [0,1] \to \mathbb{R}^k \mid \eta \in C^1,\ \eta(0) = \eta(1) = 0 \,\},$$

scaled by a real parameter h: the nudged curve $\gamma + h\eta$ moves each point $\gamma(t)$ by $h\,\eta(t)$. The endpoint conditions $\eta(0) = \eta(1) = 0$ keep it admissible, since

$$(\gamma + h\eta)(0) = p + h \cdot 0 = p, \qquad (\gamma + h\eta)(1) = q + h \cdot 0 = q$$

for *every* h; so $\mathcal{V}$ plays the role of the tangent space to $\mathcal{C}$ at γ.

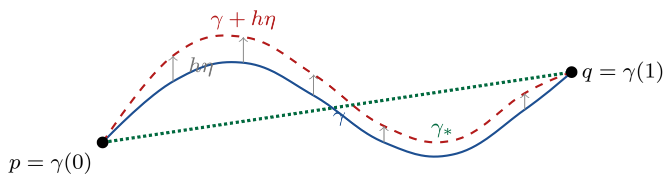

Figure 8.A.3. A CURVE AND AN ADMISSIBLE VARIATION. The variation η is a vector field along γ vanishing at the endpoints, so every perturbed curve $\gamma + h\eta$ still joins p to q (no arrows at the ends). The dotted line is the energy-minimising straight curve γ_* of 8.A.8.

Step 1: reduce to one real variable. For fixed γ and admissible η, set

$$g(h) := E(\gamma + h\eta) = \int_0^1 \|\gamma'(t) + h\,\eta'(t)\|^2\,dt, \qquad g : \mathbb{R} \to \mathbb{R},$$

using $(\gamma + h\eta)' = \gamma' + h\eta'$. This is an ordinary real function of h, which one-variable calculus can attack.

Proposition 8.A.4

The energy along a variation is quadratic in h

For fixed γ and admissible η,

$$g(h) = E(\gamma + h\eta) = E(\gamma) + 2Bh + Ch^2, \qquad B = \int_0^1 \gamma' \cdot \eta'\,dt, \quad C = \int_0^1 \|\eta'\|^2\,dt.$$

Against the abstract expansion $\varphi(x_0 + h) = \varphi(x_0) + A(h) + o(h)$ of 8.1.3: the constant $E(\gamma)$ is $\varphi(x_0)$, the linear term $2Bh$ is the derivative $A(h)$, and the quadratic Ch^2 is the error—genuinely $o(h)$, since $|Ch^2|/|h| = |C|\,|h| \to 0$.

Proof 8.A.4

Direct Proof by Expanding and Integrating.

(1) Expand the integrand with $\|v\|^2 = v \cdot v$ and bilinearity:

$$\|\gamma' + h\eta'\|^2 = (\gamma' + h\eta') \cdot (\gamma' + h\eta') = \|\gamma'\|^2 + 2h\,(\gamma' \cdot \eta') + h^2\|\eta'\|^2,$$

the two cross terms coinciding because the dot product is symmetric.

(2) Integrate over $[0, 1]$ and pull the constants $2h, h^2$ outside (linearity of $\int$):

$$E(\gamma + h\eta) = \int_0^1 \|\gamma'\|^2\,dt + 2h \int_0^1 \gamma' \cdot \eta'\,dt + h^2 \int_0^1 \|\eta'\|^2\,dt = E(\gamma) + 2Bh + Ch^2.$$

(3) For the remainder, $|Ch^2|/|h| = |C|\,|h| \to 0$ as $h \to 0$, so $Ch^2 = o(h)$.

■

[Direct] bilinearity and symmetry of $\cdot$; linearity of $\int$.

Definition 8.A.5

First variation

The *first variation* of E at γ in the direction η is the coefficient of h in 8.A.4,[9]

$$\delta E(\gamma)[\eta] := \frac{d}{dh}\Big|_{h=0} E(\gamma + h\eta) = g'(0) = 2B = 2\int_0^1 \gamma'(t)\cdot\eta'(t)\,dt,$$

the directional (Gâteaux) derivative of E at γ. (Differentiating $g(h) = E(\gamma) + 2Bh + Ch^2$ gives $g'(h) = 2B + 2Ch$, so $g'(0) = 2B$.)

Step 2: integration by parts. Assume now $\gamma \in C^2$,[10] so that γ'' is continuous, and rewrite $\int_0^1 \gamma'\cdot\eta'\,dt$ to move the derivative off the variation η and onto the curve γ.

Proposition 8.A.6

The first variation after integration by parts

For $\gamma \in C^2$ and admissible η,

$$\delta E(\gamma)[\eta] = 2\int_0^1 \gamma'\cdot\eta'\,dt = -2\int_0^1 \gamma''(t)\cdot\eta(t)\,dt.$$

Proof 8.A.6

Direct Proof by the Product Rule and the Fundamental Theorem.

(1) The product rule for $t \mapsto \gamma'(t)\cdot\eta(t) = \sum_j \gamma_j'\eta_j$ gives

$$\tfrac{d}{dt}(\gamma'\cdot\eta) = \gamma''\cdot\eta + \gamma'\cdot\eta', \qquad \text{so} \qquad \gamma'\cdot\eta' = \tfrac{d}{dt}(\gamma'\cdot\eta) - \gamma''\cdot\eta.$$

(2) Integrate over $[0,1]$; the exact-derivative term becomes a boundary value by the Fundamental Theorem of Calculus:

$$\int_0^1 \gamma'\cdot\eta'\,dt = \big[\gamma'(t)\cdot\eta(t)\big]_0^1 - \int_0^1 \gamma''\cdot\eta\,dt = \gamma'(1)\cdot\eta(1) - \gamma'(0)\cdot\eta(0) - \int_0^1 \gamma''\cdot\eta\,dt.$$

(3) The endpoint conditions $\eta(0) = \eta(1) = 0$ (8.A.2) kill the boundary term, leaving $\int_0^1 \gamma'\cdot\eta'\,dt = -\int_0^1 \gamma''\cdot\eta\,dt$. Multiplying by 2 gives the claim.

■

[Direct] uses 8.A.2; the endpoint conditions kill the boundary term.

Step 3: stationarity and the fundamental lemma. If γ minimises E, then for every admissible η the function $g(h) = E(\gamma + h\eta)$ has a minimum at $h = 0$, so $g'(0) = \delta E(\gamma)[\eta] = 0$;

[9] On the board the first variation is written *with* a factor of h, as $E'(\gamma)\cdot\eta = 2h\int_0^1 \gamma'\cdot\eta'\,dt$, straight from the $2Bh$ term; we define it instead as the coefficient of h (a derivative should not still contain h). Both give the same stationarity condition—setting either to zero, the nonzero scalar divides out. Some texts write ε for the variation parameter in place of h.

[10] The board worked in C^1. Taking $\gamma \in C^2$ is what makes the integration by parts valid as written; with only $\gamma \in C^1$ the same conclusion $\gamma'' \equiv 0$ follows from the du Bois-Reymond lemma, at the cost of a longer argument. We follow the lecture's C^2 route.

by 8.A.6 this says $\int_0^1 \gamma'' \cdot \eta \, dt = 0$ for all admissible η. One integral vanishing does not force its integrand to vanish—but vanishing against *every* test direction does:

Lemma 8.A.7

Fundamental lemma of the calculus of variations

Let $f : [0,1] \to \mathbb{R}^k$ be continuous. If $\int_0^1 f(t) \cdot \eta(t) \, dt = 0$ for every $\eta \in C^1([0,1], \mathbb{R}^k)$ with $\eta(0) = \eta(1) = 0$, then $f \equiv 0$ on $[0,1]$.

Proof 8.A.7

Proof by Contradiction with a Bump Test Field.

(1) It suffices that each component $f_j \equiv 0$: taking $\eta = \phi \, e_j$ for a scalar ϕ and the standard basis vector e_j reduces the hypothesis to $\int_0^1 f_j \, \phi \, dt = 0$.

(2) Suppose $f_j(t_0) \neq 0$ for some $t_0 \in (0,1)$; say $f_j(t_0) > 0$. By continuity, $f_j > 0$ throughout an open interval $(a,b) \subseteq (0,1)$ containing t_0.

(3) Take the bump $\phi(t) = (t-a)^2(b-t)^2$ on (a,b) and $\phi = 0$ elsewhere. Then $\phi \in C^1([0,1])$ (value and first derivative vanish at a, b), $\phi(0) = \phi(1) = 0$, and $\phi > 0$ on (a,b), so

$$\int_0^1 f_j \, \phi \, dt = \int_a^b \underbrace{f_j}_{>0} \, \underbrace{\phi}_{>0} \, dt > 0,$$

contradicting the hypothesis. Hence $f_j \equiv 0$ for each j, and $f \equiv 0$.

■

[A9+Constr] reduce to one component, then localise with an explicit bump.

Proposition 8.A.8

The unique critical curve is the straight line

If $\gamma \in C^2$ satisfies $\delta E(\gamma)[\eta] = 0$ for all admissible η, then $\gamma'' \equiv 0$, equivalently

$$\gamma_*(t) = p + t(q-p) = (1-t)\, p + t\, q, \qquad t \in [0,1],$$

the constant-speed straight line from p to q, with energy $E(\gamma_*) = \|q-p\|^2$.

Proof 8.A.8

Constructive Proof by Integrating Twice.

(1) Stationarity and 8.A.6 give $\int_0^1 \gamma'' \cdot \eta \, dt = 0$ for all admissible η; since γ'' is continuous, 8.A.7 forces $\gamma'' \equiv 0$ (the Euler–Lagrange equation).

(2) Integrate twice: $\gamma'' = 0$ gives $\gamma'(t) = a$, then $\gamma(t) = a\,t + b$, with $a, b \in \mathbb{R}^k$.

(3) The endpoints give $\gamma(0) = b = p$ and $\gamma(1) = a + b = q$, so $a = q - p$ and $\gamma_*(t) = p + t(q-p)$, with velocity $\gamma_*'(t) = q - p$ and energy $E(\gamma_*) = \int_0^1 \|q-p\|^2 \, dt = \|q-p\|^2$.

■

[Constr] stationarity + 8.A.7 give $\gamma'' \equiv 0$; the endpoints fix the constants.

Theorem 8.A.9

The straight line is the global minimiser

For every $\gamma \in \mathcal{C}$,

$$E(\gamma) \;\geq\; \|q-p\|^2,$$

with equality if and only if $\gamma = \gamma_*$. Hence $\gamma_*(t) = p + t(q-p)$ is the unique minimiser of E over $\mathcal{C}$.

Proof 8.A.9

Direct Proof by the Cauchy–Schwarz Inequality.

(1) By the Fundamental Theorem of Calculus, componentwise, $q - p = \gamma(1) - \gamma(0) = \int_0^1 \gamma'(t)\,dt$.

(2) Apply the scalar Cauchy–Schwarz inequality $\left(\int_0^1 u\,dt\right)^2 \leq \int_0^1 u^2\,dt$ (the constant 1 in the second slot) to each component $u = \gamma_j'$, then sum:

$$\|q-p\|^2 = \sum_{j=1}^{k}\Big(\int_0^1 \gamma_j'\,dt\Big)^2 \leq \sum_{j=1}^{k}\int_0^1 (\gamma_j')^2\,dt = \int_0^1 \|\gamma'\|^2\,dt = E(\gamma).$$

(3) The straight line meets the bound, $E(\gamma_*) = \|q-p\|^2$, so it is a global minimiser. Equality in Cauchy–Schwarz forces each γ_j' constant, hence $\gamma' \equiv q - p$ and $\gamma = \gamma_*$; the minimiser is unique.

■

[Direct] a lower bound holding for every admissible curve; cf. 8.A.8.

Remark 8.A.10

The general Euler–Lagrange pattern

This is the first instance of a general pattern. For a Lagrangian $L = L(t,x,v)$, the functional $I[\gamma] = \int_0^1 L\big(t,\gamma,\gamma'\big)\,dt$ has first variation given by the same argument—first variation, then integration by parts—producing the *Euler–Lagrange equation*

$$\frac{d}{dt}\left(\frac{\partial L}{\partial v}\right) - \frac{\partial L}{\partial x} = 0.$$

For the energy $L(t,x,v) = \|v\|^2$ one has $\partial L/\partial v = 2v$ and $\partial L/\partial x = 0$, so the equation reads $\frac{d}{dt}(2\gamma') = 0$, i.e. $\gamma'' = 0$—exactly 8.A.8.

Remark 8.A.11

Energy versus length

Energy is not length, but Cauchy–Schwarz links them. With $\mathrm{Length}(\gamma) = \int_0^1 \|\gamma'\|\,dt$, treating the integrand as $\|\gamma'\| \cdot 1$,

$$\mathrm{Length}(\gamma) = \int_0^1 \|\gamma'\| \cdot 1\,dt \leq \Big(\int_0^1 \|\gamma'\|^2\,dt\Big)^{1/2}\Big(\int_0^1 1\,dt\Big)^{1/2} = E(\gamma)^{1/2},$$

so $\mathrm{Length}(\gamma)^2 \leq E(\gamma)$, with equality iff $\|\gamma'\|$ is constant. With $\mathrm{Length}(\gamma) \geq \|q-p\|$ (triangle inequality for integrals) this chains to $E(\gamma) \geq \mathrm{Length}(\gamma)^2 \geq \|q - p\|^2$. Length is reparametrisation-invariant but energy is not: minimising energy both shortens the curve *and* selects the constant-speed parametrisation—which is why energy, not length, is the convenient functional for geodesics.

Index of Objects