UJEP

Department of Mathematics

Mathematics for Data Engineers

KMA/MDI

Part III

Integration and the geometry of functions

Lectures 8–9

Yaroslav Bazaikin

October 2026

Preface

This part of *Mathematics for Data Engineers*, KMA/MDI, contains Lectures 8–9: integration, changes of variables, and approximation of functions. It continues Parts I–II. Vectors become functions, finite sums become integrals, and the same Gram matrices and projection arguments determine the best model.

We assume elementary derivatives and antiderivatives and the linear algebra and optimization of Parts I–II. Each lecture ends with six review questions and eight exercises; the final two exercises, marked with an asterisk, extend the core. Sections labelled *Further perspective* provide optional connections.

The route through this part

Lecture	Mathematical tool and applied purpose
8	Integrals and iterated integration: recover totals, averages, and spatial summaries from local data.
9	Coordinate weights and function geometry: preserve totals under resampling and optimize signal representations.

The final calculations prepare Fourier and Haar expansions. Their coefficients will use the same inner product and normalization; their approximation error will use the same squared norm.

Starting point and notation

The running signal example is a new continuous record

$$s(t) = 1 + t + t^2, \qquad -1 \le t \le 1,$$

with time and amplitude expressed in fixed normalized units. We want to store two numbers, a, b, and reconstruct

$$p(t) = a + bt.$$

The fixed functions 1 and t specify the model family; a, b are its unknown coefficients. The *offset* a is the reconstructed value at $t = 0$, and the *slope* b is its change per unit of normalized time.

As before, the residual means observed or prescribed value minus model value: $r(t) = s(t) - p(t)$. Our objective is

$$E(a, b) = \int_{-1}^{1} \big(s(t) - a - bt\big)^2 \, \mathrm{d}t. \tag{1}$$

There is no factor $1/2$. Dividing by the interval length 2 gives mean squared error; taking its square root gives root mean squared error (RMSE). All three objectives have the same minimizers.

An integration variable is a *dummy variable*: consistently replacing t by u leaves an integral unchanged. For iterated integration, the limits and differential specify which variable changes:

$$\int_{x=0}^{2} \left[\int_{y=0}^{1} f(x, y) \, \mathrm{d}y \right] \mathrm{d}x.$$

The inner integral holds x fixed and returns a function of x; the outer integral accumulates that function. We write $\mathrm{d}A = \mathrm{d}x \, \mathrm{d}y$ for area.

The real function inner product will be $\langle f, g \rangle = \int_I fg$ on a specified interval I. Its squared norm is $\|f\|_2^2 = \int_I |f|^2$. This is mathematical signal energy; physical energy requires the units and conversion factor of the model. Vectors of coefficients remain columns, $^{\mathsf{T}}$ denotes transpose, and $D\Phi$ is the derivative matrix introduced in Part II.

Contents

Preface **i**

Starting point and notation **ii**

8 Accumulation in one and several dimensions **1**

8.1 From samples to a Riemann integral . . . 1
8.2 Accumulation and antiderivatives . . . 2
8.3 Totals, averages, and units . . . 3
8.4 Area accumulation and iterated integration . . . 3
8.5 A nonrectangular region and the integration order . . . 4
8.6 Further perspective: moments and long observation windows . . . 6
8.7 Questions and exercises . . . 8

9 Substitution, energy, and L2 approximation **9**

9.1 A substitution changes the differential . . . 9
9.2 Polar coordinates and radial data . . . 10
9.3 Function norms, energy, and an inner product . . . 11
9.4 What the space L^2 contains . . . 12
9.5 Best approximation in a finite function model . . . 14
9.6 Further perspective: a first harmonic and complex coordinates . . . 16
9.7 Questions and exercises . . . 17

From finite data to Fourier coordinates **18**

Lecture 8

Accumulation in one and several dimensions

A rate is amount per unit time; a spatial density is amount per unit area. Multiplying a local value by the corresponding cell size gives a cell contribution. Integration is the limit of adding these contributions.

8.1 From samples to a Riemann integral

Let $f : [\alpha, \beta] \to \mathbb{R}$ be bounded, with $\alpha < \beta$. A *tagged partition* consists of points $\alpha = t_0 < \cdots < t_n = \beta$ and sample locations $\xi_i \in [t_{i-1}, t_i]$. Set

$$\Delta t_i = t_i - t_{i-1}, \qquad |P| = \max_i \Delta t_i, \qquad S_P(f) = \sum_{i=1}^{n} f(\xi_i)\Delta t_i.$$

The number $|P|$ is the *mesh*; $S_P(f)$ is a *Riemann sum*.

Definition and an integrability criterion

The function f is *Riemann integrable* with integral A if, for every $\varepsilon > 0$, there is a $\delta > 0$ such that

$$|S_P(f) - A| < \varepsilon \quad \text{for every tagged partition with } |P| < \delta.$$

We write $A = \int_\alpha^\beta f(t)\,\mathrm{d}t$. Thus the limit is independent of both the partition and its sample choices.

Equivalently, writing $m_i = \inf_{[t_{i-1},t_i]} f$ and $M_i = \sup_{[t_{i-1},t_i]} f$, the difference $\sum_i (M_i - m_i)\Delta t_i$ can be made arbitrarily small. Continuous functions, and bounded piecewise continuous functions with finitely many pieces, are integrable.

Reasoning: continuity controls every sample choice. A continuous function on a closed bounded interval is uniformly continuous. For sufficiently small mesh, every $M_i - m_i$ is at most $\varepsilon/(\beta - \alpha)$. The upper-minus-lower sum is then at most ε, and every tagged sum lies between those bounds. □

Example 8.1. Recovering ingested data from a rate

An ingestion service has modeled input rate $q(t) = 2 + t$ MB/s over $0 \le t \le 2$, where t is the numerical time in seconds. We seek the amount received in this window. For n equal time cells, left and right endpoint sampling gives

$$S_{\text{left}} = \frac{2}{n}\sum_{k=0}^{n-1}\left(2 + \frac{2k}{n}\right) = 6 - \frac{2}{n},$$

$$S_{\text{right}} = \frac{2}{n}\sum_{k=1}^{n}\left(2 + \frac{2k}{n}\right) = 6 + \frac{2}{n}.$$

Both tend to 6 MB. With four cells they give 5.5 and 6.5 MB; monotonicity of q makes them lower and upper bounds. Midpoint sampling gives 6 exactly because this rate is affine.

Each product has units (MB/s) × s = MB. Summing rates alone has the wrong units. A finite telemetry record estimates the integral only with a model of the rate between samples; here that model is stated explicitly.

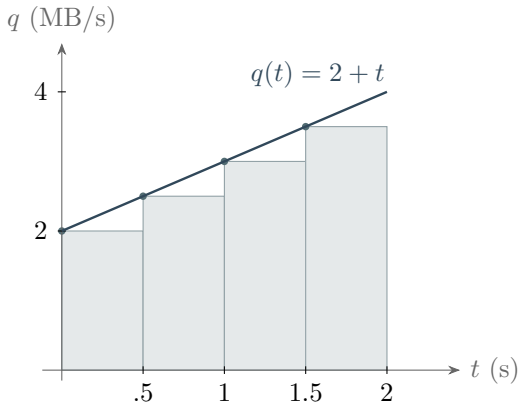


FIGURE 8.1. Four left-endpoint cells estimate 5.5 MB ingested in two seconds. Each height is a rate and each width is 0.5 s. The exact area under $q(t) = 2 + t$ is 6 MB.

8.2 Accumulation and antiderivatives

For Riemann integrable functions and fixed scalars u, v,

$$\int_\alpha^\beta (uf + vg) = u\int_\alpha^\beta f + v\int_\alpha^\beta g,$$

$$\int_\alpha^\beta f = \int_\alpha^\gamma f + \int_\gamma^\beta f \quad (\alpha \le \gamma \le \beta), \qquad f \le g \implies \int_\alpha^\beta f \le \int_\alpha^\beta g.$$

These properties follow from the corresponding weighted sums. Define oriented limits by $\int_b^a f = -\int_a^b f$ and $\int_a^a f = 0$.

A signed rate gives *net* accumulation. For normalized velocity $v(t) = 1 - t$ on $[0, 2]$, displacement is $\int_0^2 v = 0$, whereas distance travelled is $\int_0^2 |v| = 1$. The absolute value changes the quantity being measured.

Fundamental theorem of calculus

If f is continuous on $[\alpha, \beta]$, then $F(t) = \int_\alpha^t f(u)\,\mathrm{d}u$ is differentiable inside the interval and $F'(t) = f(t)$. If G is continuous on the closed interval with $G' = f$ in its interior, then

$$\int_\alpha^\beta f(t)\,\mathrm{d}t = G(\beta) - G(\alpha). \tag{8.1}$$

Reasoning: differentiation recovers the local rate. For small nonzero h,

$$\frac{F(t+h) - F(t)}{h} = \frac{1}{h}\int_t^{t+h} f(u)\,\mathrm{d}u.$$

For either sign of h, this is the average over the interval between t and $t + h$. Continuity makes it tend to $f(t)$. Since $(G - F)' = 0$, the mean value theorem makes $G - F$ constant, proving (8.1). □

For the ingestion rate, $G(t) = 2t + t^2/2$ gives $G(2) - G(0) = 6$. Antiderivatives compute the same accumulated amount as the Riemann-sum limit.

8.3 Totals, averages, and units

For an interval of length $L = \beta - \alpha > 0$, define

$$\overline{f} = \frac{1}{L}\int_{\alpha}^{\beta} f(t)\,\mathrm{d}t.$$

This *mean value* is the constant producing the same total integral value over the interval. The integral has units of f times units of t; the mean has units of f. Thus the mean ingestion rate is $6/2 = 3$ MB/s.

Example 8.2. Energy used during a processing transient

After a batch completes, a server's power above its idle baseline is modeled by $P(t) = P_0 e^{-kt}$, with $P_0 = 6$ W and $k = 2\,\mathrm{s}^{-1}$. We seek the extra energy used during the first second. Since watts are joules per second,

$$Q(T) = \int_0^T P_0 e^{-kt}\,\mathrm{d}t = \frac{P_0}{k}(1 - e^{-kT}), \qquad Q(1) = 3(1 - e^{-2})\ \mathrm{J} \approx 2.594\ \mathrm{J}.$$

The mean excess power is $Q(1)/(1\,\mathrm{s}) \approx 2.594$ W. Its numerical value agrees with the energy only because the window is one second. Also $Q'(T) = P(T)$: extending the window adds energy at the power present at its endpoint. The exponent kt is dimensionless.

For sampled data the time average is approximated by $\sum_i f(\xi_i)\Delta t_i / L$. Equal widths give the ordinary sample average; unequal widths require time weights. For example, values 2 and 4 held for 1.5 s and 0.5 s have mean 2.5, although their unweighted mean is 3.

8.4 Area accumulation and iterated integration

On $R = [a, b] \times [c, d]$, a double Riemann sum is $\sum_{i,j} f(\xi_{ij}, \eta_{ij})\Delta x_i \Delta y_j$. Its common limit as the maximum cell diameter tends to zero is the *double integral* $\iint_R f\,\mathrm{d}A$, when this limit exists independently of the partitions and sample choices. In particular, $\iint_R 1\,\mathrm{d}A = (b - a)(d - c)$. A nonnegative density ρ gives total amount $M = \iint_R \rho\,\mathrm{d}A$.

Iterated integration on a rectangle

If f is continuous on the closed rectangle R, then

$$\iint_R f\,\mathrm{d}A = \int_a^b \left[\int_c^d f(x, y)\,\mathrm{d}y\right]\mathrm{d}x = \int_c^d \left[\int_a^b f(x, y)\,\mathrm{d}x\right]\mathrm{d}y. \tag{8.2}$$

This is the continuous rectangular case of Fubini's theorem.

For fixed x, the inner integral accumulates a vertical slice per unit horizontal width. Multiplying by Δx gives a strip contribution. The finite cell sums can be grouped in either order; continuity controls their errors uniformly as the grid is refined.

Example 8.3. A spatial request map: total and average density

A telemetry model describes request locations during one fixed reporting window. On the rectangle $R = [0,2] \times [0,1]$, coordinates are numerical distances in kilometres, and $\rho(x,y) = 1 + x + 2y$ is the expected number of requests per square kilometre, in thousands. We need a total for capacity planning and a mean density for comparing regions.

$$\int_0^1 (1 + x + 2y)\,\mathrm{d}y = 2 + x, \qquad M = \int_0^2 (2 + x)\,\mathrm{d}x = 6,$$
$$\int_0^2 (1 + x + 2y)\,\mathrm{d}x = 4 + 4y, \quad M = \int_0^1 (4 + 4y)\,\mathrm{d}y = 6.$$

Both orders predict 6000 requests. The area is $2\,\mathrm{km}^2$, so the mean density is 3000 requests per square kilometre. The density is a spatial model for the specified window; it is not itself a count in a pixel or a rate per second.

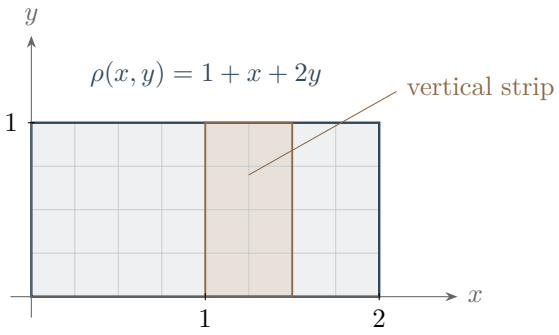


FIGURE 8.2. The density map is integrated over area cells. A vertical slice contributes $2 + x$ thousand requests per kilometre of horizontal width. The shaded strip indicates which slices are accumulated in the outer integral.

8.5 A nonrectangular region and the integration order

Define an integral over a bounded region D by extending the integrand by zero to a containing rectangle, whenever that extension is Riemann integrable. This holds for continuous f on a closed region whose boundary has area zero. In particular, for

$$D = \{(x,y) : a \le x \le b,\ \ell(x) \le y \le u(x)\},$$

where ℓ, u are continuous and $\ell \le u$, slice integration gives

$$\iint_D f\,\mathrm{d}A = \int_a^b \left[\int_{\ell(x)}^{u(x)} f(x,y)\,\mathrm{d}y\right] \mathrm{d}x.$$

Changing order requires another description of this same set. If the boundary formula changes along the slices, split the domain into pieces.

Example 8.4. Aggregating over a triangular service region

For another reporting window, suppose the request density is $\rho(x, y) = 1 + x$ in the same units. The service region has vertices $(0, 0)$, $(2, 0)$, $(0, 1)$. Its sloping edge is $y = 1 - x/2$. A vertical slice has limits $0 \le y \le 1 - x/2$, so

$$\begin{aligned} M &= \int_0^2 \left[\int_0^{1 - x/2} (1 + x)\, \mathrm{d}y \right] \mathrm{d}x \\ &= \int_0^2 (1 + x)(1 - x/2)\, \mathrm{d}x = \left[x + \frac{x^2}{4} - \frac{x^3}{6} \right]_0^2 = \frac{5}{3}. \end{aligned}$$

Solving the same boundary for x gives $0 \le y \le 1$ and $0 \le x \le 2(1 - y)$. Hence

$$M = \int_0^1 \left[\int_0^{2(1-y)} (1 + x)\, \mathrm{d}x \right] \mathrm{d}y = \int_0^1 \left(2(1 - y) + 2(1 - y)^2 \right) \mathrm{d}y = \frac{5}{3}.$$

The prediction is $5/3$ thousand requests. The region has area $1\,\mathrm{km}^2$, so its mean density is numerically $5/3$ thousand per square kilometre. Using the bounding rectangle would predict 4 thousand and include requests outside the service region.

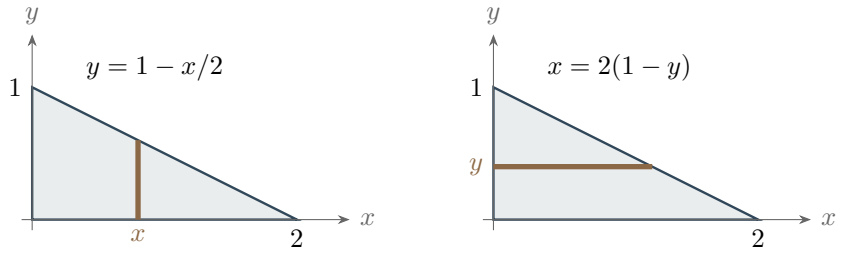


FIGURE 8.3. The same triangular region described by vertical and horizontal slices, with equal axis scales. At $x = 0.8$ the vertical slice ends at $y = 0.6$; at $y = 0.4$ the horizontal slice ends at $x = 1.2$.

The geometric check precedes the antiderivative: fix the outer coordinate, determine the entry and exit points of its slice, then integrate. Reversing the order can also simplify the integrand's antiderivative, as Exercise 8.6 shows.

8.6 Further perspective: moments and long observation windows

For a nonnegative density of total amount $M > 0$, define its *first moments* and *weighted centre* by

$$m_x = \iint_D x\rho \,\mathrm{d}A, \quad m_y = \iint_D y\rho \,\mathrm{d}A, \qquad \overline{x} = m_x/M, \quad \overline{y} = m_y/M.$$

Here m_x means the moment of the x coordinate. Each moment accumulates position times amount; dividing by total amount returns a position.

Example 8.5. One location as a compact spatial summary

Represent the requests in Example 8.4 by one location (u, v). Choose it to minimize the density-weighted squared coordinate error

$$C(u, v) = \iint_D ((x-u)^2 + (y-v)^2)\rho(x, y) \,\mathrm{d}A.$$

Dividing by M would give mean squared error and the same minimizer. The data supply

$$m_x = \int_0^2 x(1+x)(1-x/2) \,\mathrm{d}x = \frac{4}{3}, \qquad m_y = \int_0^2 \frac{1+x}{2}(1-x/2)^2 \,\mathrm{d}x = \frac{1}{2}.$$

Since $M = 5/3$, the weighted centre $(\overline{x}, \overline{y})$ is $(4/5, 3/10)$ km. Expanding the objective and completing squares gives

$$C(u, v) = C(\overline{x}, \overline{y}) + M\big((u-\overline{x})^2 + (v-\overline{y})^2\big).$$

It is therefore the unique global optimum in the weighted centre. Equivalently, the Hessian is $2MI_2$, with both eigenvalues positive, as in Part II.

The geometric centroid $(2/3, 1/3)$ would treat locations uniformly. Weighting by the modeled request density shifts the summary right and slightly down, towards the more active part of the region.

Improper integration and integration by parts

If f is integrable on every $[0, T]$, define

$$\int_0^\infty f(t)\,\mathrm{d}t := \lim_{T\to\infty} \int_0^T f(t)\,\mathrm{d}t$$

when the limit is finite. This is an *improper integral*; an infinite endpoint denotes a limit, not an antiderivative evaluation.

For $u, v \in C^1([\alpha, \beta])$, the product rule and the fundamental theorem give *integration by parts*:

$$\int_\alpha^\beta uv'\,\mathrm{d}t = [uv]_\alpha^\beta - \int_\alpha^\beta u'v\,\mathrm{d}t.$$

First apply this identity on a finite interval, then check any infinite-endpoint limit.

Example 8.6. Choosing a window for a decaying transient

After a trigger, an excess input rate is modeled in normalized units by $r(t) = e^{-kt}$, $k > 0$, $t \geq 0$. We want to stop recording while missing at most a fraction $\varepsilon \in (0, 1)$ of the total input. Integration gives

$$\int_0^T r(t)\,\mathrm{d}t = \frac{1 - e^{-kT}}{k}, \qquad \int_0^\infty r(t)\,\mathrm{d}t = \frac{1}{k}, \qquad \int_T^\infty r(t)\,\mathrm{d}t = \frac{e^{-kT}}{k}.$$

The omitted fraction is e^{-kT}, so it suffices and is necessary that $T \geq k^{-1}\log(1/\varepsilon)$. For a 1% budget, $T \geq (\log 100)/k$.

To summarize when the input arrives, compute its first time moment. Integration by parts with $u = t$, $v' = e^{-kt}$ yields

$$\int_0^T te^{-kt}\,\mathrm{d}t = -\frac{Te^{-kT}}{k} + \frac{1 - e^{-kT}}{k^2} \longrightarrow \frac{1}{k^2}.$$

Dividing by the total $1/k$ gives mean arrival time $1/k$. Also $\int_0^\infty r^2\,\mathrm{d}t = 1/(2k)$, the squared-integral quantity used for signal energy in Lecture 9.

Decay to zero alone is insufficient: $\int_0^T (1+t)^{-1}\,\mathrm{d}t = \log(1+T) \to \infty$. The defining integrals of later transforms likewise require convergence conditions.

8.7 Questions and exercises

Questions for review

T8.1. Define a tagged partition, its mesh, and a Riemann integral. Which choices must give the same limit?

T8.2. State the upper-and-lower-sum criterion. Explain how uniform continuity guarantees integrability.

T8.3. State the fundamental theorem with its hypotheses and derive it from a short-interval average.

T8.4. Distinguish net accumulation, accumulated absolute magnitude, and a mean value. How do units and unequal sampling intervals enter?

T8.5. Define a double integral through cell contributions and state the continuous rectangular case of Fubini's theorem. What does an inner integral return?

T8.6. Describe a triangular region in both integration orders. Why does changing order require new limits?

Exercises

8.1. An input rate is $q(t) = 3+2t$ MB/s for numerical time $t \in [0, 1]$ in seconds. Compute left and right sums for n equal cells and their common limit. Compare with the antiderivative and midpoint sum.

8.2. For normalized velocity $v(t) = 2t - 1$ on $[0, 1]$, find displacement, total distance, and mean velocity. Explain the role of the absolute value.

8.3. For input rate $q(t) = 4e^{-t}$ on $[0, 2]$, find $Q(T) = \int_0^T q(t)\,\mathrm{d}t$ and verify $Q'(T) = q(T)$. Find the full-window total and mean rate.

8.4. On $R = [0, 1] \times [0, 2]$, use density $\rho(x, y) = 2 + x + y$. Compute its total in both orders and find the average density.

8.5. On the triangle with vertices $(0, 0)$, $(1, 0)$, $(0, 1)$, use $\rho(x, y) = 1 + y$. Describe the region in both orders; compute its total and average density.

8.6. Let $D = \{(x, y) : 0 \le x \le 1,\ x \le y \le 1\}$. Reverse the order in $\int_0^1 \int_x^1 e^{y^2}\,\mathrm{d}y\,\mathrm{d}x$ and evaluate it. Why is one order easier?

8.7. (*) For the request map in Example 8.3, calculate both first moments and the weighted centre. Compare with $(1, 1/2)$ and interpret the displacement.

8.8. (*) For $r(t) = e^{-3t}$, calculate the improper integrals of r, r^2, and tr. Find the tail after T and the smallest $T \ge 0$ making it at most $1/300$. What fraction of the total does this omit?

Lecture 9

Substitution, energy, and L^2 approximation

A coordinate change alters the cell weights in an integral. Once those weights are correct, integration defines function lengths and inner products. The projection geometry of Part I then solves continuous approximation problems.

9.1 A substitution changes the differential

If $\phi \in C^1([a,b])$ and f is continuous on an interval containing $\phi([a,b])$, then

$$\int_{\phi(a)}^{\phi(b)} f(x)\,\mathrm{d}x = \int_a^b f(\phi(u))\phi'(u)\,\mathrm{d}u.$$

The formula follows by applying the chain rule to an antiderivative of f. The limits are oriented; monotonicity of ϕ is not needed for this identity. When ϕ is increasing or decreasing and maps onto a geometric interval, its unsigned length element is $|\phi'(u)|\,\mathrm{d}u$.

For example, $u = t^2,\ \mathrm{d}u = 2t\,\mathrm{d}t$, gives $\int_0^1 2te^{-t^2}\,\mathrm{d}t = \int_0^1 e^{-u}\,\mathrm{d}u = 1 - e^{-1}$. Both limits and integrand are expressed in the new variable.

For a planar coordinate map $\Phi(u,v) = (x(u,v), y(u,v))$, set

$$D\Phi = \begin{pmatrix} x_u & x_v \\ y_u & y_v \end{pmatrix}, \qquad J_\Phi = \det D\Phi = x_u y_v - x_v y_u.$$

The matrix is the derivative from Part II; J_Φ is its *Jacobian determinant*. They are different objects.

Change of variables for area

Let U, D be closed bounded regions with piecewise smooth boundaries. Suppose Φ extends to a C^1 map near U, maps U onto D, is one-to-one on the interior of U, and has $J_\Phi \neq 0$ there. Suppose also that the image of the boundary has area zero. For continuous f on D,

$$\iint_D f(x,y)\,\mathrm{d}x\,\mathrm{d}y = \iint_U f(\Phi(u,v))\,|J_\Phi(u,v)|\,\mathrm{d}u\,\mathrm{d}v. \tag{9.1}$$

Thus boundary coincidences are allowed; interior multiple coverage is excluded.

Reasoning: local linearization explains the area weight. The derivative maps a small rectangle's edges to $\Phi_u \Delta u$ and $\Phi_v \Delta v$. Their parallelogram has area $|\det(\Phi_u, \Phi_v)|\Delta u \Delta v$. Differentiability supplies this first-order area scale. The absolute value removes orientation: reversing a coordinate direction changes the determinant's sign but not cell area. □

Example 9.1. A resampled image needs physical cell areas

An image is stored on the parameter square $0 \le u, v \le 1$. Calibration associates these grid labels with physical coordinates

$$\Phi(u, v) = (2u + v, v), \qquad D\Phi = \begin{pmatrix} 2 & 1 \\ 0 & 1 \end{pmatrix}, \qquad J_\Phi = 2.$$

In fixed normalized units, the field per unit physical area is $f(x, y) = x + y$. We need its total and mean over the physical image. The inverse $v = y$, $u = (x - y)/2$ proves one-to-one coverage. The square's vertices map to $(0, 0), (2, 0), (3, 1), (1, 1)$.

$$|D| = \int_0^1 \int_0^1 2 \, \mathrm{d}u \, \mathrm{d}v = 2, \qquad \iint_D f \, \mathrm{d}A = \int_0^1 \int_0^1 (2u + 2v) \, 2 \, \mathrm{d}u \, \mathrm{d}v = 4.$$

The spatial mean is $4/2 = 2$. A stored sample is the value $f(\Phi(u, v)) = 2u+2v$; its contribution to the total also requires the physical cell area. Omitting the factor 2 underestimates the total by half. Here it cancels only when both numerator and area denominator of a mean are transformed consistently.

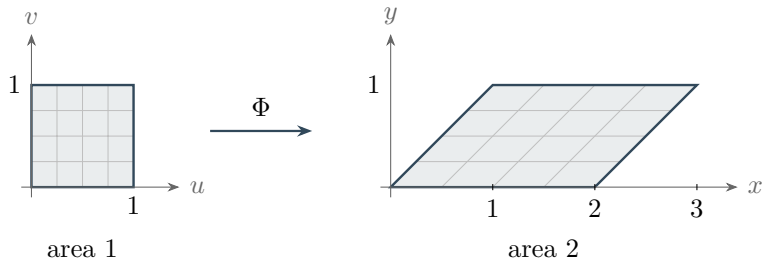


FIGURE 9.1. The stored square grid and its physical image under $\Phi(u, v) = (2u + v, v)$, with equal length scales. Every image cell has twice the parameter-cell area; its field value alone does not include this factor.

9.2 Polar coordinates and radial data

For $x = r \cos\theta$, $y = r \sin\theta$,

$$D\Phi = \begin{pmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{pmatrix}, \qquad J_\Phi = r.$$

Consequently, with $r \ge 0$ and θ in radians,

$$\mathrm{d}A = r \, \mathrm{d}r \, \mathrm{d}\theta. \tag{9.2}$$

Equal changes in angle cover wider cells at larger radii.

The rectangle $[0, R] \times [0, 2\pi]$ covers the disk of radius R. Its interior maps one-to-one onto the disk minus the centre, boundary circle, and angular seam. Those omitted sets have zero area. The entire edge $r = 0$ maps to the centre, and the two angle edges map to the same radius; these are precisely the permitted boundary exceptions.

Example 9.2. Total brightness in a circular crop

A spot image has normalized brightness $I(x, y) = e^{-(x^2+y^2)}$. Crop it to $D = \{x^2 + y^2 \le 4\}$. To aggregate brightness, use its radial formula together with the area weight:

$$M = \int_0^{2\pi} \int_0^2 e^{-r^2} r \, \mathrm{d}r \, \mathrm{d}\theta = 2\pi \left[-\frac{1}{2} e^{-r^2} \right]_0^2 = \pi(1 - e^{-4}).$$

The crop area is 4π, so its mean brightness is $(1 - e^{-4})/4 \approx 0.245$. Averaging uniformly over radius would answer a different question: equal-width annuli contain different areas.

For any continuous radial profile F,

$$\iint_{x^2+y^2 \le R^2} F(\sqrt{x^2 + y^2}) \, \mathrm{d}A = 2\pi \int_0^R F(r) r \, \mathrm{d}r.$$

For this spot, the outer annulus $1 \le r \le 2$ contributes $\pi(e^{-1} - e^{-4})$, about 36% of the crop's total brightness despite its lower intensity.

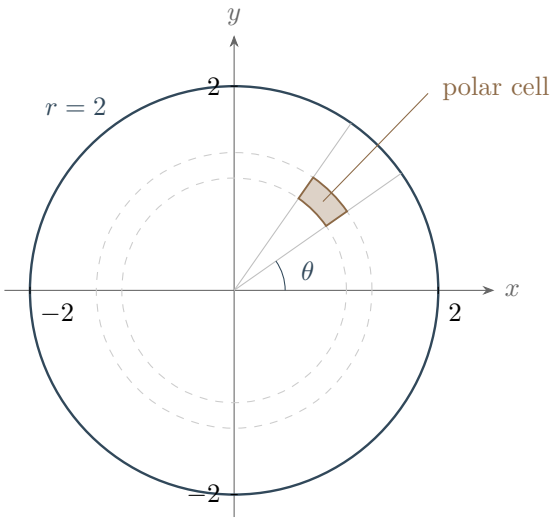


FIGURE 9.2. The cell $1.1 \le r \le 1.35$, $35° \le \theta \le 55°$ has exact area $\frac{1}{2}(1.35^2 - 1.1^2)(\pi/9)$. Angles enter the integral in radians. The first-order cell area is $r \, \Delta r \, \Delta\theta$.

9.3 Function norms, energy, and an inner product

Let $C(I)$ be the real continuous functions on $I = [\alpha, \beta]$, $\alpha < \beta$, with pointwise addition and scalar multiplication. A *norm* satisfies positivity, $\|f\| = 0$ iff $f = 0$, homogeneity $\|cf\| = |c| \, \|f\|$, and the triangle inequality $\|f + g\| \le \|f\| + \|g\|$. Three norms on $C(I)$ are

$$\|f\|_1 = \int_I |f(t)| \, \mathrm{d}t, \qquad \|f\|_2 = \left(\int_I |f(t)|^2 \, \mathrm{d}t \right)^{1/2}, \qquad \|f\|_\infty = \max_{t \in I} |f(t)|.$$

They measure accumulated absolute magnitude, square-integrated size, and peak magnitude, respectively.

The integral defines a real inner product:

$$\langle f, g \rangle = \int_I f(t) g(t) \, \mathrm{d}t, \qquad \|f\|_2^2 = \langle f, f \rangle . \tag{9.3}$$

Linearity and symmetry follow from integration. Positivity is strict for nonzero continuous f: if $f(t_0) \neq 0$, continuity bounds $|f|$ away from zero on an interval of positive length, so $\int_I f^2 > 0$.

Write $L = \beta - \alpha$. *Signal energy* is $E_f = \|f\|_2^2$; the *root mean square* (RMS) is $f_{\mathrm{RMS}} = \|f\|_2 / \sqrt{L}$. For physical time, their units are amplitude squared times time and amplitude, respectively. If f is voltage across constant resistance $R_0 > 0$, physical energy is $R_0^{-1} \int_I f^2 \, \mathrm{d}t$. In Example 8.2, the integrand was already power and required no squaring.

Example 9.3. Summarizing a record: mean, RMS, and peak

For the record $s(t) = 1 + t + t^2$ on $[-1, 1]$, a dashboard needs its mean level, typical squared magnitude, and peak. Since $s(t) = (t + 1/2)^2 + 3/4 > 0$,

$$\int_{-1}^{1} s \, \mathrm{d}t = \|s\|_1 = \frac{8}{3}, \qquad \overline{s} = \frac{4}{3}.$$

Squaring first gives $s^2 = 1 + 2t + 3t^2 + 2t^3 + t^4$. Odd terms integrate to zero, hence

$$E_s = \|s\|_2^2 = 2 + 2 + \frac{2}{5} = \frac{22}{5}, \qquad \|s\|_2 = \sqrt{\frac{22}{5}}, \qquad s_{\mathrm{RMS}} = \sqrt{\frac{11}{5}}.$$

The stationary point $t = -1/2$ is a minimum; the endpoint values are 1 and 3. Thus $\|s\|_\infty = 3$. The mean $4/3$ records the constant level, RMS ≈ 1.483 summarizes squared magnitude, and peak 3 determines the required amplitude range. A small average alone would not certify a small peak.

Reasoning: Cauchy–Schwarz and the triangle inequality. For $\|g\|_2 > 0$, choose $c = \langle f, g \rangle / \|g\|_2^2$. Then

$$0 \le \|f - cg\|_2^2 = \|f\|_2^2 - \frac{\langle f, g \rangle^2}{\|g\|_2^2}, \quad \text{so} \quad |\langle f, g \rangle| \le \|f\|_2 \, \|g\|_2 .$$

The zero case is immediate. Applying this bound to $\|f + g\|_2^2 = \|f\|_2^2 + 2\langle f, g \rangle + \|g\|_2^2$ gives $\|f + g\|_2 \le \|f\|_2 + \|g\|_2$. □

Functions are *orthogonal* if $\langle f, g \rangle = 0$; then $\|f + g\|_2^2 = \|f\|_2^2 + \|g\|_2^2$. On $[-1, 1]$, 1 and t are orthogonal because $\int_{-1}^{1} t \, \mathrm{d}t = 0$. Their supports overlap: cancellation, not separation, makes the cross term vanish.

9.4 What the space L^2 contains

On $[-1, 1]$, let $f = 0$ everywhere and let $g = 0$ except at $g(0) = 1$. These are different functions but $\int |f - g|^2 = 0$. Thus the integral norm on a space allowing such functions requires an identification.

A set has *measure zero* if, for every $\varepsilon > 0$, it can be covered by countably many intervals of total length below ε. A statement holds *almost everywhere* (a.e.) if its exceptional set has measure zero. Finite and countable sets have this property; some uncountable sets do too.

Definition of $L^2(I)$

For measurable real functions on I, put $f \sim g$ if $f = g$ a.e. Then

$$L^2(I) = \left\{ f : \int_I |f|^2 < \infty \right\} / \sim .$$

An element is an equivalence class. We write f for a representative when computing, but changing it on a null set changes neither its class nor any inner product.
Here the integral is the Lebesgue integral. Its construction is outside the course; for the continuous and bounded piecewise continuous functions used in our finite-interval calculations, it equals the Riemann integral.

The integral inner product is well-defined: $2|fg| \le |f|^2 + |g|^2$ makes the product integrable, and $\int |f|^2 = 0$ iff $f = 0$ a.e. Consequently it is positive definite on classes. The Cauchy–Schwarz argument above applies unchanged.

Convergence in L^2 means $\|f_n - f\|_2 \to 0$. The space is *complete*: if $\|f_n - f_m\|_2 \to 0$ as $m, n \to \infty$, there is an $f \in L^2$ with $\|f_n - f\|_2 \to 0$. We state completeness without proof. Continuous functions alone are not complete in this norm: continuous ramps can approach a jump. Finite projection below needs only a finite-dimensional model, while infinite expansions will use limits in L^2.

Example 9.4. A short pulse erased by compression

Suppose a compressor reconstructs zero on a record containing the pulse

$$f_n(t) = \max(1 - n|t|, 0), \qquad -1 \le t \le 1, \quad n = 1, 2, \dots .$$

The reconstruction error is exactly f_n. Its height is 1 and its support, the closure of where it is nonzero, is $[-1/n, 1/n]$. Therefore

$$\|f_n\|_2^2 = 2 \int_0^{1/n} (1 - nt)^2 \, \mathrm{d}t = \frac{2}{3n}, \qquad \mathrm{RMSE} = \sqrt{\frac{1}{3n}} \longrightarrow 0.$$

An RMSE budget of 0.1 permits erasing these pulses when $n \ge 34$. Yet $f_n(0) = 1$ and $\|f_n\|_\infty = 1$ for every n. There is convergence to zero in L^2, but neither uniform convergence nor pointwise convergence to zero at the centre. A requirement to preserve unit-height events therefore needs a peak or event-specific criterion as well.

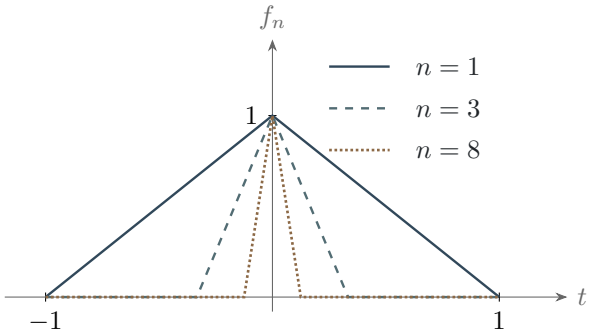


FIGURE 9.3. Erasing a pulse of height 1 leaves squared errors 2/3, 2/9, and 1/12 for $n = 1, 3, 8$. Their peaks agree; their widths and integrated errors differ.

The definition also applies on infinite intervals. The response e^{-kt}, $k > 0$, belongs to $L^2([0,\infty))$ because $\int_0^\infty e^{-2kt}\,\mathrm{d}t = 1/(2k)$. The constant 1 belongs to L^2 on a finite interval but not on this infinite interval. The domain is part of the energy specification.

9.5 Best approximation in a finite function model

Let $s \in L^2(I)$ be a real target and let $\phi_1, \dots, \phi_m \in L^2(I)$ be fixed, linearly independent functions. Independence here means $\sum_j z_j\phi_j = 0$ a.e. only when every $z_j = 0$. The unknowns are real coefficients:

$$p_{\boldsymbol{c}} = \sum_{j=1}^{m} c_j\phi_j, \qquad W = \operatorname{span}\{\phi_1, \dots, \phi_m\}, \qquad E(\boldsymbol{c}) = \|s - p_{\boldsymbol{c}}\|_2^2 .$$

Define the *Gram matrix* and *data inner products* by $G_{ij} = \langle\phi_i, \phi_j\rangle$ and $d_i = \langle s, \phi_i\rangle$. Expanding the finite square gives

$$E(\boldsymbol{c}) = \|s\|_2^2 - 2\boldsymbol{d}^{\mathsf{T}}\boldsymbol{c} + \boldsymbol{c}^{\mathsf{T}}G\boldsymbol{c}. \tag{9.4}$$

Thus $\nabla E = 2(G\boldsymbol{c} - \boldsymbol{d})$ and $H_E = 2G$. Moreover, $\boldsymbol{z}^{\mathsf{T}}G\boldsymbol{z} = \left\|\sum_j z_j\phi_j\right\|_2^2 > 0$ for $\boldsymbol{z} \neq 0$. By Parts I–II, G is invertible and E is a strictly convex quadratic.

Normal equations and residual orthogonality

There is a unique coefficient vector $\boldsymbol{c}_*$ and a unique best approximation $p = p_{\boldsymbol{c}_*}$ as an L^2 element. They are characterized by

$$G\boldsymbol{c}_* = \boldsymbol{d} \quad\Longleftrightarrow\quad \langle s - p, \phi_i\rangle = 0 \quad (i = 1, \dots, m). \tag{9.5}$$

In other words, the residual $s - p$ is orthogonal to the whole model space W.

Reasoning: a finite projection is globally optimal. The normal equations have a unique solution since G is invertible. For any $q \in W$, the decomposition $s - q = (s - p) + (p - q)$ has orthogonal terms, so

$$\|s - q\|_2^2 = \|s - p\|_2^2 + \|p - q\|_2^2 . \tag{9.6}$$

The last term is nonnegative and vanishes precisely when $q = p$ in L^2. □

For an orthogonal basis, $c_j = \langle s, \phi_j\rangle \,/\, \|\phi_j\|_2^2$. For an *orthonormal* basis q_j, the coefficients are simply $\gamma_j = \langle s, q_j\rangle$, and

$$p = \sum_{j=1}^{m} \gamma_j q_j, \qquad \|s - p\|_2^2 = \|s\|_2^2 - \sum_{j=1}^{m} |\gamma_j|^2 .$$

If the original list is dependent, reduce it to a basis. The best function remains unique, although its coefficients in the redundant list need not be.

Example 9.5. Storing a record as an offset and a slope

Compress the record $s(t) = 1 + t + t^2$ on $[-1, 1]$ to $p(t) = a + bt$, minimizing (1). The ordered basis is $\phi_1 = 1$, $\phi_2 = t$; the coefficient column is $(a, b)^{\mathsf{T}}$. Its entries are computed from the fixed functions:

$$G = \begin{pmatrix} \int_{-1}^{1} 1\,\mathrm{d}t & \int_{-1}^{1} t\,\mathrm{d}t \\ \int_{-1}^{1} t\,\mathrm{d}t & \int_{-1}^{1} t^2\,\mathrm{d}t \end{pmatrix} = \begin{pmatrix} 2 & 0 \\ 0 & 2/3 \end{pmatrix}, \qquad \boldsymbol{d} = \begin{pmatrix} \int_{-1}^{1} s\,\mathrm{d}t \\ \int_{-1}^{1} st\,\mathrm{d}t \end{pmatrix} = \begin{pmatrix} 8/3 \\ 2/3 \end{pmatrix}.$$

The basis is orthogonal but not normalized. Solving gives

$$a_* = \frac{4}{3}, \qquad b_* = 1, \qquad p(t) = \frac{4}{3} + t, \qquad r(t) = t^2 - \frac{1}{3}.$$

Thus the stored offset is $4/3$ and the reconstructed rise per unit time is 1. The checks $\int_{-1}^{1} r = 0$ and $\int_{-1}^{1} tr = 0$ confirm residual orthogonality. The model preserves the mean; it overpredicts near the middle and underpredicts near the ends.

$$\|r\|_2^2 = \int_{-1}^{1} \left(t^4 - \frac{2}{3}t^2 + \frac{1}{9} \right) \mathrm{d}t = \frac{8}{45}, \qquad \mathrm{RMSE} = \frac{2}{3\sqrt{5}} \approx 0.298.$$

Completing the parameter square gives a second optimality confirmation:

$$E(a, b) = \frac{8}{45} + 2\left(a - \frac{4}{3}\right)^2 + \frac{2}{3}(b - 1)^2. \tag{9.7}$$

In the normalized basis $q_0 = 1/\sqrt{2}$, $q_1 = \sqrt{3/2}\, t$, the same reconstruction has coefficients $\gamma_0 = 4\sqrt{2}/3$, $\gamma_1 = \sqrt{2/3}$. Indeed, $a = \gamma_0/\sqrt{2}$ and $b = \sqrt{3/2}\,\gamma_1$. The retained energy is $\gamma_0^2 + \gamma_1^2 = 38/9$; adding the error $8/45$ recovers $22/5$. Changing basis changes the stored coordinates, not the reconstructed function.

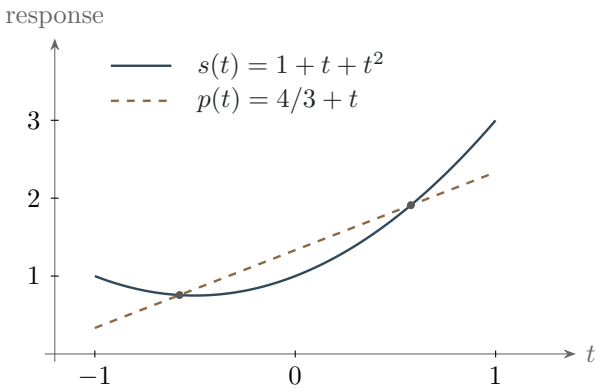


FIGURE 9.4. The two-parameter reconstruction $p(t) = 4/3 + t$ crosses the record at $t = \pm 1/\sqrt{3}$. It overestimates the middle and underestimates the ends. The signed residual has zero mean, while its squared integral is $8/45$.

Sampling this same new record at $t = -1, 0, 1$ gives $(1, 1, 3)$. An equally weighted three-point fit yields $5/3 + t$, not $4/3 + t$. It optimizes a different weighting of the record. To approximate the integral objective on a refining grid, use $\sum_i (s(\xi_i) - a - b\xi_i)^2 \Delta t_i$. The final table makes the connection with sampled model matrices explicit.

9.6 Further perspective: a first harmonic and complex coordinates

Example 9.6. Compressing a repeated ramp to its first harmonic

A cyclic record rises as $s(t) = t$ for $-\pi < t < \pi$ and resets at the period boundary; t is normalized phase. Store its mean and one sine–cosine pair, reconstructing

$$p(t) = \frac{a_0}{2} + a_1 \cos t + b_1 \sin t.$$

The model functions $1, \cos t, \sin t$ are fixed. They are orthogonal, with squared norms $2\pi, \pi, \pi$ on $[-\pi, \pi]$; for instance, $\sin^2 t = (1 - \cos 2t)/2$ gives the sine norm. Endpoint values assigned to the ramp do not affect these integrals.

Oddness gives $\int_{-\pi}^{\pi} s = 0$ and $\int_{-\pi}^{\pi} s \cos t = 0$, so $a_0 = a_1 = 0$. Integration by parts gives

$$b_1 = \frac{1}{\pi} \int_{-\pi}^{\pi} t \sin t \, \mathrm{d}t = \frac{2}{\pi} [-t \cos t + \sin t]_0^{\pi} = 2.$$

The best reconstruction is therefore $p(t) = 2 \sin t$, with squared error

$$\|s - p\|_2^2 = \|s\|_2^2 - \|p\|_2^2 = \frac{2\pi^3}{3} - 4\pi, \qquad \mathrm{RMSE} = \sqrt{\frac{\pi^2}{3} - 2} \approx 1.136.$$

This coefficient minimizes integrated error within the chosen model. Near the reset, $s(t) - p(t) \to \pi$ as $t \to \pi^-$: one smooth harmonic cannot reproduce the sharp edge. Lecture 10 adds higher harmonics and quantifies the remaining error.

For any real target, the constant multiplier is its mean: $a_0/2 = (2\pi)^{-1} \int_{-\pi}^{\pi} s(t) \, \mathrm{d}t$. The notation $a_0/2$ accommodates the constant function's squared norm 2π, twice the squared norm of each nonzero integer sine or cosine.

Complex coordinates and the coefficient sign

For complex signals we retain Part I's convention

$$\langle f, g \rangle = \int_{-\pi}^{\pi} f(t) \overline{g(t)} \, \mathrm{d}t.$$

It is linear in the first argument and conjugate-linear in the second. The functions $e_n(t) = e^{int}/\sqrt{2\pi}$, $n \in \mathbb{Z}$, satisfy

$$\langle e_n, e_m \rangle = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{i(n-m)t} \, \mathrm{d}t = \begin{cases} 1, & n = m, \\ 0, & n \neq m. \end{cases}$$

For $n \neq m$, the exponential antiderivative has equal endpoint values. Thus finite orthogonal projection uses $\gamma_n = \langle s, e_n \rangle = (2\pi)^{-1/2} \int_{-\pi}^{\pi} s(t) e^{-int} \, \mathrm{d}t$. Conjugation supplies the minus sign.

For the unnormalized basis e^{int}, use $\widehat{s}_n = (2\pi)^{-1} \int_{-\pi}^{\pi} s(t) e^{-int} \, \mathrm{d}t$. They satisfy $\widehat{s}_n = \gamma_n/\sqrt{2\pi}$, so $\widehat{s}_n e^{int} = \gamma_n e_n(t)$. Normalization changes a coefficient and its basis function together.

9.7 Questions and exercises

Questions for review

T9.1. State a one-variable substitution rule and an area change-of-variables rule with their hypotheses. Explain the signed derivative and the absolute determinant.

T9.2. Derive the polar area factor. Which parts of the parameter rectangle map to the same points, and why is this allowed?

T9.3. Define $\|f\|_1$, $\|f\|_2$, and $\|f\|_\infty$ on continuous functions. Distinguish signal energy, RMS, and the units of each.

T9.4. Define equality a.e., an L^2 element, and L^2 convergence. Explain why an integral norm cannot fix an isolated value or control a pulse's peak.

T9.5. Derive Cauchy–Schwarz and Pythagoras from the integral inner product. Does orthogonality require disjoint supports?

T9.6. Derive the Gram equations, prove global optimality by residual orthogonality, and explain how normalization and independence affect the coefficients.

Exercises

9.1. Evaluate $\int_0^2 te^{-t^2}\,\mathrm{d}t$ by substitution. State the new limits and the factor linking the differentials.

9.2. Map $[0,1]^2$ by $x = u + v$, $y = 2v$. Find the image vertices and Jacobian; compute its area, $\iint_D x\,\mathrm{d}A$, and the mean of x.

9.3. Use polar coordinates to integrate $x^2 + y^2$ over $x^2 + y^2 \le 9$. Find its spatial average and compare their units.

9.4. For $f(t) = 1 - t$ on $[0,1]$, compute its mean, $\|f\|_1$, energy, $\|f\|_2$, RMS, and maximum norm. Which numerical agreements depend on the interval length?

9.5. Find the best constant approximation to $s(t) = t^2$ on $[-1,1]$ in L^2. Verify residual orthogonality and calculate the minimum squared error.

9.6. Find the best $a + bt$ approximation to $s(t) = t^2$ on $[0,1]$. Form the full Gram matrix, solve the normal equations, and calculate the squared error. Why is G not diagonal here?

9.7. (*) On $[0,1]$, let $h_n(t) = \sqrt{n}$ for $0 \le t \le 1/n$ and zero elsewhere. Find its mean and L^2 norm. It tends to zero at each $t > 0$; does it tend to zero in L^2? Compare with Example 9.4.

9.8. (*) Verify orthonormality of $e_n(t) = e^{int}/\sqrt{2\pi}$. For $s(t) = 2\cos t - \sin t$, find the nonzero normalized coefficients γ_n and check that their squared magnitudes sum to $\|s\|_2^2$.

From finite data to Fourier coordinates

On a tagged time grid, define $y_i = s(\xi_i)$, $A_{ij} = \phi_j(\xi_i)$, and $D = \operatorname{diag}(\Delta t_1, \ldots, \Delta t_n)$. Thus column j of A is the sampled version of the fixed function ϕ_j; the unknown coefficient vector is the same in both models.

Weighted sampled model	Function model on I
$(A\boldsymbol{c})_i = \sum_j c_j \phi_j(\xi_i)$	$p_{\boldsymbol{c}}(t) = \sum_j c_j \phi_j(t)$
$\sum_i (y_i - (A\boldsymbol{c})_i)^2 \Delta t_i$	$\int_I (s - p_{\boldsymbol{c}})^2$
$G_{\text{grid}} = A^{\mathsf{T}} D A$	$G_{ij} = \int_I \phi_i \phi_j$
$\boldsymbol{d}_{\text{grid}} = A^{\mathsf{T}} D \boldsymbol{y}$	$d_i = \int_I s \phi_i$
$A^{\mathsf{T}} D(\boldsymbol{y} - A\boldsymbol{c}) = 0$	$\langle s - p_{\boldsymbol{c}}, \phi_j \rangle = 0$ for every j

For continuous s, ϕ_j, refinement makes the matrix and data entries on the left tend to those on the right by Riemann integration. Equal widths give a common positive factor in the discrete objective, so its minimizer agrees with ordinary least squares. Unequal widths generally change the fit.

The best approximation is unique in the model space; independence determines whether its coefficients are unique. For an orthonormal basis, coefficient squares measure retained energy, and the remaining energy is exactly the squared approximation error.

Lecture 10 applies this finite geometry to successively larger Fourier and Haar spaces. Fourier coefficients use global oscillations; Haar coefficients use local averages and differences. A convergence theorem is needed to pass from finite optimal models to an infinite representation. Pointwise values, maximum error, and L^2 error remain distinct.

Integral transforms later replace a discrete index by frequency, decay, or scale. The same obligations persist: specify the domain and normalization, include the coordinate weight, and establish convergence of an improper integral.