Lectures in Mathematical Modeling

Lecture 1: Modeling Change

Kenneth Sok Kin Cheng

Revised 20 August 2026

Learning goals

By the end of the lesson, readers should be able to:

1. explain why a mathematical model is a purposeful simplification of reality;
2. recognise and test a **proportional relationship**;
3. translate "future = present + change" into a **difference equation**;
4. interpret parameters, initial conditions, and equilibrium values in context;
5. distinguish unconstrained growth from constrained growth;
6. evaluate a simple model through validation, sensitivity, and limitations.

1 What is a mathematical model?

A **mathematical model** is a mathematical representation built for a particular purpose. It may help us describe a system, explain a mechanism, forecast what may happen, or compare possible decisions. A model is not a complete copy of reality: it keeps the features that matter for the question and deliberately ignores others.

Launch: What would you model? **[10 min]**

A school notices that the queue at its canteen is sometimes short and sometimes extends outside the building.
In pairs, discuss:

1. What could be measured? List at least four variables.
2. What does "a bad queue" mean: long waiting time, long physical length, or too many late students?
3. Name two simplifying assumptions you might make.
4. What decision could the model support?

Modeling lesson: Before choosing mathematics, define the question and the output that matters.

The modeling cycle

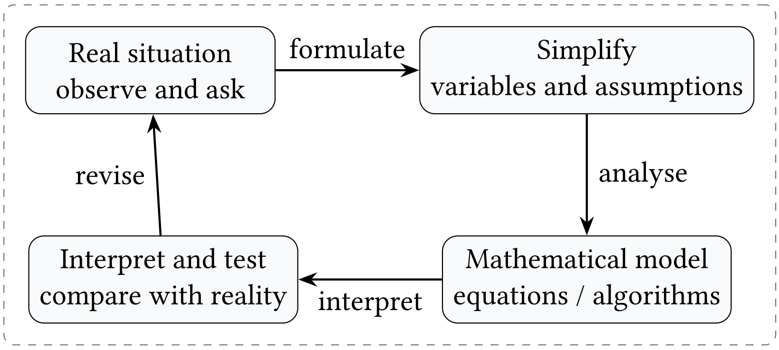


Figure 1: Modeling is iterative. A failed check is not the end; it tells us what to revise.

Four questions to ask throughout a modeling problem

1. **Purpose**: What decision or explanation is required?
2. **Structure**: Which variables interact, and how?
3. **Evidence**: What data or reasoning supports the relationship?
4. **Credibility**: How will we test the result and state limitations?

2 A first simplifying relationship: proportionality

Two variables y and x are **proportional** if

$$y = kx,$$

where k is a **constant of proportionality**. The graph is a straight line through the origin. The units of k are important: they explain how much y changes for each unit of x.

Worked Example 1: Spring–mass data

A spring is loaded with different masses. Let m be mass in grams and e be elongation in centimetres. The measured data are shown below.

m	50	100	150	200	250	300	350	400	450	500	550
e	1.000	1.875	2.750	3.250	4.375	4.875	5.675	6.500	7.250	8.000	8.750

A least-squares line forced through the origin has slope

$$k = \frac{\sum m_i e_i}{\sum m_i^2} \approx 0.01622 \text{ cm/g}.$$

Thus a simple model is

$$\boxed{e = 0.01622m.}$$

For a mass of 420 g, the model predicts $e \approx 6.81$ cm.

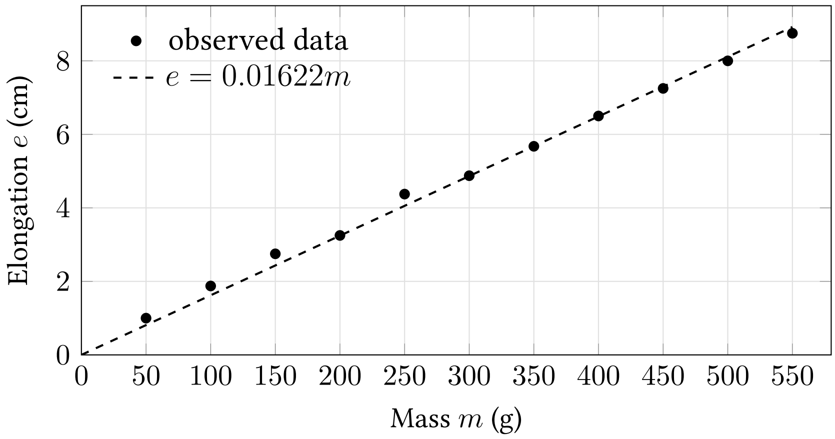


Figure 2: The points lie close to a straight line through the origin, so proportionality is plausible.

Model criticism: Is the model exact?

The points do not lie exactly on the line because measurements contain error and a real spring may not obey the same relationship under every load. We should not write "the spring follows the equation exactly." A defensible statement is: *Within the tested range, elongation is approximately proportional to mass.*

Checkpoint: Proportionality

A taxi charges a fixed flag-fall fee plus a charge per kilometre. Is total fare proportional to distance? Explain in one sentence.
Hint: What should the fare be when distance is zero?

3 Modeling discrete change

When a quantity is recorded at regular steps (days, months, generations), a useful starting point is

$$\boxed{\text{next value} = \text{current value} + \text{change during one step.}}$$

For a sequence $a_0, a_1, a_2, \ldots$, the **first difference** is

$$\Delta a_n = a_{n+1} - a_n.$$

A **difference equation** states how the next value depends on current values. Together with an **initial condition**, it defines a discrete **dynamical system**.

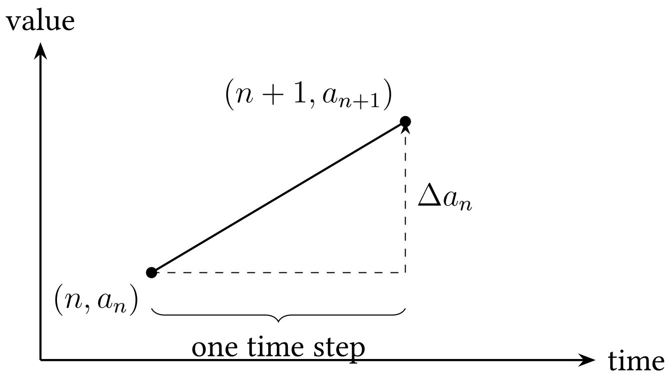


Figure 3: A first difference is the vertical change between consecutive observations.

Worked Example 2: Monthly compound interest

A certificate begins with $a_0 = 1000$ dollars and earns 1% per month. The monthly change equals 1% of the current balance:

$$\Delta a_n = 0.01 a_n.$$

Therefore

$$a_{n+1} = a_n + 0.01a_n = 1.01a_n, \qquad a_0 = 1000.$$

The first values are

$$1000,\ 1010,\ 1020.10,\ 1030.30, \ldots$$

Repeated substitution suggests the closed form

$$\boxed{a_n = 1000(1.01)^n.}$$

The recurrence explains the step-by-step mechanism; the closed form answers long-range questions quickly.

Guided practice: Drug elimination **[8 min]**

A medicine loses 20% of the amount present each hour. Initially there are 640 mg.

1. Define a_n clearly, including its units.
2. Write the change equation $\Delta a_n =$ ________________.
3. Write the recurrence and calculate a_1, a_2, a_3.
4. Without calculating every term, predict the long-term behavior.

Interpreting a multiplier

For $a_{n+1} = ra_n$:

Multiplier	Behavior
$r > 1$	positive growth
$0 < r < 1$	decay toward 0
$-1 < r < 0$	alternating signs with shrinking magnitude
$\lvert r\rvert > 1$	magnitude grows; signs may alternate

The meaning of negative values must be checked against the context. A recurrence can be mathematically valid but physically meaningless after it predicts a negative population or balance.

4 Constant input or removal: $a_{n+1} = ra_n + b$

Many systems combine proportional change with a fixed input or output:

$$a_{n+1} = ra_n + b.$$

Examples include a regular deposit, a fixed monthly payment, medication taken at fixed intervals, and a stock replenished by a constant amount.

An **equilibrium value** a^* is a value that remains unchanged:

$$a^* = ra^* + b.$$

When $r \neq 1$,

$$\boxed{a^* = \frac{b}{1-r}.}$$

Subtracting the equilibrium from the recurrence gives a particularly useful form:

$$a_{n+1} - a^* = r(a_n - a^*).$$

Thus the distance from equilibrium is multiplied by r at every step.

Worked Example 3: Repeated medication

At the end of each day, 50% of a drug remains. Immediately afterwards, a patient takes another 0.1 mg. Let a_n be the amount just after the nth dose. Then

$$a_{n+1} = 0.5a_n + 0.1.$$

The equilibrium is

$$a^* = \frac{0.1}{1-0.5} = 0.2 \text{ mg}.$$

Because $|0.5| < 1$, the distance from 0.2 is halved each day. The equilibrium is **stable**: nearby solutions move toward it.

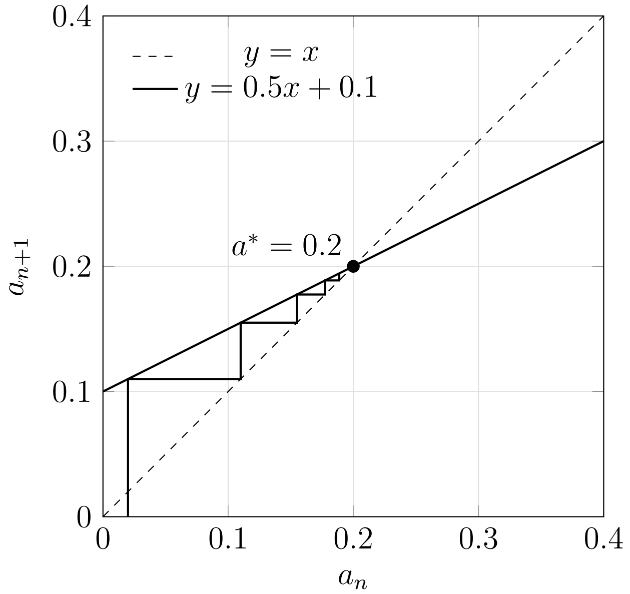


Figure 4: A cobweb plot: the iterations approach the intersection of $y = x$ and $y = 0.5x + 0.1$.

Worked Example 4: Withdrawal from an investment

An account earns 1% per month and pays out 1000 dollars at the end of each month:

$$a_{n+1} = 1.01a_n - 1000.$$

Its equilibrium is

$$a^* = \frac{-1000}{1-1.01} = 100000.$$

Because $1.01 > 1$, the equilibrium is **unstable**. Starting exactly at 100000 keeps the balance constant; starting above it causes growth, while starting below it causes the balance to move downward until the model eventually predicts a negative value. At that point the recurrence must stop, because an account cannot continue making withdrawals after it is empty.

Checkpoint: Equilibrium and stability

For each system, find the equilibrium and predict whether it is stable.

1. $x_{n+1} = 0.8x_n + 12$
2. $y_{n+1} = 1.1y_n - 5$

Explain the conclusion using the multiplier, not only a numerical table.

5 When growth slows: a logistic difference equation

Exponential growth assumes that the same percentage growth can continue indefinitely. In a limited environment, resources become scarce and growth usually slows. A simple constrained model is

$$p_{n+1} = p_n + k(M - p_n)p_n,$$

where M is the **carrying capacity** and k controls the speed of growth.

The change

$$\Delta p_n = k(M - p_n)p_n$$

contains two factors:

- p_n: more organisms can produce more new organisms;
- $M - p_n$: the remaining capacity becomes small near the limit.

Worked Example 5: Yeast culture

Suppose

$$p_{n+1} = p_n + 0.00082(665 - p_n)p_n, \qquad p_0 = 9.6.$$

The first values are approximately

$$9.6,\ 14.8,\ 22.6,\ 34.5,\ 52.4,\ 78.7,\ 116.6, \ldots$$

Growth first accelerates, then slows as the population approaches 665. The S-shaped pattern is characteristic of **logistic growth**.

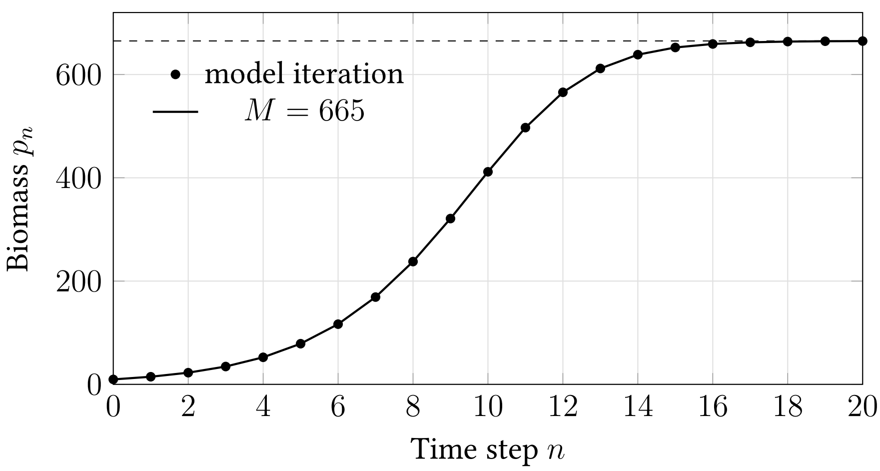


Figure 5: The population approaches the carrying capacity rather than growing without bound.

Model criticism: Sensitivity and interpretation

Changing k mainly changes *how quickly* the model approaches equilibrium. Changing M changes the predicted long-run level itself. Therefore, a conclusion about the final population is much more sensitive to an inaccurate estimate of M than to a moderate error in k.
A good competition report should state this in context rather than merely list percentages.

Model comparison **[8 min]**

Two teams model the same yeast population.

$$\text{Team A: } p_{n+1} = 1.45 p_n,$$
$$\text{Team B: } p_{n+1} = p_n + 0.00082(665 - p_n)p_n.$$

Which model is more appropriate for short-term growth? Which is more credible over a long period? Give one reason for each answer.

6 Mini-case: a discrete SIR epidemic model

This final case is a preview of a **system of difference equations**. It is included to demonstrate how the ideas above combine in a larger modeling problem; it is not intended for full derivation within this first lesson.

A dormitory contains $N = 1000$ students. At day n:

- S_n are susceptible;
- I_n are currently infected;
- R_n are removed from transmission (recovered or isolated).

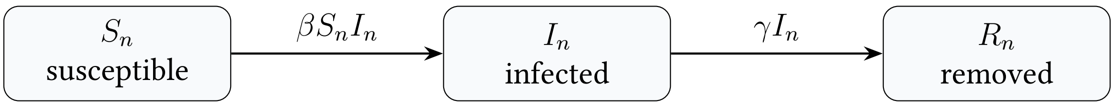


Figure 6: Compartment flow in the SIR model.

Under the simplifying assumptions of a closed population, homogeneous mixing, no reinfection, and constant rates,

$$\begin{aligned} S_{n+1} &= S_n - \beta S_n I_n, \\ I_{n+1} &= I_n + \beta S_n I_n - \gamma I_n, \\ R_{n+1} &= R_n + \gamma I_n. \end{aligned}$$

Here β is the transmission parameter and γ is the daily recovery proportion. For $\beta = 0.0005$, $\gamma = 0.1$, and $(S_0, I_0, R_0) = (999, 1, 0)$, the early reproduction indicator is

$$\mathcal{R}_0 \approx \frac{\beta S_0}{\gamma} = 4.995 > 1,$$

so infections initially increase.

Day	S_n	I_n	R_n
0	999.0	1.0	0.0
5	993.6	5.4	1.1
10	965.0	28.1	6.9
15	831.8	132.1	36.1
20	448.9	399.7	151.3
24	157.7	507.0	335.2
30	33.3	357.8	608.9

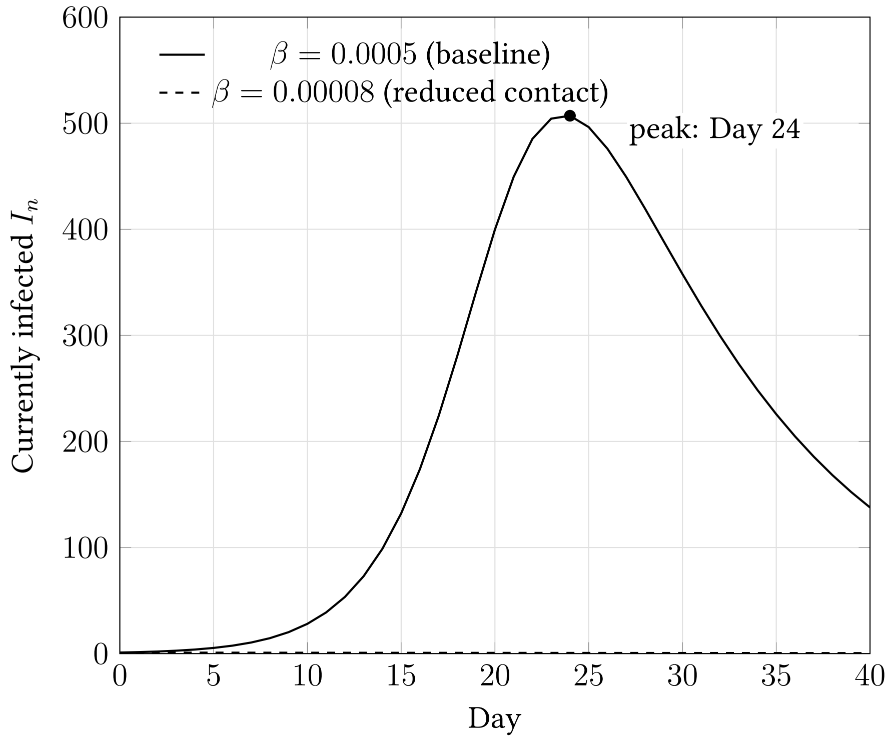


Figure 7: Reducing the transmission parameter below the threshold prevents initial growth.

Competition-style communication **[10 min]**

Using the table and graph, write a three-sentence conclusion:

1. one sentence reporting the baseline peak;
2. one sentence describing the intervention result;
3. one sentence stating a limitation of the model.

Avoid saying that the model "proves" what will happen.

7 Lesson summary

Six ideas to retain

1. A model is built for a purpose and is always a simplification.
2. Parameter units and interpretations matter as much as algebra.

3. For discrete time, start with next = current + change.
4. In $a_{n+1} = ra_n + b$, the multiplier r determines stability around equilibrium.
5. Nonlinear feedback can slow growth and create a carrying capacity.
6. A credible model includes testing, sensitivity analysis, and limitations.

Checkpoint: Exit ticket **[2 min]**

Complete the sentence:
"A mathematical model is useful when ______________________________, but it may fail when ______________________________."

8 Practice problems

Core practice

1. A water tank contains 800 L and loses 6% of its current volume each hour.
   (a) Write a recurrence for the volume V_n.
   (b) Find V_5 and the long-term behavior.
   (c) State one reason the model might eventually become unrealistic.
2. A club has 120 members. Each month, 8% of current members leave and 15 new members join.
   (a) Formulate $M_{n+1} = rM_n + b$.
   (b) Find the equilibrium membership.
   (c) Decide whether the equilibrium is stable and explain why.
3. A town's recycling programme begins with 200 households. Participation grows according to
   $$H_{n+1} = H_n + 0.002(1000 - H_n)H_n.$$
   Explain the meaning of 1000 and why the model slows near that value.
4. A sequence satisfies $x_{n+1} = 1.15x_n - 30$, $x_0 = 250$.
   (a) Find the equilibrium.
   (b) Calculate the first four terms.
   (c) Does the context need a stopping rule if x_n represents money? Explain.
5. The following data record the number of downloads after successive advertising rounds:
   $$80,\ 116,\ 169,\ 244,\ 351.$$
   Test whether a constant multiplier is a reasonable approximation. Estimate the multiplier and predict the next value. Comment on the uncertainty of your prediction.
6. In an epidemic model, β is estimated from only three days of data. Explain why sensitivity analysis with respect to β is important. Your answer should refer to decisions, not only mathematics.

Practice check and hints

For Questions 1–5, useful checks are:

1. $V_{n+1} = 0.94V_n$, $V_5 \approx 587.12$ L, and $V_n \to 0$ unless a physical floor or different loss mechanism is imposed.
2. $M_{n+1} = 0.92M_n + 15$ has stable equilibrium $M^* = 187.5$.
3. The value 1000 is the saturation level represented by the model.
4. $x^* = 200$ and $(x_1, x_2, x_3, x_4) = (257.5, 266.125, 276.044, 287.450)$. A money model still needs domain and stopping rules, even though this initial value moves upward.
5. Successive download ratios are close to 1.45; a defensible next estimate is about 508, conditional on that pattern continuing.

For Question 6, a complete response links plausible changes in β to forecast trajectories, intervention timing or allocation, and a stated sensitivity range.

Modeling challenge

Choose one of the following and prepare a one-page modeling brief containing: purpose, variables, assumptions, recurrence(s), expected behavior, validation plan, and one limitation.

1. Number of bicycles available at two sharing stations each morning.
2. Daily amount of food waste in a school canteen after a reduction campaign.
3. Growth and saturation of users joining a new school app.

You do not need a perfect numerical answer. The quality of the formulation and justification is the main criterion.

A Optional derivation: closed form for $a_{n+1} = ra_n + b$

This appendix is not part of the two-hour core lesson. It may be assigned to students who are comfortable with algebraic derivations.

Let $a^* = b/(1-r)$ for $r \neq 1$. Define $u_n = a_n - a^*$. Then

$$\begin{aligned} u_{n+1} &= a_{n+1} - a^* \\ &= ra_n + b - a^* \\ &= r(a_n - a^*) \\ &= ru_n. \end{aligned}$$

Therefore $u_n = r^n u_0$, so

$$\boxed{a_n = a^* + r^n(a_0 - a^*)} = \boxed{\frac{b}{1-r} + r^n\left(a_0 - \frac{b}{1-r}\right)}.$$

This form makes stability transparent: when $|r| < 1$, the term $r^n(a_0 - a^*)$ tends to zero.

Model audit card

Audit before trusting or communicating a model

Purpose	What question or decision does the model support?	**Data**	Which inputs are observed, derived, or simulated?
Assumptions	Which simplifications may change the answer?	**Units**	Are dimensions and scales consistent?
Verification	Were equations, code, constraints, and limiting cases checked?	**Validation**	Is performance tested against independent evidence?
Uncertainty	How do measurement, parameter, solver, and structural errors affect conclusions?	**Ethics**	Who benefits, who bears risk, and what fairness criterion matters?
Reproducibility	Are data provenance, versions, seeds, and transformations recorded?	**Communication**	Are claims conditional, traceable, and decision-relevant?

Further reading

- GAIMME: Guidelines for Assessment and Instruction in Mathematical Modeling Education (SIAM / COMAP)
- 18.03SC Differential Equations (MIT OpenCourseWare)

Sources, provenance, and licence

These notes follow the iterative modeling and assessment perspective in the SIAM/COMAP GAIMME report. Competition-specific remarks are a snapshot checked on 20 August 2026 against the official COMAP HiMCM instructions; readers should verify current rules before submission. Unless a source is explicitly identified, classroom datasets and numerical scenarios are synthetic or simulated teaching examples and are not empirical claims about a real institution. Third-party references retain their own terms.

Student response guide and checking criteria

For closed calculations, show the definition, substitution, units, and at least one independent check; the worked examples in this unit are method hints, but you must complete the arithmetic yourself. Open modeling responses are checked on five dimensions: a clear purpose and output; stated data and provenance; defensible assumptions and scope; traceable computational verification and model validation; and alignment among uncertainty, limitations, and the recommendation. Support each dimension with concrete evidence rather than writing only “completed.”

Lectures in Mathematical Modeling

Lecture 2: The Modeling Process and Scaling

Kenneth Sok Kin Cheng

Revised 20 August 2026

Learning goals

By the end of the lesson, readers should be able to:

1. turn a vague real-world question into a precise modeling objective;
2. identify variables, units, assumptions, and useful **submodels**;
3. distinguish a proportional relationship from a general linear relationship;
4. test relationships of the form $y \propto f(x)$ by transforming variables;
5. apply **geometric similarity** to derive area, volume, and weight scaling laws;
6. evaluate a model through interpretation, sensitivity, and limitations.

1 From a vague question to a modeling objective

A real-world question rarely arrives in a form that can be solved immediately. The modeler must decide what is being predicted, for whom the answer is useful, and under what conditions the conclusion is intended to hold. This first act of narrowing the question is called **problem formulation**.

Launch: Is the scale model enough? **[10 min]**

A school wants to build a shaded outdoor stage. Before construction, students make a 1:20 scale model and test it under a lamp.
In pairs, discuss:

1. Which measurements from the model can be scaled directly to the full stage?
2. Which effects might not scale in the same way?
3. What decision could the model support?
4. What evidence would make you trust the model?

Modeling lesson: A model can preserve some relationships while distorting others. We must know which quantities scale and why.

A practical modeling workflow

For most competition and classroom problems, the following six questions are more useful than immediately searching for a formula.

1. **Purpose**: What decision, explanation, or prediction is required?
2. **Variables and units:** What must be measured, and in what units?
3. **Assumptions:** What is held constant, neglected, or idealised?

4. **Structure:** Can the problem be separated into smaller submodels?
5. **Evidence:** How will parameters be estimated and results checked?
6. **Communication:** What conclusion is justified, and where might it fail?

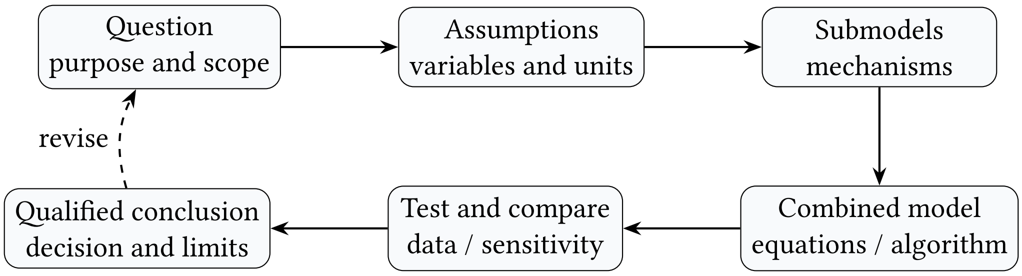


Figure 1: A model is developed through connected decisions, not by choosing a formula in isolation.

Checkpoint: A modelable question

Rewrite the vague question "How safe is this road?" as a more precise modeling question. Your version must identify an output quantity and at least one condition under which the answer applies.

2 Case study: vehicular stopping distance

Consider the question: *How much distance does a car need to stop safely?* A useful first model separates the process into two mechanisms:

$$\text{total stopping distance} = \text{reaction distance} + \text{braking distance}.$$

This is an example of **decomposition**: a complicated system is represented by simpler submodels whose outputs can be combined.

Worked Example 1: Building the stopping-distance model

Let v be the initial speed in metres per second, t_r the driver's reaction time in seconds, and a the magnitude of the braking deceleration in metres per second squared.
Reaction submodel. During the reaction interval, the car is assumed to travel at approximately constant speed:

$$d_r = t_r v.$$

Braking submodel. Once the brakes act, assume constant deceleration. From $0 = v^2 - 2ad_b$,

$$d_b = \frac{v^2}{2a}.$$

Combined model.

$$\boxed{d(v) = t_r v + \frac{v^2}{2a}.}$$

For a baseline driver with $t_r = 1.2$ s and a dry-road deceleration of $a = 6.5$ m/s^2,

$$\boxed{d(v) = 1.2v + 0.0769v^2.}$$

At 50 km/h ($v = 13.89$ m/s), the predicted stopping distance is about 31.5 m. At 100 km/h ($v = 27.78$ m/s), it is about 92.7 m. Doubling speed therefore requires almost three times the distance, not merely twice the distance.

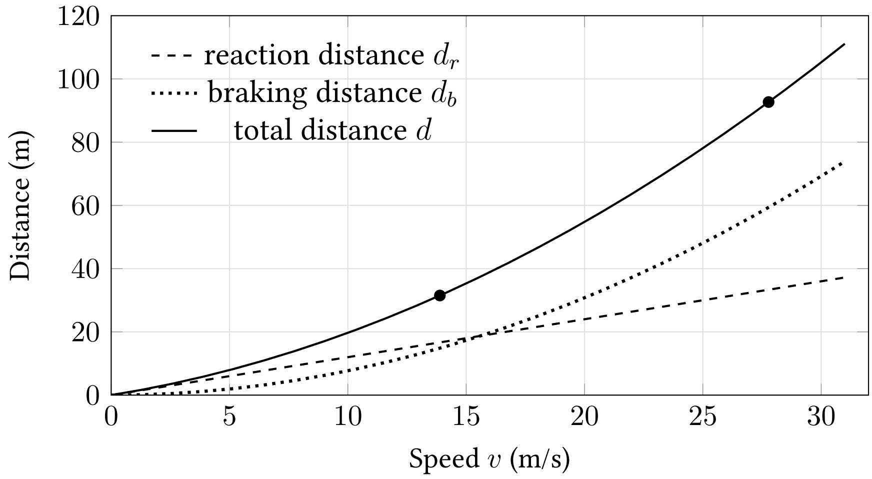


Figure 2: The quadratic braking term becomes increasingly important as speed rises.

Checkpoint: When does braking dominate?

Find the positive speed at which the braking distance first equals the reaction distance. Express the answer in both m/s and km/h. What does this threshold mean for high-speed driving?

A model is not complete when an equation has been obtained. We must ask whether its conclusions are stable under plausible changes in the assumptions.

Worked Example 2: Sensitivity at 80 km/h

At 80 km/h, $v = 22.22$ m/s. Compare four scenarios:

Scenario	t_r (s)	a (m/s^2)	Predicted distance (m)
Baseline: alert driver, dry road	1.2	6.5	64.7
Slower reaction	1.8	6.5	78.0
Wet road	1.2	3.5	97.2
Slower reaction and wet road	1.8	3.5	110.5

For this speed, the conclusion is more sensitive to the available deceleration than to a moderate change in reaction time. This does not mean reaction time is unimportant; it means road and tyre conditions dominate this particular comparison.

Model criticism: What the model leaves out

The model assumes level ground, constant deceleration, immediate full braking after the reaction interval, and no change in direction. It also treats t_r and a as known constants. In reality, they vary with fatigue, distraction, weather, tyre condition, road slope, and braking technology.
A defensible conclusion is therefore conditional: *under the stated reaction time and deceleration assumptions, the required stopping distance increases nonlinearly with speed.*

Refine the model **[8 min]**

Choose one omitted factor—road slope, rain, tyre wear, or driver fatigue.

1. Which parameter or submodel would it affect?

2. Would you represent the factor by a new variable or by changing an existing parameter?

3. What data would be needed to estimate the change?

3 Generalised proportionality and transformed variables

A proportional model need not have the form $y = kx$. More generally,

$$y \propto f(x) \qquad \Longleftrightarrow \qquad y = kf(x).$$

If the relationship is plausible, plotting y against the transformed variable $f(x)$ should produce an approximately straight line through the origin.

Linear is not always proportional

A relationship $y = mx + c$ is **linear** in the school-algebra sense, but it is proportional only when $c = 0$. A taxi fare with a flag-fall charge is linear in distance but not proportional to distance. Likewise, $y = kx^2$ is not a straight line when plotted against x, yet it is proportional to x^2 and becomes a straight line when plotted against the transformed variable x^2.

Worked Example 3: Kepler's third law

Let R be a planet's mean orbital radius measured in astronomical units (AU), and let T be its orbital period in Earth years. Kepler's third law suggests

$$T \propto R^{3/2}.$$

Using Earth, $R = 1$ and $T = 1$, so the proportionality constant is approximately 1:

$$\boxed{T \approx R^{3/2}.}$$

For Jupiter, $R = 5.203$ AU, giving

$$T \approx 5.203^{3/2} \approx 11.87 \text{ years},$$

which is close to the observed 11.86 years.

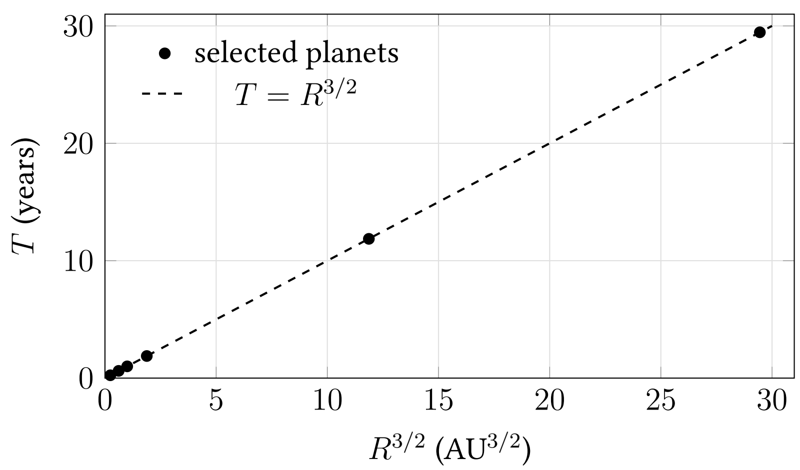


Figure 3: After transforming the horizontal variable, Kepler's relationship is approximately proportional.

Checkpoint: Choose the useful transformation

For each proposed relationship, state what should be plotted on the horizontal axis to test proportionality.

1. $F \propto v^2$
2. $P \propto 1/V$
3. $N \propto e^{rt}$, where r is known

4 Geometric similarity and scaling laws

Two objects are **geometrically similar** if they have the same shape and every corresponding length is multiplied by the same **scale factor** λ.

If all lengths are multiplied by λ, then every area is multiplied by λ^2 and every volume is multiplied by λ^3:

$$\boxed{L \propto \lambda, \qquad S \propto \lambda^2, \qquad V \propto \lambda^3.}$$

For objects of the same material and density, mass and weight also scale as volume:

$$W \propto V \propto L^3.$$

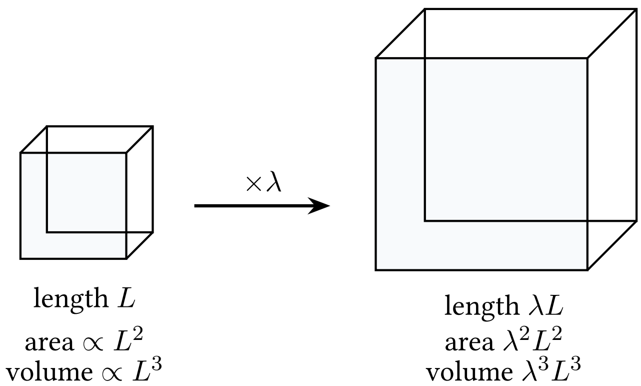


Figure 4: Length, area, and volume do not scale at the same rate.

Scaling table **[6 min]**

Complete the table without using a formula sheet.

Scale factor λ	Length factor	Area factor	Volume factor
2			
3			
$\frac{1}{2}$			

Then explain why a twice-as-tall storage tank does not hold merely twice as much water.

Worked Example 4: A submarine scale model

A 7-m model and a 70-m submarine are geometrically similar, so the full-size vessel has scale factor $\lambda = 10$.

- Corresponding lengths are multiplied by 10.

- Surface areas are multiplied by $10^2 = 100$.
- Volumes and masses are multiplied by $10^3 = 1000$, assuming equal density.

If heat loss is mainly proportional to surface area, the full-size vessel loses about 100 times as much heat under comparable temperature conditions—not 1000 times as much. If storage capacity is proportional to volume, it is about 1000 times as large.

Model criticism: Similarity is an assumption, not an observation

The area and volume laws are exact for mathematically similar shapes. Their application to real objects is only as credible as the similarity assumption. A juvenile animal may not have the same proportions as an adult; a large bridge model may require thicker supports; and a raindrop changes shape as it grows.

5 Model refinement through a scaling model

A useful model is often developed in stages. We begin with the simplest plausible relationship, identify what it misses, and then refine it.

Worked Example 5: Estimating the weight of a fish

Suppose the weight W of a bass must be estimated without a scale.
Model I: length only. If all fish are geometrically similar and have approximately equal density, then

$$W \propto L^3.$$

This model is convenient, but two fish with the same length may have very different body thicknesses.
Model II: length and girth. Approximate the fish as a sequence of similar cross-sections. Cross-sectional area scales with the square of girth G, so

$$\boxed{W \propto LG^2.}$$

Suppose a fish with $L = 40$ cm, $G = 25$ cm weighs 1.25 kg. Then

$$1.25 = k(40)(25)^2 \quad \Longrightarrow \quad k = 0.00005 \text{ kg/cm}^3.$$

A fish with $L = 50$ cm and $G = 30$ cm is predicted to weigh

$$W = 0.00005(50)(30)^2 = 2.25 \text{ kg}.$$

The second model uses one additional measurement to represent body condition more directly.

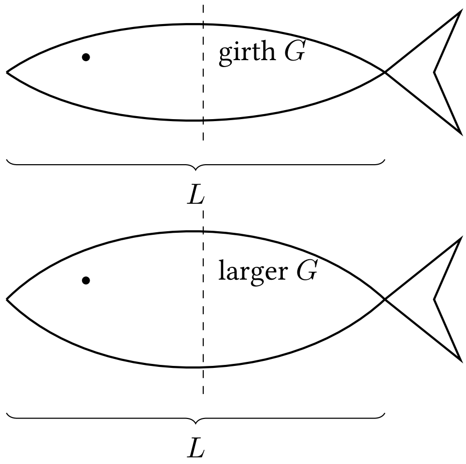


Figure 5: Length alone cannot distinguish fish with different body thicknesses.

Model criticism: Does the refined model solve everything?

The model $W = kLG^2$ still assumes comparable species, density, and body shape. Girth measurements may be inconsistent, and a single calibration fish gives no estimate of prediction error. Refinement improves relevance, but it also introduces an additional measurement and another source of uncertainty.

Checkpoint: Compare two models

For a fish whose length increases by 10% while girth remains unchanged:

1. By what percentage does Model I, $W \propto L^3$, increase its prediction?
2. By what percentage does Model II, $W \propto LG^2$, increase its prediction?
3. Why do the models disagree?

6 A final scaling consequence: strength versus weight

Scaling laws can reveal qualitative limits even when no numerical constant is known. Suppose geometrically similar animals are made of comparable material.

- Muscle strength is approximately proportional to muscle cross-sectional area: strength $\propto L^2$.
- Body weight is proportional to volume: $W \propto L^3$.

Therefore

$$\frac{\text{strength}}{\text{weight}} \propto \frac{L^2}{L^3} = \frac{1}{L} \propto W^{-1/3}.$$

Can a giant ant keep its relative strength? [10 min]

An ant is enlarged so that every length becomes 10 times as large, while shape and material remain unchanged.

1. By what factor does muscle cross-sectional area increase?
2. By what factor does body weight increase?
3. By what factor does strength per unit weight change?
4. State one reason why the geometric-similarity assumption may fail for a real giant animal.

What scaling arguments can and cannot do

A scaling argument often predicts an exponent or qualitative trend without determining the proportionality constant. It is especially useful for comparing sizes and identifying which effects grow faster. It does not replace calibration, validation, or knowledge of the underlying mechanism.

Optional extension: Terminal velocity of similar objects

Suppose geometrically similar falling objects experience quadratic drag. At terminal velocity, weight and drag balance:

$$mg \propto Av_t^2.$$

For constant density, $m \propto L^3$ and frontal area $A \propto L^2$, so

$$L^3 \propto L^2 v_t^2 \quad \Longrightarrow \quad v_t \propto L^{1/2} \propto m^{1/6}.$$

The small exponent means terminal velocity changes relatively slowly with mass. This conclusion depends on quadratic drag, geometric similarity, and comparable density; it is not a universal law for every falling object.

7 Lesson summary

Six ideas to retain

1. A vague question must be narrowed to a purpose, output, and set of conditions.
2. Complicated systems can often be built from interpretable submodels.
3. A proportional relationship $y = kf(x)$ becomes linear through the origin when y is plotted against $f(x)$.
4. In geometrically similar objects, length, area, and volume scale as λ, λ^2, and λ^3.
5. Model refinement should address a specific weakness, not merely add complexity.
6. Every conclusion should be accompanied by assumptions, sensitivity, and a statement of where the model may fail.

Checkpoint: Exit ticket **[2 min]**

Complete both statements:

1. "The most important assumption in today's stopping-distance model is ________________________ because ________________________."
2. "When every length is doubled, ______________ grows faster than ______________ because ____________________."

8 Practice problems

Core practice

1. A driver has reaction time 1.4 s and can decelerate at 5.0 m/s^2.
   (a) Write a stopping-distance model in terms of speed v in m/s.
   (b) Estimate the distance required at 72 km/h.
   (c) Identify one parameter that is uncertain and explain how it could be measured.
2. A delivery fee is 25 dollars plus 4 dollars per kilometre.
   (a) Is fee proportional to distance? Explain.
   (b) Write the linear model and interpret both constants.
3. For each relationship, identify a transformed horizontal variable that would produce a line through the origin if the model were correct:
$$y \propto x^3, \qquad y \propto \sqrt{x}, \qquad y \propto \frac{1}{x^2}.$$
4. Planet X has a mean orbital radius of 4 AU. Use $T \approx R^{3/2}$ to estimate its orbital period. If the observed period is 8.5 years, calculate the absolute relative error using the observed period as the denominator, and suggest one possible reason for the difference.
5. A cube-shaped tank is redesigned so that each side is 30% longer.
   (a) By what factor does its surface area change?
   (b) By what factor does its capacity change?
   (c) Which quantity has the larger percentage increase?
6. A $1{:}25$ model of a monument uses 0.80 kg of material. Estimate the material required for a geometrically similar full-size monument of equal density. State why the estimate may be unrealistic for an actual structure.
7. A fish model has the form $W = kLG^2$. A measured fish has $L = 45$ cm, $G = 28$ cm, and $W = 1.80$ kg.
   (a) Estimate k.
   (b) Predict the weight of a fish with $L = 52$ cm and $G = 31$ cm.
   (c) State two sources of prediction uncertainty.
8. Two geometrically similar animals have weights in the ratio $8{:}1$.
   (a) Find the ratio of their characteristic lengths.
   (b) Find the ratio of their surface areas.
   (c) Compare their surface-area-to-volume ratios.

Practice check and hints

1. $d(v) = 1.4v + v^2/10$; at 72 km/h, $v = 20$ m/s and $d = 68$ m.
2. The fee is $25 + 4d$ dollars and is not proportional because of the nonzero fixed charge.

3. Plot against x^3, $\sqrt{x}$, and $1/x^2$, respectively.
4. The model gives 8 years; relative to 8.5 years, the absolute relative error is about 5.9%.
5. Surface area changes by $1.3^2 = 1.69$ and capacity by $1.3^3 = 2.197$.
6. Volume scaling gives $0.80(25^3) = 12500$ kg; structural design and material thickness need not scale geometrically.
7. $k \approx 5.102 \times 10^{-5}$ kg/cm^3 and the prediction is about 2.55 kg; measurement error and body-shape variation are relevant uncertainties.
8. The length and area ratios are $2{:}1$ and $4{:}1$; the smaller animal has twice the surface-area-to-volume ratio.

Modeling challenge

Choose one question and prepare a one-page modeling brief.

1. How should the dimensions of a school water tank change if student population increases by 50%?
2. Can a small wind-tunnel model predict the aerodynamic force on a full-size vehicle?
3. How can the weight of a fruit be estimated from measurements that do not require a scale?

The emphasis is on the logic of the model. A fully calibrated numerical answer is not required.

One-page modeling brief planner

Purpose and output	What decision or prediction will the model support? What quantity will be reported?
Variables and units	Define the measurable quantities and distinguish inputs from outputs.
Assumptions	State what is treated as constant, neglected, or geometrically similar.
Submodels / scaling	Give at least two relationships and explain why each is plausible.
Calibration	Identify the data needed to estimate any unknown constant.
Validation	Explain what observation or comparison would test the model.
Sensitivity	Name one assumption or parameter that should be varied.
Limitation	State where the conclusion should not be used.

Checkpoint: Submission check

Before submitting, underline the sentence in your brief that gives the main conclusion. Circle every numerical value that has a unit. Put a star beside the assumption that most strongly affects the result.

Model audit card

Audit before trusting or communicating a model

Purpose	What question or decision does the model support?	**Data**	Which inputs are observed, derived, or simulated?
Assumptions	Which simplifications may change the answer?	**Units**	Are dimensions and scales consistent?
Verification	Were equations, code, constraints, and limiting cases checked?	**Validation**	Is performance tested against independent evidence?
Uncertainty	How do measurement, parameter, solver, and structural errors affect conclusions?	**Ethics**	Who benefits, who bears risk, and what fairness criterion matters?
Reproducibility	Are data provenance, versions, seeds, and transformations recorded?	**Communication**	Are claims conditional, traceable, and decision-relevant?

Further reading

- GAIMME: Guidelines for Assessment and Instruction in Mathematical Modeling Education (SIAM / COMAP)
- Guide to the SI: Values, quantities, and dimensions (NIST)
- Planetary fact sheet —ratio to Earth values (NASA NSSDCA)

Sources, provenance, and licence

These notes follow the iterative modeling and assessment perspective in the SIAM/COMAP GAIMME report. Competition-specific remarks are a snapshot checked on 20 August 2026 against the official COMAP HiMCM instructions; readers should verify current rules before submission. Unless a source is explicitly identified, classroom datasets and numerical scenarios are synthetic or simulated teaching examples and are not empirical claims about a real institution. Third-party references retain their own terms.

Student response guide and checking criteria

For closed calculations, show the definition, substitution, units, and at least one independent check; the worked examples in this unit are method hints, but you must complete the arithmetic yourself. Open modeling responses are checked on five dimensions: a clear purpose and output; stated data and provenance; defensible assumptions and scope; traceable computational verification and model validation; and alignment among uncertainty, limitations, and the recommendation. Support each dimension with concrete evidence rather than writing only “completed.”

Lectures in Mathematical Modeling

Lecture 3: Model Fitting and Residual Analysis

Kenneth Sok Kin Cheng

Revised 20 August 2026

Learning goals

By the end of the lesson, readers should be able to:

1. distinguish **model fitting** from **interpolation**;
2. calculate and interpret a **residual**;
3. fit a proportional model and a straight-line model by the **least-squares method**;
4. linearise simple exponential or power relationships using transformed variables;
5. compare fitted models using graphs, residual patterns, error measures, and withheld data;
6. communicate why a numerically good fit may still be an unsuitable model.

1 From a proposed relationship to a fitted model

In Lecture 2, physical reasoning suggested relationships such as

$$d_b \propto v^2$$

for braking distance and speed. Such a statement describes the *form* of a model, but it does not yet determine the proportionality constant. Real observations are also unlikely to lie exactly on a mathematical curve. We therefore need a systematic way to estimate parameters and judge whether the proposed relationship is credible.

Launch: Three lines, three stories **[10 min]**

A set of six data points is shown with three possible fitted lines drawn by different teams.
In pairs, discuss:

1. What could "best fit" mean?
2. Should a positive error cancel a negative error?
3. Should one very inaccurate point matter more than several small inaccuracies?
4. What information would you need before choosing among the lines?

Modeling lesson: A fitted model is never selected by appearance or a single number alone. The criterion must match the purpose of the model.

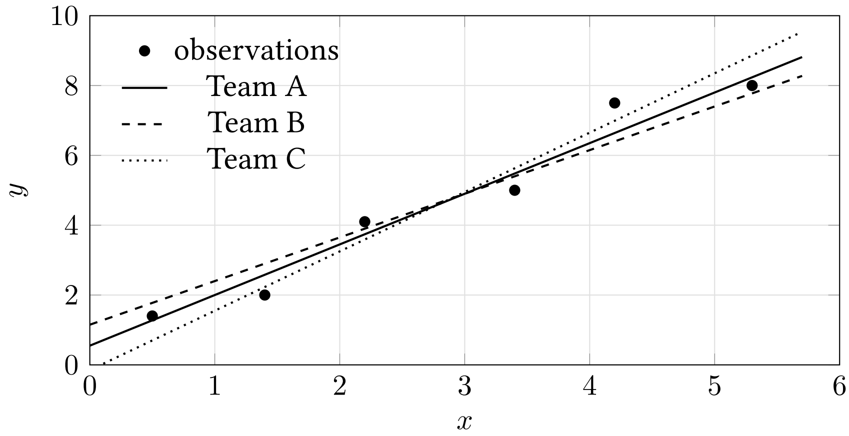


Figure 1: Different fitting criteria can prefer different models. The purpose of fitting must be stated.

Fitting, interpolation, and prediction

- In **model fitting**, a model family is proposed from theory or previous reasoning, and its parameters are estimated from data. The curve need not pass through every point.
- In **interpolation**, a curve is constructed to pass through specified data points, usually to estimate values between them. This is the main subject of Lecture 4.
- A fitted model may be used for **prediction**, but prediction outside the observed range is **extrapolation** and requires additional caution.

2 Residuals and sources of discrepancy

Suppose the observed value at x_i is y_i and the model predicts $\hat{y}_i = f(x_i)$. The **residual** is

$$\boxed{r_i = y_i - \hat{y}_i.}$$

A positive residual means the model underpredicts the observation; a negative residual means the model overpredicts it.

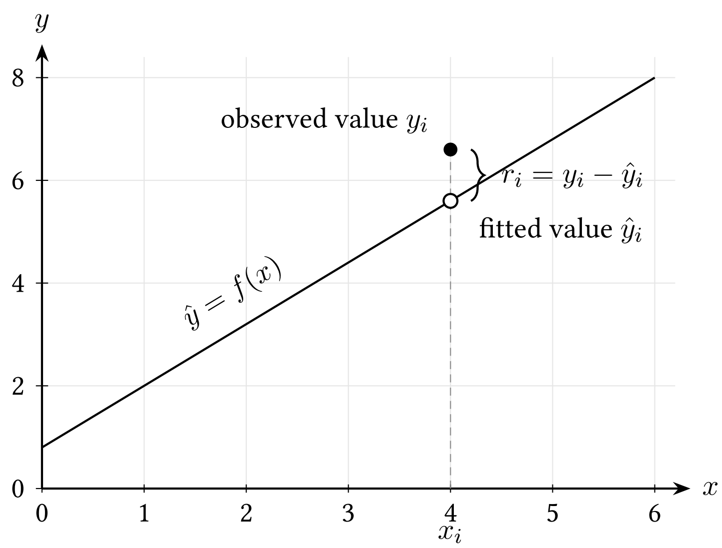


Figure 2: At x_i, the residual is the signed vertical difference between the observed value y_i and the fitted value $\hat{y}_i = f(x_i)$.

A discrepancy is not always a fitting failure

Differences between data and model can arise from several sources:

1. **measurement error**: limited precision, recording error, or natural variation;
2. **formulation error**: omitted variables or an unsuitable functional form;
3. **parameter uncertainty**: the available data do not determine the parameters exactly;
4. **numerical error**: rounding or approximation during computation.

A very precise fitting algorithm cannot repair a poor model structure or unreliable data.

Checkpoint: Signs and units

A model predicts a stopping distance of 42.3 m, while the observed distance is 45.0 m.

1. Calculate the residual.
2. Does the model overpredict or underpredict?
3. What are the units of the residual?

3 Fitting a proportional model through the origin

If theory suggests

$$y = kx,$$

we should not automatically fit a general line $y = ax + b$. A nonzero intercept would contradict the proposed mechanism. For data (x_i, y_i), the least-squares value of k minimises

$$S(k) = \sum_{i=1}^{m} (y_i - kx_i)^2.$$

Differentiating and setting $S'(k) = 0$ gives

$$\boxed{k = \frac{\sum x_i y_i}{\sum x_i^2}.}$$

Worked Example 1: Calibrating the braking-distance model

Lecture 2 proposed $d_b = kv^2$, where v is speed in m/s and d_b is braking distance in m. Let

$$X = v^2.$$

Then the model becomes $d_b = kX$, a proportional relationship in the transformed variable X.

v (m/s)	5	10	15	20	25	30
$X = v^2$	25	100	225	400	625	900
d_b (m)	2.0	7.6	17.9	30.1	49.2	68.5

Using the least-squares formula,

$$k = \frac{\sum X_i d_i}{\sum X_i^2} = \frac{109277.5}{1421875} \approx 0.07685.$$

Therefore

$$\boxed{d_b \approx 0.07685v^2.}$$

Because $d_b = v^2/(2a)$ under constant deceleration, the fitted coefficient corresponds to

$$a \approx \frac{1}{2(0.07685)} \approx 6.51 \text{ m/s}^2.$$

The fitted parameter therefore has a physical interpretation rather than being merely a numerical constant.

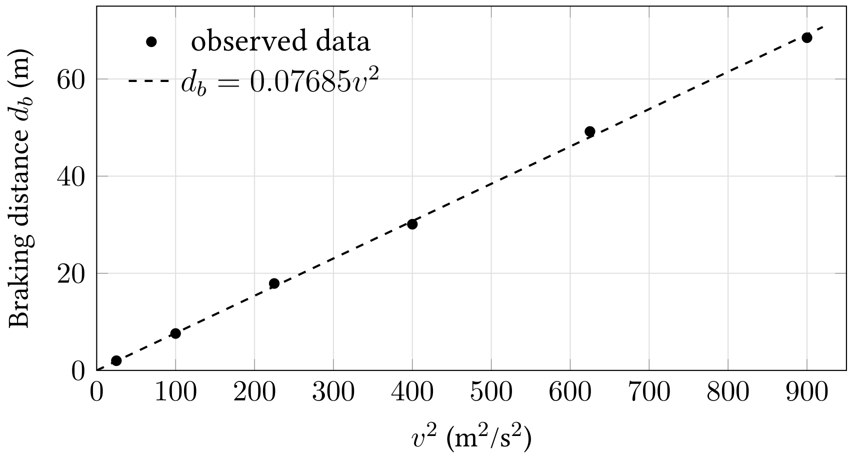


Figure 3: Plotting d_b against v^2 turns the proposed quadratic relationship into a line through the origin.

Model criticism: Should the intercept be forced to zero?

For $v = 0$, physical reasoning requires $d_b = 0$. Forcing the line through the origin is therefore justified. In another problem, a nonzero baseline may be real: an instrument can have an offset, or a delivery fee can include a fixed charge. The fitting constraint must come from the context, not from convenience.

Residual check **[8 min]**

For $v = 25$ m/s, the fitted model predicts

$$\hat{d}_b = 0.07685(25)^2 \approx 48.03 \text{ m}.$$

The observation is 49.2 m.

1. Calculate the residual.
2. Explain whether the model underpredicts or overpredicts.
3. Suggest one physical reason for the discrepancy.

4 Fitting a general straight line

When a nonzero intercept is plausible, fit

$$y = ax + b.$$

The least-squares line minimises

$$S(a,b)=\sum_{i=1}^{m}(y_i-ax_i-b)^2.$$

A convenient form of the solution is

$$\boxed{a=\frac{\sum(x_i-\bar{x})(y_i-\bar{y})}{\sum(x_i-\bar{x})^2},\qquad b=\bar{y}-a\bar{x}.}$$

Worked Example 2: Calibrating a force sensor

Known masses x in kg are placed on a sensor, producing output voltages y.

x (kg)	0	1	2	3	4	5
y (V)	0.18	1.72	3.19	4.76	6.25	7.81

Here $\bar{x}=2.5$, $\bar{y}=3.985$,

$$\sum(x_i-\bar{x})(y_i-\bar{y})=26.655,\qquad \sum(x_i-\bar{x})^2=17.5.$$

Thus

$$a\approx 1.523,\qquad b\approx 0.177,$$

and the calibration model is

$$\boxed{y\approx 1.523x+0.177.}$$

The intercept is meaningful: the unloaded sensor reads about 0.18 V rather than zero. Forcing the model through the origin would hide this offset.

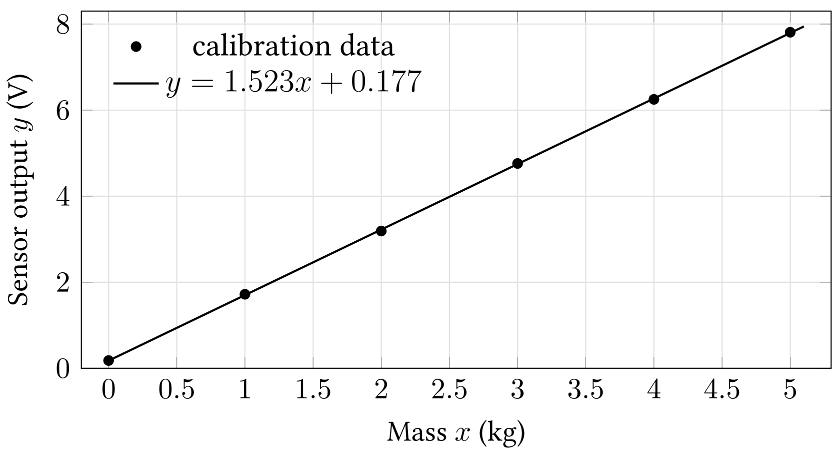


Figure 4: The fitted intercept represents sensor offset. A parameter should be interpreted in the original context.

Why squares?

Least squares minimises

$$\mathrm{SSE}=\sum r_i^2,$$

the **sum of squared errors**. Squaring prevents positive and negative residuals from cancelling and penalises large residuals strongly. This makes the method mathematically convenient, but also sensitive to **outliers**.

5 Error measures and residual plots

No single error measure answers every question. Three common summaries are

$$\text{MAE} = \frac{1}{m}\sum |r_i|, \qquad \text{RMSE} = \sqrt{\frac{1}{m}\sum r_i^2}, \qquad \max |r_i|.$$

The **mean absolute error** is easy to interpret; the **root mean squared error** penalises large errors more strongly; the maximum error focuses on the worst case. All three have the same units as the response variable.

For the braking-distance fit,

$$\text{MAE} \approx 0.54 \text{ m}, \qquad \text{RMSE} \approx 0.66 \text{ m}, \qquad \max |r_i| \approx 1.17 \text{ m}.$$

These values describe the observed calibration range only. They do not prove that the model remains accurate at higher speeds or on different roads.

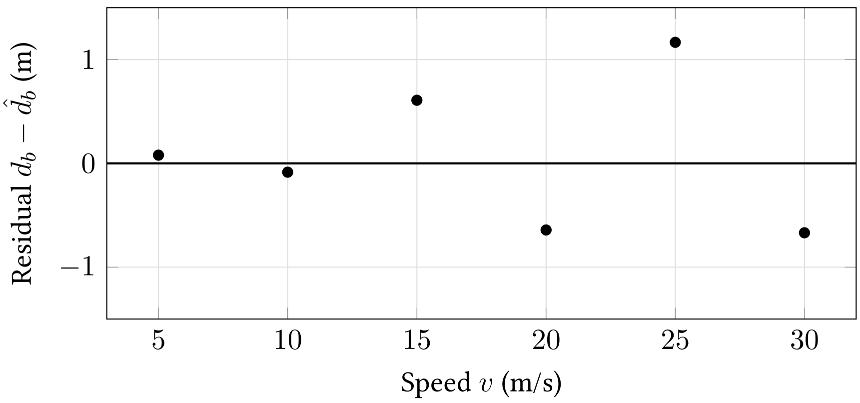


Figure 5: Residuals should be examined for systematic patterns, not merely averaged into one statistic.

Checkpoint: Reading a residual plot

Suppose residuals are positive at low x, negative in the middle, and positive again at high x.

1. Is the scatter random?
2. What does the curved pattern suggest about a fitted straight line?
3. Would collecting more decimal places solve the main problem?

6 Linearising an exponential model

Some model families become linear after a transformation. For an exponential decay model

$$C = Ae^{-kt},$$

taking natural logarithms gives

$$\ln C = \ln A - kt.$$

Thus a plot of $\ln C$ against t should be approximately linear, with slope $-k$ and intercept $\ln A$.

Worked Example 3: Battery capacity decay

A battery is repeatedly tested under the same load. Capacity C is recorded as a percentage of its initial value.

Cycle block t	0	1	2	3	4	5
Capacity C (%)	100	82	68	55	46	37

A least-squares line fitted to $(t, \ln C)$ is approximately

$$\ln C = 4.607 - 0.19765t.$$

Exponentiating gives

$$\boxed{C \approx 100.2e^{-0.19765t}.}$$

The fitted model predicts that capacity is multiplied by

$$e^{-0.19765} \approx 0.821$$

per cycle block, corresponding to an average decrease of about 17.9% of the remaining capacity each block.

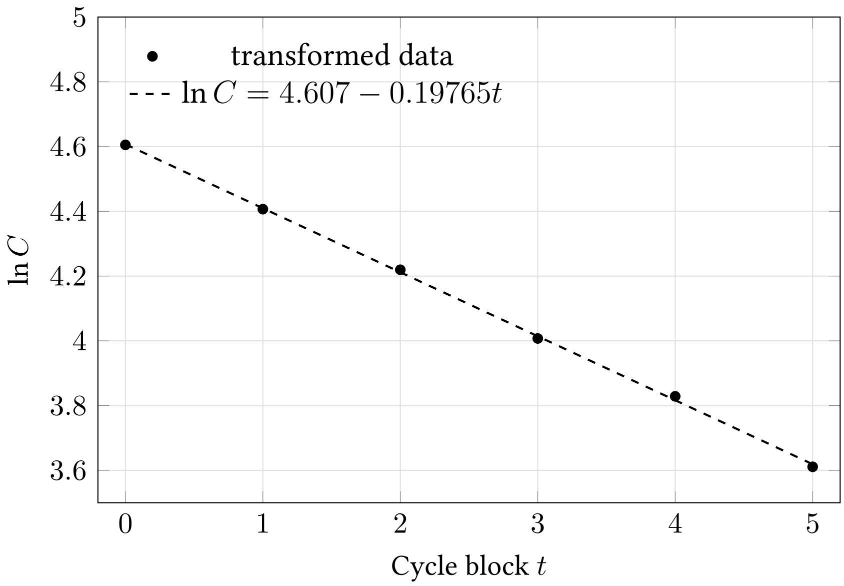


Figure 6: Linearisation makes parameter estimation convenient, but the resulting model must be checked in the original capacity scale.

Model criticism: Transformation changes the fitting criterion

Least squares in $\ln C$ minimises squared differences between logarithms, not squared differences between capacities. It gives greater relative importance to smaller values. Therefore:

1. transform only when the model form and data permit it;
2. convert the fitted model back to the original variables;
3. calculate residuals and error measures in the original scale;
4. compare alternative models using the same response variable.

Choose a transformation **[7 min]**

State the variables that should be plotted to test each proposed model.

1. $y = Ae^{bx}$
2. $y = Ax^n$
3. $y = A/x$

For each case, identify the slope or intercept that has a parameter interpretation.

7 Choosing between models: fit, structure, and validation

A model should not be selected only because it has the smallest calibration error. We also ask whether it respects the mechanism, whether residuals show structure, and whether it predicts data that were not used to fit the parameters.

Worked Example 4: Linear or exponential decay?

Using the six battery observations from $t = 0$ to $t = 5$:

- the exponential fit is $C \approx 100.2e^{-0.19765t}$ with calibration RMSE about 0.39 percentage points;
- the least-squares straight line is $C \approx 95.81 - 12.46t$ with calibration RMSE about 2.93 percentage points.

The value at $t = 6$, $C = 30.5\%$, is withheld during fitting and used for **validation**.

$$\hat{C}_{\text{exp}}(6) \approx 30.61, \qquad \hat{C}_{\text{lin}}(6) \approx 21.07.$$

The exponential model performs much better on the withheld point and never predicts negative capacity. Both the numerical evidence and the mechanism support exponential decay over this range.

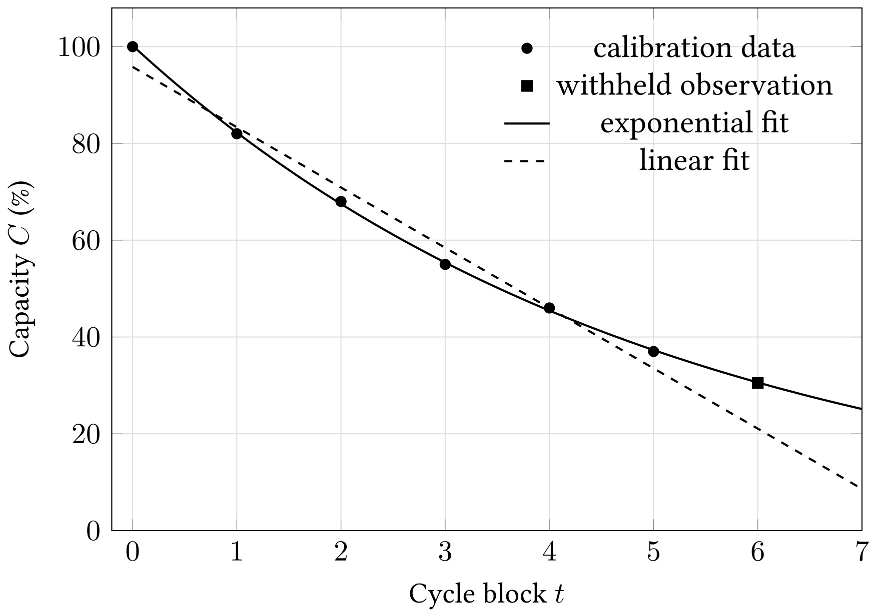


Figure 7: Calibration error, validation performance, and long-term plausibility all favour the exponential model.

A practical model-selection checklist

Before accepting a fitted model, ask:

1. **Mechanism:** Does the functional form agree with the assumptions or governing relationships?
2. **Scale:** Were errors assessed in the original response variable?
3. **Residuals:** Are they small and free of systematic pattern?
4. **Validation**: Does the model predict observations not used for fitting?
5. **Range:** Is the intended prediction inside or outside the observed range?
6. **Interpretability:** Do the parameters and conclusions make sense in context?

Competition-style communication **[10 min]**

Write a four-sentence model-selection paragraph for the battery example:

1. identify the two models compared;
2. report one calibration measure;
3. explain the withheld-data result;
4. state one limitation or condition on the conclusion.

Avoid saying that a high-quality fit "proves" the model.

Optional extension: Modern audit: measurement uncertainty and identifiability

A small residual does not show that every fitted parameter is well determined. If x_i is measured with error but treated as exact, an ordinary least-squares slope can be biased; residuals then combine response noise, input uncertainty, model-form error, and parameter uncertainty. Record instrument resolution and repeated-measurement variability, then propagate plausible perturbations through the fit.

For a model $f(x;\boldsymbol{\theta})$, the parameters are practically **identifiable** only when the available observations can distinguish their effects. A useful local diagnostic is the sensitivity matrix

$$J_{ij} = \frac{\partial f(x_i;\boldsymbol{\theta})}{\partial \theta_j}.$$

If two columns of J are nearly proportional, different parameter combinations produce almost the same fitted curve. More decimal places or a lower training error will not resolve that ambiguity. Report parameter intervals or profiles, test the fit under stated measurement uncertainty, and collect observations where the competing parameter effects differ most. Prediction may still be stable even when individual parameters are not, so audit parameter claims and prediction claims separately.

8 Lesson summary

Seven ideas to retain

1. Model fitting estimates parameters within a proposed model family; interpolation is a different task.
2. A residual is observed minus predicted, with the same units as the response variable.

3. The fitting constraint—through the origin or with an intercept—must be justified by context.
4. Least squares minimises squared residuals and is sensitive to large errors and outliers.
5. Residual plots reveal systematic structure that a single error statistic can hide.
6. Transformations simplify fitting but alter the error criterion; always return to the original variables.
7. A credible model combines mechanism, calibration, validation, interpretation, and stated limitations.

Checkpoint: Exit ticket **[2 min]**

Complete both statements:

1. "A small RMSE is not sufficient because ____________________."
2. "I would force a fitted line through the origin only when ____________________."

9 Practice problems

Core practice

1. A model predicts water demand of 82, 91, and 106 L, while the corresponding observations are 85, 88, and 110 L.
   (a) Calculate the three residuals.
   (b) Find the MAE and RMSE.
   (c) Which observation contributes most strongly to the SSE?
2. The data below relate force F to extension x for a spring:

x (cm)	1	2	3	4	5
F (N)	1.9	4.2	6.0	8.1	10.3

   Fit $F = kx$ by least squares. Interpret the units of k and calculate the residual at $x = 4$ cm.
3. A temperature sensor gives readings

x (actual °C)	0	10	20	30	40
y (sensor °C)	1.2	10.8	20.5	30.6	40.1

   Explain why an intercept should be allowed. Fit a straight line using a spreadsheet or calculator and interpret both parameters.
4. State the transformation for each model and explain what a straight transformed plot would support:
   (a) $y = Ae^{bx}$;
   (b) $y = Ax^n$;
   (c) $y = A/(B + x)$.
5. Two models have the following residuals, measured in minutes:

$$A : (-1.0, -0.8, 0.2, 0.4, 1.2), \qquad B : (-0.2, -0.1, 0.0, 0.1, 3.0).$$

Calculate and compare their MAE, RMSE, and maximum error. Which model is preferred by MAE, which by RMSE, and which for avoiding a severe worst case? Explain why "average performance" needs a stated loss measure here.

6. A fitted quadratic has residuals that alternate randomly around zero, while a fitted line has residuals forming a clear U-shape. Explain what this suggests. Why is "the quadratic has more parameters" not by itself a sufficient conclusion?

7. A population model is fitted using years 2018–2024 and then used to predict 2045. Explain why a small training RMSE does not establish that the 2045 forecast is trustworthy. Name two additional checks.

Practice check and hints

1. Residuals are $(3, -3, 4)$ L; MAE $= 3.33$ L, RMSE ≈ 3.37 L, and the third observation contributes most to SSE.

2. $k = 2.04$ N/cm; at $x = 4$ cm the residual is $8.1 - 8.16 = -0.06$ N.

3. A free-intercept fit is $y \approx 0.976x + 1.12$; the intercept represents the zero-load offset.

4. Use $(x, \ln y)$, $(\ln x, \ln y)$, and $(x, 1/y)$, respectively; a line supports the stated family but does not by itself validate it.

5. For A, MAE $= 0.72$, RMSE ≈ 0.81, maximum error $= 1.2$; for B, the values are 0.68, 1.35, and 3.0. Thus MAE prefers B, whereas RMSE and worst-case control prefer A.

6. The U-shaped line residuals indicate missed curvature. Compare withheld performance and residual structure before accepting the quadratic.

7. A credible long-horizon check includes temporal holdout testing plus at least one of mechanism review, scenario sensitivity, or comparison with an alternative model family.

Modeling challenge

Choose one dataset that your team can collect or obtain: cooling temperature, queue waiting time, plant growth, device battery decay, transport travel time, or another approved context.
Prepare a one-page fitting brief containing:

1. the question, variables, units, and proposed model family;
2. a justification for allowing or excluding an intercept;
3. fitted parameter values and one graph in the original variables;
4. a residual plot and at least two error measures;
5. one validation strategy and one limitation;
6. a short conclusion written for a non-specialist decision maker.

The goal is not to obtain the lowest possible error by adding terms. The goal is to produce a defensible model.

A Optional derivation: the least-squares line

This appendix is not part of the two-hour core lesson. It may be assigned to students who are comfortable with partial differentiation.

For $y = ax + b$, minimise

$$S(a, b) = \sum_{i=1}^{m} (y_i - ax_i - b)^2.$$

Setting both partial derivatives equal to zero gives

$$\frac{\partial S}{\partial a} = -2 \sum x_i (y_i - ax_i - b) = 0,$$
$$\frac{\partial S}{\partial b} = -2 \sum (y_i - ax_i - b) = 0.$$

Therefore

$$a \sum x_i^2 + b \sum x_i = \sum x_i y_i,$$
$$a \sum x_i + mb = \sum y_i.$$

These are the **normal equations**. Solving them is equivalent to the mean-centred formulas used in the lesson.

B Optional extension: alternatives to least squares

Least squares is not the only defensible fitting criterion.

- The L^1 criterion minimises $\sum |r_i|$ and is less sensitive to a single large outlier.
- The L^∞ or **minimax criterion** minimises $\max |r_i|$ and is useful when the worst error matters most.

The appropriate criterion depends on the decision. A medical safety model, a forecasting model, and an engineering calibration problem may not value errors in the same way.

Model audit card

Audit before trusting or communicating a model

Purpose	What question or decision does the model support?	**Data**	Which inputs are observed, derived, or simulated?
Assumptions	Which simplifications may change the answer?	**Units**	Are dimensions and scales consistent?
Verification	Were equations, code, constraints, and limiting cases checked?	**Validation**	Is performance tested against independent evidence?
Uncertainty	How do measurement, parameter, solver, and structural errors affect conclusions?	**Ethics**	Who benefits, who bears risk, and what fairness criterion matters?
Reproducibility	Are data provenance, versions, seeds, and transformations recorded?	**Communication**	Are claims conditional, traceable, and decision-relevant?

Further reading

- Engineering Statistics Handbook (NIST / SEMATECH)
- Introductory Statistics 2e (OpenStax)

Sources, provenance, and licence

These notes follow the iterative modeling and assessment perspective in the SIAM/COMAP GAIMME report. Competition-specific remarks are a snapshot checked on 20 August 2026 against the official COMAP HiMCM instructions; readers should verify current rules before submission. Unless a source is explicitly identified, classroom datasets and numerical scenarios are synthetic or simulated teaching examples and are not empirical claims about a real institution. Third-party references retain their own terms.

Student response guide and checking criteria

For closed calculations, show the definition, substitution, units, and at least one independent check; the worked examples in this unit are method hints, but you must complete the arithmetic yourself. Open modeling responses are checked on five dimensions: a clear purpose and output; stated data and provenance; defensible assumptions and scope; traceable computational verification and model validation; and alignment among uncertainty, limitations, and the recommendation. Support each dimension with concrete evidence rather than writing only “completed.”

Lectures in Mathematical Modeling

Lecture 4: Empirical Models, Interpolation, and Splines

Kenneth Sok Kin Cheng

Revised 20 August 2026

Learning goals

By the end of the lesson, readers should be able to:

1. distinguish **interpolation**, **smoothing**, and **extrapolation**;
2. construct and interpret a linear interpolant between two observations;
3. build a quadratic interpolating polynomial in Lagrange form;
4. use a difference table to recognise simple polynomial structure;
5. explain why an exact high-degree interpolant may behave poorly between data points;
6. compare piecewise linear interpolation with a natural cubic spline and select an appropriate empirical method.

1 When the data come before the formula

Lecture 3 began with a proposed model family and estimated its parameters from data. In many practical situations, however, no convincing mechanism is yet available. We may only have measurements at selected times or locations and need a reasonable value between them. This leads to an **empirical model**: a mathematical representation constructed mainly from observed data.

Launch: Should the curve pass through every point? **[10 min]**

A classroom sensor records carbon-dioxide concentration once every five minutes:

time (min)	0	5	10	15	20
CO_2 (ppm)	612	648	701	694	760

In pairs, discuss:

1. How would you estimate the concentration at minute 12?
2. Should the estimated curve pass through every recorded value?
3. Which matters more here: measurement noise or exact agreement with the observations?
4. Would the same method be defensible for predicting minute 60?

Modeling lesson: The purpose of the model and the reliability of the data determine whether exact interpolation is desirable.

Interpolation, smoothing, and extrapolation

- **Interpolation** constructs a function that agrees with specified data values and estimates values *inside* the observed range.
- **Smoothing** allows small deviations from noisy data to reveal an underlying trend.
- **Extrapolation** predicts *outside* the observed range. It depends on assumptions that the data alone cannot verify.

An interpolant can be mathematically exact and still be scientifically untrustworthy if the observations contain noise or the curve behaves implausibly between them.

Model criticism: Exact agreement is not the same as credibility

If a thermometer is accurate only to the nearest degree, forcing a complicated curve through every reported value may model rounding noise rather than temperature. Conversely, if the data are exact geometric coordinates in computer graphics, passing through every point may be essential. The data-generating process must guide the method.

2 Linear interpolation and piecewise linear models

Suppose two trusted observations are (x_0, y_0) and (x_1, y_1), with $x_0 < x_1$. For a point x between them, define the relative position

$$\lambda = \frac{x - x_0}{x_1 - x_0}, \qquad 0 \le \lambda \le 1.$$

The **linear interpolant** is

$$\boxed{L(x) = (1-\lambda)y_0 + \lambda y_1 = y_0 + \frac{y_1 - y_0}{x_1 - x_0}(x - x_0).}$$

The formula is a weighted average: when x is closer to x_0, more weight is placed on y_0.

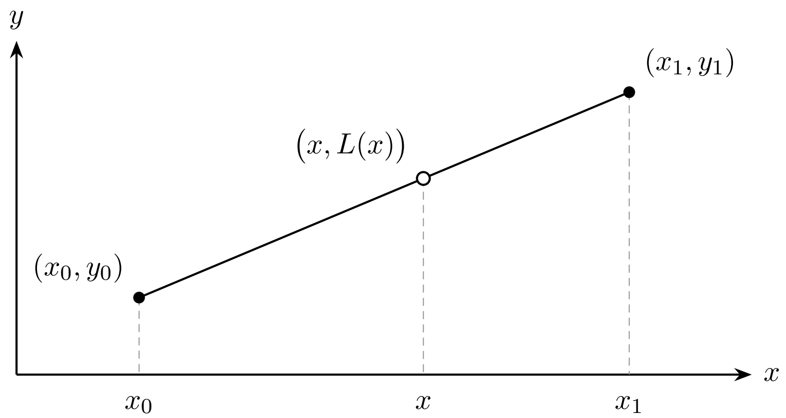


Figure 1: The interpolated point lies on the line segment joining the two observations; its horizontal position determines the weights assigned to y_0 and y_1.

Worked Example 1: Estimating temperature between readings

A heating experiment records 38.0°C at $t = 4$ minutes and 62.0°C at $t = 10$ minutes. Estimate the temperature at $t = 7.5$ minutes.

The relative position is

$$\lambda = \frac{7.5 - 4}{10 - 4} = \frac{3.5}{6}.$$

Hence

$$T(7.5) = 38.0 + \frac{3.5}{6}(62.0 - 38.0) = 38.0 + 14.0 = \boxed{52.0^\circ\text{C}}.$$

The estimate assumes that temperature changed at an approximately constant rate over this six-minute interval. It does not claim that the entire heating process is linear.

With several observations, we may connect each consecutive pair by a line. The resulting **piecewise linear interpolant** is continuous and simple to compute, but its slope changes abruptly at the data points.

Guided practice: Build a piecewise model **[10 min]**

A library records the number of occupied seats:

$$(0, 20), \qquad (2, 44), \qquad (5, 68),$$

where x is hours after opening.

1. Write the linear formula on $0 \leq x \leq 2$.
2. Write the linear formula on $2 \leq x \leq 5$.
3. Estimate occupancy at $x = 1.5$ and $x = 3.5$.
4. Compare the slopes of the two pieces. What does the change of slope mean in context?

Checkpoint: Interpolation or extrapolation?

For the temperature observations at $t = 4$ and $t = 10$ minutes, classify each request:

1. estimate $T(8)$;
2. estimate $T(15)$;
3. estimate the maximum temperature after one hour.

Explain why only one request can be answered by linear interpolation alone.

3 Quadratic interpolation through three points

Two points determine a line; three points with distinct x-coordinates determine a unique polynomial of degree at most two. The **Lagrange form** constructs this polynomial directly.

For three points (x_0, y_0), (x_1, y_1), and (x_2, y_2),

$$P_2(x) = y_0 L_0(x) + y_1 L_1(x) + y_2 L_2(x),$$

where

$$L_0(x) = \frac{(x - x_1)(x - x_2)}{(x_0 - x_1)(x_0 - x_2)},$$
$$L_1(x) = \frac{(x - x_0)(x - x_2)}{(x_1 - x_0)(x_1 - x_2)},$$
$$L_2(x) = \frac{(x - x_0)(x - x_1)}{(x_2 - x_0)(x_2 - x_1)}.$$

Each basis polynomial satisfies $L_i(x_i) = 1$ and is zero at the other two nodes.

Worked Example 2: A quadratic interpolant

Construct the polynomial through

$$(0,1), \qquad (1,3), \qquad (2,2).$$

The basis functions are

$$L_0(x) = \frac{(x-1)(x-2)}{2}, \qquad L_1(x) = -x(x-2), \qquad L_2(x) = \frac{x(x-1)}{2}.$$

Therefore

$$\begin{aligned} P_2(x) &= L_0(x) + 3L_1(x) + 2L_2(x) \\ &= -\frac{3}{2}x^2 + \frac{7}{2}x + 1. \end{aligned}$$

At $x = 1.5$,

$$P_2(1.5) = \boxed{2.875}.$$

Substitution verifies $P_2(0) = 1$, $P_2(1) = 3$, and $P_2(2) = 2$ exactly.

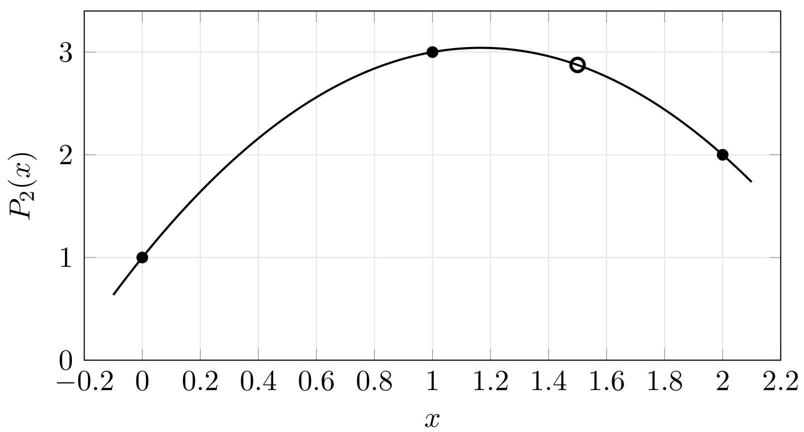


Figure 2: The solid curve is $P_2(x) = -1.5x^2 + 3.5x + 1$. Filled points are the interpolation nodes; the open point is the estimate at $x = 1.5$.

Model criticism: What does the curve between the points mean?

The interpolation conditions determine the quadratic uniquely, but they do not prove that the real process is quadratic. A different physical mechanism could produce the same three measurements and different values between them. Interpolation supplies a convenient curve; scientific interpretation still requires context.

4 Difference tables as a diagnostic

For equally spaced values of x, successive differences can reveal low-degree polynomial structure:

- constant first differences suggest a linear model;
- constant second differences suggest a quadratic model;
- constant third differences suggest a cubic model.

The result is exact for perfect polynomial data and only approximate for noisy observations.

Worked Example 3: Recognising a quadratic pattern

Consider

x	0	1	2	3	4
y	2	5	10	17	26

The difference table is

y	2	5	10	17	26
Δy		3	5	7	9
$\Delta^2 y$			2	2	2

The second differences are constant, so the data lie on a quadratic. For unit spacing, the leading coefficient is half the constant second difference, giving $a = 1$. Writing

$$y = x^2 + bx + c$$

and using $y(0) = 2$ and $y(1) = 5$ gives $c = 2$ and $b = 2$. Thus

$$\boxed{y = x^2 + 2x + 2.}$$

What a difference table can and cannot do

A difference table is a diagnostic, not a certificate of truth. With measured data, higher-order differences often amplify noise. The method is most useful for identifying a simple pattern in accurate, equally spaced data. For unequally spaced nodes, use **divided differences**; an optional introduction appears in Appendix A.

Pattern or noise? **[8 min]**

For each sequence, calculate enough differences to suggest a simple model.

1. $4,\ 9,\ 16,\ 25,\ 36$
2. $3.0,\ 6.1,\ 11.9,\ 21.2,\ 34.0$

For the second sequence, explain why “approximately quadratic” is more defensible than “exactly quadratic.”

5 Why a high-degree polynomial can fail

Given $n+1$ distinct points, a unique polynomial of degree at most n passes through all of them. This theorem guarantees existence, not good behavior. As the degree increases, a global polynomial can oscillate strongly between nodes, especially near the endpoints.

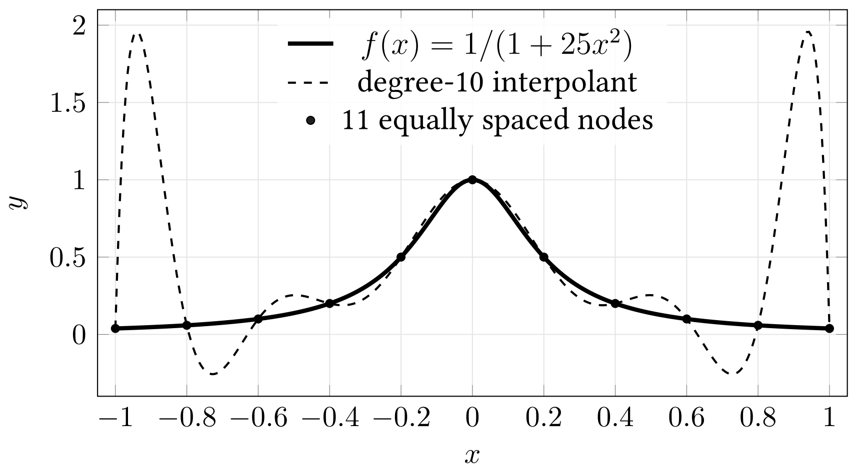


Figure 3: The polynomial matches every node but oscillates badly between them. This is **Runge's phenomenon**.

Model criticism: The zero-residual trap

At the interpolation nodes, every residual is zero. Therefore SSE, MAE, and RMSE computed only at those nodes are also zero, even though the curve is poor between them. Model assessment must consider behavior between observations, sensitivity to data perturbations, and the intended use of the curve.

Possible remedies are to lower the degree, smooth noisy data, or use better-spaced nodes. Another option is to replace the global polynomial with piecewise low-degree polynomials.

6 Piecewise interpolation and cubic splines

A **spline** uses a separate polynomial on each interval between consecutive **knots**. A linear spline joins the points by straight segments. A **cubic spline** uses cubic pieces while imposing smooth joins.

For a cubic spline S:

1. S passes through every data point;
2. adjacent pieces have the same value at each interior knot;
3. their first derivatives agree, so there is no corner;
4. their second derivatives agree, so curvature changes smoothly.

A **natural cubic spline** additionally satisfies $S'' = 0$ at the two endpoints.

Worked Example 4: Natural spline through three points

Construct the natural cubic spline through

$$(0, 0), \qquad (1, 1), \qquad (2, 0).$$

Let $M_i = S''(x_i)$. Natural endpoint conditions give $M_0 = M_2 = 0$. With equal spacing $h = 1$, the single interior spline equation is

$$M_0 + 4M_1 + M_2 = 6(y_2 - 2y_1 + y_0) = -12,$$

so $M_1 = -3$. The resulting spline is

$$S(x) = \begin{cases} -\frac{1}{2}x^3 + \frac{3}{2}x, & 0 \le x \le 1, \\ 1 - \frac{3}{2}(x-1)^2 + \frac{1}{2}(x-1)^3, & 1 \le x \le 2. \end{cases}$$

At $x = 0.5$, the spline gives $S(0.5) = 0.6875$, while linear interpolation gives 0.5. Neither value is automatically "correct"; the smoother cubic estimate encodes stronger assumptions about how the slope changes.

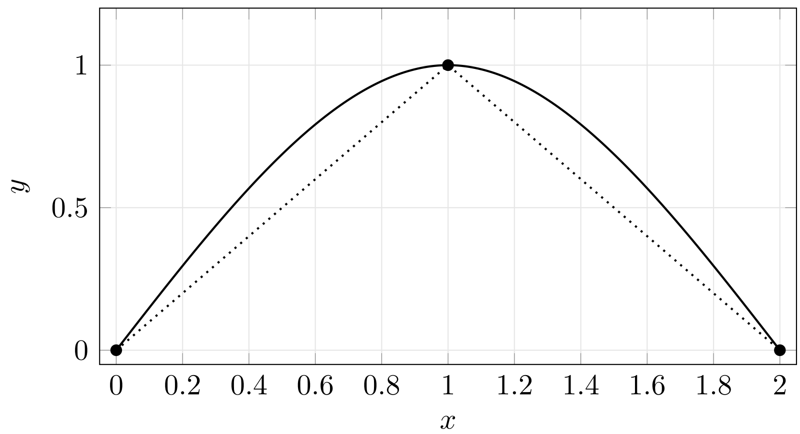


Figure 4: The dotted line is the linear spline, the solid curve is the natural cubic spline, and the filled points are the knots. Both interpolate the same data, but only the cubic spline has continuous slope and curvature at $x = 1$.

7 Choosing an empirical method

Method	Strength	Main caution
Linear interpolation	Transparent, local, easy to compute	Corners at knots; assumes constant rate within each interval
Quadratic or low-degree interpolation	Captures curvature with few points	A small dataset does not prove the polynomial mechanism
High-degree global polynomial	Passes through every point	Oscillation, sensitivity, and poor extrapolation
Low-degree smoothing model	Handles noisy data and reveals trend	Does not reproduce every observation exactly
Cubic spline	Local, smooth, accurate between many trusted points	May overshoot; boundary conditions matter

Checkpoint: Which property matters?

A road profile is measured at survey points and will be used to estimate vehicle motion. Would you prefer a linear spline or a cubic spline? State one benefit and one risk of your choice.

Model criticism: Cubic splines are not automatically shape-preserving

A cubic spline is smooth and usually stable, but it may overshoot between data points or create negative values when the modeled quantity must remain nonnegative. For monotone or bounded data, specialised shape-preserving splines may be more appropriate.

Method-selection studio **[12 min]**

Choose an appropriate method for each situation and justify it in one or two sentences.

1. Exact coordinates define the edge of a manufactured part.
2. Daily demand data contain substantial random variation, and the goal is to reveal the weekly trend.
3. A temperature sensor failed for two minutes between two reliable readings.
4. Ten trusted measurements describe a smooth road elevation profile.
5. A team wants to forecast ten years beyond a five-year dataset.

For the final situation, explain why none of the interpolation methods is sufficient by itself.

A practical empirical-modeling workflow

1. Clarify whether the task is interpolation, smoothing, or extrapolation.
2. Plot the original data and check units, ordering, missing values, and outliers.
3. Decide whether observations should be treated as exact or noisy.
4. Start with the simplest local method that meets the purpose.
5. Inspect the curve between points, not only at the observations.
6. Test sensitivity to removing or perturbing a data point.
7. State the valid interval and any shape constraints explicitly.

8 Lesson summary

Six ideas to retain

1. Interpolation estimates inside the data range; extrapolation requires additional assumptions.
2. Linear interpolation is a weighted average between two observations.
3. Three distinct nodes determine a unique quadratic, but not necessarily the true mechanism.
4. Difference tables can reveal simple polynomial patterns in equally spaced data.
5. Zero residuals at interpolation nodes do not guarantee good behavior between them.
6. Splines replace one unstable high-degree polynomial with smooth, local, low-degree pieces.

Checkpoint: Exit ticket **[2 min]**

Complete both statements:

1. "I would interpolate rather than smooth when ____________________________________."

2. "A curve passing through every point can still be poor because ________________________________."

9 Practice problems

Core practice

1. A sensor reads 18.2 units at $t = 3$ and 25.4 units at $t = 9$.
   (a) Use linear interpolation to estimate the value at $t = 5.5$.
   (b) State the assumption behind the calculation.
   (c) Explain why the same formula should not automatically be used at $t = 15$.
2. Construct the piecewise linear interpolant through $(0, 2)$, $(2, 6)$, and $(5, 3)$. Evaluate it at $x = 1$, $x = 3$, and $x = 4.5$. Identify the point where the slope changes.
3. Find the quadratic interpolating polynomial through $(0, 2)$, $(1, 1)$, and $(3, 5)$ using the Lagrange form. Evaluate the polynomial at $x = 2$ and verify all three interpolation conditions.
4. Construct a difference table for
$$7,\ 12,\ 19,\ 28,\ 39,\ 52.$$
   Suggest a polynomial degree and find a formula in terms of $n = 0, 1, 2, \ldots$.
5. A degree-12 polynomial passes through 13 experimental measurements exactly. Give three reasons why a lower-degree smoother or spline may nevertheless be more credible.
6. For the points $(0, 0)$, $(1, 2)$, and $(2, 0)$:
   (a) write the linear spline;
   (b) describe the continuity properties required of a cubic spline;
   (c) without calculating the full cubic spline, predict how its graph would differ near $x = 1$.
7. Decide whether interpolation, smoothing, or neither is the main task:
   (a) estimating rainfall at an unmonitored hour between two observations;
   (b) finding the long-run trend in noisy monthly sales;
   (c) predicting population in 2070 using records ending in 2025;
   (d) reconstructing a computer-drawn curve from exact control points.

Practice check and hints

1. Linear interpolation gives 21.2 units; it assumes an approximately constant rate between $t = 3$ and $t = 9$, not beyond the observed interval.
2. The pieces are $2 + 2x$ for $0 \le x \le 2$ and $8 - x$ for $2 \le x \le 5$; the requested values are 4, 5, and 3.5, with a slope change at $x = 2$.
3. $P(x) = x^2 - 2x + 2$, so $P(2) = 2$; direct substitution verifies all three nodes.
4. Constant second differences equal to 2 support a quadratic: $y_n = n^2 + 4n + 7$.
5. Relevant concerns include oscillation between nodes, noise amplification, endpoint behavior, sensitivity, and withheld-data performance.

6. The linear spline is $2x$ then $4 - 2x$. A cubic spline is continuous with continuous first and second derivatives, so it rounds the corner near $x = 1$.
7. The classifications are interpolation, smoothing, neither (long extrapolation), and interpolation, respectively.

Modeling challenge

A school records pedestrian flow at a corridor entrance every five minutes, but several readings are missing. Some recorded values may also contain counting errors.
Prepare a one-page empirical-modeling brief containing:

1. the purpose of reconstructing the missing data;
2. which observations you would treat as reliable or noisy;
3. a comparison of piecewise linear interpolation, cubic splines, and smoothing;
4. one diagnostic graph or calculation;
5. the interval on which your result is valid;
6. one sensitivity test and one limitation.

The quality of the justification matters more than using the most complicated method.

A Optional extension: Newton divided differences

For unequally spaced nodes, ordinary differences are replaced by **divided differences**:

$$f[x_i, x_{i+1}] = \frac{f(x_{i+1}) - f(x_i)}{x_{i+1} - x_i},$$

$$f[x_i, \ldots, x_{i+k}] = \frac{f[x_{i+1}, \ldots, x_{i+k}] - f[x_i, \ldots, x_{i+k-1}]}{x_{i+k} - x_i}.$$

The Newton interpolating polynomial is

$$P_n(x) = f[x_0] + f[x_0, x_1](x - x_0) + \cdots + f[x_0, \ldots, x_n] \prod_{j=0}^{n-1} (x - x_j).$$

Unlike the Lagrange form, the Newton form can be extended efficiently when a new data point is added.

Unequally spaced data

For $(0, 1)$, $(1, 3)$, and $(3, 7)$,

$$f[0, 1] = 2, \qquad f[1, 3] = 2,$$

so

$$f[0, 1, 3] = \frac{2 - 2}{3 - 0} = 0.$$

Hence $P_2(x) = 1 + 2x$, showing that the three points are actually collinear even though the spacing is unequal.

B Optional extension: General natural-spline system

Let $x_0 < \cdots < x_n$, $h_i = x_{i+1} - x_i$, and $M_i = S''(x_i)$. For a natural cubic spline, $M_0 = M_n = 0$, and the interior second derivatives satisfy

$$h_{i-1}M_{i-1} + 2(h_{i-1} + h_i)M_i + h_i M_{i+1} = 6\left(\frac{y_{i+1} - y_i}{h_i} - \frac{y_i - y_{i-1}}{h_{i-1}}\right).$$

This is a tridiagonal linear system, so a spline with many knots can be computed efficiently. The formula is useful for implementation but is not part of the two-hour core lesson.

Model audit card

Audit before trusting or communicating a model

Purpose	What question or decision does the model support?	**Data**	Which inputs are observed, derived, or simulated?
Assumptions	Which simplifications may change the answer?	**Units**	Are dimensions and scales consistent?
Verification	Were equations, code, constraints, and limiting cases checked?	**Validation**	Is performance tested against independent evidence?
Uncertainty	How do measurement, parameter, solver, and structural errors affect conclusions?	**Ethics**	Who benefits, who bears risk, and what fairness criterion matters?
Reproducibility	Are data provenance, versions, seeds, and transformations recorded?	**Communication**	Are claims conditional, traceable, and decision-relevant?

Further reading

- Interpolation user guide (SciPy)
- Calculus Volume 1 (OpenStax)

Sources, provenance, and licence

These notes follow the iterative modeling and assessment perspective in the SIAM/COMAP GAIMME report. Competition-specific remarks are a snapshot checked on 20 August 2026 against the official COMAP HiMCM instructions; readers should verify current rules before submission. Unless a source is explicitly identified, classroom datasets and numerical scenarios are synthetic or simulated teaching examples and are not empirical claims about a real institution. Third-party references retain their own terms.

Student response guide and checking criteria

For closed calculations, show the definition, substitution, units, and at least one independent check; the worked examples in this unit are method hints, but you must complete the arithmetic yourself. Open modeling responses are checked on five dimensions: a clear purpose and output; stated data and provenance; defensible assumptions and scope; traceable computational verification and model validation; and alignment among uncertainty, limitations, and the recommendation. Support each dimension with concrete evidence rather than writing only “completed.”

Lectures in Mathematical Modeling

Lecture 5: Monte Carlo and Discrete-Event Simulation

Kenneth Sok Kin Cheng

Revised 20 August 2026

Learning goals

By the end of the lesson, readers should be able to:

1. distinguish a **deterministic model** from a **stochastic model**;
2. map uniform random numbers to outcomes from a discrete or empirical probability distribution;
3. construct and interpret a simple **Monte Carlo estimator**;
4. explain why repeated simulation runs are needed and why simulation error decreases slowly;
5. trace a single-server queue using arrival, service-start, waiting, and departure times;
6. design a simulation experiment that includes outputs, replications, validation, sensitivity, and limitations.

1 Why simulate?

Previous lectures represented a system using equations or curves. That approach remains valuable, but some systems contain uncertain events whose exact sequence cannot be predicted: customers arrive at irregular times, service durations vary, and equipment may fail unexpectedly. A **simulation** imitates the operation of such a system and records what happens under specified assumptions.

Launch: The canteen queue **[10 min]**

At lunchtime, students arrive at one canteen counter at irregular times. Some orders take much longer than others. The school is considering opening a second counter.

In pairs, identify:

1. two quantities that vary unpredictably from student to student;
2. two quantities that describe the current *state* of the system;
3. three outputs that the school should examine before making a decision;
4. one factor that might make a simulation unrealistic.

Modeling lesson: A simulation requires more than random numbers. It needs explicit input assumptions, update rules, outputs, and a plan for checking credibility.

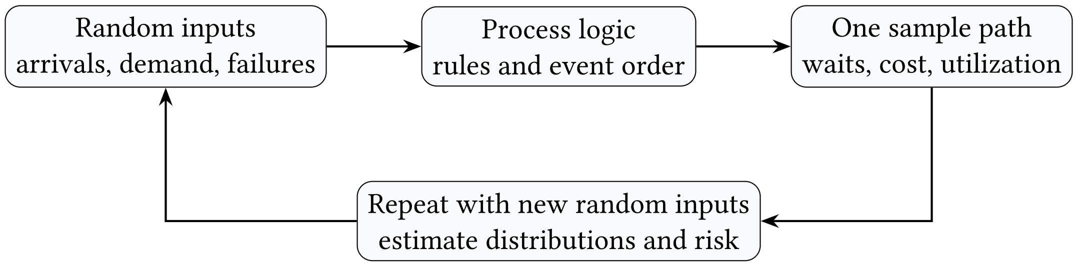


Figure 1: A simulation combines random inputs with deterministic update rules. One run gives one possible history; repeated runs reveal the range of possible outcomes.

Core simulation language

- A **state variable** describes the system at a given time, such as queue length or inventory level.
- An **event** changes the state, such as an arrival, a service completion, or a delivery.
- A **sample path** is one simulated history of the system.
- A **replication** is an independent run under the same policy and assumptions.
- A **random seed** reproduces the same pseudorandom sequence, which is useful for debugging and fair comparisons.

The computer performs deterministic calculations after the random inputs have been generated. Simulation is therefore structured experimentation, not random guessing.

2 From uniform random numbers to model inputs

A **random variable** assigns a numerical value to each possible outcome of a random experiment. To simulate a discrete random variable, partition the interval $[0,1)$ according to cumulative probabilities and generate $U \sim \text{Uniform}(0,1)$.

Suppose a service time S has possible values $s_1, \dots, s_k$ with probabilities $p_1, \dots, p_k$. Let

$$F_j = p_1 + \cdots + p_j.$$

Assign $S = s_j$ when $F_{j-1} \le U < F_j$, where $F_0 = 0$. This is the discrete **inverse transform method**.

Worked Example 1: Simulating service times

Historical observations at a snack counter suggest

service time S (min)	2	3	4	5
$P(S=s)$	0.20	0.35	0.30	0.15

The cumulative probabilities are 0.20, 0.55, 0.85, and 1.00, so

$$S = \begin{cases} 2, & 0 \le U < 0.20, \\ 3, & 0.20 \le U < 0.55, \\ 4, & 0.55 \le U < 0.85, \\ 5, & 0.85 \le U < 1. \end{cases}$$

For $U = 0.07, 0.48, 0.61, 0.93$, the simulated service times are $2, 3, 4, 5$ minutes respectively.

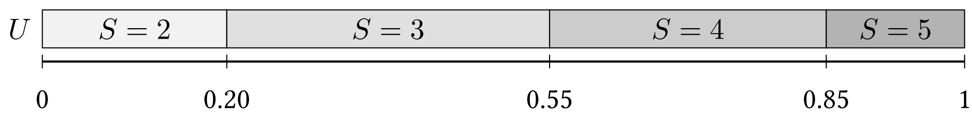


Figure 2: The length of each interval equals the probability of its outcome. A larger probability occupies more of $[0, 1)$.

Guided practice: Interarrival times **[10 min]**

Student interarrival times have the distribution

A (min)	2	3	4	5
$P(A = a)$	0.15	0.35	0.35	0.15.

1. Construct the cumulative probability table and the interval mapping.
2. Convert $U = 0.08, 0.32, 0.71, 0.94, 0.51$ into interarrival times.
3. Explain why a fixed seed is helpful when comparing one counter with two counters.

Model criticism: The input distribution is part of the model

A simulation can produce precise-looking output from a poor input model. Historical frequencies may be unrepresentative of examination days, special menus, or changing student behavior. The probabilities, dependence between variables, and time period represented by the data must all be justified.

3 Monte Carlo estimation

A **Monte Carlo simulation** estimates a quantity by repeating a random experiment. Consider points drawn uniformly from the unit square. The probability of falling inside the quarter circle $x^2 + y^2 \le 1$ is its area, $\pi/4$. If C of n points fall inside, then

$$\frac{C}{n} \approx \frac{\pi}{4}, \qquad \boxed{\widehat{\pi} = 4\frac{C}{n}.}$$

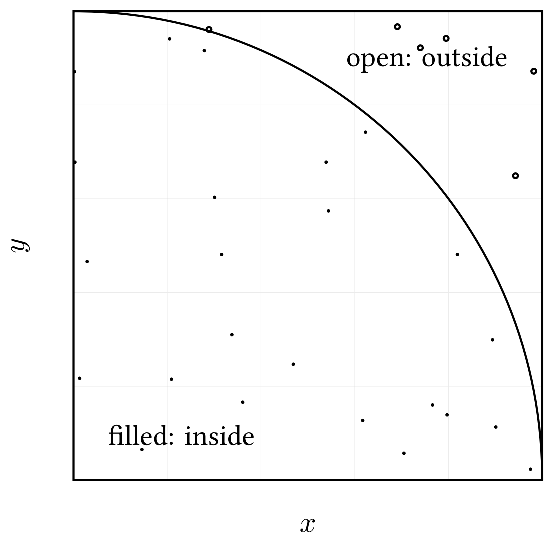


Figure 3: 30 reproducible sample points: 24 fall inside the quarter circle and 6 outside. The picture represents one sample path, not a guarantee of accuracy.

Worked Example 2: One estimate of π

In Figure 3, $C = 24$ and $n = 30$, so

$$\widehat{\pi} = 4\left(\frac{24}{30}\right) = \boxed{3.20}.$$

The estimate is plausible but rough. A different seed would produce a different count. The value 3.20 should therefore be reported together with n, the random-number procedure, and preferably results from repeated runs.

For independent samples, Monte Carlo uncertainty usually decreases at the rate

$$\text{typical error} \propto \frac{1}{\sqrt{n}}.$$

Thus reducing the typical error by a factor of 2 requires about 4 times as many samples; reducing it by a factor of 10 requires about 100 times as many.

Checkpoint: Convergence

A simulation uses $2{,}500$ independent samples. Approximately how many samples are required to reduce its typical Monte Carlo error to one-fifth of the current value? Explain using the square-root rule.

Replication and uncertainty

Increasing the number of observations within one run and increasing the number of independent replications answer different questions. A long run may estimate steady performance; several replications reveal run-to-run variability. Report at least a mean and a measure of spread, and retain the individual replication results rather than only the grand average.

4 Discrete-event simulation of a queue

In a **discrete-event simulation**, the state changes only when an event occurs. For a single-server queue, let

- A_i be the arrival time of customer i;
- S_i be the required service time;
- B_i be the service-start time;
- D_i be the departure time;
- $W_i = B_i - A_i$ be the waiting time.

The update equations are

$$\boxed{B_i = \max(A_i, D_{i-1}), \qquad D_i = B_i + S_i, \qquad W_i = B_i - A_i,}$$

with $D_0 = 0$. The maximum expresses the queue logic: service begins only after both the customer has arrived and the server is free.

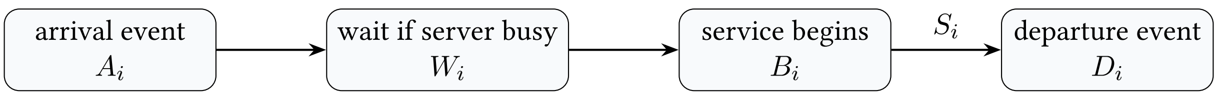


Figure 4: A queue is governed by event order. Randomness enters through arrival and service times; the update logic is deterministic.

Worked Example 3: Manual trace of one counter

Six customers have arrival times

$$A_i = (0, 3, 5, 10, 12, 16)$$

and service times

$$S_i = (4, 5, 3, 4, 2, 5)$$

in minutes. Applying the update equations gives:

Customer i	1	2	3	4	5	6
Arrival A_i	0	3	5	10	12	16
Service S_i	4	5	3	4	2	5
Start B_i	0	4	9	12	16	18
Wait W_i	0	1	4	2	4	2
Departure D_i	4	9	12	16	18	23

The average waiting time is

$$\bar{W} = \frac{0+1+4+2+4+2}{6} = \boxed{2.17 \text{ min}},$$

and the maximum waiting time is 4 minutes. These values describe only this six-customer sample path.

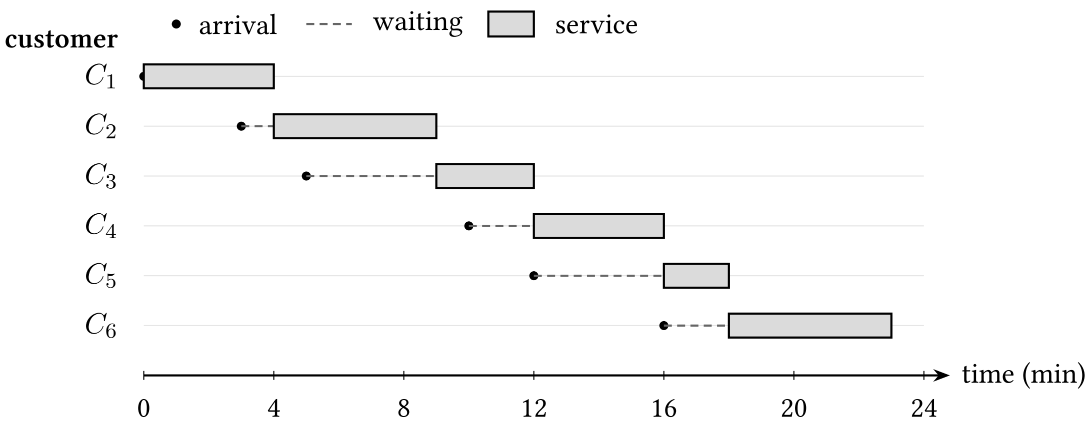


Figure 5: Customer-centred queue timeline. A filled point marks arrival, a dashed segment shows waiting, and a shaded block shows the service interval.

Policy comparison: Add a second counter **[15 min]**

Use the same arrivals and service times as Worked Example 3. Assign each arriving customer to the counter that becomes available first.

1. Complete a table containing counter number, start time, waiting time, and departure time.
2. Find the average and maximum waiting times.
3. Explain why this single sample path is insufficient evidence for permanently staffing a second counter.

5 From one run to a decision

A simulation study is an experiment on a model. Before running it, define the policy alternatives, input distributions, performance measures, simulation length, number of replications, and comparison method.

A useful preliminary check for a queue is the **traffic intensity**

$$\rho \approx \frac{\text{mean service time}}{c \times \text{mean interarrival time}},$$

where c is the number of servers. If $\rho > 1$, work arrives faster than the system can process it, so the queue tends to grow rather than settle to a stable long-run distribution.

Worked Example 4: Capacity before detailed simulation

Suppose the mean interarrival time is 3.5 minutes and the mean service time is 4.7 minutes.
For one counter,

$$\rho_1 = \frac{4.7}{1(3.5)} \approx 1.34 > 1.$$

The counter lacks sufficient average capacity. A long simulation will show increasing waits, but the key conclusion is structural: one counter cannot keep up under these assumptions.
For two counters,

$$\rho_2 = \frac{4.7}{2(3.5)} \approx 0.67 < 1.$$

A stable system is possible, although waiting-time variability still requires simulation. Capacity checks should accompany, not replace, detailed simulation.

Model criticism: Do not mistake a simulated average for a universal fact

A result such as "average wait $= 2.17$ minutes" is conditional on the input distributions, queue discipline, staffing policy, operating period, and random sample. A credible report should provide:

1. the assumptions and data source for each random input;
2. the number and length of replications;
3. the mean together with variability or percentiles;
4. sensitivity to uncertain assumptions;
5. verification of the algorithm and validation against real observations.

Simulation study design **[10 min]**

The school wants to compare one counter, two counters, and one counter with a shorter menu.
Write a short experiment plan specifying:

1. the random inputs and how they would be estimated;
2. at least four outputs;
3. the simulation horizon and number of replications;
4. one fair comparison technique using common random seeds;
5. one validation check and one sensitivity test.

Verification and validation

Verification asks, “Did we implement the model correctly?” Trace small cases by hand, check conservation rules, and reproduce results with a fixed seed.
Validation asks, “Does this model represent the real system adequately for its purpose?” Compare simulated and observed queue lengths, waits, utilization, and arrival patterns. A program can be verified but still model the wrong system.

Transient warm-up and replication-level uncertainty

A long-run simulation started from an artificial empty state usually has a **transient**: early observations reflect initialization rather than steady operation. For a steady-state question, choose and justify a warm-up (or burn-in) period b, discard observations before b, and check whether conclusions are stable under nearby choices of b. Do not discard the beginning of a finite-horizon study—for example, one actual lunch period starting with an empty queue—because that initial condition is part of the system being modeled.
For R independent replications, first compute one output Y_r per replication. With replication mean $\bar{Y}$ and sample standard deviation s_Y, an approximate two-sided $100(1-\alpha)\%$ confidence interval is

$$\bar{Y} \pm t_{1-\alpha/2,\,R-1}\frac{s_Y}{\sqrt{R}}.$$

The replication, not each correlated customer or time step within it, is the unit of uncertainty. State the initialization, warm-up rule, run length, number of replications, and interval method together.

6 Lesson summary

Six ideas to retain

1. Simulation combines random inputs with deterministic process rules.
2. Uniform random numbers can be mapped to empirical outcomes through cumulative probabilities.
3. One Monte Carlo run is one possible outcome, not the answer.
4. Typical Monte Carlo error decreases only as $1/\sqrt{n}$.
5. In a queue, event order is captured by $B_i = \max(A_i, D_{i-1})$.
6. Decisions require replications, uncertainty summaries, verification, validation, and sensitivity analysis.

Checkpoint: Exit ticket **[2 min]**

Complete both statements:
“A fixed random seed is useful because ______________________________.”
“A second seed or independent replication is necessary because ______________________________.”

7 Practice problems

Core practice

1. A repair time T has distribution

T (h)	1	2	3	5
$P(T=t)$	0.10	0.40	0.35	0.15.

   Construct the interval mapping and convert $U = 0.04, 0.36, 0.71, 0.92$ into repair times.

2. A quarter-circle simulation produces $C = 7{,}824$ inside points out of $n = 10{,}000$.
   (a) Estimate π.
   (b) If the sample size is multiplied by 25, by what factor should the typical Monte Carlo error change?
   (c) Explain why the actual error need not decrease in every individual run.

3. For a single-server queue, arrivals occur at

$$A = (0, 2, 6, 7, 11),$$

   and service times are

$$S = (3, 5, 2, 4, 3).$$

   Construct a complete table of start times, waits, and departures. Find the average wait, maximum wait, and average time in the system.

4. The mean interarrival time at a help desk is 6 minutes and the mean service time is 8 minutes.
   (a) Calculate ρ for one server and two servers.
   (b) What can be concluded before simulation?
   (c) What important information is still missing?

5. Explain why each statement is inadequate:
   (a) "The simulation was run once and gave an average wait of 4.2 minutes."
   (b) "The program has no coding errors, so the model is valid."
   (c) "The histogram fits last month's data, so it will remain correct all year."
   (d) "The two-counter system has a lower mean wait, so it is automatically the best policy."

6. A team compares two staffing policies using exactly the same sequence of arrivals and service requirements. Explain why this **common random numbers** design can make the comparison clearer. Why should the team still use several seeds?

Practice check and hints

1. The intervals are $[0, 0.10)$, $[0.10, 0.50)$, $[0.50, 0.85)$, and $[0.85, 1)$; the outputs are $(1, 2, 3, 5)$ h.
2. $\widehat{\pi} = 3.1296$. Multiplying n by 25 divides typical error by 5, although sampling variation need not improve monotonically in individual runs.
3. (B_i, W_i, D_i) equals $(0, 0, 3)$, $(3, 1, 8)$, $(8, 2, 10)$, $(10, 3, 14)$, $(14, 3, 17)$. Mean wait $= 1.8$ min, maximum wait $= 3$ min, and mean time in system $= 5.2$ min.

4. $\rho_1 = 8/6 \approx 1.33$ and $\rho_2 = 8/12 \approx 0.67$. One server lacks average capacity; two may be stable, but input distributions, dependence, queue discipline, and time variation remain relevant.

5. The four statements omit, respectively, replication uncertainty; validation; temporal stability; and cost, fairness, or other objectives.

6. Common random numbers reduce noise in paired policy differences; multiple independent seeds are still needed to estimate uncertainty and avoid a seed-specific conclusion.

Modeling challenge

Prepare a one-page simulation brief for one of the following systems:

1. a school canteen queue;
2. bicycle availability at a sharing station;
3. daily inventory of a popular stationery item;
4. patient arrivals at a small clinic.

Include: purpose, state variables, events, random inputs, input-data plan, process logic, outputs, policy alternatives, replication plan, validation, sensitivity, and one limitation. A complete computer program is not required.

A Optional algorithm: Single-server queue

This appendix is not part of the two-hour core lesson. It translates the event equations into pseudocode.

Queue simulation pseudocode

1. Set previous departure $D \leftarrow 0$ and initialize output lists.
2. For customer $i = 1, \ldots, n$:
   (a) generate or read interarrival and service times;
   (b) update the arrival time A_i;
   (c) set $B_i \leftarrow \max(A_i, D)$;
   (d) set $W_i \leftarrow B_i - A_i$;
   (e) set $D_i \leftarrow B_i + S_i$ and update $D \leftarrow D_i$;
   (f) record waiting time, time in system, queue length, and server state.
3. Compute summary statistics for the replication.
4. Repeat with independent seeds and summarize across replications.

B Optional note: Standard error for the quarter-circle estimator

Let I_j equal 1 when sample point j lies inside the quarter circle and 0 otherwise. Then I_j is Bernoulli with $p = \pi/4$, and

$$\widehat{\pi} = 4\bar{I}.$$

Therefore

$$\mathrm{SE}(\widehat{\pi}) = 4\sqrt{\frac{p(1-p)}{n}} \approx \frac{1.64}{\sqrt{n}}.$$

This is a standard deviation of the estimator, not a guaranteed error bound. For example, increasing n from 100 to $10{,}000$ reduces the standard error by a factor of 10.

Model audit card

Audit before trusting or communicating a model

Purpose	What question or decision does the model support?	**Data**	Which inputs are observed, derived, or simulated?
Assumptions	Which simplifications may change the answer?	**Units**	Are dimensions and scales consistent?
Verification	Were equations, code, constraints, and limiting cases checked?	**Validation**	Is performance tested against independent evidence?
Uncertainty	How do measurement, parameter, solver, and structural errors affect conclusions?	**Ethics**	Who benefits, who bears risk, and what fairness criterion matters?
Reproducibility	Are data provenance, versions, seeds, and transformations recorded?	**Communication**	Are claims conditional, traceable, and decision-relevant?

Further reading

- Discrete-event simulation documentation (SimPy)
- Engineering Statistics Handbook (NIST / SEMATECH)

Sources, provenance, and licence

These notes follow the iterative modeling and assessment perspective in the SIAM/COMAP GAIMME report. Competition-specific remarks are a snapshot checked on 20 August 2026 against the official COMAP HiMCM instructions; readers should verify current rules before submission. Unless a source is explicitly identified, classroom datasets and numerical scenarios are synthetic or simulated teaching examples and are not empirical claims about a real institution. Third-party references retain their own terms.

Student response guide and checking criteria

For closed calculations, show the definition, substitution, units, and at least one independent check; the worked examples in this unit are method hints, but you must complete the arithmetic yourself. Open modeling responses are checked on five dimensions: a clear purpose and output; stated data and provenance; defensible assumptions and scope; traceable computational verification and model validation; and alignment among uncertainty, limitations, and the recommendation. Support each dimension with concrete evidence rather than writing only “completed.”

Lectures in Mathematical Modeling

Lecture 6: Linear Programming and Decision Optimization

Kenneth Sok Kin Cheng

Revised 20 August 2026

Learning goals

By the end of the lesson, readers should be able to:

1. identify **decision variables**, an **objective function**, and **constraints** from a real situation;
2. formulate a two-variable **linear program**;
3. graph the **feasible region** and locate its **extreme points**;
4. use corner evaluation and objective level lines to solve both maximization and minimization models;
5. recognise infeasible, unbounded, and multiple-optimum cases from geometry;
6. explain why integrality, linearity, certainty, and divisibility are modeling assumptions rather than facts;
7. formulate a small **binary program** with budget and logical constraints;
8. interpret **binding constraints**, **slack**, **shadow prices**, and a **range of optimality** in context.

1 From a situation to a decision model

Optimization begins with a decision, not an algorithm. A modeler must first decide what can be controlled, what outcome matters, and what limits the possible choices. A general constrained optimization model has the form

$$\text{optimize } f(\mathbf{x}) \qquad \text{subject to} \qquad g_i(\mathbf{x}) \leq b_i.$$

When the objective and every constraint are linear, the model is a **linear program**.

Launch: Planning a production week **[10 min]**

A school workshop makes display stands and storage boxes for an exhibition. Each product earns a different contribution, but both use limited wood and labor.
In pairs, decide:

1. what the decision variables should represent;
2. what quantity should be maximized;
3. which resource limits must become constraints;
4. whether fractional answers such as 12.4 boxes are meaningful.

Modeling lesson: The mathematical structure should reflect the decision that must actually be implemented.

Four ingredients of an optimization model

1. **Decision variables:** quantities that the decision maker controls.
2. **Objective function:** a numerical measure of success to maximize or minimize.
3. **Constraints:** resource, policy, physical, or logical restrictions.
4. **Domain restrictions:** nonnegative, integer, binary, or bounded values required by context.

A model is incomplete if any of these ingredients is left implicit.

Worked Example 1: Formulating the carpenter's problem

A carpenter earns a contribution of $25 per table and $30 per bookcase. A table requires 20 board-feet of lumber and 5 labor hours; a bookcase requires 30 board-feet and 4 labor hours. Each week, 690 board-feet and 120 labor hours are available.
Let

$$x = \text{number of tables}, \qquad y = \text{number of bookcases}.$$

The objective is weekly contribution

$$\boxed{\text{maximize } P = 25x + 30y.}$$

The resource constraints are

$$\begin{aligned} 20x + 30y &\le 690 && \text{(lumber)},\\ 5x + 4y &\le 120 && \text{(labor)},\\ x, y &\ge 0 && \text{(nonnegativity)}. \end{aligned}$$

If only whole products can be made, the additional restrictions $x, y \in \mathbb{Z}$ should be stated. We first solve the continuous model, then examine whether the resulting decision is implementable.

Checkpoint: Formulation

A bakery makes regular cakes and premium cakes. Contributions are $18 and $26. Each regular cake requires 2 hours of preparation and 1 hour of decoration; each premium cake requires 3 preparation hours and 2 decoration hours. Weekly capacities are 36 preparation hours and 20 decoration hours. Define the variables and write the complete linear program. Include units and domain restrictions.

2 The feasible region

Each inequality describes a half-plane. The **feasible region** is the set of points satisfying all constraints simultaneously. A feasible point is an implementable plan under the assumptions of the model; an infeasible point violates at least one restriction.

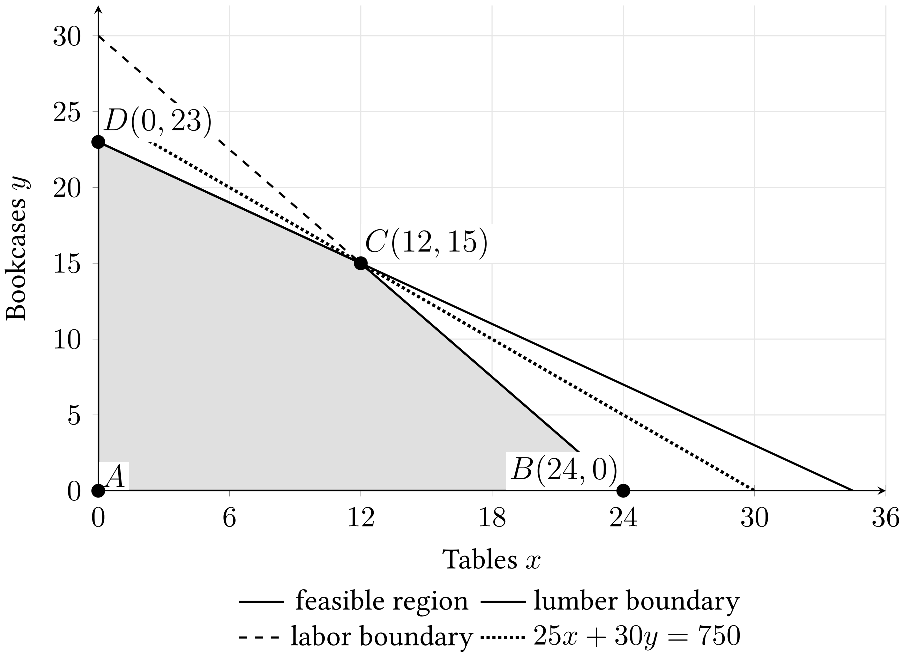


Figure 1: The feasible region is the overlap of all constraints. The dotted objective line is the highest parallel level line that still touches the feasible region.

A set is **convex** if the line segment joining any two points in the set remains inside the set. Intersections of linear half-planes are convex, so every linear-programming feasible region is convex.

Why do corners matter?

A linear objective has parallel level lines. Moving a level line in the improving direction, the last contact with a bounded convex polygon occurs at a corner, or along an entire edge when there are multiple optimal solutions.

Therefore, for a two-variable linear program, it is enough to evaluate the objective at the feasible extreme points. This conclusion depends on linearity; it is not generally true for nonlinear objectives.

Worked Example 2: Solving by extreme points

The feasible extreme points are

$$A = (0, 0), \quad B = (24, 0), \quad C = (12, 15), \quad D = (0, 23).$$

Evaluate $P = 25x + 30y$:

point	P (dollars)
A	0
B	600
C	750
D	690

Hence the continuous optimum is

$$\boxed{x = 12, \qquad y = 15, \qquad P = 750.}$$

Both resource constraints are equalities at C, so all lumber and all labor are used.

Guided graphical solution **[15 min]**

For the bakery model from the checkpoint:

1. draw each boundary line using intercepts;
2. identify the feasible side of each line;
3. list all feasible extreme points;
4. evaluate the objective at each extreme point;
5. state the optimal plan in a complete sentence.

3 Minimization and minimum requirements

Not every optimization problem seeks the largest profit. A school may minimize staffing cost, a transport company may minimize fuel use, or a diet model may minimize cost while meeting minimum nutritional requirements. Constraints of the form "at least" usually produce $\geq$ inequalities, so the feasible region may lie *above* the boundary lines and may not contain the origin.

Worked Example 3: Minimum-cost delivery fleet

A delivery service can hire trucks and vans for one day. A truck costs \$120 and carries 8 pallet units; a van costs \$70 and carries 4 pallet units. At least 48 pallet units must be carried, and at least 8 vehicles must be available for route coverage.
Let x be the number of trucks and y the number of vans. The continuous model is

$$\text{minimize } C = 120x + 70y \quad \text{subject to} \quad 8x + 4y \geq 48, \qquad x + y \geq 8, \qquad x, y \geq 0.$$

Dividing the first constraint by 4 gives $2x + y \geq 12$. The relevant lower-bound extreme points are

$$(0, 12), \qquad (4, 4), \qquad (8, 0).$$

Their costs are \$840, \$760, and \$960 respectively. Hence the continuous optimum is

$$\boxed{x = 4, \qquad y = 4, \qquad C = 760.}$$

Here the objective line is moved *toward smaller cost* until it first touches the feasible region.

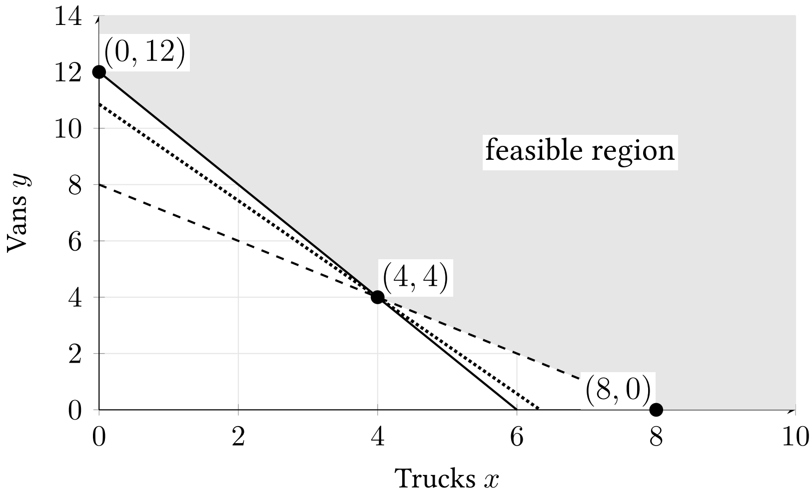


Figure 2: For minimum requirements, the feasible region lies above the constraints. The minimum-cost level line first touches the region at $(4, 4)$.

Checkpoint: Maximize or minimize?

For each situation, identify a plausible objective and state whether it should be maximized or minimized:

1. scheduling revision sessions while meeting minimum subject coverage;
2. allocating an advertising budget to increase expected registrations;
3. choosing emergency supplies while keeping total weight as small as possible.

4 Three special outcomes

A correctly formulated linear program does not always have one finite optimal corner. The geometry should be checked before an answer is reported.

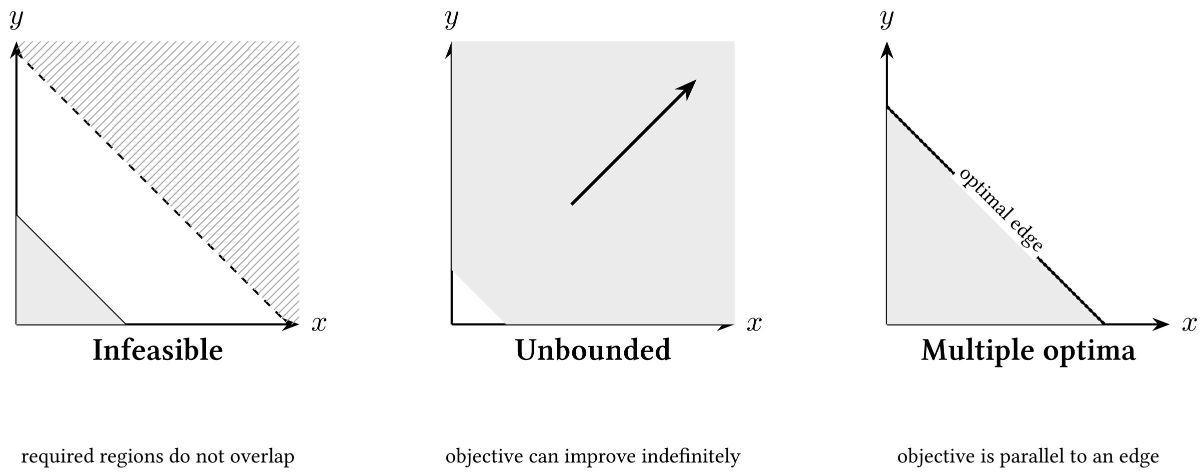


Figure 3: Before interpreting a solution, check whether the model is infeasible, unbounded, or has an entire edge of optimal solutions.

How to report special cases

- **Infeasible**: the requirements cannot all be satisfied. Recheck data or negotiate constraints.
- **Unbounded**: the model permits indefinite improvement. A missing upper limit, cost, or capacity is often the cause.
- **Multiple optima**: several decisions are mathematically equivalent, so secondary criteria such as risk, fairness, or ease of implementation can decide among them.

Model criticism: A feasible point is not automatically a sensible decision

The mathematics checks only the stated constraints. A feasible production plan may still be unacceptable because the model omitted demand limits, storage, contractual minimums, setup costs, uncertainty, or fairness. Optimization does not replace judgment; it makes the stated trade-offs explicit.

5 Binding constraints, slack, and what-if questions

At a proposed solution, a constraint is **binding** if it holds with equality. Otherwise it has **slack**. For a resource constraint

$$a_1x + a_2y \le b,$$

the slack is $b - (a_1x + a_2y)$.

Worked Example 4: Interpreting slack

At $B = (24, 0)$:

$$\begin{aligned} \text{lumber used} &= 20(24) = 480, & \text{lumber slack} &= 690 - 480 = 210,\\ \text{labor used} &= 5(24) = 120, & \text{labor slack} &= 120 - 120 = 0. \end{aligned}$$

Labor is binding, while 210 board-feet remain unused. Adding lumber alone cannot improve this plan locally because labor is already exhausted.
At the optimum $C = (12, 15)$, both slacks are zero. Both resources are bottlenecks under the current coefficients.

A **shadow price** estimates the increase in the optimal objective value from one additional unit of a binding resource, provided the change is small enough that the same pair of constraints remains active.

Worked Example 5: Marginal value of resources

At the carpenter optimum, the two binding constraints are

$$20x + 30y = b_L, \qquad 5x + 4y = b_H.$$

The shadow prices u and v for lumber and labor satisfy

$$20u + 5v = 25, \qquad 30u + 4v = 30,$$

because the value assigned to the resources used by one product must match its contribution at the margin. Solving gives

$$\boxed{u = \frac{5}{7} \approx 0.71 \text{ dollars per board-foot}}, \qquad \boxed{v = \frac{15}{7} \approx 2.14 \text{ dollars per labor hour}}.$$

Thus one extra labor hour is locally more valuable than one extra board-foot. Paying \$3 for one extra labor hour would not be worthwhile under this local model, while paying \$1 might be.

Worked Example 6: Range of optimality for table profit

Suppose the table contribution changes from \$25 to c_T, while bookcase contribution remains \$30. The objective level lines have slope

$$-\frac{c_T}{30}.$$

At the current optimum $C = (12, 15)$, the slopes of the two binding edges are $-5/4$ (labor) and $-2/3$ (lumber). The same corner remains optimal while the objective slope lies between them:

$$-\frac{5}{4} \le -\frac{c_T}{30} \le -\frac{2}{3}.$$

Reversing signs and solving gives

$$\boxed{20 \le c_T \le 37.5.}$$

Within this interval, the recommended production quantities remain $(12, 15)$ even though total contribution changes. At an endpoint, the objective line is parallel to one edge and multiple optimal solutions appear. Outside the interval, a different corner becomes optimal.

Sensitivity studio **[8 min]**

Use slopes rather than recomputing every objective value.

1. If table contribution rises to \$34, does the production mix change?
2. If it rises to \$40, which neighboring corner should become preferable?
3. Why is a **range of optimality** more useful than testing only a single alternative coefficient?

Checkpoint: Slack and marginal value

A solution leaves 40 kg of material unused but uses every available labor hour.

1. Which constraint is binding?
2. Which resource should have zero shadow price locally?
3. Why does "buy the resource with the larger shadow price" still require a validity range and a cost comparison?

Model criticism: Sensitivity is local

A shadow price is not a universal market value. If a resource changes enough, the optimal corner may change and the old shadow price no longer applies. Sensitivity analysis should report both the marginal value and the range over which the interpretation remains valid.

6 When whole-number decisions matter

A continuous linear program allows fractional values. This is appropriate for divisible quantities such as kilograms, hours, or proportions. It may be inappropriate for people, vehicles, projects, or indivisible products. Adding $x_j \in \mathbb{Z}$ gives an **integer program**; restricting $x_j \in \{0, 1\}$ gives a **binary program**.

Worked Example 7: Why rounding can fail

Consider

$$\text{maximize } 7x + 10y \quad \text{subject to} \quad 3x + 5y \le 14, \qquad x, y \ge 0,$$

where x and y count indivisible packages.
The continuous model prefers $(14/3, 0)$, with objective $98/3 \approx 32.67$. Rounding x to 5 violates the capacity constraint because $3(5) = 15 > 14$. Rounding down to $(4, 0)$ gives value 28, but the integer-feasible point $(3, 1)$ uses all capacity and gives value 31.
Conclusion: Solving continuously and rounding afterward can produce either an infeasible or a suboptimal decision.

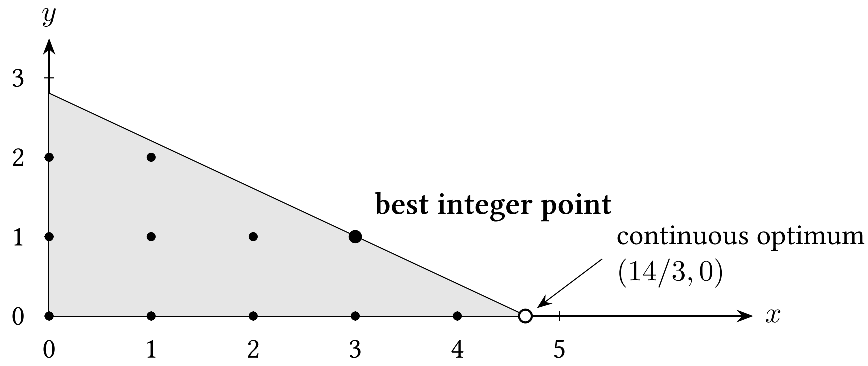


Figure 4: Integer feasibility replaces a continuous region by a set of lattice points. The best integer point need not be obtained by rounding the continuous optimum.

Worked Example 8: Binary project selection

A student council considers five projects. Costs and estimated benefit scores are:

Project	Cost	Benefit score
A: Study hub	4	8
B: Booking system	5	12
C: Peer mentoring	3	7
D: Wellbeing event	4	9
E: Equipment fund	6	11

Let $z_A, \ldots, z_E \in \{0, 1\}$ indicate selection. The council has budget 12, may choose at most three projects, Project B requires A, and C and D cannot both be chosen:

$$\text{maximize } B = 8z_A + 12z_B + 7z_C + 9z_D + 11z_E,$$

$$\begin{aligned} 4z_A + 5z_B + 3z_C + 4z_D + 6z_E &\leq 12, \\ z_A + z_B + z_C + z_D + z_E &\leq 3, \\ z_B &\leq z_A, \\ z_C + z_D &\leq 1. \end{aligned}$$

For this small problem, feasible combinations can be enumerated. The strongest candidates are

selection	cost	benefit
$A + B + C$	12	27
$D + E$	10	20
$A + E$	10	19
$C + E$	9	18
$A + D$	8	17

Hence the optimal selection is $A + B + C$, with benefit score 27.

Differentiated optimization studio **[15 min]**

Choose one route.
Core route: Verify every constraint for the proposed choice $A + B + C$, then explain why choosing B alone is prohibited.
Challenge route: Suppose the budget falls to 10. Find the new optimal combination by systematic enumeration and report which logical constraint becomes important.
Extension route: Add a condition "exactly one of D and E must be selected" and write the corresponding binary equation. Discuss whether the original optimum remains feasible.

7 Assumptions and communication

The standard linear-programming assumptions are useful but demanding:

- **Proportionality:** doubling an activity doubles its resource use and contribution.
- **Additivity:** total effects are sums; there are no interaction or setup effects.
- **Divisibility:** continuous variables may take fractional values.
- **Certainty:** all coefficients are treated as known constants.

Model criticism: When linearity becomes questionable

Bulk discounts, overtime premiums, fixed setup costs, uncertain demand, product interactions, congestion, and economies of scale all violate simple linearity. A useful first model may still be linear, but its assumptions and intended range must be stated explicitly.

Competition-style conclusion **[8 min]**

Write a four-sentence recommendation for the carpenter:

1. report the optimal production plan and contribution;
2. identify the binding resources;
3. state one useful sensitivity result;
4. acknowledge one assumption that could change the decision.

Avoid writing that the model "proves" the decision is best in reality.

Optional extension: Modern audit: multiple objectives, fairness, and stakeholders

Many public decisions cannot be represented honestly by one profit score. A transparent first model may use a weighted objective

$$\max\ \lambda P(x) + (1-\lambda)B(x), \qquad 0 \le \lambda \le 1,$$

where P measures efficiency and B measures a second stated goal, such as access or service quality. The weight λ is a value judgment, not a measured law. Vary it, display the resulting trade-off, and identify decisions that remain acceptable across a defensible range.
An aggregate optimum can conceal unequal burdens. Let $s_g(x)$ denote the service received by stakeholder group g. A fairness floor $s_g(x) \ge s_{\min}$ for every group, or a max–min objective $\max \min_g s_g(x)$, makes one distributive principle explicit. It does not establish that the principle is ethically sufficient. Before recommending a solution, state who defined the objectives, who gains, who bears cost or risk, which groups are absent from the data, and how affected stakeholders can challenge assumptions. Report subgroup outcomes alongside the total objective; do not describe a mathematically feasible

allocation as fair without a stated criterion and stakeholder review.

8 Lesson summary

Nine ideas to retain

1. Optimization translates a decision into variables, an objective, constraints, and domain restrictions.
2. A feasible solution satisfies every stated constraint; feasibility does not guarantee practical suitability.
3. A two-variable linear program can be solved by evaluating feasible extreme points, whether the objective is maximized or minimized.
4. A model may be infeasible, unbounded, or have multiple optimal solutions; each outcome has a different interpretation.
5. Binding constraints identify current bottlenecks; slack measures unused capacity.
6. Shadow prices provide local marginal values, not permanent prices.
7. A range of optimality describes how far an objective coefficient can change before the recommended corner changes.
8. Whole-number or yes/no decisions require integer or binary variables; naive rounding is unsafe.
9. Logical requirements such as dependence and incompatibility can be written as linear constraints on binary variables.
10. The optimal answer is conditional on assumptions about linearity, divisibility, certainty, and omitted factors.

Checkpoint: Exit ticket **[2 min]**

Complete both statements:
“The mathematical optimum is ________________________.”
“Before implementing it, I would check _____________________________.”

9 Practice problems

Core practice

1. A farmer plants wheat and corn. Wheat earns \$200 per acre and uses 3 labor-days and 2 fertilizer units. Corn earns \$300 per acre and uses 2 labor-days and 4 fertilizer units. At most 45 acres, 100 labor-days, and 120 fertilizer units are available.
   (a) Define the variables and formulate the linear program.
   (b) Graph the feasible region and find all extreme points.
   (c) Determine the optimal planting plan and identify binding constraints.
2. For the carpenter model, calculate the slack in each resource at A, B, C, and D. Explain why unused lumber at B does not imply that B can be improved by adding more tables.

3. A delivery plan uses trucks x and vans y:

$$\text{minimize } 120x + 70y \quad \text{subject to} \quad 8x + 4y \geq 48, \quad x + y \geq 8, \quad x, y \geq 0.$$

   Interpret every coefficient, graph the feasible region, and find the continuous optimum. Then discuss whether integrality matters.

4. A student claims: "The point with the greatest x-coordinate must maximize every profit function." Explain the error using level lines.

5. For each of the following, sketch a feasible region and explain the outcome: (a) inconsistent constraints, (b) an unbounded maximization problem, and (c) multiple optimal solutions.

6. For the carpenter model, derive the range of bookcase contribution c_B for which $(12, 15)$ remains optimal. Interpret what happens at the endpoints.

7. Re-solve the binary project model when the budget is 10. Show a compact enumeration table rather than listing all 2^5 possibilities.

8. Construct an example in which a linear program has multiple optimal solutions. State the geometric condition that causes this.

9. A report states: "One extra labor hour is worth \$2.14, so 1000 additional hours are worth \$2140." Critique the statement carefully.

Practice check and hints

1. With wheat x and corn y, maximize $200x + 300y$ subject to $x + y \leq 45$, $3x + 2y \leq 100$, $2x + 4y \leq 120$, and nonnegativity. The vertices are $(0, 0)$, $(100/3, 0)$, $(20, 20)$, and $(0, 30)$; $(20, 20)$ is optimal with value 10000, and labor and fertilizer bind.

2. Lumber/labor slacks at A, B, C, D are $(690, 120)$, $(210, 0)$, $(0, 0)$, and $(0, 28)$. At B, labor prevents another table.

3. The delivery optimum is $(4, 4)$ with cost 760; integrality matters in principle but does not change this solution.

4. The maximizing corner depends on the level-line slope, not on the largest coordinate alone.

5. Check whether the feasible set is empty, whether the objective can grow without bound, or whether an objective line is parallel to a binding edge.

6. The bookcase range is $20 \leq c_B \leq 37.5$; an endpoint gives multiple optima along an adjacent edge.

7. With budget 10, $A + B$ and $D + E$ tie at benefit 20; both are feasible and optimal.

8. For example, maximizing $x + y$ subject to $x + y \leq 4$ and nonnegativity makes every point on $x + y = 4$ optimal.

9. The shadow price 2.14 is local to the current optimal basis and resource range; extrapolating it to 1000 hours is unsupported.

Modeling challenge

A school must choose how many standard and intensive revision workshops to offer. Each type uses teacher hours, rooms, and a limited printing budget. Student benefit is estimated differently for the two formats.
Prepare a one-page optimization brief containing:

1. decision purpose and stakeholders;
2. variables with units and domain restrictions;
3. an objective and at least three constraints;
4. assumptions behind linearity and benefit scores;
5. a graphical or computational solution plan;
6. one sensitivity question and one implementation limitation.

The quality of the formulation and interpretation is more important than obtaining a particular numerical optimum.

A Optional extension: What the Simplex Method does

The **Simplex Method** is an algebraic method for larger linear programs. It starts at one feasible extreme point and repeatedly moves to an adjacent extreme point with a better objective value. **Slack variables** convert inequalities into equations, and a **pivot** changes the current basis.

For the carpenter model, Simplex follows the path

$$(0,0) \longrightarrow (0,23) \longrightarrow (12,15),$$

then stops because no adjacent move improves the objective. A full tableau treatment is best taught as a separate lesson after students understand formulation, geometry, special cases, and interpretation.

B Optional extension: Derivative-free search

A one-dimensional nonlinear objective may be optimized by **golden-section search** when it is unimodal on a known interval. This is a different topic from linear programming: it narrows an interval containing one optimum rather than moving among corners of a feasible polytope. It is reserved for a later numerical-optimization lesson.

Model audit card

Audit before trusting or communicating a model

Purpose	What question or decision does the model support?	**Data**	Which inputs are observed, derived, or simulated?
Assumptions	Which simplifications may change the answer?	**Units**	Are dimensions and scales consistent?
Verification	Were equations, code, constraints, and limiting cases checked?	**Validation**	Is performance tested against independent evidence?
Uncertainty	How do measurement, parameter, solver, and structural errors affect conclusions?	**Ethics**	Who benefits, who bears risk, and what fairness criterion matters?
Reproducibility	Are data provenance, versions, seeds, and transformations recorded?	**Communication**	Are claims conditional, traceable, and decision-relevant?

Further reading

- Linear Programming (NEOS Guide)
- GAIMME: Guidelines for Assessment and Instruction in Mathematical Modeling Education (SIAM / COMAP)

Sources, provenance, and licence

These notes follow the iterative modeling and assessment perspective in the SIAM/COMAP GAIMME report. Competition-specific remarks are a snapshot checked on 20 August 2026 against the official COMAP HiMCM instructions; readers should verify current rules before submission. Unless a source is explicitly identified, classroom datasets and numerical scenarios are synthetic or simulated teaching examples and are not empirical claims about a real institution. Third-party references retain their own terms.

Student response guide and checking criteria

For closed calculations, show the definition, substitution, units, and at least one independent check; the worked examples in this unit are method hints, but you must complete the arithmetic yourself. Open modeling responses are checked on five dimensions: a clear purpose and output; stated data and provenance; defensible assumptions and scope; traceable computational verification and model validation; and alignment among uncertainty, limitations, and the recommendation. Support each dimension with concrete evidence rather than writing only “completed.”

Lectures in Mathematical Modeling

Lecture 7: Decision Theory and Risk

Kenneth Sok Kin Cheng

Revised 20 August 2026

Learning goals

By the end of the lesson, readers should be able to:

1. distinguish a **decision alternative** from a **state of nature**;
2. compute and interpret **expected value** for repeated risky decisions;
3. build and solve a simple **decision tree** using the fold-back method;
4. use **Bayes' theorem** to update probabilities after new information;
5. compute the **expected value of perfect information**;
6. apply Laplace, Maximin, Maximax, Hurwicz, and Minimax Regret criteria when probabilities are unknown;
7. explain why the mathematically optimal decision may depend on risk tolerance, repetition, and consequences.

1 Why decision theory?

Previous lectures often assumed that the model inputs were known. In real decisions, some inputs are uncertain: the weather may change, market demand may be high or low, a bid may be accepted or rejected, and a medical test may be correct or misleading. **Decision theory** provides a disciplined way to choose actions when outcomes depend on uncertain future states.

A decision model separates three ingredients:

1. **Alternatives**: actions the decision maker can choose.
2. **States of nature**: uncertain conditions not controlled by the decision maker.
3. **Payoffs**: consequences produced by each alternative-state combination.

Launch: What should the stall sell? **[10 min]**

A school fundraising stall can sell either hot coffee or iced tea. Coffee earns more on a cold day; iced tea earns more on a warm day. Tomorrow's weather is uncertain.
In pairs, discuss:

1. What are the alternatives? What are the states of nature?
2. What information would change the decision?
3. Is this a repeated decision or a one-time decision?
4. Would you rather maximize average profit or avoid a very bad outcome?

Modeling lesson: Before computing anything, identify who chooses, what is uncertain, and what consequence matters.

Decision modeling structure

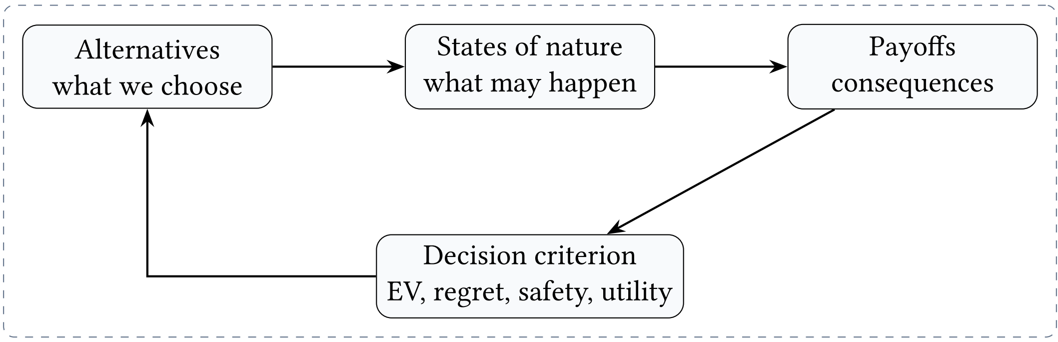


Figure 1: A decision model combines choices, uncertain states, consequences, and a criterion for judging them.

Repeated decisions versus one-time decisions

Expected value is a long-run average. It is most convincing when the same kind of decision is repeated many times, such as an insurance company selling many policies. For a one-time decision with large downside risk, the highest expected value may not be the most acceptable choice.
This distinction is not a minor philosophical issue. It determines whether we should use expected value, a risk-averse criterion, or a regret-based criterion.

2 Expected value for repeated risky decisions

If a random outcome has payoffs $w_1, w_2, \ldots, w_n$ with probabilities $p_1, p_2, \ldots, p_n$, where $\sum p_i = 1$, its **expected value** is

$$E = \sum_{i=1}^{n} p_i w_i.$$

The value E is not necessarily a possible outcome. It is the average payoff expected over many repetitions under the same probability model.

Worked Example 1: A dice game

A game costs \$1 to play. If two fair dice sum to 7, the player receives \$6 total; otherwise the player receives nothing. The net payoff is therefore \$5 for a sum of 7 and $-\$1$ otherwise.
There are 36 equally likely dice outcomes and 6 of them sum to 7, so

$$P(\text{sum} = 7) = \frac{6}{36} = \frac{1}{6}.$$

The expected net payoff is

$$E = 5\left(\frac{1}{6}\right) + (-1)\left(\frac{5}{6}\right) = 0.$$

The game is **fair** in the expected-value sense. This does not mean every player breaks even; it means that the average net gain approaches zero over many plays.

Model criticism: Expected value is not a promise

A student might say, "The expected value is zero, so I will neither win nor lose." This is wrong. A single play gives either \$5 or $-\$1$. Expected value describes the centre of the distribution over repeated trials, not the outcome of one trial.

Worked Example 2: Concession decision with known weather probability

A concession firm must choose one product for tomorrow's school sports day.

	Cold day	**Warm day**
Coffee	\$4000	\$1000
Iced tea	\$1500	\$5000

The forecast gives $P(\text{cold}) = 0.30$, so $P(\text{warm}) = 0.70$.

$$E(\text{coffee}) = 0.30(4000) + 0.70(1000) = 1900.$$

$$E(\text{iced tea}) = 0.30(1500) + 0.70(5000) = 3950.$$

The expected-value decision is to sell iced tea.
The **break-even probability** for cold weather satisfies

$$4000p + 1000(1-p) = 1500p + 5000(1-p).$$

Solving gives $p = \frac{4000}{6500} \approx 0.615$. Coffee becomes preferable only if the probability of a cold day exceeds about 61.5%.

Checkpoint: Expected value

A bid costs \$2000 to prepare. If accepted, it earns a net profit of \$18,000 after subtracting preparation cost. If rejected, the company loses the \$2000 preparation cost.

1. Write the expected value in terms of the acceptance probability p.
2. Find the break-even value of p.
3. Explain whether the company should bid if $p = 0.15$.

3 Decision trees and the fold-back method

A **decision tree** is useful when a decision is followed by one or more uncertain events. We use square nodes for decisions, circular nodes for uncertainty, and terminal values for payoffs.

Fold-back method

Solve a decision tree from right to left:

1. At each chance node, compute the expected value of its branches.
2. At each decision node, keep the branch with the best expected value.
3. Replace the solved node by its value and continue moving left.

This method is called **fold-back**, because we simplify the future part of the tree before choosing the earlier action.

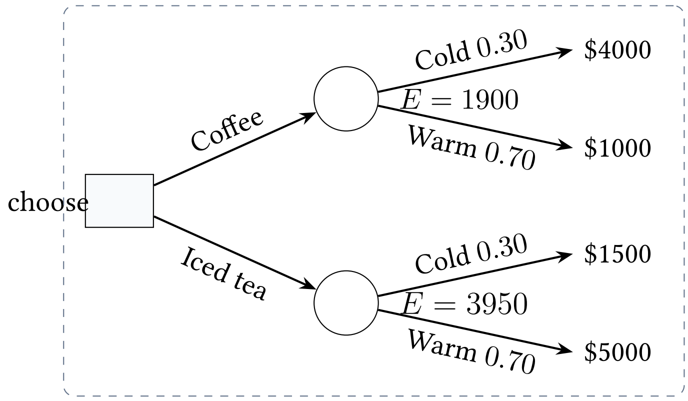


Figure 2: Decision tree for the concession example. Fold-back chooses the branch with the larger expected value.

Worked Example 3: Plant-size decision

A company may build a large plant, build a small plant, or build nothing. Demand may be high, moderate, or low.

Alternative	**High** (0.25)	**Moderate** (0.40)	**Low** (0.35)
Large plant	200000	100000	-120000
Small plant	90000	50000	-20000
No plant	0	0	0

$$E(\text{large}) = 0.25(200000) + 0.40(100000) + 0.35(-120000) = 48000.$$

$$E(\text{small}) = 0.25(90000) + 0.40(50000) + 0.35(-20000) = 35500.$$

$$E(\text{none}) = 0.$$

By expected value, the large plant is best. However, the low-demand loss is $120,000, so the recommendation should mention risk, not only the average.

Model criticism: Risk profile

Two alternatives can have similar expected values but very different downside risk. A complete decision analysis should report:

1. expected value;
2. best-case and worst-case payoff;
3. probability of loss;
4. whether the decision is repeated or one-time;
5. the decision maker's risk tolerance.

4 Information, Bayes' theorem, and EVPI

Information is valuable only if it can change the decision. Decision theory allows us to place an upper bound on the value of information and, when the information is imperfect, to compute whether it is worth buying.

Perfect information

The **expected value of perfect information** is

$$\text{EVPI} = E(\text{best action after knowing the state}) - E(\text{best action now}).$$

It is the most we should ever pay for flawless information. Any imperfect forecast or survey must be worth no more than EVPI.

Worked Example 4: EVPI for the plant-size decision

Without information, the best expected value is $48,000 from the large plant.
With perfect information, the company would choose the best action after seeing the true demand state:

State	Best action	Payoff
High	Large plant	$200000
Moderate	Large plant	$100000
Low	No plant	$0

Therefore,

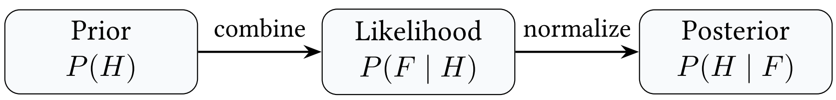
$$E(\text{perfect information}) = 0.25(200000) + 0.40(100000) + 0.35(0) = 90000.$$

$$\boxed{\text{EVPI} = 90000 - 48000 = 42000.}$$

No market research report, even a perfect one, is worth more than $42,000 for this decision.

When information is imperfect, we update probabilities using **Bayes' theorem**. If $H_1, \ldots, H_n$ are possible states and F is evidence, then

$$P(H_i \mid F) = \frac{P(F \mid H_i)P(H_i)}{\sum_j P(F \mid H_j)P(H_j)}.$$



Bayes update: new evidence changes the probability model before the decision is evaluated.

Figure 3: Bayes' theorem updates beliefs from prior probabilities to posterior probabilities using evidence.

Worked Example 5: Market research with Bayes update

For the plant-size problem, suppose a market report is either favorable F or unfavorable U. Its historical accuracy is:

State	$P(F \mid \text{state})$	$P(U \mid \text{state})$
High	0.70	0.30
Moderate	0.50	0.50
Low	0.20	0.80

First compute

$$P(F) = 0.70(0.25) + 0.50(0.40) + 0.20(0.35) = 0.445.$$

The posterior probabilities after a favorable report are

$$P(H \mid F) = \frac{0.70(0.25)}{0.445} \approx 0.393,$$
$$P(M \mid F) = \frac{0.50(0.40)}{0.445} \approx 0.449,$$
$$P(L \mid F) = \frac{0.20(0.35)}{0.445} \approx 0.157.$$

Using the unrounded posterior probabilities, re-evaluating the large plant after a favorable report gives

$$E(\text{large} \mid F) \approx 104719.$$

Small plant gives about $54,719, so a favorable report supports the large plant.
For an unfavorable report, $P(U) = 0.555$ and

$$P(H \mid U) \approx 0.135, \quad P(M \mid U) \approx 0.360, \quad P(L \mid U) \approx 0.505.$$

Using the unrounded posterior probabilities,

$$E(\text{large} \mid U) \approx 2523, \qquad E(\text{small} \mid U) \approx 20090,$$

so an unfavorable report supports the small plant. The report is useful because it changes the decision in some cases.

Checkpoint: Value of information

In the market-research example, after a favorable report the decision is large plant, and after an unfavorable report the decision is small plant.

1. Compute the expected value if the company uses the report: $0.445E_F + 0.555E_U$.
2. Subtract the best no-report expected value $48,000 to obtain EVSI.
3. Should the company buy the report if it costs $10,000? What if it costs $25,000?

5 When probabilities are unknown

Sometimes we cannot assign credible probabilities. In that case, expected value may give a false impression of precision. We use alternative criteria that represent different attitudes toward uncertainty.

Running Example: Investment payoffs

A student club has funds to invest for one year. Payoffs are in thousand dollars.

Alternative	Fast growth	Normal growth	Slow growth
Stocks	10	6.5	-4
Bonds	8	6	1
Savings	5	5	5

No reliable probabilities are available for the three economic states.

Criterion	Rule	Decision in example
Laplace	Treat all states as equally likely and average each row.	Bonds/Savings tie at 5.0.
Maximin	For each row, find the worst payoff; choose the best worst payoff.	Savings.
Maximax	For each row, find the best payoff; choose the best best payoff.	Stocks.
Hurwicz	Use $\alpha(\text{best}) + (1-\alpha)(\text{worst})$.	Depends on α; at $\alpha = 0.6$, Bonds.
Minimax regret	Build a regret matrix and minimize the largest regret.	Bonds.

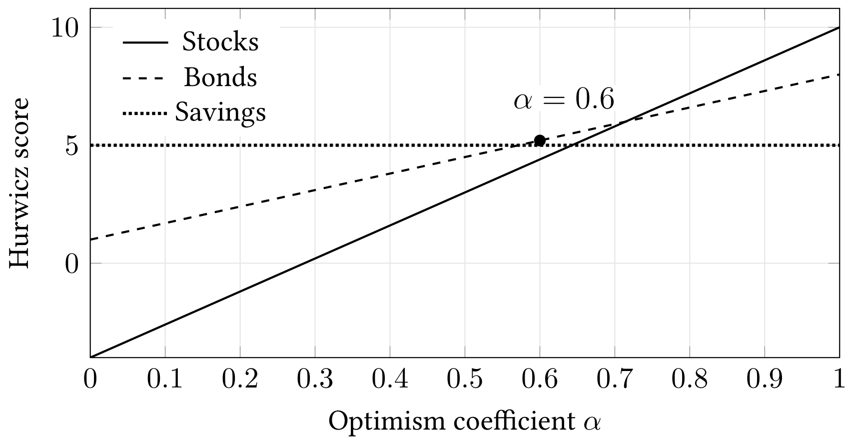


Figure 4: Hurwicz scores change with the optimism coefficient. The chosen alternative depends on the decision maker's attitude toward risk.

Worked Example 6: Minimax regret

For each state, subtract each payoff from the best payoff in that column.

	Fast	Normal	Slow	Max regret
Stocks	$10-10=0$	$6.5-6.5=0$	$5-(-4)=9$	9
Bonds	$10-8=2$	$6.5-6=0.5$	$5-1=4$	4
Savings	$10-5=5$	$6.5-5=1.5$	$5-5=0$	5

The smallest maximum regret is 4, so the minimax-regret decision is Bonds. This criterion asks: "If the true state is later revealed, which decision limits how much we will regret not having chosen the best action for that state?"

Decision studio **[15 min]**

Use the investment example.

1. **Core route:** Apply Laplace, Maximin, Maximax, and Minimax Regret. Write one sentence explaining each attitude.
2. **Challenge route:** Find the α values where Hurwicz changes from Savings to Bonds and from Bonds to Stocks.
3. **Extension route:** Suppose the club cannot tolerate any possibility of loss. Which criteria respect

this constraint? Which criterion ignores it?

6 Integrated mini-case: stocking under uncertain demand

Decision theory becomes most useful when the modeler must connect a business rule, uncertain demand, and a recommendation. The following example is intentionally small enough to compute by hand, but it contains the same structure as larger inventory decisions.

Worked Example 7: Seasonal T-shirt stocking

A school club can order 10, 20, or 30 festival T-shirts. Each shirt costs \$50 and sells for \$80. Unsold shirts can be cleared for \$20 each. If demand exceeds stock, each unmet demand unit causes a goodwill loss of \$10. Demand is estimated as:

$$P(D = 10) = 0.30, \qquad P(D = 20) = 0.40, \qquad P(D = 30) = 0.30.$$

For an order quantity Q and demand D, the payoff is

$$80 \min(Q, D) + 20 \max(Q - D, 0) - 50Q - 10 \max(D - Q, 0).$$

This gives the payoff matrix:

Order quantity	$D = 10$	$D = 20$	$D = 30$	**Expected value**
$Q = 10$	300	200	100	$0.30(300) + 0.40(200) + 0.30(100) = 200$
$Q = 20$	0	600	500	$0.30(0) + 0.40(600) + 0.30(500) = 390$
$Q = 30$	-300	300	900	$0.30(-300) + 0.40(300) + 0.30(900) = 300$

By expected value, the best order is $Q = 20$.
Different risk criteria tell a more nuanced story:

- Maximin chooses $Q = 10$, because its worst payoff is still \$100.
- Maximax chooses $Q = 30$, because it can earn \$900 if demand is high.
- Minimax regret chooses $Q = 20$: its maximum regret is \$400, smaller than the maximum regret for $Q = 10$ or $Q = 30$.

Thus the same payoff matrix can support different recommendations depending on whether the club prioritizes average profit, downside protection, or regret control.

Sensitivity check **[8 min]**

In the T-shirt example, suppose the probability of high demand $D = 30$ increases while the probability of low demand $D = 10$ decreases. Without recalculating every criterion, predict which order quantity becomes more attractive. Then compute the expected values when

$$P(D = 10) = 0.20, \qquad P(D = 20) = 0.40, \qquad P(D = 30) = 0.40.$$

Explain whether the decision changes.

Model criticism: Hidden assumptions in payoff matrices

A payoff matrix looks precise, but each number depends on assumptions: selling price, clearance price, lost goodwill cost, and the demand probabilities. If these inputs are estimated weakly, the report should include sensitivity checks rather than present the final order quantity as certain.

7 Writing a defensible decision recommendation

Decision theory does not remove judgment. It makes the judgment explicit. In a modeling report, the recommendation should not simply state the branch with the largest expected value; it should state the assumptions under which the recommendation holds.

Model criticism: Common mistakes in decision models

1. Treating guessed probabilities as if they were measured facts.
2. Using expected value for a one-time catastrophic decision without discussing risk.
3. Forgetting to subtract the cost of information from EVSI.
4. Confusing a decision node with a chance node: the modeler chooses at one; nature decides at the other.
5. Reporting a criterion result without explaining the risk attitude represented by the criterion.

Competition-style conclusion

For the plant-size case, a concise conclusion might be:
Using the prior demand probabilities, the large plant has the highest expected value, $48,000, and is therefore the recommended decision if management is willing to accept a possible $120,000 loss in the low-demand state. Perfect information would be worth at most $42,000, so any market research costing more than this should be rejected immediately. Because the large-plant recommendation is sensitive to the low-demand loss and to the estimated probability of high demand, we also recommend reporting a risk profile and considering the small plant if management is strongly risk-averse.

Optional extension:* *Modern audit: mean performance and tail risk

Expected payoff describes average performance; it does not describe how severe rare losses can be. Write loss as $L = -W$. For a confidence level α, the **value at risk** $\mathrm{VaR}_\alpha(L)$ is an α-quantile of loss, while the **conditional value at risk** summarizes the tail beyond that threshold. For a continuous loss distribution,

$$\mathrm{CVaR}_\alpha(L) = E[L \mid L \geq \mathrm{VaR}_\alpha(L)] \,.$$

Thus $\mathrm{CVaR}_{0.95}$ is the mean loss in the worst 5% of cases. In a finite simulation, estimate it by averaging the worst 5% of simulated losses and report the number of tail observations.
Expected loss answers "What happens on average?"; CVaR answers "How bad are outcomes after we enter the adverse tail?" They are complementary, not interchangeable. A defensible comparison reports expected payoff, probability of loss, a stated tail metric, and sensitivity to uncertain probabilities. CVaR is itself uncertain when extreme observations are scarce, so use confidence intervals or repeated simulation and do not present a noisy tail estimate as a guaranteed safety bound.

Lesson summary

1. A decision model separates alternatives, states of nature, probabilities, and payoffs.
2. Expected value is a long-run average and is most appropriate for repeated decisions.
3. Decision trees are solved by fold-back: expected values at chance nodes, best choices at decision nodes.
4. Bayes' theorem updates probabilities after evidence.
5. EVPI is an upper bound on the value of information; EVSI must be compared with the cost of the information.
6. When probabilities are unknown, different criteria represent different attitudes toward risk and uncertainty.

Checkpoint: Exit ticket **[2 min]**

Complete the sentence:
"Expected value is a useful decision criterion when ______________________, but it may be misleading when ______________________."

8 Practice problems

Core practice

1. A game costs \$2 to play. A player wins \$12 with probability 0.20 and wins nothing otherwise. Compute the expected net payoff. Is the game fair?
2. A food stall must choose noodles or sandwiches. Profits are:

	Rainy ($p = 0.40$)	Sunny ($p = 0.60$)
Noodles	3200	1600
Sandwiches	1800	3600

   Compute both expected values and find the break-even probability of rain.
3. For the plant-size problem in Worked Example 3, compute the probability of losing money under each alternative. Does this change how you would present the recommendation?
4. A medical screening test has $P(+ \mid D) = 0.95$, $P(+ \mid \text{no } D) = 0.08$, and disease prevalence $P(D) = 0.02$. Compute $P(D \mid +)$. Explain why a highly sensitive test can still have a moderate positive predictive value.
5. For the investment payoff matrix, apply Hurwicz with $\alpha = 0.3$ and $\alpha = 0.8$. Explain why the decision changes.
6. For the following payoff matrix, construct the regret matrix and apply minimax regret:

	S_1	S_2	S_3
A	20	5	8
B	12	12	12
C	4	18	10

Practice check and hints

1. The expected net payoff is $0.20(10) + 0.80(-2) = 0.40$ dollars, so the game is not fair to the organizer in the expected-value sense.
2. Expected profits are 2240 for noodles and 2880 for sandwiches. The break-even rain probability is $10/17 \approx 0.588$.
3. Loss probabilities are 0.35, 0.35, and 0 for the large, small, and no-plant choices; loss magnitude must also be reported.
4. $P(D \mid +) = 0.019/(0.019 + 0.0784) \approx 0.195$; low prevalence leaves many false positives relative to true positives.
5. With $\alpha = 0.3$, the Hurwicz scores for stocks, bonds, and savings are 0.2, 3.1, and 5, so savings wins. With $\alpha = 0.8$, they are 7.2, 6.6, and 5, so stocks win.
6. The regret rows are $A : (0, 13, 4)$, $B : (8, 6, 0)$, and $C : (16, 0, 2)$; maximum regrets are 13, 8, and 16, so minimax regret selects B.

Modeling challenge

Choose one of the following and prepare a one-page decision analysis brief containing: alternatives, states of nature, payoff table, probability assumptions if any, chosen criterion, recommendation, sensitivity comment, and one limitation.

1. A school club deciding whether to hold an outdoor fundraising event, indoor event, or cancel.
2. A student team deciding whether to submit a safe project, a risky innovative project, or a hybrid project to a modeling competition.
3. A small shop deciding whether to stock 10, 20, or 30 units of a seasonal item.

Your answer should explain why the chosen criterion is appropriate for the situation.

A Optional extension: expected utility

Expected value treats each dollar equally. In reality, the pain of losing \$100,000 may be more than twice the pain of losing \$50,000. **Expected utility** replaces payoff w by a utility function $u(w)$ and chooses the alternative with largest

$$E[u(W)] = \sum_i p_i u(w_i).$$

A concave utility function represents risk aversion: gaining an extra dollar matters less when the decision maker is already wealthy, while losses near a critical threshold may be unacceptable. This is why a risk-averse manager may choose the small plant even when the large plant has a larger expected monetary value.

B Optional extension: expected value of sample information

If a report can be favorable or unfavorable, the expected value with sample information is

$$E(\text{with report}) = P(F)\max_a E(a \mid F) + P(U)\max_a E(a \mid U).$$

Then

$$\text{EVSI} = E(\text{with report}) - E(\text{best decision without report}).$$

The report should be purchased only when EVSI exceeds its cost. If the cost is included in terminal payoffs, do not subtract it a second time.

Model audit card

Audit before trusting or communicating a model

Purpose	What question or decision does the model support?	**Data**	Which inputs are observed, derived, or simulated?
Assumptions	Which simplifications may change the answer?	**Units**	Are dimensions and scales consistent?
Verification	Were equations, code, constraints, and limiting cases checked?	**Validation**	Is performance tested against independent evidence?
Uncertainty	How do measurement, parameter, solver, and structural errors affect conclusions?	**Ethics**	Who benefits, who bears risk, and what fairness criterion matters?
Reproducibility	Are data provenance, versions, seeds, and transformations recorded?	**Communication**	Are claims conditional, traceable, and decision-relevant?

Further reading

- Introductory Statistics 2e (OpenStax)
- GAIMME: Guidelines for Assessment and Instruction in Mathematical Modeling Education (SIAM / COMAP)

Sources, provenance, and licence

These notes follow the iterative modeling and assessment perspective in the SIAM/COMAP GAIMME report. Competition-specific remarks are a snapshot checked on 20 August 2026 against the official COMAP HiMCM instructions; readers should verify current rules before submission. Unless a source is explicitly identified, classroom datasets and numerical scenarios are synthetic or simulated teaching examples and are not empirical claims about a real institution. Third-party references retain their own terms.

Student response guide and checking criteria

For closed calculations, show the definition, substitution, units, and at least one independent check; the worked examples in this unit are method hints, but you must complete the arithmetic yourself. Open modeling responses are checked on five dimensions: a clear purpose and output; stated data and provenance; defensible assumptions and scope; traceable computational verification and model validation; and alignment among uncertainty, limitations, and the recommendation. Support each dimension with concrete evidence rather than writing only “completed.”

Lectures in Mathematical Modeling
Lecture 8: Game Theory and Strategic Modeling

Kenneth Sok Kin Cheng

Revised 20 August 2026

Learning goals

By the end of the lesson, readers should be able to:

1. explain why strategic models differ from ordinary decision models;
2. identify players, strategies, payoffs, and assumptions in a game;
3. find a pure-strategy solution using **maximin**, **minimax**, and movement diagrams;
4. solve a 2×2 zero-sum game without a saddle point using mixed strategies;
5. interpret a game value as a guaranteed long-run payoff, not as a promise in one play;
6. analyse a partial-conflict game using **Nash equilibrium** and explain the Prisoner's Dilemma;
7. decide when game theory is an appropriate modeling tool and state its limitations.

1 Why game theory?

In Lecture 7, a decision maker faced uncertain states of nature: weather, market demand, or a research report. Nature did not try to outsmart the decision maker. In **game theory**, at least one other participant is also making a decision, and that participant may change strategy after thinking about what we will do.

A **game** is a mathematical model of strategic interaction. It must specify:

- the **players**;
- each player's possible **strategies**;
- the **payoff** produced by every combination of strategies;
- what each player knows when making a decision.

Launch: ordinary decision or game? **[8 min]**

For each situation, decide whether a decision-theory model or a game-theory model is more appropriate.

1. A shop chooses how many umbrellas to order before the rainy season.
2. Two restaurants decide whether to offer lunch discounts.
3. A goalkeeper chooses which side to dive during a penalty kick.
4. A school decides whether to hire one or two extra canteen staff.

For the game situations, identify the players and their possible strategies.

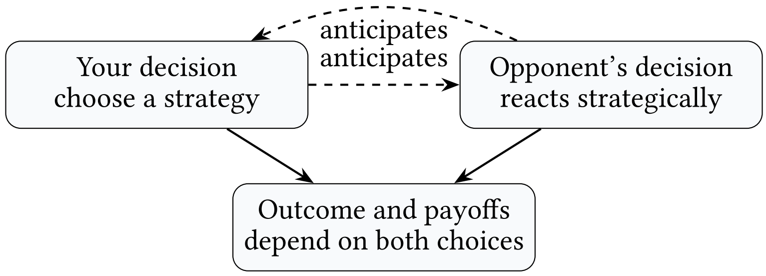


Figure 1: A game is different from an ordinary decision because the other side is also reasoning.

Total conflict and partial conflict

A **total-conflict game**, often called a **zero-sum game**, is one in which one player's gain is exactly the other player's loss. A penalty kick, rock-paper-scissors, or a search-versus-hide problem is often modeled this way.
A **partial-conflict game** allows both players to gain or lose together. Advertising wars, arms races, pricing decisions, and environmental agreements often have this structure.

2 Pure strategies in total-conflict games

In a two-player zero-sum game, we write a payoff matrix from the row player's point of view. The **row player** wants larger numbers; the **column player** wants smaller numbers.

Pure strategies, maximin, minimax

A **pure strategy** is a fixed choice, such as always choosing Row 1. For a payoff matrix $A = (a_{ij})$:

$$\text{row player's maximin} = \max_i \min_j a_{ij}, \qquad \text{column player's minimax} = \min_j \max_i a_{ij}.$$

If these two numbers are equal, the game has a **saddle point**. The corresponding cell is a stable pure-strategy solution: neither player can improve by switching alone.

Worked Example 1: Location competition with a saddle point

Two competing stores choose either a large district L or a small district S. The entry gives Store A's market share percentage.

Store A \ Store B	**B:** L	**B:** S	**Row minimum**
A: L	60	68	60
A: S	52	60	52
Column maximum	60	68	

The row player guarantees at least 60 by choosing L. The column player can hold A to at most 60 by choosing L. Hence

$$\max\{60, 52\} = 60 = \min\{60, 68\}.$$

There is a saddle point at (L, L) with game value $V = 60$. The conclusion is not merely "both choose L"; it is: A can guarantee 60% market share, and B can prevent A from exceeding 60%.

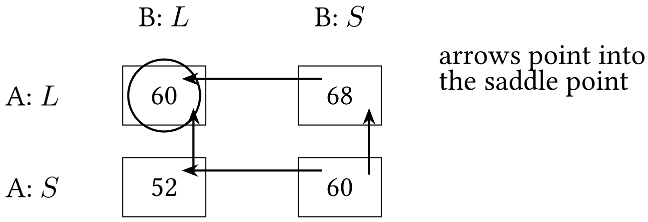


Figure 2: Movement diagram for a pure-strategy saddle point. Vertical arrows show A's preference; horizontal arrows show B's preference.

Checkpoint: Pure strategy check **[5 min]**

For the matrix below, find the row minima, column maxima, and decide whether a saddle point exists.

$$\begin{pmatrix} 4 & 1 \\ 2 & 3 \end{pmatrix}$$

Explain the result in words, not only by calculation.

Worked Example 2: Dominated strategies

Before solving a game, look for strategies that a rational player should never use. A row is **dominated** if another row gives the row player at least as much payoff in every column and strictly more in at least one column. A column is dominated if another column gives the row player at most as much payoff in every row.

Consider the row player's payoff matrix:

	C_1	C_2	C_3
R_1	4	2	5
R_2	3	1	4
R_3	1	6	2

Row R_1 dominates R_2 because $4 > 3$, $2 > 1$, and $5 > 4$. Thus the row player should never choose R_2. After removing it, compare columns C_1, C_2, C_3 from the column player's viewpoint. Since the column player wants smaller payoffs, a column with larger entries is unattractive.

Dominance does not always solve the game, but it simplifies the model and often reveals strategic structure before calculation.

Model criticism: Do not remove strategies too quickly

A dominated strategy can be removed only when dominance is clear across every possible response of the opponent. Removing a strategy because it is "usually bad" is not valid. In modeling reports, state the dominance comparison explicitly.

3 When no pure strategy is stable: mixed strategies

If there is no saddle point, any fixed strategy can be exploited. The solution is to randomize. A **mixed strategy** assigns probabilities to pure strategies.

The central idea is not "being unpredictable" in a vague sense. The mathematical purpose is to make the opponent indifferent between their pure strategies, so that they cannot exploit a predictable bias.

Worked Example 3: Batter versus pitcher

A batter guesses Fastball F or Curve C. The pitcher throws Fastball or Curve. The entries are batting averages.

Batter \ Pitcher	Pitcher: F	Pitcher: C	Row min
Guess F	.400	.200	.200
Guess C	.100	.300	.100
Column max	.400	.300	

The maximin is .200, while the minimax is .300. Since they are unequal, no pure-strategy saddle point exists.
Let x be the probability that the batter guesses Fastball. Then the batter's expected average if the pitcher throws Fastball is

$$E_F(x) = .400x + .100(1 - x) = .100 + .300x.$$

If the pitcher throws Curve,

$$E_C(x) = .200x + .300(1 - x) = .300 - .100x.$$

The batter wants to maximize the lower of these two lines. The best choice occurs where the two lines meet:

$$.100 + .300x = .300 - .100x \quad \Rightarrow \quad x = 0.5.$$

The game value is

$$V = .100 + .300(0.5) = .250.$$

So the batter should guess Fastball with probability 0.5 and Curve with probability 0.5. This guarantees a batting average of .250 in the long run.

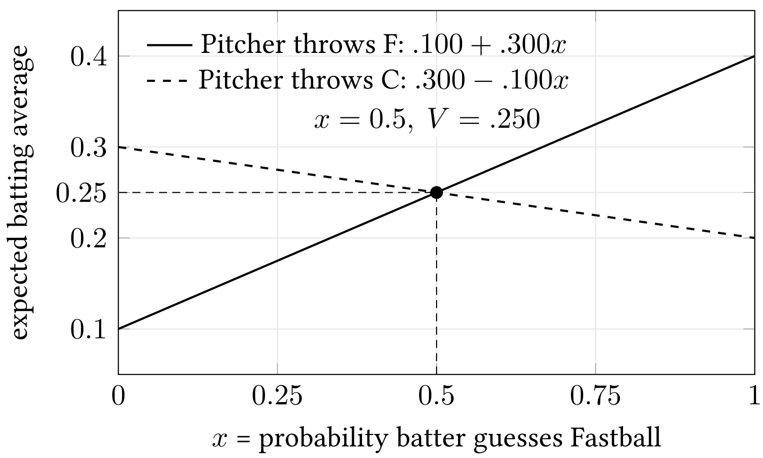


Figure 3: The batter chooses the mixture that maximizes the lower envelope of the two opponent-response lines.

Solving a 2×2 zero-sum game by indifference

For a payoff matrix
$$\begin{pmatrix} a & b \\ c & d \end{pmatrix},$$
let x be the probability that the row player chooses Row 1. To find the row player's optimal mix, equate the column player's two pure responses:

$$ax + c(1-x) = bx + d(1-x).$$

To find the column player's optimal mix, let y be the probability that the column player chooses Column 1, and equate the row player's two pure responses:

$$ay + b(1-y) = cy + d(1-y).$$

This method works after checking that no saddle point exists.

Worked Example 4: Predator and prey

A prey animal chooses Hide or Run. A predator chooses Ambush or Pursue. Entries are the prey's survival probabilities.

Prey \ Predator	Ambush	Pursue
Hide	0.30	0.60
Run	0.50	0.40

There is no saddle point. Let x be the probability the prey Hides. If the predator Ambushes,

$$E_A(x) = 0.30x + 0.50(1-x) = 0.50 - 0.20x.$$

If the predator Pursues,

$$E_P(x) = 0.60x + 0.40(1-x) = 0.40 + 0.20x.$$

Equating them gives
$$0.50 - 0.20x = 0.40 + 0.20x \quad \Rightarrow \quad x = 0.25.$$
The prey should Hide 25% and Run 75%. The guaranteed survival probability is

$$V = 0.50 - 0.20(0.25) = 0.45.$$

For the predator, let y be the probability of Ambush. Equating the prey's two rows:

$$0.30y + 0.60(1-y) = 0.50y + 0.40(1-y) \quad \Rightarrow \quad y = 0.5.$$

The predator should Ambush 50% and Pursue 50%.

Model criticism: What does a mixed strategy mean?

A mixed strategy does not mean that the player is confused. It means that the player deliberately randomizes to prevent exploitation. The result is meaningful when the interaction is repeated many times, or when unpredictability itself has strategic value. In a single high-stakes play, a player may still choose a pure action based on information not represented in the matrix.

4 Games against nature: conservative guarantees

Sometimes the "opponent" is not intelligent: market demand, economy, weather, or machine failure. If probabilities are known, Lecture 7 expected value is appropriate. If probabilities are unknown and the decision maker wants a guaranteed floor, we may treat nature as adversarial.

Worked Example 5: Small or large production

A firm chooses Small or Large production. The economy may be Poor or Good. Payoffs are in thousands of dollars.

Firm \ Economy	Poor	Good	Worst case
Small	500	300	300
Large	100	900	100

A pure maximin decision chooses Small and guarantees 300. But a mixed strategy may guarantee more. Let x be the probability of Small. The guaranteed payoff against Poor is

$$E_P(x) = 500x + 100(1-x) = 100 + 400x.$$

Against Good,

$$E_G(x) = 300x + 900(1-x) = 900 - 600x.$$

Equate them:

$$100 + 400x = 900 - 600x \quad \Rightarrow \quad x = 0.8.$$

The firm should use Small 80% and Large 20%, guaranteeing

$$V = 100 + 400(0.8) = 420.$$

This is better than the pure maximin guarantee of 300.

Checkpoint: Decision theory or game against nature? [5 min]

In the production example, suppose reliable probabilities are available: $P(\text{Poor}) = 0.30$ and $P(\text{Good}) = 0.70$. Should the firm still use the conservative mixed strategy? Compute the expected value of Small and Large and compare with the game-theoretic guarantee.

5 Partial conflict and Nash equilibrium

Zero-sum games are only one type of strategic model. In many real situations, both players may benefit from cooperation, but each also has an incentive to defect.

For partial-conflict games, we write payoffs as ordered pairs (row payoff, column payoff).

Nash equilibrium

A strategy pair is a **Nash equilibrium** if no player can improve their own payoff by switching strategy unilaterally. In a movement diagram, a Nash equilibrium is a cell into which all relevant preference arrows point.

Worked Example 6: Prisoner's Dilemma as an arms race

Two countries choose Disarm or Arm. Payoffs are preference scores, where 4 is best and 1 is worst.

Country A \ Country B	B Disarms	B Arms
A Disarms	$(3,3)$	$(1,4)$
A Arms	$(4,1)$	$(2,2)$

If B disarms, A prefers to arm because $4 > 3$. If B arms, A still prefers to arm because $2 > 1$. Thus Arm is A's dominant strategy. By symmetry, Arm is also B's dominant strategy. The Nash equilibrium is therefore (Arm, Arm) with payoff $(2,2)$.
However, (Disarm, Disarm) gives $(3,3)$, which is better for both. This is the dilemma: individually rational choices lead to a collectively worse outcome.

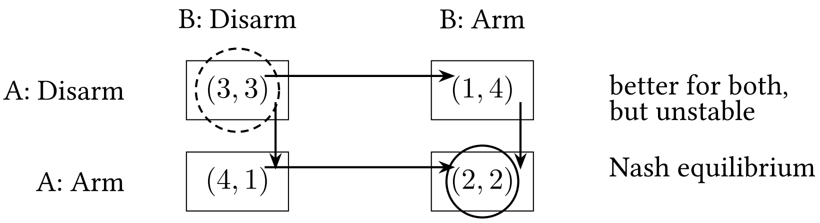


Figure 4: In the Prisoner's Dilemma, the Nash equilibrium is not the socially best outcome.

Model criticism: A common modeling mistake

Do not call every conflict a Prisoner's Dilemma. A true Prisoner's Dilemma requires: each player has a dominant strategy to defect; mutual defection is the Nash equilibrium; mutual cooperation is better for both than mutual defection. Many conflicts are coordination games, chicken games, or bargaining games instead.

6 Differentiated strategic modeling studio

Strategic modeling studio **[20 min]**

Choose one route according to readiness. All groups should finish with a short written recommendation.
Core route: Pricing competition. Two drink stalls choose Low or High prices. The payoffs are weekly profits in hundreds of dollars.

	Stall B Low	Stall B High
Stall A Low	$(6,6)$	$(10,3)$
Stall A High	$(3,10)$	$(8,8)$

Draw a movement diagram and identify any Nash equilibrium. Is the result socially efficient?
Challenge route: Inspection game. A student may Study or Not Study. A teacher may Check or Not Check. From the student's point of view, payoffs are:

	Check	Not Check
Study	4	3
Not Study	-2	5

Treat this as a zero-sum simplification. Find the student's optimal mixed strategy and the value of the game.
Extension route: General formulation. For the matrix

$$\begin{pmatrix} 5 & 2 & 4 \\ 1 & 6 & 3 \end{pmatrix},$$

write the row player's linear program to maximize the guaranteed payoff V. You do not need to solve it, but each constraint must have a clear meaning.

Checkpoint: Competition-style communication **[8 min]**

Write a four-sentence conclusion for one game in this lecture:

1. identify the recommended strategy;
2. state the game value or equilibrium outcome;
3. explain the strategic reason behind the recommendation;
4. state one limitation of the model.

Avoid writing only "we choose Row 1" without explaining what the opponent is expected to do.

7 Lesson summary

Eight ideas to retain

1. Game theory models decisions in which other rational players react.
2. A payoff matrix must state whose payoff is being recorded.
3. In a zero-sum game, the row player maximizes and the column player minimizes.
4. A pure-strategy saddle point exists when maximin equals minimax.
5. Without a saddle point, mixed strategies prevent exploitation.
6. In a 2×2 zero-sum game, optimal mixing is found by making the opponent indifferent.
7. In partial-conflict games, Nash equilibria can be stable but inefficient.
8. Game theory gives strategic insight only if the payoff matrix and rationality assumptions are credible.

Checkpoint: Exit ticket **[2 min]**

Complete the sentence:
"A game-theory model is useful when ______________________, but it may fail when ______________________."

8 Practice problems

Core practice

1. For each zero-sum matrix, find row minima, column maxima, and decide whether a saddle point exists.

$$A = \begin{pmatrix} 5 & 2 \\ 3 & 6 \end{pmatrix}, \qquad B = \begin{pmatrix} 2 & 5 \\ 3 & 2 \end{pmatrix}.$$

2. A goalkeeper dives Left or Right. A kicker shoots Left or Right. Entries are scoring probabilities:

	Goalkeeper L	Goalkeeper R
Kick L	0.55	0.90
Kick R	0.85	0.60

Check for a saddle point. If none exists, find the kicker's optimal mixed strategy and the game value.

3. A firm faces uncertain demand and chooses Conservative or Aggressive production. Payoffs are in thousands:

	Low demand	High demand
Conservative	40	60
Aggressive	10	100

Treat demand as nature. Find the best pure maximin decision. Then find whether a mixed strategy can guarantee more.

4. Draw the movement diagram for the pricing game in the studio. Is there a dominant strategy for either stall? Is the Nash equilibrium Pareto efficient?

5. In the Prisoner's Dilemma matrix from this lecture, explain why $(\text{Disarm}, \text{Disarm})$ is not a Nash equilibrium even though it is better for both countries.

6. Give one real-world situation that should *not* be modeled as zero-sum. Explain what would be lost if it were forced into a zero-sum payoff matrix.

Modeling challenge

Choose one scenario and prepare a one-page strategic modeling brief containing players, strategies, payoff table, assumptions, equilibrium or recommended strategy, and limitations.

1. Two schools decide whether to host their open day on the same weekend or different weekends.
2. Two delivery companies choose whether to enter a new district.
3. A student team chooses a competition topic while anticipating what other teams may choose.

Your brief should justify the payoff values qualitatively even if exact numerical data are unavailable.

9 Additional classroom problem bank

Extension practice

1. **Battle of the search routes.** A rescue team searches North or South. A lost hiker chooses, unknowingly, one of two paths. The payoff to the rescue team is probability of finding the hiker within one hour:

	Hiker North	Hiker South
Search North	0.80	0.30
Search South	0.40	0.70

   Treat this as a game against nature. Find the pure maximin strategy and the mixed strategy that maximizes the guaranteed probability of success.

2. **Penalty kick interpretation.** A team claims, "Our kicker should always kick right because the average scoring rate is higher." Explain why this argument may fail in a strategic game. Use the goalkeeper's possible response in your explanation.

3. **Dominance with context.** In a pricing game, one row is dominated mathematically. Give a possible real-world reason why a firm might still consider it, and explain whether that means the payoff matrix is incomplete.

4. **Repeated interaction.** Revisit the drink-stall pricing game. Explain how repeated play, reputation, or a school regulation might change the equilibrium outcome. State which assumption of the one-shot game has changed.

Strategic modeling brief template

For a competition-style response, students may use the following structure:

1. **Players:** Who makes strategic decisions?
2. **Strategies:** What actions are realistically available to each player?
3. **Payoffs:** What does each player value, and how are numerical payoffs justified?
4. **Solution concept:** Is the game zero-sum, a game against nature, or partial conflict? Which method is appropriate?
5. **Result:** State the saddle point, mixed strategy, game value, or Nash equilibrium.
6. **Interpretation:** What does the result recommend in the real situation?
7. **Limitations:** Which assumptions are most fragile?

A Optional extension: Linear programming formulation for larger zero-sum games

For an $m \times n$ zero-sum matrix $A = (a_{ij})$, let x_i be the probability that the row player chooses row i. The row player maximizes the guaranteed payoff V:

$$\begin{aligned} \text{maximize} \quad & V \\ \text{subject to} \quad & V \leq \sum_{i=1}^{m} a_{ij} x_i, \qquad j = 1, \ldots, n, \\ & \sum_{i=1}^{m} x_i = 1, \\ & x_i \geq 0. \end{aligned}$$

Each inequality says: even if the column player chooses column j, the expected payoff must be at least V.

The column player's dual problem minimizes V subject to analogous constraints. The equality of the two optimal values is closely related to the minimax theorem.

B Optional extension: Repeated Prisoner's Dilemma

If the Prisoner's Dilemma is played once, defection is individually rational. If it is repeated many times, cooperation can become stable. A simple strategy is **Tit-for-Tat**: cooperate first, then copy the opponent's previous move. It rewards cooperation and punishes defection, but only when players expect future interaction and can observe one another's actions reliably.

Model audit card

Audit before trusting or communicating a model

Purpose	What question or decision does the model support?	**Data**	Which inputs are observed, derived, or simulated?
Assumptions	Which simplifications may change the answer?	**Units**	Are dimensions and scales consistent?
Verification	Were equations, code, constraints, and limiting cases checked?	**Validation**	Is performance tested against independent evidence?
Uncertainty	How do measurement, parameter, solver, and structural errors affect conclusions?	**Ethics**	Who benefits, who bears risk, and what fairness criterion matters?
Reproducibility	Are data provenance, versions, seeds, and transformations recorded?	**Communication**	Are claims conditional, traceable, and decision-relevant?

Further reading

- ECON 159: Game Theory (Open Yale Courses)
- GAIMME: Guidelines for Assessment and Instruction in Mathematical Modeling Education (SIAM / COMAP)

Sources, provenance, and licence

These notes follow the iterative modeling and assessment perspective in the SIAM/COMAP GAIMME report. Competition-specific remarks are a snapshot checked on 20 August 2026 against the official COMAP HiMCM instructions; readers should verify current rules before submission. Unless a source is explicitly identified, classroom datasets and numerical scenarios are synthetic or simulated teaching examples and are not empirical claims about a real institution. Third-party references retain their own terms.

Student response guide and checking criteria

For closed calculations, show the definition, substitution, units, and at least one independent check; the worked examples in this unit are method hints, but you must complete the arithmetic yourself. Open modeling responses are checked on five dimensions: a clear purpose and output; stated data and provenance; defensible assumptions and scope; traceable computational verification and model validation; and alignment among uncertainty, limitations, and the recommendation. Support each dimension with concrete evidence rather than writing only "completed."

Lectures in Mathematical Modeling

Lecture 9: Differential Equations and Dynamic Models

Kenneth Sok Kin Cheng

Revised 20 August 2026

Learning goals

By the end of the lesson, readers should be able to:

1. explain when a **differential equation** is more appropriate than a difference equation;
2. build first-order models from the idea "rate in minus rate out" or "rate proportional to current amount";
3. solve and interpret exponential, logistic, cooling, and simple pollution models;
4. use a **phase line** to classify **equilibria** as stable or unstable;
5. apply Euler, improved Euler/Heun, and RK4 ideas to approximate dynamic models;
6. read a basic two-population **phase plane** using nullclines;
7. communicate assumptions, parameters, validation evidence, and limitations of a dynamic model.

1 Why differential equations?

A **difference equation** updates a system at fixed steps:

$$P_{n+1} = P_n + kP_n.$$

A **differential equation** describes the instantaneous rate of change:

$$\frac{dP}{dt} = kP.$$

The second form is appropriate when the change is naturally continuous, or when the time step is so small that a continuous approximation is easier to interpret.

Launch: choosing the time language **[12 min]**

For each situation, decide whether a difference equation or a differential equation is more natural. Give one reason.

1. A bank account receives interest once per month.
2. A hot drink cools continuously after being placed on a table.
3. A school records the number of absent students each morning.
4. A pollutant is continuously flushed out of a lake by a river.

Modeling lesson: The mathematics should match the timing of the mechanism, not merely the data table we happen to have.

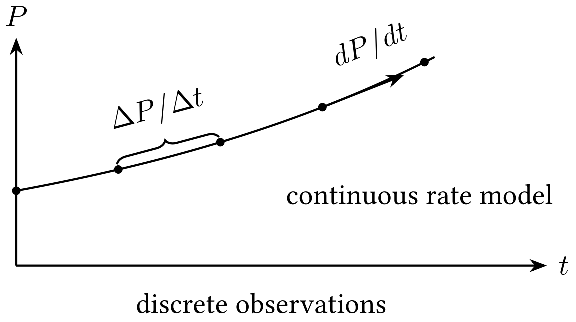


Figure 1: A differential equation idealizes many small updates as an instantaneous rate law.

Core vocabulary

A **first-order ordinary differential equation** has the form

$$\frac{dy}{dt} = f(t, y).$$

An **initial value problem** adds a starting value $y(t_0) = y_0$. A value y^* with $f(y^*) = 0$ is an **equilibrium**. If nearby solutions move toward it, it is **stable**; if they move away, it is **unstable**.

2 Exponential growth and decay

The simplest rate law says that the rate of change is proportional to the present amount:

$$\frac{dP}{dt} = kP.$$

If $k > 0$, the quantity grows exponentially; if $k < 0$, it decays exponentially.

Worked Example 1: Population growth and the danger of extrapolation

Suppose a population is 203.2 million in 1970 and 248.7 million in 1990. Let t be years after 1970 and assume

$$\frac{dP}{dt} = kP, \qquad P(0) = 203.2.$$

Separation gives

$$\frac{dP}{P} = k\,dt \quad \Rightarrow \quad \ln P = kt + C \quad \Rightarrow \quad P(t) = 203.2e^{kt}.$$

Use the 1990 value to estimate k:

$$248.7 = 203.2e^{20k} \quad \Rightarrow \quad k = \frac{1}{20}\ln\left(\frac{248.7}{203.2}\right) \approx 0.0101.$$

The 2000 prediction is

$$P(30) = 203.2e^{0.0101(30)} \approx 274.9 \text{ million}.$$

This is a reasonable short-term estimate, but the same model predicts unlimited growth. The equation is useful only while resources, policy, and demographics remain close to the calibration period.

Model criticism: Model criticism: what does k hide?

The parameter k compresses birth rates, death rates, immigration, age structure, economic conditions, policy, and culture into one number. A good report should say not only that $k \approx 0.0101$, but also why a constant k is a simplification and how far into the future the assumption remains credible.

3 Logistic growth and phase-line reasoning

Exponential growth ignores limited resources. A common refinement is the **logistic model**:

$$\frac{dP}{dt} = rP\left(1 - \frac{P}{M}\right),$$

where M is the **carrying capacity**. The factor $1 - P/M$ reduces growth as P approaches M.

Worked Example 2: Choosing between exponential and logistic models

A fish population in a closed pond is initially $P(0) = 80$. Ecologists believe the pond can support at most $M = 500$ fish. They estimate $r = 0.40$ per month.

$$\frac{dP}{dt} = 0.40P\left(1 - \frac{P}{500}\right).$$

The logistic solution is

$$P(t) = \frac{M}{1 + Ae^{-rt}}, \qquad A = \frac{M - P_0}{P_0}.$$

Here $A = (500 - 80)/80 = 5.25$, so

$$\boxed{P(t) = \frac{500}{1 + 5.25e^{-0.40t}}.}$$

At $t = 6$ months,

$$P(6) = \frac{500}{1 + 5.25e^{-2.4}} \approx 339.$$

The exponential model with the same initial growth rate gives $80e^{0.40(6)} \approx 882$, which already exceeds the pond's capacity. The logistic model is more credible because it encodes the physical constraint.

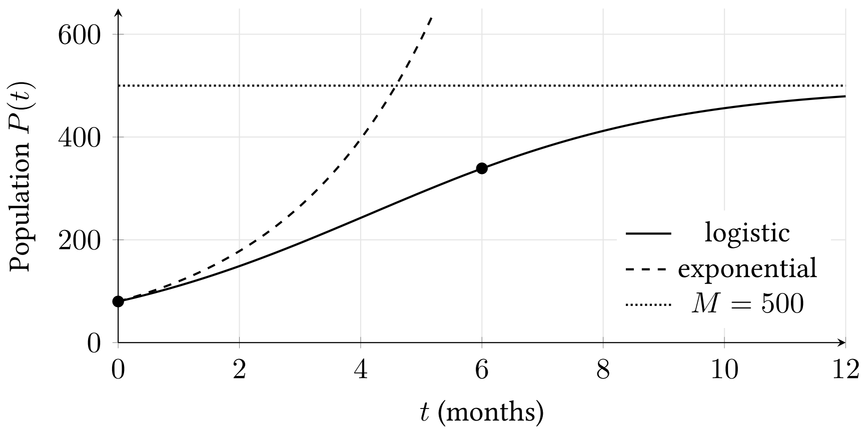
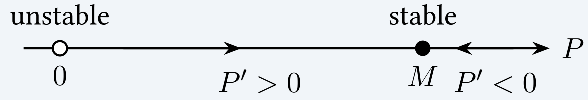


Figure 2: The logistic model behaves like exponential growth at first, but bends toward the carrying capacity.

Phase line for the logistic model

For the autonomous equation $P' = rP(1 - P/M)$, equilibria occur at $P = 0$ and $P = M$. If $0 < P < M$, then $P' > 0$; if $P > M$, then $P' < 0$. Therefore 0 is unstable and M is stable.



Checkpoint: Phase-line check

For $y' = y(3 - y)$:

1. find all equilibria;
2. determine the sign of y' on each interval;
3. classify each equilibrium as stable or unstable;
4. explain the long-run behavior if $y(0) = 1$, $y(0) = 5$, and $y(0) = -1$.

4 Rate-balance models: cooling and pollution

Many differential equation models are built from the principle

$$\text{rate of change} = \text{rate in} - \text{rate out}.$$

This principle is easier for students to remember than a list of formula types.

Worked Example 3: Newton's law of cooling

A cup of tea is $90°$C when placed in a room at $25°$C. Newton's law of cooling assumes that the cooling rate is proportional to the temperature difference:

$$\frac{dT}{dt} = -k(T - 25), \qquad T(0) = 90.$$

The solution is

$$T(t) = 25 + 65e^{-kt}.$$

If after 10 minutes the temperature is $55°$C, then

$$55 = 25 + 65e^{-10k} \quad \Rightarrow \quad e^{-10k} = \frac{30}{65} \quad \Rightarrow \quad k \approx 0.0773.$$

To find when the tea reaches $40°$C:

$$40 = 25 + 65e^{-0.0773t} \quad \Rightarrow \quad t \approx 19.0 \text{ minutes.}$$

The stable equilibrium is the ambient temperature $25°$C.

Worked Example 4: Lake cleanup as rate in minus rate out

A lake has volume $V = 10^9$ L. Clean water flows in and out at $r = 5 \times 10^7$ L/year. Let $p(t)$ be the amount of pollutant in grams. If no new pollutant enters, the pollutant leaves at concentration p/V:

$$\frac{dp}{dt} = 0 - r\frac{p}{V} = -\frac{r}{V}p = -0.05p.$$

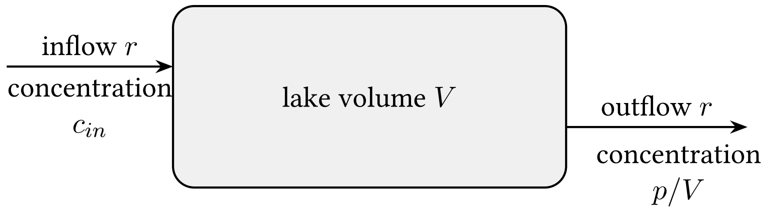
So

$$p(t) = p_0 e^{-0.05t}.$$

The time to reach 10% of the initial pollutant is

$$0.10p_0 = p_0 e^{-0.05t} \quad \Rightarrow \quad t = \frac{\ln 10}{0.05} \approx 46.1 \text{ years}.$$

The model explains why environmental cleanup can be slow even after pollution input stops: the relevant time scale is the **flushing time** $V/r = 20$ years.



$$p'(t) = rc_{in} - r\,p(t)/V$$

Figure 3: A lake pollution model follows directly from rate in minus rate out.

Model criticism: Validation and limitations

A rate-balance model assumes perfect mixing: the pollutant concentration is the same everywhere in the lake. In a large lake, this may be false. A report should discuss whether the lake is well mixed, whether the inflow and outflow rates vary seasonally, and whether pollutant can settle into sediment and later re-enter the water.

Worked Example 5: Repeated drug dosage

A medicine is effective above $L = 2$ mg/L and unsafe above $H = 8$ mg/L. After a dose, the blood concentration decays according to

$$C' = -kC, \qquad k = 0.18 \text{ hr}^{-1}.$$

If each dose raises concentration by $H - L = 6$ mg/L, the residual concentration just before the next dose should approach L. For repeated dosing at interval T, the limiting residual is

$$R = \frac{H - L}{e^{kT} - 1}.$$

Setting $R = L$ gives

$$L = \frac{H - L}{e^{kT} - 1} \quad \Rightarrow \quad e^{kT} = \frac{H}{L} \quad \Rightarrow \quad \boxed{T = \frac{1}{k} \ln\left(\frac{H}{L}\right)}.$$

Here

$$T = \frac{1}{0.18} \ln 4 \approx 7.70 \text{ hours}.$$

Interpretation: an 8-hour schedule is close but slightly slower than the ideal schedule; it may be acceptable only if clinical safety margins and patient variation are considered.

Checkpoint: Dosage interpretation

If k increases because the body eliminates the drug faster, should the recommended interval T increase or decrease? Explain using the formula and in ordinary language.

5 Numerical solution: from Euler to reliable computation

Analytic formulas are useful, but many modeling ODEs cannot be solved neatly. In competitions and applied projects, it is often more realistic to state the differential equation clearly and then compute the solution numerically. The issue is not merely "press a button". Students must understand what the numerical method is doing, how accurate it is likely to be, and how to check whether the output is trustworthy.

Numerical solution as a modeling tool

A **numerical solution** approximates the values of an unknown solution at a grid of times

$$t_0, t_1 = t_0 + h, t_2 = t_0 + 2h, \ldots$$

where h is the **step size**. A numerical method turns the rate law

$$y' = f(t, y)$$

into an update rule for y_{n+1} from information near (t_n, y_n). The smaller the step size, the more often the model is updated, but the more computation is required.

Euler's method as the baseline

Euler's method follows the tangent line at the current point:

$$\boxed{t_{n+1} = t_n + h, \qquad y_{n+1} = y_n + hf(t_n, y_n).}$$

It is valuable because it reveals the mechanism of numerical integration. However, it is only first-order accurate: the **global error** is usually $O(h)$, so halving h often roughly halves the error.

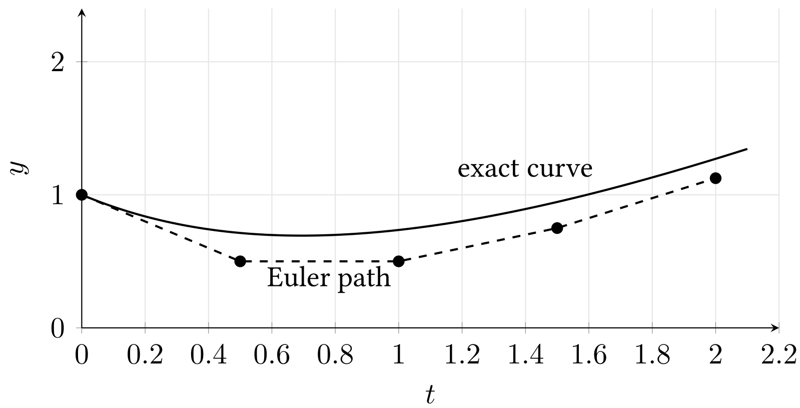


Figure 4: Euler's method replaces a curved solution by short tangent-line steps.

Worked Example 6: Euler approximation and step-size check

Approximate the solution of

$$y' = t - y, \qquad y(0) = 1,$$

on $0 \le t \le 2$ using $h = 0.5$.

$$y_{n+1} = y_n + 0.5(t_n - y_n).$$

n	t_n	y_n	slope $t_n - y_n$	y_{n+1}
0	0.0	1.000	-1.000	0.500
1	0.5	0.500	0.000	0.500
2	1.0	0.500	0.500	0.750
3	1.5	0.750	0.750	1.125

The exact solution is $y = t - 1 + 2e^{-t}$. At $t = 2$, exact $y(2) \approx 1.271$, so Euler gives error about 0.146. **Step-size check.** If we repeat Euler with $h = 0.25$, the estimate at $t = 2$ is approximately 1.200, with error about 0.070. The error is roughly halved, which is consistent with first-order convergence.

Checkpoint: Numerical method check

Why is a single Euler computation not enough evidence? State two checks a modeler should perform before trusting a numerical ODE result.

Improved Euler / Heun's method

Euler uses only the slope at the beginning of the step. If the slope changes across the interval, this can be crude. The **improved Euler method**, also called **Heun's method**, first predicts a trial endpoint, then averages the starting and ending slopes:

$$\begin{aligned} k_1 &= f(t_n, y_n), \\ y_{\text{pred}} &= y_n + hk_1, \\ k_2 &= f(t_n + h, y_{\text{pred}}), \\ y_{n+1} &= y_n + \frac{h}{2}(k_1 + k_2). \end{aligned}$$

This is a second-order method: its global error is typically $O(h^2)$.

Worked Example 7: Euler versus Heun versus RK4

Use the same IVP $y' = t - y$, $y(0) = 1$, with $h = 0.5$ up to $t = 2$. The exact value is $y(2) = 1.27067$.

Method	Update idea	Estimate at $t = 2$	Error
Euler	starting slope only	1.12500	0.14567
Improved Euler / Heun	average of start and predicted-end slopes	1.30518	0.03451
RK4	weighted four-slope average	1.27110	0.00043

The main lesson is not that one method is always best, but that numerical method choice matters. Euler is excellent for explaining the idea; Heun and RK4 are better default tools for producing values in a report.

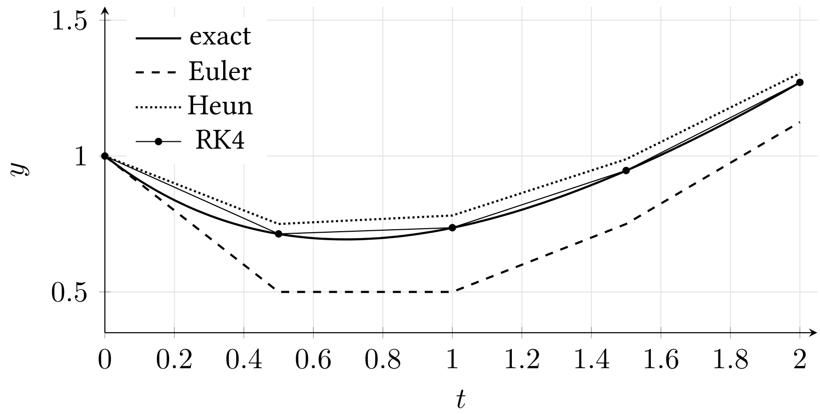


Figure 5: Different numerical methods may produce visibly different paths even with the same step size.

Fourth-order Runge-Kutta: the practical default

The classical **fourth-order Runge-Kutta method** balances accuracy and simplicity. It samples four slopes inside one step:

$$\begin{aligned}
k_1 &= f(t_n, y_n),\\
k_2 &= f\left(t_n + \frac{h}{2},\, y_n + \frac{h}{2}k_1\right),\\
k_3 &= f\left(t_n + \frac{h}{2},\, y_n + \frac{h}{2}k_2\right),\\
k_4 &= f(t_n + h,\, y_n + hk_3),\\
y_{n+1} &= y_n + \frac{h}{6}(k_1 + 2k_2 + 2k_3 + k_4).
\end{aligned}$$

RK4 has global error $O(h^4)$ for sufficiently smooth problems. In student modeling reports, it is often enough to state that the ODE was solved by RK4 or by a standard adaptive solver, and then report a step-size or tolerance check.

A numerical ODE workflow for modeling reports

1. **Write the rate law clearly.** Define variables, units, parameters, and initial conditions.
2. **Choose a method.** Use Euler for classroom illustration; use Heun, RK4, or an adaptive solver for reported numerical results.
3. **Check convergence.** Repeat with h and $h/2$, or with a tighter solver tolerance. The conclusion should not change materially.
4. **Check against known behavior.** The numerical path should respect equilibria, signs, conservation, non-negativity, or carrying capacities.
5. **Separate numerical error from modeling error.** A better solver cannot fix a wrong mechanism.

Stability of a numerical method

Accuracy is not the only issue. A method can be unstable even for a simple stable differential equation. Consider

$$y' = -10y, \qquad y(0) = 1.$$

The true solution $y = e^{-10t}$ decays smoothly to 0. Euler's method gives

$$y_{n+1} = y_n + h(-10y_n) = (1 - 10h)y_n.$$

For the numerical solution to decay rather than explode, we need

$$|1 - 10h| < 1 \quad \Rightarrow \quad 0 < h < 0.2.$$

Thus a large step size may create artificial oscillation or blow-up even though the true model is stable.

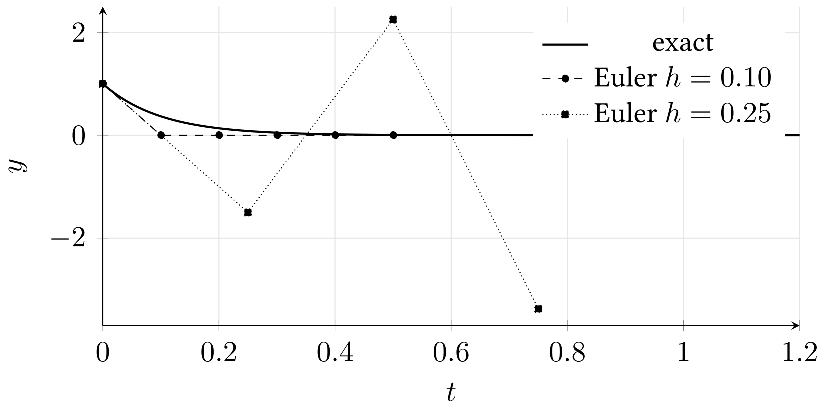


Figure 6: For a stable decay equation, too large a step size can create artificial numerical instability.

Model criticism: What to write in a competition report

Avoid writing only "we used Euler's method". A stronger statement is:

> We solved the ODE numerically using RK4 with step size $h = 0.05$. Repeating the computation with $h = 0.025$ changed the predicted peak by less than 1%, so numerical error is small relative to parameter uncertainty. The model preserves non-negative populations and approaches the expected stable equilibrium.

This sentence shows method, convergence, qualitative validation, and limitation.

6 Systems and phase planes: predator and prey

For two interacting quantities, a single phase line is not enough. We use a **phase plane**. The key visual tools are **nullclines**: curves where one derivative is zero.

Worked Example 8: Predator-prey nullclines

Let $x(t)$ be prey and $y(t)$ be predator. A simple Lotka-Volterra model is

$$x' = x(2 - 0.5y), \qquad y' = y(-1 + 0.1x).$$

The x-nullclines satisfy $x' = 0$, so $x = 0$ or $y = 4$. The y-nullclines satisfy $y' = 0$, so $y = 0$ or $x = 10$. The positive equilibrium is

$$(x^*, y^*) = (10, 4).$$

If $y < 4$, prey increase; if $y > 4$, prey decrease. If $x > 10$, predators increase; if $x < 10$, predators decrease.

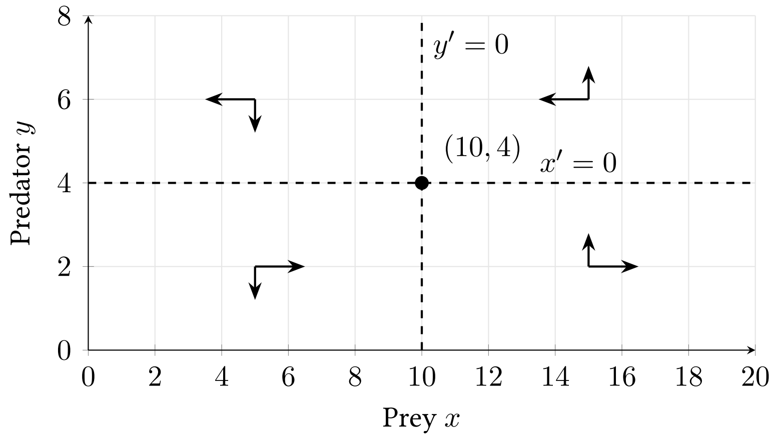


Figure 7: Nullclines divide the predator-prey phase plane into regions where each population increases or decreases.

Differentiated modeling studio **[20 min]**

Choose one route.

1. **Core route: dosing interval.** A medicine has $L = 2$, $H = 8$, and $k = 0.18$ hr^{-1}. Use $T = (1/k)\ln(H/L)$ and explain the result in plain language.
2. **Challenge route: lake cleanup.** A lake has $V = 3 \times 10^9$ L and $r = 10^8$ L/year. Find the time to reduce pollutant to 25% of its initial amount, then discuss one assumption that may fail.
3. **Extension route: predator-prey.** For $x' = x(1-y)$ and $y' = y(-2+x)$, find the nullclines, the positive equilibrium, and the signs of x' and y' in the four regions.

Choosing a differential-equation method

Situation	**Recommended method**	**Modeling question answered**
$y' = ky$ or $y' = -ky$	separation / known formula	How fast does the amount grow or decay?
$y' = rP(1 - P/M)$	logistic formula + phase line	What long-run capacity is approached?
$y' = f(y)$	phase line	Which equilibria are stable?
$y' + P(t)y = Q(t)$	integrating factor	What is the balance between input and removal?
No clean formula, classroom calculation	Euler	What does one tangent-line approximation do?
No clean formula, report-quality computation	Heun / RK4 / adaptive solver	Is the numerical prediction stable under step-size refinement?
Two interacting variables	nullclines / phase plane	Which population increases or decreases in each region?

Model criticism: Numerical error versus modeling error

If Euler's method gives a poor estimate, there are two possible causes. **Numerical error** means the step size is too large for the differential equation. **Modeling error** means the differential equation itself is not a good representation of the real system. Reducing h helps only with numerical error; it cannot fix a bad assumption such as constant temperature, perfect mixing, or unlimited growth.

7 Writing dynamic-model conclusions

Model criticism: Competition-style communication

A dynamic-model paragraph should contain four parts:

1. **Mechanism:** what rate law or balance law was assumed?
2. **Calibration**: how were parameters such as k, r, M, or V/r obtained?
3. **Prediction and interpretation:** what does the model imply in context?
4. **Credibility check**: what evidence supports the model, and where does it likely fail?

Weak conclusion: "The logistic model predicts 339 fish."
Stronger conclusion: "Using a carrying capacity of 500 fish and an estimated monthly growth rate of 0.40, the logistic model predicts about 339 fish after six months. The exponential model exceeds the pond capacity over the same period, so it is unsuitable for long-range prediction. The main uncertainty is the carrying capacity estimate, which should be checked against pond area, food supply, and historical population counts."

Lesson summary

Concept	Modeling meaning
Differential equation	Relates a quantity to its instantaneous rate of change.
Initial condition	Selects one solution curve from a family.
Exponential model	Rate proportional to amount; useful locally, dangerous for long extrapolation.
Logistic model	Growth slows near carrying capacity; stable equilibrium at M.
Phase line	Visual tool for one-variable autonomous DEs; shows stability.
Rate-balance model	rate in − rate out; used for pollution, mixing, and flow problems.
Euler's method	First-order tangent-line approximation; useful as a baseline but often too crude for final results.
Heun / RK4	Higher-order numerical solvers; compare step sizes before reporting quantitative predictions.
Phase plane	Visual tool for systems; nullclines locate equilibria and direction regions.
Model criticism	State assumptions, parameter sources, validation evidence, and limits.

Checkpoint: Exit ticket

In three sentences, explain the difference between an exponential model and a logistic model. Then give one real-world situation where the logistic model is more credible.

Core practice

1. Solve $y' = 0.3y$, $y(0) = 12$. Find $y(5)$ and interpret the parameter 0.3.
2. A chemical decays according to $C' = -0.12C$, $C(0) = 40$. When does it fall below 5?
3. For $P' = 0.2P(1 - P/1000)$ and $P(0) = 100$, write the logistic solution and estimate $P(10)$.
4. Draw the phase line for $y' = y(y - 4)$ and classify the equilibria.
5. A room is kept at 22°C. An object cools from $80°\mathrm{C}$ according to $T' = -0.05(T - 22)$. Find $T(t)$ and $T(20)$.
6. Apply Euler's method with $h = 1$ to $y' = 2 - y$, $y(0) = 0$ for three steps. Then repeat with Heun's method and compare.

Modeling challenge

A small reservoir initially contains 12 million litres of water with pollutant concentration 0.08 g/L. Clean water enters and exits at 0.6 million litres per day. A nearby factory may continue to add pollutant at an unknown constant rate s grams/day.

1. Define $p(t)$ and write a differential equation for pollutant amount.
2. Solve the model when $s = 0$.
3. Find the long-run pollutant level when $s > 0$.
4. Suppose measurements after 10 days show concentration 0.052 g/L. Estimate whether s is closer to 0 or significantly positive.
5. Write a short recommendation for environmental officers. Include limitations of the model.

Extension practice

1. Use separation to solve $y' = ty^2$, $y(0) = 1$. Identify whether the solution exists for all $t > 0$.
2. For $y' = y(1 - y)(y - 3)$, draw the phase line and classify all equilibria.
3. For $x' = x(3 - y)$, $y' = y(-2 + x)$, find the positive equilibrium and the nullclines. Give the direction of motion in each region.
4. Explain why Euler's method may become inaccurate or unstable if h is too large, even when the differential equation is correct. Use $y' = -10y$ as an example.

Optional appendix: integrating factor method

A first-order linear equation has standard form

$$y' + P(t)y = Q(t).$$

The integrating factor is

$$\mu(t) = e^{\int P(t)\,dt}.$$

Multiplying the equation by $\mu(t)$ turns the left side into a product derivative:

$$(\mu y)' = \mu Q.$$

Thus

$$\mu(t)y(t) = \int \mu(t)Q(t)\,dt + C.$$

This method is useful for pollution, mixing, heating/cooling with input, and many first-order engineering models.

Additional problem bank

1. **Coffee cooling.** Coffee cools from $85°$C to $60°$C in 12 minutes in a $24°$C room. Estimate k and predict when it reaches $45°$C.
2. **Choosing a model.** A population grows from 100 to 180 to 250 to 310 over four equal time intervals, and experts believe the maximum is about 400. Explain why an exponential model may overpredict and propose a logistic model structure.
3. **Parameter sensitivity.** For the lake cleanup model $p(t) = p_0 e^{-rt/V}$, explain how the cleanup time changes if the inflow-outflow rate r doubles.
4. **Euler design.** A student uses $h = 5$ days to simulate a fast-changing epidemic model. What checks should be performed before trusting the output?
5. **Model comparison.** Write two sentences comparing a phase-line conclusion and a numerical simulation conclusion. What can each reveal that the other may hide?

Model audit card

Audit before trusting or communicating a model

Purpose	What question or decision does the model support?	**Data**	Which inputs are observed, derived, or simulated?
Assumptions	Which simplifications may change the answer?	**Units**	Are dimensions and scales consistent?
Verification	Were equations, code, constraints, and limiting cases checked?	**Validation**	Is performance tested against independent evidence?
Uncertainty	How do measurement, parameter, solver, and structural errors affect conclusions?	**Ethics**	Who benefits, who bears risk, and what fairness criterion matters?
Reproducibility	Are data provenance, versions, seeds, and transformations recorded?	**Communication**	Are claims conditional, traceable, and decision-relevant?

Further reading

- 18.03SC Differential Equations (MIT OpenCourseWare)
- GAIMME: Guidelines for Assessment and Instruction in Mathematical Modeling Education (SIAM / COMAP)

Sources, provenance, and licence

These notes follow the iterative modeling and assessment perspective in the SIAM/COMAP GAIMME report. Competition-specific remarks are a snapshot checked on 20 August 2026 against the official COMAP HiMCM instructions; readers should verify current rules before submission. Unless a source is explicitly identified, classroom datasets and numerical scenarios are synthetic or simulated teaching examples and are not empirical claims about a real institution. Third-party references retain their own terms.

Student response guide and checking criteria

For closed calculations, show the definition, substitution, units, and at least one independent check; the worked examples in this unit are method hints, but you must complete the arithmetic yourself. Open modeling responses are checked on five dimensions: a clear purpose and output; stated data and provenance; defensible assumptions and scope; traceable computational verification and model validation; and alignment among uncertainty, limitations, and the recommendation. Support each dimension with concrete evidence rather than writing only "completed."

Lectures in Mathematical Modeling

Lecture 10: Continuous Optimization and Model Design

Kenneth Sok Kin Cheng

Revised 20 August 2026

Learning goals

By the end of the lesson, readers should be able to:

1. distinguish **continuous optimization** from the discrete and linear-programming models in Lecture 6;
2. construct a single-variable objective function from submodels, especially in the **economic order quantity** model;
3. use first- and second-derivative tests to find and interpret an optimum;
4. solve a two-variable unconstrained optimization problem using **partial derivatives** and the **Hessian**;
5. explain the geometric idea behind **Lagrange multipliers**;
6. interpret a multiplier as a **shadow price**;
7. describe when numerical optimization is needed and how to report solver checks in a modeling paper.

1 From choosing among options to choosing a best value

Lecture 6 optimized decisions over a polygon or a list of integer choices. Here the decision variable can vary continuously. A delivery interval might be 2.8 days, a price might be \$37.40, and a mixing proportion might be 0.62. The modeling task is not simply to differentiate a function: it is to build the function that deserves to be optimized.

Launch: a decision is not yet a model **[12 min]**

A school shop wants to decide how many bottles of water to order at a time. Ordering too often creates administrative cost; ordering too many creates storage cost and risk of waste.

1. What is the decision variable?
2. What should the objective be: total cost per order, cost per day, profit, or service reliability?
3. Which assumptions would make the problem continuous rather than integer?
4. What data would you need before trusting the recommendation?

Modeling lesson: continuous optimization starts before calculus. The most important step is deciding what quantity is being optimized and what simplifications make the objective computable.

Core vocabulary

A **decision variable** is a controllable quantity. An **objective function** measures what we want to maximize or minimize. A **constraint** restricts feasible decisions. A **critical point** is a point where the derivative or gradient is zero. A **global optimum** is best among all feasible choices; a **local optimum** is only best nearby.

2 Single-variable optimization: the EOQ model

The **economic order quantity** model is useful because it is simple, teachable, and rich in modeling interpretation. It balances two costs that move in opposite directions: frequent ordering and excessive storage.

Worked Example 1: Deriving the EOQ formula

A station sells fuel at a constant rate r gallons per day. Every delivery has fixed cost d, and holding one gallon for one day costs s. Let T be the delivery cycle length in days.
Step 1: connect the decision to the order size. If the station orders exactly enough fuel for T days, then

$$Q = rT.$$

Step 2: build cost per cycle. Delivery cost per cycle is d. The inventory decreases approximately linearly from Q to 0, so the average inventory is $Q/2 = rT/2$. The storage cost per cycle is

$$s \cdot \frac{rT}{2} \cdot T = \frac{srT^2}{2}.$$

Step 3: optimize average daily cost.

$$c(T) = \frac{d + srT^2/2}{T} = \frac{d}{T} + \frac{srT}{2}, \qquad T > 0.$$

Then

$$c'(T) = -\frac{d}{T^2} + \frac{sr}{2}.$$

Setting $c'(T) = 0$ gives

$$T^2 = \frac{2d}{sr}, \qquad \boxed{T^* = \sqrt{\frac{2d}{sr}}}.$$

Since

$$c''(T) = \frac{2d}{T^3} > 0,$$

this critical point is a minimum. The optimal order quantity is

$$\boxed{Q^* = rT^* = \sqrt{\frac{2dr}{s}}}.$$

Interpretation: high delivery cost pushes the model toward larger, less frequent orders; high storage cost pushes it toward smaller, more frequent orders.

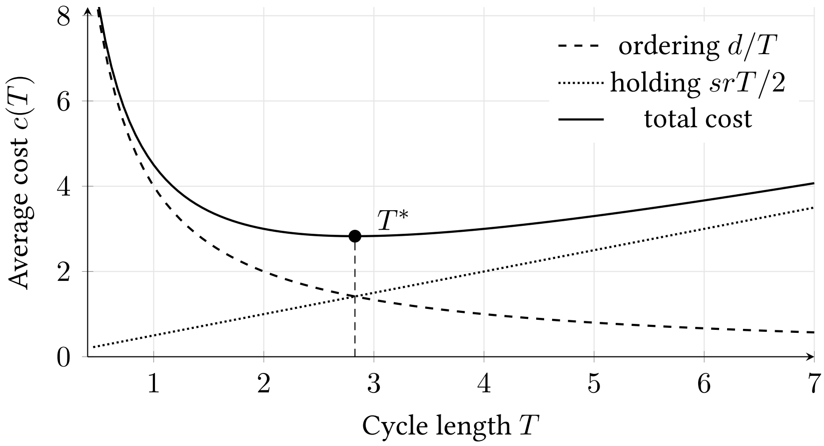


Figure 1: The EOQ optimum occurs where the decreasing order-cost curve and increasing holding-cost curve balance.

Checkpoint: Equal-cost principle

Show that at T^*,

$$\frac{d}{T^*} = \frac{srT^*}{2}.$$

Then explain why this is a useful way to check whether a numerical answer is reasonable.

Worked Example 2: Applying EOQ and checking sensitivity

A bookstore sells $r = 200$ textbooks per month. Each order costs $d = \$120$. Holding cost is $s = \$0.50$ per book per month.

$$T^* = \sqrt{\frac{2(120)}{0.50(200)}} = \sqrt{2.4} \approx 1.55 \text{ months},$$

so

$$Q^* = rT^* = 200(1.55) \approx 310 \text{ books}.$$

The minimum monthly cost is

$$c^* = \frac{120}{1.55} + \frac{0.50(200)(1.55)}{2} \approx \$154.9.$$

Sensitivity: if storage cost doubles to $s = 1.00$, then

$$T^*_{new} = \sqrt{\frac{2(120)}{1.00(200)}} = \sqrt{1.2} \approx 1.10 \text{ months}.$$

The cycle length is multiplied by $1/\sqrt{2} \approx 0.707$, not cut in half. Square-root laws are less sensitive than linear laws.

Model criticism: Model criticism: what EOQ assumes

The EOQ model assumes constant demand, no stockouts, instant delivery, no quantity discount, and divisible order sizes. In a modeling report, the EOQ result should be treated as a baseline. If demand is seasonal, delivery has delay, or order quantity must be an integer carton size, the final recommendation may require simulation or integer optimization.

3 Two-variable optimization: gradient and Hessian

For a function $f(x, y)$, the **gradient**

$$\nabla f = (f_x, f_y)$$

points in the direction of fastest increase. A critical point satisfies

$$f_x = 0, \qquad f_y = 0.$$

The **Hessian**

$$H = \begin{pmatrix} f_{xx} & f_{xy} \\ f_{yx} & f_{yy} \end{pmatrix}$$

helps classify the point. For two variables, define

$$D = f_{xx} f_{yy} - f_{xy}^2.$$

If $D > 0$ and $f_{xx} < 0$, the point is a local maximum. If $D > 0$ and $f_{xx} > 0$, it is a local minimum. If $D < 0$, it is a saddle point.

Worked Example 3: Two-product profit optimization

A company produces two products. Let x and y be daily production quantities. Suppose the estimated profit is

$$P(x, y) = 100x + 120y - x^2 - 2y^2 - xy - 500.$$

The negative quadratic terms represent price pressure, congestion, or diminishing returns.
Find the critical point.

$$P_x = 100 - 2x - y, \qquad P_y = 120 - 4y - x.$$

Set both to zero:

$$2x + y = 100, \qquad x + 4y = 120.$$

Solving gives

$$x^* = 40, \qquad y^* = 20.$$

Classify.

$$P_{xx} = -2, \quad P_{yy} = -4, \quad P_{xy} = -1, \quad D = (-2)(-4) - (-1)^2 = 7 > 0.$$

Since $D > 0$ and $P_{xx} < 0$, the point is a local maximum. Because the quadratic form is concave, it is the global maximum.
Interpretation. At $(40, 20)$, marginal profit from one more unit of either product is zero. Producing more of one product reduces the marginal profitability of the other through the cross-term $-xy$.

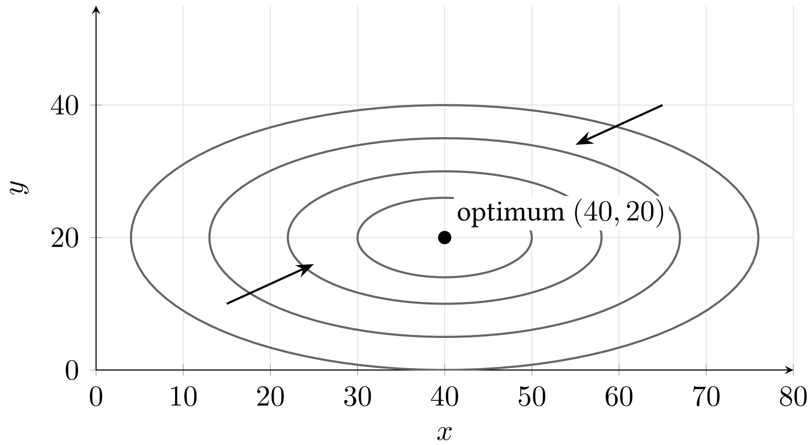


Figure 2: Contour lines of the profit function. The arrows indicate uphill directions toward the maximum.

Checkpoint: Hessian check

For

$$f(x, y) = 48xy - 32x^3 - 24y^2,$$

find the interior critical point and use the Hessian test to classify it. Then explain why a boundary check would still be needed if the feasible domain were a closed rectangle.

4 Constraints and Lagrange multipliers

Many continuous optimization problems have a constraint. The geometry is simple: at a constrained optimum, the best achievable level curve of the objective is tangent to the constraint curve. This means the gradients are parallel.

Lagrange multiplier method

To optimize $f(x, y)$ subject to $g(x, y) = c$, form

$$\mathcal{L}(x, y, \lambda) = f(x, y) + \lambda\,[g(x, y) - c].$$

Solve

$$\mathcal{L}_x = 0, \qquad \mathcal{L}_y = 0, \qquad \mathcal{L}_\lambda = 0.$$

Equivalently,

$$\nabla f = -\lambda \nabla g.$$

The multiplier λ measures marginal value: how much the optimal objective changes when the constraint is relaxed by one unit, with sign depending on convention and whether we maximize or minimize.

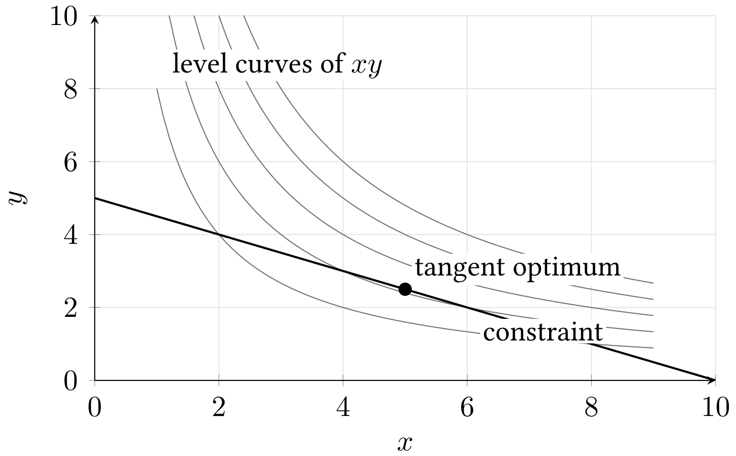


Figure 3: At a constrained optimum, the objective level curve just touches the feasible constraint curve.

Worked Example 4: Maximum area with fixed fencing

A farmer has 100 meters of fencing for three sides of a rectangular garden against a wall. Let x be the side perpendicular to the wall and y the side parallel to the wall. The constraint is

$$2x + y = 100.$$

The area is

$$A = xy.$$

Use Lagrange multipliers:

$$\mathcal{L} = xy + \lambda(2x + y - 100).$$

Then

$$\mathcal{L}_x = y + 2\lambda = 0, \qquad \mathcal{L}_y = x + \lambda = 0, \qquad 2x + y = 100.$$

From $\lambda = -x$ and $y + 2\lambda = 0$, we get

$$y = 2x.$$

Substitute into the constraint:

$$2x + 2x = 100 \Rightarrow x = 25, \qquad y = 50.$$

The maximum area is

$$A^* = 25 \cdot 50 = 1250 \text{ m}^2.$$

Shadow-price interpretation: if the total fencing increases slightly, the maximum area increases at approximately the marginal value of the constraint. From the single-variable form $A(x) = x(100 - 2x)$, the optimal value is $A^*(F) = F^2/8$, so $dA^*/dF = F/4 = 25$ at $F = 100$. One extra meter of fencing adds about 25 m^2 of area near the optimum.

Worked Example 5: Budget-constrained production

A small workshop has a production function

$$Q(L, K) = L^{0.6}K^{0.4},$$

where L is labor units and K is capital units. Labor costs \$50 per unit and capital costs \$80 per unit. The budget is \$4000:

$$50L + 80K = 4000.$$

Maximize Q subject to this budget.

Use

$$\mathcal{L} = L^{0.6}K^{0.4} + \lambda(50L + 80K - 4000).$$

The first-order conditions imply

$$\frac{Q_L}{Q_K} = \frac{50}{80}.$$

Since

$$Q_L = 0.6L^{-0.4}K^{0.4}, \qquad Q_K = 0.4L^{0.6}K^{-0.6},$$

we get

$$\frac{Q_L}{Q_K} = \frac{0.6}{0.4}\frac{K}{L} = 1.5\frac{K}{L} = \frac{50}{80} = 0.625.$$

Thus

$$K = \frac{0.625}{1.5}L \approx 0.4167L.$$

Use the budget:

$$50L + 80(0.4167L) = 4000 \Rightarrow 83.33L = 4000 \Rightarrow L^* = 48,$$

so

$$K^* = 20.$$

Interpretation: the optimal mix does not spend equally on labor and capital. It balances marginal product per dollar.

Model criticism: Important limitation: equality constraints are not the whole story

Lagrange multipliers handle smooth equality constraints. Many real models also have inequalities, such as $x \geq 0$, budget caps, capacity limits, and minimum service requirements. If an inequality is binding, it behaves like an equality; if it is slack, its multiplier is zero. The full theory is called the **Karush–Kuhn–Tucker conditions**, which is useful but too technical for the core lesson.

5 Numerical optimization: when equations do not solve neatly

Calculus gives equations, but not always a closed-form solution. In competition work, students often minimize a nonlinear error function or maximize a model score using software. The key is not only to run a solver, but to know how to check and report it.

Gradient search as a modeling tool

For maximizing $f(\mathbf{x})$, a basic gradient ascent step is

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k \nabla f(\mathbf{x}_k),$$

where α_k is a step size. For minimizing an error function, use gradient descent:

$$\mathbf{x}_{k+1} = \mathbf{x}_k - \alpha_k \nabla f(\mathbf{x}_k).$$

In practice, α_k may be chosen by a **line search**, such as golden-section search, or by a solver routine.

Worked Example 6: Gradient ascent for a profit model

Use the profit model from Worked Example 3:

$$P(x, y) = 100x + 120y - x^2 - 2y^2 - xy - 500.$$

Starting from $(0, 0)$,

$$\nabla P = (100 - 2x - y,\ 120 - 4y - x).$$

With step size $\alpha = 0.10$, the first iteration is

$$(x_1, y_1) = (0, 0) + 0.10(100, 120) = (10, 12).$$

At $(10, 12)$,

$$\nabla P = (100 - 20 - 12,\ 120 - 48 - 10) = (68, 62),$$

so

$$(x_2, y_2) = (10, 12) + 0.10(68, 62) = (16.8, 18.2).$$

The path moves toward $(40, 20)$, but a fixed step size may overshoot or converge slowly. A good numerical study should compare step sizes or use a reliable line search.

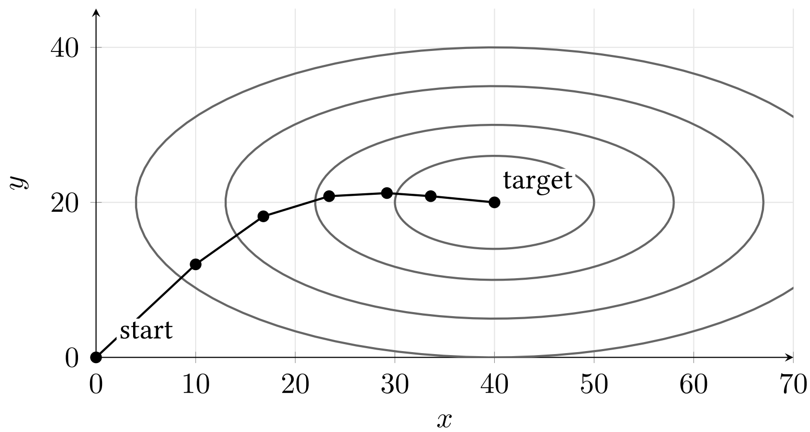


Figure 4: A numerical optimizer follows improvement directions, but convergence must be checked.

Model criticism: How to report numerical optimization

A competition report should not simply say "we used a solver." A stronger description is:

We optimized the nonlinear objective using gradient-based search. To reduce dependence on initial values, we used five starting points and obtained the same optimum to three decimal places. We also checked the result by plotting objective contours and by verifying that the gradient norm at the solution was below 10^{-5}*. A step-size refinement check changed the objective by less than* 0.2%*.*

This tells the reader that the numerical result is not a black box.

6 Applied extension: renewable-resource optimization

This section is suitable for the final part of the lesson or for homework. It connects optimization to sustainability.

Worked Example 7: Biological versus economic fishery optimum

Let N be fish population and let annual natural growth be

$$g(N) = aN(N_u - N), \qquad 0 < N < N_u.$$

The **biological optimum** maximizes sustainable yield $g(N)$. Since

$$g'(N) = a(N_u - 2N),$$

we get

$$N_b = \frac{N_u}{2}.$$

Suppose price per fish is p and harvest cost per fish $c(N)$ decreases as fish become easier to catch in a larger population. Profit from sustainable harvest is

$$\text{TP}(N) = g(N)[p - c(N)].$$

At N_b,

$$\text{TP}'(N_b) = g'(N_b)[p - c(N_b)] - g(N_b)c'(N_b).$$

Because $g'(N_b) = 0$, $g(N_b) > 0$, and $c'(N_b) < 0$,

$$\text{TP}'(N_b) = -g(N_b)c'(N_b) > 0.$$

Therefore the profit function is still increasing at N_b, so the **economic optimum** N_p satisfies

$$N_p > N_b.$$

Interpretation: maximizing profit may mean keeping a larger stock and harvesting less than the maximum sustainable yield, because catching each fish is cheaper when the stock is abundant.

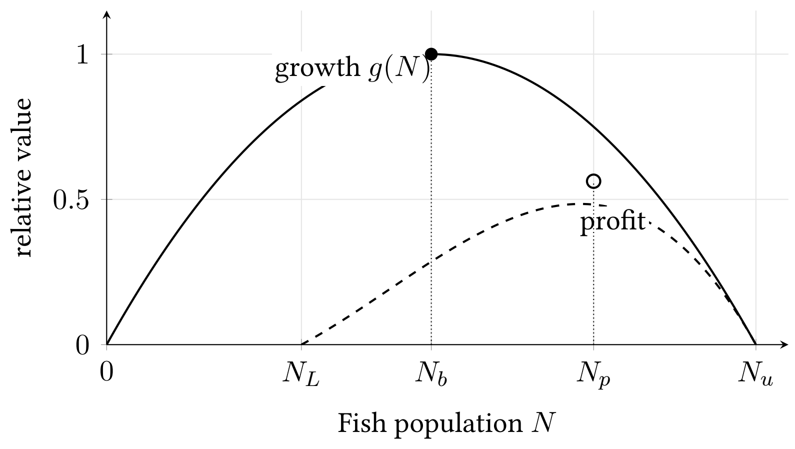


Figure 5: The biological optimum maximizes yield; the economic optimum may occur at a larger population.

Model criticism: Policy interpretation

A mathematical optimum is not automatically a policy recommendation. A fishery model should consider uncertainty in the stock estimate, illegal catch, ecosystem effects, enforcement cost, and risk of collapse. A cautious policy may prefer a population above the economic optimum when measurement error is large.

7 Classroom studio and communication

Differentiated optimization studio **[18 min]**

Choose one route.

1. **Core route: EOQ.** For $d = 80$, $r = 120$, $s = 0.40$, compute T^* and Q^*. Check the equal-cost principle.
2. **Challenge route: constrained area.** Maximize xy subject to $x + 2y = 12$. Solve by Lagrange multipliers and interpret the geometry.
3. **Extension route: numerical optimization.** Starting from $(0, 0)$, perform three gradient-ascent steps with $\alpha = 0.05$ for $P(x, y) = 100x + 120y - x^2 - 2y^2 - xy - 500$. Compare with the exact optimum $(40, 20)$.

Model criticism: Competition-style conclusion

A continuous-optimization conclusion should include:

1. **Decision:** state the recommended value of the decision variable(s).
2. **Reason:** identify the objective and why the optimum occurs there.
3. **Sensitivity:** describe how the recommendation changes when key parameters change.
4. **Feasibility:** check integer, capacity, safety, or practical constraints.
5. **Limitations:** state assumptions that may break in real use.

Weak conclusion: "The optimal order quantity is 310 books."
Stronger conclusion: "Under constant monthly demand of 200 books, an ordering cost of $120 and holding cost of $0.50 per book-month, the EOQ model recommends ordering about 310 books every 1.55 months. This balances ordering and holding costs at about $77.5 each per month. Because Q^* depends on a square root, moderate cost estimation errors have limited effect, but the recommendation should be rounded to carton size and reconsidered if demand is seasonal."

Lesson summary

Concept	Modeling meaning
Continuous optimization	Choose the best value from a continuum, not only among listed alternatives.
EOQ	Balances fixed order cost and holding cost; optimum follows a square-root law.
Second derivative	Confirms whether a critical point is a minimum or maximum.
Gradient	Points in the direction of fastest increase.
Hessian	Uses curvature to classify multivariable critical points.
Lagrange multiplier	Finds constrained optima where level curves are tangent to constraints.
Shadow price	Marginal value of relaxing a constraint.
Numerical optimization	Needed when equations cannot be solved neatly; requires convergence and robustness checks.
Model criticism	The mathematical optimum must be checked against assumptions, data quality, and practical constraints.

Checkpoint: Exit ticket

In three sentences, explain the difference between an unconstrained optimum and a constrained optimum. Then give one example where the calculus optimum is not directly usable because of a practical constraint.

Core practice

1. For $c(T) = 90/T + 30T$, find T^* and verify that $c''(T^*) > 0$.
2. A shop has $d = \$60$, $r = 150$ units/week, and $s = \$0.20$ per unit-week. Find the EOQ cycle length and order quantity.
3. Find and classify the critical point of $f(x, y) = 60x + 80y - x^2 - y^2 - xy$.
4. Maximize $A = xy$ subject to $2x + y = 40$ using Lagrange multipliers.
5. Explain why the point where $f_x = f_y = 0$ might fail to be a global maximum.

Modeling challenge

A school canteen sells a snack. If it orders too little, it loses potential sales; if it orders too much, leftovers have no resale value. A simplified continuous model estimates daily profit as

$$\Pi(q) = 8q - 0.03q^2 - 120 - 0.5 \max(q - 180, 0),$$

where q is the number of snacks prepared. The term $8q - 0.03q^2$ models demand saturation; the fixed cost is 120; the final term penalizes leftovers beyond 180.

1. Explain why this is a continuous optimization approximation, even though snacks are discrete.
2. Optimize $\Pi(q)$ separately on $q \leq 180$ and $q \geq 180$.
3. Determine the best feasible q and interpret.

4. State two practical checks before giving the recommendation to the canteen manager.

Extension practice

1. For $f(x, y) = x^2 - y^2$, show that $(0, 0)$ is a saddle point using both the Hessian test and a contour argument.
2. A container must have volume $V = xyz = 1000$ and rectangular dimensions $x, y, z > 0$. Set up the Lagrange multiplier problem to minimize surface area $2xy + 2xz + 2yz$. What shape do you expect?
3. For the EOQ model, show that a 10% error in d changes T^* by only about 5%. Use either calculus or the square-root relationship.
4. Describe a numerical optimization validation plan for a nonlinear model with many local optima.

Optional appendix: KKT idea for inequalities

For a maximization problem with inequality constraint $g(x) \le c$, the multiplier is active only when the constraint binds. Informally:

$$\lambda \ge 0, \qquad g(x) \le c, \qquad \lambda[g(x) - c] = 0.$$

The last condition says: either the constraint is binding and may have positive value, or it is slack and its shadow price is zero. This is the bridge from Lagrange multipliers to modern nonlinear programming.

Model audit card

Audit before trusting or communicating a model

Purpose	What question or decision does the model support?	**Data**	Which inputs are observed, derived, or simulated?
Assumptions	Which simplifications may change the answer?	**Units**	Are dimensions and scales consistent?
Verification	Were equations, code, constraints, and limiting cases checked?	**Validation**	Is performance tested against independent evidence?
Uncertainty	How do measurement, parameter, solver, and structural errors affect conclusions?	**Ethics**	Who benefits, who bears risk, and what fairness criterion matters?
Reproducibility	Are data provenance, versions, seeds, and transformations recorded?	**Communication**	Are claims conditional, traceable, and decision-relevant?

Further reading

- Nonlinear Constrained Optimization (NEOS Guide)
- Calculus Volume 3 (OpenStax)

Sources, provenance, and licence

These notes follow the iterative modeling and assessment perspective in the SIAM/COMAP GAIMME report. Competition-specific remarks are a snapshot checked on 20 August 2026 against the official COMAP HiMCM instructions; readers should verify current rules before submission. Unless a source is explicitly identified, classroom datasets and numerical scenarios are synthetic or simulated teaching examples and are not empirical claims about a real institution. Third-party references retain their own terms.

Student response guide and checking criteria

For closed calculations, show the definition, substitution, units, and at least one independent check; the worked examples in this unit are method hints, but you must complete the arithmetic yourself. Open modeling responses are checked on five dimensions: a clear purpose and output; stated data and provenance; defensible assumptions and scope; traceable computational verification and model validation; and alignment among uncertainty, limitations, and the recommendation. Support each dimension with concrete evidence rather than writing only "completed."

Lectures in Mathematical Modeling

Lecture 11: Dimensional Analysis and Similitude

Kenneth Sok Kin Cheng

Revised 20 August 2026

Learning goals

By the end of the lesson, students should be able to:

1. explain why **dimensional analysis** reduces experimental complexity;
2. check whether an equation is **dimensionally compatible**;
3. represent physical quantities using the MLT system: **mass**, **length**, and **time**;
4. construct a **dimensionless product** by solving exponent equations;
5. apply **Buckingham's Pi Theorem** to reduce a variable list;
6. use dimensional analysis to derive scaling laws such as wind force, pendulum period, and roasting time;
7. explain **similitude** and why model experiments must match key dimensionless numbers;
8. write a modeling paragraph that states variables, assumptions, Π-groups, validation, and limitations.

1 Why dimensional analysis?

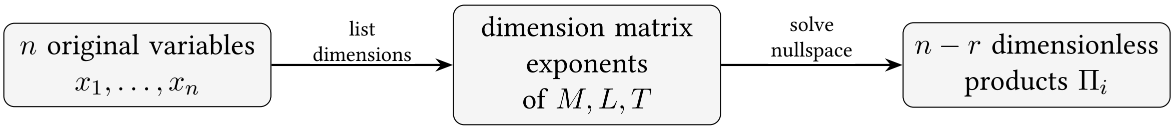
A physical model often begins with a long list of possible variables. If an experimenter studies n variables and tests k values of each, a full grid requires k^n trials. This is the **curse of dimensionality** in experimental design.

Dimensional analysis does not find the whole law by magic. Instead, it says: physical laws cannot depend on whether we measure in metres or feet, kilograms or pounds. Therefore the important relationships can be expressed using *dimensionless* combinations of variables. This can reduce n original variables to far fewer dimensionless groups.

Launch: the experimental burden **[12 min]**

A team wants to study the drag force on a small underwater robot. They suspect that drag depends on velocity v, length L, water density ρ, viscosity μ, and frontal area A.

1. If each variable is tested at 5 values, how many experiments are required in a full grid?
2. Why might this be unrealistic for a school or competition project?
3. What kind of mathematical trick would help reduce the number of experiments?



Experiments vary Π-groups, not every original variable independently.

Figure 1: Dimensional analysis converts a long variable list into a smaller set of dimensionless groups.

2 Dimensions and compatibility

In mechanics, most quantities can be expressed using three primary dimensions:

$$M = \text{mass}, \qquad L = \text{length}, \qquad T = \text{time}.$$

A velocity has dimension LT^{-1}; a force has dimension MLT^{-2}; an energy has dimension ML^2T^{-2}.

Dimensional compatibility

An equation is **dimensionally compatible** if every term being added or subtracted has the same dimension. Products and powers may have different dimensions, but sums must match.
For example, in

$$s = s_0 + v_0 t - \frac{1}{2} g t^2,$$

each term has dimension L. Therefore the equation passes the compatibility check.

Quantity	Dimension	Typical units
Velocity v	LT^{-1}	m/s
Acceleration a	LT^{-2}	m/s^2
Force F	MLT^{-2}	N
Density ρ	ML^{-3}	kg/m^3
Pressure p	$ML^{-1}T^{-2}$	Pa
Dynamic viscosity μ	$ML^{-1}T^{-1}$	Pa s
Energy E	ML^2T^{-2}	J
Power P	ML^2T^{-3}	W

Worked Example 1: catching a wrong model before collecting data

A student proposes the model

$$F = mv + v^2$$

for the force needed to move an object.

$$[F] = MLT^{-2}, \qquad [mv] = M \cdot LT^{-1} = MLT^{-1},$$

while

$$[v^2] = L^2T^{-2}.$$

The two terms on the right cannot be added, and neither has the dimension of force. The model is structurally wrong before any numerical fitting begins.

Modeling lesson: dimensional checks are a cheap error detector. They do not prove a model is correct, but they quickly reject many impossible models.

Checkpoint: Quick compatibility check

For each equation, decide whether it is dimensionally compatible.

1. $E = \frac{1}{2}mv^2$.
2. $p = \rho g h$.
3. $F = \rho A + mv$.
4. $t = 2\pi\sqrt{L/g}$.

3 Constructing dimensionless products

A **dimensionless product** is a product of powers of variables whose total dimension is $M^0L^0T^0$. Suppose variables $x_1, \ldots, x_n$ have dimensions. We search for exponents $a_1, \ldots, a_n$ such that

$$x_1^{a_1}x_2^{a_2}\cdots x_n^{a_n}$$

is dimensionless. Equating powers of M, L, T gives a homogeneous linear system.

Worked Example 2: the simple pendulum

Let the period t of a pendulum depend on length ℓ, mass m, gravity g, and release angle θ. The variables are:

$$t(T), \qquad \ell(L), \qquad m(M), \qquad g(LT^{-2}), \qquad \theta(1).$$

Consider

$$t^a\ell^bm^cg^d\theta^e.$$

Its dimension is

$$T^aL^bM^c(LT^{-2})^d = M^cL^{b+d}T^{a-2d}.$$

For this to be dimensionless,

$$c = 0, \qquad b + d = 0, \qquad a - 2d = 0.$$

There are two free choices. Taking $d = 1$ gives $a = 2$, $b = -1$, $c = 0$, so

$$\Pi_1 = \frac{gt^2}{\ell}.$$

Taking $e = 1$ gives

$$\Pi_2 = \theta.$$

Therefore the period law must have the form

$$\frac{gt^2}{\ell} = H(\theta), \qquad \text{or} \qquad \boxed{t = \sqrt{\frac{\ell}{g}}\,h(\theta).}$$

Interpretation: mass disappears. Dimensional analysis predicts that pendulum period does not depend on the bob's mass.

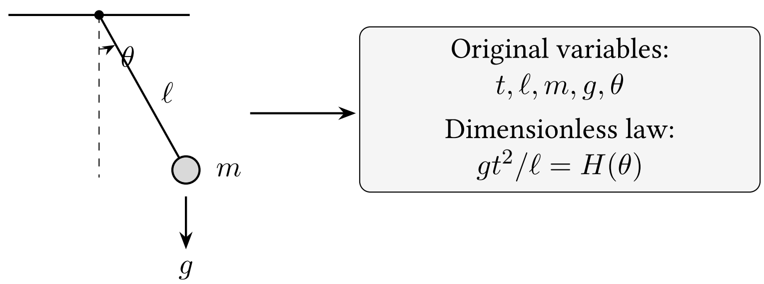


Figure 2: For the pendulum, dimensional analysis removes mass and identifies the correct scale $\sqrt{\ell/g}$.

Worked Example 3: wind force and a missing variable

A first attempt might say wind force on a sign depends only on wind speed v and area A:

$$F = kv^a A^b.$$

But

$$[v^a A^b] = (LT^{-1})^a (L^2)^b = L^{a+2b} T^{-a},$$

which has no mass dimension. Since $[F] = MLT^{-2}$, a mass-bearing variable is missing. Include air density ρ with dimension ML^{-3}:

$$F = k\rho^c v^a A^b.$$

Equating dimensions gives

$$M : 1 = c, \qquad T : -2 = -a \Rightarrow a = 2,$$

$$L : 1 = -3c + a + 2b = -3 + 2 + 2b \Rightarrow b = 1.$$

Thus

$$\boxed{F = k\rho v^2 A.}$$

Interpretation: doubling wind speed quadruples the force; doubling exposed area doubles the force. The dimensionless constant k must be found experimentally.

4 Buckingham's Pi theorem

Buckingham's Pi theorem

If a model uses n dimensional variables and the dimension matrix has rank r, then the physical relationship can be expressed using

$$n - r$$

independent dimensionless products:

$$F(\Pi_1, \Pi_2, \ldots, \Pi_{n-r}) = 0.$$

For most mechanics problems using M, L, T, the rank is often $r = 3$, so the number of groups is usually $n - 3$.

A teachable seven-step workflow

1. State the dependent variable and plausible independent variables.
2. Write the MLT dimension of every variable.
3. Build products of powers and solve for dimensionless combinations.
4. Check that every Π-group is truly dimensionless.
5. Use Buckingham's theorem to write the reduced relationship.
6. Interpret the scaling law and identify what still needs data.
7. Validate using plots of dimensionless variables, not only original variables.

Worked Example 4: projectile range

The horizontal range R of a projectile depends on initial speed v, gravity g, and launch angle θ.

$$R(L), \qquad v(LT^{-1}), \qquad g(LT^{-2}), \qquad \theta(1).$$

There are four variables and rank two here because only L and T appear. Therefore there are $4-2=2$ dimensionless products.
Let

$$R^a v^b g^c \theta^d.$$

Its dimension is

$$L^a(LT^{-1})^b(LT^{-2})^c = L^{a+b+c}T^{-b-2c}.$$

For dimensionless products:

$$a+b+c=0, \qquad -b-2c=0.$$

Choose $a=1$. Then $b=-2$, $c=1$, giving

$$\Pi_1 = \frac{Rg}{v^2}.$$

The angle is already dimensionless, so $\Pi_2=\theta$. Hence

$$\frac{Rg}{v^2} = H(\theta), \qquad \text{or} \qquad \boxed{R = \frac{v^2}{g}H(\theta).}$$

The exact mechanics result $R = v^2\sin(2\theta)/g$ is consistent with the dimensional result. Dimensional analysis finds the scale v^2/g but not the particular function $\sin(2\theta)$.

Model criticism: What dimensional analysis can and cannot do

Dimensional analysis can reveal missing variables, impossible equations, scaling exponents, and useful dimensionless combinations. It cannot determine dimensionless constants or arbitrary functions such as $H(\theta)$. Those require theory, data, or simulation.

5 Scaling laws and model refinement

Worked Example 5: turkey roasting and the $w^{2/3}$ law

Suppose roasting time t depends on a characteristic length ℓ, thermal diffusivity κ with dimension L^2T^{-1}, and a dimensionless temperature ratio. The only way to combine t, ℓ, and κ dimensionlessly is

$$\Pi = \frac{\kappa t}{\ell^2}.$$

Therefore, for fixed oven conditions,

$$t \propto \frac{\ell^2}{\kappa}.$$

If turkeys are geometrically similar, weight w is proportional to volume, so

$$w \propto \ell^3, \qquad \ell \propto w^{1/3}.$$

Thus

$$\boxed{t \propto w^{2/3}.}$$

This differs from the folk rule "minutes per pound", which implies $t \propto w$. The folk rule tends to overcook large turkeys.

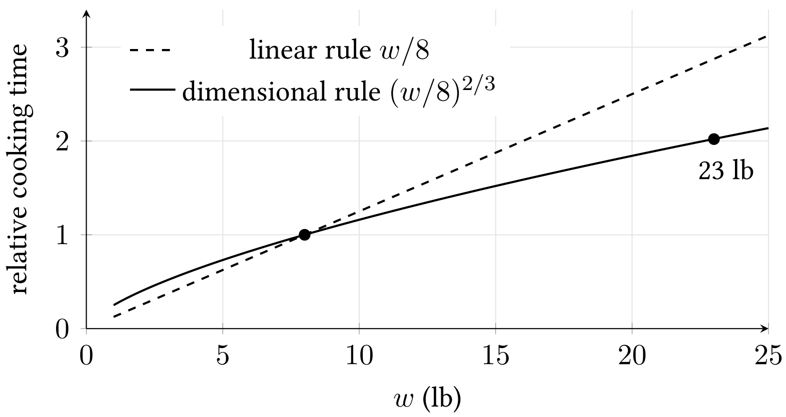


Figure 3: The $w^{2/3}$ law grows more slowly than a linear minutes-per-pound rule.

Worked Example 6: raindrop terminal velocity

Let terminal velocity v depend on radius r, gravity g, air density ρ, and dynamic viscosity μ:

$$v(LT^{-1}), \quad r(L), \quad g(LT^{-2}), \quad \rho(ML^{-3}), \quad \mu(ML^{-1}T^{-1}).$$

There are 5 variables and rank 3, so we expect 2 dimensionless products. One convenient set is

$$\Pi_1 = \frac{v}{\sqrt{rg}}, \qquad \Pi_2 = \frac{r^{3/2}\sqrt{g}\,\rho}{\mu}.$$

Therefore

$$\boxed{v = \sqrt{rg}\,H\left(\frac{r^{3/2}\sqrt{g}\,\rho}{\mu}\right).}$$

Interpretation: terminal velocity is not just proportional to $\sqrt{rg}$; viscosity and density determine the drag regime through the second dimensionless group. This is why small mist droplets, raindrops, and hailstones do not follow exactly the same simple law.

Checkpoint: Scaling interpretation

In the wind-force model $F = k\rho v^2 A$, what happens to force if:

1. wind speed triples while ρ and A are unchanged?
2. a sign is replaced by a geometrically similar sign twice as tall and twice as wide?
3. the same car drives at the same speed at high altitude where air density is lower?

6 Similitude: from model to prototype

Similitude means that a laboratory model and the real prototype have the same relevant dimensionless products. Geometric similarity alone is not enough. A small model submarine is not dynamically similar to a real submarine unless the key force ratios also match.

Classical dimensionless numbers

Name	Formula	Physical meaning
Reynolds number Re	$\frac{vL\rho}{\mu}$	inertia / viscosity
Froude number Fr	$\frac{v^2}{gL}$	inertia / gravity
Mach number Ma	$\frac{v}{c}$	speed / sound speed
Weber number We	$\frac{\rho v^2 L}{\sigma}$	inertia / surface tension
Drag coefficient C_D	$\frac{F_D}{\frac{1}{2}\rho v^2 A}$	dimensionless drag

Worked Example 7: Froude similarity for a spillway model

A 1:20 scale model of a dam spillway is tested in a lab. The prototype flow speed is $v_p = 8$ m/s. Gravity effects dominate, so match the Froude number

$$Fr = \frac{v^2}{gL}.$$

Let $L_m = L_p/20$. Froude similarity requires

$$\frac{v_m^2}{gL_m} = \frac{v_p^2}{gL_p}.$$

Thus

$$v_m = v_p\sqrt{\frac{L_m}{L_p}} = 8\sqrt{\frac{1}{20}} \approx 1.79 \text{ m/s}.$$

Interpretation: model velocity is not $8/20 = 0.4$ m/s. For gravity-controlled free-surface flow, velocity scales with the square root of length.

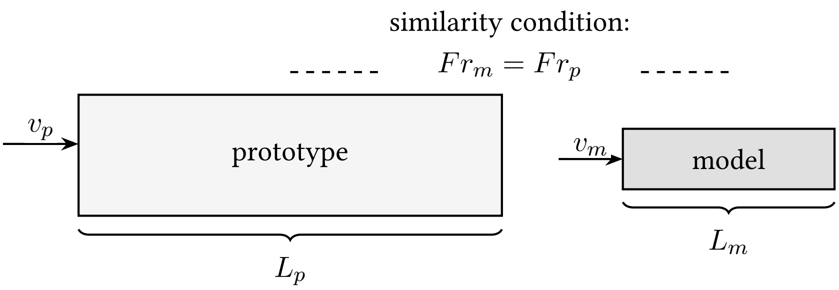


Figure 4: A scaled model predicts a prototype only when the relevant dimensionless numbers match.

Model criticism: Partial similitude

In real experiments, it may be impossible to match all dimensionless numbers at once. A ship model may need Froude similarity for waves and Reynolds similarity for viscosity, but the required model speeds may conflict. A strong report should state which Π-groups were matched, which were not, and how this affects confidence in the prediction.

7 Differentiated modeling studio

Dimensional-analysis studio **[18 min]**

Choose one route.

1. **Core route: projectile range.** Derive $R = (v^2/g)H(\theta)$ using dimensions, then explain what dimensional analysis leaves unknown.
2. **Challenge route: Stokes drag.** Assume drag force F on a slow sphere depends on radius r, speed v, and viscosity μ. Show that $F = k\mu rv$.
3. **Extension route: blast-wave scaling.** Suppose blast radius R depends on energy E, air density ρ, and time t. Derive the scaling law $R \propto (Et^2/\rho)^{1/5}$.

Model criticism: Competition-style communication

A good dimensional-analysis paragraph should include:

1. the dependent variable and the assumed independent variables;
2. the dimension table and the number of resulting Π-groups;
3. the reduced dimensionless relationship;
4. the physical interpretation of the scaling;
5. a validation plan and a limitation.

Weak: "The answer is $t \propto w^{2/3}$."

Stronger: "Assuming geometrically similar turkeys and fixed oven conditions, the cooking time must scale as $t \propto \ell^2/\kappa$. Since weight is proportional to volume, $\ell \propto w^{1/3}$ and hence $t \propto w^{2/3}$. This predicts that large turkeys need less time per pound than small turkeys. The main limitation is that real turkeys are not perfectly similar and ovens do not heat uniformly."

Lesson summary

Concept	Modeling meaning
MLT system	Expresses physical quantities using mass, length, and time.
Compatibility	Additive terms must have the same dimension; a fast model-error check.
Dimensionless product	A product of powers whose total dimension is $M^0L^0T^0$.
Buckingham Π theorem	Reduces n dimensional variables to $n-r$ dimensionless groups.
Scaling law	Reveals how changing size, speed, density, or time changes outcomes.
Similitude	Model and prototype must match relevant Π-groups.
Partial similitude	Match the dominant physics; disclose unmatched effects.
Validation	Fit or compare dimensionless variables, not only raw variables.

Checkpoint: Exit ticket

In three sentences, explain why dimensional analysis is useful before collecting data. Include one example of a dimensionless product.

Practice problems

Core practice

1. Find the dimension of pressure, power, density, viscosity, and surface tension in the MLT system.
2. Check whether $p = \rho gh$, $F = ma$, and $E = mv + gh$ are dimensionally compatible.
3. For a pendulum, verify that gt^2/ℓ is dimensionless.
4. Use dimensional analysis to show that the period of a mass-spring oscillator must have the form $t = C\sqrt{m/k}$, where k is spring stiffness with dimension MT^{-2}.
5. If $F = k\rho v^2 A$, by what factor does F change when v is doubled and A is tripled?
6. For Froude similarity, a 1:25 scale model represents a river flow with prototype speed 5 m/s. Find the model speed.

Modeling challenge

A team wants to design a small-scale experiment for the drag force F_D on a toy car moving through air. They believe F_D depends on speed v, frontal area A, air density ρ, viscosity μ, and characteristic length L.

1. List the dimensions of all variables.
2. Predict how many dimensionless groups should appear.
3. One likely group is the Reynolds number $Re = vL\rho/\mu$. Verify it is dimensionless.

4. Propose a second dimensionless group involving F_D.
5. Explain how the team should use these groups to compare a small model car and a real car.
6. State one limitation of the experiment.

Extension practice

1. **Projectile range.** Derive $R = (v^2/g)H(\theta)$ and verify the exact formula $R = v^2 \sin(2\theta)/g$ is consistent.
2. **Terminal velocity.** Starting with v, r, g, ρ, μ, derive two independent dimensionless products.
3. **Blast wave.** Assume blast radius R depends on energy E, air density ρ, and time t. Show that $R = C(Et^2/\rho)^{1/5}$.
4. **Choosing variables.** A student models cooking time using only turkey weight w. Explain why this variable list is incomplete, then propose a better list.
5. **Partial similitude.** Explain why matching both Reynolds and Froude numbers may be impossible in a small water-channel experiment.

Optional appendix: choosing Π-groups

Different complete sets of Π-groups are possible. For example, $Re = vL\rho/\mu$ and $1/Re$ contain the same information. A good set should be:

- dimensionless;
- independent;
- physically interpretable;
- convenient for plotting and experimentation;
- chosen so the dependent variable appears in only one group when possible.

Model audit card

Audit before trusting or communicating a model

Purpose	What question or decision does the model support?	**Data**	Which inputs are observed, derived, or simulated?
Assumptions	Which simplifications may change the answer?	**Units**	Are dimensions and scales consistent?
Verification	Were equations, code, constraints, and limiting cases checked?	**Validation**	Is performance tested against independent evidence?
Uncertainty	How do measurement, parameter, solver, and structural errors affect conclusions?	**Ethics**	Who benefits, who bears risk, and what fairness criterion matters?
Reproducibility	Are data provenance, versions, seeds, and transformations recorded?	**Communication**	Are claims conditional, traceable, and decision-relevant?

Further reading

- Guide to the SI: Values, quantities, and dimensions (NIST)
- Planetary fact sheet —ratio to Earth values (NASA NSSDCA)

Sources, provenance, and licence

These notes follow the iterative modeling and assessment perspective in the SIAM/COMAP GAIMME report. Competition-specific remarks are a snapshot checked on 20 August 2026 against the official COMAP HiMCM instructions; readers should verify current rules before submission. Unless a source is explicitly identified, classroom datasets and numerical scenarios are synthetic or simulated teaching examples and are not empirical claims about a real institution. Third-party references retain their own terms.

Student response guide and checking criteria

For closed calculations, show the definition, substitution, units, and at least one independent check; the worked examples in this unit are method hints, but you must complete the arithmetic yourself. Open modeling responses are checked on five dimensions: a clear purpose and output; stated data and provenance; defensible assumptions and scope; traceable computational verification and model validation; and alignment among uncertainty, limitations, and the recommendation. Support each dimension with concrete evidence rather than writing only “completed.”

Lectures in Mathematical Modeling

Lecture 12: Graphical and Qualitative Modeling

Kenneth Sok Kin Cheng

Revised 20 August 2026

Learning goals

By the end of the lesson, readers should be able to:

1. use the shape of a graph as a mathematical model when exact equations are unknown;
2. identify and interpret **equilibria** as intersections of response curves;
3. reason from **slope**, **concavity**, and curve shifts to predict policy effects;
4. build a simple **discrete dynamical system** and test stability;
5. explain the economic condition $MR = MC$ and connect it to supply curves;
6. analyze a per-unit tax using supply and demand graphs;
7. write a qualitative modeling conclusion with assumptions, diagrams, and limitations.

1 Graphs as models

A model does not always have to begin with a fully specified formula. In many policy and competition problems, we know qualitative features before we know numbers. A demand curve is decreasing. A cost curve may first flatten and then steepen. A deterrence requirement may increase but with diminishing additional effect. These features are mathematical information.

A **qualitative graphical model** uses the shape of one or more functions to infer behavior. The central questions are:

- Where do curves intersect?
- Is an intersection stable or unstable?
- How does a parameter shift a curve?
- Which conclusion is robust even if the exact formula changes?

Launch: what can a graph tell us? **[12 min]**

Consider a school canteen. Let q be the number of lunch sets prepared and let $C(q)$ be total cost.

1. Sketch a plausible cost curve. Should it be increasing? Should it be linear?
2. Sketch a plausible demand curve between price and quantity demanded.
3. State one conclusion you can make from the shapes, even without exact equations.

Modeling lesson: a graph records assumptions. A line says constant marginal effect; a curved graph says the marginal effect changes.

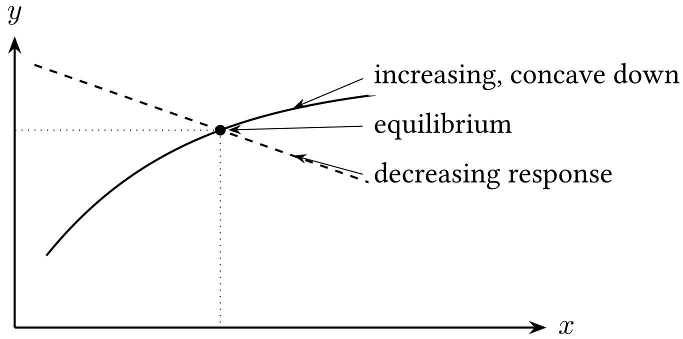


Figure 1: A graphical model turns qualitative assumptions about direction and curvature into predictions about intersections and comparative statics.

Core vocabulary

A **response curve** shows the value one part of a system requires in response to another part. An **equilibrium** is an intersection where all response requirements are simultaneously satisfied. A **comparative-static analysis** asks how the equilibrium changes after a parameter shifts a curve.

2 Case study I: an arms-race graphical model

The arms-race model is not included because of its political content, but because it is an excellent example of qualitative modeling. The modeler does not need the exact military formula. The shape of the response curves already gives useful conclusions.

Let x be the number of weapons held by Country X and y be the number held by Country Y. Suppose

$$y = f(x)$$

means: the minimum number of Y weapons needed to maintain deterrence when X has x weapons. Similarly,

$$x = g(y)$$

means: the minimum number of X weapons needed when Y has y weapons.

Deterrence curve assumptions

For the core lesson we use three qualitative assumptions.

1. f and g are increasing: more weapons by one side require a stronger response by the other.
2. They are concave down: the first enemy weapons matter a lot; later additions may have smaller marginal impact.
3. Their intercepts represent minimum retaliation requirements even if the opponent has very few weapons.

These assumptions are weaker than a full formula, but strong enough to support policy reasoning.

Worked Example 1: Reading the deterrence equilibrium

Suppose the two response curves intersect at (x_m, y_m). At this point, Country X has enough weapons relative to Y, and Country Y has enough weapons relative to X. The point is therefore a mutual deterrence equilibrium.

If the current state lies below $y = f(x)$, then Y judges its arsenal insufficient and builds more weapons.
If the current state lies left of $x = g(y)$, then X judges its arsenal insufficient and builds more weapons.

Repeated responses tend to move the system toward the intersection if each reaction is moderate.

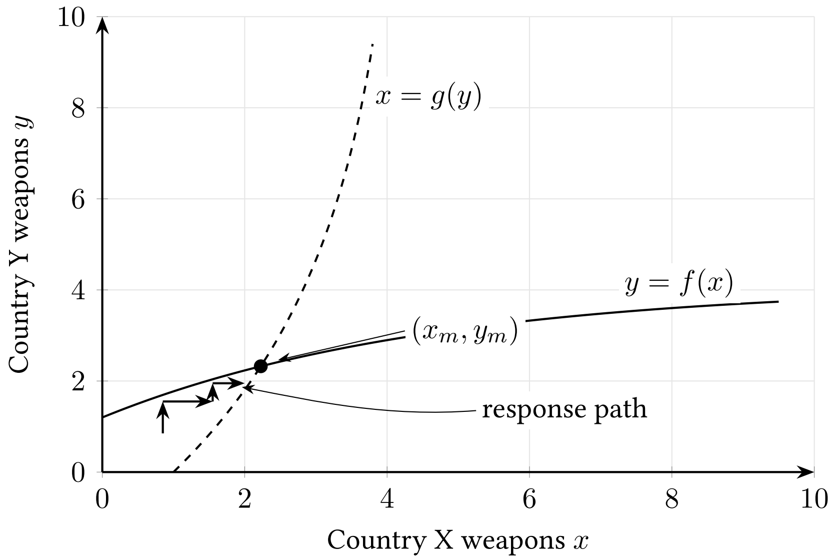


Figure 2: Two response curves determine a mutual deterrence equilibrium. The diagram is qualitative: the conclusions depend on shape and shifts, not exact numerical coordinates.

Worked Example 2: Policy shifts

The graphical model becomes useful when we ask how a policy changes a curve.

1. **Civil defense by X.** If X protects its population better, Y needs more surviving weapons to create the same deterrent effect. Y's response curve $y = f(x)$ shifts upward. The new equilibrium has a larger y, and usually a larger x as X responds to Y's buildup.
2. **Mobile missiles by X.** If X's weapons are harder to destroy, X needs fewer weapons for the same survivability. X's response curve $x = g(y)$ shifts left or becomes flatter. The equilibrium can move downward for both sides.
3. **Multiple warheads.** If one weapon can attack many targets, the exchange ratio changes. The intercept effect may reduce required launchers, but the offensive effect may steepen the response curves. A purely qualitative diagram may become ambiguous; the modeler must state that more data or a more explicit formula is needed.

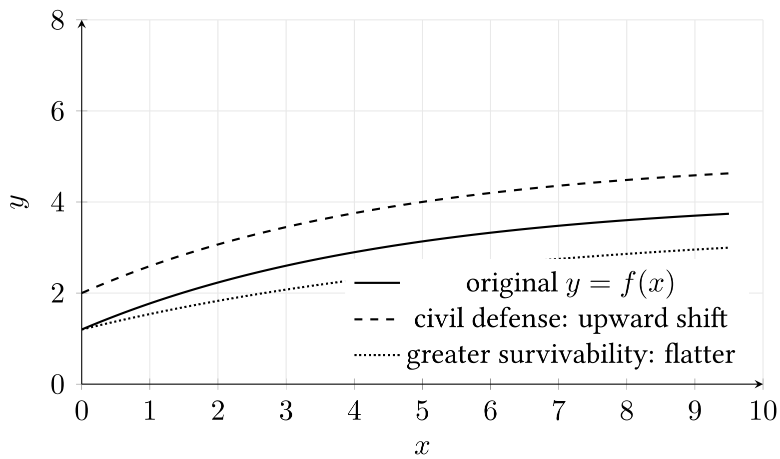


Figure 3: Policy analysis often reduces to curve shifting. The critical question is which parameter changes and in which direction.

Checkpoint: Graphical policy check

A country builds better missile shelters, so a larger fraction of its weapons survive a first strike. Which response curve shifts: its own curve or the opponent's curve? Does the curve move upward, downward, leftward, or become flatter? Explain in two sentences.

3 A discrete arms-race update

A graph gives qualitative direction. A **discrete dynamical system** gives a sequence of numerical updates.

Worked Example 3: A staged arms-race model

Suppose Country Y believes it needs 120 weapons even when X has none, and then adds one additional weapon for every two weapons X has. Country X believes it needs 60 weapons even when Y has none, and then adds one additional weapon for every three weapons Y has:

$$y_{n+1} = 120 + \frac{1}{2}x_n, \qquad x_{n+1} = 60 + \frac{1}{3}y_n.$$

At equilibrium,

$$y^* = 120 + \frac{1}{2}x^*, \qquad x^* = 60 + \frac{1}{3}y^*.$$

Substitute the second equation into the first:

$$y^* = 120 + \frac{1}{2}\left(60 + \frac{1}{3}y^*\right) = 150 + \frac{1}{6}y^*.$$

Thus $\frac{5}{6}y^* = 150$, so

$$\boxed{y^* = 180}, \qquad \boxed{x^* = 120}.$$

The product of reaction coefficients is $\frac{1}{2} \cdot \frac{1}{3} = \frac{1}{6} < 1$, so repeated updates converge to the equilibrium.

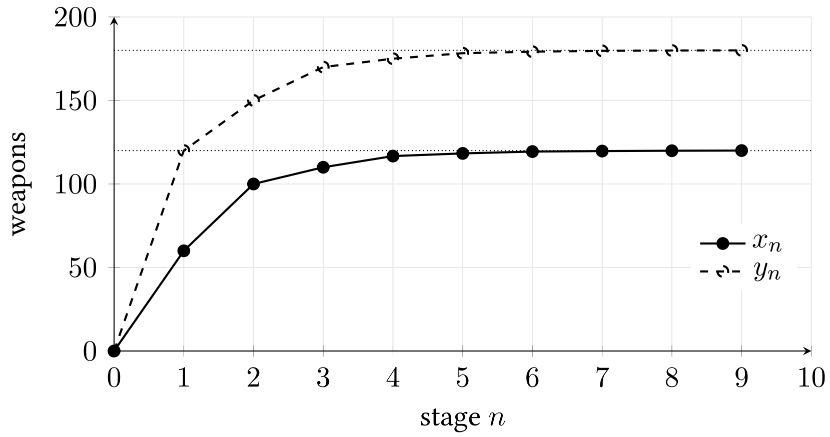


Figure 4: When the product of reaction coefficients is below one, the staged arms-race model converges geometrically to equilibrium.

Model criticism: Model criticism: what does the update hide?

The update equations assume that each side reacts only to the opponent's previous stockpile, with fixed coefficients and no negotiation, budget limit, domestic politics, or uncertainty. A good modeling report should present the recurrence as a simplified warning model, not as a complete theory of political behavior.

4 Case study II: firm behavior and $MR = MC$

Graphical modeling also explains economic decisions. Let q be output quantity. A competitive firm accepts the market price p. Its **marginal revenue** is therefore $MR = p$. Its **marginal cost** is $MC(q) = TC'(q)$.

Profit-maximizing rule

Profit is
$$\Pi(q) = TR(q) - TC(q).$$
The first-order condition is
$$\Pi'(q) = MR(q) - MC(q) = 0.$$
For a competitive firm, $MR = p$, so the firm produces where
$$\boxed{MR = MC.}$$
The firm should produce more while $MR > MC$, and stop increasing output once the next unit costs more than it earns.

Worked Example 4: A firm's optimal output

A small producer sells at price $p = 10$ dollars per unit and has total cost
$$TC(q) = 100 + 2q + 0.1q^2.$$
Then
$$TR(q) = 10q, \qquad \Pi(q) = 10q - (100 + 2q + 0.1q^2) = 8q - 0.1q^2 - 100.$$
The marginal cost is
$$MC(q) = TC'(q) = 2 + 0.2q.$$
Set $MR = MC$:
$$10 = 2 + 0.2q \quad \Rightarrow \quad q^* = 40.$$
The maximum profit is
$$\Pi(40) = 8(40) - 0.1(40)^2 - 100 = 60.$$
The graph shows the same conclusion: produce until the horizontal price line meets the marginal-cost curve.

Checkpoint: Economic interpretation

If a per-unit tax of 2 dollars is imposed on the producer, the marginal cost becomes $MC(q) = 4+0.2q$. What is the new optimal output? Explain why the output falls.

5 Gasoline tax: supply, demand, and tax incidence

Now move from one firm to a market. The **supply curve** gives the price at which producers are willing to supply quantity q. The **demand curve** gives the price at which consumers are willing to buy quantity q. The intersection is the **market equilibrium**.

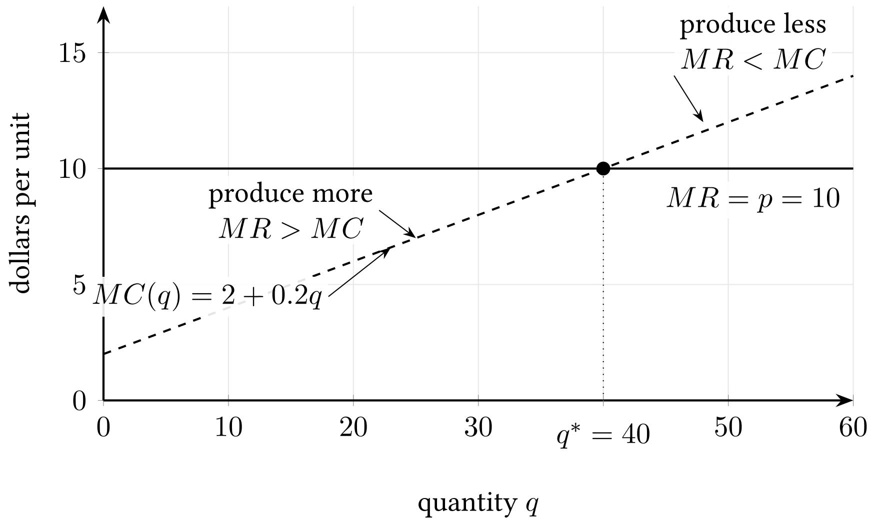


Figure 5: For a competitive firm, the profit-maximizing output occurs where price equals marginal cost.

A per-unit tax shifts the supply curve upward by the tax amount. The price paid by consumers rises, the quantity falls, and the tax burden is shared between consumers and producers.

Worked Example 5: Linear tax incidence

Suppose

$$S(q) = 1 + 0.5q, \qquad D(q) = 5 - 0.5q.$$

The original equilibrium solves

$$1 + 0.5q = 5 - 0.5q \quad \Rightarrow \quad q^* = 4, \qquad p^* = 3.$$

A per-unit tax $\tau = 1$ shifts supply to

$$S_1(q) = S(q) + 1 = 2 + 0.5q.$$

The new equilibrium solves

$$2 + 0.5q = 5 - 0.5q \quad \Rightarrow \quad q_1 = 3, \qquad p_1 = 3.5.$$

Consumers pay $p_1 - p^* = 0.5$ more per unit. Producers receive $p_1 - \tau = 2.5$, which is 0.5 less than before. The one-dollar tax is split equally in this symmetric example.

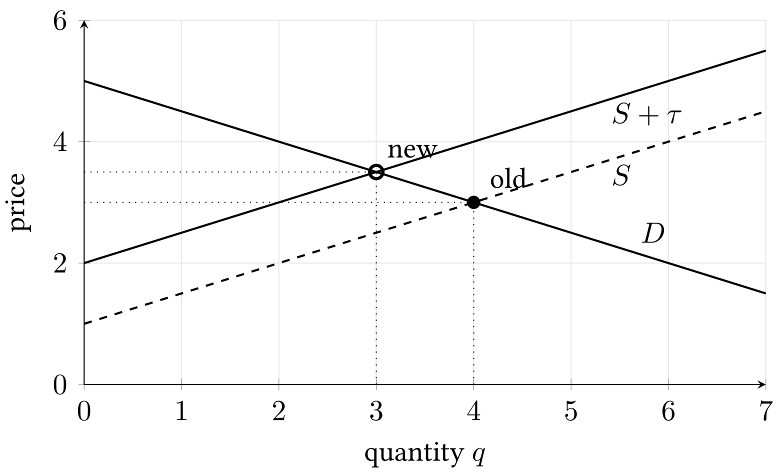


Figure 6: A per-unit tax shifts the supply curve upward. Price rises by less than the tax; quantity falls.

Tax incidence and elasticity

The side of the market with less flexibility bears more of the tax. If demand is steep, consumers cannot reduce quantity easily, so consumers pay most of the tax. If demand is flat, consumers can switch away easily, so producers absorb more of the tax.

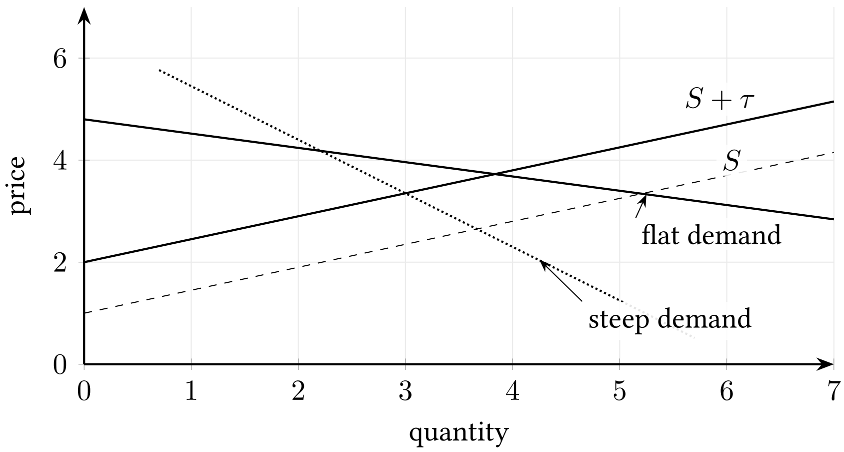


Figure 7: The same tax has different effects depending on the slope of demand. This is a qualitative conclusion that does not require exact functional forms.

Model criticism: Short-run versus long-run gasoline demand

Immediately after an oil shock, demand is often inelastic: people still need to commute, cars cannot be replaced instantly, and public transport may be unavailable. A gasoline tax then raises prices but may reduce quantity only slightly. Over several years, demand becomes more elastic as people buy efficient cars, change routes, carpool, or use public transport. The same tax can then reduce consumption more effectively. The policy conclusion depends on the time horizon.

6 Differentiated studio: choosing a graph-based model

Differentiated modeling studio **[18 min]**

Choose one route.

1. **Core route: tax graph.** Draw supply and demand. Shift supply upward by a tax. Label old and new equilibria. State who pays more when demand is steep.

2. **Challenge route: staged update.** For

$$y_{n+1} = 200 + \frac{1}{3}x_n, \qquad x_{n+1} = 80 + \frac{1}{4}y_n,$$

find the equilibrium and test stability.

3. **Extension route: curve-shift argument.** Explain, using deterrence curves, why increasing survivability may reduce equilibrium weapons, while increasing offensive exchange ratio may increase them.

Choosing a graph-based modeling tool

Situation	Graphical tool	Question answered
Two sides respond to each other	Response curves	Where is mutual satisfaction?
Repeated strategic reaction	Difference equations	Does the process converge?
Firm chooses output	MR and MC curves	What quantity maximizes profit?
Market clears	Supply and demand	What price and quantity occur?
Policy intervention	Curve shift	How does equilibrium change?
Unknown exact formula	Shape and slope reasoning	Which conclusion is robust?

7 Writing qualitative-model conclusions

Model criticism: Competition-style communication

A qualitative graphical model should not sound vague. A strong paragraph should include:

1. **Variables**: what the axes represent.
2. **Shape assumptions**: increasing/decreasing, concavity, intercepts, or stability.
3. **Mechanism**: why those shapes are reasonable.
4. **Conclusion**: what happens to the equilibrium after a policy change.
5. **Limitation**: what cannot be decided without data.

Weak conclusion: "The tax makes price go up."
Stronger conclusion: "A per-unit gasoline tax shifts the supply curve upward. The new equilibrium has higher consumer price and lower quantity. Since short-run gasoline demand is steep, most of the tax burden is borne by consumers and quantity falls only modestly; therefore the tax is poor as an immediate conservation policy but may be more effective in the long run when demand becomes more elastic."

Lesson summary

Concept	Modeling meaning
Qualitative graphical model	Uses direction, curvature, and intersections when exact equations are unknown.
Response curve	Shows what one part of a system requires in response to another.
Equilibrium	Intersection where all requirements are satisfied simultaneously.
Curve shift	Represents a change in policy, technology, cost, or preference.
Staged arms race	Difference-equation version of response curves; stable if reaction product is below one.
$MR = MC$	Competitive firm produces until the next unit costs what it earns.
Supply-demand equilibrium	Intersection of willingness to supply and willingness to buy.
Tax incidence	Tax burden falls more on the less flexible side of the market.
Model criticism	State shape assumptions and identify when a conclusion is only qualitative.

Checkpoint: Exit ticket

In three sentences, explain why graphical modeling can be useful even when exact formulas are unknown. Use either the arms-race model or the gasoline-tax model as your example.

Core practice

1. A response curve $y = f(x)$ shifts upward. Explain what happens to the intersection with an unchanged increasing curve $x = g(y)$, using a diagram.
2. For $y_{n+1} = 50 + 0.4x_n$, $x_{n+1} = 30 + 0.2y_n$, find the equilibrium and test stability.
3. A firm has $TC(q) = 200 + 4q + 0.05q^2$ and sells at price $p = 12$. Find q^* and maximum profit.
4. Supply and demand are $S(q) = 2 + 0.3q$, $D(q) = 8 - 0.7q$. Find equilibrium. Then impose a tax $\tau = 1$ and find the new equilibrium.
5. Draw two demand curves, one steep and one flat. For the same supply shift, explain which case gives a larger quantity reduction.

Modeling challenge

A city wants to reduce private-car use near a school. It can either impose a parking fee, improve bus service, or redesign the school gate area. Build a qualitative graphical model.

1. Define the axes for your graph. Possible choices: cost of driving vs number of cars, waiting time vs number of students, or supply-demand for parking spaces.
2. State the shape assumptions for each curve.
3. Show how one policy shifts one curve.

4. Predict the direction of change in equilibrium.
5. Write a short recommendation and state what data would be needed to make the prediction quantitative.

Extension practice

1. For the staged arms-race system

$$y_{n+1} = a + bx_n, \qquad x_{n+1} = c + dy_n,$$

derive the equilibrium and show that convergence requires $|bd| < 1$.
2. For linear supply and demand $S(q) = \alpha + \beta q$, $D(q) = \gamma - \delta q$, derive the consumer share of a unit tax τ.
3. Construct a graphical argument for why a subsidy to public transport may reduce gasoline use more in the long run than in the short run.
4. Explain one limitation of the arms-race response-curve model that cannot be fixed by drawing a more detailed graph.

Optional appendix: smooth deterrence formula

One possible analytic form for a deterrence curve is

$$y = \frac{y_0}{s^{x/y}}, \qquad 0 < s < 1.$$

It says that if X can target Y's missiles more times, Y needs more weapons to retain enough survivors. The formula is less important than its qualitative properties: it is increasing and concave down. In a student report, this formula should be introduced only after the assumptions have been explained in words.

Model audit card

Audit before trusting or communicating a model

Purpose	What question or decision does the model support?	**Data**	Which inputs are observed, derived, or simulated?
Assumptions	Which simplifications may change the answer?	**Units**	Are dimensions and scales consistent?
Verification	Were equations, code, constraints, and limiting cases checked?	**Validation**	Is performance tested against independent evidence?
Uncertainty	How do measurement, parameter, solver, and structural errors affect conclusions?	**Ethics**	Who benefits, who bears risk, and what fairness criterion matters?
Reproducibility	Are data provenance, versions, seeds, and transformations recorded?	**Communication**	Are claims conditional, traceable, and decision-relevant?

Modern modeling bridge

Contemporary modeling often joins several traditions rather than choosing a single equation. Network models represent relational structure; agent-based models expose heterogeneous local rules; robust optimization protects decisions against uncertain inputs; causal models distinguish intervention from association; and regularized prediction controls complexity when candidate features are numerous. The same cycle still applies: define purpose, represent mechanisms and evidence, verify computation, validate for the intended use, quantify uncertainty, and revise. Interactive examples are available through the modeling laboratories at skcKenneth.github.io.

Further reading

- Principles of Economics 3e (OpenStax)
- GAIMME: Guidelines for Assessment and Instruction in Mathematical Modeling Education (SIAM / COMAP)

Sources, provenance, and licence

These notes follow the iterative modeling and assessment perspective in the SIAM/COMAP GAIMME report. Competition-specific remarks are a snapshot checked on 20 August 2026 against the official COMAP HiMCM instructions; readers should verify current rules before submission. Unless a source is explicitly identified, classroom datasets and numerical scenarios are synthetic or simulated teaching examples and are not empirical claims about a real institution. Third-party references retain their own terms.

Student response guide and checking criteria

For closed calculations, show the definition, substitution, units, and at least one independent check; the worked examples in this unit are method hints, but you must complete the arithmetic yourself. Open modeling responses are checked on five dimensions: a clear purpose and output; stated data and provenance; defensible assumptions and scope; traceable computational verification and model validation; and alignment among uncertainty, limitations, and the recommendation. Support each dimension with concrete evidence rather than writing only "completed."

Supplementary Notes for Mathematical Modeling
Supplementary 1: Sets, Functions, and Mathematical Language

Kenneth Sok Kin Cheng

Revised 20 August 2026

Learning goals

By the end of this supplementary lesson, readers should be able to:

1. use the language of **sets**, **elements**, **subsets**, **union**, **intersection**, and **complement**;
2. decide whether a relation is a **function** by checking whether each input has exactly one output;
3. determine the **domain** and **range** of common real functions;
4. interpret a graph using the **vertical line test**;
5. compute sums, products, quotients, and **compositions** of functions with attention to domain restrictions;
6. explain why precise mathematical language matters in mathematical modeling.

1 Why this supplement matters for modeling

A mathematical model begins before any equation is written. We must decide what objects are under discussion, which values are allowed, what counts as an input, and what output is produced. These decisions are expressed using the language of sets and functions.

Launch: define the modeling universe **[10 min]**

A school wants to model student lunch waiting time. Discuss:

1. What is the **universal set** of people under discussion: all students, all diners, or everyone on campus?
2. What input could define a function: arrival time, queue length, meal type, or payment method?
3. What output should the function return: waiting time, service time, total time, or probability of waiting more than 10 minutes?

Modeling lesson: Vague sets create vague models. A good model states the population, variables, and valid inputs explicitly.

2 Sets as modeling categories

Core definitions

A **set** is a collection of objects. The objects are called **elements**. If x is an element of A, write $x \in A$; if not, write $x \notin A$.
If every element of A is also in B, then A is a **subset** of B, written $A \subseteq B$. If $A \subseteq B$ but $A \neq B$, then A is a **proper subset**, written $A \subsetneq B$.

The **empty set** is written $\varnothing$. It contains no elements.

Worked Example 1: Classifying students using sets

Let U be the set of students in a modeling club. Define

$$A = \{\text{students who know Python}\}, \qquad B = \{\text{students who have joined a modeling contest}\}.$$

Then:

- $A \cap B$ is the set of students who both know Python and have contest experience.
- $A \cup B$ is the set of students who know Python or have contest experience, or both.
- A^c is the set of students in U who do not know Python.
- $B - A$ is the set of students with contest experience but without Python experience.

These are not abstract symbols only; they are categories a teacher may use to form balanced teams.

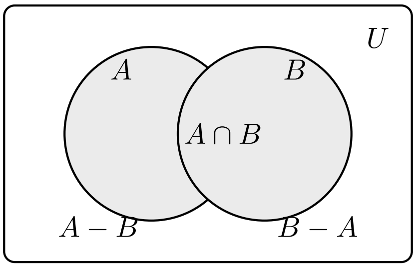


Figure 1: A Venn diagram is a compact way to reason about categories, overlap, and exclusions.

Set operations

Given sets A and B inside a universal set U,

$$A \cup B = \{x \mid x \in A \text{ or } x \in B\}, \qquad A \cap B = \{x \mid x \in A \text{ and } x \in B\},$$

$$B - A = \{x \mid x \in B \text{ and } x \notin A\}, \qquad A^c = U - A.$$

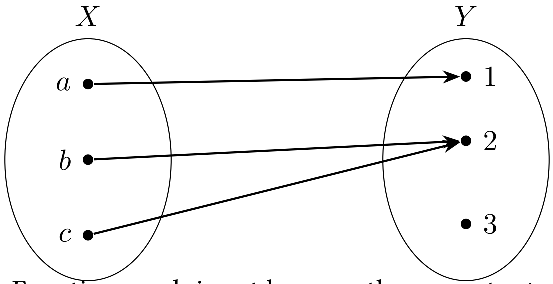
The word "or" in mathematics is usually inclusive: $x \in A \cup B$ allows x to be in both sets.

Checkpoint: Set-language check

Let $U = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$, $A = \{2, 4, 6, 8, 10\}$, and $B = \{1, 2, 3, 5, 8\}$. Find $A \cap B$, $A \cup B$, $B - A$, and A^c. Then write one sentence interpreting each set if A means "even-numbered students" and B means "students selected for a pilot test."

3 Relations and functions

A **relation** from X to Y is any set of ordered pairs (x, y) with $x \in X$ and $y \in Y$. A **function** is a special relation: each input is paired with at most one output. In most calculus contexts, we require every input in the stated domain to have exactly one output.



Function: each input has exactly one output.
Different inputs may share the same output.

Figure 2: A many-to-one relation can still be a function. The forbidden pattern is one input producing two different outputs.

Worked Example 2: Is it a function?

Let $X = \{1, 2, 3\}$ and $Y = \{a, b, c\}$.

1. $R_1 = \{(1, a), (2, a), (3, b)\}$ is a function from X to Y because each input has exactly one output.
2. $R_2 = \{(1, a), (1, b), (2, c)\}$ is not a function because input 1 has two different outputs.
3. $R_3 = \{(1, a), (2, b)\}$ is a function on domain $\{1, 2\}$, but not a function with domain X if every element of X must be used.

Modeling interpretation: If the input is "student ID", then a model for "class level" should not output two different class levels for the same student.

Domain, codomain, and range

For a function $f : X \to Y$,

- the **domain** is the input set where the function is defined;
- the **codomain** is the declared output universe Y;
- the **range** is the set of outputs that actually occur.

Symbolically,

$$\text{Dom}(f) = \{x \mid f(x) \text{ is defined}\}, \qquad \text{Range}(f) = \{f(x) \mid x \in \text{Dom}(f)\}.$$

4 Real functions, domain, and graph

In calculus and modeling, a common case is a **real function**, where both inputs and outputs are real numbers. If a formula is given without a stated domain, the implied domain is the largest set of real numbers for which the formula makes sense.

Worked Example 3: Domain and range from a formula

Consider

$$f(x) = \sqrt{x - 2}.$$

For a real-valued output, the expression inside the square root must be non-negative:

$$x - 2 \geq 0 \quad \Rightarrow \quad x \geq 2.$$

Thus

$$\text{Dom}(f) = [2, \infty), \qquad \text{Range}(f) = [0, \infty).$$

The starting point of the graph is $(2, 0)$.

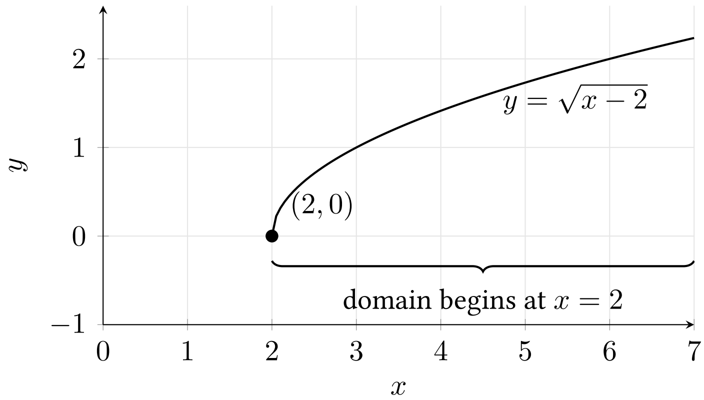


Figure 3: The graph of $f(x) = \sqrt{x-2}$ begins only where the formula produces real values.

Vertical line test

A curve is the graph of a function of x if every vertical line intersects it at most once. This is the geometric version of the rule: one input, one output.

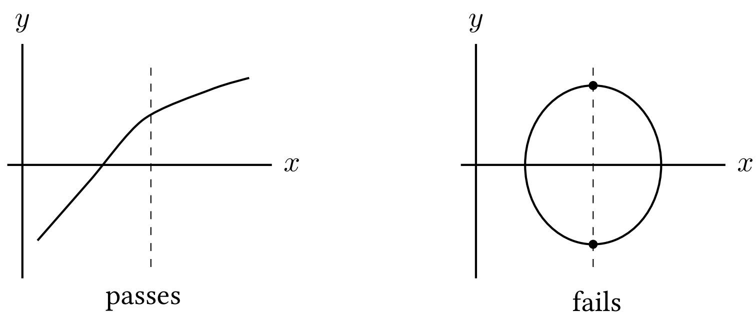


Figure 4: The curve on the left represents a function; the circle on the right does not, because one x gives two y values.

Checkpoint: Domain check

Find the domain of each real function:

1. $f(x) = \dfrac{1}{\sqrt{x-2}}$;
2. $g(x) = \sqrt{9-x^2}$;
3. $h(x) = \dfrac{x+1}{x^2-4}$.

For each, state the restriction in words before writing the interval or set notation.

5 Operations and composition

Functions can be added, subtracted, multiplied, divided, and composed. These operations are useful because many models are built from submodels.

Algebra of functions

If f has domain A and g has domain B, then the sum, difference, and product are defined on $A \cap B$:

$$(f+g)(x) = f(x) + g(x), \quad (f-g)(x) = f(x) - g(x), \quad (fg)(x) = f(x)g(x).$$

The quotient is defined where both functions are defined and $g(x) \neq 0$:

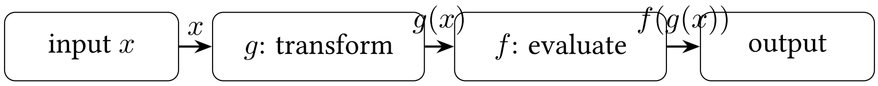
$$\left(\frac{f}{g}\right)(x) = \frac{f(x)}{g(x)}.$$

Composition

The **composite function** $f \circ g$ is

$$(f \circ g)(x) = f(g(x)).$$

The input x must first be valid for g, and then the output $g(x)$ must be valid as an input for f.



A composition is a modeling pipeline: output from one step becomes input to the next.

Figure 5: Composition represents a sequence of transformations.

Worked Example 4: Composition and domain

Let

$$f(u) = \sqrt{u}, \qquad g(x) = 4 - x^2.$$

Then

$$(f \circ g)(x) = f(4 - x^2) = \sqrt{4 - x^2}.$$

The domain is not all real numbers. Since the square root requires $4 - x^2 \geq 0$,

$$-2 \leq x \leq 2.$$

Therefore

$$\text{Dom}(f \circ g) = [-2, 2].$$

Modeling interpretation: If $g(x)$ computes a predicted quantity that must be non-negative before another formula can use it, domain checking prevents meaningless outputs.

Worked Example 5: Iterating a function

Let

$$f(x) = \frac{1}{1 + x}.$$

Compute $f(f(f(2)))$:

$$f(2) = \frac{1}{3}, \qquad f(f(2)) = f\left(\frac{1}{3}\right) = \frac{1}{1 + 1/3} = \frac{3}{4},$$

$$f(f(f(2))) = f\left(\frac{3}{4}\right) = \frac{1}{1 + 3/4} = \frac{4}{7}.$$

Iteration is important in modeling because a difference equation is just repeated function application:

$$x_{n+1} = f(x_n).$$

Model criticism: Common mistake: ignoring the domain

A formula may look algebraically correct but fail as a real function. For example,

$$\frac{1}{\sqrt{x - |x|}}$$

is never defined for any real x: if $x \geq 0$, then $x - |x| = 0$ and the denominator is zero; if $x < 0$, then $x - |x| = 2x < 0$, so the square root is not real.

6 Short extension: induction and binomial coefficients

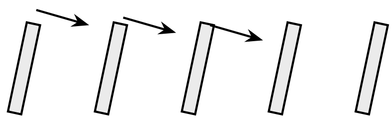
This supplementary note originally contains induction and binomial coefficients. They are valuable, but in a two-hour class they should be treated as a short extension or homework pathway. They support discrete models, counting, probability, and binomial distributions later in modeling.

Mathematical induction

To prove a statement $P(n)$ for all positive integers n:

1. Prove the **base case**: $P(1)$ is true.
2. Prove the **inductive step**: if $P(k)$ is true, then $P(k+1)$ is true.

Then $P(n)$ is true for every $n \in \mathbb{Z}^+$.



Base case starts the chain; inductive step passes truth from one case to the next.

Figure 6: Induction is a logical chain reaction, not a numerical pattern check.

Worked Example 6: Sum of the first n integers

Prove

$$1 + 2 + \cdots + n = \frac{n(n+1)}{2}.$$

Base case: for $n = 1$, left side $= 1$ and right side $= 1(2)/2 = 1$.
Inductive step: assume

$$1 + 2 + \cdots + k = \frac{k(k+1)}{2}.$$

Then

$$1 + 2 + \cdots + k + (k+1) = \frac{k(k+1)}{2} + (k+1) = \frac{(k+1)(k+2)}{2}.$$

This is exactly the formula for $n = k + 1$. Hence the formula holds for all $n \in \mathbb{Z}^+$.

Binomial coefficient

The **factorial** is

$$n! = n(n-1)\cdots 2 \cdot 1, \qquad 0! = 1.$$

The **binomial coefficient**

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

counts the number of ways to choose k objects from n objects. The **binomial theorem** is

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^{n-k} b^k.$$

Checkpoint: Extension check

1. Compute $\binom{8}{3}$ and explain what it counts.
2. Expand $(1+x)^4$ using the binomial theorem.
3. Explain why $\sum_{k=0}^{n} \binom{n}{k} = 2^n$ by interpreting both sides as counting subsets of an n-element set.

7 Using these ideas in modeling reports

Model criticism: Mathematical-language checklist

When writing a model, check the following:

1. What is the universal set or population under study?
2. What is the domain of each function or variable?
3. Are there values that must be excluded, such as negative quantities or zero denominators?
4. Does each input determine exactly one output?
5. If two formulas are composed, is the output of the first formula valid as the input of the second?
6. Does the graph match the claimed function and domain?

Lesson summary

Concept	Modeling meaning
Set	Defines the population, category, or collection being modeled.
Subset	Describes a narrower category within a larger modeling universe.
Union/intersection	Combines or overlaps categories, useful for grouping and filtering data.
Complement	Represents everything in the universe not satisfying a condition.
Function	A deterministic input-output rule; each input has one output.
Domain	States which input values are valid and meaningful.
Range	States which output values the model can actually produce.
Graph	Visual representation of ordered pairs $(x, f(x))$.
Composition	Represents a sequence of model transformations.
Induction/binomial	Supports discrete reasoning, counting, and probability models.

Checkpoint: Exit ticket

A school models total cafeteria time by

$$T(n) = \sqrt{25 - n} + \frac{60}{n},$$

where n is the number of open service counters.

1. What real-number restrictions does the formula impose on n?
2. What additional real-world restrictions should be imposed?
3. Why is domain checking important before optimizing this model?

Core practice

1. Let $U = \{1, 2, \ldots, 12\}$, $A = \{2, 4, 6, 8, 10, 12\}$, and $B = \{3, 6, 9, 12\}$. Find $A \cup B$, $A \cap B$, $A - B$, and $(A \cup B)^c$.
2. Decide whether each relation is a function: (a) $\{(1, 2), (2, 2), (3, 4)\}$; (b) $\{(1, 2), (1, 3), (2, 4)\}$; (c) $\{(x, y) \in \mathbb{R}^2 : y = x^2\}$; (d) $\{(x, y) \in \mathbb{R}^2 : x = y^2\}$.
3. Find the domain of each real function: $f(x) = \sqrt{x + 5}$, $g(x) = 1/(x^2 - 9)$, $h(x) = \sqrt{4 - x}/(x + 1)$.
4. Let $f(x) = x^2 + 1$ and $g(x) = 2x - 3$. Find $f \circ g$, $g \circ f$, $(f + g)(x)$, and $(fg)(x)$.
5. Use induction to prove $1 + 3 + 5 + \cdots + (2n - 1) = n^2$.

Answer check: core practice

For the set problem,

$$A \cup B = \{2, 3, 4, 6, 8, 9, 10, 12\}, \qquad A \cap B = \{6, 12\},$$
$$A - B = \{2, 4, 8, 10\}, \qquad (A \cup B)^c = \{1, 5, 7, 11\}.$$

The four relations are, respectively, a function, not a function, a function, and not a function. The domains are $[-5, \infty)$, $\mathbb{R} \setminus \{-3, 3\}$, and $(-\infty, -1) \cup (-1, 4]$. For the given f and g,

$$f \circ g = (2x - 3)^2 + 1, \quad g \circ f = 2x^2 - 1, \quad f + g = x^2 + 2x - 2, \quad fg = (x^2 + 1)(2x - 3).$$

For induction, verify the base case and show that the statement for k implies the statement for $k + 1$.

Modeling challenge

A school wants to form modeling teams. Let U be all students in the training program. Define sets for students who know Python, students who are strong in writing, students with previous contest experience, and students available on Saturday.

1. Define four sets using clear notation.
2. Write an expression for the set of students who know Python and are available on Saturday.
3. Write an expression for students who have writing strength but no previous contest experience.
4. Suppose each student is assigned exactly one role: coder, writer, analyst, or presenter. Is "student $\mapsto$ role" a function? Explain.
5. Suppose each student may have several strengths. Is "student $\mapsto$ strength" a function? Explain.
6. Write a short paragraph explaining how this set-and-function language can help a teacher form balanced teams.

Answer check: modeling challenge

A complete response uses correct set notation and states all membership conditions. "Student $\mapsto$ role" is a function exactly when every student has one role; "student $\mapsto$ strength" is not single-valued unless the codomain is redesigned as a set of strengths. The recommendation should connect team balance to the stated availability and skill constraints.

Extension practice

1. Prove by induction that

$$1^2 + 2^2 + \cdots + n^2 = \frac{n(n + 1)(2n + 1)}{6}.$$

2. Prove that $\binom{n}{0} + \binom{n}{1} + \cdots + \binom{n}{n} = 2^n$ using the binomial theorem and then explain it using subsets.
3. Let $f(x) = 1/(1 + x)$. Find a formula for the first five iterates $f(1)$, $f^{(2)}(1)$, ..., $f^{(5)}(1)$. What pattern do you notice?
4. Find the domain of $F(x) = \sqrt{\dfrac{x - 1}{x + 2}}$. Give your answer in interval notation and justify it with a sign chart.

Answer check: extension practice

Use the induction hypothesis for the sum of squares and the binomial theorem at $1+1$ for the subset identity. The first five iterates are

$$\frac{1}{2}, \quad \frac{2}{3}, \quad \frac{3}{5}, \quad \frac{5}{8}, \quad \frac{8}{13}.$$

For the final expression, the domain is

$$(-\infty, -2) \cup [1, \infty).$$

Model audit card

Audit before trusting or communicating a model

Purpose	What question or decision does the model support?	**Data**	Which inputs are observed, derived, or simulated?
Assumptions	Which simplifications may change the answer?	**Units**	Are dimensions and scales consistent?
Verification	Were equations, code, constraints, and limiting cases checked?	**Validation**	Is performance tested against independent evidence?
Uncertainty	How do measurement, parameter, solver, and structural errors affect conclusions?	**Ethics**	Who benefits, who bears risk, and what fairness criterion matters?
Reproducibility	Are data provenance, versions, seeds, and transformations recorded?	**Communication**	Are claims conditional, traceable, and decision-relevant?

Further reading

- Precalculus 2e (OpenStax)

Sources, provenance, and licence

These notes follow the iterative modeling and assessment perspective in the SIAM/COMAP GAIMME report. Competition-specific remarks are a snapshot checked on 20 August 2026 against the official COMAP HiMCM instructions; readers should verify current rules before submission. Unless a source is explicitly identified, classroom datasets and numerical scenarios are synthetic or simulated teaching examples and are not empirical claims about a real institution. Third-party references retain their own terms.

Student response guide and checking criteria

For closed calculations, show the definition, substitution, units, and at least one independent check; the worked examples in this unit are method hints, but you must complete the arithmetic yourself. Open modeling responses are checked on five dimensions: a clear purpose and output; stated data and provenance; defensible assumptions and scope; traceable computational verification and model validation; and alignment among uncertainty, limitations, and the recommendation. Support each dimension with concrete evidence rather than writing only “completed.”

Lectures in Mathematical Modeling

Supplementary Two: Numbers, Coordinates, and Geometric Models

Kenneth Sok Kin Cheng

Revised 20 August 2026

Learning goals

By the end of this supplementary lesson, readers should be able to:

1. distinguish **rational numbers**, **irrational numbers**, and **real numbers**;
2. interpret **absolute value** as distance and use it to describe tolerances or errors;
3. represent intervals on the **real number line**;
4. use the coordinate plane to model location, distance, midpoint, and coverage;
5. derive and interpret equations of lines and circles;
6. connect slope, distance, and geometry to data fitting, routing, optimization, and modeling constraints.

1 Why numbers and coordinates matter in modeling

A mathematical model turns a real situation into objects that can be computed, compared, and communicated. Before calculus, optimization, or simulation can be used, students need a precise language for three simple ideas:

$$\text{quantity}, \qquad \text{location}, \qquad \text{distance}.$$

Numbers describe quantities, coordinates describe location, and distance formulas describe closeness, error, travel, or coverage. This supplementary note is therefore not a separate algebra review; it is the coordinate language behind many modeling tasks.

Launch: choosing the right number system **[10 min]**

For each modeling quantity, decide whether integers, rational numbers, or real numbers are the most natural language. Give one sentence of explanation.

1. Number of students assigned to a team.
2. Proportion of students who chose bus travel.
3. Distance between two locations on a map.
4. Error between an observed value and a predicted value.
5. Number of delivery robots required for a school route.

Modeling lesson: Choosing the wrong number system can create wrong decisions. For example, a linear program may output 2.7 robots, but the actual decision must be an integer.

2 Number systems: from counting to the real line

Core vocabulary

Symbol	Name	Modeling meaning
$\mathbb{N}$	natural numbers	counting objects: $1, 2, 3, \ldots$
$\mathbb{Z}$	integers	signed counts or net changes: $\ldots, -2, -1, 0, 1, 2, \ldots$
$\mathbb{Q}$	rational numbers	ratios, fractions, exact proportions: m/n with $n \neq 0$
$\mathbb{R}$	real numbers	continuous quantities: length, time, mass, temperature, error

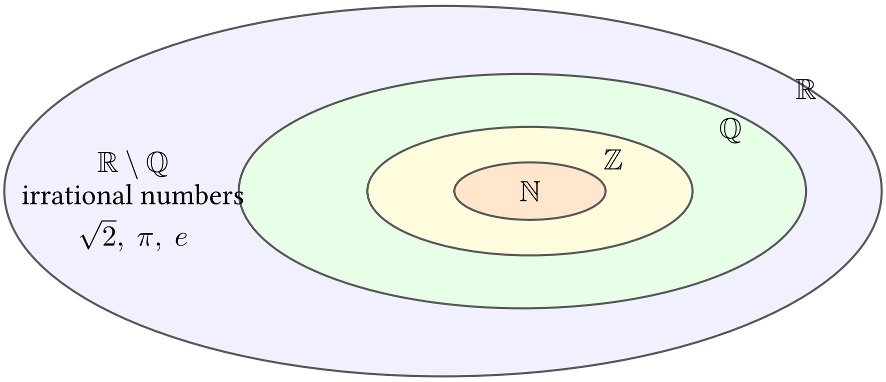


Figure 1: The hierarchy of number systems. The natural numbers are contained in the integers, the integers are contained in the rational numbers, and the rational numbers are contained in the real numbers. The irrational numbers are the real numbers that are not rational.

Worked Example 1: Classifying modeling quantities

Classify each quantity.

1. A school needs 7 buses. This is in $\mathbb{N}$ because it counts objects.
2. A team completed $\frac{3}{4}$ of its survey. This is in $\mathbb{Q}$ because it is a ratio of integers.
3. The diagonal of a 1×1 square is $\sqrt{2}$. This is irrational, hence real but not rational.
4. A fitted parameter $k = 0.03716\ldots$ from regression is usually treated as real.

Modeling interpretation: Some models are continuous approximations of discrete decisions. That is allowed, but the report must explain how the continuous result is converted into a feasible decision.

Rational numbers and the gap problem

A **rational number** is a number of the form m/n, where $m, n \in \mathbb{Z}$ and $n \neq 0$. Rational numbers are useful for exact proportions, but they do not fill the number line completely. For example, there is no rational number whose square is 2.

$$\sqrt{2} \notin \mathbb{Q}, \qquad \sqrt{2} \in \mathbb{R}.$$

The real number system is needed because geometry, distance, limits, and calculus naturally produce irrational values.

Worked Example 2: Why $\sqrt{2}$ is not rational

Suppose $\sqrt{2} = m/n$ in lowest terms. Then

$$m^2 = 2n^2.$$

Thus m^2 is even, so m is even. Write $m = 2k$. Then

$$4k^2 = 2n^2 \quad \Rightarrow \quad n^2 = 2k^2,$$

so n is also even. This contradicts the assumption that m/n was in lowest terms. Hence $\sqrt{2}$ is irrational.
Modeling meaning: Even the distance between $(0,0)$ and $(1,1)$ is irrational. A coordinate model therefore requires real numbers, not only fractions.

Checkpoint: Number-system check

A route-planning model returns a shortest distance of $5\sqrt{5}$ km, a travel time of 11.2 minutes, and a need for 3.4 vans. Which quantities are continuous model outputs? Which must be converted into integer decisions?

3 Absolute value, distance, and intervals

Absolute value as distance

For a real number a,

$$|a| = \begin{cases} a, & a \geq 0, \\ -a, & a < 0. \end{cases}$$

Geometrically, $|a|$ is the distance from a to 0 on the number line. More generally, $|x - a|$ is the distance between x and a.

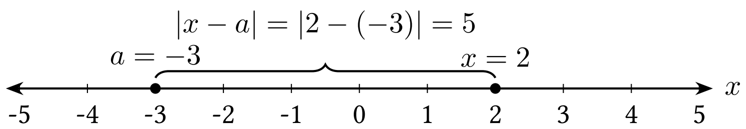


Figure 2: Absolute value measures distance on the real line.

Worked Example 3: Measurement tolerance

A temperature sensor is acceptable if its reading x is less than 0.4°C from the true value 25°C. Write this condition using absolute value and interval notation.

$$|x - 25| < 0.4.$$

This means

$$24.6 < x < 25.4,$$

so the acceptable interval is $(24.6, 25.4)$.
Modeling interpretation: Absolute value inequalities are a compact way to state tolerance, error bounds, residual thresholds, and safety margins.

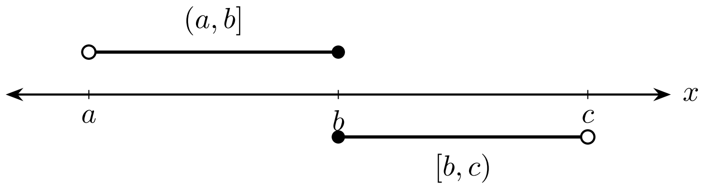


Figure 3: Open endpoints are drawn as hollow circles; closed endpoints are filled.

Checkpoint: Interval check

Rewrite each condition in interval notation.

1. $|x - 6| \leq 2$.
2. A residual r satisfies $|r| < 0.1$.
3. A model parameter k is positive but less than 1.

4 Coordinates, distance, and midpoint

A point in the plane is represented by an ordered pair (x, y). In modeling, coordinates may represent a physical map, a feature space, or two variables measured from data.

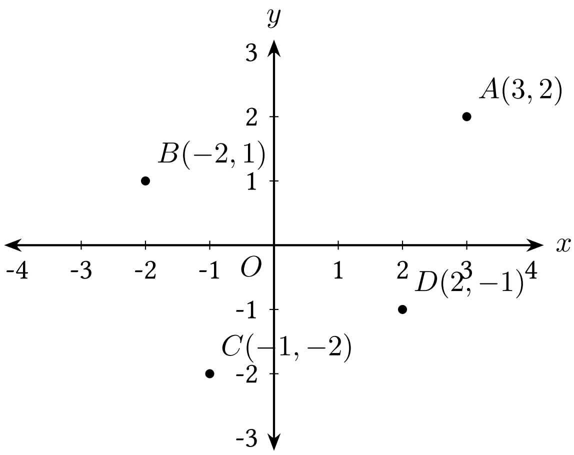


Figure 4: Coordinates locate points relative to the x- and y-axes.

Distance and midpoint formulas

For $P_1 = (x_1, y_1)$ and $P_2 = (x_2, y_2)$,

$$d(P_1, P_2) = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}.$$

The midpoint is

$$M = \left(\frac{x_1 + x_2}{2}, \frac{y_1 + y_2}{2} \right).$$

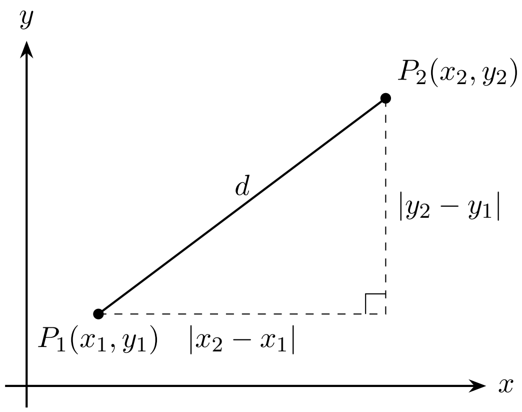


Figure 5: The distance formula is the Pythagorean theorem applied to coordinate differences.

Worked Example 4: Choosing a meeting point

Two students live at $A = (1, 2)$ and $B = (7, 6)$ on a simplified map measured in km. A fair meeting point halfway between them is

$$M = \left(\frac{1+7}{2}, \frac{2+6}{2}\right) = (4, 4).$$

The direct distance between them is

$$d(A, B) = \sqrt{(7-1)^2 + (6-2)^2} = \sqrt{36+16} = \sqrt{52} \approx 7.21 \text{ km}.$$

Modeling interpretation: The midpoint is fair under straight-line distance. If actual walking routes follow roads, the Euclidean midpoint may no longer be fair.

Model criticism: Model criticism: Euclidean distance versus real travel distance

The distance formula assumes a flat coordinate plane and straight-line movement. In route planning, road networks, stairs, slopes, traffic lights, rivers, gates, and restricted areas may make Euclidean distance a poor proxy. A modeler should say whether Euclidean distance is a reasonable approximation or only a first estimate.

5 Lines as simple models

A line is the simplest model of constant rate of change.

$$\text{slope} = \frac{\text{change in output}}{\text{change in input}} = \frac{\Delta y}{\Delta x}.$$

In modeling, slope carries units. If x is time in minutes and y is distance in metres, then slope is metres per minute.

Forms of a line

Form	Equation	Use
Slope-intercept	$y = mx + b$	slope and starting value are known
Point-slope	$y - y_1 = m(x - x_1)$	one point and slope are known
Two-point	$m = \dfrac{y_2 - y_1}{x_2 - x_1}$	two observations are known
General form	$Ax + By + C = 0$	useful for point-to-line distance

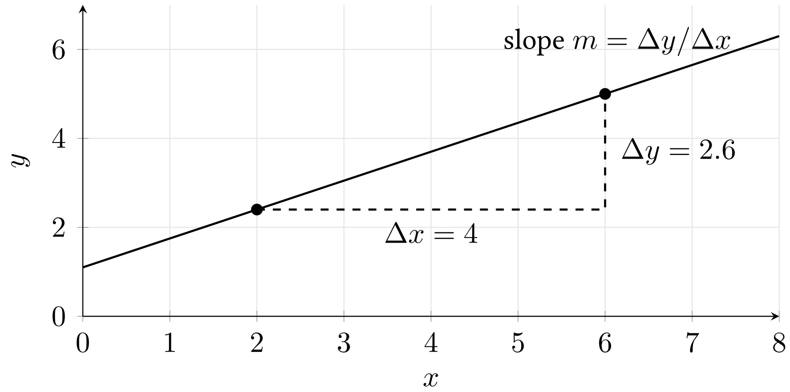


Figure 6: The slope of a line measures constant change per unit input.

Worked Example 5: A linear travel-time model

A delivery robot takes 4 minutes to travel 120 m and 10 minutes to travel 300 m along a corridor. Let x be distance in metres and y be travel time in minutes. Assuming a linear model,

$$m = \frac{10 - 4}{300 - 120} = \frac{6}{180} = \frac{1}{30} \text{ min/m.}$$

Using point-slope form through $(120, 4)$,

$$y - 4 = \frac{1}{30}(x - 120),$$

so

$$y = \frac{x}{30}.$$

The intercept is 0, suggesting that the robot has negligible fixed start-up delay in this simplified case. **Interpretation:** The speed is the reciprocal of the slope: 30 m/min. If later data show a nonzero intercept, that may represent loading time or waiting time.

Checkpoint: Line check

Find the line through $(2, 5)$ and $(8, 17)$. Interpret the slope if x is the number of hours studied and y is a test score.

Parallel, perpendicular, and point-to-line distance

For non-vertical lines, parallel lines have the same slope. Perpendicular lines have slopes satisfying

$$m_1 m_2 = -1.$$

The distance from (x_0, y_0) to the line $Ax + By + C = 0$ is

$$d = \frac{|Ax_0 + By_0 + C|}{\sqrt{A^2 + B^2}}.$$

This formula measures the shortest perpendicular distance from a point to a linear boundary.

Worked Example 6: Distance to a safety boundary

A restricted area is approximated by the line

$$3x - 4y + 12 = 0.$$

A drone is at $P = (2, 5)$. Its shortest distance to the boundary is

$$d = \frac{|3(2) - 4(5) + 12|}{\sqrt{3^2 + (-4)^2}} = \frac{|6 - 20 + 12|}{5} = \frac{2}{5} = 0.4.$$

If coordinates are measured in kilometres, the drone is 0.4 km from the boundary.

6 Circles as distance constraints

A circle is a set of points at a fixed distance from a centre. This makes circles useful for coverage, range, safety, and service-radius models.

Equation of a circle

A circle with centre (h, k) and radius r has equation

$$(x - h)^2 + (y - k)^2 = r^2.$$

The general form

$$x^2 + y^2 + Dx + Ey + F = 0$$

can be converted to standard form by completing the square.

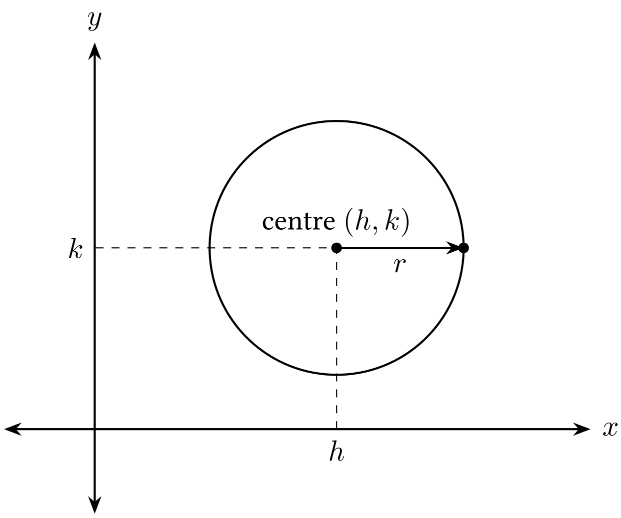


Figure 7: A circle models all locations within a fixed distance from a centre.

Worked Example 7: Coverage radius

A Wi-Fi router is located at $(3, -2)$ and has effective range 5 m. The coverage boundary is

$$(x-3)^2 + (y+2)^2 = 25.$$

A device at $(7, 1)$ has squared distance

$$(7-3)^2 + (1+2)^2 = 16 + 9 = 25,$$

so it lies exactly on the boundary.
Modeling interpretation: In real life, walls and interference make the boundary non-circular. The circle is a first approximation based on distance alone.

Worked Example 8: Completing the square

Find the centre and radius of

$$x^2 + y^2 - 6x + 4y - 12 = 0.$$

Group terms and complete the square:

$$(x^2 - 6x) + (y^2 + 4y) = 12,$$

$$(x-3)^2 + (y+2)^2 = 12 + 9 + 4 = 25.$$

Hence the centre is $(3, -2)$ and the radius is 5.

Differentiated geometric-modeling studio **[13 min]**

Choose one route.

1. **Core route: facility distance.** A help desk is at $(2, 1)$. Find which of $A = (5, 5)$, $B = (-1, 3)$, and $C = (4, -2)$ is closest.
2. **Challenge route: coverage model.** A sensor at $(4, 3)$ covers radius 3. Write the coverage inequality and test whether $(6, 5)$ and $(8, 3)$ are covered.
3. **Extension route: boundary distance.** A road is modeled by $2x - y + 1 = 0$. A school at $(3, 4)$ must be at least 1 km from the road. Check whether the location is acceptable and suggest a direction to move.

7 Optional appendix: completeness and why calculus works

Completeness in one page

A set $S \subset \mathbb{R}$ is **bounded above** if there is a number M with $x \leq M$ for all $x \in S$. The **supremum** is the least upper bound.
The real numbers satisfy the **completeness axiom**: every nonempty subset of $\mathbb{R}$ that is bounded above has a supremum in $\mathbb{R}$.
This is deeper than what is needed for ordinary coordinate geometry, but it explains why limits and calculus are possible. Rational numbers have gaps; real numbers fill those gaps.

Model criticism: Competition-style communication

When using coordinates in a report, do not merely write formulas. State:

1. what the axes represent and what units are used;
2. why Euclidean distance is acceptable, or what it ignores;
3. whether variables are continuous or must be rounded to integers;
4. how distance, slope, or circles support the final recommendation.

Weak statement: "The distance is 5."
Stronger statement: "Using a coordinate map measured in kilometres, the Euclidean distance between the station and Building A is 5 km. This is a lower-bound estimate because actual walking routes must follow campus roads."

Lesson summary

Concept	Modeling meaning
$\mathbb{N}, \mathbb{Z}, \mathbb{Q}, \mathbb{R}$	Different number systems support counting, ratios, and continuous measurement.
Irrational number	Natural geometric quantities such as $\sqrt{2}$ cannot be expressed as fractions.
Absolute value	Measures distance from zero or error from a target.
Interval	Describes acceptable ranges, constraints, and uncertainty bands.
Coordinate plane	Converts location into computable ordered pairs.
Distance formula	Measures straight-line closeness, travel lower bounds, and error.
Slope	Constant rate of change; units matter.
Line	Simplest model of linear relationship or boundary.
Circle	Set of points at fixed distance; useful for coverage and range models.
Completeness	The hidden property of $\mathbb{R}$ that supports limits and calculus.

Checkpoint: Exit ticket

A school map uses coordinates in metres. In three sentences, explain how you would model whether a new water station covers all classrooms within 80 m. Mention the centre, the radius, and one limitation of the model.

Core practice

1. Classify each number as natural, integer, rational, irrational, or real: $-4, 0, 7/3, \sqrt{9}, \sqrt{7}, \pi$.
2. Solve and write in interval notation: $|x-4| < 1.5$ and $|2x+1| \leq 5$.
3. Find the distance and midpoint between $A = (-2, 3)$ and $B = (4, -1)$.
4. Find the equation of the line through $(1, 5)$ and $(4, 11)$.
5. Find the line through $(2, -1)$ perpendicular to $y = 3x + 4$.

6. Find the centre and radius of $x^2 + y^2 + 2x - 8y + 8 = 0$.

Answer check: core practice

Using $\mathbb{N} = \{1, 2, 3, \ldots\}$, the classifications follow directly from the set definitions. The remaining numerical answers are

$$(2.5, 5.5), \quad [-3, 2]; \qquad d(A, B) = 2\sqrt{13}, \quad M = (1, 1);$$

$$y = 2x + 3; \qquad y + 1 = -\frac{1}{3}(x - 2); \qquad \text{circle centre } (-1, 4),\ r = 3.$$

For the induction item, check the base case and then show that the statement for $n = k$ implies the statement for $n = k + 1$.

Modeling challenge

A campus planner wants to place a temporary information booth. Buildings are represented by points

$$A = (0, 0), \quad B = (6, 2), \quad C = (3, 7), \quad D = (-2, 4).$$

1. Compute the midpoint of A and B. Explain why this may or may not be a good booth location.
2. Find the distance from the proposed booth $P = (2, 3)$ to each building.
3. If the booth can serve students within radius 5, write the coverage circle and determine which buildings are covered.
4. A main walkway is approximated by the line $x + y = 5$. Find the distance from P to the walkway.
5. Write a short recommendation. Include one limitation of using coordinates and Euclidean distance.

Answer check: modeling challenge

The midpoint of A and B is $(3, 1)$. From $P = (2, 3)$, the four distances are

$$\sqrt{13}, \quad \sqrt{17}, \quad \sqrt{17}, \quad \sqrt{17}.$$

The coverage disk is $(x - 2)^2 + (y - 3)^2 \leq 25$, which covers all four buildings. The distance from P to $x + y = 5$ is 0. A sound recommendation should also note that straight-line distance can underestimate travel distance along a walkway network.

Extension practice

1. Prove that the product of a nonzero rational number and an irrational number is irrational.
2. Prove the triangle inequality $|a + b| \leq |a| + |b|$ using the fact that $-|a| \leq a \leq |a|$.
3. A set $S = (0, 1)$ has no maximum but has supremum 1. Explain the difference between maximum and supremum.
4. Find all points on the x-axis that are equidistant from $(2, 3)$ and $(-1, 4)$.
5. A circle passes through $(0, 0)$, $(4, 0)$, and $(0, 3)$. Find its equation.

Answer check: extension practice

Use contradiction for the rational-times-irrational claim and the bounds $-|a| \le a \le |a|$ for the triangle inequality. The set $(0,1)$ has supremum 1 but no maximum because $1 \notin (0,1)$. The equidistant point on the x-axis has $x = -2/3$. The circle through the three given points is

$$(x-2)^2 + \left(y - \frac{3}{2}\right)^2 = \frac{25}{4}.$$

Model audit card

Audit before trusting or communicating a model

Purpose	What question or decision does the model support?	**Data**	Which inputs are observed, derived, or simulated?
Assumptions	Which simplifications may change the answer?	**Units**	Are dimensions and scales consistent?
Verification	Were equations, code, constraints, and limiting cases checked?	**Validation**	Is performance tested against independent evidence?
Uncertainty	How do measurement, parameter, solver, and structural errors affect conclusions?	**Ethics**	Who benefits, who bears risk, and what fairness criterion matters?
Reproducibility	Are data provenance, versions, seeds, and transformations recorded?	**Communication**	Are claims conditional, traceable, and decision-relevant?

Further reading

- Precalculus 2e (OpenStax)
- Guide to the SI: Values, quantities, and dimensions (NIST)

Sources, provenance, and licence

These notes follow the iterative modeling and assessment perspective in the SIAM/COMAP GAIMME report. Competition-specific remarks are a snapshot checked on 20 August 2026 against the official COMAP HiMCM instructions; readers should verify current rules before submission. Unless a source is explicitly identified, classroom datasets and numerical scenarios are synthetic or simulated teaching examples and are not empirical claims about a real institution. Third-party references retain their own terms.

Student response guide and checking criteria

For closed calculations, show the definition, substitution, units, and at least one independent check; the worked examples in this unit are method hints, but you must complete the arithmetic yourself. Open modeling responses are checked on five dimensions: a clear purpose and output; stated data and provenance; defensible assumptions and scope; traceable computational verification and model validation; and alignment among uncertainty, limitations, and the recommendation. Support each dimension with concrete evidence rather than writing only “completed.”

Lectures in Mathematical Modeling

Supplementary 3: Graphs, Transformations, and Periodic Models

Kenneth Sok Kin Cheng

Revised 20 August 2026

Learning goals

By the end of this supplementary lesson, readers should be able to:

1. read a **graph** as a model of a relationship between two quantities;
2. distinguish a **relation** from a **function** using the vertical-line test;
3. determine **domain**, **range**, intercepts, symmetry, and key features from formulas and graphs;
4. use graph transformations to sketch related functions without re-plotting from scratch;
5. interpret elementary graph families as modeling shapes: linear, quadratic, reciprocal, absolute value, square-root, and sinusoidal;
6. build a simple **periodic model** using amplitude, period, phase shift, and vertical shift;
7. explain what a graph reveals, what it hides, and what assumptions are needed before using it for prediction.

1 Graphs as modeling language

A graph is not merely a picture of a formula. In mathematical modeling, a graph is a compact way to communicate a relationship: increase, decrease, saturation, threshold, periodicity, outliers, and breakdown of assumptions can often be seen before they are calculated.

Launch: what can a graph say? **[12 min]**

A school records the waiting time at a canteen from 12:00 to 12:40. The curve rises quickly, reaches a peak, then slowly returns to nearly zero.

1. What does the peak mean in context?
2. What would a flatter curve mean?
3. What information is missing if we only know the curve shape but not the units?
4. What intervention could move the peak downward or earlier?

Modeling lesson: a graph should be read as a statement about a system, not only as a drawing.

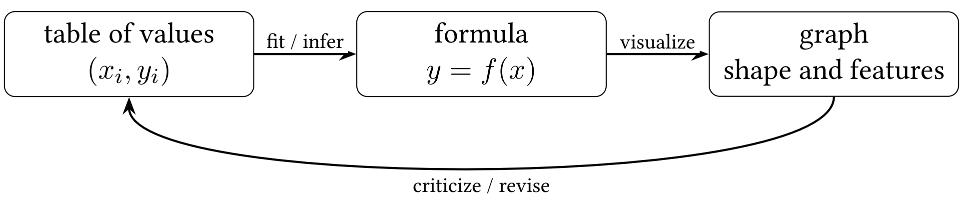


Figure 1: In modeling, data, formula, and graph should support one another.

Core vocabulary

A **relation** is a set of ordered pairs (x, y). A **function** is a special relation in which each input x corresponds to exactly one output y. The **domain** is the set of allowed inputs; the **range** is the set of outputs produced. The graph of $y = f(x)$ is the set of all points $(x, f(x))$.

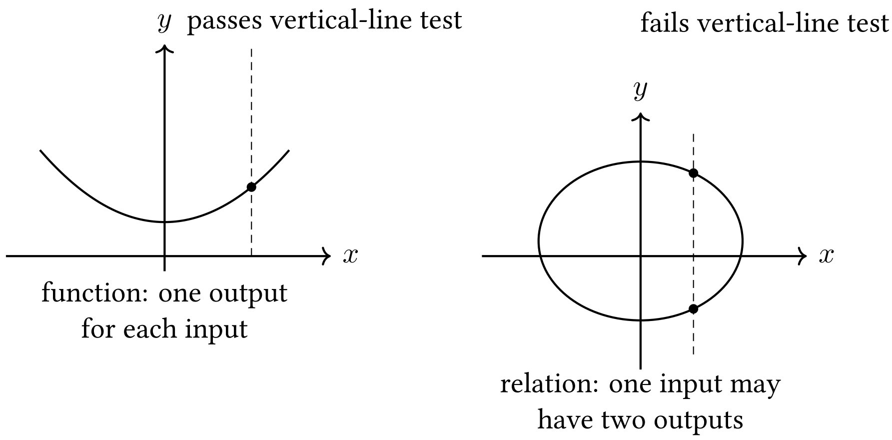


Figure 2: The vertical-line test checks whether a relation can be written as $y = f(x)$.

2 Domain, range, intercepts, and symmetry

Before fitting or interpreting a graph, a modeler must decide which values are meaningful. In school mathematics, a formula may have a natural algebraic domain. In modeling, the realistic domain may be smaller.

Worked Example 1: Natural domain versus modeling domain

A simplified stopping-distance model is

$$D(v) = 0.75v + 0.006v^2,$$

where v is speed in km/h and D is stopping distance in metres.
Algebraic domain: the formula accepts all real v. **Modeling domain**: speed cannot be negative, and the model was calibrated only for ordinary road speeds, say $0 \le v \le 120$.
The y-intercept is $D(0) = 0$, meaning no stopping distance when speed is zero. There is no meaningful negative-speed branch, even though the algebraic formula could be evaluated there. The graph should therefore be drawn only on the modeling domain.

Worked Example 2: Domain and range from a formula

Find the natural domain and range of

$$f(x) = \sqrt{4 - x^2}.$$

The square root requires $4 - x^2 \ge 0$, so $-2 \le x \le 2$. Hence

$$\mathrm{Dom}(f) = [-2, 2].$$

Since $f(x)$ is the upper half of the circle $x^2 + y^2 = 4$, the output satisfies $0 \le y \le 2$:

$$\mathrm{Range}(f) = [0, 2].$$

This is a function because each x in $[-2, 2]$ gives exactly one nonnegative y.

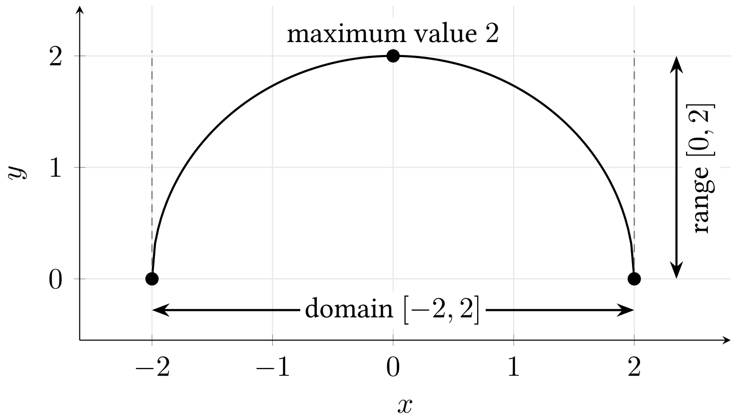


Figure 3: The graph of $y = \sqrt{4 - x^2}$ is the upper semicircle of radius 2, so its domain and range are visible.

Checkpoint: Feature check

For each function, state the natural domain, one intercept, and whether the graph is even, odd, or neither.

1. $f(x) = x^4 - 3x^2$
2. $g(x) = \dfrac{1}{x-2}$
3. $h(x) = \sqrt{x+1}$

3 Graph transformations and standard shapes

Graph transformations allow us to sketch related functions quickly. This is important in modeling because fitted or theoretical curves often differ from a standard graph only by scaling or shifting.

Transformations of $y = f(x)$

New function	Effect on the graph
$y = f(x) + c$	shift upward by c
$y = f(x - c)$	shift right by c
$y = af(x)$	vertical stretch by factor $\lvert a \rvert$; reflect in x-axis if $a < 0$
$y = f(bx)$	horizontal compression by factor b if $b > 1$
$y = -f(x)$	reflect in the x-axis
$y = f(-x)$	reflect in the y-axis

Worked Example 3: Transforming a graph

Describe the graph of

$$y = -2\sqrt{x+3} + 4.$$

Start with $y = \sqrt{x}$.

1. $\sqrt{x+3}$ shifts the graph left by 3.
2. $-2\sqrt{x+3}$ reflects it in the x-axis and stretches vertically by factor 2.

3. $-2\sqrt{x+3}+4$ shifts the graph upward by 4.
The starting point moves from $(0,0)$ to $(-3,4)$. The domain is $[-3,\infty)$ and the range is $(-\infty,4]$.

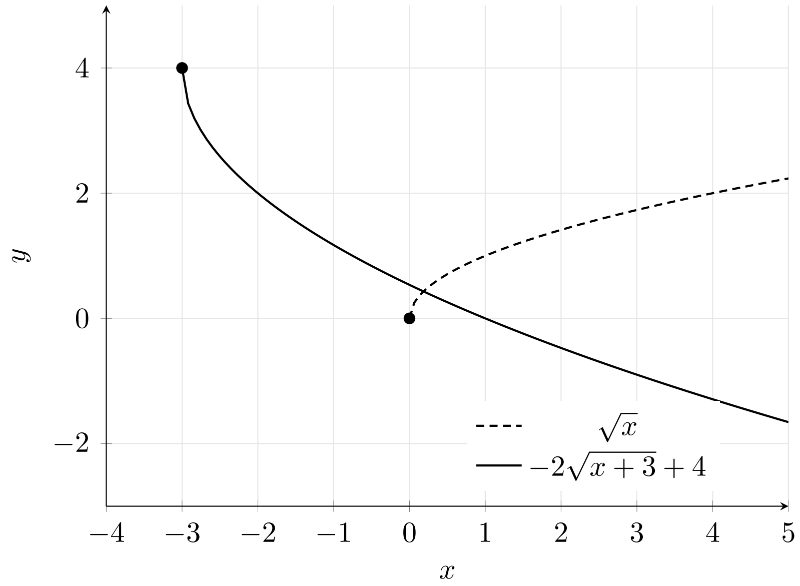


Figure 4: Transformations let us sketch a new model from a familiar parent graph.

Standard graph families for modeling

Family	Typical formula	Modeling meaning
Linear	$mx+b$	constant rate of change; calibration line
Quadratic	ax^2+bx+c	accelerating change, projectile-like shapes, cost curves
Absolute value	$a\|x-h\|+k$	distance from a target, penalty around a desired value
Square root	$a\sqrt{x-h}+k$	fast initial growth then slowing; diminishing gains
Reciprocal	$a/(x-h)+k$	inverse relationship, asymptote, capacity or congestion effects
Sinusoidal	$A\sin(B(t-C))+D$	periodic or seasonal behaviour

Model criticism: Choosing a graph family is an assumption

If a student chooses a quadratic graph only because it fits six points well, the model may extrapolate badly. A graph family should be justified by a mechanism: constant rate, saturation, inverse dependence, periodic forcing, or geometric constraint. A competition report should state why the chosen shape is meaningful, not only that it fits the data.

4 Trigonometric graphs as periodic models

The most useful part of trigonometry for modeling is not memorizing identities; it is recognizing repeated behaviour. Many systems have a repeating pattern: daylight, tides, traffic cycles, temperature, electricity demand, and school attendance.

Sinusoidal model

A basic periodic model has the form

$$y = A\sin(B(t-C)) + D.$$

Here $|A|$ is the **amplitude**, $2\pi/|B|$ is the **period**, C is the **phase shift**, and D is the **midline**. The same interpretation applies to cosine models.

Worked Example 4: A seasonal attendance model

Suppose the average number of students absent from a school follows a yearly cycle. The minimum is about 8 students in September, the maximum is about 28 students in March, and the period is one year. Let t be months after September.
The midline is

$$D = \frac{28+8}{2} = 18,$$

and the amplitude is

$$A = \frac{28-8}{2} = 10.$$

Because the minimum occurs at $t = 0$, a convenient cosine model is

$$\boxed{A(t) = 18 - 10\cos\left(\frac{2\pi}{12}t\right).}$$

At $t = 6$, the model gives $A(6) = 28$, representing the early-spring peak in March. This model should undergo **validation** against several years of records before being used for staffing decisions.

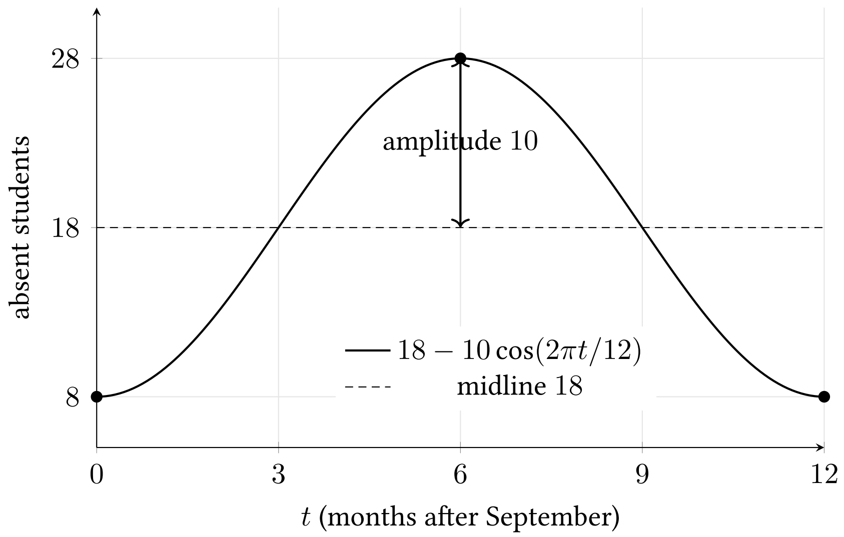


Figure 5: A sinusoidal graph turns a repeating pattern into interpretable parameters.

Checkpoint: Periodic model check

A tide height ranges from 1.2 m to 3.8 m and repeats every 12 hours. At $t = 0$, the tide is at its maximum.

1. Find the amplitude and midline.

2. Write a cosine model.

3. Explain one limitation of this model.

5 Asymptotes and conics as reserve modeling shapes

Some graphs are useful because they show boundaries rather than ordinary turning points. An **asymptote** describes a value the curve approaches but does not cross in a simple way. A **conic section** describes geometric constraints such as fixed distance, fixed sum of distances, or fixed difference of distances.

Worked Example 5: Reading asymptotes in a congestion model

Suppose a simple travel-time model is

$$T(q) = 12 + \frac{40}{100-q}, \qquad 0 \le q < 100,$$

where q is the number of vehicles entering a short road segment per minute. The graph has a vertical asymptote at $q = 100$. As q approaches capacity, the predicted travel time increases rapidly. The asymptote is a warning: operating close to capacity is unstable and small demand changes can cause large delays.

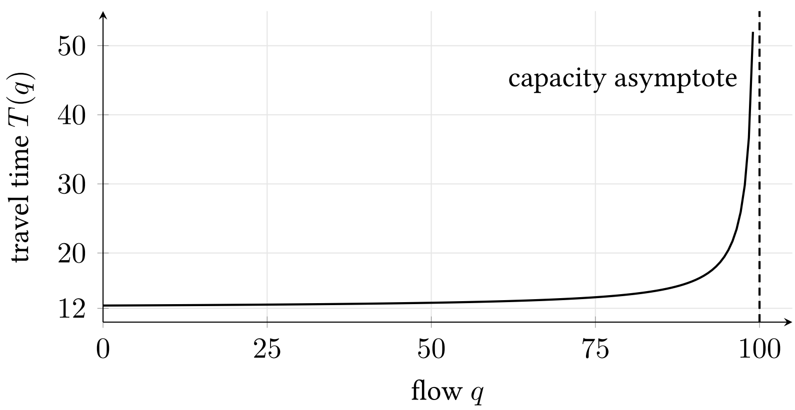


Figure 6: An asymptote can represent a practical capacity boundary.

Conics in one paragraph

Conic	Defining idea	Modeling use
Circle	fixed distance from a centre	coverage, safety radius, sensor range
Parabola	equal distance from focus and directrix	projectile shape, reflector geometry
Ellipse	fixed sum of distances to two foci	orbit-like shapes, service regions
Hyperbola	fixed difference of distances to two foci	localization from time-difference signals

6 Writing graph-based modeling conclusions

Model criticism: Competition-style communication

A graph-based conclusion should include four parts:

1. **Feature:** What graph property matters? For example: peak, intercept, slope, asymptote, period.
2. **Meaning:** What does that feature represent in the real system?
3. **Decision:** What action follows from this interpretation?
4. **Limitation:** What would make the graph unreliable?

Weak conclusion: "The graph goes up and then down."
Stronger conclusion: "The waiting-time curve peaks at about 12:18, so the canteen bottleneck is concentrated in the first half of lunch. A second service counter should be opened before the peak rather than after it. This conclusion assumes that the recorded day is typical and that waiting time was measured consistently."

Differentiated graph-modeling studio **[14 min]**

Choose one route.

1. **Core route:** Given the graph of a quadratic cost model, identify intercepts, vertex, and the meaningful domain.
2. **Challenge route:** Fit a sinusoidal model from maximum, minimum, and period; then explain the phase shift.
3. **Extension route:** Choose between a reciprocal model and a square-root model for a congestion or learning-curve dataset. Justify by mechanism and residual pattern.

Lesson summary

Concept	Modeling meaning
Graph	Visual representation of a relation or function; reveals qualitative structure.
Vertical-line test	Checks whether each input has only one output.
Domain and range	Specify realistic inputs and possible outputs.
Intercepts	Describe starting values, thresholds, or break-even points.
Symmetry	Reduces work and reveals structural assumptions.
Transformation	Reuses known graph shapes through shifts, stretches, and reflections.
Sinusoidal model	Represents periodic behaviour through amplitude, period, phase, and midline.
Asymptote	Indicates a limiting value or practical boundary.
Conic	Encodes geometric constraints such as distance, sum of distances, or coverage.

Checkpoint: Exit ticket

Pick one graph feature from today's lesson and explain how it could affect a real decision in a modeling problem. Your answer must include the feature, its context meaning, and one limitation.

Core practice

1. Find the natural domain of $f(x) = \dfrac{\sqrt{9 - x^2}}{x - 1}$.
2. Determine whether each function is even, odd, or neither: $x^5 - x$, $x^4 + 2x^2$, $x^2 + x$.
3. Describe the transformations from $y = x^2$ to $y = -3(x + 2)^2 + 5$.
4. State the amplitude, period, phase shift, and midline of $y = 4\sin(2t - \pi) + 3$.
5. Find the vertical and horizontal asymptotes of $f(x) = \dfrac{2x + 1}{x - 3}$.

Answer check: core practice

The answers are

$$[-3, 1) \cup (1, 3]; \qquad \text{odd, even, neither;}$$

a shift 2 units left, reflection in the x-axis, vertical stretch by 3, and shift 5 units up; amplitude 4, period π, phase shift $\pi/2$, and midline $y = 3$; vertical asymptote $x = 3$ and horizontal asymptote $y = 2$.

Modeling challenge

A small theme park records the number of visitors entering per hour from 9:00 to 18:00. The data rise slowly in the morning, peak around 14:00, then decline. A student proposes a quadratic model.

1. Define a time variable and a visitor-count function.
2. Explain why a quadratic graph may be reasonable over one day.
3. State the modeling domain and explain why extrapolating beyond it may be misleading.
4. Suggest one graph feature that would help staffing decisions.
5. Write a short graph-based conclusion with one limitation.

Answer check: modeling challenge

A complete response defines time from 9:00 and a visitor-count function, justifies a one-day quadratic as a local rise–peak–fall approximation, restricts the domain to the observed operating day, and identifies the peak time or peak height as a staffing quantity. The conclusion must warn that extrapolation beyond 9:00–18:00 has no support from this model.

Extension practice

1. Prove that the graph of $x^2 + y^2 = 4$ is not the graph of a function of x, but can be split into two functions.
2. A signal is modeled by $S(t) = 2.5\cos(\pi t/6) + 7$. Find its maximum, minimum, period, and value at $t = 3$.

3. Explain how the asymptote in $T(q) = 12 + 40/(100 - q)$ changes if the road capacity increases from 100 to 130 vehicles per minute.
4. Use a circle and its interior to model the coverage points within 50 m of a Wi-Fi router at $(20, 15)$. Write the corresponding inequality.

Answer check: extension practice

The circle splits into $y = \pm\sqrt{4 - x^2}$. For the signal, the maximum is 9.5, the minimum is 4.5, the period is 12, and $S(3) = 7$. Raising road capacity to 130 moves the vertical asymptote from $q = 100$ to $q = 130$ in the correspondingly revised model. The Wi-Fi coverage disk is

$$(x - 20)^2 + (y - 15)^2 \leq 2500.$$

Optional appendix: trigonometric identities

The identities below are useful for algebra and calculus, but they need not all be taught in the two-hour core.

$$\begin{aligned}
\sin^2\theta + \cos^2\theta &= 1,\\
\sin(\alpha + \beta) &= \sin\alpha\cos\beta + \cos\alpha\sin\beta,\\
\cos(\alpha + \beta) &= \cos\alpha\cos\beta - \sin\alpha\sin\beta,\\
\sin(2\theta) &= 2\sin\theta\cos\theta,\\
\cos(2\theta) &= 2\cos^2\theta - 1 = 1 - 2\sin^2\theta.
\end{aligned}$$

Model audit card

Audit before trusting or communicating a model

Purpose	What question or decision does the model support?	**Data**	Which inputs are observed, derived, or simulated?
Assumptions	Which simplifications may change the answer?	**Units**	Are dimensions and scales consistent?
Verification	Were equations, code, constraints, and limiting cases checked?	**Validation**	Is performance tested against independent evidence?
Uncertainty	How do measurement, parameter, solver, and structural errors affect conclusions?	**Ethics**	Who benefits, who bears risk, and what fairness criterion matters?
Reproducibility	Are data provenance, versions, seeds, and transformations recorded?	**Communication**	Are claims conditional, traceable, and decision-relevant?

Further reading

- Precalculus 2e (OpenStax)

Sources, provenance, and licence

These notes follow the iterative modeling and assessment perspective in the SIAM/COMAP GAIMME report. Competition-specific remarks are a snapshot checked on 20 August 2026 against the official COMAP HiMCM instructions; readers should verify current rules before submission. Unless a source is explicitly identified, classroom datasets and numerical scenarios are synthetic or simulated teaching examples and are not empirical claims about a real institution. Third-party references retain their own terms.

Student response guide and checking criteria

For closed calculations, show the definition, substitution, units, and at least one independent check; the worked examples in this unit are method hints, but you must complete the arithmetic yourself. Open modeling responses are checked on five dimensions: a clear purpose and output; stated data and provenance; defensible assumptions and scope; traceable computational verification and model validation; and alignment among uncertainty, limitations, and the recommendation. Support each dimension with concrete evidence rather than writing only "completed."

Lectures in Mathematical Modeling

Supplementary 4: Limits, Continuity, and Asymptotic Thinking

Kenneth Sok Kin Cheng

Revised 20 August 2026

Learning goals

By the end of this supplementary lesson, readers should be able to:

1. explain a **limit** as the value approached by a function near a point, not necessarily at the point;
2. evaluate common limits by substitution, factorisation, rationalisation, and standard trigonometric limits;
3. use **one-sided limits** to decide whether a two-sided limit exists;
4. interpret **infinite limits**, **limits at infinity**, and **asymptotes**;
5. classify removable, jump, and infinite discontinuities;
6. apply **continuity**, the **Intermediate Value Theorem**, and the **Extreme Value Theorem** in modeling contexts;
7. explain numerical and graphical evidence for a limit without confusing it with proof.

1 Limits as local behaviour

A **limit** describes what a function is approaching. It is local: it looks near a point, not necessarily at the point. This idea is central to calculus, but it is also a modeling idea. A sensor may fail at one time, a formula may be undefined at one value, or a process may approach a capacity without ever reaching it exactly.

Launch: the missing sensor reading **[12 min]**

A temperature sensor records readings every minute. At $t = 5$, the sensor fails, but nearby readings suggest the curve should pass near $23.8°C$.

1. Should we say the temperature at $t = 5$ is known?
2. Should we say the *trend* near $t = 5$ is known?
3. What evidence would make the estimated limiting value more credible?

Modeling lesson: a limit is about approaching behaviour. It can be meaningful even when the value at the exact input is missing or unreliable.

Core definition, stated informally

We write

$$\lim_{x \to a} f(x) = L$$

when $f(x)$ can be made as close as we like to L by choosing x sufficiently close to a, with $x \neq a$. The value $f(a)$ may equal L, differ from L, or fail to exist.

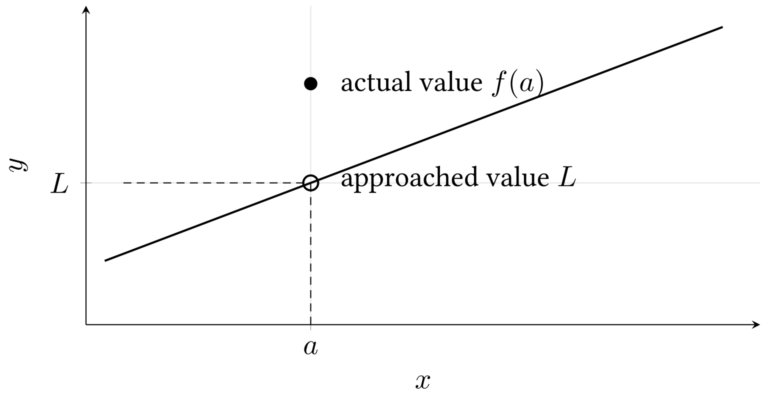


Figure 1: A limit describes the value approached by the graph near $x = a$. The actual value at a can be different.

Worked Example 1: A removable hole

Evaluate
$$\lim_{x\to 1} \frac{x^2-1}{x-1}.$$
At $x = 1$, the expression is undefined. But for $x \neq 1$,
$$\frac{x^2-1}{x-1} = \frac{(x-1)(x+1)}{x-1} = x+1.$$
Therefore
$$\lim_{x\to 1} \frac{x^2-1}{x-1} = \lim_{x\to 1}(x+1) = 2.$$
Interpretation: the original formula has a hole at $x = 1$, but the nearby trend approaches 2.

Worked Example 2: Rationalisation

Evaluate
$$\lim_{x\to 0} \frac{\sqrt{x+4}-2}{x}.$$
Direct substitution gives $0/0$, so we rationalise:
$$\frac{\sqrt{x+4}-2}{x} \cdot \frac{\sqrt{x+4}+2}{\sqrt{x+4}+2} = \frac{x}{x(\sqrt{x+4}+2)} = \frac{1}{\sqrt{x+4}+2}, \qquad x \neq 0.$$
Hence
$$\lim_{x\to 0} \frac{\sqrt{x+4}-2}{x} = \frac{1}{2+2} = \frac{1}{4}.$$

Checkpoint: Algebraic limit check

Evaluate:

1. $\lim_{x\to 2} \dfrac{x^2-4}{x-2}$;
2. $\lim_{x\to 9} \dfrac{\sqrt{x}-3}{x-9}$;
3. $\lim_{x\to 1} \dfrac{x^3-1}{x^2-1}$.

For each one, state the obstacle first: direct substitution, $0/0$, square root, or factorisation.

2 From informal limits to epsilon-delta language

In a first modeling course, students do not need to master every ε-δ **proof**. They should, however, understand the meaning of the definition: output tolerance controls input tolerance.

The formal definition

We say

$$\lim_{x\to a} f(x) = L$$

if for every $\varepsilon > 0$ there exists $\delta > 0$ such that

$$0 < |x-a| < \delta \implies |f(x) - L| < \varepsilon.$$

Here ε is the allowed error in the output, while δ is the allowed distance from the input a.

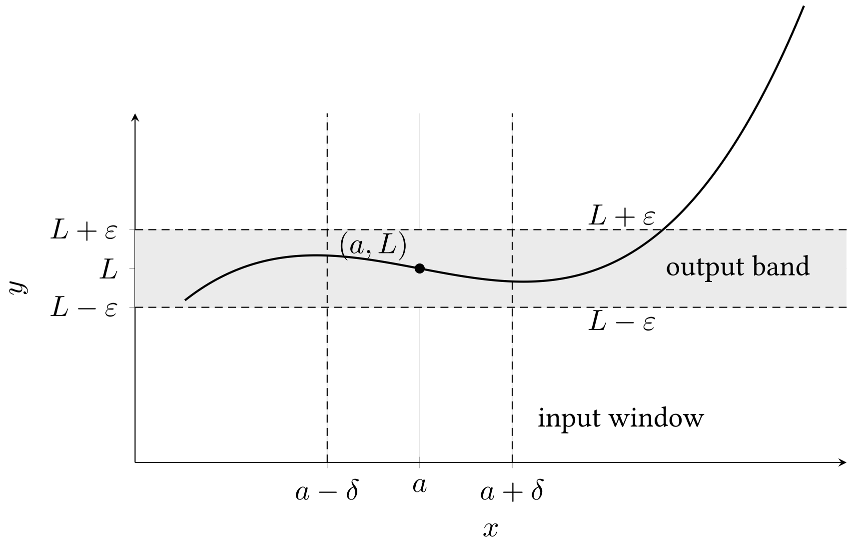


Figure 2: The epsilon-delta definition of a limit: if x is sufficiently close to a, then $f(x)$ is within ε of L.

Worked Example 3: A simple epsilon-delta proof

Prove that $\lim_{x\to 3}(2x-1) = 5$.
We want

$$|(2x-1) - 5| = |2x-6| = 2|x-3| < \varepsilon.$$

So it is enough to require $|x-3| < \varepsilon/2$. Let $\delta = \varepsilon/2$. Then whenever $0 < |x-3| < \delta$,

$$|(2x-1) - 5| = 2|x-3| < 2\delta = \varepsilon.$$

Therefore $\lim_{x\to 3}(2x-1) = 5$.

Model criticism: Proof versus evidence

A numerical table can suggest a limit; a graph can illustrate a limit; an epsilon-delta proof establishes the limit. In mathematical modeling reports, students usually need to explain the trend and justify the computation. Formal proof is useful when the result will be used as a theorem or when the graph is misleading.

3 One-sided and infinite limits

Some functions approach different values from the left and from the right. A two-sided limit exists only when the two one-sided limits agree.

One-sided limits

The **right-hand limit** is $\lim_{x\to a^+} f(x)$; it approaches a using $x > a$. The **left-hand limit** is $\lim_{x\to a^-} f(x)$; it approaches a using $x < a$.

$$\lim_{x\to a} f(x) = L \iff \lim_{x\to a^-} f(x) = L \text{ and } \lim_{x\to a^+} f(x) = L.$$

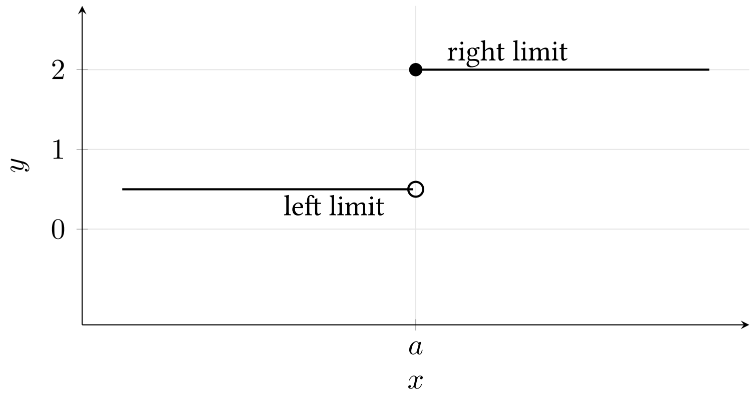


Figure 3: A jump occurs when the left-hand and right-hand limits are different.

Worked Example 4: Absolute value quotient

Evaluate

$$\lim_{x\to 0} \frac{|x|}{x}.$$

If $x > 0$, then $|x| = x$, so $|x|/x = 1$. If $x < 0$, then $|x| = -x$, so $|x|/x = -1$. Therefore

$$\lim_{x\to 0^+} \frac{|x|}{x} = 1, \qquad \lim_{x\to 0^-} \frac{|x|}{x} = -1.$$

Since the two one-sided limits are unequal, the two-sided limit does not exist.

Infinite limits and vertical asymptotes

Writing $\lim_{x\to a} f(x) = +\infty$ means that $f(x)$ grows without bound as x approaches a. It does not mean that $+\infty$ is an ordinary real-number limit. If at least one one-sided limit is infinite, then the line $x = a$ is a **vertical asymptote**.

Worked Example 5: Capacity blow-up

A simple congestion model for waiting time is

$$W(q) = \frac{1}{1-q}, \qquad 0 \le q < 1,$$

where q is demand as a fraction of capacity. Then

$$\lim_{q \to 1^-} W(q) = +\infty.$$

Interpretation: as demand approaches full capacity, the model predicts waiting time grows extremely fast. The vertical asymptote $q = 1$ is a warning that operating close to capacity is unstable.

4 Limits at infinity and asymptotic thinking

A **limit at infinity** describes long-run behaviour. In modeling, it answers questions such as: Does the prediction stabilise? Does a cost approach a baseline? Does a probability approach one?

Worked Example 6: Long-run behaviour of a rational model

Evaluate

$$\lim_{x \to \infty} \frac{3x^2 - 5x + 2}{2x^2 + x - 1}.$$

Divide numerator and denominator by x^2:

$$\frac{3 - 5/x + 2/x^2}{2 + 1/x - 1/x^2} \to \frac{3}{2}.$$

Therefore the model has horizontal asymptote $y = 3/2$.

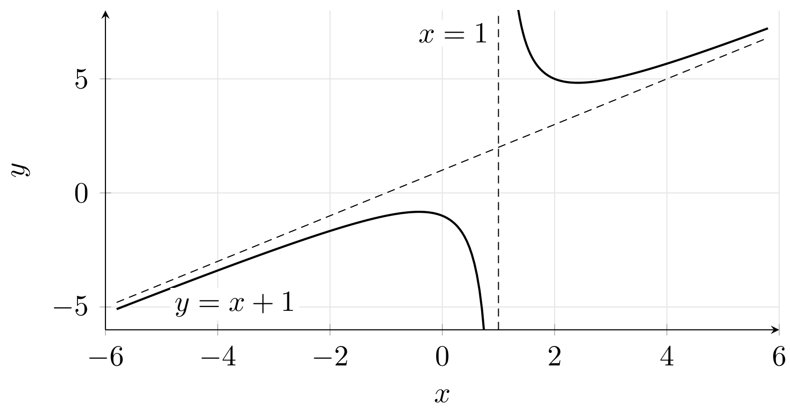


Figure 4: Asymptotes summarise behaviour near forbidden inputs and far from the origin.

Limit techniques for class use

Situation	Technique	Example
Continuous expression	Substitute directly	polynomial, square root on domain
$0/0$ with polynomial	Factor and cancel	$(x^2-4)/(x-2)$
$0/0$ with radicals	Rationalise	$(\sqrt{x+4}-2)/x$
Oscillating bounded factor	Squeeze theorem	$x^2 \sin(1/x)$
Piecewise or absolute value	Check one-sided limits	$\lvert x\rvert/x$
Rational function at infinity	Divide by highest power	leading coefficient ratio
Vertical blow-up	One-sided infinite limits	$1/(x-a)$ or $1/(x-a)^2$

5 Continuity: no jump, no hole, no blow-up

Continuity means that the function value matches the limiting trend.

Continuity at a point

A function f is **continuous** at $x = a$ if all three conditions hold:

1. $f(a)$ is defined;
2. $\lim_{x\to a} f(x)$ exists;
3. $\lim_{x\to a} f(x) = f(a)$.

If any condition fails, the function is discontinuous at a.

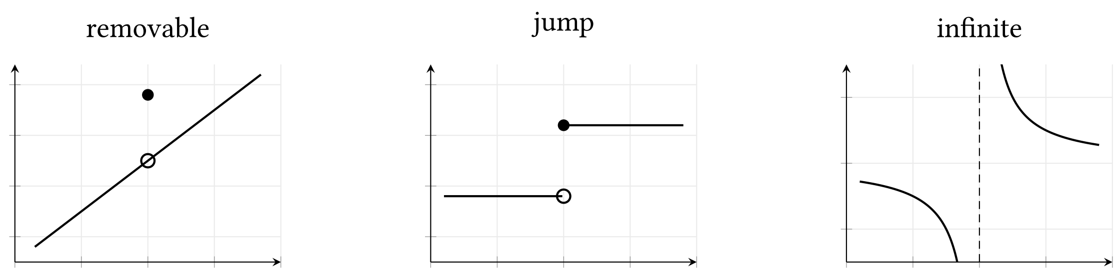


Figure 5: Three common discontinuities: a hole, a jump, and a vertical blow-up.

Worked Example 7: Make a piecewise model continuous

Find k so that

$$f(x) = \begin{cases} kx^2 - 1, & x < 2, \\ 3x + k, & x \geq 2 \end{cases}$$

is continuous at $x = 2$.
Left-hand limit:

$$\lim_{x\to 2^-} f(x) = 4k - 1.$$

Right-hand value and limit:

$$\lim_{x \to 2^+} f(x) = f(2) = 6 + k.$$

Set them equal:

$$4k - 1 = 6 + k \quad \Rightarrow \quad 3k = 7 \quad \Rightarrow \quad k = \frac{7}{3}.$$

Model criticism: Modeling interpretation of continuity

Continuity is an assumption. A physical temperature curve is often continuous; a fare system, school grade boundary, or traffic-light rule may have jumps. Before applying a continuous model, ask whether the real system changes smoothly or by thresholds.

6 Existence theorems: IVT and EVT

The **Intermediate Value Theorem** and **Extreme Value Theorem** are important because they guarantee existence even when we cannot solve exactly.

Intermediate Value Theorem

If f is continuous on $[a, b]$ and N lies between $f(a)$ and $f(b)$, then there exists $c \in (a, b)$ such that $f(c) = N$.
In particular, if $f(a)$ and $f(b)$ have opposite signs, then f has a root in (a, b).

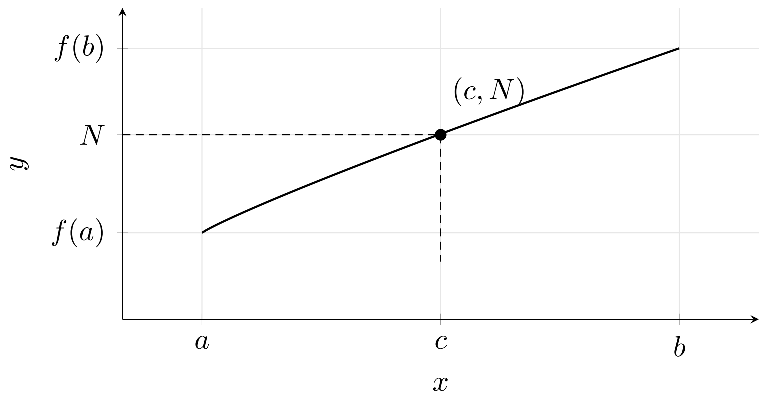


Figure 6: A continuous graph cannot move from $f(a)$ to $f(b)$ without taking every intermediate value.

Worked Example 8: Existence of a solution

Show that

$$x^3 - x - 1 = 0$$

has a solution between 1 and 2.
Let $f(x) = x^3 - x - 1$. Since f is a polynomial, it is continuous. Compute

$$f(1) = 1 - 1 - 1 = -1 < 0, \qquad f(2) = 8 - 2 - 1 = 5 > 0.$$

By the IVT, there exists $c \in (1, 2)$ such that $f(c) = 0$.

Extreme Value Theorem

If f is continuous on a closed bounded interval $[a, b]$, then f attains an absolute maximum and an absolute minimum on $[a, b]$.
For modeling, this theorem justifies the phrase "there is an optimal value" before we try to find it.
The hypotheses matter: continuous function, closed interval, bounded interval.

7 Squeeze theorem and trigonometric limits

The **Squeeze Theorem** handles functions trapped between two simpler functions.

Squeeze theorem

If

$$g(x) \leq f(x) \leq h(x)$$

near a, and

$$\lim_{x \to a} g(x) = \lim_{x \to a} h(x) = L,$$

then $\lim_{x \to a} f(x) = L$.

Worked Example 9: Oscillation controlled by a vanishing factor

Evaluate

$$\lim_{x \to 0} x^2 \sin\left(\frac{1}{x}\right).$$

Since $-1 \leq \sin(1/x) \leq 1$ for $x \neq 0$,

$$-x^2 \leq x^2 \sin\left(\frac{1}{x}\right) \leq x^2.$$

Both outer functions approach 0, so by the Squeeze Theorem,

$$\lim_{x \to 0} x^2 \sin\left(\frac{1}{x}\right) = 0.$$

Two standard trigonometric limits

$$\lim_{x \to 0} \frac{\sin x}{x} = 1, \qquad \lim_{x \to 0} \frac{1 - \cos x}{x} = 0.$$

These are foundational for derivatives of trigonometric functions. They also show why radians, not degrees, are the natural angle unit in calculus.

Worked Example 10: Using the standard trig limit

Evaluate

$$\lim_{x \to 0} \frac{\sin(3x)}{x}.$$

Rewrite:

$$\frac{\sin(3x)}{x} = 3 \cdot \frac{\sin(3x)}{3x}.$$

As $x \to 0$, also $3x \to 0$, so

$$\lim_{x\to 0} \frac{\sin(3x)}{x} = 3 \cdot 1 = 3.$$

8 Differentiated limit studio

Differentiated studio **[10 min]**

Choose one route.

1. **Core route: algebraic limits.** Evaluate

$$\lim_{x\to -2} \frac{x^3+8}{x+2}, \qquad \lim_{x\to 0} \frac{\sqrt{x+9}-3}{x}.$$

2. **Challenge route: one-sided and continuity.** For

$$f(x) = \begin{cases} ax+1, & x<1, \\ x^2+a, & x \geq 1, \end{cases}$$

find a so that f is continuous at $x = 1$.

3. **Extension route: asymptotic model.** A cost model is

$$C(n) = \frac{5n^2+30n}{2n^2+1}.$$

Find the long-run cost and explain what the horizontal asymptote means.

Model criticism: Competition-style communication

When using limits in a modeling report, write in context.

Weak statement: "The limit is infinity."

Stronger statement: "As demand approaches full capacity from below, the waiting-time model $W(q) = 1/(1-q)$ grows without bound. This suggests that operating near $q = 1$ is unsafe: small increases in demand may produce very large waiting times. The conclusion depends on the model's assumption that service capacity remains fixed."

Lesson summary

Concept	Modeling meaning
Limit	Describes approaching behaviour near a point.
Hole	Formula may be undefined at one input while the trend remains clear.
One-sided limit	Detects thresholds, jumps, and directional behaviour.
Infinite limit	Signals blow-up or capacity failure.
Limit at infinity	Describes long-run behaviour and horizontal asymptotes.
Continuity	Assumes smooth change; may fail in threshold systems.
IVT	Guarantees existence of a solution between two signs or values.
EVT	Guarantees a maximum/minimum when continuity and closed interval hold.
Squeeze theorem	Controls a complicated function between simpler ones.

Checkpoint: Exit ticket

In three sentences, explain the difference between $f(a)$ and $\lim_{x\to a} f(x)$. Then give one modeling situation where the limit is useful even if $f(a)$ is missing.

Core practice

1. Evaluate $\lim\limits_{x\to 4}(5x-3)$ using direct substitution.
2. Evaluate $\lim\limits_{x\to 2}\dfrac{x^2-x-2}{x-2}$.
3. Evaluate $\lim\limits_{x\to 0}\dfrac{\sqrt{x+1}-1}{x}$.
4. Evaluate $\lim\limits_{x\to 0}\dfrac{\tan x}{x}$.
5. Find the one-sided limits of $f(x)=\lfloor x\rfloor$ at $x=3$.
6. Classify the discontinuity of $f(x)=(x^2-9)/(x-3)$ at $x=3$.

Practice check and hints

1. Direct substitution gives 17.
2. Factor and cancel $x-2$ for $x\neq 2$; the limit is 3.
3. Rationalisation gives the limit $1/2$.
4. Write $\tan x/x=(\sin x/x)/\cos x$; the limit is 1.
5. The left- and right-hand limits are 2 and 3, respectively, so the two-sided limit does not exist.
6. The discontinuity is removable; the limiting value at $x=3$ is 6.

Modeling challenge

A school cafeteria uses

$$W(q) = \frac{4q}{1-q}, \qquad 0 \le q < 1,$$

to estimate average waiting time in minutes, where q is demand divided by service capacity.

1. Find $\lim_{q\to 0^+} W(q)$ and interpret.
2. Find $\lim_{q\to 1^-} W(q)$ and interpret.
3. Find the value of q at which the model predicts $W = 12$ minutes.
4. State one limitation of this model.
5. Write a short recommendation for cafeteria managers based on the limiting behaviour.

Modeling check

The first three numerical checks are 0, $+\infty$, and $q = 3/4$. For the final two parts, state the capacity-domain assumption, distinguish the asymptotic warning from empirical validation, name at least one omitted mechanism, and make a recommendation that is consistent with the limiting behaviour.

Extension practice

1. Prove using the epsilon-delta definition that $\lim_{x\to 1}(3x+2) = 5$.
2. Use the sequential criterion to show that $\lim_{x\to 0} \sin(1/x)$ does not exist.
3. Evaluate $\displaystyle\lim_{x\to\infty}\left(\sqrt{x^2+2x+3} - x\right)$.
4. Find all vertical and horizontal asymptotes of $f(x) = \dfrac{3x^2-2}{x^2-4}$.
5. Use the IVT to prove that $e^x = 3 - 2x$ has at least one real solution.

Extension hints

For Question 1, $\delta = \varepsilon/3$ is sufficient. For Question 2, use two sequences approaching 0 for which the sine values are 1 and -1. Rationalising in Question 3 gives the limit 1. In Question 4 the vertical asymptotes are $x = \pm 2$ and the horizontal asymptote is $y = 3$. For Question 5, apply the IVT to $h(x) = e^x + 2x - 3$: $h(0) = -2$ and $h(1) = e - 1 > 0$.

Optional appendix: L'Hopital's Rule preview

Some limits produce indeterminate forms such as $0/0$ or ∞/∞. After derivatives are introduced, one powerful tool is L'Hopital's Rule: under suitable conditions,

$$\lim_{x\to a}\frac{f(x)}{g(x)} = \lim_{x\to a}\frac{f'(x)}{g'(x)}.$$

This rule should not replace algebraic thinking. Factorisation, rationalisation, and standard trigonometric limits remain important because they reveal the structure of the expression.

Model audit card

Audit before trusting or communicating a model

Purpose	What question or decision does the model support?	**Data**	Which inputs are observed, derived, or simulated?
Assumptions	Which simplifications may change the answer?	**Units**	Are dimensions and scales consistent?
Verification	Were equations, code, constraints, and limiting cases checked?	**Validation**	Is performance tested against independent evidence?
Uncertainty	How do measurement, parameter, solver, and structural errors affect conclusions?	**Ethics**	Who benefits, who bears risk, and what fairness criterion matters?
Reproducibility	Are data provenance, versions, seeds, and transformations recorded?	**Communication**	Are claims conditional, traceable, and decision-relevant?

Further reading

- Calculus Volume 1 (OpenStax)

Sources, provenance, and licence

These notes follow the iterative modeling and assessment perspective in the SIAM/COMAP GAIMME report. Competition-specific remarks are a snapshot checked on 20 August 2026 against the official COMAP HiMCM instructions; readers should verify current rules before submission. Unless a source is explicitly identified, classroom datasets and numerical scenarios are synthetic or simulated teaching examples and are not empirical claims about a real institution. Third-party references retain their own terms.

Student response guide and checking criteria

For closed calculations, show the definition, substitution, units, and at least one independent check; the worked examples in this unit are method hints, but you must complete the arithmetic yourself. Open modeling responses are checked on five dimensions: a clear purpose and output; stated data and provenance; defensible assumptions and scope; traceable computational verification and model validation; and alignment among uncertainty, limitations, and the recommendation. Support each dimension with concrete evidence rather than writing only “completed.”

Lectures in Mathematical Modeling

Supplementary 5: Derivatives, Rates, and Local Linear Models

Kenneth Sok Kin Cheng

Revised 20 August 2026

Learning goals

By the end of this supplementary lesson, readers should be able to:

1. interpret the **derivative** as a local slope, instantaneous rate, and sensitivity coefficient;
2. compute simple derivatives from the **difference quotient**;
3. use the power, product, quotient, trigonometric, exponential, logarithmic, and **chain rule** correctly;
4. identify common reasons why a function is not differentiable;
5. use implicit, parametric, and related-rate differentiation in geometric or physical models;
6. use the **Mean Value Theorem** to connect average and instantaneous rates;
7. explain derivative calculations in a modeling report instead of presenting them as isolated algebra.

1 Derivative as slope, rate, and sensitivity

In modeling, a derivative is not just a calculus operation. It answers a practical question: *how fast does the output change when the input changes slightly?* If $y = f(x)$, then $f'(a)$ measures the local change in y per unit change in x near $x = a$.

Launch: average speed versus speed at one instant **[12 min]**

A robot travels along a corridor. Its distance from the start after t seconds is modeled by

$$s(t) = 0.4t^2 + 1.2t \qquad (0 \le t \le 10),$$

where s is in metres.

1. Compute the average speed from $t = 3$ to $t = 5$.
2. Compute the average speed from $t = 4$ to $t = 4.1$.
3. What value should the speedometer display at exactly $t = 4$?

Discussion point: smaller time windows give more local information. The derivative is the limiting value of these average rates.

Core definition

For a function f, the **derivative at** a is

$$f'(a) = \lim_{h \to 0} \frac{f(a+h) - f(a)}{h},$$

provided the limit exists. The fraction is the **difference quotient**. It is the slope of the secant line through $(a, f(a))$ and $(a+h, f(a+h))$.

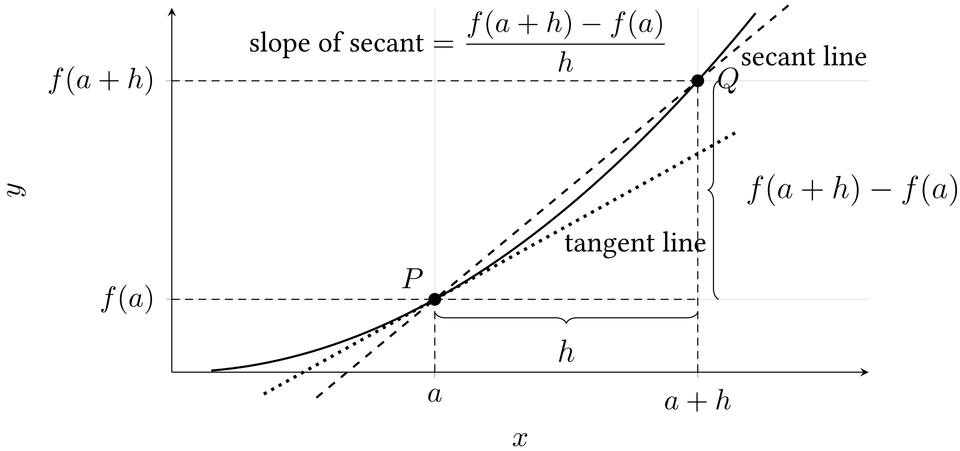


Figure 1: The derivative at a is the limit of the slopes of secant lines: $f'(a) = \lim_{h\to 0} \frac{f(a+h)-f(a)}{h}$. As h becomes smaller, the secant line through P and Q approaches the tangent line at P.

Worked example 1: speed from a position model

For $s(t) = 0.4t^2 + 1.2t$, the instantaneous speed at $t = 4$ is

$$s'(4) = \lim_{h\to 0} \frac{s(4+h)-s(4)}{h}.$$

Compute

$$\begin{aligned} s(4+h) &= 0.4(4+h)^2 + 1.2(4+h) \\ &= 0.4(16+8h+h^2) + 4.8 + 1.2h \\ &= 11.2 + 4.4h + 0.4h^2, \end{aligned}$$

while $s(4) = 11.2$. Hence

$$\frac{s(4+h)-s(4)}{h} = 4.4 + 0.4h \to 4.4.$$

So the robot's speed at $t = 4$ is 4.4 m/s.

Checkpoint: Checkpoint 1

Explain the units of $s'(4)$. Why is it not measured in metres?

2 First-principles derivatives

Before using rules, students should experience why the rules are true. This prevents derivatives from becoming a memorised table detached from modeling.

Worked example 2: derivative of x^2 from the definition

Let $f(x) = x^2$. Then

$$f'(x) = \lim_{h\to 0} \frac{(x+h)^2 - x^2}{h} = \lim_{h\to 0} \frac{2xh + h^2}{h} = \lim_{h\to 0}(2x+h) = 2x.$$

At $x = 3$, the local slope is 6. This means that near $x = 3$, increasing x by 0.01 increases x^2 by about 0.06.

Worked example 3: derivative of $\sqrt{x}$

For $x > 0$,

$$\begin{aligned}\frac{d}{dx}\sqrt{x} &= \lim_{h\to 0}\frac{\sqrt{x+h}-\sqrt{x}}{h}\\ &= \lim_{h\to 0}\frac{(x+h)-x}{h(\sqrt{x+h}+\sqrt{x})}\\ &= \lim_{h\to 0}\frac{1}{\sqrt{x+h}+\sqrt{x}} = \frac{1}{2\sqrt{x}}.\end{aligned}$$

The derivative is larger near $x = 0$, so the square-root graph is steep near the origin and gradually flattens.

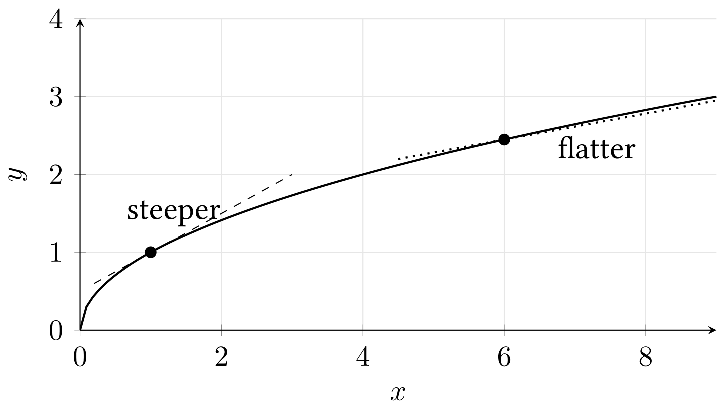


Figure 2: The derivative of $\sqrt{x}$ decreases as x increases. This is a useful visual interpretation of $1/(2\sqrt{x})$.

Model criticism: Model critique: derivative as local linear model

A derivative gives a local approximation:

$$f(a + \Delta x) \approx f(a) + f'(a)\Delta x.$$

This is powerful, but it is local. It may be poor when Δx is large or when the graph bends sharply. In a modeling report, always state the interval over which the local approximation is being used.

3 Differentiation rules for efficient calculation

Core derivative rules

If f and g are differentiable, then

$$\begin{aligned}&\frac{d}{dx}(c) = 0, && \frac{d}{dx}(x^n) = nx^{n-1}, && \frac{d}{dx}[cf] = cf',\\ &\frac{d}{dx}(f+g) = f'+g', && \frac{d}{dx}(fg) = f'g+fg', && \frac{d}{dx}\left(\frac{f}{g}\right) = \frac{f'g-fg'}{g^2}.\end{aligned}$$

The product rule is not $(fg)' = f'g'$. The quotient rule is not the derivative of numerator divided by derivative of denominator.

Worked example 4: product and quotient rules

Differentiate

$$F(x) = x^2 \sin x + \frac{x^2+1}{x-1}.$$

For the product term,

$$\frac{d}{dx}(x^2 \sin x) = 2x \sin x + x^2 \cos x.$$

For the quotient term,

$$\frac{d}{dx}\left(\frac{x^2+1}{x-1}\right) = \frac{2x(x-1)-(x^2+1)}{(x-1)^2} = \frac{x^2-2x-1}{(x-1)^2}.$$

Therefore

$$F'(x) = 2x \sin x + x^2 \cos x + \frac{x^2-2x-1}{(x-1)^2}.$$

Worked example 5: trigonometric and exponential derivatives

For a seasonal demand model

$$D(t) = 120 + 30 \sin\left(\frac{2\pi t}{12}\right) + 5e^{0.03t},$$

where t is measured in months, compute the instantaneous rate of change:

$$D'(t) = 30 \cos\left(\frac{2\pi t}{12}\right) \cdot \frac{2\pi}{12} + 5e^{0.03t} \cdot 0.03.$$

Thus

$$D'(t) = 5\pi \cos\left(\frac{\pi t}{6}\right) + 0.15e^{0.03t}.$$

The first term describes seasonal increase/decrease; the second term describes long-term growth.

Checkpoint: Checkpoint 2

Why is the derivative of $\sin(\pi t/6)$ not simply $\cos(\pi t/6)$? Name the rule being used.

4 The chain rule as a modeling pipeline

The **chain rule** is essential for modeling because many models are pipelines: an input affects an intermediate quantity, which affects the output.

Chain rule

If $y = f(u)$ and $u = g(x)$, then

$$\frac{dy}{dx} = \frac{dy}{du} \cdot \frac{du}{dx}.$$

In function notation,

$$\frac{d}{dx} f(g(x)) = f'(g(x))g'(x).$$

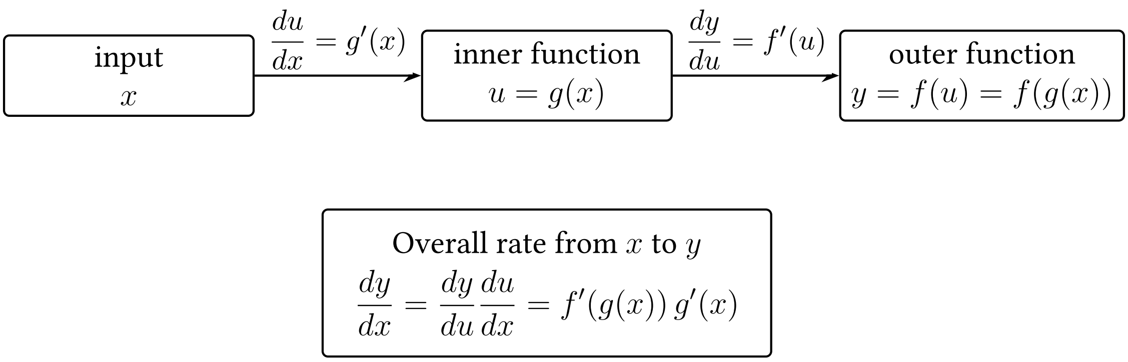


Figure 3: The chain rule for a composite function $y = f(g(x))$. The input x first changes the intermediate variable $u = g(x)$, and then u changes the final output $y = f(u)$. Thus the overall rate is the product of the two rates.

Worked example 6: pollution concentration through a transformation

A concentration index is modeled by

$$I(t) = \sqrt{1 + 0.4C(t)}, \qquad C(t) = 20e^{-0.08t}.$$

Find $I'(t)$ and interpret the sign.

$$I'(t) = \frac{1}{2\sqrt{1 + 0.4C(t)}} \cdot 0.4C'(t).$$

Since $C'(t) = -1.6e^{-0.08t}$,

$$I'(t) = \frac{0.2(-1.6e^{-0.08t})}{\sqrt{1 + 8e^{-0.08t}}} < 0.$$

The index is decreasing. The derivative becomes closer to zero over time because the exponential decay slows in absolute size.

Worked example 7: logarithmic differentiation

Differentiate $y = x^x$ for $x > 0$.

$$\ln y = x \ln x.$$

Differentiate both sides:

$$\frac{y'}{y} = \ln x + 1.$$

Therefore

$$y' = x^x(\ln x + 1).$$

This method is useful when both the base and exponent vary.

5 When derivatives do not exist

A derivative may fail to exist even for a simple-looking graph. In a model, that failure may mark a switch, threshold, collision, capacity boundary, or policy change.

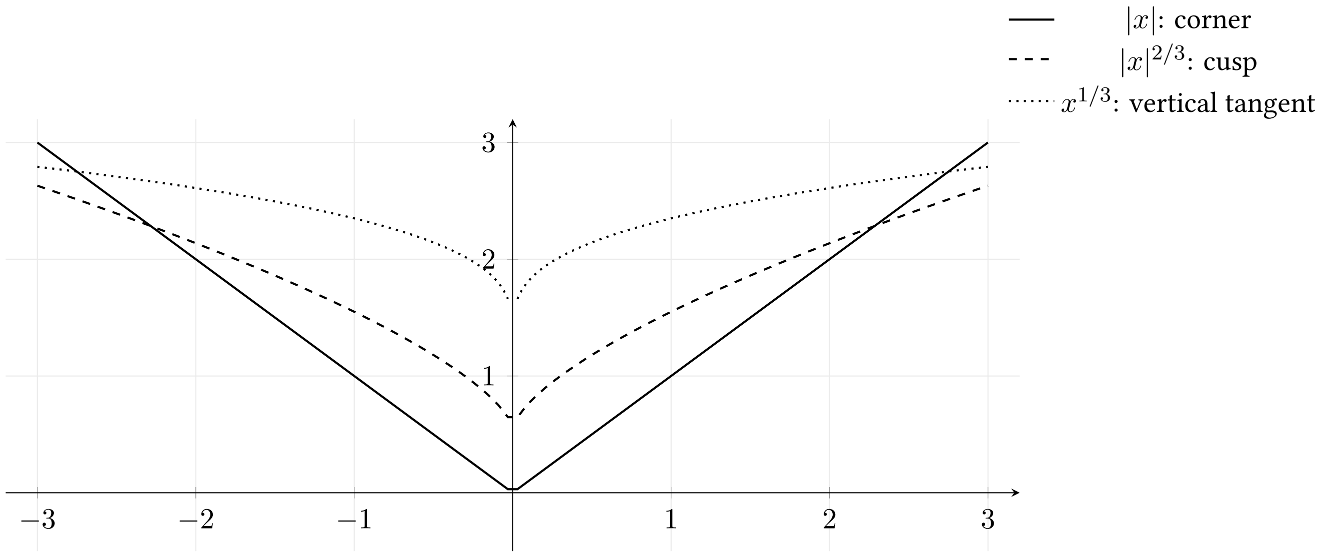


Figure 4: Three common non-differentiable behaviours. Each can have a modeling interpretation.

Checkpoint: Checkpoint 3

For $f(x) = |x|$, compute the right-hand derivative and left-hand derivative at 0. Why does the derivative not exist there?

Model criticism: Model critique: smoothing versus preserving a threshold

In applications, replacing a kink by a smooth curve may make calculus easier, but it may erase a real threshold. For example, a penalty fee may start exactly when demand exceeds capacity. A good modeler should ask: is the corner a mathematical inconvenience, or is it the phenomenon we need to preserve?

6 Implicit, parametric, and related-rate differentiation

Not all models are naturally written as $y = f(x)$. Circles, ellipses, supply-demand equilibria, and geometric constraints often appear as equations involving several variables.

Worked example 8: implicit differentiation on a circle

For a circle

$$x^2 + y^2 = 25,$$

we differentiate both sides with respect to x:

$$2x + 2y\frac{dy}{dx} = 0,$$

so

$$\frac{dy}{dx} = -\frac{x}{y}.$$

At $(3, 4)$, the tangent slope is $-3/4$, so the tangent line is

$$y - 4 = -\frac{3}{4}(x - 3).$$

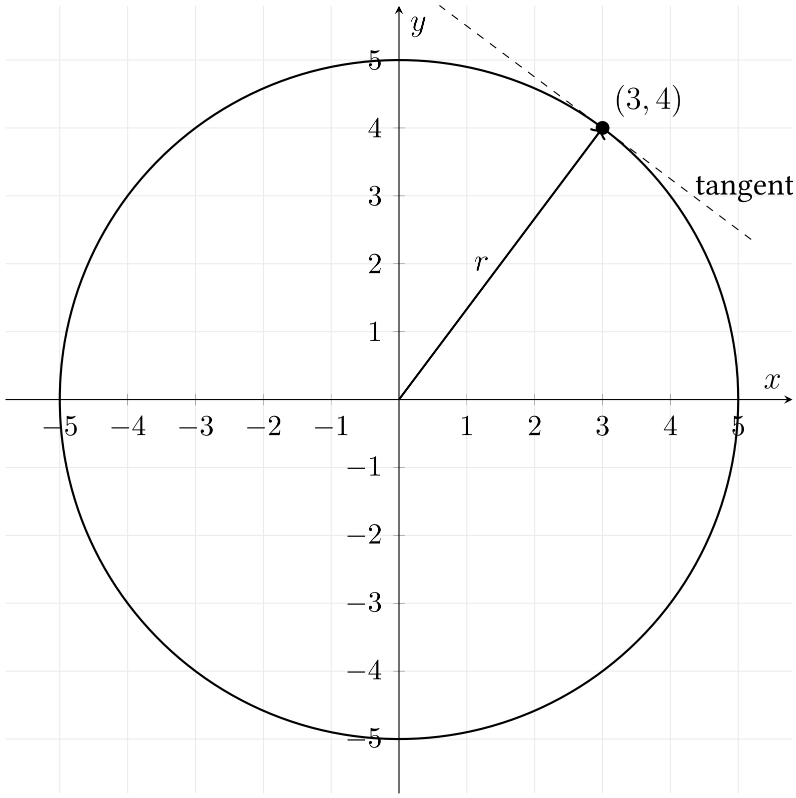


Figure 5: Implicit differentiation finds tangent slopes even when the curve is not a function of x globally.

Parametric derivative

If a curve is given by $x = g(t)$ and $y = h(t)$, then

$$\frac{dy}{dx} = \frac{dy/dt}{dx/dt}, \qquad dx/dt \neq 0.$$

This is useful for motion models because t is the natural independent variable.

Worked example 9: related rates for a sliding ladder

A 10 m ladder leans against a wall. Its base moves away at 0.5 m/s. How fast is the top moving down when the base is 6 m from the wall?
Let x be the distance of the base from the wall and y the height of the top. Then

$$x^2 + y^2 = 100.$$

Differentiate with respect to time:

$$2x\frac{dx}{dt} + 2y\frac{dy}{dt} = 0.$$

When $x = 6$, $y = 8$. Hence

$$\frac{dy}{dt} = -\frac{x}{y}\frac{dx}{dt} = -\frac{6}{8}(0.5) = -0.375 \text{ m/s}.$$

The negative sign means the height is decreasing.

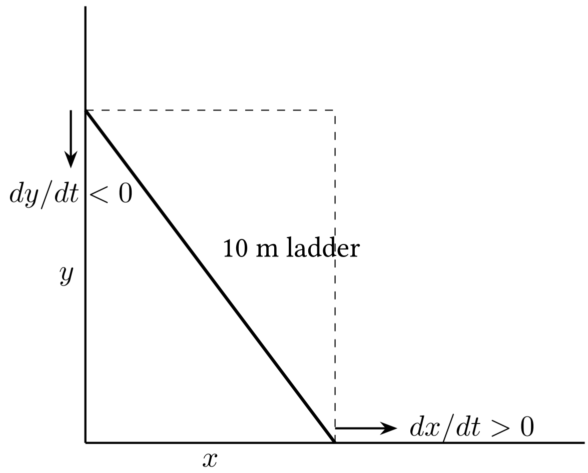


Figure 6: Related rates translate a geometric constraint into a relationship between rates.

7 Mean Value Theorem and local linear thinking

Mean Value Theorem

If f is continuous on $[a, b]$ and differentiable on (a, b), then there exists $c \in (a, b)$ such that

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

The instantaneous rate somewhere equals the average rate over the interval.

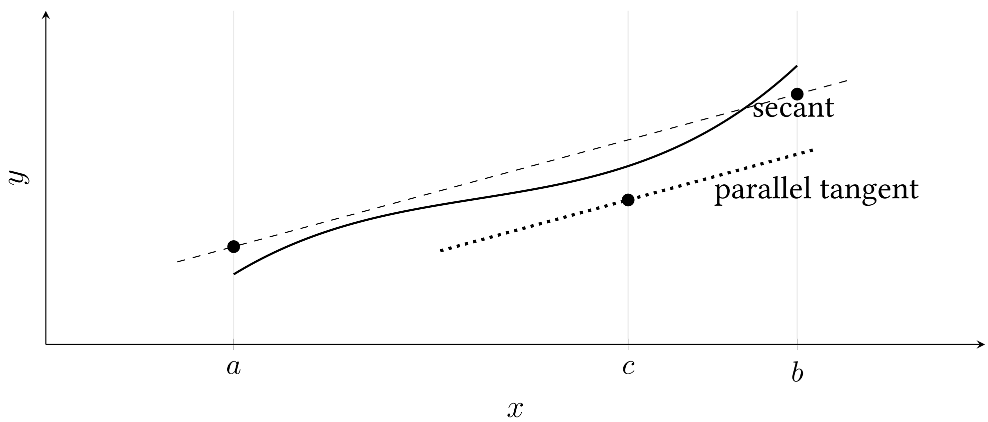


Figure 7: The Mean Value Theorem links an average rate to at least one instantaneous rate.

Worked example 10: interpreting an average-velocity claim

A car's position along a straight route increases by 120 km over 1.5 hours. If its position function is continuous on the time interval and differentiable in the interior, the Mean Value Theorem says that at some instant its velocity was exactly

$$\frac{120}{1.5} = 80 \text{ km/h}.$$

This does not say the velocity was always 80 km/h. It says that an instantaneous velocity equal to the average velocity occurred at least once. If the car moved only forwards, its speed at that instant was also 80 km/h.

What students should take away

- A derivative is a local rate, slope, and sensitivity.
- The difference quotient is the bridge from average rate to instantaneous rate.
- Rules speed up calculation, but the chain rule is the main rule for composite models.
- Non-differentiability is often meaningful, not merely a technical failure.
- Implicit and related-rate methods are essential when the model is constrained geometrically.
- The MVT connects local and average behaviour, which is why derivatives support real-world interpretation.

8 Optional extension: L'Hôpital's Rule

Reserve material

L'Hôpital's Rule belongs after students understand derivatives. If a limit has indeterminate form $0/0$ or ∞/∞, and the relevant derivative quotient limit exists, then

$$\lim_{x\to a}\frac{f(x)}{g(x)}=\lim_{x\to a}\frac{f'(x)}{g'(x)}.$$

It should not be used before checking the indeterminate form.

Optional example: exponential dominates polynomial

Evaluate

$$\lim_{x\to\infty}\frac{x^2}{e^x}.$$

The form is ∞/∞. Applying L'Hôpital twice gives

$$\lim_{x\to\infty}\frac{2x}{e^x}\to\lim_{x\to\infty}\frac{2}{e^x}=0.$$

Thus exponential growth eventually dominates quadratic growth.

9 Exercises

Core practice

1. Use the definition of derivative to find the derivative of $f(x)=3x^2-2x$.
2. Differentiate:
$$f(x)=5x^4-3x^{-2}+7\sqrt{x}-9.$$
3. Differentiate $g(x)=(x^2+1)(\sin x+x)$.
4. Differentiate $h(x)=\dfrac{x^2+3x}{x-2}$.
5. Differentiate $p(x)=\sin(3x^2+1)$ and explain which part is the inner function.
6. Find the tangent line to $x^2+y^2=13$ at $(2,3)$.

Practice check and hints

1. The derivative from the definition is $6x - 2$.
2. The result is $20x^3 + 6x^{-3} + 7/(2\sqrt{x})$ on the original real domain.
3. Use the product rule to obtain $2x(\sin x + x) + (x^2 + 1)(\cos x + 1)$.
4. The quotient-rule result simplifies to $(x^2 - 4x - 6)/(x - 2)^2$.
5. The inner function is $3x^2 + 1$, and $p'(x) = 6x\cos(3x^2 + 1)$.
6. The tangent has slope $-2/3$: $y - 3 = -\frac{2}{3}(x - 2)$.

Modeling practice

1. A queue waiting-time model is $W(\rho) = \dfrac{4\rho}{1-\rho}$ for $0 < \rho < 1$. Compute $W'(\rho)$ and explain what happens as $\rho \to 1^-$. Why is this important for capacity planning?
2. A pollutant concentration follows $C(t) = 60e^{-0.12t} + 8$. Compute $C'(t)$ and determine when the magnitude of the rate first falls below 1 unit/day.
3. A circular infection front has radius $r(t)$ and area $A(t) = \pi r(t)^2$. If $dr/dt = 0.3$ km/day, find dA/dt when $r = 5$ km.
4. A seasonal sales model is $S(t) = 200 + 40\cos(\pi t/6)$. Find when sales are increasing fastest and interpret your answer.

Modeling checks

For Question 1, $W'(\rho) = 4/(1-\rho)^2$, which diverges as $\rho \to 1^-$. For Question 2, $|C'(t)|$ crosses 1 at $t = \ln(7.2)/0.12 \approx 16.45$ days and is below 1 thereafter. Question 3 gives $dA/dt = 3\pi$ km^2/day. In Question 4, the fastest increase occurs at $t = 9 + 12k$; within one 12-month cycle, use $t = 9$. Include units, the stated domain, and a decision-relevant interpretation.

Challenge and extension

1. Show that if $f'(x) = 0$ for all x in an interval, then f is constant on that interval. State clearly where the Mean Value Theorem is used.
2. Differentiate $y = x^x \sin x$ for $x > 0$.
3. For the parametric curve $x = t^2 + 1$, $y = t^3 - t$, find dy/dx and the slope at $t = 2$.
4. Use L'Hôpital's Rule to evaluate
$$\lim_{x\to 0} \frac{e^x - 1 - x}{x^2}.$$
5. Write a short modeling paragraph explaining why a derivative-based approximation may fail near a kink or threshold.

Challenge checks

For Question 1, apply the Mean Value Theorem on any closed subinterval of the given interval. For Question 2, $y' = x^x\left[(\ln x + 1)\sin x + \cos x\right]$ for $x > 0$. For Question 3, $dy/dx = (3t^2 - 1)/(2t)$ for $t \neq 0$, and the slope at $t = 2$ is $11/4$. Question 4 has limit $1/2$. Question 5 should identify the kink or threshold, explain why one local slope cannot cross it, and suggest a piecewise or one-sided model.

Model audit card

Audit before trusting or communicating a model

Purpose	What question or decision does the model support?	**Data**	Which inputs are observed, derived, or simulated?
Assumptions	Which simplifications may change the answer?	**Units**	Are dimensions and scales consistent?
Verification	Were equations, code, constraints, and limiting cases checked?	**Validation**	Is performance tested against independent evidence?
Uncertainty	How do measurement, parameter, solver, and structural errors affect conclusions?	**Ethics**	Who benefits, who bears risk, and what fairness criterion matters?
Reproducibility	Are data provenance, versions, seeds, and transformations recorded?	**Communication**	Are claims conditional, traceable, and decision-relevant?

Further reading

- Calculus Volume 1 (OpenStax)

Sources, provenance, and licence

These notes follow the iterative modeling and assessment perspective in the SIAM/COMAP GAIMME report. Competition-specific remarks are a snapshot checked on 20 August 2026 against the official COMAP HiMCM instructions; readers should verify current rules before submission. Unless a source is explicitly identified, classroom datasets and numerical scenarios are synthetic or simulated teaching examples and are not empirical claims about a real institution. Third-party references retain their own terms.

Student response guide and checking criteria

For closed calculations, show the definition, substitution, units, and at least one independent check; the worked examples in this unit are method hints, but you must complete the arithmetic yourself. Open modeling responses are checked on five dimensions: a clear purpose and output; stated data and provenance; defensible assumptions and scope; traceable computational verification and model validation; and alignment among uncertainty, limitations, and the recommendation. Support each dimension with concrete evidence rather than writing only "completed."

Lectures in Mathematical Modeling

Supplementary 6: Applications of Derivatives and Optimization

Kenneth Sok Kin Cheng

Revised 20 August 2026

Learning goals

By the end of this two-hour supplementary lesson, students should be able to:

- use the sign of f' to determine **increasing/decreasing behaviour**;
- locate and classify **critical points**;
- use f'' to discuss **concavity** and **inflection points**;
- sketch a curve from domain, asymptotes, derivative sign, and concavity;
- formulate and solve a one-variable **optimization** model;
- communicate derivative-based conclusions as modeling evidence, not just computation.

1 Derivative evidence for optimization and model design

In a modeling report, a derivative is rarely an end in itself. It is evidence for a statement such as:

> "The waiting time increases slowly at first, then rises sharply near capacity."

The first derivative tells us the *direction and rate of change*; the second derivative tells us whether the change is accelerating or slowing down. These ideas turn calculus into a practical language for decisions: when to stop production, where a maximum occurs, whether a policy has diminishing returns, and whether a model becomes unstable near a boundary.

Launch activity: reading a derivative without calculating

A robot delivery system has travel-time function $T(q)$, where q is the number of orders per hour. Suppose that at $q = 40$, $T'(40) > 0$ and $T''(40) > 0$.
What can you say in words? A good answer is not "the derivative is positive." It should say:

> At 40 orders per hour, increasing demand will increase travel time, and the increase is itself accelerating. The system is moving toward congestion.

2 Extreme Values and Critical Points

Extreme values

Let f be defined on a domain D.

- $f(c)$ is an **absolute maximum** if $f(c) \geq f(x)$ for all $x \in D$.
- $f(c)$ is an **absolute minimum** if $f(c) \leq f(x)$ for all $x \in D$.

- $f(c)$ is a **local maximum** or **local minimum** if the comparison only needs to hold near c.

A point c is a **critical point** if $f'(c) = 0$ or $f'(c)$ does not exist.

Closed-interval method

If f is continuous on $[a, b]$, then the **Extreme Value Theorem** guarantees an absolute maximum and an absolute minimum. To find them:

1. find all critical points in (a, b);
2. evaluate f at the critical points and endpoints;
3. compare the values.

Worked example 1: absolute extrema on a closed interval

Find the absolute maximum and minimum of

$$f(x) = 2x^3 - 3x^2 - 12x + 5 \qquad \text{on } [-2, 3].$$

First,

$$f'(x) = 6x^2 - 6x - 12 = 6(x - 2)(x + 1).$$

Critical points in $(-2, 3)$ are $x = -1$ and $x = 2$. Now compare the candidate values:

x	-2	-1	2	3
$f(x)$	1	12	-15	-4

Therefore the absolute maximum is 12 at $x = -1$, and the absolute minimum is -15 at $x = 2$.

Checkpoint: Checkpoint

Why do endpoints matter? Because an absolute maximum or minimum on a closed interval may occur at the boundary even when $f'(x) \neq 0$ there.

3 First Derivative Test: Direction of Change

Monotonicity from the sign of f'

If $f'(x) > 0$ on an interval, then f is **increasing** there. If $f'(x) < 0$, then f is **decreasing**. This follows from the **Mean Value Theorem**.

First derivative test

Let c be a critical point where f is continuous.

- If f' changes from $+$ to $-$ at c, then $f(c)$ is a local maximum.
- If f' changes from $-$ to $+$ at c, then $f(c)$ is a local minimum.
- If f' does not change sign, then c is not a local extremum.

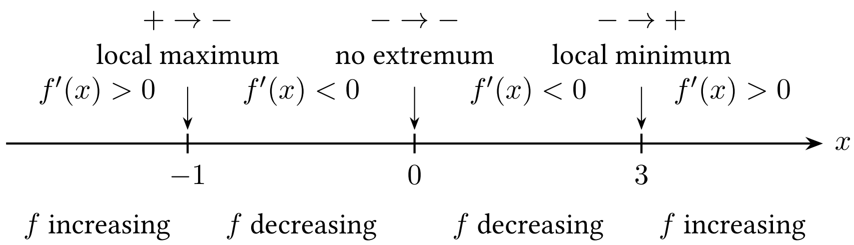


Figure 1: A first-derivative sign chart. Since $f'(x) > 0$ means f is increasing and $f'(x) < 0$ means f is decreasing, a change from $+$ to $-$ gives a local maximum, while a change from $-$ to $+$ gives a local minimum. At $x = 0$, the sign of f' does not change, so there is no extremum.

Worked example 2: first derivative test

Analyse local extrema of

$$f(x) = x^4 - 4x^3.$$

We compute

$$f'(x) = 4x^3 - 12x^2 = 4x^2(x - 3).$$

Critical points are $x = 0$ and $x = 3$. Since $4x^2 \geq 0$, the sign of $f'(x)$ away from 0 is determined by $x - 3$:

x	$(-\infty, 0)$	0	$(0, 3)$	3	$(3, \infty)$
$f'(x)$	$-$	0	$-$	0	$+$
f	$\searrow$		$\searrow$		$\nearrow$

At $x = 0$, f' does not change sign, so there is no extremum. At $x = 3$, f' changes from negative to positive, so f has a local minimum, with $f(3) = -27$.

4 Concavity and the Second Derivative

Concavity and inflection

The second derivative measures how the slope changes.

- If $f''(x) > 0$, then f' is increasing and the graph is **concave up**.
- If $f''(x) < 0$, then f' is decreasing and the graph is **concave down**.
- An **inflection point** occurs where concavity changes.

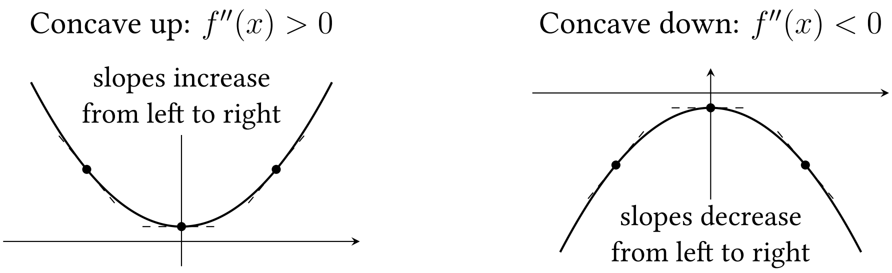


Figure 2: Concavity describes how tangent slopes change. For a concave-up graph, $f''(x) > 0$, the tangent slopes increase from left to right. For a concave-down graph, $f''(x) < 0$, the tangent slopes decrease from left to right.

Second derivative test

Suppose $f'(c) = 0$ and f'' is continuous near c.

- If $f''(c) > 0$, then f has a local minimum at c.
- If $f''(c) < 0$, then f has a local maximum at c.
- If $f''(c) = 0$, the test is inconclusive. Use the first derivative test.

Worked example 3: concavity and inflection

Find the inflection points of

$$f(x) = x^4 - 6x^2 + 1.$$

We have

$$f'(x) = 4x^3 - 12x, \qquad f''(x) = 12x^2 - 12 = 12(x-1)(x+1).$$

The possible inflection points are $x = -1$ and $x = 1$. Check the sign of f'':

x	$(-\infty, -1)$	$(-1, 1)$	$(1, \infty)$
$f''(x)$	$+$	$-$	$+$
concavity	up	down	up

Concavity changes at both $x = -1$ and $x = 1$. Since $f(-1) = f(1) = -4$, the inflection points are $(-1, -4)$ and $(1, -4)$.

5 Curve Sketching as Model Diagnosis

A curve sketch is not only a drawing exercise. In modeling, it diagnoses where a formula is valid, where it has unrealistic blow-up, where the response saturates, and where the model predicts a turning point.

Curve sketching checklist

For $y = f(x)$, analyse:

1. domain and excluded values;
2. intercepts and symmetry;
3. vertical, horizontal, or oblique asymptotes;
4. sign chart for f' and local extrema;
5. sign chart for f'' and inflection points;
6. interpretation in the original context.

Worked example 4: rational curve sketch

Sketch and interpret

$$f(x) = \frac{x^2}{x^2 - 1}.$$

The domain is $x \neq \pm 1$. The function is even. There are vertical asymptotes at $x = -1$ and $x = 1$, and

$$\lim_{x \to \pm\infty} \frac{x^2}{x^2 - 1} = 1,$$

so $y = 1$ is a horizontal asymptote. Also,

$$f'(x) = \frac{-2x}{(x^2-1)^2},$$

so f increases for $x < 0$ and decreases for $x > 0$ on each domain interval. The point $(0, 0)$ is a local maximum on the middle branch.

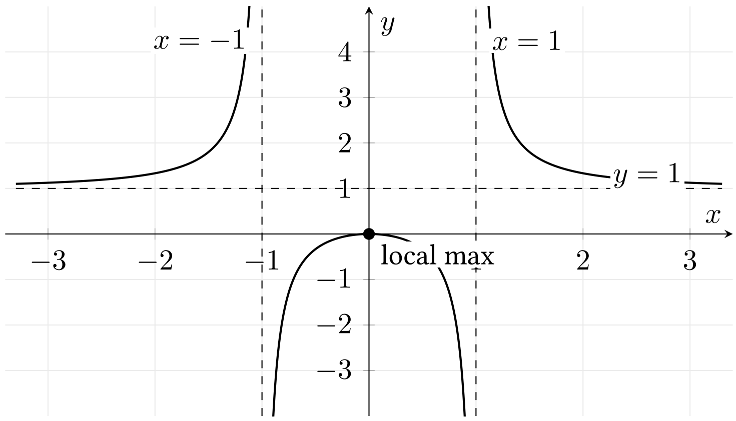


Figure 3: A clean curve sketch for $f(x) = x^2/(x^2 - 1)$. Labels are placed outside dense regions and boxed with white background to avoid overlap with the branches and asymptotes.

Model interpretation. If this formula represented a response ratio, then $x = \pm 1$ would be dangerous thresholds: the model predicts unbounded behaviour there. The horizontal asymptote $y = 1$ suggests the response stabilises far from the thresholds.

Model criticism: Model critique: derivative evidence

A derivative-based sketch is only as meaningful as the formula and domain. When using calculus in a modeling report, state:

- the domain used in the real problem;
- whether excluded points are physical boundaries or mathematical artefacts;
- whether maxima/minima occur inside the feasible range or at a boundary;
- whether a vertical asymptote indicates real danger, capacity failure, or a model breakdown.

6 Applied Optimization

Optimization workflow

A good optimization solution has four layers:

1. define variables with units;
2. write the objective and constraints;
3. reduce to a one-variable function on a realistic domain;
4. verify the optimum and interpret it in context.

Skipping the domain is one of the most common modeling mistakes.

Worked example 5: fencing beside a river

A farmer has 200 m of fencing and wants to enclose a rectangular field beside a straight river. No fence is needed along the river. Find the dimensions that maximize the area.
Let x be the side parallel to the river and y the width. Then

$$x + 2y = 200, \qquad A = xy.$$

Using $x = 200 - 2y$,

$$A(y) = (200 - 2y)y = 200y - 2y^2, \qquad 0 < y < 100.$$

Now

$$A'(y) = 200 - 4y,$$

so $A'(y) = 0$ gives $y = 50$. Since $A''(y) = -4 < 0$, this gives a maximum. Therefore

$$y = 50 \text{ m}, \qquad x = 100 \text{ m}, \qquad A_{\max} = 5000 \text{ m}^2.$$

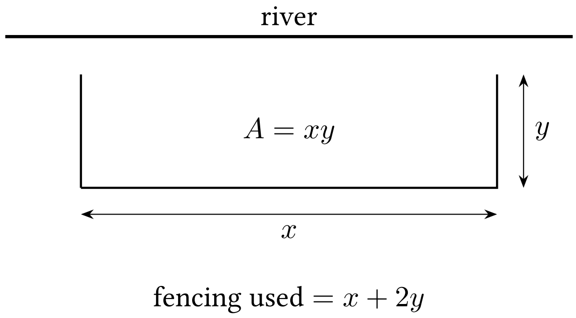


Figure 4: A rectangular field is built against a river, so only three sides need fencing. Thus $A = xy$ and the fencing constraint is $x + 2y = 200$.

Worked example 6: profit-maximizing production

A manufacturer sells q units per week. Cost is

$$C(q) = 0.01q^2 + 10q + 400,$$

and the price-demand equation is

$$p(q) = 50 - 0.02q.$$

Find the weekly production level maximizing profit.
Revenue is

$$R(q) = qp(q) = 50q - 0.02q^2.$$

Profit is

$$P(q) = R(q) - C(q) = (50q - 0.02q^2) - (0.01q^2 + 10q + 400) = 40q - 0.03q^2 - 400.$$

Then

$$P'(q) = 40 - 0.06q.$$

Set $P'(q) = 0$:

$$q = \frac{40}{0.06} \approx 666.7.$$

Since $P''(q) = -0.06 < 0$, this is a maximum for the continuous model. In practice, choose either $q = 666$ or $q = 667$ and compare integer profit.

Checkpoint: Checkpoint

Why might a continuous optimum be inappropriate for production quantity? Because real production may require an integer number of units, batch sizes, machine capacity, or demand uncertainty. The calculus result is a guide, not automatically the final decision.

7 Local Approximation and Newton's Method

For a differentiable function, the tangent line gives a local linear approximation:

$$f(x) \approx f(a) + f'(a)(x - a).$$

This is useful when exact calculation is difficult, or when a model needs a quick sensitivity estimate.

Worked example 7: local linear approximation

Approximate $\sqrt{4.1}$ using $f(x) = \sqrt{x}$ near $a = 4$.

$$f(4) = 2, \qquad f'(x) = \frac{1}{2\sqrt{x}}, \qquad f'(4) = \frac{1}{4}.$$

Therefore

$$\sqrt{x} \approx 2 + \frac{1}{4}(x - 4).$$

At $x = 4.1$,

$$\sqrt{4.1} \approx 2 + \frac{1}{4}(0.1) = 2.025.$$

This is close to the actual value $2.0248\ldots$.

Newton's method

To solve $f(x) = 0$, start with a guess x_0 and iterate

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}.$$

Geometrically, x_{n+1} is the x-intercept of the tangent line at x_n.

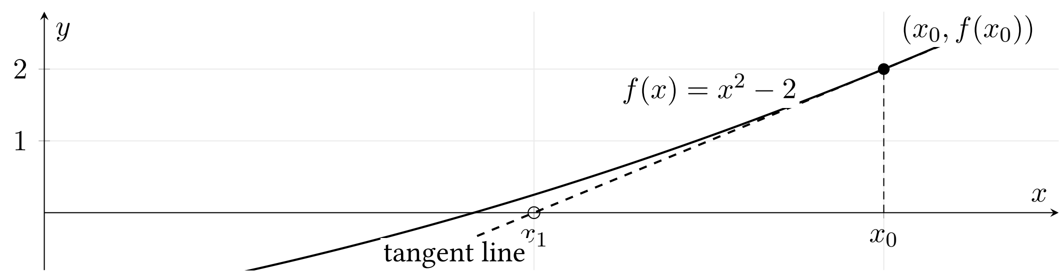


Figure 5: One step of Newton's method for $f(x) = x^2 - 2$. Starting from x_0, the tangent line meets the x-axis at the next approximation x_1.

Worked example 8: approximating $\sqrt{2}$

Solve $f(x) = x^2 - 2 = 0$ with $x_0 = 1$.

$$x_{n+1} = x_n - \frac{x_n^2 - 2}{2x_n} = \frac{1}{2}\left(x_n + \frac{2}{x_n}\right).$$

n	x_n
0	1.000000
1	1.500000
2	1.416667
3	1.414216
4	1.414214

Newton's method converges very quickly here because the initial guess is reasonable and the root is simple.

8 Differentiated Optimization Studio

15-minute studio

Choose one route.
Core route. A box with square base and open top must have volume 32 m^3. Find the dimensions minimizing surface area.
Challenge route. Find the largest rectangle that can be inscribed in the ellipse

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1.$$

Extension route. Use Newton's method to approximate the positive solution of $\cos x = x$ starting from $x_0 = 0.7$. Explain why your iteration is credible.

Model criticism: How to write the modeling conclusion

A strong derivative-based conclusion should contain:

1. the objective being optimized;
2. the feasible domain;
3. the critical point or boundary comparison;
4. the verification method;
5. the practical interpretation and limitation.

For example:

> Within the continuous production model, the profit function is concave because $P''(q) < 0$, so the critical point gives the global maximum. The optimal continuous production level is about 667 units per week. In an operational plan, this should be rounded only after checking integer profit and capacity constraints.

Summary

Key takeaways

Derivative information	Modeling interpretation
$f'(x) > 0$	output increases as input increases.
$f'(x) < 0$	output decreases as input increases.
f' changes $+ \to -$	local maximum.
f' changes $- \to +$	local minimum.
$f''(x) > 0$	slope is increasing; response accelerates; concave up.
$f''(x) < 0$	slope is decreasing; diminishing returns or concave down.
Vertical asymptote	possible capacity limit, singularity, or model breakdown.
Closed interval comparison	absolute optimum may occur at a boundary.
Newton iteration	tangent-line method for solving equations numerically.

Exercises

Core practice

1. Find the absolute maximum and minimum of $f(x) = x^3 - 3x^2 + 1$ on $[-1/2, 4]$.
2. For $f(x) = 2x^3 + 3x^2 - 36x + 5$, find critical points, intervals of increase/decrease, and local extrema.
3. Find all inflection points of $f(x) = x^5 - 5x^4 + 5x^3$.
4. Sketch $f(x) = \dfrac{x^2 - 4x + 3}{x - 2}$ using the curve-sketching checklist.
5. A manufacturer has cost $C(q) = 0.01q^2 + 10q + 400$ and price $p = 50 - 0.02q$. Rework the profit example and compare the integer choices $q = 666$ and $q = 667$.

Core-practice checks

For Question 1, compare the endpoints with the critical points $x = 0, 2$: the absolute minimum is -3 at $x = 2$ and the absolute maximum is 17 at $x = 4$. For Question 2, the critical points are $x = -3, 2$, giving a local maximum of 86 and a local minimum of -39. In Question 3, test concavity on the four intervals cut by $x = 0$ and $x = (3 \pm \sqrt{3})/2$. For Question 4, polynomial division gives $f(x) = x - 2 - 1/(x - 2)$; use this form to identify both asymptotes. For Question 5, the integer profits are $P(666) = 12933.32$ and $P(667) = 12933.33$ in the stated monetary units.

Modeling challenge

A cafeteria counter can serve at most 60 students per minute. A simple congestion model for average waiting time is

$$W(q) = \frac{q}{60 - q}, \qquad 0 < q < 60,$$

where q is arrival rate.

1. Find $W'(q)$ and $W''(q)$.
2. Interpret the signs of W' and W''.
3. What happens as $q \to 60^-$?
4. Write a short recommendation explaining why operating near capacity is risky.

Modeling-challenge check

You should obtain

$$W'(q) = \frac{60}{(60-q)^2} > 0, \qquad W''(q) = \frac{120}{(60-q)^3} > 0, \qquad \lim_{q \to 60^-} W(q) = +\infty.$$

A defensible recommendation names the feasible domain, explains both the increasing delay and its acceleration, and treats the blow-up as a warning about capacity or model breakdown rather than a literal infinite prediction.

Extension practice

1. Use Newton's method to find the positive root of $x^3 + 2x - 5 = 0$ starting from $x_0 = 1$.
2. Find the Maclaurin polynomial $P_4(x)$ for $\cos x$ and use it to approximate $\cos(0.2)$.
3. Find the dimensions of the largest rectangle inscribed under the semicircle $x^2 + y^2 = R^2$, $y \geq 0$.

Extension hints

For Question 1, use $x_{n+1} = x_n - (x_n^3 + 2x_n - 5)/(3x_n^2 + 2)$; the positive root is approximately 1.32827. For Question 2, $P_4(x) = 1 - x^2/2 + x^4/24$, so $P_4(0.2) \approx 0.9800667$. For Question 3, write the rectangle area as $A(x) = 2x\sqrt{R^2 - x^2}$: the maximizing dimensions are $\sqrt{2}R$ by $R/\sqrt{2}$.

Model audit card

Audit before trusting or communicating a model

Purpose	What question or decision does the model support?	**Data**	Which inputs are observed, derived, or simulated?
Assumptions	Which simplifications may change the answer?	**Units**	Are dimensions and scales consistent?
Verification	Were equations, code, constraints, and limiting cases checked?	**Validation**	Is performance tested against independent evidence?
Uncertainty	How do measurement, parameter, solver, and structural errors affect conclusions?	**Ethics**	Who benefits, who bears risk, and what fairness criterion matters?
Reproducibility	Are data provenance, versions, seeds, and transformations recorded?	**Communication**	Are claims conditional, traceable, and decision-relevant?

Further reading

- Calculus Volume 1 (OpenStax)
- Nonlinear Constrained Optimization (NEOS Guide)

Sources, provenance, and licence

These notes follow the iterative modeling and assessment perspective in the SIAM/COMAP GAIMME report. Competition-specific remarks are a snapshot checked on 20 August 2026 against the official COMAP HiMCM instructions; readers should verify current rules before submission. Unless a source is explicitly identified, classroom datasets and numerical scenarios are synthetic or simulated teaching examples and are not empirical claims about a real institution. Third-party references retain their own terms.

Student response guide and checking criteria

For closed calculations, show the definition, substitution, units, and at least one independent check; the worked examples in this unit are method hints, but you must complete the arithmetic yourself. Open modeling responses are checked on five dimensions: a clear purpose and output; stated data and provenance; defensible assumptions and scope; traceable computational verification and model validation; and alignment among uncertainty, limitations, and the recommendation. Support each dimension with concrete evidence rather than writing only “completed.”

Lectures in Mathematical Modeling

Supplementary 7: Definite Integrals, Accumulation, and Numerical Integration

Kenneth Sok Kin Cheng

Revised 20 August 2026

Learning goals

By the end of this two-hour supplementary lesson, students should be able to:

- interpret a **definite integral** as signed area, accumulated change, and average value;
- construct and read a **Riemann sum** using left, right, and midpoint rectangles;
- use the **Fundamental Theorem of Calculus** to evaluate basic definite integrals;
- apply **substitution** and **integration by parts** in simple modeling cases;
- recognise when an integral is **improper** and whether it converges;
- compare rectangle, trapezoidal, Simpson, and Monte Carlo thinking as numerical integration tools.

1 From area to accumulated change

A definite integral is not just a way to compute geometric area. In modeling it often represents an accumulated quantity: total water entering a tank, total cost over time, average inventory, probability mass, or total pollutant load. The same notation

$$\int_a^b f(x)\,dx$$

may mean area, total change, or average effect depending on what f represents.

Launch activity: what does the integral measure?

Suppose $r(t)$ is the arrival rate of students at a cafeteria, measured in students per minute. What does

$$\int_0^{20} r(t)\,dt$$

represent? It is not a rate. It is the total number of students who arrive during the first 20 minutes.

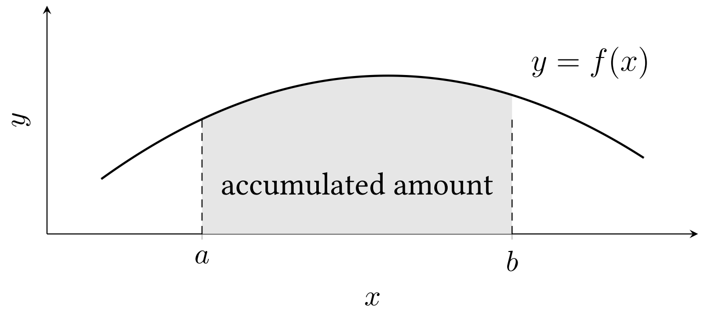


Figure 1: The definite integral accumulates the vertical quantity $f(x)$ across an interval. If f is a rate, the integral is total change. If f is a density, the integral is total mass.

2 Riemann sums: building the integral

Partition and Riemann sum

A **partition** of $[a, b]$ is a list

$$a = x_0 < x_1 < \cdots < x_n = b.$$

The width of the k-th subinterval is $\Delta x_k = x_k - x_{k-1}$. If x_k^* is a sample point in the subinterval, the corresponding **Riemann sum** is

$$\sum_{k=1}^{n} f(x_k^*)\Delta x_k.$$

For a regular partition, $\Delta x = (b - a)/n$. Common choices are left endpoint, right endpoint, and midpoint.

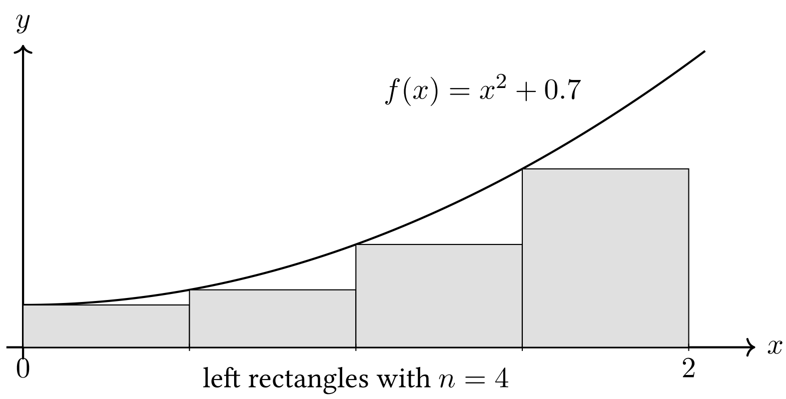


Figure 2: A left Riemann sum uses the height at the left endpoint of each subinterval. For an increasing function, it underestimates the area.

Worked example 1: a left Riemann sum

Approximate

$$\int_0^2 (x^2 + 1)\,dx$$

using a left Riemann sum with $n = 4$ equal subintervals.
Here $\Delta x = 0.5$, and the left sample points are $0, 0.5, 1, 1.5$. Thus

$$S_L = 0.5\,[f(0) + f(0.5) + f(1) + f(1.5)] = 0.5(1 + 1.25 + 2 + 3.25) = 3.75.$$

The exact value will later be $14/3 \approx 4.667$, so the left sum underestimates.

The definite integral

If the Riemann sums approach a unique number as the mesh of the partition tends to zero, we define

$$\int_a^b f(x)\,dx = \lim_{\|P\|\to 0} \sum_{k=1}^{n} f(x_k^*)\Delta x_k.$$

When $f \geq 0$, this is area. When f changes sign, it is **signed area**: area above the x-axis contributes positively, and area below contributes negatively.

Checkpoint: Checkpoint

For an increasing positive function on $[a, b]$, which is usually larger: the left sum or the right sum? Explain using rectangles, not formulas.

3 The Fundamental Theorem of Calculus

The Fundamental Theorem of Calculus connects the two ideas from Supplementary 5 and this lesson: derivative and integral.

Fundamental Theorem of Calculus

Let f be continuous on $[a, b]$.

- **Part I:** If $A(x) = \int_a^x f(t)\,dt$, then $A'(x) = f(x)$.
- **Part II:** If $F'(x) = f(x)$, then

$$\int_a^b f(x)\,dx = F(b) - F(a).$$

The first part says accumulation has rate $f(x)$; the second part says definite integrals can be evaluated by antiderivatives.

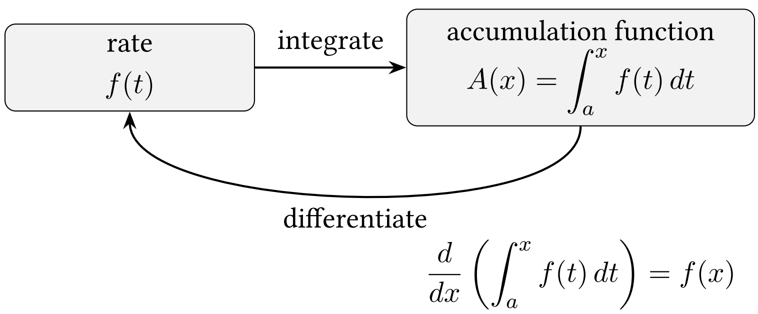


Figure 3: FTC Part I says that integration and differentiation undo each other in the accumulation-function setting.

Worked example 2: evaluating by the FTC

Evaluate

$$\int_0^2 (x^2 + 1)\,dx.$$

An antiderivative is $F(x) = x^3/3 + x$. Hence

$$\int_0^2 (x^2+1)\,dx = \left[\frac{x^3}{3} + x\right]_0^2 = \frac{8}{3} + 2 = \frac{14}{3}.$$

This explains why the $n = 4$ left sum 3.75 was only an approximation.

Worked example 3: differentiating an integral

Compute

$$\frac{d}{dx}\left(\int_2^{x^2} \sin(t^2)\,dt\right).$$

Let $u = x^2$. By FTC Part I and the chain rule,

$$\frac{d}{dx}\int_2^u \sin(t^2)\,dt = \sin(u^2)\frac{du}{dx} = 2x\sin(x^4).$$

The integrand need not have an elementary antiderivative; the derivative is still easy.

4 Substitution and integration by parts

Substitution as reverse chain rule

If $u = g(x)$ and $du = g'(x)\,dx$, then

$$\int f(g(x))g'(x)\,dx = \int f(u)\,du.$$

For a definite integral, change the limits: $x = a \mapsto u = g(a)$ and $x = b \mapsto u = g(b)$.

Worked example 4: substitution with definite limits

Evaluate

$$\int_0^2 xe^{x^2}\,dx.$$

Let $u = x^2$, so $du = 2x\,dx$ and $x\,dx = \frac{1}{2}\,du$. The limits become $0 \mapsto 0$ and $2 \mapsto 4$. Therefore

$$\int_0^2 xe^{x^2}\,dx = \frac{1}{2}\int_0^4 e^u\,du = \frac{1}{2}(e^4 - 1).$$

Integration by parts as reverse product rule

From $(uv)' = u'v + uv'$, we obtain

$$\int u\,dv = uv - \int v\,du.$$

A useful choice rule is LIATE: logarithmic, inverse-trigonometric, algebraic, trigonometric, exponential. The function earlier in the list is often a good choice for u.

Worked example 5: integration by parts

Evaluate

$$\int_0^1 xe^x\,dx.$$

Choose $u = x$ and $dv = e^x\,dx$. Then $du = \,dx$ and $v = e^x$. Thus

$$\int_0^1 xe^x\,dx = [xe^x]_0^1 - \int_0^1 e^x\,dx = e - (e-1) = 1.$$

Checkpoint: Checkpoint

In substitution, the main question is "what inner function has its derivative nearby?" In integration by parts, the main question is "which part becomes simpler after differentiating?"

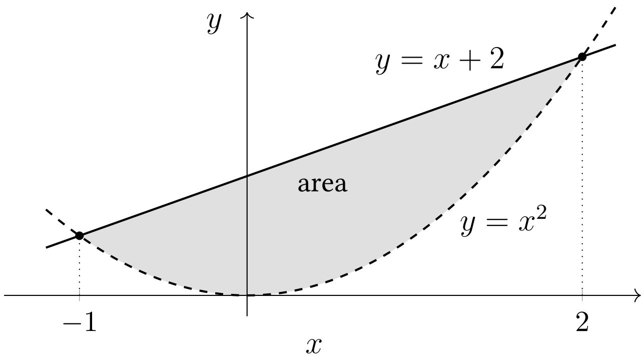
5 Modeling applications of definite integrals

Area between two curves

If $f(x) \geq g(x)$ on $[a, b]$, the area between the curves is

$$A = \int_a^b [f(x) - g(x)]\,dx.$$

The order matters: top curve minus bottom curve.



$$\text{Area} = \int_{-1}^2 \left[(x+2) - x^2\right]dx$$

Figure 4: For area between curves, integrate upper minus lower. The shaded region is bounded by $y = x + 2$ and $y = x^2$ from $x = -1$ to $x = 2$.

Worked example 6: area between curves

Find the area enclosed by $y = x^2$ and $y = x + 2$.
Intersections satisfy $x^2 = x + 2$, so $x = -1$ or $x = 2$. On this interval, $x + 2 \geq x^2$. Hence

$$A = \int_{-1}^2 [(x+2) - x^2]\,dx = \left[\frac{x^2}{2} + 2x - \frac{x^3}{3}\right]_{-1}^2 = \frac{9}{2}.$$

Average value and accumulated change

For a continuous function f on $[a, b]$, its **average value** is

$$\bar{f} = \frac{1}{b-a} \int_a^b f(x)\,dx.$$

If $Q'(t) = r(t)$, then

$$Q(b) - Q(a) = \int_a^b r(t)\,dt.$$

This is the form used in flow, pollution, population, inventory, and energy models.

Worked example 7: average inventory

In the EOQ model, suppose inventory decreases linearly from Q to 0 over a cycle of length T:

$$I(t) = Q\left(1 - \frac{t}{T}\right), \qquad 0 \le t \le T.$$

The average inventory is

$$\bar{I} = \frac{1}{T}\int_0^T Q\left(1 - \frac{t}{T}\right)\,dt = \frac{Q}{T}\left[t - \frac{t^2}{2T}\right]_0^T = \frac{Q}{2}.$$

This familiar result is an integral statement, not only a geometric triangle argument.

Model criticism: Model criticism: units and sign

Before reporting an integral, check its units. If $r(t)$ is litres/hour and t is measured in hours, then $\int r(t)\,dt$ is litres. Also check whether signed area makes sense: negative values may represent outflow, loss, or correction rather than literal negative area.

6 Improper integrals and convergence

Some modeling integrals extend over an infinite interval or include an unbounded integrand. These are **improper integrals**. They must be defined using limits.

Two common improper integrals

For an infinite interval,

$$\int_a^\infty f(x)\,dx = \lim_{b\to\infty} \int_a^b f(x)\,dx.$$

For a vertical blow-up at a,

$$\int_a^b f(x)\,dx = \lim_{c\to a^+} \int_c^b f(x)\,dx.$$

If the limit exists, the integral **converges**. Otherwise it **diverges**.

Worked example 8: convergence versus divergence

Compare

$$\int_1^\infty \frac{1}{x^2}\,dx \qquad \text{and} \qquad \int_1^\infty \frac{1}{x}\,dx.$$

For the first,

$$\int_1^\infty \frac{1}{x^2}\,dx = \lim_{b\to\infty}\left[-\frac{1}{x}\right]_1^b = 1,$$

so it converges. For the second,

$$\int_1^\infty \frac{1}{x}\,dx = \lim_{b\to\infty} \ln b = \infty,$$

so it diverges. Infinite interval does not automatically mean infinite area.

Checkpoint: Checkpoint

Why is the computation $\int_{-1}^1 1/x^2\,dx = [-1/x]_{-1}^1$ invalid? Identify the hidden discontinuity.

7 Numerical integration

Many integrals in modeling do not have elementary antiderivatives. For example, $\int e^{-x^2}\,dx$ appears in statistics, but cannot be expressed using elementary functions. We then approximate.

Trapezoidal and Simpson rules

For a regular partition of $[a, b]$ with n subintervals and $h = (b - a)/n$:

$$T_n = \frac{h}{2}\left[f(x_0) + 2f(x_1) + \cdots + 2f(x_{n-1}) + f(x_n)\right].$$

If n is even, Simpson's rule is

$$S_n = \frac{h}{3}\left[f(x_0) + 4f(x_1) + 2f(x_2) + \cdots + 4f(x_{n-1}) + f(x_n)\right].$$

Trapezoidal uses straight-line panels; Simpson uses quadratic panels.

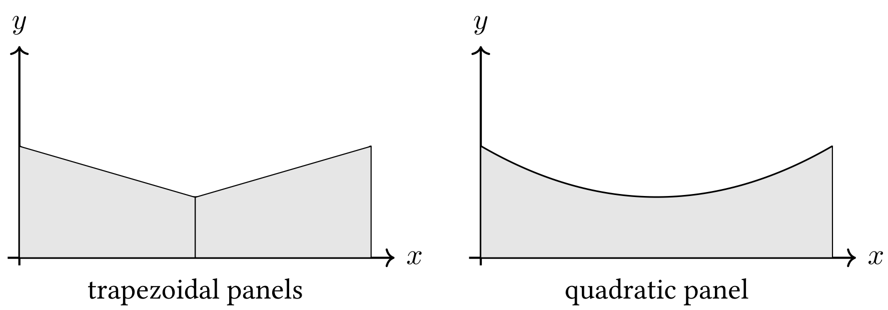


Figure 5: Trapezoidal panels approximate the curve by straight segments. Simpson's rule approximates over pairs of subintervals by a quadratic curve.

Worked example 9: comparing numerical rules

Estimate

$$\int_0^1 e^{-x^2}\,dx$$

using $n = 4$ subintervals. With $h = 0.25$, the function values are approximately

x_k	0	0.25	0.50	0.75	1.00
$e^{-x_k^2}$	1.00000	0.93941	0.77880	0.56978	0.36788

Thus

$$T_4 = 0.74298, \qquad S_4 = 0.74685.$$

The true value is approximately 0.746824, so Simpson's rule is much more accurate here.

Model criticism: Numerical integration in a modeling report

A good report should not only state "we used Simpson's rule." It should specify the interval, step size, stopping criterion, and a validation check. For example:

> We estimated the integral using Simpson's rule with step size $h = 0.01$. Doubling the number of subintervals changed the value by less than 10^{-4}, so numerical error is small relative to data uncertainty.

8 Summary and practice

Key ideas

- Riemann sums approximate integrals by finite pieces.
- The definite integral is an accumulation operator: area, total change, average value, probability, or mass.
- FTC connects accumulated change and instantaneous rate.
- Substitution is reverse chain rule; integration by parts is reverse product rule.
- Improper integrals require limits and may converge or diverge.
- Numerical integration is a modeling choice; accuracy and cost should be justified.

Core practice

1. Compute left, right, and midpoint sums for $\int_0^2 x^2\,dx$ with $n = 4$. Compare with the exact value.
2. Evaluate $\int_1^4 (3x^2 - 2x + 5)\,dx$ using the FTC.
3. Evaluate $\int_0^1 2x(1 + x^2)^5\,dx$ using substitution.
4. Evaluate $\int_1^e \ln x\,dx$ using integration by parts.
5. Determine whether $\int_1^\infty x^{-3/2}\,dx$ converges.
6. Find the average value of $f(x) = x^2$ on $[0, 3]$.

Practice check and hints

1. The left, right, and midpoint sums are $7/4$, $15/4$, and $21/8$; the exact value is $8/3$.
2. The FTC gives 63.
3. With $u = 1 + x^2$, the integral is $21/2$.
4. Integration by parts gives 1.
5. The improper integral converges to 2.
6. The average value is 3.

Modeling challenge: cafeteria inflow

During the first 30 minutes of lunch, the arrival rate of students is modeled by

$$r(t) = 18 + 0.9t - 0.03t^2, \qquad 0 \le t \le 30,$$

where t is measured in minutes and $r(t)$ in students per minute.

1. Compute the total number of students arriving in the first 30 minutes.
2. Find the average arrival rate.
3. Explain one assumption behind using a smooth rate function for student arrivals.
4. Suggest one way to validate the model using observed queue data.

Modeling checks

The model predicts 675 arrivals and an average rate of 22.5 students per minute. A complete response also distinguishes a smooth expected rate from discrete arrivals, names a data-collection interval and measurement rule, and proposes an out-of-sample comparison between integrated predictions and observed counts.

Extension practice

1. Show that if f is odd and integrable on $[-a, a]$, then $\int_{-a}^{a} f(x)\,dx = 0$.
2. Use Simpson's rule with $n = 4$ to approximate $\int_0^1 \frac{1}{1+x^2}\,dx$ and compare with $\pi/4$.
3. Verify that $\int_0^\infty \lambda e^{-\lambda x}\,dx = 1$ for $\lambda > 0$, and explain why this matters for probability density models.

Extension hints

For Question 1, substitute $u = -x$ on one half of the interval. For Question 2, Simpson's rule gives $8011/10200 \approx 0.785392$, close to $\pi/4 \approx 0.785398$. For Question 3, evaluate the truncated integral first and then let its upper limit tend to infinity; the value is 1, as required for a probability density.

Model audit card

Audit before trusting or communicating a model

Purpose	What question or decision does the model support?	**Data**	Which inputs are observed, derived, or simulated?
Assumptions	Which simplifications may change the answer?	**Units**	Are dimensions and scales consistent?
Verification	Were equations, code, constraints, and limiting cases checked?	**Validation**	Is performance tested against independent evidence?
Uncertainty	How do measurement, parameter, solver, and structural errors affect conclusions?	**Ethics**	Who benefits, who bears risk, and what fairness criterion matters?
Reproducibility	Are data provenance, versions, seeds, and transformations recorded?	**Communication**	Are claims conditional, traceable, and decision-relevant?

Further reading

- Calculus Volume 1 (OpenStax)
- Calculus Volume 2 (OpenStax)

Sources, provenance, and licence

These notes follow the iterative modeling and assessment perspective in the SIAM/COMAP GAIMME report. Competition-specific remarks are a snapshot checked on 20 August 2026 against the official COMAP HiMCM instructions; readers should verify current rules before submission. Unless a source is explicitly identified, classroom datasets and numerical scenarios are synthetic or simulated teaching examples and are not empirical claims about a real institution. Third-party references retain their own terms.

Student response guide and checking criteria

For closed calculations, show the definition, substitution, units, and at least one independent check; the worked examples in this unit are method hints, but you must complete the arithmetic yourself. Open modeling responses are checked on five dimensions: a clear purpose and output; stated data and provenance; defensible assumptions and scope; traceable computational verification and model validation; and alignment among uncertainty, limitations, and the recommendation. Support each dimension with concrete evidence rather than writing only "completed."

Lectures in Mathematical Modeling

Supplementary 8: Probability Essentials for Modeling and Simulation

Kenneth Sok Kin Cheng

Revised 20 August 2026

Learning goals

By the end of this two-hour supplementary lesson, students should be able to:

- describe a random experiment using a **sample space**, **event**, and probability model;
- compute simple probabilities using counting, complements, and inclusion–exclusion;
- distinguish **conditional probability**, **independence**, and disjointness;
- use **Bayes' theorem** to update probabilities from evidence;
- model uncertainty using **random variables**, PMF/PDF, CDF, expectation, and variance;
- recognise when binomial, Poisson, exponential, normal, or Markov-chain models are appropriate;
- explain how the Law of Large Numbers and Central Limit Theorem justify Monte Carlo simulation.

1 Why probability belongs in mathematical modeling

A deterministic model gives one output when the input is fixed. Real systems often require a different language. Customers do not arrive exactly every five minutes; a sensor reading contains noise; a production line occasionally creates defective products; weather changes from day to day. A modeling report should therefore be able to say not only "what will happen", but also "how likely", "with what variability", and "how confident are we?"

Launch activity: deterministic or stochastic?

Classify each situation as mainly deterministic, stochastic, or mixed.

1. A robot travels 120 m at a constant speed of 1.5 m/s.
2. Customers enter a canteen during lunch time.
3. A component has a 2% chance of failing during one month.
4. A delivery route is planned, but travel time depends on weather and crowding.

A strong modeling answer usually separates the predictable structure from the random variation.

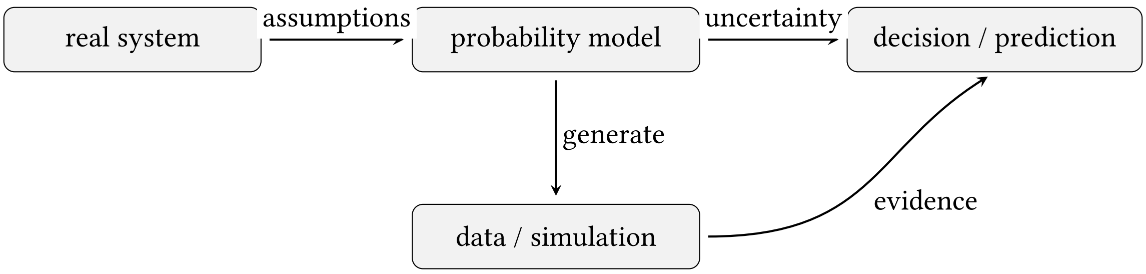


Figure 1: Probability enters modeling when assumptions must describe uncertainty, variability, and evidence rather than a single fixed outcome.

2 Sample spaces, events, and counting

Sample space and event

A **random experiment** is a procedure whose outcome is uncertain but whose possible outcomes can be described. The **sample space** Ω is the set of all possible outcomes. An **event** is a subset $A \subseteq \Omega$.

Worked example 1: two dice

For rolling two fair dice,

$$\Omega = \{(i,j) : 1 \le i,j \le 6\}, \qquad |\Omega| = 36.$$

The event "sum is 7" is

$$\{(1,6),(2,5),(3,4),(4,3),(5,2),(6,1)\},$$

so $\mathbb{P}(\text{sum} = 7) = 6/36 = 1/6$.

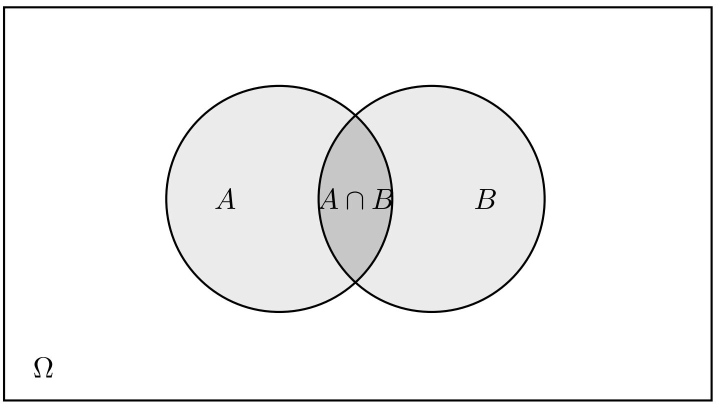
Probability rules

For events A and B,

$$\mathbb{P}(A^c) = 1 - \mathbb{P}(A), \qquad \mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B).$$

If all outcomes are equally likely and Ω is finite, then

$$\mathbb{P}(A) = \frac{|A|}{|\Omega|}.$$



$$\mathrm{P}(A \cup B) = \mathrm{P}(A) + \mathrm{P}(B) - \mathrm{P}(A \cap B)$$

Figure 2: Inclusion–exclusion subtracts the overlap once because it is counted in both A and B.

Worked example 2: birthday problem

Assume birthdays are uniformly distributed over 365 days. For $n = 23$ people, compute the probability that at least two people share a birthday.
It is easier to use the complement. The probability all birthdays are distinct is

$$\frac{365 \cdot 364 \cdot 363 \cdots (365 - 22)}{365^{23}} \approx 0.4927.$$

Therefore

$$\mathbb{P}(\text{at least one shared birthday}) \approx 1 - 0.4927 = 0.5073.$$

Only 23 people are enough for a probability slightly above one half.

Checkpoint: Checkpoint

A bag has 5 red balls and 3 blue balls. Two balls are drawn without replacement. Is "first ball red" independent of "second ball red"? Explain using conditional probability.

3 Conditional probability, independence, and Bayes

Conditional probability

For events A and B with $\mathbb{P}(B) > 0$,

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}.$$

This means: restrict the sample space to B, then renormalise.

Independence

Events A and B are **independent** if

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B),$$

or equivalently $\mathbb{P}(A \mid B) = \mathbb{P}(A)$ when $\mathbb{P}(B) > 0$. Independence is not the same as disjointness. If two positive-probability events are disjoint, knowing one happened makes the other impossible, so they are dependent.

Total probability and Bayes' theorem

If $B_1, \dots, B_n$ partition Ω, then

$$\mathbb{P}(A) = \sum_{i=1}^{n} \mathbb{P}(A \mid B_i)\mathbb{P}(B_i).$$

Bayes' theorem updates the probability of a cause after observing evidence:

$$\mathbb{P}(B_k \mid A) = \frac{\mathbb{P}(A \mid B_k)\mathbb{P}(B_k)}{\sum_{i=1}^{n} \mathbb{P}(A \mid B_i)\mathbb{P}(B_i)}.$$

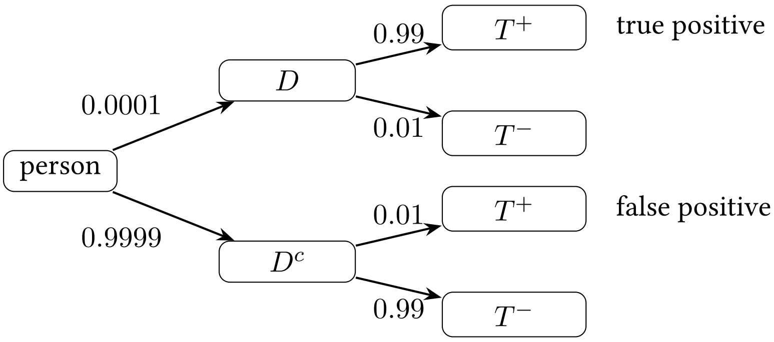


Figure 3: A probability tree separates prior probability from test reliability. The false-positive branch can dominate when the base rate is very small.

Worked example 3: medical testing and base rate

A disease affects 1 in 10,000 people. A test has 99% sensitivity and 99% specificity. A randomly selected person tests positive. What is the probability they have the disease?
Let D be "has disease" and T^+ be "positive test". Then

$$\mathbb{P}(D) = 10^{-4}, \quad \mathbb{P}(T^+ \mid D) = 0.99, \quad \mathbb{P}(T^+ \mid D^c) = 0.01.$$

Therefore

$$\mathbb{P}(D \mid T^+) = \frac{0.99(10^{-4})}{0.99(10^{-4}) + 0.01(0.9999)} \approx 0.0098.$$

Even after a positive test, the probability is under 1%. This is not because the test is useless; it is because the disease is rare.

Model criticism: Model criticism: Bayes in reports

When using Bayes' theorem, a report should state the prior probabilities, evidence reliability, and what changes after observing evidence. Many wrong conclusions come from ignoring the base rate.

4 Random variables, distributions, expectation, and variance

Random variable

A **random variable** is a function $X : \Omega \to \mathbb{R}$ that assigns a numerical value to each outcome. It is **discrete** if its possible values are finite or countable, and **continuous** if probabilities are described by intervals and integrals.

PMF, PDF, and CDF

For a discrete random variable, the **probability mass function** is

$$p_X(x) = \mathbb{P}(X = x), \qquad \sum_x p_X(x) = 1.$$

For a continuous random variable, the **probability density function** satisfies

$$\mathbb{P}(a \le X \le b) = \int_a^b f_X(x)\, dx.$$

For any random variable, the **cumulative distribution function** is

$$F_X(x) = \mathbb{P}(X \le x).$$

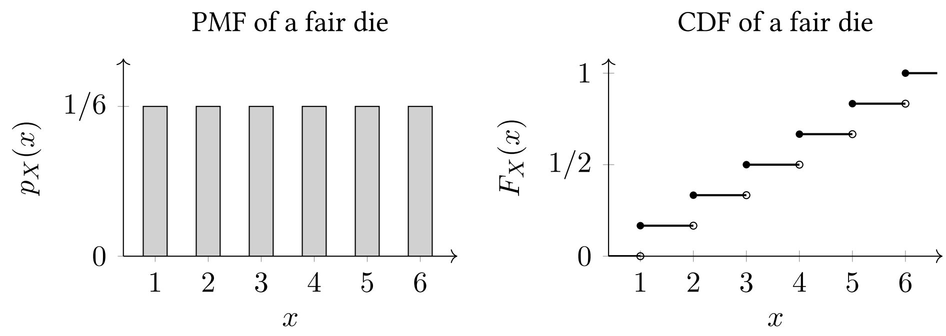


Figure 4: A discrete distribution has mass at points. Its CDF increases in jumps.

Expectation and variance

The **expected value** is the long-run average:

$$\mathbb{E}[X] = \sum_x x p_X(x) \quad \text{(discrete)}, \qquad \mathbb{E}[X] = \int_{-\infty}^{\infty} x f_X(x)\, dx \quad \text{(continuous)}.$$

The **variance** measures spread:

$$\text{Var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2.$$

The standard deviation is $\text{SD}(X) = \sqrt{\text{Var}(X)}$.

Worked example 4: die roll mean and variance

Let X be the result of a fair six-sided die. Then

$$\mathbb{E}[X] = \frac{1+2+3+4+5+6}{6} = 3.5,$$

$$\mathbb{E}[X^2] = \frac{1+4+9+16+25+36}{6} = \frac{91}{6},$$

so

$$\text{Var}(X) = \frac{91}{6} - (3.5)^2 = \frac{35}{12} \approx 2.917.$$

The standard deviation is about 1.708.

Worked example 5: expected profit

An insurer sells a one-year policy for a premium of \$1,500. The claim is \$200,000 if death occurs, and the probability of death in the year is $p = 0.01$. The insurer's expected profit is

$$0.99(1500) + 0.01(1500 - 200000) = 1485 - 1985 = -500.$$

At this premium the policy loses \$500 in expectation. This connects directly to decision theory and risk modeling.

5 Named distributions for modeling

When to use which distribution

Distribution	Modeling question	Key quantities
Bernoulli(p)	one success/failure trial	mean p, variance $p(1-p)$
Binomial(n, p)	number of successes in n independent trials	mean np, variance $np(1-p)$
Poisson(λ)	number of random arrivals in a fixed interval	mean λ, variance λ
Exponential(λ)	waiting time between Poisson arrivals	mean $1/\lambda$, memoryless
Normal(μ, σ^2)	measurement error, averages, natural variability	mean μ, variance σ^2

Worked example 6: binomial quality control

A factory has defect rate $p = 0.02$. In a batch of $n = 100$ items, let X be the number of defective items. Then $X \sim \text{Bin}(100, 0.02)$ and

$$\mathbb{E}[X] = 100(0.02) = 2.$$

The probability of at most 2 defects is

$$\mathbb{P}(X \leq 2) = \sum_{k=0}^{2} \binom{100}{k} (0.02)^k (0.98)^{100-k} \approx 0.677.$$

Worked example 7: Poisson arrivals and exponential waiting

Customers arrive at a service window at average rate 12 per hour. If arrivals are modeled as Poisson, then the number of arrivals in one hour is $X \sim \text{Pois}(12)$:

$$\mathbb{P}(X = 10) = e^{-12}\frac{12^{10}}{10!} \approx 0.105.$$

The waiting time until the next customer is $T \sim \text{Exp}(12)$ hours, so

$$\mathbb{E}[T] = \frac{1}{12}\text{ hour} = 5\text{ minutes}.$$

This is the probability foundation of queue simulation.

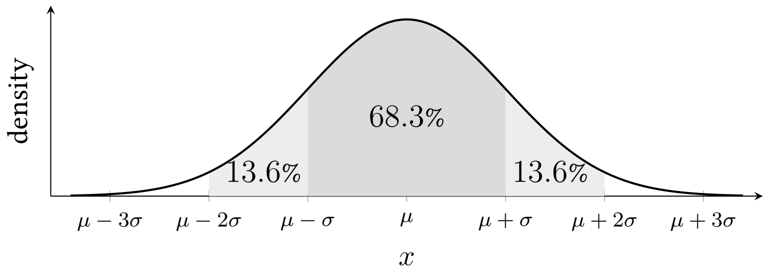


Figure 5: The normal distribution is used for measurement error, natural variation, and sampling distributions. About 68.3% of values lie within one standard deviation of the mean.

Checkpoint: Checkpoint

Which distribution would you use for each case?

1. Number of students arriving in the next 10 minutes.
2. Whether a randomly selected component is defective.
3. Waiting time until the next customer.
4. Measurement error of repeated sensor readings.

6 Monte Carlo, LLN, and CLT

The link between probability and simulation is simple: if we cannot compute a quantity exactly, we can sometimes estimate it by repeated random trials. The theory tells us why the estimate improves.

Law of Large Numbers

Let $X_1, X_2, \dots$ be independent and identically distributed with mean μ. The sample mean

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

converges to μ in probability. In modeling language: repeated simulation averages approach the expected value.

Central Limit Theorem

If $X_1, \dots, X_n$ are i.i.d. with mean μ and variance σ^2, then for large n,

$$\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \approx Z \sim \mathcal{N}(0,1).$$

Therefore an approximate 95% confidence interval for μ is

$$\bar{X}_n \pm 1.96\frac{s}{\sqrt{n}},$$

where s is the sample standard deviation.

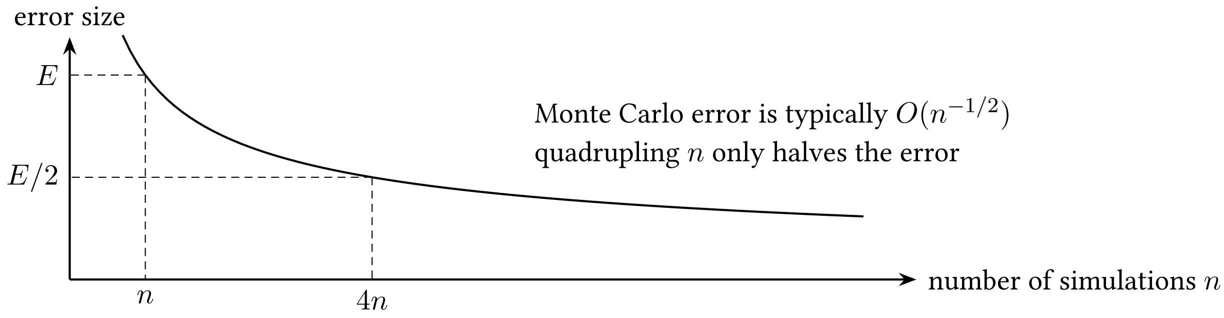


Figure 6: The $1/\sqrt{n}$ convergence rate explains why Monte Carlo is flexible but often slow.

Worked example 8: sample size for Monte Carlo

Suppose a Monte Carlo estimator is an average of Bernoulli indicators, so its standard deviation is at most $1/2$. How large should n be for a normal-approximation 95% confidence interval half-width at most 0.005?

$$1.96\frac{1/2}{\sqrt{n}} \le 0.005 \quad\Longrightarrow\quad \sqrt{n} \ge 196 \quad\Longrightarrow\quad n \ge 38416.$$

Thus the normal-approximation design requires at least 38416 samples under this worst-case variance bound. Finite-sample coverage should be checked separately when the estimated probability is near 0 or 1.

Model criticism: Reporting simulation uncertainty

A competition report should not only give a simulated estimate. It should also report the number of replications, the standard error, a confidence interval, and whether the conclusion changes under reasonable parameter perturbations.

7 A first Markov-chain model

Discrete-time Markov chain

A sequence $X_0, X_1, X_2, \ldots$ on a finite state space is a **Markov chain** if the next state depends on the present state but not on the full past:

$$\mathbb{P}(X_{n+1} = j \mid X_n = i, X_{n-1}, \ldots, X_0) = \mathbb{P}(X_{n+1} = j \mid X_n = i) = P_{ij}.$$

The matrix $P = (P_{ij})$ is the **transition matrix**; each row sums to 1.

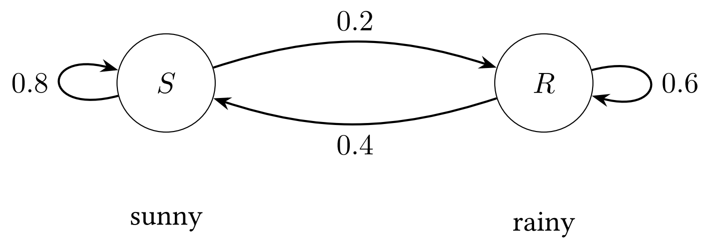


Figure 7: A two-state Markov chain for weather. The probabilities leaving each state add to 1.

Worked example 9: two-day weather prediction

States are S = sunny and R = rainy. Suppose

$$P = \begin{pmatrix} 0.8 & 0.2 \\ 0.4 & 0.6 \end{pmatrix},$$

where rows are today's weather and columns are tomorrow's weather. If today is sunny, the distribution after two days is

$$(1,0)P^2 = (1,0)\begin{pmatrix} 0.72 & 0.28 \\ 0.56 & 0.44 \end{pmatrix} = (0.72, 0.28).$$

Thus the probability of rain two days from now is 0.28.

Worked example 10: stationary distribution

A stationary distribution $\pi = (\pi_S, \pi_R)$ satisfies $\pi P = \pi$ and $\pi_S + \pi_R = 1$. From the weather matrix,

$$\pi_S = 0.8\pi_S + 0.4\pi_R \quad \Longrightarrow \quad 0.2\pi_S = 0.4\pi_R \quad \Longrightarrow \quad \pi_S = 2\pi_R.$$

Together with $\pi_S + \pi_R = 1$, we get

$$\pi = (2/3, 1/3).$$

In the long run, about two-thirds of days are sunny.

8 Differentiated modeling studio

Studio task: canteen queue uncertainty

A canteen receives students at an average rate of 18 per 10 minutes. During peak time, arrivals are modeled as Poisson.

1. Compute the probability of exactly 20 arrivals in the next 10 minutes.
2. Compute or approximate the probability of more than 24 arrivals.
3. If one server can handle 20 students per 10 minutes, what risk does this create?
4. Write one paragraph explaining whether the Poisson assumption is reasonable.

Competition-style reporting paragraph

We model arrivals in each 10-minute period as $X \sim \text{Pois}(18)$, assuming independent arrivals at an approximately constant rate during the short peak interval. This gives $\mathbb{E}[X] = 18$ and $\text{Var}(X) = 18$, so the typical fluctuation is about $\sqrt{18} \approx 4.24$ students. Since the server capacity is 20 students per 10 minutes, the system may become overloaded in a non-negligible fraction of intervals. We therefore recommend evaluating service capacity using not only the mean arrival rate, but also the upper tail probability $\mathbb{P}(X > 20)$ and sensitivity to the estimated rate.

Model criticism: Model criticism

Probability models are assumptions, not facts. A Poisson arrival model assumes independent arrivals and a roughly constant rate. If students arrive in groups after a bell rings, the independence assumption is weak. If the rate changes sharply during lunch, the constant-rate assumption is weak. Validation should compare both the mean and the variability of observed arrivals against the model.

9 Core practice and extension problems

Core practice

1. Two fair dice are rolled. Find $\mathbb{P}(\text{sum} = 8)$ and $\mathbb{P}(\text{sum} \geq 10)$.
2. A bag contains 5 red and 3 blue balls. Two are drawn without replacement. Find the probability both are red.
3. A factory has two machines. Machine A produces 70% of items with defect rate 1%; machine B produces 30% with defect rate 4%. A randomly chosen item is defective. Find the probability it came from B.
4. Let X be the number of heads in 4 tosses of a fair coin. Write its distribution and compute $\mathbb{E}[X]$.
5. If $X \sim \text{Pois}(5)$, compute $\mathbb{P}(X = 3)$ and $\mathbb{P}(X = 0)$.
6. If $T \sim \text{Exp}(6)$ hours, find $\mathbb{E}[T]$ and $\mathbb{P}(T > 0.5)$.

Practice check and hints

1. The probabilities are $5/36$ and $1/6$.
2. The probability is $\binom{5}{2}/\binom{8}{2} = 5/14$.
3. Bayes' rule gives $0.30(0.04)/[0.70(0.01) + 0.30(0.04)] = 12/19$.
4. $X \sim \text{Bin}(4, 1/2)$, so $\mathbb{P}(X = k) = \binom{4}{k}/16$ and $\mathbb{E}[X] = 2$.
5. The probabilities are $e^{-5}5^3/3! \approx 0.14037$ and $e^{-5} \approx 0.006738$.
6. $\mathbb{E}[T] = 1/6$ hour $= 10$ minutes, and $\mathbb{P}(T > 0.5) = e^{-3} \approx 0.0498$.

Modeling challenge

A small help desk receives requests according to a Poisson process with rate $\lambda = 10$ per hour. Service time is approximately exponential with mean 5 minutes.

1. Convert the mean service time into a service rate μ per hour.
2. Compute the utilisation ratio $\rho = \lambda/\mu$.
3. Explain what it means if ρ is close to 1.
4. Suggest one data source that could validate the arrival-rate assumption and one source that could validate the service-time assumption.

Modeling checks

The service rate is $\mu = 12$ per hour and $\rho = 10/12 = 5/6 \approx 0.833$. Interpret this as a fitted load indicator, not proof that Poisson arrivals or exponential service times are valid. A complete validation plan specifies timestamped arrival data, observed service durations, diagnostics for time variation and dependence, and a held-out comparison.

Extension practice

1. Prove that if A and B are independent, then A and B^c are independent.
2. Let $X \sim U(0,1)$ and $Y = -\ln X$. Derive the CDF of Y and identify its distribution.
3. Let $X_1, \ldots, X_n$ be independent Bernoulli(p). Use linearity of expectation to show $\mathbb{E}[\sum X_i] = np$.
4. In the weather Markov chain, if today is rainy, compute the probability of a sunny day three days from now.
5. Give an example where Cov$(X, Y) = 0$ but X and Y are not independent.

Extension hints

Use $\mathbb{P}(A \cap B^c) = \mathbb{P}(A) - \mathbb{P}(A \cap B)$ in Question 1. In Question 2, $F_Y(y) = 1 - e^{-y}$ for $y \geq 0$, so $Y \sim$ Exp(1). Question 3 follows by summing expectations. The three-day sunny probability in Question 4 is 0.624. For Question 5, one example is $X \sim U(-1,1)$ with $Y = X^2$: the covariance is 0, but the variables are dependent.

Model audit card

Audit before trusting or communicating a model

Purpose	What question or decision does the model support?	**Data**	Which inputs are observed, derived, or simulated?
Assumptions	Which simplifications may change the answer?	**Units**	Are dimensions and scales consistent?
Verification	Were equations, code, constraints, and limiting cases checked?	**Validation**	Is performance tested against independent evidence?
Uncertainty	How do measurement, parameter, solver, and structural errors affect conclusions?	**Ethics**	Who benefits, who bears risk, and what fairness criterion matters?
Reproducibility	Are data provenance, versions, seeds, and transformations recorded?	**Communication**	Are claims conditional, traceable, and decision-relevant?

Further reading

- Introductory Statistics 2e (OpenStax)

Sources, provenance, and licence

These notes follow the iterative modeling and assessment perspective in the SIAM/COMAP GAIMME report. Competition-specific remarks are a snapshot checked on 20 August 2026 against the official COMAP HiMCM instructions; readers should verify current rules before submission. Unless a source is explicitly identified, classroom datasets and numerical scenarios are synthetic or simulated teaching examples and are not empirical claims about a real institution. Third-party references retain their own terms.

Student response guide and checking criteria

For closed calculations, show the definition, substitution, units, and at least one independent check; the worked examples in this unit are method hints, but you must complete the arithmetic yourself. Open modeling responses are checked on five dimensions: a clear purpose and output; stated data and provenance; defensible assumptions and scope; traceable computational verification and model validation; and alignment among uncertainty, limitations, and the recommendation. Support each dimension with concrete evidence rather than writing only “completed.”

Lectures in Mathematical Modeling

Supplementary 9: Linear Algebra Essentials for Modeling

Kenneth Sok Kin Cheng

Revised 20 August 2026

Learning goals

By the end of this two-hour supplementary lesson, students should be able to:

- represent data, decisions, and states using **vectors**;
- compute and interpret **dot products**, norms, linear combinations, and span;
- read a **matrix** as a table, a linear transformation, or a system of equations;
- solve small systems by **Gaussian elimination**;
- interpret **rank**, free variables, and the solution structure of homogeneous systems;
- connect linear algebra to least squares, Markov chains, game theory, and dimensional analysis;
- explain why eigenvectors and stationary distributions matter in dynamic modeling.

1 Why linear algebra belongs in modeling

Calculus describes how one quantity changes. Linear algebra describes how many quantities combine. In mathematical modeling, this is unavoidable: a delivery route has many travel times, a regression model has many residuals, a Markov chain has many transition probabilities, and a dimensional-analysis problem has many physical exponents.

Launch activity: where are the vectors?

For each modeling situation, identify a natural vector.

1. A robot route visits five buildings and records five travel times.
2. A school canteen records arrivals in six ten-minute intervals.
3. A company tracks market share among three brands over many weeks.
4. A regression model compares predicted and observed values for ten data points.

The goal is not to make notation heavier. It is to make multi-variable information readable and computable.

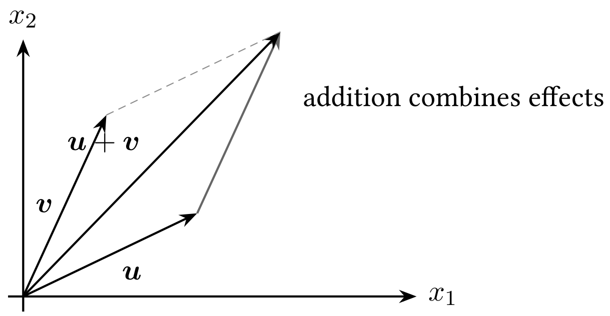


Figure 1: Vector addition models combined effects, such as displacement, resource use, or accumulated changes across components.

2 Vectors as data packages

Vectors, dot products, and norms

A vector in $\mathbb{R}^n$ is an ordered list

$$\boldsymbol{v} = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix}.$$

For $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^n$,

$$\boldsymbol{u} \cdot \boldsymbol{v} = u_1 v_1 + \cdots + u_n v_n, \qquad \|\boldsymbol{v}\| = \sqrt{\boldsymbol{v} \cdot \boldsymbol{v}}.$$

The dot product measures alignment. If $\boldsymbol{u} \cdot \boldsymbol{v} = 0$, the vectors are **orthogonal**.

Worked example 1: resource-use vector

A delivery robot uses battery, time, and staff attention. For one route, suppose

$$\boldsymbol{x} = \begin{pmatrix} 18 \\ 42 \\ 3 \end{pmatrix}$$

means 18 battery units, 42 minutes, and 3 staff interventions. A cost vector

$$\boldsymbol{c} = \begin{pmatrix} 0.4 \\ 1.2 \\ 8 \end{pmatrix}$$

assigns weights to each component. The total weighted cost is

$$\boldsymbol{c} \cdot \boldsymbol{x} = 0.4(18) + 1.2(42) + 8(3) = 81.6.$$

This is the same idea as an LP objective $\boldsymbol{c} \cdot \boldsymbol{x}$ in Lecture 6.

Linear combination and span

A **linear combination** of $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k$ is

$$c_1 \boldsymbol{v}_1 + \cdots + c_k \boldsymbol{v}_k.$$

The set of all such combinations is the **span**, written

$$\mathrm{span}(\boldsymbol{v}_1, \dots, \boldsymbol{v}_k).$$

Vectors are **linearly independent** if the only way to make $\mathbf{0}$ is to use all zero coefficients.

Worked example 2: detecting dependence

Let

$$\boldsymbol{v}_1 = \begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix}, \quad \boldsymbol{v}_2 = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}, \quad \boldsymbol{v}_3 = \begin{pmatrix} 1 \\ 4 \\ 2 \end{pmatrix}.$$

Since

$$\boldsymbol{v}_1 + 2\boldsymbol{v}_2 = \begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix} + 2\begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 4 \\ 2 \end{pmatrix} = \boldsymbol{v}_3,$$

the three vectors are linearly dependent. In modeling terms, the third feature adds no new direction; it is already explained by the first two.

Checkpoint: Checkpoint

In $\mathbb{R}^3$, can four vectors be linearly independent? Explain without computation.

3 Matrices as structured models

Matrix multiplication

An $m \times n$ matrix $\mathbf{A}$ maps vectors in $\mathbb{R}^n$ to vectors in $\mathbb{R}^m$ by

$$\boldsymbol{y} = \mathbf{A}\boldsymbol{x}.$$

If $\mathbf{A} \in \mathbb{R}^{m\times n}$ and $\mathbf{B} \in \mathbb{R}^{n\times p}$, then

$$(\mathbf{AB})_{ij} = \sum_{k=1}^{n} a_{ik} b_{kj}.$$

The product $\mathbf{AB}$ represents composition: do $\mathbf{B}$ first, then $\mathbf{A}$.

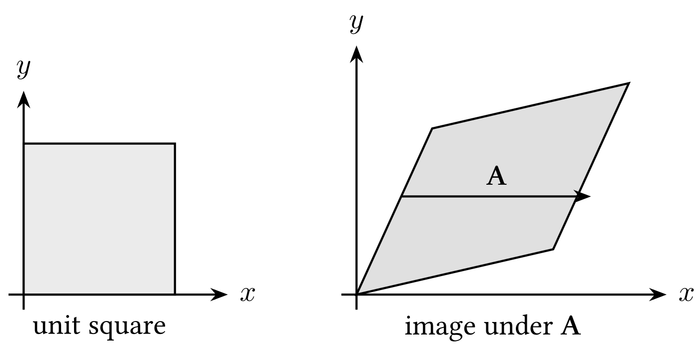


Figure 2: A matrix can be viewed as a transformation. The columns of $\mathbf{A}$ show where the unit basis vectors move.

Worked example 3: matrix product as a transition

Suppose a two-state weather model uses the transition matrix

$$\mathbf{P} = \begin{pmatrix} 0.8 & 0.2 \\ 0.4 & 0.6 \end{pmatrix},$$

where rows represent today's state and columns represent tomorrow's state: sunny S or rainy R. If today is certainly sunny, the row vector is

$$\boldsymbol{x}_0^\top = (1, 0).$$

After one day,

$$\boldsymbol{x}_1^\top = \boldsymbol{x}_0^\top \mathbf{P} = (0.8, 0.2).$$

After two days,

$$\boldsymbol{x}_2^\top = \boldsymbol{x}_0^\top \mathbf{P}^2 = (0.72, 0.28).$$

Matrix powers describe repeated transitions, which is the core of Markov-chain modeling.

Model criticism: Model critique: what does a matrix assume?

A transition matrix assumes that the next state depends on the current state in a fixed way. For weather, customer movement, or disease progression, ask:

- Are transition probabilities stable over time?
- Does tomorrow depend only on today, or also on the previous week?
- Are all relevant states included?

A matrix is compact, but compact notation can hide strong assumptions.

4 Solving systems by Gaussian elimination

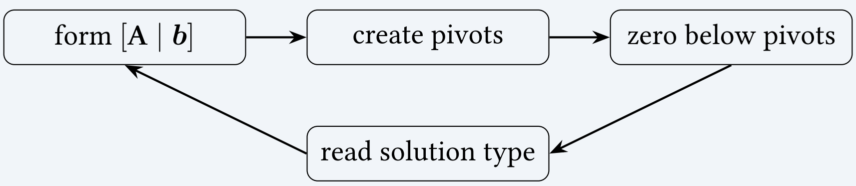
Linear system

A system of m linear equations in n unknowns can be written as

$$\mathbf{A}\boldsymbol{x} = \boldsymbol{b}.$$

Gaussian elimination uses row operations on the augmented matrix $[\mathbf{A} \mid \boldsymbol{b}]$ to reveal pivots, free variables, and consistency.

Gaussian elimination route



Worked example 4: a 3×3 system

Solve

$$\begin{cases} x + 2y + z = 6, \\ 2x + 3y - z = 4, \\ -x + y + 4z = 9. \end{cases}$$

The augmented matrix reduces as follows:

$$\left[\begin{array}{ccc|c} 1 & 2 & 1 & 6 \\ 2 & 3 & -1 & 4 \\ -1 & 1 & 4 & 9 \end{array}\right] \xrightarrow{R_2 - 2R_1,\ R_3 + R_1} \left[\begin{array}{ccc|c} 1 & 2 & 1 & 6 \\ 0 & -1 & -3 & -8 \\ 0 & 3 & 5 & 15 \end{array}\right]$$

$$\xrightarrow{R_3 + 3R_2} \left[\begin{array}{ccc|c} 1 & 2 & 1 & 6 \\ 0 & -1 & -3 & -8 \\ 0 & 0 & -4 & -9 \end{array}\right].$$

Back substitution gives $z = 9/4$, $y = 5/4$, and $x = 5/4$. Therefore

$$(x, y, z) = \left(\frac{5}{4}, \frac{5}{4}, \frac{9}{4}\right).$$

Rank and solution type

The **rank** of a matrix is the number of pivots in row echelon form. A system has:

- no solution if row reduction gives $[0\ 0\ \cdots\ 0 \mid c]$ with $c \neq 0$;
- exactly one solution if every variable column has a pivot;
- infinitely many solutions if the system is consistent but has free variables.

For $\mathbf{A} \in \mathbb{R}^{m \times n}$,

$$\dim(\text{null space}) = n - \text{rank}(\mathbf{A}).$$

Worked example 5: null space and dimensionless products

For a pendulum model, suppose the period T may depend on length L, mass m, and gravitational acceleration g. A dimensionless product has the form

$$T^a L^b m^c g^d.$$

Using dimensions $[T] = (0, 0, 1)$, $[L] = (0, 1, 0)$, $[m] = (1, 0, 0)$, and $[g] = (0, 1, -2)$ in rows M, L, T, the dimension equations are

$$\begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & -2 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \\ d \end{pmatrix} = \mathbf{0}.$$

The equations are $c = 0$, $b + d = 0$, and $a - 2d = 0$. Taking $d = 1$ gives

$$(a, b, c, d) = (2, -1, 0, 1),$$

so one dimensionless product is

$$\Pi = \frac{T^2 g}{L}.$$

This is exactly the null-space calculation behind Buckingham's Π theorem.

5 Least squares as linear algebra

Residual vector and normal equations

For data (x_i, y_i) and a line $y = a + bx$, define

$$\mathbf{X} = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_m \end{pmatrix}, \quad \boldsymbol{\beta} = \begin{pmatrix} a \\ b \end{pmatrix}, \quad \boldsymbol{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix}.$$

Predictions are $\hat{\boldsymbol{y}} = \mathbf{X}\boldsymbol{\beta}$ and residuals are

$$\boldsymbol{r} = \boldsymbol{y} - \mathbf{X}\boldsymbol{\beta}.$$

Least squares chooses $\boldsymbol{\beta}$ to minimise $\|\boldsymbol{r}\|^2$. The normal equations are

$$\mathbf{X}^\top \mathbf{X}\boldsymbol{\beta} = \mathbf{X}^\top \boldsymbol{y}.$$

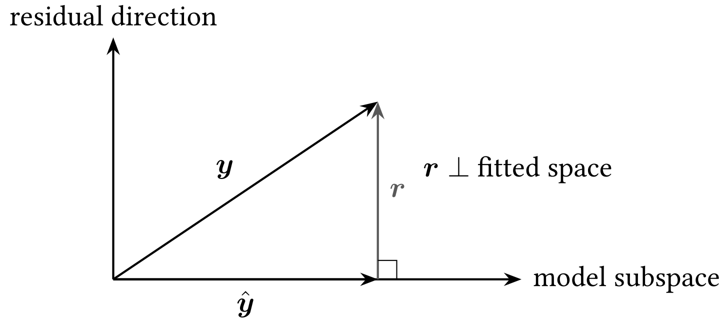


Figure 3: Least squares chooses the fitted vector $\hat{\boldsymbol{y}}$ so that the residual vector is orthogonal to the model subspace. This gives $\mathbf{X}^\top \boldsymbol{r} = \mathbf{0}$.

Worked example 6: fitting a line

Fit $y = a + bx$ to the data

$$(0, 1),\ (1, 2),\ (2, 2),\ (3, 4).$$

Then

$$\mathbf{X} = \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{pmatrix}, \qquad \boldsymbol{y} = \begin{pmatrix} 1 \\ 2 \\ 2 \\ 4 \end{pmatrix}.$$

Compute

$$\mathbf{X}^\top \mathbf{X} = \begin{pmatrix} 4 & 6 \\ 6 & 14 \end{pmatrix}, \qquad \mathbf{X}^\top \boldsymbol{y} = \begin{pmatrix} 9 \\ 18 \end{pmatrix}.$$

Solve

$$\begin{pmatrix} 4 & 6 \\ 6 & 14 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} 9 \\ 18 \end{pmatrix}.$$

The solution is $a = 0.9$, $b = 0.9$. The fitted model is

$$\hat{y} = 0.9 + 0.9x.$$

Model criticism: Model critique: matrix calculation is not the whole model

The normal equations compute the best line under a squared-error criterion. A modeling report still needs to discuss residual patterns, units, outliers, whether the relationship should be linear, and whether the prediction range is appropriate.

6 Eigenvalues, stationary states, and long-run behavior

Eigenvalue and eigenvector

For a square matrix $\mathbf{A}$, a non-zero vector $\boldsymbol{v}$ is an **eigenvector** with **eigenvalue** λ if

$$\mathbf{A}\boldsymbol{v} = \lambda\boldsymbol{v}.$$

The direction of $\boldsymbol{v}$ is preserved by the transformation; only its scale changes.

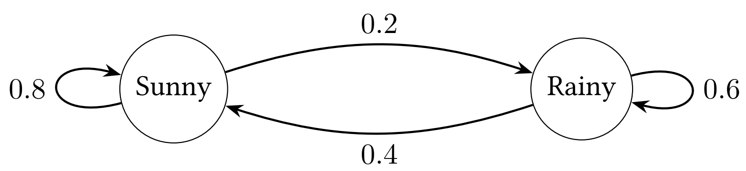


Figure 4: A two-state Markov chain. The stationary distribution is a left eigenvector of the transition matrix with eigenvalue 1.

Worked example 7: stationary distribution

For

$$\mathbf{P} = \begin{pmatrix} 0.8 & 0.2 \\ 0.4 & 0.6 \end{pmatrix},$$

a stationary distribution $\boldsymbol{\pi} = (\pi_1, \pi_2)$ satisfies

$$\boldsymbol{\pi}\mathbf{P} = \boldsymbol{\pi}, \qquad \pi_1 + \pi_2 = 1.$$

The first equation gives

$$0.8\pi_1 + 0.4\pi_2 = \pi_1 \quad \Rightarrow \quad 0.4\pi_2 = 0.2\pi_1 \quad \Rightarrow \quad \pi_1 = 2\pi_2.$$

Together with $\pi_1 + \pi_2 = 1$, we get

$$\boldsymbol{\pi} = \left(\frac{2}{3}, \frac{1}{3}\right).$$

In the long run, about two thirds of days are sunny in this model.

Worked example 8: population matrix

Let J_n and A_n be juvenile and adult populations in year n. Suppose each adult produces 0.6 juveniles, 30% of juveniles mature, and 80% of adults survive. Then

$$\begin{pmatrix} J_{n+1} \\ A_{n+1} \end{pmatrix} = \underbrace{\begin{pmatrix} 0 & 0.6 \\ 0.3 & 0.8 \end{pmatrix}}_{\mathbf{M}} \begin{pmatrix} J_n \\ A_n \end{pmatrix}.$$

The dominant eigenvalue of $\mathbf{M}$ describes the long-run growth factor. Solving

$$\det(\mathbf{M} - \lambda\mathbf{I}) = \det\begin{pmatrix} -\lambda & 0.6 \\ 0.3 & 0.8 - \lambda \end{pmatrix} = \lambda^2 - 0.8\lambda - 0.18 = 0$$

gives $\lambda \approx 0.983$. The model predicts a slow long-run decline, because the dominant growth factor is slightly below 1.

7 Differentiated modeling studio

Modeling studio

Choose one route according to readiness.

Core route. Solve a 3×3 linear system by elimination and explain the solution type.

Challenge route. Fit a line to five data points using the normal equations and compute the residual vector.

Extension route. Find the stationary distribution of a three-state transition matrix and interpret the long-run state proportions.

Each group should write one short paragraph explaining what the matrix represents before doing the computation.

Competition-style communication

A concise modeling report paragraph might read:

We represented the system state as a vector and encoded transitions by a matrix. Repeated behavior was computed using matrix multiplication, while the long-run distribution was obtained by solving the stationary equation $\boldsymbol{\pi} P = \boldsymbol{\pi}$ with $\sum_i \pi_i = 1$. This interpretation assumes transition probabilities are stable and that the current state contains enough information to predict the next step.

Checkpoint: Exit ticket

Answer in three sentences:

1. What does a pivot tell you in Gaussian elimination?
2. Why is a residual vector useful in least squares?
3. What does a stationary distribution mean in a Markov-chain model?

8 Exercises

Exercise 1: Vectors

1. Find a unit vector in the direction of $(3, 4, 0)$.
2. Compute the dot product of $(1, 2, 3)$ and $(4, -1, 2)$, then find the angle between them.
3. Determine whether $(1, 0, 2)$, $(3, 1, 4)$, and $(2, 1, 0)$ are linearly independent.

4. In a delivery model, $\boldsymbol{x} = (8, 20, 2)$ records distance, time, and number of delays. If $\boldsymbol{c} = (1.5, 0.8, 10)$, interpret $\boldsymbol{c} \cdot \boldsymbol{x}$.

Vector checks

The unit vector is $(3/5, 4/5, 0)$. The dot product in Question 2 is 8, so $\cos\theta = 8/\sqrt{294}$ and $\theta \approx 62.19°$. The determinant formed from the three vectors in Question 3 is -2, so they are independent. The weighted delivery cost is 48 in the cost scale defined by $\boldsymbol{c}$.

Exercise 2: Matrices and systems

1. Compute **AB** and **BA** for

$$\mathbf{A} = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}, \qquad \mathbf{B} = \begin{pmatrix} 5 & 0 \\ 1 & 2 \end{pmatrix}.$$

2. Solve by Gaussian elimination:

$$\begin{cases} 2x + y - z = 3, \\ x - y + 2z = 1, \\ 3x + 2y + z = 10. \end{cases}$$

3. Explain why the system

$$x + 2y + 3z = 1, \qquad 2x + 4y + 6z = 3$$

has no solution.

4. Find all solutions of

$$\begin{cases} x_1 + x_2 - x_3 + 2x_4 = 0, \\ 2x_1 - x_2 + x_3 + x_4 = 0. \end{cases}$$

Matrix and system checks

$$\mathbf{AB} = \begin{pmatrix} 7 & 4 \\ 19 & 8 \end{pmatrix}, \qquad \mathbf{BA} = \begin{pmatrix} 5 & 10 \\ 7 & 10 \end{pmatrix}.$$

Question 2 has $(x, y, z) = (4/5, 3, 8/5)$. In Question 3 the left side of the second equation is twice the first, but $3 \neq 2(1)$, so the system is inconsistent. For Question 4, with free parameters $s, t \in \mathbb{R}$,

$$(x_1, x_2, x_3, x_4) = s(0, 1, 1, 0) + t(-1, -1, 0, 1).$$

Exercise 3: Modeling connections

1. Fit $y = a + bx$ to $(1, 2), (2, 3), (3, 5), (4, 4), (5, 6)$ using $\mathbf{X}^\top \mathbf{X} \boldsymbol{\beta} = \mathbf{X}^\top \boldsymbol{y}$.

2. For

$$P = \begin{pmatrix} 0.7 & 0.2 & 0.1 \\ 0.3 & 0.5 & 0.2 \\ 0.2 & 0.3 & 0.5 \end{pmatrix},$$

find the stationary distribution.

3. A zero-sum game has payoff matrix

$$\mathbf{A} = \begin{pmatrix} 3 & -1 \\ 0 & 2 \end{pmatrix}.$$

Find Row's mixed strategy by making Column indifferent.

4. In dimensional analysis, explain why the number of independent dimensionless groups equals "number of variables minus rank of the dimension matrix."

Modeling checks

The fitted line is $\hat{y} = 1.3 + 0.9x$. The stationary distribution is $(19/41, 13/41, 9/41)$. Row's mixed strategy is $(1/3, 2/3)$, obtained by equating the two column payoffs. For dimensional analysis, connect the number of free exponent directions to the nullity of the dimension matrix via rank–nullity.

Exercise 4: Extension

1. Find eigenvalues and eigenvectors of $\begin{pmatrix} 4 & 1 \\ 2 & 3 \end{pmatrix}$.
2. Let $\mathbf{R}(\theta) = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$. Explain geometrically why $\mathbf{R}(\alpha)\mathbf{R}(\beta) = \mathbf{R}(\alpha+\beta)$.
3. Prove that if $\boldsymbol{v}$ is an eigenvector of $\mathbf{A}$ with eigenvalue λ, then $\boldsymbol{v}$ is an eigenvector of $\mathbf{A}^k$ with eigenvalue λ^k.
4. For a square matrix, list three equivalent ways to say that it is invertible.

Extension hints

The eigenpairs are $\lambda = 5$ with direction $(1,1)$ and $\lambda = 2$ with direction $(1,-2)$. For rotations, compose the geometric actions or use angle-addition identities. Question 3 follows by induction on k. Three valid invertibility equivalents are nonzero determinant, full rank, and trivial null space.

Appendix A: Algorithm summary

Task	Practical method
Solve $\mathbf{A}\boldsymbol{x} = \boldsymbol{b}$	Row-reduce $[\mathbf{A} \mid \boldsymbol{b}]$, then back-substitute.
Find a null space	Row-reduce $\mathbf{A}$, assign parameters to free variables, write basis vectors.
Test independence	Put vectors as columns; check whether every column has a pivot.
Find least-squares line	Form $\mathbf{X}$, compute $\mathbf{X}^\top\mathbf{X}$ and $\mathbf{X}^\top\boldsymbol{y}$, solve normal equations.
Find stationary distribution	Solve $\boldsymbol{\pi}P = \boldsymbol{\pi}$ plus $\sum_i \pi_i = 1$.
Find eigenvalues	Solve $\det(\mathbf{A} - \lambda\mathbf{I}) = 0$.
Find eigenvectors	For each eigenvalue, solve $(\mathbf{A} - \lambda\mathbf{I})\boldsymbol{v} = \mathbf{0}$.

Appendix B: Connection map to the main lectures

Tool	Where it appears
Dot product and norm	Lecture 3 residual vector; Lecture 6 linear objective $\boldsymbol{c} \cdot \boldsymbol{x}$.
Matrix multiplication	Lecture 5 transition matrices; Lecture 8 mixed strategies.
Gaussian elimination	Lecture 6 simplex thinking; Lecture 11 dimensionless products.
Rank and null space	Lecture 11 Buckingham Π theorem; Lecture 3 design matrix.
Least squares	Lecture 3 model fitting and residual analysis.
Eigenvalues	Lecture 5 Markov chains; Lecture 9 linear dynamical systems; Lecture 12 discrete systems.

Model audit card

Audit before trusting or communicating a model

Purpose	What question or decision does the model support?	**Data**	Which inputs are observed, derived, or simulated?
Assumptions	Which simplifications may change the answer?	**Units**	Are dimensions and scales consistent?
Verification	Were equations, code, constraints, and limiting cases checked?	**Validation**	Is performance tested against independent evidence?
Uncertainty	How do measurement, parameter, solver, and structural errors affect conclusions?	**Ethics**	Who benefits, who bears risk, and what fairness criterion matters?
Reproducibility	Are data provenance, versions, seeds, and transformations recorded?	**Communication**	Are claims conditional, traceable, and decision-relevant?

Further reading

- 18.06SC Linear Algebra (MIT OpenCourseWare)

Sources, provenance, and licence

These notes follow the iterative modeling and assessment perspective in the SIAM/COMAP GAIMME report. Competition-specific remarks are a snapshot checked on 20 August 2026 against the official COMAP HiMCM instructions; readers should verify current rules before submission. Unless a source is explicitly identified, classroom datasets and numerical scenarios are synthetic or simulated teaching examples and are not empirical claims about a real institution. Third-party references retain their own terms.

Student response guide and checking criteria

For closed calculations, show the definition, substitution, units, and at least one independent check; the worked examples in this unit are method hints, but you must complete the arithmetic yourself. Open modeling responses are checked on five dimensions: a clear purpose and output; stated data and provenance; defensible assumptions and scope; traceable computational verification and model validation; and alignment among uncertainty, limitations, and the recommendation. Support each dimension with concrete evidence rather than writing only "completed."

Lectures in Mathematical Modeling

Supplementary 10: Multivariable Calculus and Lagrange Multipliers

Kenneth Sok Kin Cheng

Revised 20 August 2026

Learning goals

By the end of this two-hour supplementary lesson, students should be able to:

- interpret a function of several variables through formulas, level curves, and modeling context;
- compute and interpret **partial derivatives** as one-variable-at-a-time sensitivities;
- use the **gradient** to identify steepest change and directional rates;
- form a local linear approximation and state the units of each sensitivity coefficient;
- locate critical points and classify them using the **Hessian matrix**;
- solve one-constraint optimization problems using **Lagrange multipliers**;
- interpret the multiplier as a local **shadow price**;
- communicate an optimization result with assumptions, verification, and limitations.

1 Why several variables change the modeling problem

A one-variable model asks how an outcome changes when one input moves. A multivariable model asks which input moved, in which direction, and under which constraints. For example, a canteen manager may choose staffing x, number of service counters y, and preparation level z. Waiting time, cost, and service quality all depend on the vector (x, y, z) rather than on one number.

Launch activity: what is being controlled?

A school event has expected net benefit

$$B(x, y) = 80x + 60y - 2x^2 - y^2,$$

where x and y are activity levels. The total staff-time constraint is $x + y = 20$.

1. Which expression is the objective? Which expression is the constraint?
2. Why can we not maximize B by examining x alone?
3. What does it mean to increase x while "holding y fixed"?

This lesson builds the tools needed to answer these questions systematically.

Functions and level curves

A function of two variables is a rule

$$f : D \subseteq \mathbb{R}^2 \to \mathbb{R}, \qquad (x, y) \mapsto f(x, y).$$

A **level curve** is the set $f(x, y) = c$. Points on the same level curve give the same output. Contour maps use many level curves to describe a surface without drawing a fragile three-dimensional picture.

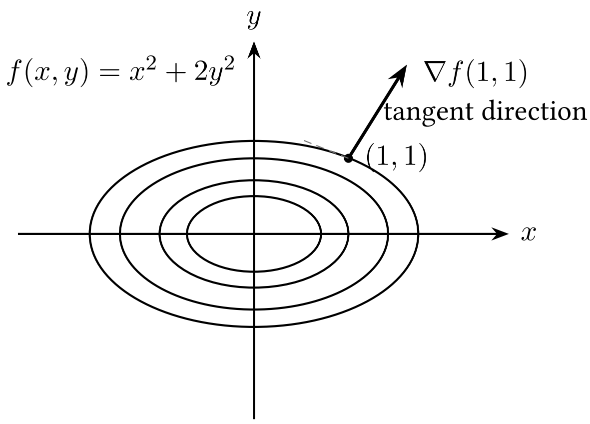


Figure 1: The gradient crosses a level curve at right angles. Moving along the curve changes direction but not the function value.

2 Partial derivatives and local sensitivity

Partial derivatives

For a function $f(x, y)$,

$$f_x(x, y) = \frac{\partial f}{\partial x}, \qquad f_y(x, y) = \frac{\partial f}{\partial y}.$$

To compute f_x, hold y fixed and differentiate with respect to x. To compute f_y, hold x fixed. In modeling language, these are **marginal effects**.

Worked example 1: interpreting units

Suppose the daily operating cost of a canteen is

$$C(x, y) = 400 + 30x + 45y + 0.8x^2 + 0.5xy,$$

where x is staff-hours and y is the number of prepared meal batches. Then

$$C_x = 30 + 1.6x + 0.5y, \qquad C_y = 45 + 0.5x.$$

At $(x, y) = (10, 8)$,

$$C_x(10, 8) = 50, \qquad C_y(10, 8) = 50.$$

Interpretation: near this operating point, one extra staff-hour adds about 50 currency units to daily cost, while one extra batch also adds about 50. The two numbers are equal here, but their units are different: cost per staff-hour versus cost per batch.

Local linear approximation

For small changes Δx and Δy,

$$f(x + \Delta x, y + \Delta y) \approx f(x, y) + f_x(x, y)\Delta x + f_y(x, y)\Delta y.$$

The change is approximated by

$$\Delta f \approx f_x \Delta x + f_y \Delta y.$$

This is the multivariable version of a tangent-line approximation and is often the first sensitivity calculation in a modeling report.

Worked example 2: a sensitivity estimate

Using the canteen cost model at $(10, 8)$, estimate the cost change if staffing rises by 0.5 hour while meal batches fall by 1:

$$\Delta C \approx C_x(0.5) + C_y(-1) = 50(0.5) - 50 = -25.$$

The model predicts a decrease of about 25 currency units. This is a local estimate; it should not be extrapolated to large changes without checking curvature.

Checkpoint: Checkpoint

If $f_x(a, b) = 0$ but $f_y(a, b) > 0$, is (a, b) necessarily a local minimum? Explain geometrically.

3 Gradient and directional change

Gradient and directional derivative

The gradient collects all first partial derivatives:

$$\nabla f(x, y) = \begin{pmatrix} f_x(x, y) \\ f_y(x, y) \end{pmatrix}.$$

For a unit vector $\boldsymbol{u}$, the **directional derivative** is

$$D_{\boldsymbol{u}} f = \nabla f \cdot \boldsymbol{u}.$$

The gradient points in the direction of steepest increase, $-\nabla f$ in the direction of steepest decrease, and $\|\nabla f\|$ is the maximum instantaneous rate of increase.

Worked example 3: moving through a temperature field

Let

$$T(x, y) = 30 - 0.5x^2 - y^2 + 2x$$

be temperature in °C, with x, y measured in kilometres. At $(1, 1)$,

$$\nabla T = (-x + 2, -2y) = (1, -2).$$

The steepest warming direction is

$$\frac{(1, -2)}{\sqrt{5}},$$

and the maximum warming rate is $\sqrt{5}$ °C per kilometre. If a student walks east, $\boldsymbol{u} = (1, 0)$, then $D_{\boldsymbol{u}} T = 1$ °C/km.

Model criticism: Model criticism: the gradient is local

A gradient is computed from the model at one point. It does not promise that the same direction remains best far away. For a long route, update the gradient along the path or solve a path-planning problem rather than following one fixed arrow.

Chain rule along a path

If a moving point follows $\boldsymbol{r}(t) = (x(t), y(t))$, then

$$\frac{d}{dt} f(x(t), y(t)) = f_x \frac{dx}{dt} + f_y \frac{dy}{dt} = \nabla f \cdot \boldsymbol{r}'(t).$$

This separates the field sensitivity ∇f from the movement $\boldsymbol{r}'(t)$.

4 Unconstrained optimization and the Hessian

Critical points

At an interior local maximum or minimum of a differentiable function,

$$\nabla f = \mathbf{0}.$$

Such points are **critical points**. As in one variable, the equation is necessary but not sufficient: a critical point may be a minimum, maximum, or saddle.

Hessian test in two variables

The Hessian is

$$\mathbf{H}f = \begin{pmatrix} f_{xx} & f_{xy} \\ f_{yx} & f_{yy} \end{pmatrix}.$$

At a critical point, let $D = f_{xx} f_{yy} - f_{xy}^2$.

Condition	Classification
$D > 0$ and $f_{xx} > 0$	local minimum
$D > 0$ and $f_{xx} < 0$	local maximum
$D < 0$	saddle point
$D = 0$	inconclusive

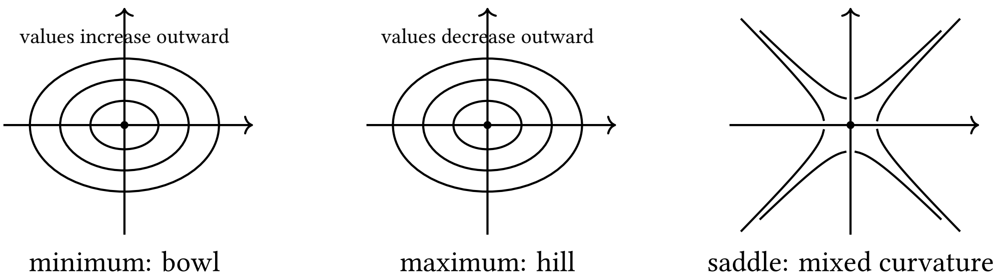


Figure 2: Contour patterns distinguish a local minimum, local maximum, and saddle. The Hessian records this local curvature algebraically.

Worked example 4: classify a critical point

For

$$f(x, y) = x^2 + 2y^2 - 2x - 8y + 5,$$

$$\nabla f = (2x - 2, 4y - 8) = \mathbf{0} \quad \Rightarrow \quad (x, y) = (1, 2).$$

The Hessian is

$$\mathbf{H}f = \begin{pmatrix} 2 & 0 \\ 0 & 4 \end{pmatrix},$$

so $D = 8 > 0$ and $f_{xx} = 2 > 0$. Therefore $(1, 2)$ is a local minimum. Completing squares shows it is also global:

$$f = (x - 1)^2 + 2(y - 2)^2 - 4.$$

Worked example 5: a saddle point

For $g(x, y) = x^2 - y^2$, $\nabla g = (2x, -2y)$, so the only critical point is $(0, 0)$. The Hessian is diag$(2, -2)$ and $D = -4 < 0$, hence the origin is a saddle: moving along the x-axis increases g, while moving along the y-axis decreases it.

5 Lagrange multipliers: optimizing on a constraint

Geometric principle

Suppose we optimize $f(x, y)$ subject to $g(x, y) = c$. At a regular constrained extremum, a level curve of f is tangent to the constraint. Their normal vectors are parallel:

$$\nabla f = \lambda \nabla g, \qquad g(x, y) = c.$$

The scalar λ is the Lagrange multiplier.

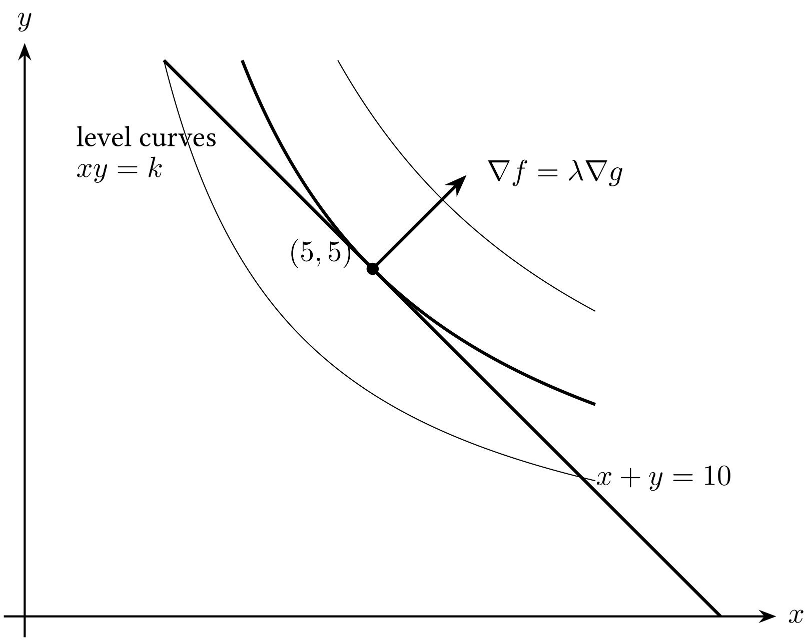


Figure 3: The maximum of $f(x, y) = xy$ subject to $x + y = 10$ occurs where the highest attainable level curve is tangent to the constraint, namely at $(5, 5)$.

Lagrange method

To optimize $f(x, y)$ subject to $g(x, y) = c$:

1. compute ∇f and ∇g;
2. solve $\nabla f = \lambda \nabla g$ together with $g = c$;
3. evaluate f at all candidates;
4. check domain boundaries, positivity restrictions, and whether the constraint is regular;
5. interpret the result in context, including units and sensitivity.

Worked example 6: maximum area with fixed perimeter

Let a rectangle have side lengths $x, y > 0$ and perimeter 20. Maximize

$$f(x, y) = xy \quad \text{subject to} \quad g(x, y) = x + y = 10.$$

Since

$$\nabla f = (y, x), \qquad \nabla g = (1, 1),$$

the Lagrange equations are

$$y = \lambda, \qquad x = \lambda, \qquad x + y = 10.$$

Hence $x = y = 5$ and the maximum area is 25. The positivity boundary gives area tending to 0, confirming the interior candidate is the constrained maximum.

Checkpoint: Checkpoint

Why is solving $\nabla f = \mathbf{0}$ inappropriate for the rectangle problem? What would it miss about the feasible set?

6 Shadow prices and a production-allocation model

Worked example 7: allocating a limited resource

A workshop chooses production levels x and y. Weekly benefit is

$$P(x, y) = 60x + 40y - 2x^2 - y^2,$$

subject to a resource limit

$$x + y = 20, \qquad x, y \geq 0.$$

Set $g(x, y) = x + y$. Then

$$\nabla P = (60 - 4x, 40 - 2y), \qquad \nabla g = (1, 1).$$

The equations

$$60 - 4x = \lambda, \qquad 40 - 2y = \lambda, \qquad x + y = 20$$

give $x = 10$, $y = 10$, and $\lambda = 20$. The optimal benefit is

$$P(10, 10) = 700.$$

Since the objective is concave (Hessian $\mathrm{diag}(-4, -2)$), this is the global constrained maximum.

Multiplier as a shadow price

Write the resource constraint as $x + y = c$ and let $P^*(c)$ be the optimal value. Near $c = 20$,

$$\frac{dP^*}{dc} = \lambda \approx 20.$$

Therefore one additional resource unit is worth about 20 benefit units locally. This does not mean every extra unit is always worth 20: the multiplier changes as the operating point changes.

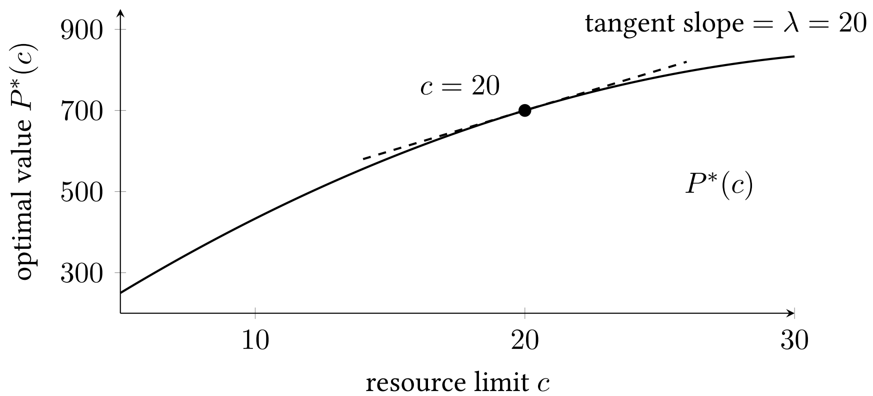


Figure 4: The shadow price is the slope of the optimal-value curve with respect to the resource limit. It is a local sensitivity, not a permanent price.

Model criticism: What the optimization has not proved

Before recommending $(x, y) = (10, 10)$, check:

- Are x and y divisible, or must they be integers?
- Is the resource relation exactly $x + y = 20$, or only $x + y \leq 20$?
- Were benefit coefficients estimated from data, and how uncertain are they?
- Does the quadratic model remain realistic near the proposed solution?

- Are there omitted constraints such as minimum service levels or staff capacity?

A mathematically optimal point can still be a poor real-world recommendation if the model is incomplete.

7 Differentiated optimization studio

Core route: gradient and Hessian

For

$$f(x, y) = 2x^2 + y^2 - 8x + 4y + 10,$$

find the critical point, classify it using the Hessian, and write f in completed-square form to verify the classification.

Challenge route: constrained geometry

Find the maximum and minimum values of

$$f(x, y) = 3x + 4y$$

on the circle $x^2 + y^2 = 25$. Explain the answer using both Lagrange multipliers and the geometry of the dot product.

Extension route: modeling and sensitivity

A farm allocates land x to crop A and y to crop B. Profit is

$$\Pi(x, y) = 100x + 80y - x^2 - 2y^2,$$

with $x + y = 50$ and $x, y \geq 0$.

1. Find the continuous optimum and multiplier.
2. Interpret the multiplier with units.
3. Test what happens when the land limit changes from 50 to 51.
4. Name one assumption that could make the mathematical optimum impractical.

8 Communicating the result

Competition-style conclusion template

A strong conclusion should state:

1. **Decision:** the recommended variable values and objective value;
2. **Feasibility:** the constraints satisfied by the recommendation;
3. **Verification:** Hessian, concavity, boundary comparison, or numerical confirmation;
4. **Sensitivity:** the meaning of the multiplier or a small perturbation test;
5. **Limitation:** at least one assumption that may affect implementation.

Example: "Under the continuous quadratic model and the binding 20-unit resource constraint, the recommended allocation is $(10, 10)$, giving benefit 700. The negative-definite Hessian confirms global concavity. The multiplier 20 indicates that one additional resource unit is worth about 20 benefit units locally. Integer restrictions and parameter uncertainty should be checked before implementation."

Checkpoint: Exit ticket

In two sentences, explain the difference between $\nabla f = \mathbf{0}$ and $\nabla f = \lambda \nabla g$. Then state one reason why a Lagrange candidate may fail to be a valid final recommendation.

9 Practice problems

Exercise 1: partial derivatives and local change

1. For $f(x, y) = x^2 y + e^{xy}$, compute f_x and f_y.
2. For $T(x, y) = 25 - 0.2x^2 - 0.5y^2$, find the gradient at $(2, 1)$ and the directional derivative toward $(5, 5)$.
3. Use local linearization to estimate $\sqrt{x^2 + y^2}$ near $(3, 4)$ when (x, y) changes to $(3.03, 3.98)$.
4. Explain the units of f_x and f_y for a cost function whose inputs are labour-hours and kilograms of material.

Partial-derivative checks

For Question 1,
$$f_x = 2xy + ye^{xy}, \qquad f_y = x^2 + xe^{xy}.$$
At $(2, 1)$, $\nabla T = (-0.8, -1)$; the unit direction toward $(5, 5)$ is $(3/5, 4/5)$, giving directional derivative -1.28. Linearization of $\sqrt{x^2 + y^2}$ about $(3, 4)$ gives 5.002. The final answer should attach cost per labour-hour and cost per kilogram to the two partial derivatives.

Exercise 2: unconstrained optimization

Find and classify all critical points:

1. $f(x, y) = x^2 + 3y^2 - 4x + 6y$;
2. $g(x, y) = x^2 - y^2 + 2x$;
3. $h(x, y) = x^3 + y^3 - 3xy$;
4. $q(x, y) = x^4 + y^4 - 4xy$.

For each, explain whether the Hessian test is conclusive.

Critical-point checks

The critical points and classifications are: $(2, -1)$, a minimum; $(-1, 0)$, a saddle; for h, $(0, 0)$ is a saddle and $(1, 1)$ is a local minimum; for q, $(0, 0)$ is a saddle and $(1, 1)$ and $(-1, -1)$ are local minima. In every listed case the Hessian determinant is nonzero at the critical point, so the test is conclusive.

Exercise 3: Lagrange multipliers

Use Lagrange multipliers and verify the result:

1. maximize xy subject to $x^2 + 4y^2 = 8$;
2. find the point on $x + 2y + 3z = 6$ closest to the origin;
3. maximize xyz subject to $x + y + z = 12$, with $x, y, z > 0$;
4. maximize $5x + 8y - x^2 - y^2$ subject to $2x + y = 12$ and $x, y \geq 0$.

Lagrange-multiplier checks

The maxima in Question 1 are $(2, 1)$ and $(-2, -1)$, with value 2. The closest point in Question 2 is $(3/7, 6/7, 9/7)$. Question 3 has $(x, y, z) = (4, 4, 4)$ and maximum 64. Question 4 has $(x, y) = (37/10, 23/5)$ and maximum $409/20$; compare with the feasible endpoints when verifying it.

Exercise 4: modeling challenge

A manufacturer chooses two continuous production levels. Profit is

$$\Pi(x, y) = 90x + 70y - 1.5x^2 - y^2 - 0.5xy,$$

subject to $2x + y \leq 60$ and $x, y \geq 0$.

1. Find the unconstrained optimum and test whether it is feasible.
2. If the resource constraint binds, solve the equality-constrained problem.
3. Compare the objective values at all relevant candidates and boundaries.
4. Interpret the multiplier and discuss whether integer production changes the recommendation.
5. Write a five-sentence modeling conclusion using the template above.

Modeling checks

The unconstrained point $(580/23, 660/23)$ violates $2x + y \leq 60$. On the binding line, the continuous optimum is

$$(x, y) = \left(\frac{160}{9}, \frac{220}{9}\right), \qquad \Pi = \frac{18200}{9} \approx 2022.22.$$

It exceeds the best axis candidates, $(0, 35)$ with value 1225 and $(30, 0)$ with value 1350. With the convention $\nabla\Pi = \lambda\nabla(2x + y)$, $\lambda = 110/9$ is the local value of one extra resource unit. For integer production, checking nearby feasible points selects $(18, 24)$ with profit 2022. A complete conclusion states the continuous assumption, boundary comparison, multiplier locality, integer check, and one validation or uncertainty limitation.

Optional extension A: several constraints

For constraints $g_1(\boldsymbol{x}) = c_1, \ldots, g_k(\boldsymbol{x}) = c_k$, the necessary condition becomes

$$\nabla f = \lambda_1 \nabla g_1 + \cdots + \lambda_k \nabla g_k.$$

The constraint gradients must be linearly independent at the candidate. Each multiplier measures local sensitivity to its corresponding constraint level, subject to regularity conditions.

Optional extension B: inequality constraints

If the feasible set contains inequalities such as $g(\boldsymbol{x}) \leq c$, the constraint may be active or inactive. When active, the solution can often be studied by setting $g = c$ and using Lagrange multipliers. A systematic treatment uses the Karush–Kuhn–Tucker conditions, which are beyond this supplementary. In a report, explicitly check whether the inequality binds rather than assuming equality from the start.

Model audit card

Audit before trusting or communicating a model

Purpose	What question or decision does the model support?	**Data**	Which inputs are observed, derived, or simulated?
Assumptions	Which simplifications may change the answer?	**Units**	Are dimensions and scales consistent?
Verification	Were equations, code, constraints, and limiting cases checked?	**Validation**	Is performance tested against independent evidence?
Uncertainty	How do measurement, parameter, solver, and structural errors affect conclusions?	**Ethics**	Who benefits, who bears risk, and what fairness criterion matters?
Reproducibility	Are data provenance, versions, seeds, and transformations recorded?	**Communication**	Are claims conditional, traceable, and decision-relevant?

Further reading

- Calculus Volume 3 (OpenStax)
- Nonlinear Constrained Optimization (NEOS Guide)

Sources, provenance, and licence

These notes follow the iterative modeling and assessment perspective in the SIAM/COMAP GAIMME report. Competition-specific remarks are a snapshot checked on 20 August 2026 against the official COMAP HiMCM instructions; readers should verify current rules before submission. Unless a source is explicitly identified, classroom datasets and numerical scenarios are synthetic or simulated teaching examples and are not empirical claims about a real institution. Third-party references retain their own terms.

Student response guide and checking criteria

For closed calculations, show the definition, substitution, units, and at least one independent check; the worked examples in this unit are method hints, but you must complete the arithmetic yourself. Open modeling responses are checked on five dimensions: a clear purpose and output; stated data and provenance; defensible assumptions and scope; traceable computational verification and model validation; and alignment among uncertainty, limitations, and the recommendation. Support each dimension with concrete evidence rather than writing only “completed.”

Writing Classes for Mathematical Modeling Competitions
W1: Paper Architecture and Judge Navigation

Kenneth Sok Kin Cheng

Revised 20 August 2026

Learning goals

By the end of this two-hour writing class, students should be able to:

- distinguish contest rules from writing folklore and verify the correct rule set before submission;
- explain the job of each major section in a modeling report;
- organize a report so that a judge can locate the answer, assumptions, evidence, and limitations quickly;
- connect problem requirements to models, results, validation, and conclusions through a **traceability map**;
- evaluate a paper using evidence-based questions rather than invented score percentages;
- rewrite weak assumptions using the pattern *what* – *why* – *effect* – *test*;
- build a complete section blueprint for a competition problem before drafting prose;
- conduct a short **judge-navigation audit** of a draft.

1 The paper is the model's interface

A team may have correct mathematics, useful code, and a sensible recommendation. None of these helps unless the report allows a reader to reconstruct the reasoning. A competition paper is therefore not a diary of what the team did. It is a designed path from the question to a defensible decision.

Launch activity: the 60-second search test

Imagine that you are judging an unfamiliar 20-page report. You have sixty seconds to locate:

1. the main recommendation;
2. the central model equation or algorithm;
3. the most important assumption;
4. one validation result;
5. one limitation.

What visual or structural features would help you find each item? List at least four: section headers, result tables, equation references, captions, bold conclusions, or a task-to-section map.

The evidence chain

A strong report makes the following chain visible:

$$\text{question} \longrightarrow \text{assumptions and data} \longrightarrow \text{model} \longrightarrow \text{results} \longrightarrow \text{checks} \longrightarrow \text{decision.}$$

Every arrow is a claim that must be explained. For example, an assumption should affect the model; a numerical result should answer a task; a validation check should change how confidently the recommendation is stated.

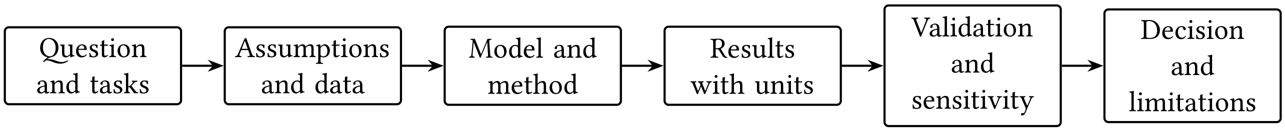


Figure 1: A modeling report is an evidence chain. A missing link is more damaging than a missing decorative section.

Model criticism: A paper is not the chronological story

Teams often write in the order they worked: failed attempt, code debugging, new idea, another failed attempt, final model. The judge needs the logical order instead. Failed approaches may be mentioned briefly when they justify a design choice, but the main narrative should present the clearest defensible route.

2 Rule snapshot: verify before writing

Formatting rules are part of the problem. They change by contest and year, so a team should create a one-page compliance sheet before writing. The table below records the official 2026 rules relevant to report design; always re-check the current official instructions before a future contest.

2026 report-format comparison

	HiMCM 2026	IM^2C International 2026
Team and work window	Up to four students; teams may work during the published 14-day contest window.	Up to four students from the same school; one consecutive five-day period.
First pages	Page 1 is the Summary Sheet. A table of contents is encouraged.	One-page Summary Sheet, one-page Table of Contents, then the solution.
Page accounting	Summary plus up to 24 further pages: 25 pages total. TOC, references, notes, and appendices count.	Solution maximum 24 pages; the one-page summary and one-page TOC are separate. References and appendices at the end do not count toward the 24-page solution limit.
Typography	English PDF; readable font of at least 12 point.	English; A4 with margins at least 1.5 cm or Letter equivalent; at least 12 point.
Identification	Control number and page number in the header; no student, advisor, or school names.	Control number and page number in the header; no student, advisor, or school names.
AI documentation	Use of LLMs or other AI tools must be identified and cited; a separate AI-use report is appended after the 25-page solution and is not counted in that limit.	Follow the current IM^2C and regional instructions; do not assume that HiMCM's AI-report format transfers automatically.

Checkpoint: Checkpoint: which statements are unsafe?

Mark each statement as safe or unsafe, then repair the unsafe ones.

1. "All modeling contests allow 25 pages."
2. "The Summary should be the first page."
3. "Appendices never count toward the page limit."
4. "A 10-point font is acceptable if the paper is long."
5. "The team should verify the current rules before creating the template."

Model criticism: Compliance layer versus writing layer

A contest rule tells you what must be submitted. A writing principle tells you how to make the reasoning easy to follow. For example, HiMCM requires a Summary Sheet, but the rule does not guarantee that a summary is informative. Compliance is necessary; communication quality still has to be designed.

3 A reader-first paper architecture

The original draft treated one section order as compulsory. That is too rigid. Good papers share recognizable functions, but adjacent functions may be merged when the logic is clearer. The test is not "Did we use exactly these headings?" The test is "Can the reader find every necessary piece of evidence?"

Recommended architecture

Layer	Typical sections	Reader's question
Front matter	Summary; Table of Contents	What did the team do and find? Where is everything?
Orientation	Problem restatement; objectives; assumptions; notation; data	What exactly is being answered, under what interpretation?
Model core	Baseline model; refined model; algorithms; derivations	Why this model, and how does it work?
Results	Numerical or analytical results; figures; decision table	What is the answer, with units and context?
Evidence	Verification; validation; uncertainty; sensitivity	Why should the reader trust the result, and when does it change?
Decision	Recommendation; strengths; limitations; conclusion	What action follows, and how cautious should we be?
Back matter	References; appendices; required disclosures	Where did information come from, and where are supporting details?

Model criticism: Useful flexibility

- Results may follow each model immediately instead of appearing in one distant "Results" section.
- Validation may be integrated beside a model when the check is essential to understanding it.
- Strengths and limitations may sit in the conclusion if they are specific and visible.
- A separate literature-review section is often unnecessary; cite the relevant source where the modeling choice is introduced.

What should not disappear is the function: assumptions, results, checks, limitations, and references must remain findable.

Card-sort activity: build the logical order

Arrange the following cards into a coherent paper route for a school-bus routing problem:

distance data | route algorithm | final routes | problem tasks | assumptions | benchmark | sensitivity to travel time | recommendation

There may be more than one defensible order. For every adjacent pair, write one sentence explaining why the second follows from the first.

Worked example 1: a coherent queue-paper skeleton

For a school canteen queue problem, a concise section plan might be:

1. **Problem interpretation.** Define the performance measures: average wait, 90th-percentile wait,

and staffing cost.

2. **Data and assumptions.** Explain how arrival and service times are estimated and when a Poisson model is plausible.
3. **Baseline capacity check.** Compare arrival rate λ and service capacity $c\mu$ before simulation.
4. **Discrete-event simulation.** State events, state variables, algorithm, warm-up, and replications.
5. **Results by staffing level.** Present one table containing wait, queue length, utilization, and cost.
6. **Validation.** Compare the one-server simulation with an analytical M/M/1 result under matching assumptions.
7. **Sensitivity.** Vary peak arrival rate and service-time distribution.
8. **Recommendation.** Choose the smallest staffing level meeting the service target; state the limitation during exceptional peaks.

The structure mirrors the evidence chain rather than the team's work diary.

4 Rubric decoded as an evidence map

Official HiMCM descriptions emphasize modeling and problem solving, analysis, conclusions, communication, clarity, support, and organization. IM^{2}C guidance asks for variables, assumptions, justified model design, testing, sensitivity, and strengths or weaknesses. These sources do not publish the percentage weights used in the original draft. Therefore, use the following as a *teaching lens*, not as an official scoring formula.

Five questions a strong report answers

1. **Task coverage**: Did the team answer every requirement in the problem statement?
2. **Model reasoning:** Are assumptions, data choices, and methods appropriate and justified?
3. **Mathematical execution:** Are equations, algorithms, calculations, and interpretations correct?
4. **Evidence and robustness:** Is the model checked against data, theory, benchmarks, limiting cases, or alternative assumptions?
5. **Communication and compliance:** Can a reader follow the argument, locate results, verify sources, and confirm that the rules were followed?

Evidence map

Question	Strong visible evidence	Weak signal
Task coverage	Task list, matching subsections, final answer table	One task is answered only implicitly or not at all
Model reasoning	Assumption rationale; baseline-to-refinement logic; method choice	A fashionable method appears with no comparison or need
Execution	Defined symbols; derived equations; reproducible algorithm; units	Final formula appears without explanation or dimensions
Evidence	Benchmark, data comparison, residuals, sensitivity table, uncertainty	"The model is accurate and robust" with no numerical check
Communication	Informative captions, result callouts, cross-references, consistent notation	Walls of text, hidden answer, unlabeled figures, notation drift
Compliance	Correct format, anonymity, citations, disclosure, file checks	Team names in paper, wrong page count, missing source documentation

Evidence hunt

Read a two-page excerpt from any modeling paper. For each of the five questions above, highlight one sentence, equation, figure, or table that supplies evidence. If you cannot find evidence, write the exact addition that would repair the gap.

Checkpoint: Checkpoint

A paper uses an advanced genetic algorithm and reports a very low objective value. Which evidence questions remain unanswered? Name at least three specific checks the paper still needs.

5 Assumptions are design decisions

An assumption is not a ceremonial sentence. It changes the feasible model, the interpretation of parameters, and the strength of the recommendation.

The four-part assumption pattern

For each important assumption, state:

1. **What:** the precise simplification or claim;
2. **Why:** evidence, domain reasoning, or a clearly stated practical necessity;
3. **Effect:** what becomes possible, and what distortion may be introduced;
4. **Test:** how the assumption will be checked or varied later.

Not every minor assumption needs four separate paragraphs, but every high-impact assumption should make these four ideas recoverable.

Worked example 2: weak and improved assumption

Weak: "We assume customer arrivals are Poisson."

Improved: "During the 20-minute peak interval, we model customer arrivals as a Poisson process with constant rate λ. This is plausible when arrivals are approximately independent and the observed counts show no strong trend within the interval. The assumption permits exponential inter-arrival sampling and comparison with queueing formulas. Because student arrivals may occur in groups after a bell, we also rerun the simulation using burst arrivals and compare the staffing recommendation."

Rewrite studio

Rewrite two of the following assumptions using *what – why – effect – test*:

1. "Traffic speed is constant."
2. "The population does not migrate."
3. "All customers behave the same."
4. "The data are accurate."

Underline the sentence that admits a possible weakness. A strong assumption section does not pretend that simplification has no cost.

Model criticism: Do not separate assumptions from their consequences

A long list at the beginning is easy to forget. Refer back to assumptions when they enter an equation: "Under Assumption A2, the arrival rate is constant, so…" Later, sensitivity analysis should identify which assumptions were actually tested. This creates traceability instead of a decorative list.

6 Results, validation, and sensitivity must be visible

A result is more than a number. A useful result unit contains a question, a metric, a value with units, a comparison, and a decision implication.

Answer-first result unit

Use the pattern:

$$\text{task} \longrightarrow \text{metric} \longrightarrow \text{result} \longrightarrow \text{comparison} \longrightarrow \text{decision.}$$

Example: "To keep the 90th-percentile wait below 6 minutes, three counters are required. The simulation estimate is 5.4 minutes (95% interval 5.1–5.7), compared with 8.9 minutes for two counters. We therefore recommend three counters during the peak period."

Worked example 3: compact evidence table

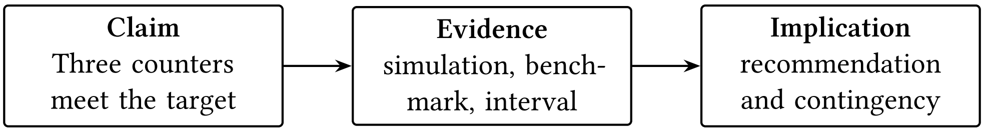
Scenario	Counters	Mean wait (min)	90th percentile	Utilization
Baseline	2	4.8	8.9	0.91
Recommended	3	2.1	5.4	0.64
Arrival rate $+20\%$	3	3.6	7.2	0.77
Slower service -10%	3	3.0	6.4	0.71

The table does not merely show output. It reveals that the recommendation misses the target when demand rises by 20%, so the conclusion should include a peak-demand contingency.

Validation and sensitivity are different

- **Verification**: Did we solve or code the model correctly? Examples: hand calculation, conservation check, unit test, convergence test.
- **Validation**: Does the model represent the real system well enough? Examples: held-out data, external benchmark, expert expectation, observed trend.
- **Sensitivity analysis**: How does the conclusion change when inputs or assumptions change?
- **Uncertainty analysis**: How uncertain is the reported output because inputs or simulation are uncertain?

A paper may use different section names, but it should not confuse these jobs.



A caption or paragraph should explain the relation among the three boxes; the reader should not have to infer it from a figure alone.

Figure 2: A result becomes persuasive when the claim, supporting evidence, and decision implication are linked explicitly.

Checkpoint: Caption test

Which caption is stronger, and why?

1. "Figure 4. Sensitivity analysis."
2. "Figure 4. Increasing peak arrival rate by 20% raises the 90th-percentile wait above the 6-minute target even with three counters; the recommendation therefore requires an overflow plan."

7 Traceability and judge navigation

A report should make it possible to trace every problem requirement to a result and every recommendation to evidence.

Task-to-evidence matrix

Before drafting, construct a table like this:

Problem task	Model / method	Result location	Check / limitation
Estimate current performance	Queue simulation	Table 3, Section 5.1	Compare with observed sample
Choose staffing level	Cost-service trade-off	Figure 6, Section 5.2	Arrival-rate sensitivity
Plan exceptional peaks	Scenario model	Table 5, Section 6	Peak timing is uncertain

If one row has an empty cell, the paper is not ready.

Judge-navigation audit

Exchange blueprints or drafts. The reviewer has 90 seconds and may not ask the author questions. Record the page or section where you found:

1. each problem task;
2. the primary model;
3. the main numerical answer;
4. the strongest validation check;
5. the most decision-changing sensitivity result;
6. the final recommendation and limitation.

Anything not found within 90 seconds needs a structural repair, not merely a grammar edit.

Model criticism: Navigation is not decoration

A table of contents, numbered equations, descriptive headings, captions, and cross-references reduce the reader's search cost. Decorative colour, complex diagrams, and excessive boxes do not automatically improve navigation. Use visual hierarchy only when it reveals the logical hierarchy.

8 Integrated blueprint workshop

Workshop: design the paper before solving it

A school wants to reduce congestion at one entrance between 7:30 and 8:00. Students arrive by foot, car, and bus. The school may change lane allocation, staff deployment, and arrival guidance. The goal is to reduce delay without creating an unsafe queue outside the gate.

In teams, produce a one-page report blueprint containing:

1. a three-sentence problem interpretation with measurable outputs;
2. a task list and task-to-evidence matrix;
3. three high-impact assumptions in *what – why – effect – test* form;
4. a baseline model and one justified refinement;
5. a proposed results table with column headings and units;
6. one verification check, one validation check, and two sensitivity scenarios;
7. a section outline with one-sentence purpose statements;
8. a final recommendation sentence template containing a condition and limitation.

Do not calculate the final answer. The product is the architecture of a defensible paper.

Differentiated routes

- **Core route**: use a deterministic arrival-flow model and a simple capacity comparison.
- **Challenge route**: add time-varying arrivals and a discrete-event simulation.

- **Extension route:** define a multi-objective score balancing delay, cost, and safety; explain how weights would be justified and tested.

All routes must preserve the same evidence chain and task traceability.

9 What to carry into W2

Class W1 takeaways

1. Rules are contest- and year-specific. Build the template only after checking the official instructions.
2. A standard architecture is useful, but section names and placement should serve the logical evidence chain.
3. Do not treat unofficial percentage weights as an official rubric. Evaluate visible evidence instead.
4. Every major assumption should reveal what is assumed, why, its effect, and how it will be tested.
5. Results must be easy to locate and must connect a number to a comparison and a decision.
6. Verification, validation, sensitivity, and uncertainty answer different questions.
7. A task-to-evidence matrix and a 90-second navigation audit expose structural gaps early.

Checkpoint: Exit ticket

In no more than four sentences:

1. name one structural change you will make to your next report;
2. identify one assumption you would test rather than merely state;
3. explain the difference between validation and sensitivity;
4. state what W2 must teach you about the one-page Summary.

In **W2: Writing the Summary**, we will construct the first page through four functions: problem, method, quantitative results, and decision insight. Students will diagnose weak summaries and draft progressively shorter versions without losing evidence.

Appendix A: Paper-architecture checklist

Audit question	Evidence to locate
Does the paper answer every task?	Task list, matching result sections, task-to-evidence matrix
Can the reader understand the model?	Defined symbols, assumptions, derivation or algorithm, units
Can the reader find the answer?	Summary, highlighted result table, explicit conclusion
Was the implementation checked?	Hand case, unit test, conservation or convergence check
Was the representation of reality checked?	Data comparison, benchmark, held-out case, expected trend
Is the recommendation robust?	Parameter and structural sensitivity; uncertainty interval
Are limitations decision-relevant?	Conditions under which the recommendation may fail
Is the document compliant?	Current page limit, font, anonymity, headers, citations, disclosure
Can a judge navigate it quickly?	Descriptive headings, captions, cross-references, consistent notation

Appendix B: Independent paper-reading task

Choose one recent Outstanding HiMCM paper or published IM^{2}C solution. Do not begin by judging the mathematics. First produce:

1. its section map;
2. a task-to-evidence matrix;
3. the first location of each main numerical answer;
4. one example of a strong assumption, caption, validation check, and limitation;
5. one structural decision you would copy and one you would change.

Then compare the report against the five evidence questions in this class.

Appendix C: Official sources checked for this lesson

- COMAP, 2026 HiMCM/MidMCM Contest Instructions and Rules.
- COMAP, HiMCM/MidMCM Tips Article.
- IM^{2}C challenge overview and modeling expectations.

Rules may change. Verify the current official documents immediately before a contest.

Model audit card

Audit before trusting or communicating a model

Purpose	What question or decision does the model support?	**Data**	Which inputs are observed, derived, or simulated?
Assumptions	Which simplifications may change the answer?	**Units**	Are dimensions and scales consistent?
Verification	Were equations, code, constraints, and limiting cases checked?	**Validation**	Is performance tested against independent evidence?
Uncertainty	How do measurement, parameter, solver, and structural errors affect conclusions?	**Ethics**	Who benefits, who bears risk, and what fairness criterion matters?
Reproducibility	Are data provenance, versions, seeds, and transformations recorded?	**Communication**	Are claims conditional, traceable, and decision-relevant?

Further reading

- HiMCM / MidMCM Contest Instructions (COMAP)
- IM²C challenge overview and modeling expectations (International Mathematical Modeling Challenge)
- GAIMME: Guidelines for Assessment and Instruction in Mathematical Modeling Education (SIAM / COMAP)

Sources, provenance, and licence

These notes follow the iterative modeling and assessment perspective in the SIAM/COMAP GAIMME report. Competition-specific remarks are a snapshot checked on 20 August 2026 against the official COMAP HiMCM instructions; readers should verify current rules before submission. Unless a source is explicitly identified, classroom datasets and numerical scenarios are synthetic or simulated teaching examples and are not empirical claims about a real institution. Third-party references retain their own terms.

Student response guide and checking criteria

For closed calculations, show the definition, substitution, units, and at least one independent check; the worked examples in this unit are method hints, but you must complete the arithmetic yourself. Open modeling responses are checked on five dimensions: a clear purpose and output; stated data and provenance; defensible assumptions and scope; traceable computational verification and model validation; and alignment among uncertainty, limitations, and the recommendation. Support each dimension with concrete evidence rather than writing only “completed.”

Writing Classes for Mathematical Modeling Competitions

W2: The Summary – Claim, Evidence, and Decision

Kenneth Sok Kin Cheng

Revised 20 August 2026

Learning goals

By the end of this two-hour writing class, students should be able to:

- distinguish a competition **Summary Sheet** from an introduction, abstract, methods list, or team diary;
- build an **evidence inventory** before drafting prose;
- map every major task in the problem statement to a model, result, and decision;
- organize a summary through four functional blocks: task, strategy, findings, and trust/decision;
- select a small set of decision-relevant numbers and report them with units, baselines, and conditions;
- distinguish verification, validation, sensitivity, and uncertainty in one-sentence summary claims;
- write conditional recommendations without weakening the paper;
- compress a draft while preserving task coverage and evidence;
- conduct a rapid peer audit using answer visibility, evidence, consistency, and precision.

1 What a competition summary must accomplish

The Summary is not a miniature version of every section. It is a decision brief: a reader should be able to identify the task, the modeling route, the main findings, and the confidence or limitation attached to the recommendation without searching the full report.

Official guidance supports this emphasis. The current HiMCM instructions require the solution to begin with the Summary Sheet and ask teams to present major ideas and results clearly. The current IM^{2}C USA instructions state that judges place considerable weight on the summary and that the most important conclusions should be prominent. Neither source supports invented claims such as "judges spend exactly seven minutes" or "the summary determines 60% of the score." Use official expectations, not folklore.

Launch activity: the 20-second answer test

Read a summary page for twenty seconds, then close it. Without looking back, answer:

1. What decision or quantity was the team asked to produce?
2. What was the team's main recommendation or numerical answer?
3. What model or method generated it?
4. What evidence makes the recommendation credible?

If a reader cannot recover at least three answers, the summary is not yet functioning as a summary.

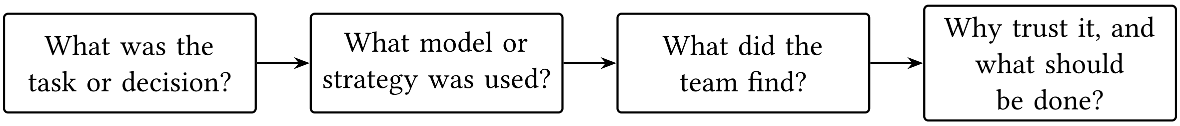


Figure 1: The summary should answer four reader questions. The wording and order may vary, but the functions should remain visible.

Model criticism: Summary, introduction, and abstract are not identical

- An **introduction** develops context, problem interpretation, and report organization. It may postpone numerical answers.
- A research **abstract** often follows disciplinary conventions and may emphasize objective, method, result, and conclusion in compressed form.
- A competition **Summary Sheet** is closer to an executive decision brief: it should state the answer explicitly, identify the modeling architecture, and indicate why the answer is defensible.

Copying the introduction into the Summary usually produces too much background and too few results.

2 Do not begin with prose: build the evidence inventory

A weak summary is often not a sentence-level problem. It is an evidence-selection problem. Teams begin writing before deciding which tasks, results, comparisons, and limitations must survive the one-page constraint.

The five-column evidence inventory

Before drafting, complete one row for every important task:

Task	Headline answer	Method	Evidence/check	Condition/caveat
What must be answered?	Number, ranking, route, threshold, or recommendation	Model or algorithm that produced it	Baseline, validation, uncertainty, or sensitivity	When might the answer change?

A row with no headline answer signals unfinished analysis. A row with no evidence signals an unsupported claim. A row with no condition may signal overconfidence.

Worked example 1: evidence inventory for an illustrative school-bus case

A school must redesign routes for 226 students using buses of capacity 60. The target is to reduce ride time while keeping the longest scheduled ride below 45 minutes.

Task	Answer	Method	Evidence/check	Condition
Choose fleet size	Four buses	Capacity lower bound plus routing model	$226/60 = 3.77$, so at least four are required	Assumes all registered riders travel that day
Design routes	Four routes, 87.4 km total	Cluster-first vehicle-routing heuristic with local search	Existing routes total 103.6 km	Road closures excluded
Reduce ride time	Mean 28.2 min; maximum 44.1 min	Time-dependent shortest paths	Existing mean 42.9 min; all planned routes below 45 min	Uses morning travel-time estimates
Test demand changes	Four buses sufficient in 18 of 20 tested demand scenarios	Reassignment under ± 10 students per district	Maximum ride remains below 45 min in 17 scenarios	A fifth bus is needed in two high-demand scenarios

The Summary does not need every number in this table. It needs the numbers that answer the task, establish improvement, and reveal the decision boundary.

Checkpoint: Checkpoint: what is missing?

A team writes: "Our optimized route length is 87.4 km." Identify at least three missing pieces from the evidence inventory. Possible questions include: compared with what, for how many buses, under which capacity and time constraints, and how stable is the result?

Model criticism: Task coverage comes before elegance

A beautifully written summary that omits one required task is incomplete. Copy the task verbs from the problem statement – estimate, design, compare, recommend, test – and check that each appears either directly or through an unmistakable result sentence.

3 Four functional blocks – with useful flexibility

The original draft's four-block idea is useful, but it should be treated as a design tool rather than a rigid paragraph law.

Recommended four-block architecture

Block	Main job	Must contain
1. Task	Orient the reader	Decision context, required outputs, important constraints
2. Strategy	Explain the modeling architecture	Baseline/refinement sequence, named methods, data role
3. Findings	Deliver the answer	Headline result first, units, comparison, task coverage
4. Trust and decision	Calibrate confidence and action	Validation/sensitivity evidence, threshold or caveat, recommendation

A typical one-page summary may devote roughly 2–3 sentences to Task, 3–5 to Strategy, 4–6 to

Findings, and 2–4 to Trust/Decision. These are drafting ranges, not rules.

Task: name the decision, outputs, and constraints – not the contest history.

Strategy: show how the models connect – not a list of software and equations.

Findings: lead with the answer, then the improvement, threshold, or uncertainty.

Trust/Decision: state the check, the tested range, and the action that follows.

Figure 2: A stable drafting architecture. A results-first opening is also possible when the answer is urgent and immediately interpretable.

Model criticism: When a results-first opening works

A summary may open with the answer when the problem is strongly decision-oriented:

> *We recommend four buses on ordinary school days, an early fleet review at 236 projected riders, and a fifth contingency bus from 241 confirmed riders.*

The next sentence must then identify the task and conditions. Results-first is not permission to remove context; it changes the order of evidence.

4 Writing the Task and Strategy blocks

4.1 Task block: compress the question, not the background

A strong Task block should make the required deliverable visible. It does not need a history lesson unless the background changes the modeling choice.

Worked example 2: weak and improved Task blocks

Weak

> *School buses are important in many places, and many students use them every day. This paper studies school bus routing, which is a complicated problem.*

Improved

> *The school must transport 226 students with buses of capacity 60 while reducing ride time and keeping every scheduled journey below 45 minutes. We therefore seek the smallest feasible fleet, a set of capacity-respecting routes, and a contingency rule for uncertain daily demand.*

The improved version names the decision variables, constraints, and outputs in two sentences.

Task-block sentence frame

[Decision maker] must [decision] for [system/population] under [main constraints]. We estimate/design/compare [required outputs], with [performance metric] as the primary criterion.

4.2 Strategy block: architecture before technical detail

The Strategy block should explain why multiple methods exist and how they hand off to one another. A software list is not a modeling strategy.

Worked example 3: method list versus model architecture

Weak

> *We used Python, clustering, Dijkstra's algorithm, optimization, simulation, and sensitivity analysis.*

Improved

> *We first derive a fleet-size lower bound from capacity, then construct feasible routes by clustering nearby stops and solving shortest paths on a time-weighted road graph. A local-search stage exchanges stops between routes to reduce the longest ride rather than total distance alone. Finally, we rerun the assignment under perturbed district demand to identify when four buses cease to be sufficient.*

The improved version presents a sequence of modeling decisions. Software is implementation detail and may be omitted from the Summary.

Checkpoint: Checkpoint: name the repair

For each transition, state what weakness is repaired:

1. capacity lower bound → feasible routing;
2. distance-minimizing routing → longest-ride local search;
3. nominal demand → perturbed-demand reruns.

A refinement should exist because the baseline misses something, not because a second algorithm sounds advanced.

5 Writing findings that carry evidence

The Findings block is not a storage place for every output. It should contain a hierarchy of claims.

Three levels of result

1. **Decision result**: the route, number, ranking, threshold, or recommendation the problem requested.
2. **Performance evidence**: improvement against a baseline or requirement.
3. **Robustness boundary**: the tested range within which the decision holds, and the point at which it changes.

A summary usually needs all three levels, but only a few carefully chosen numbers.

Decision result
What should be chosen?

Performance evidence
How much better, safer, cheaper, or faster?

Boundary and uncertainty
Under which scenarios does the recommendation remain valid or change?

Figure 3: Result selection should move from the decision to its quantitative support and then to its tested boundary.

A quantitative claim needs context

A useful result sentence normally includes several of the following:

- **value and unit**: 28.2 minutes, 87.4 km, four buses;
- **baseline or target**: down from 42.9 minutes; below the 45-minute limit;
- **scope**: morning route, 226 registered students, ordinary-day demand;
- **uncertainty or tested range**: 18 of 20 scenarios; ± 10 students by district;
- **decision implication**: keep four buses; activate a fifth under the trigger condition.

Worked example 4: progressively stronger result sentences

Version	Sentence
Too vague	Our optimized routes are much better than the current routes.
Value only	The mean ride time is 28.2 minutes.
Compared	The mean ride time falls from 42.9 to 28.2 minutes, a 34.3% reduction.
Decision-ready	With four buses, the mean ride time falls from 42.9 to 28.2 minutes and the longest route is 44.1 minutes, meeting the 45-minute requirement.
Calibrated	With four buses, the mean ride time falls from 42.9 to 28.2 minutes and the longest route is 44.1 minutes; four buses remain feasible in 18 of 20 perturbed-demand scenarios, while two high-demand scenarios require a fifth bus.

Model criticism: Number economy

More numbers do not automatically create more evidence. Select approximately three to six headline numbers that serve different jobs: the answer, the improvement, the constraint, and the robustness boundary. Move route-by-route details, parameter tables, and secondary metrics to the body.

Checkpoint: Checkpoint: repair the denominator

"The method succeeds 90% of the time" is incomplete. Rewrite it by specifying: succeeds at what, over how many scenarios or replications, under what input range, and relative to which threshold.

6 Trust language without overclaiming

The final block should tell the reader why the findings deserve confidence and where that confidence ends.

Four distinct evidence verbs

Term	Question answered	Summary language
verification	Did we implement or solve the model correctly?	"A hand-calculated four-stop case matched the program output."
validation	Does the model reproduce relevant reality or an accepted benchmark?	"Predicted ride times differed from two held-out mornings by a mean of 2.4 minutes."
sensitivity	How do results change when inputs or assumptions change?	"The four-bus plan remained feasible in 18 of 20 perturbed-demand scenarios."
uncertainty	What range or probability surrounds the estimate?	"The simulated 90th-percentile wait was 7.8–9.1 minutes across replications."

Do not use "validated" when the only check was that code ran without errors. Do not use "robust" without naming the tested range and outcome metric.

Worked example 5: caveat without self-sabotage

Weak

> *Our model has many unrealistic assumptions, so the results may not be accurate.*

Improved

> *The four-bus recommendation is stable under the tested* ± 10*-student district perturbations, but it excludes road closures and unusually concentrated absences. We therefore recommend four scheduled buses, an early fleet review at 236 projected riders, and a fifth-bus contingency from 241 confirmed riders or when a major route becomes unavailable.*

The improved version identifies the tested domain, the untested condition, and the operational response.

Precision language bank

Avoid	Use when supported
Our model proves...	Our model predicts... / Under the stated assumptions, the analysis implies...
The solution is optimal.	The solution is optimal for the stated linear model. / It was the best found among the tested candidates.
The result is robust.	The recommendation did not change when p varied from 0.8 to 1.2 times its baseline value.
The model is accurate.	The mean absolute validation error was 2.4 minutes on the held-out observations.
The method is efficient.	Runtime fell from 48 to 7 seconds for the 250-node test case.

7 Complete worked transformation

This section turns the evidence inventory from Worked Example 1 into a complete competition-style Summary. The case is illustrative; the purpose is sentence design, not the bus-routing mathematics.

Poor first draft

This paper studies a school bus routing problem. We use clustering, shortest paths, an optimization algorithm, Python, and sensitivity analysis. We find that our model works well and can reduce the routes. The results show that four buses are good, and the school should use our model. There are some limitations, but future work could improve them.

Model criticism: Diagnosis

The draft mentions a topic but not the actual constraints. It lists tools but not the modeling architecture. It gives no units, baseline, threshold, or task-by-task result. "Works well," "good," and "limitations" are unsupported labels. The recommendation has no trigger condition.

Final worked Summary

TASK. The school must transport 226 registered students using buses of capacity 60 while reducing travel time and keeping every scheduled morning ride below 45 minutes. We determine the minimum feasible fleet, construct a route plan, and identify when demand uncertainty requires additional capacity.
STRATEGY. We first obtain a capacity lower bound, showing that at least four buses are necessary. We then cluster nearby stops and connect them through time-weighted shortest paths on the road network. Because minimizing total distance can leave one excessively long route, a local-search stage exchanges stops between buses to reduce the maximum ride while preserving capacity. Finally, we rerun the assignment under twenty perturbed-demand scenarios in which each district gains or loses up to ten riders.
FINDINGS. **We recommend four ordinary-day routes totalling 87.4 km**. Relative to the existing 103.6-km plan, total route length falls by 15.6%. The predicted mean student ride falls from 42.9 to 28.2 minutes, while the longest scheduled ride is 44.1 minutes, below the 45-minute requirement. Four buses remain feasible in 18 of 20 perturbed-demand scenarios; in the two most concentrated high-demand cases, at least one route exceeds capacity and a fifth bus is required.
TRUST AND DECISION. A hand-traced small instance reproduces the route-cost calculation, and modeled ride times are compared with recorded morning travel times before optimization. The plan is therefore suitable for ordinary demand, but it should not be treated as unconditional. We recommend scheduling four buses, beginning an early fleet review at 236 projected riders, and activating a fifth-bus contingency from 241 confirmed riders, when district demand becomes unusually concentrated, or when a major road closure invalidates the travel-time matrix.

Checkpoint: Sentence-function audit

For the final Summary, underline:

1. the sentence that states the headline answer;
2. the sentence that establishes improvement against a baseline;
3. the sentence that reports the constraint test;
4. the sentence that reports sensitivity;
5. the sentence that changes the recommendation from unconditional to conditional.

If two different required tasks rely on the same vague sentence, add specificity.

8 Compression and revision

A one-page limit creates a selection problem. The goal is not to shorten every sentence equally; it is to remove low-value material while protecting the evidence chain.

Three-pass revision

1. **Completeness pass:** Does every major task have an answer? Are the recommendation and important conditions explicit?
2. **Evidence pass:** Does each major claim have a number, comparison, check, or tested range?
3. **Compression pass:** Remove history, process narrative, duplicated method detail, generic adjectives, software names, equations, and future-work filler.

Do not begin with compression. A short incomplete summary is still incomplete.

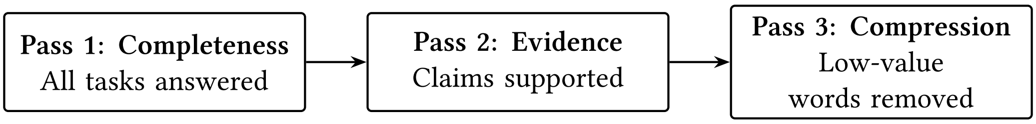


Figure 4: The revision order matters. Cutting before checking task coverage can delete the only sentence that answers a requirement.

Worked example 6: compression without losing evidence

Before – 44 words

> *In order to determine whether our proposed routing model would continue to function effectively when there were changes in the number of students assigned to each of the different geographic districts, we carried out a sensitivity analysis involving a number of alternative demand situations.*

After – 19 words

> *We reran the assignment under twenty scenarios with district demand perturbed by up to* ± 10 *students.*

The shorter sentence preserves the action, sample size, changed input, and perturbation range.

Model criticism: High-value details to protect

Protect task answers, decision thresholds, units, baselines, validation metrics, tested ranges, and recommendation conditions. Cut contest-year references, team workflow, code language, equations, exhaustive parameter lists, and generic phrases such as “future work could improve the model.”

9 Writing studio and peer audit

Core route: evidence-to-summary studio

Using the school-bus evidence inventory, write a 12–16 sentence summary without copying the worked version. Requirements:

- answer all four tasks;
- include at least four units or quantitative comparisons;
- include one verification or validation sentence;
- include one sensitivity boundary;
- end with a conditional recommendation.

Challenge route: two openings

Write two versions of the first four sentences:

1. a task-first opening;
2. a results-first opening.

Exchange drafts with a partner. Which version makes the answer easier to identify without making the context ambiguous?

Extension route: summary under conflicting metrics

Suppose the shortest route plan reduces total distance to 82 km but gives one student a 51-minute ride, while the balanced plan uses 87.4 km and keeps all rides below 45 minutes. Write two sentences explaining why the balanced plan is recommended. Name the objective, the constraint, and the trade-off without calling either solution "best" in an undefined sense.

Two-minute peer audit – classroom tool, not an official rubric

Score each item 0, 1, or 2.

Criterion	Question
Task coverage	Can every major requirement be matched to an answer?
Answer visibility	Can the headline recommendation be found in ten seconds?
Model architecture	Does the reader understand how the methods connect and why refinements exist?
Quantitative evidence	Do key claims include units, baselines, thresholds, or uncertainty?
Trust calibration	Are verification/validation/sensitivity stated accurately, with a boundary or caveat?
Compression and consistency	Is every sentence useful, and do all numbers agree with the body?

After scoring, give only two comments: the highest-value sentence and the single most important repair.

Checkpoint: Exit ticket

Write one sentence for each prompt:

1. The headline answer to a modeling problem you have solved.
2. The baseline or threshold that makes the answer meaningful.
3. The evidence that supports trust in the answer.
4. The condition under which the recommendation would change.

These four sentences form the core of a future Results and Trust/Decision block.

10 What to carry into W3

Class W2 take-aways

1. Build an evidence inventory before writing prose.
2. Treat the four blocks as functions: Task, Strategy, Findings, Trust/Decision.
3. Put the headline answer before secondary outputs.
4. Report numbers with units, baselines, thresholds, and scope.
5. Use verification, validation, sensitivity, and uncertainty accurately.
6. A limitation is useful when it produces a condition, trigger, or caution.
7. Revise in the order completeness $\rightarrow$ evidence $\rightarrow$ compression.
8. The Summary must agree exactly with the body, tables, and conclusion.

In **W3: Assumptions, Validation, and Sensitivity** we develop the evidence behind the trust sentences introduced here. Students will design assumptions that can be tested, distinguish model verification from real-world validation, and turn parameter changes into decision-relevant sensitivity conclusions.

Appendix A: Evidence-inventory worksheet

Task	Headline answer	Method	Evidence/check	Condition/caveat

Appendix B: Sentence frames

- **Task:** *[Decision maker] must [decision] under [constraints]; we therefore estimate/design/compare [outputs].*

- **Architecture**: *We first [baseline], then [refinement] to address [named weakness], and finally [test/decision stage].*
- **Headline result**: *We recommend [action/quantity], which produces [metric with units] under [scope].*
- **Comparison**: *Relative to [baseline], [metric] changes from [old] to [new], a [percentage or absolute] improvement.*
- **Validation**: *Against [held-out data/benchmark], the model achieved [error or agreement metric].*
- **Sensitivity**: *The recommendation remained unchanged when [input] varied over [range], but changed when [threshold].*
- **Conditional decision**: *Use [action A] under [condition]; switch to [action B] when [trigger].*

Appendix C: Final consistency checklist

- Every required task appears in the Summary and body.
- The headline numbers match the final results table exactly.
- Units, denominators, time horizons, and scenario counts are visible.
- "Optimal," "validated," "robust," and "accurate" are supported by the correct evidence.
- No software name appears unless it matters to reproducibility or interpretation.
- No equation is included unless the equation itself is a headline result.
- The recommendation includes a decision maker, action, and relevant condition.
- The Summary contains no citation or claim absent from the reference list or body.
- The page follows the current official contest template and identification rules.

Official-source note. Contest-specific statements were checked in July 2026 against the COMAP 2026 HiMCM/MidMCM Instructions and the 2026 IM^{2}C USA Regional Instructions. Rules, templates, and regional requirements may change; verify the current official pages before each contest.

Model audit card

Audit before trusting or communicating a model

Purpose	What question or decision does the model support?	**Data**	Which inputs are observed, derived, or simulated?
Assumptions	Which simplifications may change the answer?	**Units**	Are dimensions and scales consistent?
Verification	Were equations, code, constraints, and limiting cases checked?	**Validation**	Is performance tested against independent evidence?
Uncertainty	How do measurement, parameter, solver, and structural errors affect conclusions?	**Ethics**	Who benefits, who bears risk, and what fairness criterion matters?
Reproducibility	Are data provenance, versions, seeds, and transformations recorded?	**Communication**	Are claims conditional, traceable, and decision-relevant?

Further reading

- HiMCM / MidMCM Contest Instructions (COMAP)
- Structure Your Article (IEEE Author Center)

Sources, provenance, and licence

These notes follow the iterative modeling and assessment perspective in the SIAM/COMAP GAIMME report. Competition-specific remarks are a snapshot checked on 20 August 2026 against the official COMAP HiMCM instructions; readers should verify current rules before submission. Unless a source is explicitly identified, classroom datasets and numerical scenarios are synthetic or simulated teaching examples and are not empirical claims about a real institution. Third-party references retain their own terms.

Student response guide and checking criteria

For closed calculations, show the definition, substitution, units, and at least one independent check; the worked examples in this unit are method hints, but you must complete the arithmetic yourself. Open modeling responses are checked on five dimensions: a clear purpose and output; stated data and provenance; defensible assumptions and scope; traceable computational verification and model validation; and alignment among uncertainty, limitations, and the recommendation. Support each dimension with concrete evidence rather than writing only "completed."

Writing Classes for Mathematical Modeling Competitions

W3: Building Trust – Assumptions, Validation, and Sensitivity

Kenneth Sok Kin Cheng

Revised 20 August 2026

Learning goals

By the end of this two-hour writing class, students should be able to:

- explain how assumptions, verification, validation, sensitivity, and uncertainty play different roles in a modeling report;
- write a four-part **assumption card**: statement, rationale/evidence, model consequence, and stress test;
- distinguish a modeling assumption from a definition, parameter value, scope choice, or mathematical requirement;
- build an **assumption ledger** that links each assumption to an equation, output, and test;
- distinguish **verification** from **validation**, and distinguish both from calibration;
- select validation evidence appropriate to the model's purpose and available data;
- design one-at-a-time, scenario, structural, and simulation-based sensitivity checks;
- report sensitivity using input ranges, output metrics, decision changes, and operational triggers;
- separate parameter sensitivity from uncertainty in estimated outputs;
- write a concise trust section that supports rather than merely praises a recommendation.

1 Trust is an evidence structure, not a confidence adjective

A model becomes useful only when a reader can see why its conclusions deserve attention. A sentence such as "our model is accurate and robust" adds no trust by itself. Trust comes from a visible chain linking the real-world interpretation to tests and decision consequences.

Official competition guidance supports this emphasis. Current COMAP guidance asks teams to state assumptions with justifications and to analyse their work, including sensitivity to changed or removed assumptions. IM^{2}C resources describe modeling as a cycle in which solutions are tested and evaluated against the real situation. Exact section names may vary, but the evidence obligations do not.

Launch activity: rank the trust claims

A team recommends four buses. Rank the following claims from least to most useful, then explain what evidence is missing.

1. "Our model is highly reliable."
2. "All buses satisfy capacity 60 in the baseline data."
3. "Predicted route times have mean absolute error 2.4 minutes on 12 held-out school days."

4. “Four buses remain sufficient in 18 of 20 demand scenarios; a fifth is required when total ridership exceeds 240.”

The last three answer different questions. None replaces the others.

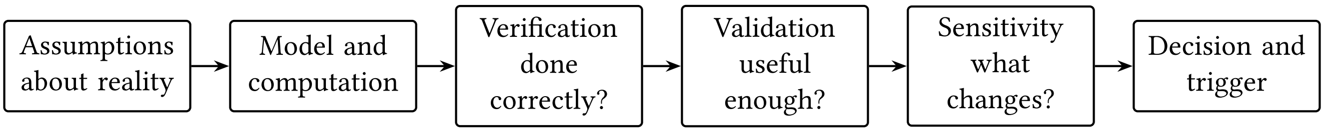


Figure 1: A report earns trust through a chain of evidence. A weakness in one link should appear as a qualification or further test, not be hidden.

Five terms that must not be merged

Term	Question	Typical evidence
Assumption	What claim about the system makes the model possible?	Domain evidence, data pattern, scope rationale, and a stated trade-off.
Calibration	What parameter values make the model fit the chosen data?	Estimation procedure, training data, objective function, fitted values.
Verification	Did we implement and solve the stated model correctly?	Units, invariants, hand cases, code tests, convergence, independent implementation.
Validation	Is the model adequate for its intended purpose?	Held-out data, external comparison, field evidence, cross-method prediction, domain plausibility.
Sensitivity / uncertainty	How do inputs and imperfect knowledge affect outputs or decisions?	Perturbations, scenarios, distributions, intervals, structural alternatives, decision switches.

Checkpoint: Checkpoint: classify before you write

Classify each sentence.

1. “We estimate service rate μ from 40 observed service times.”
2. “When arrivals are set to zero, the simulated queue remains empty.”
3. “Predicted average wait differs from observed wait by 0.8 minutes.”
4. “Replacing exponential service with fixed service reduces the recommendation from three servers to two.”
5. “We treat morning demand as stationary from 7:15 to 8:00.”

Suggested labels: calibration, verification, validation, structural sensitivity, assumption.

2 Write assumptions as testable modeling choices

An assumption is neither an apology nor a list of obvious facts. It is a deliberate claim about the system that changes the mathematics or the scope of the conclusion.

The four-part assumption card

For every material assumption, write:

1. **Statement – what?** A precise claim that could be false.
2. **Rationale/evidence – why?** Data, literature, domain logic, or a transparent simplification reason.
3. **Model consequence – effect?** Which equation, algorithm, parameter, or computational burden changes because of it?
4. **Risk/stress test – then what?** How might the assumption fail, and which sensitivity or validation check addresses that risk?

The fourth part closes the loop between the Assumptions section and later evidence.

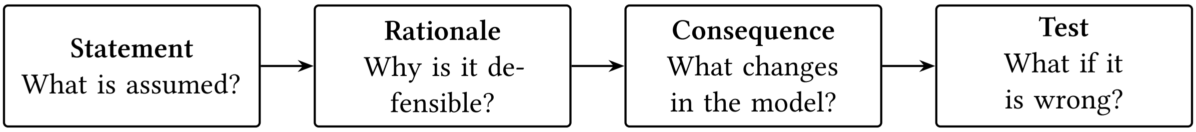


Figure 2: A complete assumption card connects interpretation, mathematics, and later testing.

Worked example 1: assumption cards for the school-bus case

Assumption A – registered riders represent daily demand.
Statement. The baseline routing problem includes all 226 registered riders and treats each as travelling on the modeled morning.
Rationale. Registration data are the only complete location data available and provide a conservative ordinary-day demand estimate.
Model consequence. The capacity constraints use 226 fixed demand points, producing the lower bound $\lceil 226/60 \rceil = 4$ buses.
Risk/test. Absence and late-registration patterns can change district loads. We therefore rerun the model under 20 district-demand scenarios with changes of up to ± 10 riders.

Assumption B – observed morning travel times are reusable.
Statement. Link travel times estimated from 12 normal school mornings are used for future ordinary weekdays.
Rationale. The data exclude holidays, roadworks, and severe-weather days, and the within-link variation is smaller than the between-link variation.
Model consequence. Routes can be evaluated through a fixed time matrix rather than a time-dependent traffic simulator.
Risk/test. Peak congestion may increase all route times. We test multiplicative travel-time factors from 0.90 to 1.10 and state a delay trigger.

Model criticism: Do not force every sentence into the Assumptions section

- "Bus capacity is 60" may be an input or policy constraint, not an assumption, if it is specified by the problem.
- "Let $x_{ij} = 1$ when bus i serves stop j" is a definition.
- "Every route must begin and end at school" is a feasibility rule.
- "Travel times are stable across ordinary mornings" is an assumption because reality may violate it.
- "We only study the morning journey" is a scope decision; explain it where the scope is introduced and record its effect on conclusions.

The section should contain choices that matter, not every fact used in the report.

The assumption ledger

A ledger prevents assumptions from disappearing after page 3.

Assumption	Where it enters	Risk if false	Evidence or test
Daily demand $= 226$	Capacity and route allocation	Overload or unnecessary fleet	Demand scenarios; threshold at 240 riders
Stable link times	Time matrix and ride limit	Underestimated maximum ride	Holdout GPS error; 0.90–1.10 time factors
No road closure	Shortest-path network	Infeasible planned route	Scenario removing the most-used road
One pickup cluster per stop	Demand aggregation	Extra walking burden	Maximum walking-distance check

An assumption with no downstream effect is probably irrelevant. A high-risk assumption with no test is a visible evidence gap.

Checkpoint: Assumption audit

For one assumption in your current project, underline the exact equation, data-processing step, or algorithm that depends on it. Then write the sentence that will report its stress test. If neither can be identified, revise or remove the assumption.

3 Verification: did we solve the stated model correctly?

Verification concerns the mathematics and implementation of the model as written. A model can be perfectly verified and still be invalid for reality; it can also be conceptually suitable but incorrectly coded.

A practical verification menu

1. **Dimensional/unit check:** both sides of every equation have compatible units.
2. **Constraint check:** every reported solution satisfies capacity, non-negativity, budget, conservation, and logical rules.
3. **Limiting or hand case:** a tiny case with a known answer is reproduced.
4. **Invariant check:** conserved totals remain conserved; probabilities sum to one; flows balance.
5. **Numerical refinement:** reducing step size or tightening tolerance changes outputs negligibly.
6. **Independent implementation:** a second method, package, or teammate reproduces key results.
7. **Reproducibility record:** data version, random seed, solver settings, and stopping rules are stated when material.

Choose checks that match the model. Do not write a generic paragraph claiming that "the code was checked."

Worked example 2: verification paragraph

Before interpreting the optimized routes, we verified feasibility and implementation. All 226 students are assigned exactly once, no bus exceeds capacity 60, every route starts and ends at school, and computed route distance equals the sum of its constituent links. On a six-stop hand case, the program reproduces the enumerated optimum of 18.6 km. Running the local-search phase from 50 random initial solutions produces the same best objective in 43 runs and values within 0.7% in all runs. These checks support the claim that the reported solution is a consistent output of the stated routing model; they do not yet establish that the travel-time assumptions represent future mornings.

Model criticism: Calibration is not validation

A fitted regression can match its training data because its parameters were chosen to do so. A queue model can reproduce a sample mean after its service rate is calibrated to that mean. This establishes calibration, not independent validation.

Reserve some evidence for testing: a later time period, held-out locations, external cases, a different observable, or an independent method. When data are scarce, state the limitation rather than calling in-sample fit "validation."

4 Validation: is the model adequate for the decision?

Validation is purpose-dependent. A route-time model may be adequate for choosing between two route plans even if it cannot predict every individual journey to the nearest minute. State the purpose, metric, comparator, and acceptance logic.

The validation claim template

A complete validation claim answers four questions:

1. **Target:** What observable or decision-relevant quantity is tested?
2. **Comparator:** Against what data, benchmark, alternative method, or established range?
3. **Metric:** Error, agreement rate, rank correlation, coverage, or qualitative pattern?
4. **Judgment:** Why is this level of agreement adequate for the model's intended use?

Example: "On 12 held-out mornings, route-time predictions have MAE 2.4 minutes and maximum error 5.1 minutes. Because route comparisons differ by 8–19 minutes, this accuracy is adequate for ranking the candidate plans, although not for promising exact arrival times."

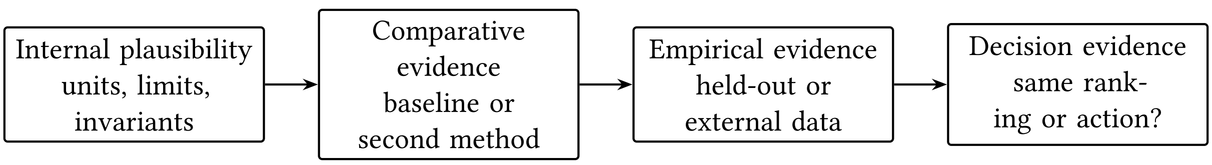


Figure 3: Validation strength usually increases as evidence becomes more independent and more closely tied to the intended decision. Not every project can reach every level.

A model-specific validation menu

Model type	Useful validation evidence	Weak substitute to avoid
Regression / prediction	Held-out error, residual pattern, baseline comparison, interval coverage	Reporting only training R^2
Optimization	Feasibility, lower/upper bounds, comparison with baseline or benchmark, scenario performance	Calling the solver output "optimal" without gap or scope
Simulation	Hand case, analytical special case, convergence across replications, output-data comparison	One run with no random-seed or uncertainty report
Dynamic model	Initial-condition reproduction, conservation, qualitative trend, out-of-sample trajectory	A visually similar curve fitted to the same data
Decision index / ranking	Expert or external ranking, stability of top choices, back-test on known cases	High internal consistency with no external criterion

Worked example 3: validation with a benchmark and held-out data

The routing model is validated at two levels.
Travel-time prediction. Twelve school mornings are held out from model construction. Across the 48 route-day combinations, the mean absolute error is 2.4 minutes and the largest error is 5.1 minutes. Errors show no systematic increase with route length.
Decision comparison. The model ranks the existing route plan worse than the proposed plan on all 12 mornings. The observed mean difference is 13.6 minutes per rider, compared with a predicted difference of 14.7 minutes. The recommendation therefore survives the validation data even though exact route times remain uncertain.

Checkpoint: Repair the validation sentence

Weak: "Our model agrees closely with the real data and is therefore valid."
Rewrite it by adding the target, comparator, metric, and decision judgment. Avoid the absolute phrase "the model is valid"; validation supports adequacy for a stated purpose, not universal truth.

5 Sensitivity and uncertainty: will the decision survive change?

Sensitivity analysis varies inputs or modeling choices and measures their effect. **Uncertainty analysis** represents imperfect knowledge about inputs or stochastic outputs and reports a range or distribution. They are related but not identical.

Four useful levels of sensitivity analysis

1. **Local/OAT sensitivity:** vary one input over a plausible range while holding the others fixed. Fast and interpretable, but misses interactions.
2. **Scenario analysis:** vary several inputs together in coherent situations such as ordinary, peak, and disruption days.

3. **Structural sensitivity**: replace a distribution, objective, functional form, aggregation rule, or model family.
4. **Global/simulation-based sensitivity**: sample uncertain inputs jointly and study output distributions or importance measures.

The method should match the decision. A small two-parameter model may need only OAT plus a structural check; a nonlinear stochastic system may need joint sampling.

Report sensitivity in decision language

For each varied input, state:

1. nominal value and plausible range;
2. output metric and baseline value;
3. magnitude or direction of output change;
4. whether the recommendation, ranking, feasibility, or threshold changes;
5. what the finding implies for data collection or contingency planning.

A result can be numerically sensitive but decision-stable, or numerically stable but close to a hard constraint. Report both the output and the action.

Worked example 4: OAT table and decision boundary

Input/test	Baseline	Lower case	Upper case	Decision effect
Total ridership	226	206	246	Four buses below 241; five at 246
Travel-time factor	1.00	0.90	1.10	Distance plan unchanged; 45-min limit fails above 1.02
Bus capacity	60	54	66	Five buses at 54; four at 60 or 66
Distance weight in objective	1.00	0.75	1.25	Same fleet; route order changes slightly
Road availability	all links	–	remove busiest link	Four buses remain feasible; distance rises 6.8 km

The fleet recommendation is stable to moderate travel-time and network changes but changes at two clear thresholds: ridership above 240 or effective capacity below 57. The ride-time service standard is more fragile than the fleet size and therefore needs a scheduling buffer.

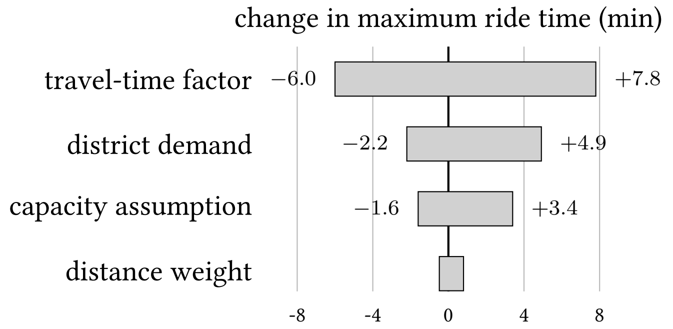


Figure 4: Illustrative tornado diagram. Bars show output change over stated input ranges; they do not by themselves show probability or decision thresholds.

Model criticism: A tornado diagram needs a table behind it

A bar chart without stated input ranges is not reproducible. A sorted bar chart also does not show parameter interactions or likelihood. Include the nominal values and ranges in a table or caption, then use prose to identify the decision-relevant threshold.

Optional normalized sensitivity

When inputs have different units, a local **elasticity** can help compare relative influence:

$$E_{y,x} \approx \frac{\Delta y/y_0}{\Delta x/x_0}.$$

An elasticity of 1.5 means a 1% input change produces approximately a 1.5% output change near the baseline. This is a local summary, not proof of global linearity.

Worked example 5: structural sensitivity changes the recommendation

The shortest-total-distance formulation produces an 82.0-km plan, but one student group remains on a bus for 51.0 minutes. Replacing the objective with a balanced objective that penalizes maximum ride time gives 87.4 km and a 44.1-minute maximum.
The recommendation is therefore structurally sensitive to the objective function. We do not call the 82.0-km plan "better" merely because its total distance is lower. We recommend the balanced plan because it satisfies the stated service standard, while reporting the 5.4-km efficiency cost of that choice.

6 Integrated case: write the complete trust story

The following example reuses the school-bus evidence from W2. The purpose is to show how the same numerical work becomes three distinct report sections.

Worked example 6: competition-ready Assumptions paragraph

A1: ordinary-day demand. We model all 226 registered riders as travelling on the target morning. Registration data provide the only complete geocoded demand set and give a conservative baseline. This assumption fixes the capacity requirement and implies a minimum of four 60-seat buses. Because absence and new registration can redistribute loads, Section 7 tests 20 district-demand scenarios and reports the fleet-change threshold.
A2: reusable travel times. Link times are estimated from 12 normal school mornings and treated as representative of future ordinary weekdays. This avoids a separate traffic-flow simulation and allows

candidate routes to be compared through one time matrix. The assumption is weaker on disruption days, so validation uses held-out mornings and sensitivity multiplies all link times by factors from 0.90 to 1.10.

Worked example 7: competition-ready Verification and Validation paragraph

Verification. Every rider is assigned exactly once; capacities, school start/end requirements, and route continuity are satisfied. The program reproduces a six-stop enumerated optimum and returns the same best objective in 43 of 50 random starts.
Validation. On 12 held-out mornings, predicted route times have MAE 2.4 minutes and maximum error 5.1 minutes. The model predicts that the proposed plan improves mean ride time by 14.7 minutes; the observed improvement is 13.6 minutes. This agreement is sufficient for comparing plans, although individual arrival times should retain a buffer.

Worked example 8: competition-ready Sensitivity and Decision paragraph

Demand perturbations of up to ± 10 riders per district leave the four-bus plan feasible in 18 of 20 scenarios. The two failures occur when total ridership exceeds the aggregate capacity of 240. Travel-time perturbation does not change fleet size, but the 45-minute service limit is crossed when ordinary-day link times rise by more than approximately 2%. Removing the most-used road increases distance by 6.8 km but does not change the fleet recommendation.
We therefore recommend four buses for ordinary days, a three-minute timetable buffer, and an early review at 236 projected riders and a fifth-bus trigger from 241 confirmed riders or a major disruption is forecast. This is a conditional operational recommendation, not a claim that four buses are universally sufficient.

Why the three sections should not repeat each other

- **Assumptions** explain the choices and identify risks before results are presented.
- **Verification/validation** show whether the implemented model and its predictions are credible.
- **Sensitivity/uncertainty** show where the output or decision changes and convert that boundary into a trigger.

Repeating "the model is reasonable" in all three sections wastes space and leaves the evidence invisible.

7 Writing patterns that weaken trust

Model criticism: Repair these common overclaims

Weak wording	Evidence-based repair
"The assumption is reasonable."	State the source, observed range, simplification benefit, and failure mode.
"The model is validated."	State the tested target, data or benchmark, metric, and intended-use judgment.
"The result is robust."	State the input range, output change, and whether the decision changed.
"The error is small."	Give the metric, units, baseline scale, and comparator.
"Monte Carlo confirms the model."	State what was sampled, how many replications, uncertainty, and which analytical claim was reproduced.
"Changing assumptions has little effect."	Name the changed assumption and quantify the change in output or decision.

Checkpoint: The evidence sentence test

Circle every adjective in your current trust sections: reasonable, accurate, realistic, reliable, robust, stable, negligible, significant. For each, add the evidence needed to make it measurable, or delete it.

8 Differentiated writing studio

Core route: build the assumption ledger

For a cafeteria queue model, complete four ledger rows using these candidate assumptions: stationary arrival rate, independent arrivals, fixed number of servers, and no customer abandonment. For each row, identify where it enters, the risk if false, and one stress test.

Challenge route: design the validation plan

A regression predicts bicycle usage from temperature, rainfall, day type, and route length. You have 180 daily observations. Propose:

1. one calibration procedure;
2. two verification checks;
3. two validation metrics on held-out data;
4. one baseline model;
5. one criterion for whether the model is adequate for staffing a bicycle-storage facility.

Do not use training R^2 as your only evidence.

Extension route: turn uncertainty into a decision rule

A reservoir model predicts that a release of 24 ML/day keeps storage above the safety threshold in 91 of 100 sampled climate-demand scenarios. At 22 ML/day, success rises to 98 of 100 but water supply to farms falls by 8%. Write a recommendation that states the trade-off, uncertainty, and trigger for switching policies.

Peer audit: no author explanation

Exchange drafts and answer:

1. Which assumption creates the greatest decision risk?
2. What was verified, and what was validated?
3. Is any calibration evidence incorrectly called validation?
4. Which sensitivity range is most decision-relevant?
5. At what threshold does the recommendation change?
6. Which sentence overclaims beyond the evidence?

A trust section must survive without the author explaining what they intended.

9 Exit ticket and preparation for W4

Exit ticket

Write four sentences for your current project:

1. one complete assumption card compressed into one or two sentences;
2. one verification claim with a concrete check;
3. one validation claim with a comparator and metric;
4. one sensitivity claim ending with the condition that changes the decision.

If the four sentences sound interchangeable, the distinctions are not yet clear.

Class W3 take-aways

1. Assumptions should be connected to equations, risks, and later tests.
2. Verification asks whether the stated model was solved correctly; validation asks whether it is adequate for a stated purpose.
3. Calibration uses data to set parameters and should not be presented as independent validation.
4. Sensitivity analysis needs an input range, output metric, and decision consequence.
5. Structural sensitivity often reveals more than repeatedly perturbing numerical parameters.
6. Uncertainty should become an interval, probability, scenario frequency, or operational trigger.
7. Trust is earned by traceable evidence, not by adjectives.

In **W4: Academic English for Modeling Reports**, we move from evidence architecture to sentence and paragraph control: claim strength, hedging, active voice, equation integration, transitions, and common translation patterns.

Appendix A: Assumption-ledger worksheet

Assumption statement	Rationale/evidence	Model consequence	Risk and planned test

Appendix B: Validation-plan matrix

Claim to test	Independent evidence	Metric	Comparator	Adequacy rule

Appendix C: Sensitivity-report sentence frames

- *Varying [input] from [lower] to [upper] changes [output] from [value] to [value], while [decision] remains unchanged.*
- *The recommendation switches from [A] to [B] when [input] crosses approximately [threshold].*
- *Among the tested inputs, [parameter] has the greatest effect on [metric]; a $\pm X\%$ change produces [output range].*
- *Replacing [model assumption] with [alternative] changes [output/decision], indicating structural sensitivity to [feature].*
- *Across [number] jointly sampled scenarios, [decision] is feasible in [number or percentage]; failures occur primarily when [condition].*
- *The numerical output is sensitive, but the ranking remains stable because [margin or threshold].*

Appendix D: Before calling a result robust

Check that the report states:

- ☐ the input or assumption varied;
- ☐ the plausible range and its source or rationale;
- ☐ the output metric and baseline;
- ☐ the numerical change in the output;
- ☐ whether feasibility, ranking, or recommendation changed;
- ☐ interactions or structural alternatives considered, where material;
- ☐ the operational threshold or contingency implied by the test.

Appendix E: Official resources used in this lesson

- COMAP HiMCM/MidMCM Tips Article: emphasizes assumptions with justifications, clear modeling processes, sensitivity, strengths/weaknesses, conclusions, and readable presentation.
- COMAP problem guidance example: describes sensitivity as adjusting or removing an assumption to examine its impact.
- IM^{2}C Mathematical Modelling Framework: presents modeling as a cyclic process involving interpretation, mathematical work, evaluation, and refinement.
- IM^{2}C Writing a Modelling Report: stresses clear, complete communication so others can evaluate and use the model.

Contest-specific requirements can change. Re-check the current official rules and problem instructions before each competition.

Model audit card

Audit before trusting or communicating a model

Purpose	What question or decision does the model support?	**Data**	Which inputs are observed, derived, or simulated?
Assumptions	Which simplifications may change the answer?	**Units**	Are dimensions and scales consistent?
Verification	Were equations, code, constraints, and limiting cases checked?	**Validation**	Is performance tested against independent evidence?
Uncertainty	How do measurement, parameter, solver, and structural errors affect conclusions?	**Ethics**	Who benefits, who bears risk, and what fairness criterion matters?
Reproducibility	Are data provenance, versions, seeds, and transformations recorded?	**Communication**	Are claims conditional, traceable, and decision-relevant?

Further reading

- GAIMME: Guidelines for Assessment and Instruction in Mathematical Modeling Education (SIAM / COMAP)
- HiMCM / MidMCM Contest Instructions (COMAP)

Sources, provenance, and licence

These notes follow the iterative modeling and assessment perspective in the SIAM/COMAP GAIMME report. Competition-specific remarks are a snapshot checked on 20 August 2026 against the official COMAP HiMCM instructions; readers should verify current rules before submission. Unless a source is explicitly identified, classroom datasets and numerical scenarios are synthetic or simulated teaching examples and are not empirical claims about a real institution. Third-party references retain their own terms.

Student response guide and checking criteria

For closed calculations, show the definition, substitution, units, and at least one independent check; the worked examples in this unit are method hints, but you must complete the arithmetic yourself. Open modeling responses are checked on five dimensions: a clear purpose and output; stated data and provenance; defensible assumptions and scope; traceable computational verification and model validation; and alignment among uncertainty, limitations, and the recommendation. Support each dimension with concrete evidence rather than writing only “completed.”

Writing Classes for Mathematical Modeling Competitions
W4: Academic Prose – Precision, Flow, and Evidence

Kenneth Sok Kin Cheng

Revised 20 August 2026

Learning goals

By the end of this two-hour writing class, students should be able to:

- treat academic writing as evidence navigation rather than decoration or translation;
- assign a clear rhetorical role to every sentence: claim, evidence, interpretation, qualification, or transition;
- build paragraphs using a claim–evidence–interpretation–implication structure;
- calibrate verbs and hedges to the strength and scope of the available evidence;
- choose active or passive voice according to information focus rather than habit;
- use tense consistently for model statements, completed procedures, reported results, and recommendations;
- create cohesion through the **given–new principle**, controlled repetition, and clear reference chains;
- introduce equations, tables, and figures as grammatical parts of the argument;
- revise wordy, ambiguous, or translation-shaped sentences without deleting necessary reasoning;
- identify high-frequency transfer patterns in English written by Chinese- and Cantonese-speaking students;
- perform macro-, paragraph-, sentence-, and proof-reading passes in the correct order;
- produce a polished competition-style results paragraph whose claims remain traceable to evidence.

1 Academic prose helps a reader verify an argument

Scientific writing is not successful because it sounds formal. It is successful because a reader can locate the claim, identify the evidence, understand the inference, and see the condition under which the conclusion applies. A complicated sentence can hide weak reasoning; a plain sentence often exposes it.

Launch activity: same result, different prose

All three versions use the same facts from an illustrative synthetic teaching case: four buses, 87 km of daily routes, mean ride time reduced from 43 to 28 minutes, and a fifth bus needed above 240 riders.
A. "After a comprehensive and sophisticated analysis, our highly robust model perfectly determines an optimal arrangement of buses."
B. "We used optimization. The answer is four buses. The result is good and can solve the problem."
C. "The optimized plan uses four buses and 87 km of daily routes, reducing mean ride time from 43 to 28 minutes. Four buses remain feasible in 18 of 20 demand scenarios; a fifth is required when ridership exceeds 240."

Underline the claim, the evidence, the comparison, and the condition in Version C. Then explain why Version A sounds confident but supplies less information.

Three standards for competition prose

1. **Precise**: quantities, comparisons, conditions, and referents are identifiable.
2. **Traceable**: the reader can connect a claim to an equation, table, figure, dataset, or test.
3. **Economical**: every sentence performs a necessary function; brevity does not remove evidence or interpretation.

Formal vocabulary is optional. These three standards are not.

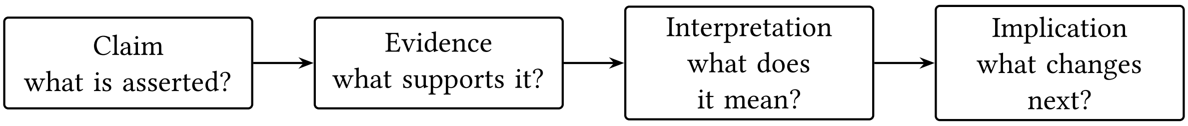


Figure 1: A strong paragraph moves from a claim to evidence, interprets the evidence, and states its modeling or decision implication. Not every paragraph needs four separate sentences, but it needs the four functions.

Checkpoint: Checkpoint: informative or merely formal?

Replace the empty formality in each sentence with information.

1. “It is worth mentioning that our model is very effective.”
2. “A comprehensive analysis was conducted on the obtained results.”
3. “Figure 4 clearly shows an obvious trend.”
4. “Our method has strong practical significance.”

A useful revision supplies a metric, comparison, condition, or decision consequence.

2 Give every sentence a job

A sentence can be grammatically correct and still weaken a paragraph because it performs no identifiable function. Before polishing vocabulary, decide what each sentence is doing.

Five common sentence roles

Role	Reader question	Typical form
Claim	What are you asserting?	“The four-bus plan satisfies all baseline capacity constraints.”
Evidence	What supports the claim?	“Maximum occupancy is 57 passengers against capacity 60.”
Interpretation	Why does the evidence matter?	“The margin is small, so the plan may be fragile under demand growth.”
Qualification	Under what scope or condition?	“This conclusion applies to the observed 7:00–8:00 demand window.”
Transition	How does this lead to the next step?	“We therefore test higher-ridership scenarios in Section 7.”

Example 1: a paragraph with visible sentence roles

Claim. The four-bus solution is feasible but operates close to its capacity limit. **Evidence.** In the baseline assignment, the most heavily loaded bus carries 57 passengers against a capacity of 60. **Interpretation.** The three-seat margin is unlikely to absorb substantial enrollment growth or clustered absences from alternative routes. **Transition.** We therefore examine demand scenarios up to 260 riders before making the final recommendation.

Method: the sentence-function audit

Write one letter beside every sentence in a draft: C (claim), E (evidence), I (interpretation), Q (qualification), or T (transition).

- A paragraph containing only C sentences is unsupported.
- A paragraph containing only E sentences is a data dump.
- A paragraph containing no I sentence forces the reader to perform the modeling interpretation.
- Repeated T sentences create signposting without content.
- A Q sentence should narrow a claim, not make it impossible to understand.

Model criticism: One paragraph, one controlling question

A paragraph should answer one main question, such as:

- Why is this assumption reasonable?
- How is this parameter estimated?
- What does the result show?
- Is the model adequate for the intended decision?
- Which parameter changes the recommendation?

If a paragraph answers two unrelated questions, split it. If two consecutive paragraphs answer the same question, combine or differentiate them.

3 Match claim strength to evidence strength

Hedging is not the insertion of words such as "perhaps" or "might." It is the accurate matching of a claim's scope and force to the evidence available.

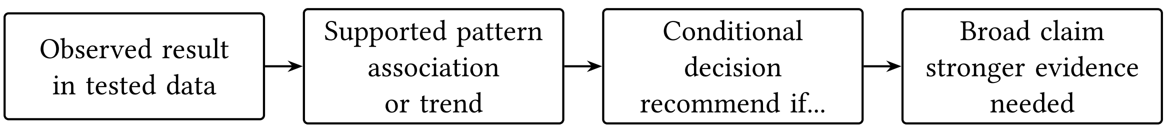


Figure 2: As a claim becomes broader or more causal, the required evidence becomes stronger. Competition papers often justify conditional recommendations more convincingly than universal claims.

A reporting-verb scale

Verb or phrase	Appropriate use	Example
shows / gives	Direct mathematical or tabulated result	"Equation (8) gives an early-review trigger of 236 projected riders."
indicates / suggests	Pattern supported but not established universally	"The scenario results suggest that route 3 is the main bottleneck."
is consistent with	Agreement without claiming confirmation	"The estimate is consistent with the 2025 school count."
supports	Evidence strengthens a claim but does not prove it	"Held-out errors below 3 minutes support use for scheduling."
recommends	A decision derived under stated criteria	"Under the tested scenarios, the model recommends four buses."
proves / guarantees	Reserve for a mathematical proof or explicit guarantee	Usually inappropriate for empirical validity or real-world success.

Example 2: overclaim, over-hedge, calibrated claim

Overclaim: "Our model proves that four buses are optimal for every future year."
Over-hedge: "The results might perhaps indicate that four buses could possibly be enough in some situations."
Calibrated: "Under the observed road network and tested demand scenarios up to 240 riders, the shortest feasible plan found uses four buses and satisfies the capacity and ride-time constraints."
The final version specifies the evidence domain, the objective, and the constraints. It needs no vague adverb.

Method: quantitative hedging

Prefer a measurable boundary to a vague qualifier.

- "approximately 28 minutes" → "28.1 minutes after rounding to the nearest minute";
- "fairly accurate" → "MAE 2.4 minutes on 12 held-out days";
- "robust to demand" → "unchanged for total ridership from 205 to 240";
- "usually feasible" → "feasible in 18 of 20 tested scenarios";
- "small error" → state the error and explain why it is adequate for the decision.

Checkpoint: Checkpoint: choose the strongest defensible verb

For each evidence statement, choose among *shows*, *suggests*, *is consistent with*, *supports*, and *proves*.

1. A symbolic derivation establishes that the objective is convex.
2. A simulation with 20 scenarios places route 3 at the capacity limit in 16 cases.
3. A predicted count of 232 differs from an external school count of 228.
4. A held-out MAE of 2.4 minutes is below the 5-minute scheduling tolerance.

4 Use voice and tense to control information focus

Active and passive voice are not competing styles. They place different information in the subject position.

Voice choice by purpose

- Use **active voice** when the team's choice, interpretation, or responsibility matters: "We define," "we exclude," "we compare," "we recommend."
- Use **passive voice** when the procedure or object should occupy the subject position and the agent is already known or irrelevant: "Travel times were estimated from GPS records."
- Avoid passive voice when it hides responsibility: "Three data points were removed" should become "We removed three duplicated records" and should state the criterion.
- Avoid forced active voice when "we" becomes repetitive: several consecutive sentences beginning with "We" can often be reorganized around the model, data, or result.

Example 3: information focus, not mechanical alternation

Choice: "We restrict the morning period to 7:00–8:00 because demand is approximately stationary within this window."
Procedure: "Link travel times were estimated from 30 GPS runs and rounded to the nearest 0.1 minute."
Interpretation: "We interpret the 2.4-minute held-out error as adequate because bus departures are scheduled at 5-minute intervals."
Result focus: "A fifth bus is required once total ridership exceeds 240."

A practical tense system

Tense	Main use	Example
Present	Equations, figures, general model properties, current argument	"Equation (6) enforces capacity." "Figure 4 shows the threshold."
Past	Completed data collection, fitting, simulation, or experiment	"We collected 30 GPS runs and fitted the link times."
Present perfect	Prior work or changes continuing to the present; use sparingly	"Several studies have modeled school transport as a routing problem."
Future / conditional	Future work, operational consequence, or scenario	"If ridership exceeds 240, the school will require a fifth bus."

Do not change tense merely because a new sentence begins. Change tense when the time or rhetorical function changes.

Model criticism: Self-reference

"We" is normally the clearest way to identify a team choice. Avoid "the authors," "one," or "it was decided" when they hide agency. The reader should know who chose a threshold, removed a record, selected a method, or interpreted a result.

5 Build paragraph flow with given and new information

A paragraph feels coherent when each sentence begins from information already available and ends with the information that should receive emphasis. This is the **given–new principle**.

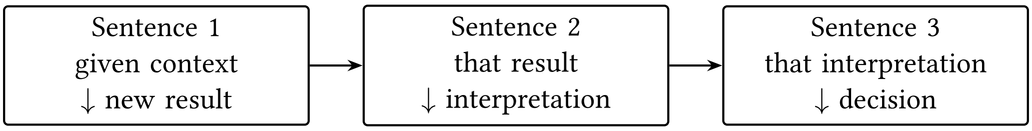


Figure 3: End a sentence with the information that the next sentence can treat as given. This creates a visible reference chain without excessive connectives.

Example 4: broken flow and repaired flow

Broken flow: "The model minimizes total distance. Capacity is 60. We used 20 scenarios. Route 3 is important. A fifth bus may be needed."
Repaired flow: "The model minimizes total distance subject to a capacity of 60 passengers per bus. This capacity becomes binding on route 3 in 16 of the 20 demand scenarios. Because route 3 absorbs most enrollment growth, the four-bus plan fails once total ridership exceeds 240. We therefore begin an early fleet review at 236 projected riders and add a fifth bus from 241 confirmed riders."

Method: create cohesion without a thesaurus

- Repeat the same technical term when it refers to the same object. Do not alternate among "model," "framework," "mechanism," and "scheme" merely to avoid repetition.
- Use a noun phrase rather than an unclear pronoun: "This threshold" is clearer than "This" when several previous facts are possible antecedents.
- Put contrast where the contrast occurs: "The analytical estimate is 15.8 minutes, whereas the simulation estimate is 16.1 minutes."
- Use connectives only when the logical relation is not already visible from sentence structure.
- Keep notation stable. A symbol change is a cohesion failure, not only a mathematical error.

Checkpoint: Checkpoint: repair the reference chain

Revise the following so that every "this," "it," and "they" has a single clear referent:
"The demand estimate was compared with the route plan, and it was inaccurate. This affected it, so they were changed. This shows the model should be improved."

6 Make equations and visuals part of the argument

An equation, table, or figure does not explain itself. The surrounding prose should tell the reader why it appears, what pattern matters, and how that pattern changes the modeling decision.

The four-step visual callout

When introducing a table or figure:
1. **Purpose**: state the question the visual answers.
2. **Reference**: direct the reader to the correct figure or table.
3. **Pattern**: identify the important comparison, trend, or threshold.
4. **Implication**: connect the pattern to the model or recommendation.

Do not write only "The results are shown in Figure 3."

Example 5: equation as a grammatical sentence

Weak:

$$D = \sum_{i,j} d_{ij} x_{ij}.$$

"This is our objective function."
Stronger: "We minimize total daily route length,

$$D = \sum_i \sum_j d_{ij} x_{ij}, \tag{1}$$

where d_{ij} is the distance from stop i to stop j and $x_{ij} = 1$ when that link is used. The objective in Eq. (1) penalizes unnecessary detours but does not by itself control ride-time equity; we impose that requirement separately as a constraint."
The equation is introduced, punctuated, defined, and interpreted.

Example 6: table commentary that adds value

Weak: "Table 4 shows the sensitivity results. It can be seen that demand changes the answer."
Stronger: "Table 4 compares the recommended fleet size across demand scenarios. The four-bus plan remains feasible up to 240 riders but fails at 245 because route 3 exceeds capacity. Demand therefore changes the decision discontinuously at a clear operational threshold rather than gradually across the entire range."

Method: report numbers so they can be checked

- Include units and a comparison baseline: "28 minutes, down from 43," not merely "28."
- State the denominator: "18 of 20 scenarios," not "90%" when the scenario count matters.
- Use precision supported by the inputs. A travel-time recommendation of 28.137846 minutes is false precision.
- Distinguish an estimate from a target or constraint: predicted wait, required wait, and maximum allowed wait are different quantities.
- Keep conditions next to the number: "MAE 2.4 minutes on 12 held-out days."
- Use "respectively" only when the paired order is unmistakable; two separate clauses are often clearer.

Model criticism: Avoid visual filler

Phrases such as "as can be clearly seen," "obviously," and "the figure below" rarely add information. Name the figure, state the pattern, and explain its consequence. If the prose repeats every number in a table, summarize only the comparison the reader should notice.

7 Revise for concision without removing reasoning

Concision means compressing language while preserving the claim, evidence, interpretation, and qualification. It does not mean deleting the evidence because the page limit is tight.

Six high-yield sentence repairs

Draft pattern	Repair
"In order to" / "due to the fact that"	Use "to" / "because."
"We conducted an analysis of"	Use a verb: "We analysed."
Repeated framing: "It should be noted that"	Delete it and state the fact.
Several actions joined by commas and "which"	Give each main action its own sentence or coordinate verbs directly.
Empty adjective: "robust," "effective," "significant"	Replace with a metric, comparison, or decision boundary.
Long noun chain: "route demand growth threshold analysis"	Rebuild with prepositions: "analysis of the demand threshold for route growth."

Example 7: sentence surgery

Draft: "In order to carry out the validation of the model, we conducted a comparison of the predicted travel time results with the actual observed travel time data, which were collected by us during twelve different school days."
Revision: "We validated predicted travel times against GPS observations from 12 school days."
The revision reduces 37 words to 12 while retaining the action, target, comparator, source, and sample size.

High-frequency transfer patterns for Chinese/Cantonese writers

These patterns are language-transfer tendencies, not defects in the writer. They are useful because they can be checked systematically.

1. **Topic-comment opening without a clear subject**: "For our model, the assumption is important." → "Our model relies on this assumption."
2. **Comma-chain sentence**: split multiple independent actions or use coordinated verbs.
3. **Repeated "which"**: replace a relative-clause chain with direct verbs or separate sentences.
4. **Generic connective**: "Besides" → "In addition," "However," or no connective, depending on the logic.
5. **Article and number mismatch**: check every singular countable noun and every plural technical term.
6. **Noun-heavy translation**: "make a determination of" → "determine."
7. **Unclear "this"**: use "this error," "this threshold," or "this comparison."
8. **Overuse of "can"**: distinguish mathematical possibility, empirical tendency, and recommendation.

Example 8: transfer-shaped paragraph to scientific paragraph

Draft: "For the route result, we can see that it has a relatively good performance, and besides, the average time is decreased, which can prove our method is effective, which is very meaningful for the school."

Revision: "The optimized routes reduce mean ride time from 43 to 28 minutes. This 15-minute reduction meets the school's 30-minute target, supporting use of the plan for the current demand level."

8 Integrated case: polish the school-bus argument

The same school-bus evidence used in W2 and W3 can now be turned into a coherent, publication-style passage.

Example 9: raw draft

"We use genetic algorithm to solve this bus problem. The algorithm gives four buses, which has 87 km route. It makes the time become 28 minutes from 43 minutes, so it is very good. In order to prove our model, we compare the data, and the error is only 2.4 minutes. Besides, after doing many sensitivity tests, we can know that the result is robust, but when the number is more than 240, it needs five buses. Therefore our model can effectively solve the school problem."

Studio: diagnose before rewriting

Mark at least eight issues in the raw draft. Include:

- one grammar issue;
- one unsupported or over-strong verb;
- one missing condition;
- one unclear reference;
- one evidence/interpretation gap;
- one tense or voice issue;
- one connective issue;
- one phrase that can be shortened.

Do not begin line editing until the paragraph's claim and evidence sequence is clear.

Example 10: competition-ready revision

We solve the capacitated school-bus routing problem using a genetic algorithm with route length as the primary objective and capacity and ride time as constraints. The best solution found uses four buses and 87 km of daily routes, reducing mean student ride time from 43 to 28 minutes. On 12 held-out school days, predicted route times have a mean absolute error of 2.4 minutes, which is below the school's 5-minute scheduling tolerance. The four-bus decision remains feasible in 18 of 20 demand scenarios; however, route 3 exceeds capacity when total ridership rises above 240. We therefore recommend four buses for current enrollment, an early review at 236 projected riders, and a fifth bus from 241 confirmed riders.

Sentence-role map for the revision

1. Method and model scope.
2. Headline result with baseline comparison.
3. Validation evidence plus intended-use interpretation.

4. Sensitivity boundary and failure mechanism.
5. Conditional recommendation and operational trigger.

The paragraph is concise because each sentence performs a different job, not because it omits evidence.

Differentiated writing studio

Core route. Rewrite a six-sentence results paragraph using the five roles above. Preserve every number and condition.
Challenge route. Produce two versions of the bus paragraph: one for the Summary and one for the Results section. Explain what detail moves to the body and why.
Extension route. Add a limitation: GPS observations cover normal traffic only. Revise the recommendation so it remains useful without overstating performance during road closures or severe congestion.

9 Use a four-pass revision sequence

Proofreading grammar before the argument is stable wastes time. Revision should move from large decisions to small ones.

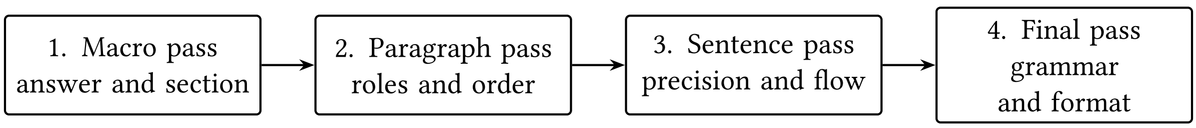


Figure 4: Revise from macro-logic to surface correctness. A polished sentence cannot repair a missing result or unsupported recommendation.

Method: the four passes

1. **Macro pass – coverage.** Does the section answer its assigned question? Are the headline result, evidence, limitation, and decision present?
2. **Paragraph pass – logic.** Does each paragraph have one controlling question and a visible claim–evidence–interpretation sequence?
3. **Sentence pass – precision.** Are subjects, verbs, referents, units, comparisons, conditions, and reporting verbs accurate?
4. **Final pass – surface and consistency.** Check grammar, articles, plurals, spelling, notation, equation/figure references, punctuation, capitalization, and formatting.

Never perform only Pass 4.

Model criticism: Tool-assisted editing

Grammar checkers and language models can propose clearer wording, but they must not invent evidence, alter numbers, strengthen claims, or change technical meaning. Use them to generate alternatives, then compare every revision against the model, data, and intended scope. Follow the disclosure requirements of the relevant competition.

Checkpoint: Exit ticket

Rewrite one paragraph from your current project and submit:

1. the original paragraph;
2. sentence-role labels;
3. the revised paragraph;
4. one claim you weakened or strengthened and the evidence-based reason;
5. one deleted phrase and why it was unnecessary.

10 Core practice and take-aways

Core practice 1: role repair

The following paragraph contains claims and evidence but no interpretation or decision link. Add two sentences without repeating the numbers.
"The baseline simulation gives a mean wait of 7.8 minutes. With a second server, mean wait falls to 2.6 minutes. The 95th-percentile waits are 21.4 and 6.9 minutes, respectively."

Core practice 2: claim calibration

Rewrite each claim so that it is as strong as the evidence permits.

1. Evidence: 18 of 20 scenarios support the same decision. Claim: "The decision is always robust."
2. Evidence: correlation $r = 0.81$ in one observational dataset. Claim: "Temperature causes demand."
3. Evidence: a convexity proof. Claim: "The numerical solver probably found a minimum."
4. Evidence: held-out MAE 2.4 minutes against a 5-minute tolerance. Claim: "The model is perfect."

Core practice 3: visual commentary

Write three sentences about a sensitivity table:

1. identify the parameter with the largest output effect;
2. identify the first parameter to change the decision;
3. state the measurement or operational action implied by these findings.

Do not repeat every cell of the table.

Modeling-writing challenge

Choose a 150–220 word passage from one completed modeling paper. Perform all four revision passes and provide a short audit table with these columns:

Problem found	Revision made	Evidence or principle used

The goal is not maximum word reduction. The goal is a passage in which every claim is traceable and every sentence has a job.

Class W4 take-aways

1. Academic prose guides evidence; it does not decorate weak reasoning.
2. Give every sentence a role and every paragraph one controlling question.
3. Match claim strength to evidence strength; prefer quantitative boundaries to vague hedges.
4. Choose active or passive voice according to information focus and responsibility.
5. Build flow through given–new order and stable technical terms, not excessive connectives.
6. Introduce equations and visuals, identify the pattern, and state its modeling implication.
7. Revise from macro-logic to paragraphs to sentences to surface correctness.
8. Plain, measurable, and traceable English is stronger than inflated formal language.

In **W5: Figures, Tables, and LaTeX Workflow**, we move from verbal evidence to visual evidence: how to design figures and tables that carry an argument, integrate them into a report, and complete a reliable submission workflow.

Appendix A: Phrase bank by rhetorical function

Function	**Useful sentence frames**
Define scope	"We restrict the analysis to... because..."; "The model applies when..."
Introduce method	"To estimate X, we..."; "We formulate the problem as..."
State a result	"The model gives..."; "The optimized plan uses..."
Compare	"Compared with the baseline,..."; "Whereas Model 1..., Model 2..."
Interpret	"This pattern indicates..."; "The difference matters because..."
Qualify	"Under the tested scenarios,..."; "Within the observed range,..."
Validate	"On held-out data,..."; "The estimate is consistent with..."
Report sensitivity	"The decision remains unchanged until..."; "The recommendation switches when..."
Recommend	"We therefore recommend... provided that..."; "If X exceeds..., then..."
Limit	"This conclusion does not cover..."; "The main limitation is..."

Appendix B: High-frequency replacements

Long or weak form	Preferred form
in order to	to
due to the fact that	because
conduct an analysis of	analyse
make a comparison between	compare
it should be noted that	delete, or "Note that" when essential
a large number of	many, or give the number
has the ability to	can
with regard to	regarding / about / on
according to the above results	the results / Table X / Eq. (Y)
we can clearly see that	the figure shows / the data indicate
very robust / highly accurate	state range, error, scenario count, or threshold
prove the real-world model	support / validate for the intended use
obtain the result of	obtain / estimate / compute
carry out the calculation of	calculate
make use of	use

Appendix C: Paragraph audit worksheet

For each paragraph, complete:

1. **Controlling question**: What single reader question does this paragraph answer?
2. **Claim**: What is the strongest assertion?
3. **Evidence**: Which number, equation, figure, source, or test supports it?
4. **Interpretation**: Why does the evidence matter?
5. **Scope**: Under which conditions does the claim apply?
6. **Implication**: What model choice, limitation, or decision follows?
7. **Flow**: Does each sentence begin from information the reader already has?
8. **Surface**: Are subjects, verbs, referents, units, and cross-references unambiguous?

Appendix D: Final language checklist

☐ Every paragraph has one controlling question.

☐ Every major claim is followed by evidence or a reference to evidence.

☐ Interpretation is not left entirely to the reader.

☐ Reporting verbs match evidence strength.

☐ Conditions and limitations sit near the claims they qualify.

☐ "This," "it," and "they" have clear referents.

☐ Equations are introduced, punctuated, and interpreted.

- ☐ Figure and table callouts identify the important pattern and implication.
- ☐ Numbers include units, baselines, denominators, and appropriate precision.
- ☐ Empty adjectives and throat-clearing phrases have been removed.
- ☐ Technical terms and symbols remain consistent.
- ☐ The final proofreading pass did not change mathematical meaning or evidence.

Model audit card

Audit before trusting or communicating a model

Purpose	What question or decision does the model support?	**Data**	Which inputs are observed, derived, or simulated?
Assumptions	Which simplifications may change the answer?	**Units**	Are dimensions and scales consistent?
Verification	Were equations, code, constraints, and limiting cases checked?	**Validation**	Is performance tested against independent evidence?
Uncertainty	How do measurement, parameter, solver, and structural errors affect conclusions?	**Ethics**	Who benefits, who bears risk, and what fairness criterion matters?
Reproducibility	Are data provenance, versions, seeds, and transformations recorded?	**Communication**	Are claims conditional, traceable, and decision-relevant?

Further reading

- Structure Your Article (IEEE Author Center)

Sources, provenance, and licence

These notes follow the iterative modeling and assessment perspective in the SIAM/COMAP GAIMME report. Competition-specific remarks are a snapshot checked on 20 August 2026 against the official COMAP HiMCM instructions; readers should verify current rules before submission. Unless a source is explicitly identified, classroom datasets and numerical scenarios are synthetic or simulated teaching examples and are not empirical claims about a real institution. Third-party references retain their own terms.

Student response guide and checking criteria

For closed calculations, show the definition, substitution, units, and at least one independent check; the worked examples in this unit are method hints, but you must complete the arithmetic yourself. Open modeling responses are checked on five dimensions: a clear purpose and output; stated data and provenance; defensible assumptions and scope; traceable computational verification and model validation; and alignment among uncertainty, limitations, and the recommendation. Support each dimension with concrete evidence rather than writing only "completed."

Writing Classes for Mathematical Modeling Competitions

W5: Visual Evidence, LaTeX, and Submission Engineering

Kenneth Sok Kin Cheng

Revised 20 August 2026

Learning goals

By the end of this two-hour writing class, students should be able to:

- treat equations, tables, figures, captions, and page layout as parts of an evidence argument rather than decoration;
- choose deliberately among prose, equation, table, plot, schematic, map, and algorithm display;
- create a visible page hierarchy that supports both rapid scanning and careful verification;
- typeset equations with definitions, punctuation, numbering, and cross-references that remain stable during revision;
- design tables with units, meaningful precision, uncertainty, and an explicit comparison target;
- select a plot type that matches the analytical question and avoid misleading axes, encodings, and scales;
- write captions that identify the display, explain its encoding, state the main finding, and connect it to the decision;
- distinguish a reproducible figure pipeline from late-stage manual formatting;
- organise a standalone, Overleaf-safe LaTeX project with predictable file names and no fragile external dependencies;
- construct a page budget and a contest-specific rule card before writing begins;
- perform content, visual, technical, anonymity, provenance, and submission preflight checks;
- package one modeling result into a competition-ready page containing prose, mathematics, a visual, and a decision statement.

1 A visual is an evidence claim, not a decoration

The reader never sees the team's calculations directly. The reader sees a sequence of selected objects on a page: sentences, equations, tables, plots, captions, and references. Each object should help the reader answer one of three questions:

1. What did the team model or compute?
2. What evidence supports the claim?
3. What decision follows from the evidence?

A polished page is therefore not merely attractive. It is an interface for verifying the argument.

Launch activity: same evidence, different page

Both drafts report the same result from an illustrative synthetic school-bus case: four buses, 87 km of daily routes, mean ride time reduced from 43 to 28 minutes, and a fifth bus required above 240 riders.
Draft A contains a dense paragraph, an unlabeled screenshot, and a table with columns named "A," "B," and "C."
Draft B opens with the headline result, displays a route-time comparison with units, labels the threshold at 240 riders, and uses a caption to explain why the threshold matters.
In pairs, list the evidence that is technically present in Draft A but practically inaccessible. Then identify the page features that make Draft B faster to inspect without changing the underlying mathematics.

Question what must be decided? → Evidence what was computed? → Display what form reveals it? → Claim what pattern matters? → Action what changes next?

Figure 1: The visual evidence chain. A display is successful only when its form reveals the evidence needed for a claim and the claim supports a decision.

Visual communication standards

A competition display should be:

1. **Necessary**: it answers a question more efficiently than prose alone.
2. **Interpretable**: labels, units, symbols, scales, and encodings are defined.
3. **Traceable**: the text refers to it and the display can be linked to data, equations, or code.
4. **Honest**: the scale and selection do not exaggerate or conceal differences.
5. **Decision-connected**: the caption or surrounding prose states why the pattern matters.

2 Choose the representation before formatting it

A common failure is to choose a display because it is visually impressive. The correct choice begins with the reader's task.

Representation by reader task

Reader task	Usually use	Reason
Understand one relationship	Equation plus prose	Exposes variables, assumptions, and mathematical structure.
Compare exact values	Table	Preserves numbers, units, rankings, and thresholds.
See a trend or shape	Line/scatter plot	Reveals direction, curvature, clusters, and departures.
Compare categories	Ordered dot/bar plot	Supports magnitude comparison from a common baseline.
Understand a process	Flowchart or state diagram	Shows sequence, dependencies, states, or information flow.
Understand geography/network	Map or network diagram	Preserves location, connectivity, and route structure.
Inspect uncertainty	Interval plot, distribution, fan chart	Shows range, variation, or probability rather than a single estimate.
Follow an algorithm	Pseudocode or numbered method	Makes logic auditable without exposing implementation detail.

Method: the representation test

Before creating a visual, complete four sentences:

1. The reader needs to compare / locate / understand / decide ____________________.
2. The evidence consists of ______________________________.
3. A _______________ is better than prose because _________________________.
4. After seeing it, the reader should conclude ______________________________.

If sentence 4 cannot be completed, the display has no defined analytical purpose.

Example 1: table or figure?

Suppose five routing methods are compared using total distance, maximum ride time, computation time, and number of constraint violations.

- Use a **table** when the reader must inspect all four exact metrics and identify whether each method satisfies a threshold.
- Use a **plot** when the argument concerns one trade-off, such as total distance versus maximum ride time.
- Use both only when they perform different jobs: the plot reveals the trade-off; the compact table records the selected method's exact values.

Repeating the same data twice merely consumes the page budget.

Checkpoint: Checkpoint: choose, do not decorate

Choose one representation and justify it.

1. Four models, each with MAE, RMSE, and training time.
2. A queue length recorded every minute for four hours.

3. A six-stage simulation procedure.
4. Route connections among 18 school stops.
5. A recommendation that changes when demand exceeds 240 riders.

More than one answer may be defensible, but the reader task must be explicit.

3 Design a page for scanning and verification

Competition readers alternate between two modes. In **scan mode**, they locate the headline answer, section purpose, key visual, and caption. In **verification mode**, they inspect definitions, equations, numerical details, and limitations. A strong page supports both.

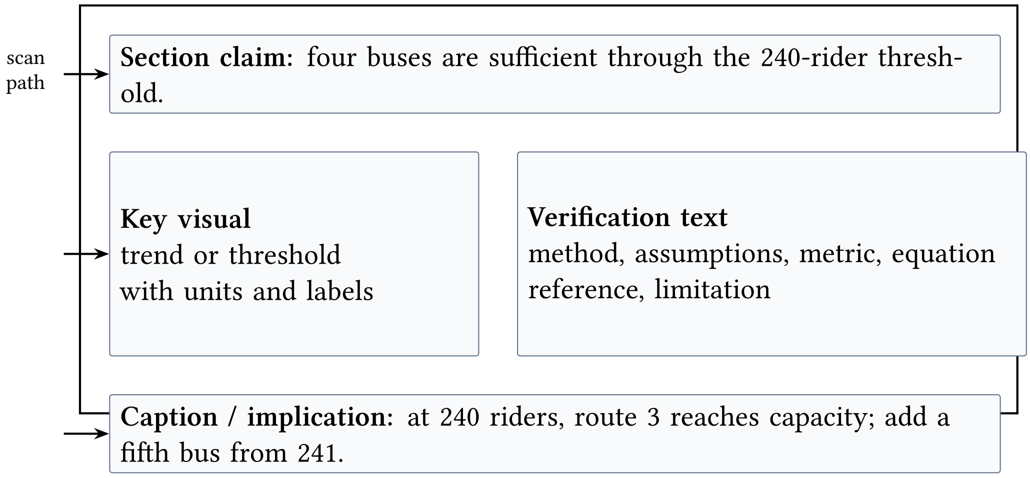


Figure 2: A page hierarchy that supports two reading modes. The section claim and caption carry the scan path; the display and explanatory text support verification.

Method: page-level hierarchy

Use hierarchy through meaning, not through decoration.

1. Give each section a claim-oriented title when possible: "Demand uncertainty changes the fleet decision" is more informative than "Sensitivity Analysis."
2. Place the key display near the paragraph that introduces it; avoid orphan figures several pages later.
3. Use bold text for a small number of decision-critical results, not for entire paragraphs.
4. Keep heading levels consistent. A paper rarely needs more than three levels.
5. Preserve white space around sections and displays; do not fill every centimetre with text.
6. Keep the same visual grammar throughout: font, line weight, caption style, decimal precision, and notation.

Model criticism: Do not confuse page density with mathematical depth

Shrinking margins, font size, captions, or plots does not create a stronger solution. It increases search cost and may violate the rules. Page limits should be managed by removing duplication, compressing derivations, and prioritising decision evidence – not by making the paper physically difficult to read.

4 Equations must function as sentences

An equation is not self-explanatory. The prose should state why it is introduced, define its symbols, and explain what feature of the real system it represents.

Method: purpose–equation–definition–interpretation

A reliable equation paragraph has four parts:

1. **Purpose**: what quantity or mechanism is being modeled?
2. **Equation**: display the relationship and number it if later referenced.
3. **Definition**: define all new variables, parameters, indices, and units near first use.
4. **Interpretation**: state what the equation captures, and when useful, what it omits.

Example 2: a readable model equation

To evaluate a routing plan, we combine total distance with a penalty for capacity violations:

$$J(\mathcal{R}) = \sum_{r\in\mathcal{R}} d_r + \alpha \sum_{r\in\mathcal{R}} \max\{0, n_r - C\}, \tag{1}$$

where $\mathcal{R}$ is the set of bus routes, d_r is route distance in kilometres, n_r is assigned ridership, $C = 60$ is bus capacity, and α is a penalty measured in kilometres per excess passenger. The first term rewards short routes; the second makes capacity violations costly. Because α is a modeling choice rather than a measured physical constant, we test its influence in the sensitivity analysis.

Equation and notation rules

- Number only equations that will be referenced or form part of a derivation.
- Use stable labels such as `eq:route-objective`, not manual numbers.
- Refer to equation (1), not "the equation above."
- Do not reuse a symbol for different concepts. If r denotes a route index, do not later use r as a growth rate.
- State units when a quantity first appears and preserve dimensional consistency.
- Punctuate displayed equations as parts of sentences.
- Move long algebraic manipulations to an appendix when the body only needs the setup and the result.

Listing 1: A stable equation and cross-reference pattern.

```
To penalize infeasible routes, we define
\begin{equation}
  J(\mathcal{R}) = D(\mathcal{R}) + \alpha V(\mathcal{R}),
  \label{eq:objective}
\end{equation}
where $D$ is total distance and $V$ is the number of
capacity violations. As shown in \cref{eq:objective}, ...
```

Checkpoint: Checkpoint: repair the equation paragraph

"We use the following formula: $Q = ab/c$. Here are the results."
Revise it so that the reader knows the purpose of Q, the meaning and units of a, b, c, the origin of the form, and the implication of the computed value.

5 Tables should expose comparisons and thresholds

A table is strongest when exact numbers matter. The design should make the intended comparison visible without forcing the reader to decode formatting.

Method: table design checklist

1. State the comparison question in the caption.
2. Put units in headers and define the evaluation set or scenario.
3. Align numbers consistently; use meaningful rather than excessive decimal places.
4. Include a baseline, target, tolerance, or benchmark when a decision depends on comparison.
5. Display uncertainty when the values are estimates, not exact constants.
6. Order rows by analytical meaning: chronology, rank, model complexity, or influence.
7. Use notes for exceptions, not unexplained symbols inside cells.
8. Do not reproduce every raw output. Select the evidence needed for the claim.

Example 3: a decision table

Table 1: Fleet requirement across ridership scenarios. The four-bus plan remains feasible up to 240 riders; the recommendation switches to five buses above that threshold.

Ridership	Buses required	Maximum load	Decision interpretation
220	4	54	Four buses feasible with six-seat margin.
230	4	57	Feasible, but limited reserve capacity.
240	4	60	Capacity boundary; monitor enrollment.
250	5	51	Fifth bus required to restore capacity margin.
260	5	54	Five-bus plan remains feasible.

The table supports a conditional recommendation rather than the vague statement that "four buses are robust."

Precision is part of the claim

Report precision that the data and method can support.

- A commute estimate of 28.37 min suggests false precision if input travel times were recorded to the nearest minute.
- A relative error of 2.1% is more interpretable when the comparator and sample are stated.
- A mean without a spread can hide instability. Consider a range, standard deviation, confidence interval, or scenario count.
- A percentage requires a denominator: "12% lower than the baseline four-route plan," not merely "12% lower."

Model criticism: Table failure modes

Avoid vertical rules, merged-cell mazes, tiny fonts, unexplained abbreviations, units repeated in every cell, inconsistent decimal places, and bolding every row. These features make the table look busy without improving the comparison.

6 Figures should answer one analytical question

Figures are efficient for patterns, relationships, uncertainty, and spatial structure. They become misleading when the encoding is selected after seeing which form makes the result look most impressive.

Plot selection by analytical question

Question	Preferred display	What to show
How does a quantity change over time?	Line plot	Time units, intervention markers, uncertainty if estimated.
Are two variables associated?	Scatter plot	Individual observations, fitted relation, residual/uncertainty information.
Which category is larger?	Ordered dot/bar plot	Common zero baseline when magnitude is the claim.
Where is a decision threshold?	Line/step plot with threshold	Boundary value, feasible and infeasible regions, direction of recommendation.
How variable is an output?	Histogram, density, interval, box plot	Sample size, centre, spread, tails, and relevant target.
How do two parameters interact?	Contour/heatmap	Legend, scale, feasible region, selected operating point.
How is a system connected?	Network or state diagram	Node/edge meaning, direction, weight, and selected path/state.

Example 4: an honest threshold figure

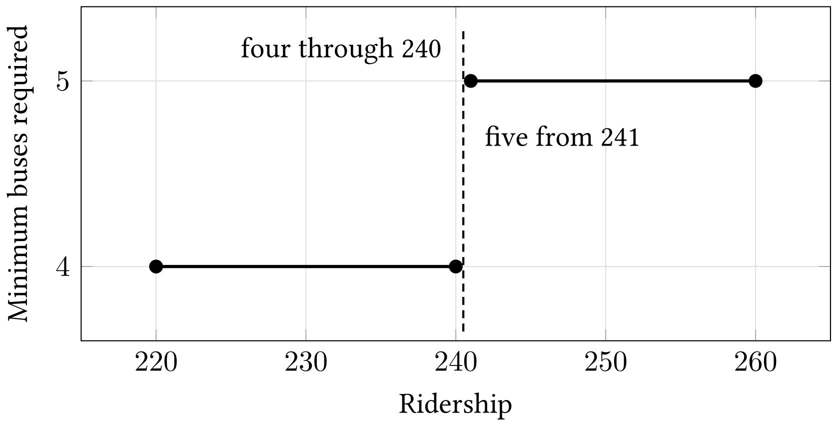


Figure 3: Minimum fleet size across tested demand scenarios. Four buses satisfy capacity through 240 riders; the requirement increases to five above the boundary. The figure therefore supports an enrollment-trigger recommendation rather than a single unconditional fleet size.

Method: figure integrity audit

Ask six questions before export:

1. Are axes, units, categories, symbols, and sample sizes defined?
2. Does the scale reveal the pattern without exaggerating it?
3. Is uncertainty shown when the line or point is estimated?
4. Can curves or categories still be distinguished in greyscale?
5. Is the important comparison visually direct, or must the reader subtract values mentally?
6. Does the caption state the finding and decision implication?

Model criticism: Frequent misleading choices

- Truncating a bar-chart axis to magnify a small difference.
- Using a dual y-axis that creates an apparent correlation through arbitrary scaling.
- Connecting unordered observations with a line.
- Showing only a fitted curve and hiding the underlying observations.
- Using a pie chart for many similar categories that require precise comparison.
- Using a rainbow scale whose order is unclear in greyscale.
- Presenting one simulation path as if it were the distribution of possible outcomes.

Checkpoint: Checkpoint: caption surgery

Revise “Figure 5. Results of the simulation.” A complete caption should answer:

1. What is displayed and under which scenario?
2. What do lines, points, shading, or panels encode?
3. What is the main numerical or structural finding?

4. What model or decision implication follows?
5. What condition or limitation should the reader remember?

7 Build a reproducible document, not a fragile final file

The document should be regenerable from source. Manual copying, unnamed screenshots, and late-stage editing inside the PDF destroy traceability and create inconsistencies.

Method: a reliable project structure

```
project/
|-- main.tex
|-- references.bib
|-- sections/
|   |-- 01_introduction.tex
|   |-- 02_model.tex
|   |-- 03_results.tex
|   ‘-- 04_validation.tex
|-- figures/
|   |-- fleet_threshold.pdf
|   ‘-- route_map.pdf
|-- data/
|   ‘-- processed_routes.csv
|-- code/
|   |-- build_tables.py
|   ‘-- make_figures.py
‘-- submission/
    ‘-- CONTROLNUMBER.pdf
```

Each final table or figure should have a known source. Keep raw data separate from processed data, and keep the submission PDF separate from working files.

A reproducible figure pipeline

1. Store the cleaned data or documented inputs.
2. Generate the figure with a named script or a stable TikZ block.
3. Export to vector PDF for line art whenever possible; use high-resolution raster only for inherently raster images.
4. Insert the exported file at a controlled width.
5. Record the script, parameters, and date or commit that generated it.
6. Regenerate all displays after any numerical change; do not edit the labels manually afterward.

Listing 2: A short, stable figure insertion pattern.

```
\begin{figure}[tbp]
  \centering
  \includegraphics[width=0.78\textwidth]{figures/fleet_threshold.pdf}
  \caption{Minimum fleet size across tested demand scenarios. ...}
  \label{fig:fleet-threshold}
\end{figure}
```

```
As shown in \cref{fig:fleet-threshold}, the decision changes ...
```

Model criticism: TikZ and competition reliability

TikZ is excellent for small mathematical diagrams whose text must match the paper. It is not automatically the best tool for every map, network, or data plot. Complex TikZ can slow compilation and is difficult to repair under deadline pressure. Prefer:

- simple fixed-node diagrams in TikZ;
- data plots generated from code and exported as PDF;
- no labels placed by trial-and-error over curves when a legend or external annotation is clearer;
- a rendered-page inspection after every substantial visual change.

A figure that compiles is not necessarily a figure that communicates.

Cross-references and bibliography

- Label every referenced equation, table, figure, and section.
- Use `\cref` or equivalent commands instead of typed numbers.
- Cite a source at the claim it supports; do not place a list of unrelated citations at paragraph end.
- Record authorship, title, year, publisher/journal, DOI or stable source, and access date when appropriate.
- Keep one citation style throughout. Consistency matters more than choosing between APA and IEEE.
- Check that every citation appears in the bibliography and every bibliography entry is cited.

Provenance and AI-use disclosure

A submission should make the origin of data, code, images, external text, and AI-assisted work clear. Current contest rules can differ, and the team must read the official instructions for the exact year and round. For example, the 2026 HiMCM instructions require teams that use LLMs or other AI tools to identify the model and purpose, cite the use in the report and references, and append a separate AI-use report. The report is not a substitute for checking every generated statement, citation, calculation, and code result.

8 Submission engineering begins before the competition

Submission failure is often a process failure rather than a mathematical failure. Teams should convert the official rules into a one-page **rule card** before creating the template.

Method: build a contest rule card

Record only information taken from the official instructions for the current year and round:

- paper size, font minimum, margin minimum, language, and anonymity requirements;
- required first pages, letters, graphics, or summary forms;
- page limit and whether contents, references, appendices, and AI reports count;
- required header/footer, team number, and page numbering;

- accepted file type, size limit, file name, and number of uploads;
- work-stop time, submission deadline, time zone, and who may submit;
- AI-use, data, code, and external-resource disclosure rules.

Do not copy a rule card from a different competition or year. In 2026 HiMCM, for example, the Summary Sheet plus solution, references, and appendices are limited to 25 pages, whereas the required AI-use report is appended outside that limit. IM^{2}C requirements may be round- and problem-specific, so the official problem statement must also be checked.

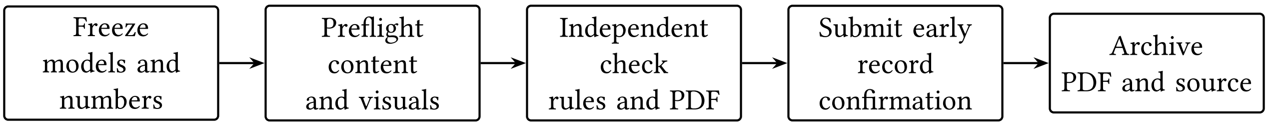


Figure 4: Submission engineering is a controlled pipeline. New analysis should not be introduced after the freeze unless it repairs a critical error.

Method: six-part preflight

1. **Content:** every task answered; headline results agree across summary, body, tables, and conclusion.
2. **Mathematics:** symbols defined; units consistent; equations, code outputs, and tables cross-checked.
3. **Visual:** captions complete; labels readable at page size; no overlap, clipping, or unexplained encodings.
4. **Technical:** clean compile; no unresolved references; correct page count; PDF opens on another device.
5. **Compliance:** anonymity, header, file name, AI disclosure, and current rule card satisfied.
6. **Submission:** correct portal and file; upload confirmed before the deadline; confirmation saved.

Model criticism: Final-hour anti-patterns

Do not introduce a new model, change the notation globally without recompiling, paste figures from chat screenshots, rebuild the bibliography manually, compress the PDF without checking glyphs, or wait for the final minutes to upload. A team should submit the best verified version, not the newest unverified version.

9 Integrated studio: build one competition-ready results page

Use the continuing school-bus evidence package from W2–W4:

- baseline mean ride time: 43.00 min; optimized mean: 28.00 min;
- four buses and 87.00 km of daily routes;
- four buses feasible through 240 riders; five required above 240;
- held-out route-time MAE: 2.40 min;
- route 3 is the binding route near the threshold.

Studio task: page blueprint

Create a one-page blueprint containing:
1. a claim-oriented subsection heading;
2. a two- to four-sentence results paragraph using the W4 paragraph arc;
3. one equation or compact method statement explaining the fleet decision;
4. one display: either table 1 or figure 3;
5. a caption that states the threshold and decision implication;
6. one limitation or condition;
7. one cross-reference to validation or sensitivity evidence.

The result should be understandable in scan mode and defensible in verification mode.

Example 5: model page text

Four buses remain sufficient through 240 riders, with an early review at 236. The optimized plan uses four buses and 87.00 km of daily routes, reducing mean ride time from 43.00 min to 28.00 min. Capacity tests show that the plan remains feasible through 240 riders but requires a fifth bus above this level (figure 3). Because route 3 reaches 60 passengers at the boundary, we recommend four buses for current demand and an early fleet review at 236 projected riders and a mandatory fifth bus from 241 riders.

The threshold is supported by capacity constraints rather than by the route-time prediction alone. Held-out route-time MAE is 2.40 min, but clustered changes in stop-level demand could move the boundary; the school should therefore update the assignment after major enrollment or residential-pattern changes.

Checkpoint: Peer audit: 90 seconds

Exchange page blueprints. Without asking the writer, record:
1. the recommendation;
2. the numerical evidence;
3. the decision threshold;
4. the main limitation;
5. where the reader would verify the method.

Anything your partner cannot locate in 90 seconds requires redesign, not merely proofreading.

10 Differentiated writing and production studio

Core route: repair a weak visual package

You are given a screenshot of an unlabeled line plot and a table with no units. Rebuild the package by supplying:
1. an analytical question;
2. a suitable display type;

3. labels and units;
4. a complete caption;
5. a two-sentence interpretation;
6. the source or script needed to reproduce it.

Challenge route: choose between competing displays

For a sensitivity analysis with five parameters and a binary recommendation, design two alternatives:

1. a tornado-style ranking of local effects;
2. a threshold/scenario table showing when the recommendation switches.

Explain which belongs in the main body and which, if either, belongs in an appendix. Your decision must refer to the paper's claim, not visual preference.

Extension route: design a reproducible production pipeline

Create a project plan that turns raw data into the final PDF. Specify:

- input files and validation checks;
- scripts that create tables and figures;
- naming conventions and output folders;
- how numerical updates propagate to the Summary and Conclusion;
- versioning, backup, and final-archive procedure;
- who performs the independent PDF and rules check.

Test the pipeline by changing one input value and regenerating every affected output.

Exercise W5.1: caption laboratory

Write a complete caption for each item:

1. observed and predicted commute times for 12 held-out days;
2. a sensitivity heatmap whose recommendation changes near $\lambda/\mu = 1$;
3. a route map with three candidate corridors and one selected route;
4. a convergence plot comparing Euler, Heun, and RK4;
5. a residual plot showing increasing variance at large fitted values.

Each caption must identify the display, state the main finding, and give the modeling implication.

Exercise W5.2: page-budget audit

Take a previous practice paper and classify every half-page as one of:

- task completion;
- model development;
- evidence/results;
- validation/sensitivity;

- recommendation/communication;
- duplication or low-priority detail.

Remove or move at least one page of duplication without shrinking the font or margins. Record what evidence, if any, was lost.

Exercise W5.3: mock submission

Run a 45-minute drill. Starting from a working project:

1. change one result and regenerate the affected table and figure;
2. update the Summary and Conclusion;
3. compile and inspect every page;
4. apply the rule card and preflight checklist;
5. rename the final PDF using a mock control number;
6. open it on a second device and place it in the submission folder;
7. record the time required for each stage.

Any stage that cannot be completed reliably becomes a training target.

11 Closing the writing sequence

W1–W5 evidence architecture

1. **W1 – Architecture:** help the reader locate the problem, model, evidence, and conclusion.
2. **W2 – Summary:** compress the task, strategy, findings, trust evidence, and decision.
3. **W3 – Trust:** expose assumptions and test the model through verification, validation, sensitivity, and uncertainty.
4. **W4 – Prose:** turn claims and evidence into calibrated, coherent paragraphs.
5. **W5 – Visual production:** present the evidence in inspectable displays and deliver a compliant, reproducible PDF.

The classes are not separate cosmetic layers. Together they form a single evidence system.

Class W5 take-aways

1. Begin every display with a reader question and end it with a decision implication.
2. Use tables for exact comparison and figures for pattern, relationship, uncertainty, process, or space.
3. Design for scan mode and verification mode on the same page.
4. Introduce equations through purpose, definition, and interpretation.
5. Precision, units, baselines, uncertainty, and thresholds are part of the claim.
6. A caption is a miniature evidence paragraph, not a label.
7. Reproducibility and cross-references reduce final-hour errors.

8. Contest rules change; create the rule card from current official instructions.
9. Compile, render, inspect, and independently verify the final PDF.
10. Submit the best verified version early, then archive both PDF and source.

Appendix A: Standalone competition-paper starter

```
\documentclass[12pt,a4paper]{article}
\usepackage[margin=2.5cm]{geometry} % replace only if rules require
\usepackage{iftex}
\ifPDFTeX\errmessage{Compile with XeLaTeX}\fi
\usepackage{fontspec,xeCJK}
\setmainfont{TeX Gyre Termes}
\setCJKmainfont{Noto Serif CJK TC}
\usepackage{amsmath,amssymb,graphicx,booktabs,array}
\usepackage{caption,subcaption,float,hyperref,cleveref}
\usepackage{siunitx}
\hypersetup{colorlinks=false,pdfborder={0 0 0}}

\begin{document}
% Insert the official current-year Summary Sheet or required first page.
% Do not include names or institution if anonymity is required.

\section{Introduction and Task Interpretation}
\section{Assumptions and Notation}
\section{Model Development}
\section{Results and Recommendation}
\section{Verification and Validation}
\section{Sensitivity and Limitations}
\section{Conclusion}

% Bibliography and appendices must follow the current contest rules.
\end{document}
```

Appendix B: Caption and display checklist

- ☐ The display answers one identifiable analytical question.
- ☐ Every axis, unit, category, symbol, panel, line, and shading convention is defined.
- ☐ The scale and data selection are honest.
- ☐ Estimated quantities show appropriate uncertainty or scenario scope.
- ☐ Text remains readable at final page size and in greyscale.
- ☐ The body refers to the display by number and states why it is introduced.
- ☐ The caption identifies the object and scenario.
- ☐ The caption states the main finding, not merely the axes.
- ☐ The caption links the finding to a model choice or decision.
- ☐ The display can be regenerated from a known source.

Appendix C: Final PDF and submission checklist

- ☐ Every task in the problem statement is answered and easy to locate.
- ☐ Summary, main text, tables, figures, and conclusion use the same final numbers.
- ☐ Units and notation are consistent; important equations are checked independently.
- ☐ No unresolved citations or cross-references remain.
- ☐ No text, table, figure, caption, or label is clipped or overlapping.
- ☐ Fonts are embedded and Chinese/English glyphs render correctly.
- ☐ Page count and all current-year rules are satisfied.
- ☐ Team number, page numbers, anonymity, and required first pages are correct.
- ☐ References, data sources, code, figures, and AI use are disclosed as required.
- ☐ File type, file size, and file name match the official instructions.
- ☐ The PDF opens and renders correctly on a second device.
- ☐ A non-author has completed the rule-card and visual inspection.
- ☐ The correct file has been submitted early and confirmation has been saved.
- ☐ Final PDF, source, data, code, and confirmation are archived together.

Appendix D: Current-rule source note

The 2026 HiMCM examples in this class are based on COMAP's official HiMCM/MidMCM instructions. The official 2026 IM^{2}C international-round page likewise states that teams must also follow the current problem-specific requirements. Teams in any competition should consult both the current rules and the current problem statement because required document components can differ by round and problem. Rules should be rechecked immediately before each competition.

References

[1] COMAP, "2026 HiMCM/MidMCM Instructions," checked 20 August 2026.

[2] IM^{2}C, "Challenge Overview," checked 20 August 2026, `https://immchallenge.org/immc-challenge/`.

Model audit card

Audit before trusting or communicating a model

Purpose	What question or decision does the model support?	**Data**	Which inputs are observed, derived, or simulated?
Assumptions	Which simplifications may change the answer?	**Units**	Are dimensions and scales consistent?
Verification	Were equations, code, constraints, and limiting cases checked?	**Validation**	Is performance tested against independent evidence?
Uncertainty	How do measurement, parameter, solver, and structural errors affect conclusions?	**Ethics**	Who benefits, who bears risk, and what fairness criterion matters?
Reproducibility	Are data provenance, versions, seeds, and transformations recorded?	**Communication**	Are claims conditional, traceable, and decision-relevant?

Further reading

- HiMCM / MidMCM Contest Instructions (COMAP)
- IM²C challenge overview and modeling expectations (International Mathematical Modeling Challenge)
- Create Graphics for Your Article (IEEE Author Center)
- LaTeX documentation (The LaTeX Project)

Sources, provenance, and licence

These notes follow the iterative modeling and assessment perspective in the SIAM/COMAP GAIMME report. Competition-specific remarks are a snapshot checked on 20 August 2026 against the official COMAP HiMCM instructions; readers should verify current rules before submission. Unless a source is explicitly identified, classroom datasets and numerical scenarios are synthetic or simulated teaching examples and are not empirical claims about a real institution. Third-party references retain their own terms.

Student response guide and checking criteria

For closed calculations, show the definition, substitution, units, and at least one independent check; the worked examples in this unit are method hints, but you must complete the arithmetic yourself. Open modeling responses are checked on five dimensions: a clear purpose and output; stated data and provenance; defensible assumptions and scope; traceable computational verification and model validation; and alignment among uncertainty, limitations, and the recommendation. Support each dimension with concrete evidence rather than writing only "completed."