Control engineering

Model a plant, design and test a feedback loop in time and frequency, then design estimators and controllers that handle sampling, actuator limits and interacting channels.

Read the subject in depth

Each volume develops its arguments in sequence, with worked problems and links to experiments.

- ### System dynamics, feedback and estimation

  Start from a physical plant and its measured signals. Derive a feedback or estimation model and test which dynamical behaviour can be controlled or inferred.

Advanced study across the subject

Each sequence connects the course to detailed chapters in neighbouring subjects. The complete course PDF includes their derivations, examples and exercises.

1. ### From a moving plant to identified dynamics

   Derive a plant from force and energy, then compare its modes with measured shifts and decay. Examine the uncertainty that remains when two parameters produce similar responses.

2. ### Time stepping and robust computation

   Investigate state growth in apparently stable systems, propagate derivatives and judge whether numerical error can account for a switching event.

How the course fits together

1. **Chapters 1–4**

   Identify a dynamic plant from measurements. Connect input, error, controller and plant to obtain a transfer function. Its poles and frequency response predict settling and the phase lag of the output.

2. **Chapters 5–8**

   Move from a measured output to a state description and sample the loop in time. Trace poles as gain changes and ask how sensitive the proposed controller is to an uncertain plant.

3. **Chapters 9–12**

   Choose feedback using an explicit cost and estimate hidden states from noisy data. Several channels can interact, and actuator constraints make a finite-horizon control problem depend on the current state.

Contents

1. Chapter 1 · From a balance law to a measured plant
2. Chapter 2 · Transfer functions and connections between systems
3. Problem sheet 1 · Modelling a plant and interpreting its signals
4. Chapter 3 · Closed-loop poles, damping and performance
5. Chapter 4 · Frequency response, disturbance rejection and robustness
6. Problem sheet 2 · Pole placement, gain and phase
7. Chapter 5 · States, reachability and estimation
8. Chapter 6 · Sampling, saturation and commissioning
9. Problem sheet 3 · What a controller can learn and deliver
10. Chapter 7 · Following poles as a controller gain changes
11. Chapter 8 · Uncertainty envelopes and robust stability
12. Problem sheet 4 · Pole geometry and plant uncertainty
13. Chapter 9 · Optimal state feedback and the Riccati equation
14. Chapter 10 · Estimating hidden states from noisy measurements
15. Problem sheet 5 · Regulate a state and estimate it from data
16. Chapter 11 · Interacting actuators, pairing and decoupling
17. Chapter 12 · Finite-horizon control with real actuator constraints
18. Problem sheet 6 · Coupled plants and constrained predictions
19. Chapter 13 · Proving a region of attraction with the energy of a pendulum
20. Chapter 14 · Why integral action rejects an unknown constant load
21. Chapter 15 · Separating reference tracking from disturbance rejection
22. Chapter 16 · How much transport delay a feedback loop can tolerate
23. Chapter 17 · Why a stable plant can initially move in the wrong direction
24. Chapter 18 · Designing a lead network at a specified gain crossover
25. Chapter 19 · The period and duty cycle of a thermostat with hysteresis
26. Chapter 20 · Learn a repeatable positioning move from its previous error
27. Problem sheet 7 · A speed regulator with an unknown load and a measured delay
28. Problem sheet 8 · A saturated PI loop remembers an impossible set point
29. System dynamics, feedback and estimation
30. From a moving plant to identified dynamics
31. Time stepping and robust computation
32. Applied study
33. Notation and assumptions
34. Further reading

Chapter 1 · From a balance law to a measured plant

Predict a transient from conservation of energy, then use measurements to decide whether a first-order model is credible.

FROM FIRST PRINCIPLES

Turn a thermal energy balance into a step response

A well-stirred body of heat capacity C stores the energy increment $C\,dT$ for a small temperature increase dT. Assume the ambient temperature T_a is constant, the heater supplies power u, and heat loss is proportional to temperature excess with coefficient $H > 0$. Conservation of energy then gives a differential equation, with every term measured in watts.

$$C\dot{T} = u - H(T - T_a).$$

Hold the power at a constant value after a step. The new equilibrium requires zero stored-energy change, hence $T_\infty = T_a + u/H$. Subtracting that constant solution leaves an equation proportional to the deviation itself, which integrates exponentially.

$$\frac{d(T - T_\infty)}{dt} = -\frac{H}{C}(T - T_\infty),$$

$$T(t) = T_\infty + [T(0) - T_\infty]e^{-t/\tau},$$

$$\tau = C/H.$$

The steady temperature rise per watt identifies $1/H$. The time required for the residual deviation to fall to $1/e$ identifies τ. Together these observations determine the stored heat capacity as $C = H\tau$ under the assumed single-temperature model.

DERIVATION AND APPLICATIONS

Identify a first-order time constant without knowing the final level

A thermometer in a stirred bath often stops recording before the new equilibrium is reached. Three equally spaced readings can still determine that equilibrium if a single relaxation dominates. Assume a fixed step in heat input, constant heat capacity and a linear heat-loss coefficient over the observed temperature range. Take the first observation after the input has finished changing. If the sampling interval is h, the unknown remaining difference from equilibrium is multiplied by the same factor a over every interval.

$$T_j = T_\infty + (T_0 - T_\infty)a^j,$$

$$a = e^{-h/\tau}.$$

Subtract successive readings. The unknown equilibrium cancels and the ratio of consecutive increments measures a. The inversion uses differences of readings, so temperature offsets cancel as well. Once a is known, rearrange the one-step update to recover equilibrium in the original temperature scale.

$$a = \frac{T_2 - T_1}{T_1 - T_0},$$

$$\tau = -\frac{h}{\ln a},$$

$$T_\infty = \frac{T_1 - aT_0}{1 - a}.$$

A fourth reading tests the model against data excluded from the estimate. It can be predicted directly from the last increment. The condition for a stable monotonic first-order response is that a lie strictly between zero and one. Ratios near unity make the extrapolated final level very sensitive to measurement error.

$$T_{3,\mathrm{pred}} = T_2 + a(T_2 - T_1).$$

WORKED EXAMPLE

Recover thermal capacity from a short record

A heater adds 6.00 W to a bath initially at equilibrium. At 20.0 s intervals the thermometer reads 20.0, 24.0 and 26.0 °C. The increments give $a = 2/4 = 0.500$, so $\tau = 28.9$ s and $T_\infty = 28.0$ °C. The heat-loss coefficient is $6.00/8.00 = 0.750$ W/K. Multiplying by the time constant gives a lumped heat capacity of 21.6 J/K. The next predicted reading is 27.0 °C.

Change the inputs and inspect this calculation below

SECOND APPLICATION

Test whether one relaxation fits a cooling record

An independent 5.00 s interval record reads 90.0, 80.0, 74.0 and 68.0 °C. The first three values give $a = 0.600$ and $\tau = 9.79$ s. They predict a fourth value of $74.0 - 0.600(6.00) = 70.4$ °C. The residual is -2.4 K. With sensor repeatability of 0.1 K this discrepancy calls for checking the assumption of constant ambient temperature or a single thermal state.

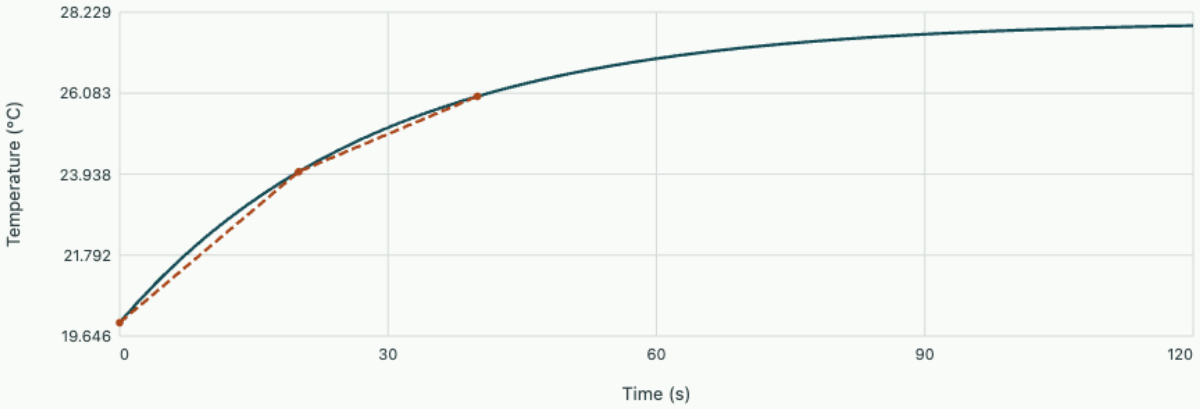
Interactive calculation · Change the inputs of this example

Use the adjustable quantities to test the two worked calculations above. The plot recomputes this derivation's balance or constitutive relation; its numerical readout reports the quantities needed for the forward calculation and the inverse or design question. Keep the fixed assumptions stated in the readout when comparing with the example values.

Open this live calculation

Try the calculation

Second reading T1 (°C)	24
Increment ratio a	0.5
Sampling interval (s)	20



τ=28.854 s; limiting temperature 28.000°C; loss coefficient 0.750 W/K and inferred capacity 21.640 J/K for a6 W step. A fourth observation tests the assumed model.

Start from this chapter. Read the time constant and gain of the first-order plant, then predict how the fitted closed-loop time scale changes when proportional gain increases.

OneSimulation laboratory · Predict the settling time and offset of a feedback loop

Choose variables that describe stored energy

A heater warms a well-stirred tank. A sensor measures only the liquid temperature, and a controller will eventually change the heating power. Before designing the controller, ask what information fixes the future of the plant when the power is given. If the tank, its contents and the thermometer can be represented by one uniform temperature, a single initial temperature suffices. A slow thermometer or a

heavy tank wall may require its own initial condition. This is a physical way of deciding how many states to put in a model.

Let C in J/K be the combined heat capacity, H in W/K the thermal conductance to the room, T_a the ambient temperature and u the delivered heater power. Over an interval dt the heater adds $u\,dt$, the room receives $H(T - T_a)\,dt$, and storage increases by $C\,dT$. Taking the limit of a short interval gives

$$C\dot{T} = u - H(T - T_a).$$

The signs can be checked without solving anything. Set $u = 0$ while $T > T_a$. Stored energy must decrease. At equilibrium $\dot{T} = 0$, so a constant 20 W input and $H = 5$ W/K sustain a 4 K excess over the surroundings. This check will later reveal a faulty actuator calibration even if the transient looks smooth.

Separate gain, time scale and initial condition

Choose an operating point (u_0, T_0) in equilibrium, keep the ambient temperature fixed and write $v = u - u_0$, $y = T - T_0$. Subtracting the equilibrium balance cancels the constant terms. Dividing by H gives

$$\tau \dot{y} + y = K v,$$

where $K = 1/H$ is the steady gain and $\tau = C/H$ is a time. An input step of amplitude v_s moves the equilibrium to $K v_s$. Subtract this new equilibrium to leave a homogeneous decay equation. Integrating from zero time yields

$$y(t) = K v_s + \big[y(0) - K v_s\big] e^{-t/\tau}.$$

After one time constant, 36.8% of the initial distance to the new value remains. Reaching a specified fraction of the final temperature and reaching a tolerance relative to the final change are equivalent only if the output is initially zero in the chosen coordinates. Working in temperature deviations makes the convention explicit.

Background · How can three samples reveal a time constant without knowing the final level?

For a stable, noise-free first-order response after a single input step, take three measurements y_0, y_1, y_2 equally spaced by h. Each difference contains the same unknown amplitude. Their ratio isolates the exponential decay.

$$\rho = \frac{y_2 - y_1}{y_1 - y_0} = e^{-h/\tau},$$

$$\tau = -\frac{h}{\ln \rho}.$$

With one second spacing and readings 2.00, 3.20 and 3.92 V, the ratio is 0.60 and the inferred time constant is 1.96 s. A constant sensor offset cancels. Near the final level both differences become small, so quantisation may dominate the ratio. In real data, evaluate multiple triples or fit all samples with a noise model.

Recover a model from a step experiment

Start from equilibrium and increase the power by 30 W. The final rise is 6 K. At 20 s the rise is 2.36 K. The final measurement identifies the gain and heat-loss conductance.

$$K = \frac{6\ \mathrm{K}}{30\ \mathrm{W}} = 0.20\ \mathrm{K/W},$$
$$H = \frac{1}{K} = 5.0\ \mathrm{W/K}.$$

Next divide the distance still to go at 20 s by the initial distance to the new equilibrium. The logarithm of this dimensionless fraction gives the time constant, which in turn gives the effective heat capacity.

$$r = \frac{6 - 2.36}{6} = 0.607,$$
$$\tau = -\frac{20\ \mathrm{s}}{\ln r} \simeq 40\ \mathrm{s},$$
$$C = H\tau \simeq 200\ \mathrm{J/K}.$$

At 60 s the predicted rise is

$$6(1 - e^{-1.5}) = 4.66\,\mathrm{K}.$$

This third measurement tests a prediction made using the earlier data.

The same experiment can expose a missing state. Plot the logarithm of the remaining fraction against time. A single time constant gives a straight line with slope $-1/\tau$. If the early trace has a shoulder but the late trace is straight, the sensor or a second thermal mass might be responsible. In the last few

samples, the temperature difference in the logarithm may be comparable to measurement noise. Use those samples to estimate the final temperature rather than relying on their individual log slopes.

Transfer the argument to a rotating shaft

With shaft inertia J, viscous drag b and driving torque u, conservation of angular momentum gives

$$J\dot{\omega} = u - b\omega.$$

A suspended flywheel with $J = 0.12$ kg·m² and $b = 0.030$ N·m·s/rad has a 4.0 s time constant. A 0.30 N·m step from rest eventually gives 10 rad/s. The equilibrium gain and transient have the same mathematical form as the thermal problem. Their units and the physical meaning of a disturbance differ. A load torque enters the shaft equation directly. A change in ambient temperature enters the tank balance through H. This distinction matters once the loop must reject disturbances.

Linearise a gravity-drained tank around an operating level

A model identified with one small power step can be accurate locally and fail after the operating point shifts. A tank with a gravity outlet illustrates this without invoking an unknown internal mechanism. Let its horizontal cross-sectional area A be constant and inlet flow q be directly controlled. Approximate outlet flow by $c\sqrt{h}$, where h is liquid level above the outlet datum and c includes effective opening area and gravity. With incompressible liquid, mass conservation reduces to

$$A\dot{h} = q - c\sqrt{h}.$$

Both sides of this equation have units of volume per time.

At an equilibrium $h_0 > 0$ the inlet is $q_0 = c\sqrt{h_0}$. Set $h = h_0 + \eta$ and $q = q_0 + u$, retaining only terms linear in the deviations. Since the tangent of $c\sqrt{h}$ at h_0 is $c/(2\sqrt{h_0})$, the constant terms cancel and the incremental model is

$$A\dot{\eta} + k_h\eta = u,$$

$$k_h = \frac{c}{2\sqrt{h_0}},$$

$$K = 1/k_h,$$

$$\tau = A/k_h.$$

The gain K converts inlet-flow change into equilibrium level change, while the time constant contains the additional storage area. Increasing level makes the outlet characteristic flatter, increasing both gain and time constant. At an empty tank the square-root derivative is singular; the small-signal model around a positive level cannot be continued through that point.

Compare a small step with a finite change

Let $A = 0.20\,\mathrm{m}^2$, $h_0 = 0.25\,\mathrm{m}$, and $q_0 = 0.0010\,\mathrm{m}^3/\mathrm{s}$. Then $c = 0.0020\,\mathrm{m}^{5/2}/\mathrm{s}$, $k_h = 0.0020\,\mathrm{m}^2/\mathrm{s}$, $K = 500\,\mathrm{s}/\mathrm{m}^2$ and $\tau = 100\,\mathrm{s}$. Increasing inlet flow by $0.00010\,\mathrm{m}^3/\mathrm{s}$ predicts a 0.050 m final rise in the linear model. The exact steady level is $(0.00110/0.0020)^2 = 0.3025\,\mathrm{m}$, a rise of 0.0525 m. Initially both models predict $\dot{h} = 0.00010/0.20 = 0.00050\,\mathrm{m}/\mathrm{s}$, since the level has not yet changed. Immediate slope agreement therefore does not validate the predicted final gain.

Repeat the thought experiment for a 50% inlet increase. The linear result is 0.50 m absolute level, whereas the nonlinear equilibrium is 0.5625 m. The difference now matters for a level limit. An informative commissioning test uses small upward and downward steps of equal amplitude at more than one operating point. The tangent model predicts symmetric final increments; the measured curvature contributes a positive quadratic term to both absolute levels. Comparing that asymmetry with sensor uncertainty can distinguish a genuine level-dependent gain from a drifting measurement zero.

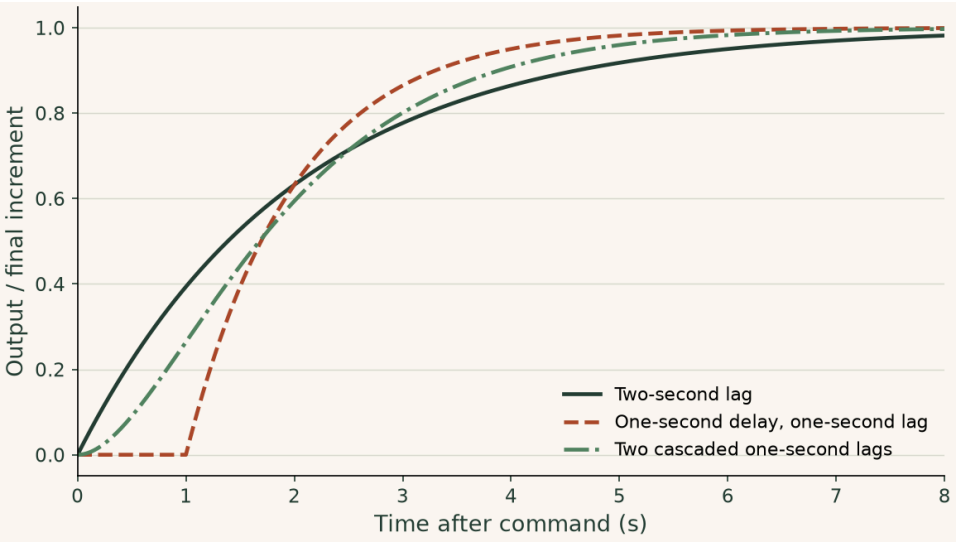
One can reconstruct the physical parameters independently from two parts of a measured step record. If a flow step of $0.00010\,\mathrm{m}^3/\mathrm{s}$ produces an initial level slope of $0.00050\,\mathrm{m}/\mathrm{s}$, the immediate mass balance gives $A = \Delta q/\dot{h}(0) = 0.20\,\mathrm{m}^2$, without knowing the outlet law. A sufficiently small steady rise of 0.050 m would separately give

$$k_h = \Delta q/\Delta h = 0.0020\,\mathrm{m}^2/\mathrm{s}.$$

Their quotient predicts the 100 s time scale. A sensor mounted behind a narrow side passage could delay the observed initial rise, making this area estimate too large without changing the final level. Comparing a direct sight-glass record with the remote sensor helps locate that additional lag before adding another state to the tank model.

Distinguish delay from a slow first-order response

Three devices have the same DC gain and the same area under their step-error curve, two seconds, but represent different mechanisms. A transport delay contributes a dead interval; two identical first-order states give the factor $(1 + t)e^{-t}$ in the step error.



The delayed model initially remains at zero. The cascade rises smoothly with zero initial slope; the single lag begins with nonzero slope. A single recorded settling time is insufficient to distinguish these plant models.

▶ **Read values from the plot**

Interactive calculation · change the inputs of this example

Allocate more of the total error-area time to pure delay. Compare the dead interval with the initial slopes of the lag models.

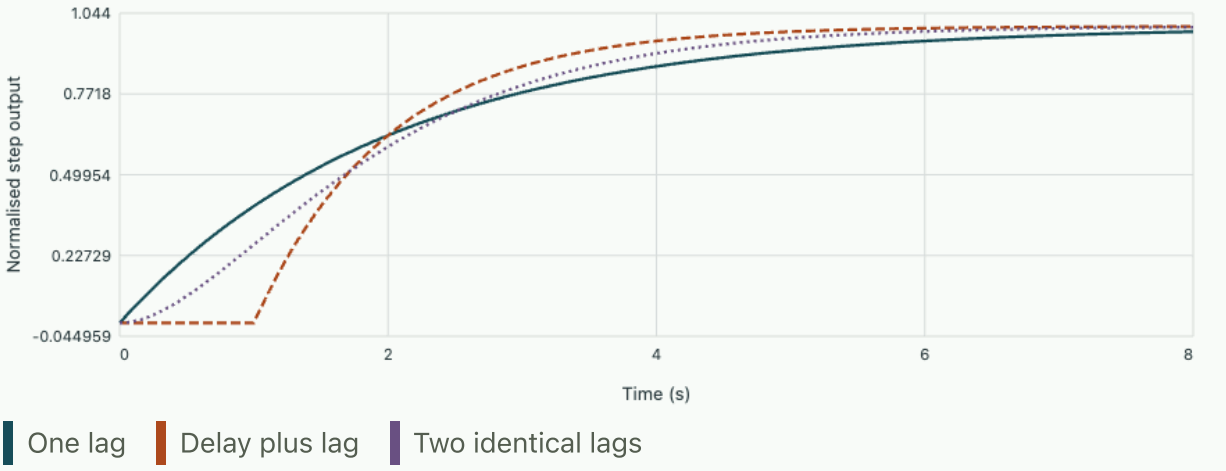
The curves below are recalculated from this section's equations in your browser. The fixed reference plot above remains available for comparison.

Open this live calculation

Try the calculation

Total error-area time (s) **2**

Delay fraction **0.5**



All three have unit final gain and error area 2.000 s. Their initial slope and dead interval remain different.

Study the full derivation

Estimating velocity from position with a model-based observer · System dynamics, feedback and estimation

An optical encoder returns position and the actuator drive supplies a commanded acceleration. Reconstruct an unmeasured velocity without differentiating successive noisy samples. Derive the observer, explain what a pair of chosen poles guarantees, and predict the tradeoff visible when the encoder carries periodic interference.

Chapter 2 · Transfer functions and connections between systems

Build a model of a sensor-actuator chain from individual balances and determine which responses a measurement can reveal.

FROM FIRST PRINCIPLES

Eliminate the feedback signal to find a closed-loop transfer

A linear transfer function relates an input perturbation to an output perturbation with zero initial deviations. For example, a first-order balance $\tau\dot{y} + y = Ku$ transforms to $(\tau s + 1)Y = KU$ because the time derivative transforms to sY under that initial condition. Thus the plant transfer P is

$$P(s) = \frac{Y(s)}{U(s)} = \frac{K}{1 + \tau s}.$$

Add a controller C acting on negative feedback from a sensor with transfer H. The error compares reference R with measured output HY. Write each intermediate relation before eliminating it; feedback makes the command depend on the very output being solved for.

$$E = R - HY,$$
$$U = CE,$$
$$Y = PC(R - HY).$$

Collect the terms containing Y and divide by the reference. The common denominator is the characteristic factor of the connected loop.

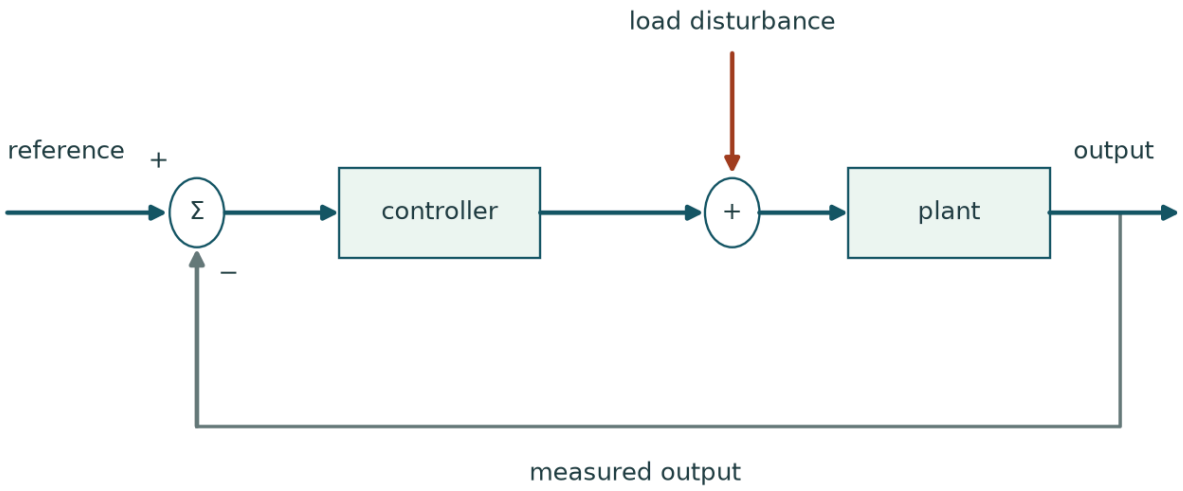
$$\frac{Y}{R} = \frac{PC}{1 + PCH}.$$

This signal algebra assumes the subsystems' transfers remain valid after connection. In an actual mechanical or hydraulic interconnection, loading may change the plant equations themselves and must be included before forming the blocks.

Background · Transfer function

$$G(s) = Y(s)/U(s)$$

State the input, output and zero initial condition. Interconnection signs and unmodelled dynamics affect what a measured closed-loop response represents.



The point at which a disturbance enters matters

Use u = C(r − y) and y = P(u + d). Substitution gives (1 + PC)y = PCr + Pd. The reference and the load share the characteristic denominator. Their numerators differ by the controller because only the reference travels through it before reaching the plant.

DERIVATION AND APPLICATIONS

Eliminate the states of a motor with electrical back reaction

A voltage-controlled motor resembles a circuit followed by a mechanical stage until the rotor is allowed to generate a back voltage. Consider a rigid rotor with inertia J, viscous drag b, winding resistance R and inductance L. With constant torque and back-emf coefficients k_t, k_e, use positive armature current for positive torque. Ignore load torque for the voltage-to-speed transfer and start with zero perturbations about a steady operating point. The transformed balances are

$$(Ls + R)I = V - k_e\Omega, \qquad (Js + b)\Omega = k_t I.$$

Multiply the second balance by the electrical impedance and substitute the first. The back voltage is proportional to the output and opposes the applied voltage. It therefore contributes a positive constant to the denominator.

$$\frac{\Omega(s)}{V(s)} = \frac{k_t}{LJs^2 + (Lb + RJ)s + Rb + k_t k_e}.$$

If the electrical time L/R is short relative to the mechanical transient of interest, current approximately follows its instantaneous resistive balance. The resulting speed gain and time constant both reflect extra damping supplied by current flowing in response to a speed change.

$$K_\omega = \frac{k_t}{Rb + k_t k_e},$$

$$\tau_\omega = \frac{J}{b + k_t k_e / R}.$$

The reduced model has one state fewer and retains the exact DC speed gain. It predicts an instantaneous current change at a voltage step, so estimates of the initial current derivative still require the full electrical state.

WORKED EXAMPLE

Predict the unloaded speed and rise time

Take $R = 2.00$ Ω, $L = 10.0$ mH, $J = 0.0100$ kg m², $b = 0.0200$ N m s/rad and $k_t = k_e = 0.200$ in consistent SI units. The feedback adds $0.200^2/2 = 0.0200$ N m s/rad. The mechanical time becomes 0.250 s against an electrical time of 0.00500 s. The gain is 2.50 rad/s/V, so a 4.00 V step produces a final 10.0 rad/s.

Change the inputs and inspect this calculation below

SECOND APPLICATION

Infer viscous drag from a DC test

A second motor has $R = 4.00$ Ω and $k_t = k_e = 0.100$ in SI units. A steady 6.00 V produces 30.0 rad/s with no external load. Back voltage consumes 3.00 V, leaving 0.750 A of current and 0.0750 N m torque. Balancing viscous drag gives $b = 0.0750/30.0 = 0.00250$ N m s/rad. The exact transfer expression then gives $K_\omega = 5.00$ rad/s/V, consistent with the measured voltage and speed.

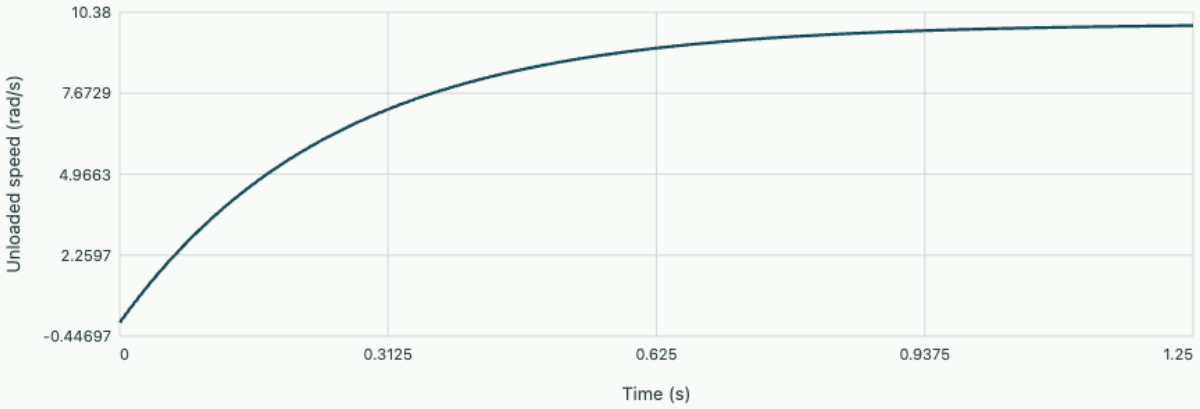
Interactive calculation · Change the inputs of this example

Use the adjustable quantities to test the two worked calculations above. The plot recomputes this derivation's balance or constitutive relation; its numerical readout reports the quantities needed for the forward calculation and the inverse or design question. Keep the fixed assumptions stated in the readout when comparing with the example values.

Open this live calculation

Try the calculation

Viscous drag (N m s/rad)	**0.02**
Rotor inertia (kg m²)	**0.01**
Voltage step (V)	**4**



Slow mechanical approximation

DC speed 10.000 rad/s; mechanical time 0.250 s versus electrical time0.005 s. R=2Ω, L=10 mH, kt=ke=0.2; the fast electrical state is neglected in this curve.

Start from this chapter. Eliminate the error and actuator signals algebraically. Compare the static gain of the resulting closed loop with the final value sampled by the OneSimulation signal graph.

OneSimulation laboratory · Predict the settling time and offset of a feedback loop

Eliminate a state in the transform domain

Chapter 1 gave a differential equation and an initial condition for a single energy store. Cascading several components leads to simultaneous equations. The Laplace transform changes the time derivative into $sY(s) - y(0)$. Applied to $\tau\dot{y} + y = Kv$, this gives

$$(\tau s + 1)Y = KV + \tau y(0).$$

With zero initial deviation the ratio $G(s) = Y/V = K/(\tau s + 1)$ describes the forced response. For nonzero initial state the extra term predicts free decay. This is why an experimental transfer function should identify a particular operating point and initialisation procedure.

The pole at $s = -1/\tau$ locates the decaying exponential. The DC gain is $G(0) = K$. The high-frequency magnitude of this model tends to zero because the energy store cannot change its state instantaneously in response to bounded power. A transfer function can thus be assessed from dimensions, a static balance and a short-time thought experiment before any curve fitting.

Connect the sensor and plant

A sensor may have its own lag. Let the actual plant output be y and the indicated value z. With plant $G_p = 1/(4s + 1)$ and sensor $G_s = 1/(s + 1)$, the individual relations $Y = G_p U$ and $Z = G_s Y$ give

$$Z/U = 1/[(4s + 1)(s + 1)].$$

For a unit step divide by s; partial fractions or substitution in the differential equation gives

$$z(t) = 1 - \frac{4}{3}e^{-t/4} + \frac{1}{3}e^{-t}.$$

Check $z(0) = 0$, $z(\infty) = 1$ and $\dot{z}(0) = 0$. The actual plant slope begins at $1/4$ per second, but the indicated output starts with a horizontal tangent. At 4 s the plant has reached 0.632 and the indicator only 0.516. A designer who assigns the entire delay to the plant may demand an unnecessarily powerful actuator. A designer who ignores the sensor will underestimate the phase lag in feedback.

Close a loop algebraically

Let $C(s)$ be a controller and use the indicated output in negative feedback, so $U = C(R - G_s Y)$. Substitute this in $Y = G_p U$ and collect all terms multiplying Y. The result is

$$Y/R = G_p C/(1 + G_p C G_s).$$

The same denominator occurs in other responses, but the numerator depends on where the signal enters. A disturbance D added at the plant input gives $Y/D = G_p/(1 + G_p C G_s)$ with zero reference. Measurement noise added to the feedback signal gives

$$Y/N = -G_p C/(1 + G_p C G_s).$$

Forcing every disturbance to enter at the reference would predict the wrong amplitudes.

Recognise an invisible mode

A rational expression may simplify by cancellation. Consider a cascade of an integrator-like unstable stage $1/(s-1)$ and a downstream stage $(s-1)/(s+2)$. The algebraic transfer from external input to final output is $1/(s+2)$ at zero initial conditions. The state between stages can still grow as e^{t}. A small mismatch in either factor also leaves an unstable component in the observed output. Physically realising two factors gives internal variables that still need checking even when their product has a well-behaved step response.

For a stable example, the measured lag above can be estimated with a first-order fit from late observations alone. The resulting single pole describes the slow tail but loses the zero initial slope. Choosing the level of model detail therefore requires stating the quantity of interest: a long-time energy estimate and a rapid feedback correction need different bandwidths. The next chapter uses the true characteristic polynomial to determine stability before assessing accuracy of tracking.

Derive two interacting liquid levels before drawing blocks

It is tempting to represent every pair of connected tanks as a cascade of two independent first-order blocks. If the pipe between the tanks is submerged at both ends, the downstream level applies a back pressure to the upstream tank. Consider small deviations h_1, h_2 from equilibrium, with positive forward intertank flow $q_{12} = (h_1 - h_2)/R_1$ and outlet flow $q_o = h_2/R_2$. The incremental hydraulic resistances measure level difference per flow and have units s/m². They can represent local tangent slopes of nonlinear valve characteristics. Take constant tank areas A_1, A_2 and a perturbation u of inlet flow.

The two balances are $A_1\dot{h}_1 = u - q_{12}$ and $A_2\dot{h}_2 = q_{12} - q_o$. For zero initial deviations, write $H_1 = (1 + R_1/R_2 + A_2R_1 s)H_2$ from the second balance and substitute it into the first. Collecting equal powers of s leads to

$$\frac{H_2}{U} = \frac{R_2}{A_1A_2R_1R_2s^2 + [A_1(R_1+R_2) + A_2R_2]s + 1}.$$

The static gain equals R_2, since at equilibrium all three volume rates are equal and $h_2 = R_2u$. The coefficient of s contains an interaction term absent if tank 1 drains freely to atmosphere. It describes the return action of the downstream level on intertank flow. An individual transfer block measured with downstream level fixed does not capture this additional path unless it is reconnected explicitly.

Distinguish a loaded connection from a cascade in a step test

Use $A_1 = A_2 = 0.10\,\mathrm{m}^2$ and $R_1 = R_2 = 100\,\mathrm{s/m}^2$. Each isolated area–resistance product is 10 s. The interacting transfer function is $100/(100s^2 + 30s + 1)$. Its decay rates are $(-30 \pm \sqrt{900 - 400})/200 = -0.0382$ and $-0.2618\,\mathrm{s}^{-1}$, giving time constants 26.18 and 3.82 s. Two independent identical drains would instead give $100/(10s+1)^2$, with repeated time constant 10 s. Both have the same final gain and initial zero slope at the downstream tank.

For a sustained flow increase of $0.00050\,\mathrm{m^3/s}$, both models predict a 0.050 m final rise. Using the two interacting time constants, the normalised step value at 20 s is $1 - [26.18e^{-20/26.18} - 3.82e^{-20/3.82}]/(26.18 - 3.82) \simeq 0.455$, giving 22.8 mm. The independent cascade predicts $1 - (1 + 2)e^{-2} = 0.594$, or 29.7 mm. Recording intermediate levels therefore distinguishes two plant structures that an equilibrium calibration cannot separate. As R_1 approaches zero the two levels equalise rapidly; the limiting first-order time constant becomes $(A_1 + A_2)R_2$, consistent with a single tank of combined storage area.

There is also a check that uses both level sensors without differentiating noisy data. Add the original tank balances and integrate over the first 20 s:

$$A_1[h_1(20) - h_1(0)] + A_2[h_2(20) - h_2(0)] = \int_0^{20} (u - q_o)\,dt.$$

The intertank flow disappears identically. In this experiment the added inlet volume is 10.0 litres, so added liquid stored plus additional outlet delivery must equal that quantity. If the two measured storage changes already exceed 10 litres, no passive choice of intertank resistance can explain the data. Conversely, outlet and storage agreement validates the external balance but leaves the internal connection undetermined. The intermediate time reading and knowledge of whether the interconnecting pipe is submerged provide the extra evidence needed for choosing the internal transfer model.

Problem sheet 1 · Modelling a plant and interpreting its signals

Use balances, local linearisation and block diagrams to decide which parameters a commissioning test actually measures.

1. A step on a heated plate

A plate initially in equilibrium at 22 °C receives a further 40 W. Its final temperature is 32 °C and it reaches 28.32 °C after 50 s. (a) Estimate the incremental static gain and time constant. (b) Recover heat capacity and heat-loss conductance in a one-capacity model. (c) Describe a second experiment that could expose a temperature-dependent loss.

▼ **Recover the physical coefficients**

Write $C\dot{\theta} + G\theta = \Delta P$ about the initial equilibrium. The 10 K final increment gives $K = 0.25$ K/W and $G = 4$ W/K. The recorded increment is 63.2% of the final value, giving $\tau = 50$ s and $C = \tau G = 200$ J/K. Repeat a small equal step about a higher initial temperature with the same airflow. A change in incremental gain indicates a changed local loss slope; a second slow exponential could instead indicate another stored-energy state. One measurement at 50 s cannot distinguish these hypotheses.

---

2. A tank drained by gravity

An open tank of cross-section $A = 0.50$ m² discharges $q_{\text{out}} = c\sqrt{H}$ with $c = 0.020\ \text{m}^{5/2}/\text{s}$ and level H in metres. (a) Find the inflow at a 1 m operating level. (b) Derive the incremental transfer function and predict the rise for a sustained additional 0.002 m³/s. (c) Check the new equilibrium exactly and identify the sign of the approximation error.

▼ **Linearise the outflow at the operating level**

The balance is

$$A\dot{H} = q - c\sqrt{H}.$$

At $H_0 = 1$ m, $q_0 = 0.020$ m³/s. The slope of the outflow is 0.010 m²/s, so $\delta H/\delta q = 1/(0.50s + 0.010)$ in the Laplace domain. Its time constant is 50 s and its final incremental prediction is 0.20 m. The exact new level is

$$(0.022/0.020)^2 = 1.21\,\text{m}.$$

The local model underestimates the rise by 0.01 m because the outflow curve becomes flatter as the level rises.

---

3. A slow sensor on a fast drive

The transfer functions from torque to speed and from speed to the recorded signal are $P = 2/(1 + 0.5s)$ and $H = 1/(1 + 2s)$. (a) Determine the transfer from torque to the recording and its DC gain. (b) For a unit step derive the recording in time. (c) Would the dominant time visible in this record be a good plant-only identification?

▼ **Distinguish dynamics by tracing the signal**

The recording has transfer $PH = 2/[(1 + 0.5s)(1 + 2s)]$ and DC gain 2. Partial fractions for a unit step give

$$y_m = 2[1 - \tfrac{4}{3}e^{-t/2} + \tfrac{1}{3}e^{-2t}].$$

Its value and slope are zero at switch-on, whereas physical speed has initial slope 4 in the chosen units. The late transient is dominated by the sensor's 2 s time constant. An independently calibrated sensor model or a faster sensor separates its effect from the 0.5 s plant response.

Explore the governing equations after solving the sheet: From a balance law to a measured plant; Transfer functions and connections between systems. Compare the assumptions and units before transferring your chosen parameters; the controls may cover a different numerical range from the written exercise.

Chapter 3 · Closed-loop poles, damping and performance

Translate a time-domain requirement into acceptable poles and find when additional gain destabilises a higher-order plant.

FROM FIRST PRINCIPLES

Connect characteristic roots to decay and oscillation

Near an equilibrium, many closed-loop positioning stages reduce to a constant-coefficient second-order equation. Consider the free response after the reference has become constant. Let x be deviation from its final value, $\omega_n > 0$ the natural frequency and ζ the dimensionless damping ratio. Try an exponential response because time derivatives of an exponential reproduce the same waveform.

$$\ddot{x} + 2\zeta\omega_n\dot{x} + \omega_n^2 x = 0,$$
$$x = e^{st} \implies s^2 + 2\zeta\omega_n s + \omega_n^2 = 0.$$

For $0 < \zeta < 1$ solving the quadratic produces conjugate roots. Their real part sets the exponential decay rate and their imaginary part sets the observed angular oscillation frequency.

$$s_{1,2} = -\zeta\omega_n \pm i\omega_n\sqrt{1-\zeta^2},$$
$$x = e^{-\zeta\omega_n t}(A\cos\omega_d t + B\sin\omega_d t).$$

Here $\omega_d = \omega_n\sqrt{1-\zeta^2}$ and the initial displacement and velocity set A, B. A measured period gives ω_d while the decaying peak envelope gives $\zeta\omega_n$. Stable poles describe decay of free modes. The numerator of a forced response also influences how prominently each mode appears.

DERIVATION AND APPLICATIONS

Separate a cubic stability boundary from oscillation frequency

Feedback gain can move an oscillatory pair across the imaginary axis even with every characteristic coefficient positive. Take a continuous linear unity-feedback loop whose open-loop transfer is $K/[s(s+a)(s+b)]$ with positive a, b, K. The integral mode can represent a position obtained from velocity; the other poles can represent actuator and measurement dynamics. Add unity to the loop transfer and clear its denominator to obtain the closed-loop polynomial.

$$p(s) = s^3 + (a+b)s^2 + ab\,s + K.$$

A marginal oscillation has $s = i\omega$ with real nonzero ω. Separate the real and imaginary parts. The imaginary equation sets the frequency without requiring the gain, while the real equation selects the gain that supports it.

$$K - (a+b)\omega^2 = 0,$$
$$\omega(ab - \omega^2) = 0,$$
$$K_{\text{crit}} = ab(a+b).$$

At this gain the cubic factors exactly into a real stable pole and a pair of purely imaginary roots. The cubic Routh first column determines which side of the boundary is stable. Its first two entries are positive, making the sign of the third entry decisive for positive gain.

$$p_{\text{crit}}(s) = (s + a + b)(s^2 + ab), \qquad \text{first column: } 1,\ a+b,\ ab - \frac{K}{a+b},\ K.$$

Below the threshold all entries have the same sign. Above it there are two sign changes and therefore two right-half-plane roots. The model treats actuator authority as unlimited and omits nonlinear friction; a finite observed oscillation amplitude requires additional physics beyond the pole crossing.

WORKED EXAMPLE

Predict the onset period of a positioning loop

Set $a = 2.00$ s⁻¹ and $b = 8.00$ s⁻¹. Marginal gain is $K_{\text{crit}} = 160$ s⁻³. The oscillation has frequency 4.00 rad/s, hence period $2\pi/4.00 = 1.57$ s. At $K = 120$ s⁻³ the third entry is $16 - 12 = 4.00$, giving stability. At $K = 180$ s⁻³ it is -2.00 and the pair grows.

Change the inputs and inspect this calculation below

SECOND APPLICATION

Recover the two lag rates from a marginal test

Another loop reaches marginality at $K = 60.0$ s⁻³ with a measured oscillation of $\sqrt{10}$ rad/s. The rates have product $ab = 10.0$ s⁻² and sum $K/(ab) = 6.00$ s⁻¹. These observations cannot describe two positive real lag rates because the maximum product at fixed sum is $6^2/4 = 9.00$ s⁻². The failed consistency check flags model structure or calibration before attributing a unique time constant to either stage.

Interactive calculation · Change the inputs of this example

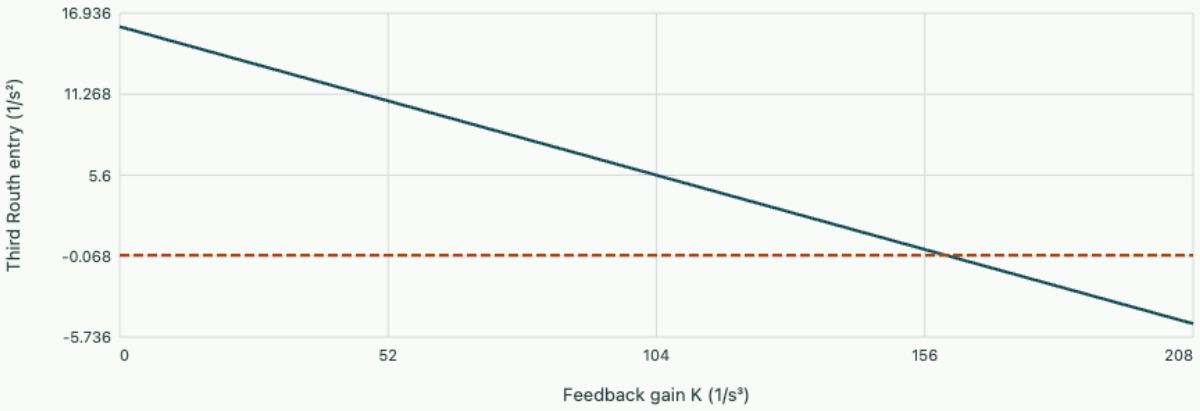
Use the adjustable quantities to test the two worked calculations above. The plot recomputes this derivation's balance or constitutive relation; its numerical readout reports the quantities needed for the forward calculation and the inverse or design question. Keep the fixed assumptions stated in the readout when comparing with the example values.

Open this live calculation

Try the calculation

First lag rate a (1/s) **2**

Second lag rate b (1/s) **8**



Critical gain 160.000 /s³; onset period 1.571 s. The third pole at onset is −10.000 /s.

Start from this chapter. Locate the real closed-loop pole for two controller gains. Compare its inverse magnitude with the time constant fitted to the numerical step response.

OneSimulation laboratory · Predict the settling time and offset of a feedback loop

Relate polynomial coefficients to observed motion

For a normalised second-order system, the free response satisfies $\ddot{y} + 2\zeta\omega_n\dot{y} + \omega_n^2 y = 0$. Substitution of e^{st} gives

$$s = -\zeta\omega_n \pm \omega_n\sqrt{\zeta^2 - 1}.$$

The real part sets the decay rate. For $0 < \zeta < 1$, the imaginary part $\omega_d = \omega_n\sqrt{1-\zeta^2}$ sets the time between successive extrema. Doubling ω_n at fixed ζ halves all characteristic times but leaves the fraction of overshoot unchanged.

Assume a unit DC gain, a constant numerator and zero initial displacement and velocity. The step response is

$$y = 1 - e^{-\zeta\omega_n t}[\cos(\omega_d t) + (\zeta\omega_n/\omega_d)\sin(\omega_d t)].$$

Differentiate the bracket together with its envelope. The cosine terms cancel and the remaining derivative is proportional to $e^{-\zeta\omega_n t}\sin(\omega_d t)$. The first peak occurs at $t_p = \pi/\omega_d$. Substituting that time gives the fractional overshoot

$$M_p = \exp\left(-\frac{\pi\zeta}{\sqrt{1-\zeta^2}}\right).$$

Design for a specified peak

Suppose a positioning axis may overshoot a 10 mm command by at most 1 mm. Solving the previous expression for the limiting damping ratio gives

$$\zeta = -\ln M_p / \sqrt{\pi^2 + (\ln M_p)^2} = 0.591.$$

At $\omega_n = 5$ rad/s the corresponding poles are approximately $-2.96 \pm 4.03i$ per second and the first peak occurs at 0.779 s. A nominal oscillatory second-order system with this damping just meets the 1 mm bound. A zero near the dominant poles, a saturated drive or another slow pole changes the transient, so the result should be checked against the full model after the poles are chosen.

A settling specification differs from a peak specification. The absolute normalised error stays below the envelope $e^{-\zeta\omega_n t}/\sqrt{1-\zeta^2}$. To guarantee a 2% settling band using this bound, solve $t \geq -\ln[0.02\sqrt{1-\zeta^2}]/(\zeta\omega_n)$. This conservative time for the example is about 1.40 s. The commonly used $4/(\zeta\omega_n)$ is a convenient estimate. The envelope bound spells out both the tolerance and its assumptions.

Follow a gain change on a third-order plant

A plant with two lags and an integrator has open-loop transfer $K/[s(s+1)(s+2)]$ in units normalised to a common time base. Close unity negative feedback. The characteristic polynomial is $s^3 + 3s^2 + 2s + K$. Its coefficients are all positive for $K > 0$, which rules out positive real roots but allows an unstable complex pair. Form the first column of the Routh array: $1,\ 3,\ (6-K)/3,\ K$. All entries must have the same sign for strict stability. Thus $0 < K < 6$.

At $K = 6$ the polynomial factors as $(s + 3)(s^2 + 2)$. The step excites an undamped pair at angular frequency $\sqrt{2}$, alongside a decaying component. At $K = 7$ the first-column signs change twice, so two roots have crossed into the right half-plane. The input-output gain alone did not reveal this upper bound. Additional lags allow feedback to arrive out of phase with motion. Chapter 4 reaches the same type of conclusion by examining sine-wave responses, which can often be measured even when coefficients of a polynomial are poorly known.

Find the effect of a finite closed-loop zero

Pole locations specify the available exponential terms of a causal linear response. Their coefficients depend on the numerator and initial condition. To isolate this distinction, choose two positive times $\tau_1 \neq \tau_2$ and a transfer function $G(s) = (1 + T_z s)/[(1 + \tau_1 s)(1 + \tau_2 s)]$. It has unit static gain for every T_z, so changing the numerator neither moves the poles nor changes the final unit-step response. For zero initial state and a unit step $Y(s) = G(s)/s$. Write $Y = 1/s + A/(s + 1/\tau_1) + B/(s + 1/\tau_2)$.

The initial value $y(0^+) = 0$ gives $A + B = -1$. The initial-slope theorem gives

$$\dot{y}(0^+) = \lim_{s \to \infty} sG(s) = T_z/(\tau_1 \tau_2);$$

equivalently $-A/\tau_1 - B/\tau_2 = T_z/(\tau_1 \tau_2)$. Solving and transforming back gives

$$y(t) = 1 - \frac{\tau_1 - T_z}{\tau_1 - \tau_2} e^{-t/\tau_1} + \frac{\tau_2 - T_z}{\tau_1 - \tau_2} e^{-t/\tau_2}.$$

For $T_z = \tau_1$ one mode has zero coefficient in this particular input–output map. A physical internal state with that mode can still respond to a nonzero initial condition or another input. With $T_z < 0$ the zero lies in the right half-plane and the initial output slope opposes the eventual response. That inverse response can be a serious limitation when interpreting an early excursion as a sign error in the actuator.

Time the undershoot of two competing actuator paths

Consider an actuator that changes a measured quantity through a slow positive effect and an immediate faster opposing effect. Let $\tau_1 = 4$ s, $\tau_2 = 1$ s and $T_z = -2$ s. The poles are both stable, and the output approaches one under a unit command. Yet the response is

$$y = 1 - 2e^{-t/4} + e^{-t},$$

with initial slope $-0.50\,\mathrm{s}^{-1}$. Set $\dot{y} = 0.5e^{-t/4} - e^{-t} = 0$. It follows that $e^{-3t/4} = 0.5$, so the minimum occurs after $t = 4 \ln 2/3 = 0.924\,\mathrm{s}$. Its value is

$$1 - 2(0.794) + 0.397 = -0.191.$$

For a 10 mm final position change this model briefly produces about 1.91 mm motion in the wrong direction. A physical travel limit 1.0 mm behind the starting point would therefore be violated even

though the real parts of both poles meet a settling-time target. Compare instead $T_z = 0$ with identical poles and DC gain: the initial slope vanishes and the response rises monotonically. Before changing controller gain to correct an observed undershoot, measure which pathway acts first. Inverting the right-half-plane zero with a causal stable feedforward filter is impossible, since exact inversion would introduce an unstable pole.

For this response the first crossing back through the initial position is later than the minimum. Let $u = e^{-t/4}$; solving $1 - 2u + u^4 = 0$ gives the trivial root $u = 1$ at release and a further root near 0.544. Thus the measured position remains on the wrong side for about $-4 \ln(0.544) = 2.44$ s. A simple diagnostic that samples only once after two seconds would label the overall gain as negative. At one minute the same test reads nearly the full positive change. Recording the full waveform distinguishes an initial inverse response from an overall sensor polarity error. A disturbance entering the two paths at a different location can exhibit yet another zero, which is why a zero should always be quoted for a specified input and output.

OneSimulation · Which path is responsible for the initial undershoot?

Two stable first-order actuators contribute to the same output. Give the slow path gain $+2$ and a two-second time constant, the fast path gain -1 and a quarter-second time constant. Both start at rest and receive the same unit step. Predict the direction of the first and last movements from the initial slope and final value.

$$y(t) = 1 - 2e^{-t/2} + e^{-4t},$$

$$\dot{y}(0) = 1 - 4 = -3 \ \mathrm{s}^{-1},$$

$$y(\infty) = 1.$$

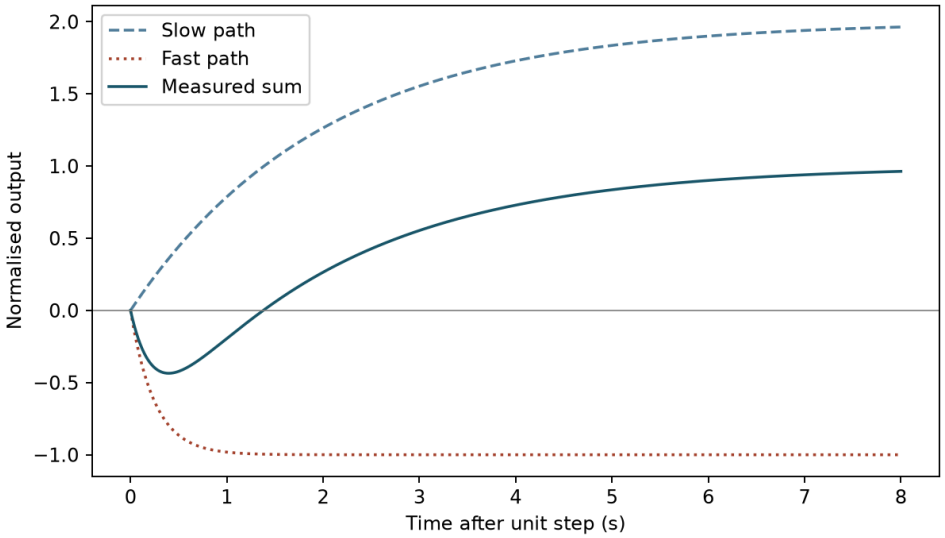
Integrate the two state equations using the fourth-order OneSimulation time stepper and compare every sample with the exponential solution. The grid minimum is –0.436 at 0.40 s. The component curves identify the mechanism. Initially the negative path builds up fastest; once it has nearly settled the slow positive path continues to grow. Vary the two time constants while keeping their gains fixed to see when the initial slope changes sign.

```
"""A positive step drives two stable paths with opposite output signs."""
import matplotlib.pyplot as plt
import numpy as np
from onesim.numerics.time.methods.backends.fixed_step import FixedStepRungeKutta

tau = np.array([2., 0.25])  # seconds, slow and fast paths
weights = np.array([2., -1.])
solver = FixedStepRungeKutta("rk4")
run = solver.integrate(lambda t, x: (1-x)/tau, np.zeros(2), 0., 8., step=0.02)
t = run.times
terms = run.states * weights
output = terms.sum(axis=1)
reference = (1-np.exp(-t[:, None]/tau)) @ weights
assert np.max(abs(output-reference)) < 1e-6
index = np.argmin(output)
print(f"First minimum on the grid: {output[index]:.3f} at {t[index]:.2f} s")

fig, ax = plt.subplots(figsize=(7, 4), constrained_layout=True)
```

```
ax.plot(t, terms[:, 0], "--", color="#51809c", label="Slow path")
ax.plot(t, terms[:, 1], ":", color="#a5442e", label="Fast path")
ax.plot(t, output, color="#19596a", label="Measured sum")
ax.axhline(0, color="0.5", linewidth=0.7)
ax.set(xlabel="Time after unit step (s)", ylabel="Normalised output")
ax.legend()
fig.savefig("control-engineering.png", dpi=170)
```



The ability to plot both internal states separates an inverse response caused by competing paths from a sign mistake in the output sensor.

Download the Python example

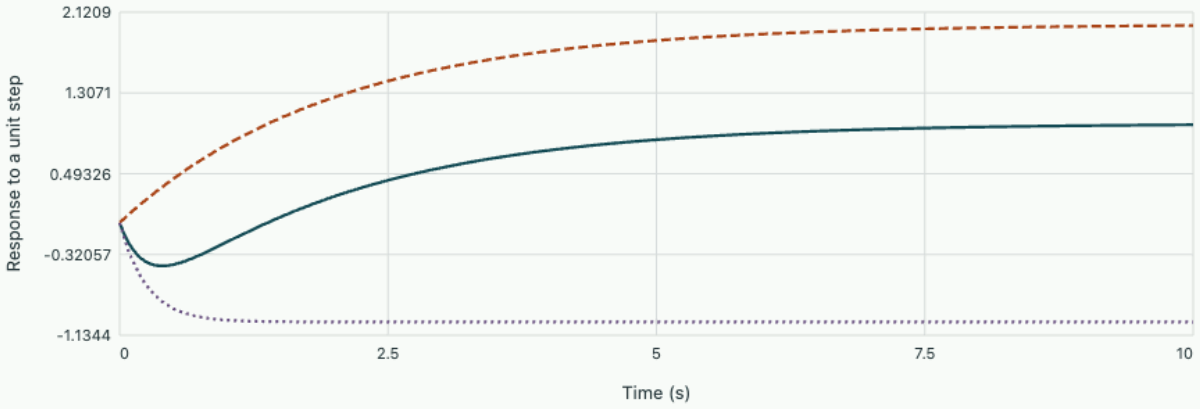
Change the inputs of this exact example and compare the recalculated curves with the recorded OneSimulation output above. The browser evaluates the stated analytical or modal reference.

Open this live calculation

Try the calculation

Slow-path time constant (s) **2**

Fast-path time constant (s) **0.25**



Initial slope -3.000 s⁻¹; final value1. Minimum -0.436 at 0.396 s. This is the exponential reference for the two states integrated in the OneSimulation example.

Assign the poles of a spring-mounted positioning stage with PID

A positioning loop can measure displacement, accumulate position error and damp measured motion. Let a stage of mass m, viscous damping c and return spring k obey $m\ddot{y} + c\dot{y} + ky = u + d$, with actuator force u and disturbance force d. The controller uses $e = r - y$, $\dot{z} = e$, and $u = K_p e + K_i z - K_d \dot{y}$. Differentiating the measured position supplies damping without turning a step in commanded position into an impulsive force command.

To find the natural motion, set $r = d = 0$ and substitute $u = -K_p y + K_i z - K_d \dot{y}$. Move the feedback terms to the left and differentiate once to eliminate z, using $\dot{z} = -y$. The resulting characteristic polynomial is

$$m s^3 + (c + K_d) s^2 + (k + K_p) s + K_i .$$

The derivative gain adds viscous damping, the proportional gain adds stiffness, and closing an integral loop adds a state. At any stable equilibrium with a constant reference, $\dot{z} = 0$ requires $y = r$. The integrator can then supply whatever constant force the spring and disturbance require, within actuator limits. The transient still depends on how the reference enters the controller.

Take $m = 2.0\,\mathrm{kg}$,

$$c = 3.0\,\mathrm{N\,s\,m^{-1}},$$

$$k = 5.0\,\mathrm{N\,m^{-1}}.$$

Suppose the desired poles are $-2, -3, -4\,\mathrm{s^{-1}}$. Multiplying the three factors gives the desired polynomial $2(s+2)(s+3)(s+4) = 2s^3 + 18s^2 + 52s + 48$. Comparing coefficients produces $K_d = 15\,\mathrm{N\,s\,m^{-1}}$,

$$K_p = 47\,\mathrm{N\,m^{-1}},$$

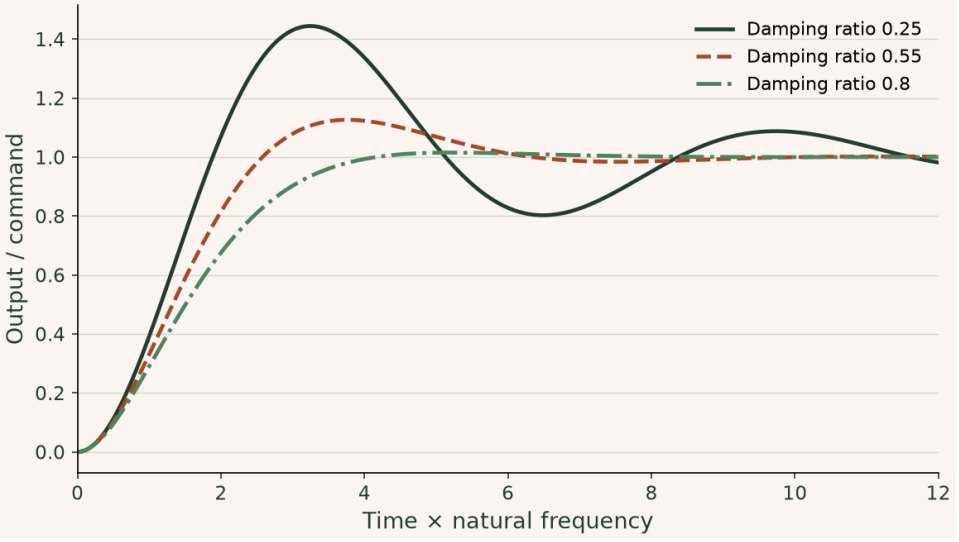
$$K_i = 48\,\mathrm{N\,m^{-1}s^{-1}}.$$

For a 20 mm command from rest the initial proportional force is $0.940\,\mathrm{N}$. At the final resting position with $d = 0$, the spring needs $0.100\,\mathrm{N}$. $K_i z$ supplies it and the proportional contribution is zero. A filtered derivative introduces an additional pole, so this ideal coefficient match must be checked again against the implemented controller bandwidth.

Practise on sheet 2 · Assign a repeated pole and check a holding-force constraint. Further reading · MIT · 16.06 Principles of Automatic Control · Lecture 10, PID Control.

Settle a tracking step without hiding its peak

Apply a unit step to a unity-DC second-order transfer function with no zeros. Keep natural frequency fixed. The first maximum occurs when $\omega_d t = \pi$; use the waveform both to check the overshoot formula and to see how late oscillations affect a settling band.



At $\zeta = 0.55$ the first overshoot is about 12.6%. A design with a smaller damping ratio may cross the command sooner while taking much longer to stay within a narrow tolerance.

▶ **Read values from the plot**

Interactive calculation · change the inputs of this example

Move through critical damping and compare overshoot, peak time and the duration of the tail.

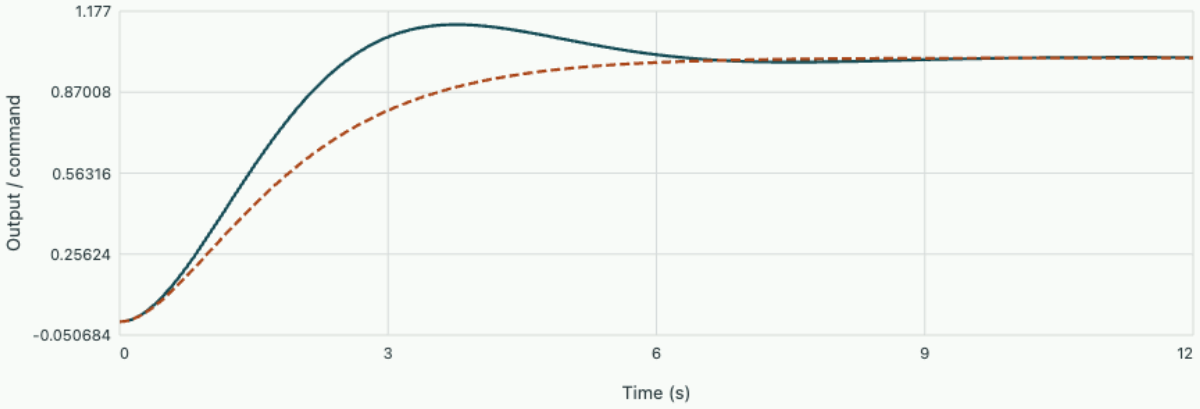
The curves below are recalculated from this section's equations in your browser. The fixed reference plot above remains available for comparison.

Open this live calculation

Try the calculation

Damping ratio **0.55**

Natural frequency (rad/s) **1**



Selected poles Critical damping

First overshoot 12.63%; first peak at 3.762 s.

Chapter 4 · Frequency response, disturbance rejection and robustness

A loop that tracks slowly changing commands may amplify a disturbance at another frequency. Work from the sinusoidal response to a defensible choice of bandwidth.

FROM FIRST PRINCIPLES

Read sinusoidal tracking and disturbance rejection from one loop

For a stable linear time-invariant plant driven at angular frequency ω, time derivatives of $e^{i\omega t}$ multiply it by $i\omega$. After transients decay, the ratio of complex output and input amplitudes is therefore the transfer function evaluated at $s = i\omega$. Its magnitude measures the amplitude ratio; its argument measures phase shift. With an ideal negative-feedback sensor define loop transfer $L = PC$ for plant P and controller C.

$$U = C(R - Y),$$
$$Y = PU + D = L(R - Y) + D.$$

Here D is a disturbance added to the physical output. Rearrangement separates the contributions of the command and disturbance at the same frequency.

$$Y = \frac{L}{1+L}R + \frac{1}{1+L}D,$$
$$T = \frac{L}{1+L},$$
$$S = \frac{1}{1+L}.$$

Adding the two ratios gives $S + T = 1$. Large loop magnitude generally permits tracking and disturbance reduction at low frequency. When the loop phasor approaches -1, the common denominator becomes small; phase information is therefore essential when interpreting a measured gain curve.

DERIVATION AND APPLICATIONS

Infer damping and bandwidth from a resonant frequency sweep

A measured sinusoidal peak need not occur at the undamped natural frequency. Consider a stable second-order plant with positive static gain K, no finite zero and transfer $G(s) = K\omega_n^2/(s^2 + 2\zeta\omega_n s + \omega_n^2)$. A slow swept-sine test measures the steady response after transients have decayed. With the frequency ratio $r = \omega/\omega_n$, the squared magnitude has a denominator that is a quadratic in r^2.

$$\frac{K^2}{|G(i\omega)|^2} = (1-r^2)^2 + 4\zeta^2 r^2 = 1 + (4\zeta^2 - 2)r^2 + r^4.$$

Find the minimum of the denominator at a positive frequency by differentiating with respect to r^2. A positive root requires $\zeta < 1/\sqrt{2}$. Substitution at this root gives the peak gain relative to the measured static gain.

$$r_p^2 = 1 - 2\zeta^2,$$
$$M_p = \frac{|G(i\omega_p)|}{K} = \frac{1}{2\zeta\sqrt{1-\zeta^2}}.$$

A different frequency occurs where the gain has fallen to $K/\sqrt{2}$. Set the squared denominator to two and select the positive root in r^2. This bandwidth is referenced to the DC gain, even when the system has a large resonant peak.

$$r_{\mathrm{bw}}^2 = 1 - 2\zeta^2 + \sqrt{(1-2\zeta^2)^2 + 1}.$$

The location of the measured peak approaches ω_n only for small damping. As damping approaches the threshold from below, the peak moves to zero frequency and becomes difficult to distinguish from the static response.

WORKED EXAMPLE

Interpret a peak in a stage measurement

A stage has $\zeta = 0.300$ and $\omega_n = 20.0$ rad/s. The predicted peak occurs at $20\sqrt{0.820} = 18.1$ rad/s. Its amplitude is $1/[0.600\sqrt{0.910}] = 1.75$ times the static response. The DC-referenced bandwidth is

$$20\sqrt{0.820 + \sqrt{1.6724}} = 29.1\,\mathrm{rad/s}.$$

A test recording only the peak frequency would therefore underestimate the bandwidth.

Change the inputs and inspect this calculation below

SECOND APPLICATION

Estimate damping without measuring phase

A second stage shows a peak twice its static gain at 12.0 Hz. Solving $4\zeta^2(1-\zeta^2) = 1/4$ on the resonant branch gives $\zeta^2 = (1 - \sqrt{3/4})/2$ and $\zeta = 0.259$. Its natural frequency is

$$12.0/\sqrt{1-2\zeta^2} = 12.9\,\mathrm{Hz}.$$

The phase should be -90° there, offering a separate frequency-response check on the no-zero model.

Interactive calculation · Change the inputs of this example

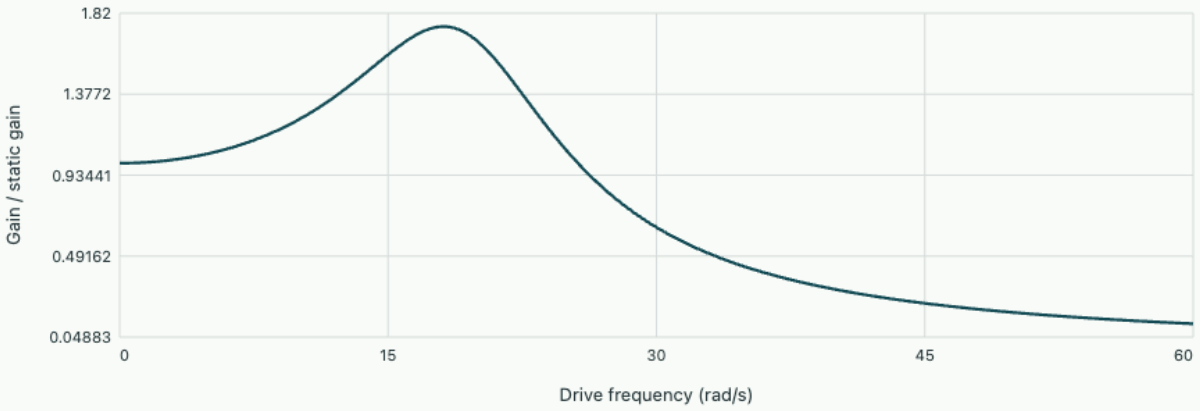
Use the adjustable quantities to test the two worked calculations above. The plot recomputes this derivation's balance or constitutive relation; its numerical readout reports the quantities needed for the forward calculation and the inverse or design question. Keep the fixed assumptions stated in the readout when comparing with the example values.

Open this live calculation

Try the calculation

Damping ratio **0.3**

Natural frequency (rad/s) **20**



Second-order frequency response

Peak frequency 18.111 rad/s; DC-referenced −3dB bandwidth 29.074 rad/s. Above ζ=1/√2 the response has no nonzero-frequency gain peak.

Read a complex gain as an experiment

Drive a stable linear plant with a small sinusoid about its operating point. Once the transient dies away, the output is a sinusoid at the same angular frequency. Inserting $u = Ue^{i\omega t}$ and $y = Ye^{i\omega t}$ in the differential equation gives

$$Y = P(i\omega)U.$$

The modulus gives the amplitude ratio and the argument gives the time shift as a fraction of a cycle. The units on a Bode plot matter: for a dimensional plant gain its magnitude in decibels presupposes a reference gain. The product around a feedback loop is dimensionless.

For $P = K/(1 + s\tau)$, the modulus falls to $K/\sqrt{2}$ at $\omega = 1/\tau$, where its phase is -45°. Above this break, doubling frequency roughly halves the output amplitude. Adding a transport delay multiplies the frequency response by $e^{-i\omega T}$; the magnitude is unchanged while phase falls by ωT radians. A magnitude measurement alone cannot detect this mechanism.

Background · How does a timing offset appear in a measured frequency response?

If the output data acquisition channel lags the input clock by a fixed Δt, every recorded sinusoid has the correct amplitude but an additional phase lag. Multiplying the physical response by the time-delay factor gives

$$G_{\text{measured}}(i\omega) = e^{-i\omega\Delta t} G_{\text{plant}}(i\omega),$$
$$\Delta\varphi = -\omega\Delta t.$$

At 20 Hz a 5 ms offset produces a −36° phase error. At 2 Hz it produces only −3.6°. If the magnitude agrees with a plant model but the unwrapped phase discrepancy grows nearly linearly with frequency, repeat the sweep with a common input pulse on both acquisition channels. A real process delay and a clock offset have the same transfer factor but different remedies.

Derive the sensitivity functions

Let $u = C(r - y - n)$ for command r and sensor noise n, with disturbance d added directly to the physical plant output $y = Pu + d$. Substitution gives

$$S = (1 + PC)^{-1},$$
$$T = PC(1 + PC)^{-1},$$
$$y = Tr + Sd - Tn.$$

The identity $S + T = 1$ holds for complex numbers at each frequency. A slow output disturbance is strongly rejected when the loop gain is large and far from -1. Sensor noise at the same frequency is then nearly reproduced by the physical output with opposite sign. At high frequencies a deliberately small loop gain protects the output against measurement noise. If the disturbance enters at the actuator instead, its output transfer function is PS. Mark its point of entry before reading any plot.

How much delay can a loop tolerate?

For a unity feedback loop take $L(s) = 4/[s(1 + s)]$, with time in seconds. A unity-gain crossover satisfies $\omega^2(1 + \omega^2) = 16$. The positive root gives

$$\omega_c = \sqrt{(\sqrt{65} - 1)/2} = 1.88\,\text{rad/s}.$$

At that frequency the phase is

$$-90^\circ - \tan^{-1} 1.88 = -152.0^\circ.$$

A further 28.0° of lag would bring the Nyquist curve through -1, where the closed-loop denominator vanishes. For pure delay alone the first corresponding delay is about $0.489/1.88 = 0.260$ s. A 0.10 s delay consumes about 10.8° at crossover.

This estimate uses this plant's unique gain crossover and its initially stable closed loop. Multiple crossovers, unstable plant poles and poorly measured resonances require inspection of the entire Nyquist curve. Even before a pole crosses into instability, the distance to -1 controls amplification: $|S| = 1/|1 + L|$. For example $L = -0.9 + 0.1i$ gives $|S| = 7.07$. A loop with low steady error may therefore magnify a narrow band of disturbance by a factor of seven.

Translate specifications into a measurement

Suppose a speed drive must reject load variations below 0.5 rad/s while its encoder contains noise above 20 rad/s. First locate the disturbance and noise on the block diagram. Specify maximum allowed output amplitude per unit amplitude at each point. Shape the gain, then inspect the two resulting transfer functions across the entire frequency range, including the transition between bands. Raising proportional gain alone shifts the crossover into regions of additional lag. A measured resonance there can set the achievable speed even when the motor model at low frequency is excellent.

Derive the phase contributed by a lead compensator

Suppose a plant has enough low-frequency authority to reject a slowly varying disturbance, but its phase at the desired gain crossover leaves too little damping margin. A real lead compensator $C(s) = K(1 + s/\omega_z)/(1 + s/\omega_p)$ with $0 < \omega_z < \omega_p$ contributes phase $\phi = \arctan(\omega/\omega_z) - \arctan(\omega/\omega_p)$. It also raises gain towards high frequencies. A useful design therefore fixes the desired crossover, calculates the new phase there and then accounts for the gain change rather than assuming crossover stays where it was.

Differentiate phase with respect to $\ln \omega$. Its stationary point satisfies $(\omega/\omega_z)/[1 + (\omega/\omega_z)^2] = (\omega/\omega_p)/[1 + (\omega/\omega_p)^2]$, giving $\omega_m = \sqrt{\omega_z \omega_p}$. Let $\alpha = \omega_z/\omega_p < 1$. At this frequency the magnitude ratio excluding K is $1/\sqrt{\alpha}$ and $\sin \phi_{\max} = (1 - \alpha)/(1 + \alpha)$. One can consequently choose a target phase, infer α, centre the lead at crossover and determine K from $|CP| = 1$.

The increased gain at high frequency can pass measurement noise into the actuator. Check the control-signal sensitivity as well as the output’s disturbance attenuation, particularly when the plant attenuates high frequencies and the sensor has its own noise floor.

Design a compensator from a measured frequency point

Take the normalised positioning plant $P(s) = 1/[s(1 + 0.2s)]$ and target crossover 5 rad/s. Its magnitude there is $1/[5\sqrt{2}] = 0.1414$; phase is

$$-90^\circ - \arctan 1 = -135^\circ.$$

A constant gain chosen for that crossover would leave 45° phase margin. Allow an extra 15° of phase for a target 60° nominal margin before measurement delay. Choosing 15° peak lead gives

$$\alpha = (1 - \sin 15^\circ)/(1 + \sin 15^\circ) = 0.589.$$

The zero and pole are placed at $5\sqrt{\alpha} = 3.84$ and $5/\sqrt{\alpha} = 6.52$ rad/s.

The lead factor's magnitude at crossover is $1/\sqrt{0.589} = 1.303$, so

$$K = 1/[0.1414(1.303)] = 5.43.$$

The loop crosses with phase –120°, leaving the intended nominal margin. A previously omitted 20 ms pure delay adds $-5(0.020) = -0.100$ rad, or –5.73°, at this point without changing magnitude. The resulting margin is approximately 54.3°. The compensator gain tends to $K/\alpha = 9.22$ at high frequency, compared with the proportional-only crossover gain 7.07. A noisy encoder can therefore change the actuator demand when this seemingly modest phase correction is installed. Verify the measured return ratio near crossover and observe the actuator spectrum above it after implementation.

Estimate the timing allowance explicitly by converting the margin reserve back into seconds. If the specification requires at least 45°, a nominal 60° design has 15° available for a pure delay at its unchanged crossover. The maximum delay under this model is $15\pi/[180(5)] = 0.0524\,\mathrm{s}$, leaving about 32 ms beyond the omitted 20 ms. This is a local phase-margin budget and assumes a single relevant gain crossover and a stable Nyquist configuration. An unmodelled resonance can create another crossover with a smaller margin. The frequency measured by a signal generator may be labelled in hertz; the relevant crossover here is

$$5/(2\pi) = 0.796\,\mathrm{Hz}.$$

Mistakenly using five hertz in the delay calculation would consume more than six times the actual phase and lead to an unnecessarily severe timing constraint.

Use a winding count when the uncontrolled plant is unstable

A gain or phase margin read at a single crossover can hide unstable modes if the open loop already has right-half-plane poles. For negative unity feedback the closed-loop characteristic condition is

$$F(s) = 1 + L(s) = 0.$$

Trace the boundary of the right half-plane up the imaginary axis from $-i\infty$ to $+i\infty$ and return on a large semicircle to the right. This contour runs clockwise. Assume a strictly proper rational loop with no poles or characteristic roots on the contour, and no hidden unstable pole cancellations.

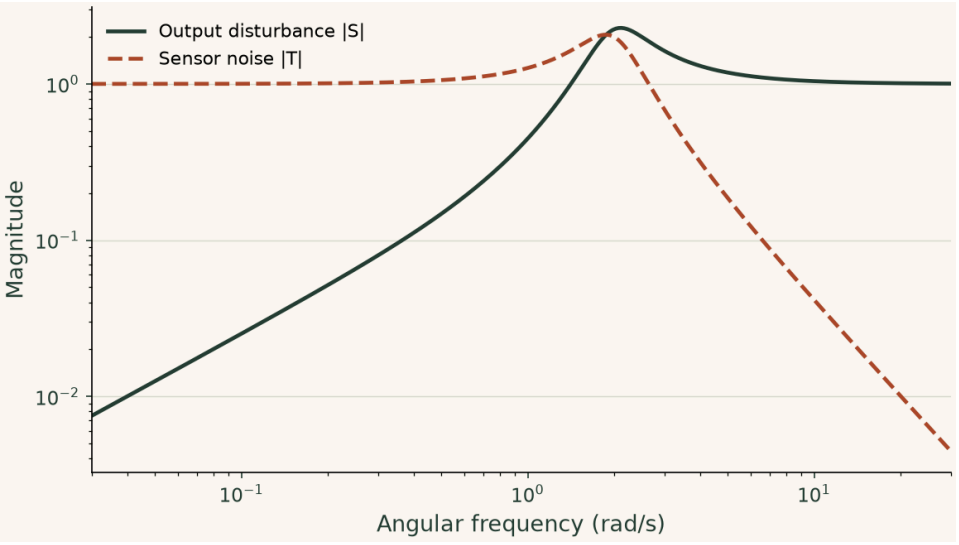
A factor $s - z$ with z inside contributes one clockwise turn of F. An enclosed denominator factor contributes a turn in the opposite direction. If Z is the number of right-half-plane zeros of F and P its number of right-half-plane poles, the net counterclockwise winding of F about zero is $P - Z$. Equivalently, count the turns N of L about -1 with counterclockwise positive. Then $Z = P - N$. Closed-loop stability requires $N = P$ under this stated contour and sign convention.

Consider a linearized temperature imbalance with loop transfer $L(s) = K/(s - a)$, where $a = 2\,\mathrm{s}^{-1}$ and $K > 0$ has the same unit. The unregulated imbalance grows exponentially. Along $s = i\omega$, the real and imaginary parts of L are $-Ka/(a^2 + \omega^2)$ and $-K\omega/(a^2 + \omega^2)$. Completing the square shows a circle centred at $-K/(2a)$ on the real axis with radius $K/(2a)$. Increasing frequency traces its upper half first and then the lower half, one counterclockwise turn. At $K = 3\,\mathrm{s}^{-1}$ the circle spans $[-1.5, 0]$ and encloses -1. Hence $P = 1$, $N = 1$, $Z = 0$. At $K = 1.5\,\mathrm{s}^{-1}$ the circle spans only $[-0.75, 0]$, giving $Z = 1$. The explicit characteristic root $s = a - K$ confirms both conclusions. At $K = a$ the locus passes through -1 at zero frequency and the closed loop has a pole at the origin.

Practise on sheet 2 · Explain why a stable three-lag plant can acquire two unstable poles. Further reading · University of Washington · X. Chen, Introduction to Control Theory (ME 545) · Lecture 24, Nyquist Stability: Encirclements and the Nyquist Criterion.

Where does load rejection give way to noise sensitivity?

For unity feedback with loop transfer $4/[s(1 + s)]$, inject the disturbance at the plant output and the noise at the sensor. Plot the magnitudes of the resulting two closed-loop maps. The identity $S + T = 1$ involves phase and does not imply that their magnitudes add to one.



At low frequency the feedback rejects an output disturbance. Near the lightly damped closed-loop resonance the sensitivity can exceed one. Calling the loop a disturbance rejector at every frequency would miss that amplification.

▶ **Read values from the plot**

Interactive calculation · change the inputs of this example

Increase gain and observe both low-frequency rejection and resonance amplification. The magnitudes do not obey a scalar sum rule.

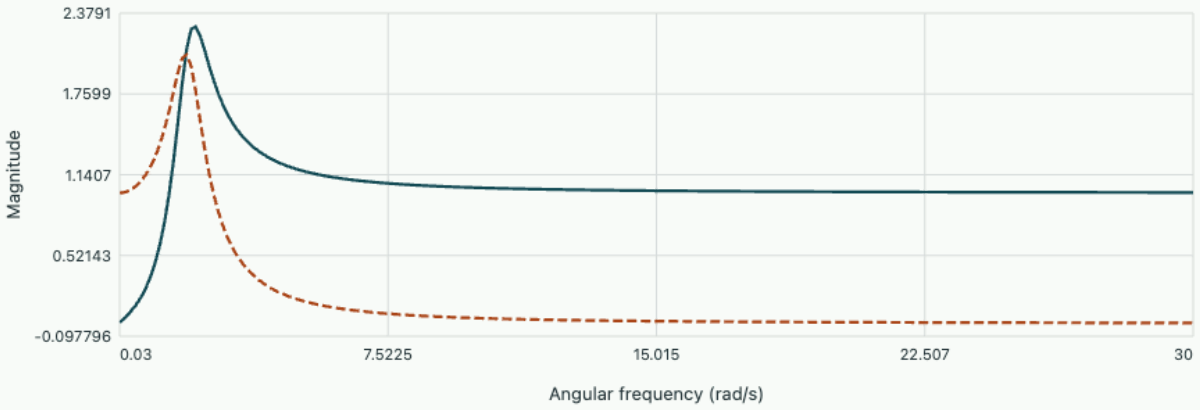
The curves below are recalculated from this section's equations in your browser. The fixed reference plot above remains available for comparison.

Open this live calculation

Try the calculation

Loop gain K **4**

Plant lag time (s) **1**



|S| output disturbance |T| sensor noise

The loop is K/[s(1+τs)]. At ω = √(K/τ), |T| = 2.000. S+T=1 is a complex identity.

Study the full derivation

Sensitivity, bandwidth and the integral price of disturbance rejection · System dynamics, feedback and estimation

A regulator attenuates a slowly varying output load yet amplifies another disturbance at a higher frequency. Separate the paths by which reference, measurement noise and physical disturbance reach the recorded variable, then derive an integral identity that explains why a stable loop can redistribute sensitivity across frequency.

Problem sheet 2 · Pole placement, gain and phase

Design a second-order transient, examine a hidden stability limit, and test a delay budget against a frequency response.

1. Select damping from a peak

A unity-static-gain second-order system has no zeros. Specify 10% overshoot and first peak at 1 s after a step. (a) Derive damping ratio and damped frequency. (b) Find the desired characteristic polynomial. (c) State which prediction fails if a numerator zero is added and the poles are held fixed.

▼ Start with the logarithm of the overshoot

The positive logarithmic decrement to the first peak is $-\ln(0.10) = 2.303 = \pi\zeta/\sqrt{1-\zeta^2}$, giving $\zeta = 0.591$. The damped frequency is $\omega_d = \pi$ rad/s, and the exponential decay rate is 2.303 s^{-1}. The poles are $-2.303 \pm 3.142i$, so the monic polynomial is $s^2 + 4.605s + 15.17$. A zero changes the residues and the step shape. The poles still give the natural frequencies and decay rates, but first-peak time and overshoot need recalculation.

---

2. Three time constants and one gain

For $L = g/[s(s+1)(s+3)]$ under unity negative feedback, (a) derive the closed-loop characteristic polynomial, (b) find the positive gain range for stability and the crossing frequency, and (c) compare the pole result with the open-loop frequency response at the crossing.

▼ Locate the imaginary-axis pair

The polynomial is $s^3 + 4s^2 + 3s + g$. The first column of the Routh array is $1, 4, (12-g)/4, g$; all signs agree for $0 < g < 12$. At $g = 12$ factor the polynomial as $(s+4)(s^2+3)$. The crossing frequency is $\sqrt{3}$ rad/s, with the third pole still at -4. Substituting $s = i\sqrt{3}$ in the open-loop denominator gives -12, so $L = -1$ at the same gain and frequency.

---

3. Delay on a positioning stage

With $L = 2/[s(1 + s/2)]$, (a) find the gain crossover, (b) estimate the delay that would use all phase margin, and (c) evaluate the remaining margin after including 0.10 s delay. Count angles consistently.

▼ Convert delay to a phase at crossover

Writing $x = \omega_c^2$ gives $x(1 + x/4) = 4$, so

$$\omega_c = \sqrt{-2 + 2\sqrt{5}} = 1.572\,\mathrm{rad/s}.$$

The phase margin is $90^\circ - \arctan(1.572/2) = 51.8^\circ$, or 0.905 rad. A pure delay leaves the crossover magnitude fixed, so the first crossing occurs at about $0.905/1.572 = 0.576$ s. A 0.10 s delay costs 0.157 rad or 9.0°, leaving approximately 42.8°. This calculation assumes the given loop is otherwise accurately modelled.

Assign a repeated pole and check a holding-force constraint

A second stage has $m = 1.0\,\mathrm{kg}$,

$$c = 2.0\,\mathrm{N\,s\,m^{-1}},$$

$$k = 6.0\,\mathrm{N\,m^{-1}}.$$

Select the PID gains for poles at $-2, -2, -5\,\mathrm{s^{-1}}$. A constant $+0.50\,\mathrm{N}$ force then acts on the stage while its position command is $+0.020\,\mathrm{m}$. Determine the holding force and whether an actuator limited to $\pm 0.30\,\mathrm{N}$ can achieve zero steady position error.

▼ **Show the worked solution**

$$(s+2)^2(s+5) = s^3 + 9s^2 + 24s + 20.$$

Subtract the passive coefficients to obtain $K_d = 7.0\,\mathrm{N\,s\,m^{-1}}$,

$$K_p = 18\,\mathrm{N\,m^{-1}},$$

$$K_i = 20\,\mathrm{N\,m^{-1}s^{-1}}.$$

At rest the plant requires $u = kr - d = 6(0.020) - 0.50 = -0.38\,\mathrm{N}$, outside the range. Even with $u = -0.30\,\mathrm{N}$, the equilibrium is

$$y = (u + d)/k = 0.0333\,\mathrm{m}.$$

The pole assignment describes unsaturated dynamics. Tuning the three gains cannot provide the missing holding authority.

Revisit chapter 3 · Assign the poles of a spring-mounted positioning stage with PID.

Explain why a stable three-lag plant can acquire two unstable poles

Three identical first-order lags with one-second time constants give $L(s) = K/(1+s)^3$, using seconds as the time unit. Find the frequency at which the locus reaches the negative real axis and the

corresponding critical positive gain. For $K = 27$, determine the unstable pole count and the signed Nyquist winding.

▼ **Show the worked solution**

The loop phase is $-3\arctan\omega$, so $-\pi$ is reached at $\omega = \tan(\pi/3) = \sqrt{3}\,\mathrm{rad\,s^{-1}}$. The magnitude there is $K/(1+3)^{3/2} = K/8$. The critical gain is 8. At $K = 27$, solve $(s+1)^3 = -27$. The three roots are -4 and $0.5 \pm i(3\sqrt{3}/2)$, so $Z = 2$. Since the three open-loop poles all lie at -1, $P = 0$ and $N = P - Z = -2$, two net clockwise turns. At the critical gain a conjugate pair lies on the imaginary axis and the winding through the critical point is undefined.

Revisit chapter 4 · Use a winding count when the uncontrolled plant is unstable.

Explore the governing equations after solving the sheet: Closed-loop poles, damping and performance; Frequency response, disturbance rejection and robustness. Compare the assumptions and units before transferring your chosen parameters; the controls may cover a different numerical range from the written exercise.

Identify a controller from a finite-time observation

A zero-initial-state, first-order proportional-feedback loop has plant static gain 2. A unit step gives error 0.11331 at six fitted closed-loop time constants. The fitted time constant is 2.22222 s. Infer the controller gain and the original plant time constant, retaining the residual exponential at the sample time.

▼ **Work through the solution**

The first-order error is

$$e(t) = e_\infty + (1 - e_\infty)e^{-t/\tau_c}.$$

Thus $e_\infty = (0.11331 - e^{-6})/(1 - e^{-6}) = 0.11111$ to rounding. With negative unity feedback, $e_\infty = 1/(1 + 2K_p)$ gives $K_p \simeq 4.00$. The closed-loop denominator gives $\tau_c = \tau_p/(1 + 2K_p)$, so $\tau_p \simeq 20.0$ s. Treating the sampled error as the asymptotic value would slightly underestimate gain.

Use the OneSimulation laboratory to check your answer

Vary the corresponding laboratory reference interactively. Compare the recalculated observable with the recorded OneSimulation runs and the approximation used in this exercise.

Chapter 5 · States, reachability and estimation

Two mechanisms can produce the same output at one instant and different outputs an instant later. Determine what the actuator can change and what a sensor history can reveal.

FROM FIRST PRINCIPLES

Build a position and velocity state model from force balance

For a moving mass m acted on by commanded force u, a spring k and viscous drag b, Newton's second law relates acceleration to present displacement and velocity. Assume constant coefficients, a rigid base and small linear motions. Defining x_1 as position and x_2 as velocity turns the second-order relation into two first-order state equations.

$$\dot{x}_1 = x_2,$$
$$m\dot{x}_2 = u - kx_1 - bx_2.$$

State feedback chooses force from the measured or estimated state. For $u = -k_p x_1 - k_v x_2 + nr$, the command r also supplies a reference input. Collecting terms yields the closed-loop position equation.

$$m\ddot{x}_1 + (b + k_v)\dot{x}_1 + (k + k_p)x_1 = nr.$$

Comparing its characteristic polynomial with a desired $s^2 + 2\zeta\omega_n s + \omega_n^2$ determines both feedback coefficients directly.

$$k_p = m\omega_n^2 - k,$$
$$k_v = 2m\zeta\omega_n - b.$$

With $n = k + k_p$, a constant position reference is reproduced at equilibrium in the nominal unloaded model. Position measurements alone cannot specify the instantaneous force law until velocity is also measured or reconstructed from the sensor history.

Background · Observability

$$\mathcal{O} = [C^{\mathsf{T}}, (CA)^{\mathsf{T}}, \ldots]^{\mathsf{T}}$$

Full column rank in the standard finite-dimensional linear model allows the initial state to be inferred from ideal output history. Noise still controls achievable precision.

DERIVATION AND APPLICATIONS

Separate observer error from the state-feedback poles

When only position is measured, feedback must operate on estimated position and velocity. Consider a continuous linear plant $\dot{x} = Ax + Bu$, $y = Cx$, with constant matrices and a known applied input. Use a full-order observer with correction proportional to the difference between measured and predicted output. Subtract its state equation from the plant with estimation error $e = x - \hat{x}$.

$$\dot{\hat{x}} = A\hat{x} + Bu + L(y - C\hat{x}),$$
$$\dot{e} = (A - LC)e.$$

The known input cancels exactly. Now apply $u = -K\hat{x}$ and use $\hat{x} = x - e$. The physical plant sees a perturbing term proportional to the residual estimation error. Writing the pair of equations together exposes an upper triangular block structure.

$$\frac{d}{dt}\begin{pmatrix} x \\ e \end{pmatrix} = \begin{pmatrix} A - BK & BK \\ 0 & A - LC \end{pmatrix}\begin{pmatrix} x \\ e \end{pmatrix}.$$

Taking the determinant of the characteristic matrix gives the product of the two diagonal determinants. For a controllable and observable nominal model the feedback poles and observer poles can therefore be chosen separately. Faster estimation may amplify sensor noise, which this deterministic pole calculation does not quantify.

$$p(s) = \det(sI - A + BK)\,\det(sI - A + LC).$$

WORKED EXAMPLE

Place poles for a position stage

Use the unit-mass double integrator $\dot{x}_1 = x_2$, $\dot{x}_2 = u$, $y = x_1$. The feedback polynomial is $s^2 + k_2 s + k_1$ and the observer polynomial is $s^2 + l_1 s + l_2$. For feedback poles $-2, -3$ s⁻¹ select $K = (6, 5)$. For observer poles $-8, -9$ s⁻¹ select $L = (17, 72)^{\mathsf{T}}$ in consistent units. The complete system has all four requested poles. The observer correction on the velocity state is large because it must reconstruct velocity from a position discrepancy.

Change the inputs and inspect this calculation below

SECOND APPLICATION

Predict the effect of an unknown constant acceleration

For a separate double integrator take $L = (10, 24)^{\mathsf{T}}$ and an unmodelled constant acceleration $d = 0.120$ m/s². At steady state $0 = e_2 - 10e_1$ and $0 = -24e_1 + d$, giving position error 5.00 mm and velocity error 0.0500 m/s. Stable observer poles remove initial-state errors, while this persistent

disturbance supplies a continuing input to the error dynamics. A sensor recording the innovation should see the 5.00 mm bias.

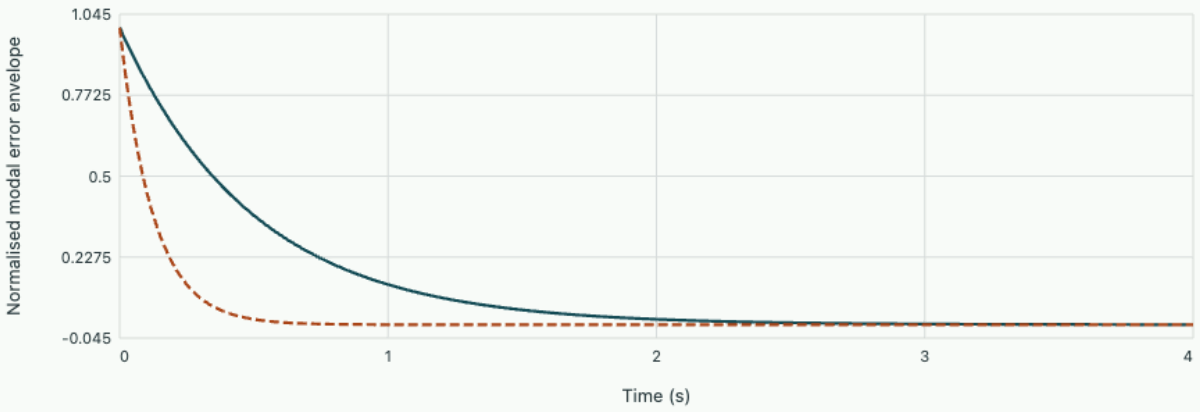
Interactive calculation · Change the inputs of this example

Use the adjustable quantities to test the two worked calculations above. The plot recomputes this derivation's balance or constitutive relation; its numerical readout reports the quantities needed for the forward calculation and the inverse or design question. Keep the fixed assumptions stated in the readout when comparing with the example values.

Open this live calculation

Try the calculation

Feedback pole-rate scale	1
Observer pole-rate scale	1
Unknown acceleration	1



K=(6.000,5.000), L=(17.000,72.000). Under constant unknown acceleration, steady position offset=0.377. The observer does not remove an unmodelled constant input.

From balances to independent states

A cart of mass m moves against a spring k and damper b. Choose $x = (q, v)^{\mathsf{T}}$, displacement and velocity. Newton's law gives

$$\dot{q} = v,\ m\dot{v} = -kq - bv + u.$$

A state contains the initial information needed to propagate the motion with specified input. Measuring displacement at one instant supplies only half of that information. The state choice is not unique: replacing velocity by momentum changes the matrix entries, yet the predicted displacement history is the same.

$$\dot{x} = Ax + Bu, \quad y = Cx, \qquad A = \begin{pmatrix} 0 & 1 \\ -k/m & -b/m \end{pmatrix},\ B = \begin{pmatrix} 0 \\ 1/m \end{pmatrix},\ C = (1 \ \ 0).$$

With zero input the transition e^{At} propagates initial conditions. Integrating the effect of forces applied at each earlier time gives

$$x(t) = e^{At}x(0) + \int_0^t e^{A(t-s)}Bu(s)\,ds.$$

This decomposition explains why both the autonomous motion and the input direction enter control design.

Test an actuator and a sensor

Over a short time, the vectors $e^{At}B = B + tAB + \cdots$ show the directions that a force can excite. In two dimensions B, AB suffice by the characteristic equation of A. For the cart their columns have determinant $-1/m^2$; a finite mass allows the actuator to change both position and velocity over time. This rank test assumes arbitrary inputs and sufficient time. Getting to a distant position in a very short time can still exceed an actuator limit.

With a known input, output and its first derivative contain Cx and CAx. Here they are position and velocity, so the observability matrix has full rank. As a counterexample, consider two uncoupled tanks with equal decay rates, driven equally, and one sensor measuring the sum of levels. The difference evolves autonomously and never enters the sum. Recording a longer noiseless trace with that sensor cannot recover the initial difference. Moving a sensor or adding a distinct excitation has to break the symmetry responsible for the missing information.

Place poles and restore the reference scale

Take $m = 1,\ b = 1,\ k = 2$ in compatible SI units. Apply $u = -f_q q - f_v v + nr$. The homogeneous characteristic polynomial is $s^2 + (1 + f_v)s + (2 + f_q)$. Desired poles at $-2, -3$ give $f_v = 4,\ f_q = 4$ by coefficient matching. At constant reference the velocity vanishes and $(2 + f_q)q = nr$, so $n = 6$ gives $q = r$ in the nominal model. A constant unmodelled force shifts this equilibrium. State feedback with a static reference multiplier alone supplies no mechanism to learn that force.

Reconstruct velocity without differentiating noisy measurements

An observer runs the model and uses the disagreement with the measured output to correct its state. Write $\dot{\hat{x}} = A\hat{x} + Bu + H(y - C\hat{x})$. In the ideal deterministic model the estimation error obeys $\dot{e} = (A - HC)e$. Here its polynomial is $s^2 + (1 + h_1)s + (h_1 + 2 + h_2)$. Desired observer poles $-4, -5$ give

$$h_1 = 8,\ h_2 = 10.$$

Thus an initial state error decays even though the sensor supplies only displacement.

With the control law evaluated at $\hat{x}$, the combined state and estimation-error dynamics are block triangular. The two sets of assigned poles therefore survive in the ideal linear model. Faster observer poles also give sensor noise larger correction gains. Compare simulated or measured velocity estimates during a quiet hold and during a sharp manoeuvre before accepting the faster design. A missing mode, a noisy estimate and the inability to measure an initial state exactly have distinct engineering implications.

For additional examples of state estimation and output feedback, see Åström and Murray's Feedback Systems.

Add a state that estimates a constant load through tracking error

A feedforward reference gain can remove steady offset for an exact plant model. When an unknown constant load enters at an actuator, the equilibrium input changes. Integrating the tracking error gives the controller a memory with which to supply the missing steady effort. For a state model $\dot{x} = Ax + Bu + Ed,\ y = Cx$, augment the state by z with $\dot{z} = r - Cx$. If the augmented closed loop is stable and reaches equilibrium, then $\dot{z} = 0$ enforces $y = r$. This argument assumes constant feasible r, d, no actuator saturation and an unbiased measurement.

Take a velocity servo $\dot{v} = -av + bu + d$ with $u = -k_v v + k_i z$; use v in m/s, u in volts and z in metres. The augmented characteristic polynomial follows by differentiating the velocity equation for a constant disturbance, or directly from its two-by-two state matrix:

$$\dot{v} = -(a + bk_v)v + bk_i z + d,$$

$$\dot{z} = r - v,$$

$$p(s) = s^2 + (a + bk_v)s + bk_i.$$

The augmented controllability test is important in more general plants: a servo integrator at zero frequency cannot arbitrarily impose an output that the input cannot sustain. Even here $b = 0$ would make the augmented system uncontrollable. Physical travel or voltage limits can also restrict the equilibrium although the ideal linear test passes.

Place the integral servo poles and check the required voltage

Assume $a = 2\,\mathrm{s}^{-1}$, $b = 3\,(\mathrm{m/s^2})/\mathrm{V}$ and target poles at −3 and −4 s⁻¹. Matching $p(s) = s^2 + 7s + 12$ gives

$$k_v = (7 - 2)/3 = 1.667\,\mathrm{V\,s/m},$$

$$k_i = 12/3 = 4.0\,\mathrm{V/m}.$$

At command $r = 1.0\,\mathrm{m/s}$ and constant opposing disturbance $d = -0.60\,\mathrm{m/s^2}$, equilibrium requires

$$u = (ar - d)/b = 0.867\,\mathrm{V}.$$

The integrator settles at $z = (u + k_v r)/k_i = 0.633\,\mathrm{m}$. The unit assigned to this internal state is integrated velocity error, not physical servo displacement.

For a pure disturbance step with reference held fixed, eliminating z gives $V(s)/D(s) = s/[(s+3)(s+4)]$ in deviation variables. A −0.60 m/s² step therefore produces $\Delta v(t) = -0.60(e^{-3t} - e^{-4t})\,\mathrm{m/s}$. Its minimum occurs at $t = \ln(4/3) = 0.288\,\mathrm{s}$, with a speed error of about −0.063 m/s before recovery. An actuator limited to ±0.75 V could not maintain the stated final speed even though the assigned poles are stable. Integral feedback enforces the available equilibrium, and a command governor or anti-windup logic becomes necessary when the physical equilibrium lies outside the reachable input range.

The disturbance response also gives a way to separate a constant accelerative load from a sensor bias. Once a real opposing load has been rejected, an independent speed measurement still reads the requested 1.0 m/s while the final actuator voltage has risen by $-d/b = 0.20$ V above its no-load value. A constant positive sensor bias of 0.05 m/s instead drives the measured velocity to the set point with a true speed of 0.95 m/s. If both effects are allowed but only the controller’s own sensor is logged, a final reading exactly at set point does not identify the source of error. An independent tachometer or a measured load torque supplies the missing observation. Integral convergence cannot by itself validate sensor calibration.

Study the full derivation

Estimating velocity from position with a model-based observer · System dynamics, feedback and estimation

An optical encoder returns position and the actuator drive supplies a commanded acceleration. Reconstruct an unmeasured velocity without differentiating successive noisy samples. Derive the observer, explain what a pair of chosen poles guarantees, and predict the tradeoff visible when the encoder carries periodic interference.

Chapter 6 · Sampling, saturation and commissioning

Carry a continuous-time design onto a sampled actuator, predict where its operating assumptions fail, and distinguish an observed final value from an equilibrium.

FROM FIRST PRINCIPLES

Integrate a held command across one sample

A sampled actuator holds its command fixed until the next update. Consider a stable first-order plant $\tau\dot{y} + y = Ku$ with constant coefficients and sample interval h. Let y_k be the value at the start of interval k and u_k the command held throughout it. During that interval the equilibrium for the applied command is the constant Ku_k. Subtract it and integrate the homogeneous decay over exactly one interval.

$$\frac{d}{dt}(y - Ku_k) = -\frac{y - Ku_k}{\tau},$$
$$y_{k+1} - Ku_k = e^{-h/\tau}(y_k - Ku_k).$$

Set $a = e^{-h/\tau}$. The resulting discrete plant is an exact sample-to-sample model under a zero-order hold. For proportional feedback with unit reference scale, substitute $u_k = g(r_k - y_k)$ to see where the sampled free response pole lies.

$$y_{k+1} = ay_k + K(1 - a)u_k,$$
$$z_{\mathrm{cl}} = a - K(1 - a)g.$$

A sample-to-sample deviation is multiplied repeatedly by z_{cl}, so decay requires $|z_{\mathrm{cl}}| < 1$. This condition depends on the time at which the command is applied. A full sample of computational delay changes the state update and its characteristic equation.

Background · Discrete pole

$$z = e^{sT_s}$$

This maps a continuous modal exponent into a sampled multiplier over interval Ts. Different continuous frequencies can give the same sampled phase.

DERIVATION AND APPLICATIONS

Quantify the phase and attenuation introduced by a hold

A digital controller sends a piecewise-constant command to a continuous actuator. Even ideal sampling and instantaneous arithmetic do not make the resulting fundamental sinusoid identical to the requested

one. Consider samples of a complex sinusoid at period h. A zero-order hold outputs each sample unchanged over the following interval. For times $t = kh + s$ within interval k, the ratio of held signal to the desired continuous sinusoid is periodic in the within-sample time s.

$$u_h(kh + s) = e^{i\omega kh},$$
$$\frac{u_h(kh + s)}{e^{i\omega(kh+s)}} = e^{-i\omega s}, \quad 0 \le s < h.$$

The zero-frequency Fourier coefficient of this periodic ratio gives the component at the original sinusoidal frequency. The remaining coefficients produce spectral images spaced by the sample frequency. Integrate the ratio over one hold interval and factor out the midpoint phase.

$$H_0(i\omega) = \frac{1}{h}\int_0^h e^{-i\omega s}\,ds = e^{-i\omega h/2}\,\frac{\sin(\omega h/2)}{\omega h/2}.$$

Below Nyquist the sine factor is positive. The fundamental therefore has phase equivalent to a delay of half a sample and a frequency-dependent gain. For small ωh, the sine expansion makes the leading attenuation quadratic while phase lag is linear.

$$\arg H_0 = -\frac{\omega h}{2},$$
$$|H_0| = 1 - \frac{(\omega h)^2}{24} + O((\omega h)^4).$$

This calculation concerns the fundamental component of a held command. Additional computation latency and sensor filtering accumulate their own phase. Frequencies approaching Nyquist also require examining the sampled closed-loop dynamics directly.

WORKED EXAMPLE

Check a servo's phase allowance

At 50.0 Hz crossover and 1.00 kHz sampling,

$$\omega h/2 = \pi(50/1000) = 0.1571.$$

The hold contributes -9.00° and its fundamental magnitude is 0.9959. A further full sample of computation delay contributes -18.0°, giving -27.0° combined at that frequency. The small gain change would conceal this phase cost in a magnitude-only measurement.

Change the inputs and inspect this calculation below

SECOND APPLICATION

Choose a sampling interval from a phase budget

At a proposed crossover of 80.0 Hz, allocate at most 6.00° lag to the hold. Since the magnitude of the phase in degrees is $180fh$, the interval must satisfy $h \leq 6/(180 \times 80) = 0.417$ ms, or a sample rate of at least 2.40 kHz. At this limit the fundamental magnitude is

$$\sin(6^\circ)/(6\pi/180) = 0.9982.$$

The resulting condition addresses the hold alone.

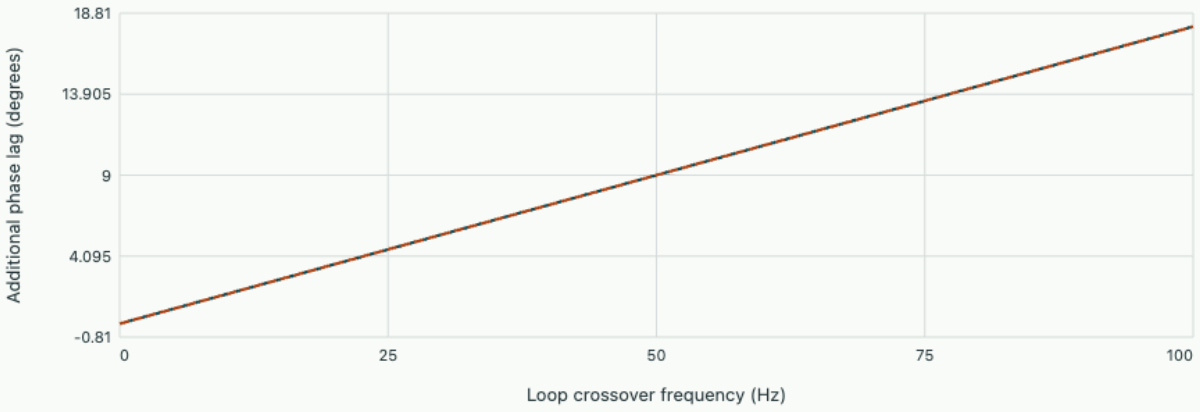
Interactive calculation · Change the inputs of this example

Use the adjustable quantities to test the two worked calculations above. The plot recomputes this derivation's balance or constitutive relation; its numerical readout reports the quantities needed for the forward calculation and the inverse or design question. Keep the fixed assumptions stated in the readout when comparing with the example values.

Open this live calculation

Try the calculation

Sampling frequency (Hz)	**1000**
Computation delay / sample	**0**



At50 Hz, additional lag=9.000° and hold magnitude=0.99589. Use only the first hold lobe for this phase interpretation.

Why one sample of latency can change the stability margin

A desktop position servo has a first-order response with time constant τ. A digital controller samples position every h seconds and holds each command constant until another is applied. Two firmware versions use the same sample period and proportional gain, yet one applies the command one whole sample later. Derive their closed-loop recurrences before attributing an observed oscillation to a different physical plant.

For the unit-gain plant $\tau\dot{x} + x = u$, solve the differential equation over a single interval with constant input u_k. The result is

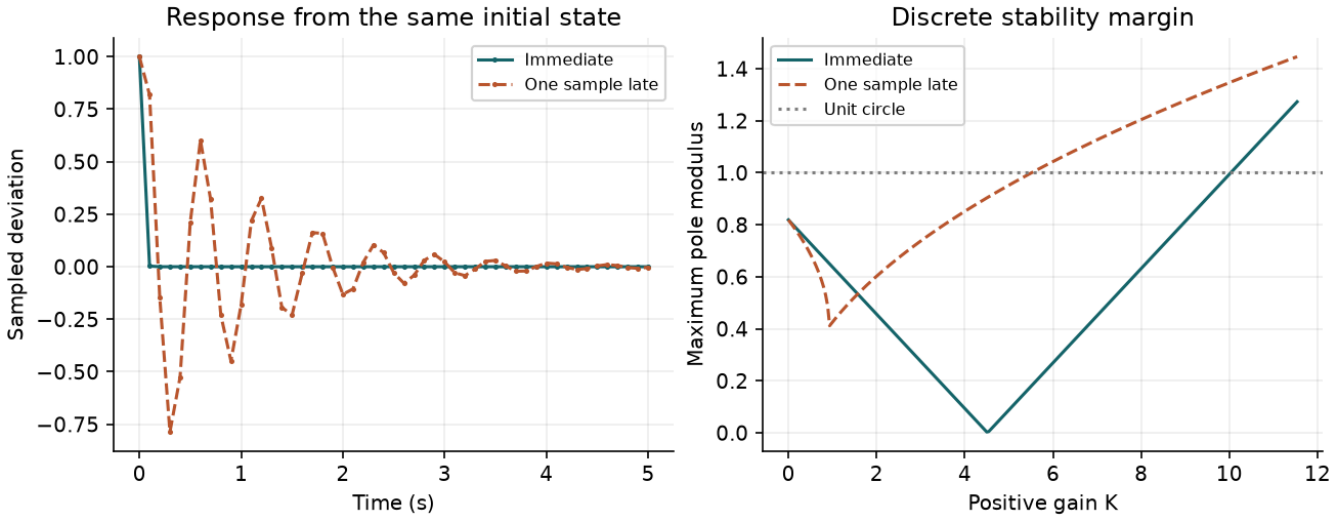
$$x_{k+1} = a x_k + b u_k,$$

where $a = e^{-h/\tau}$ and $b = 1 - a$. Under immediate regulation to zero the command is $u_k = -K x_k$. Then there is one closed-loop pole, $a - bK$, and the positive gain range is $0 \le K < (1 + a)/b$. At equality the pole sits at minus one and a step-to-step sign alternation persists.

With a one-sample delay the command is $u_k = -K x_{k-1}$. A trial solution $x_k = z^k$ now gives

$$z^2 - a z + bK = 0.$$

For $0 < a < 1$ and $K \ge 0$ the two roots both lie inside the unit circle exactly when $bK < 1$. One can see the source of the upper limit by following the roots. At low gain they are real, lie between zero and a, and meet when $bK = a^2/4$. Beyond that value they are conjugate and their common modulus is $\sqrt{bK}$, since their product equals bK.



The nearly deadbeat immediate update has one small pole. Adding a remembered sample gives two oscillatory poles for exactly the same plant and feedback gain.

Take $\tau = 0.50$ s, $h = 0.10$ s and $K = 4.5$. The exact hold gives $a = 0.81873$, $b = 0.18127$. Immediate application produces a pole near 0.0030, while the delayed poles are about $0.4094 \pm 0.8051i$, of modulus 0.903. Starting at unit displacement with the old stored sample equal to zero, the delayed plant first decays naturally. If the old sample had instead already equalled one, the first two data points would differ even though the subsequent recurrence and its poles were unchanged.

The complex argument of a delayed pole supplies another quantity to measure. At the default gain the pole angle is approximately 1.10 radians per sample. The damped oscillation therefore takes $2\pi/1.10$ samples per cycle, approximately 0.57 s of physical time. Its envelope shrinks to about $0.903^{10} = 0.36$ of its original value over one second. Estimate both quantities from a simulated sequence. Divide the measured period by the sampling interval before comparing with the pole angle. When the response is sampled sparsely, looking only at the step-to-step change of sign can obscure an oscillation that takes nearly six samples to repeat.

Choose a gain for which immediate feedback is stable and delayed feedback is unstable, and quantify the growth.

▼ **Work through the solution**

The two limiting gains are 10.033 and 5.517. At $K = 7$, the immediate pole is

$$0.81873 - 7(0.18127) \simeq -0.450.$$

The delayed conjugate pair has modulus $\sqrt{7(0.18127)} \simeq 1.126$, so its envelope grows by about 12.6% per sample. In the notebook, compare these predictions using the same physical-time horizon. If an actual processor has a fixed time delay, shortening h changes the delay in samples; the one-sample derivation must then be replaced with the actual timing model.

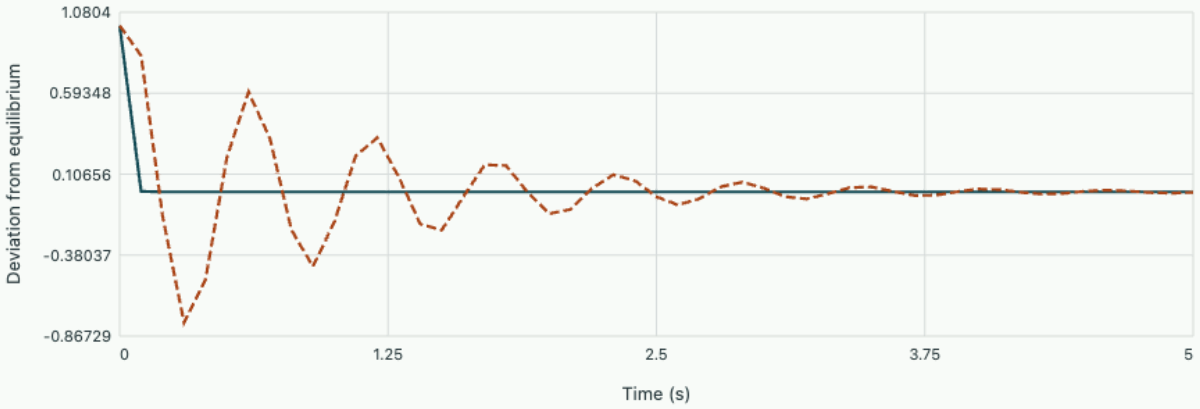
Interactive calculation · Sample delay motion

Vary the notebook inputs and compare both plotted observables. The same equations generate the browser curves and the notebook reference; fixed material, geometry and model assumptions are stated in the readout.

Open this live calculation

Try the calculation

Feedback gain	**4.5**
Plant time constant (s)	**0.5**
Sample period (s)	**0.1**



Immediate One sample late

Selected pole moduli: immediate 0.003, delayed 0.903. Immediate gain limit 10.033; delayed limit 5.517. Both share the same hold map; one extra sample changes the characteristic equation.

Download the editable notebook for this example

Interactive calculation · Sample delay poles

Vary the notebook inputs and compare both plotted observables. The same equations generate the browser curves and the notebook reference; fixed material, geometry and model assumptions are stated in the readout.

Open this live calculation

Try the calculation

Feedback gain	4.5
Plant time constant (s)	0.5
Sample period (s)	0.1



Selected pole moduli: immediate 0.003, delayed 0.903. Immediate gain limit 10.033; delayed limit 5.517. Both share the same hold map; one extra sample changes the characteristic equation.

Download the editable notebook for this example

Open editable notebook Download .ipynb

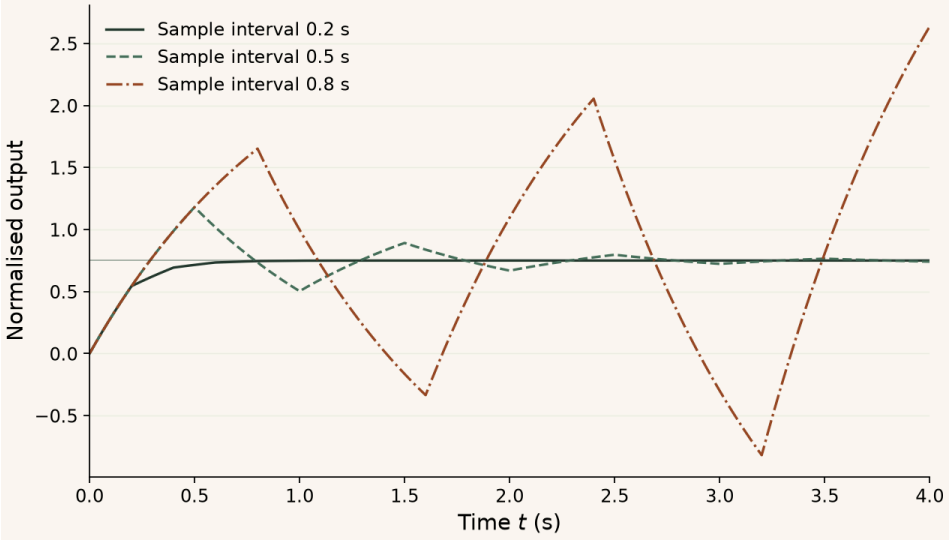
Derive the held-input model

A processor reads a sensor every h seconds and holds its most recent command until the next sample. Solve $\tau\dot{y} + y = Ku_k$ exactly from kh to $(k+1)h$ with this constant input. The result is

$$y_{k+1} = ay_k + K(1-a)u_k,$$

where $a = e^{-h/\tau}$. This recursion is exact at the sampling instants for the assumed hold. Taking a large Euler step through the continuous closed-loop equation describes a different experiment. A real controller has only the sampled information available when it computes u_k.

With proportional control $u_k = K_p(r_k - y_k)$ the discrete pole is $z = a - KK_p(1 - a)$. Stability requires $|z| < 1$. A negative pole of modulus less than one alternates sign each update while decaying. For $K = 1,\ \tau = 1$ s and $K_p = 3$, $h = 0.2$ s gives $z = 0.275$; $h = 0.8$ s gives $z = -1.203$. The same nominal continuous design gives radically different sampled records. For positive gain the upper gain limit is $(1 + a)/[K(1 - a)]$. Additional computation and transport delays lower the available margin further.



Same plant and gain, three sample intervals. The lines show the exact continuous output between updates; the held quantity is the actuator command.

Locate the sampling limit in a computed trace

Keep the plant integration at 40 RK4 substeps per hold. Predict which sample intervals give a monotone response, a decaying oscillation or a growing one. Compare the stored OneSimulation traces on either side of 0.69 s. Estimate the discrete pole from two successive output increments; their ratio removes the unknown equilibrium.

At 0.65 s the measured pole is −0.9118, while at 0.70 s it is −1.0137. The latter increasingly overshoots even though the internal integration resolves each hold. Try 0.69 s and 0.70 s with `python sampled-feedback.py --period 0.69`, then double `--substeps` to check that the stability change is a sampling effect. Download the runnable OneSimulation model.

Why a saturated integrator remembers the past

Let $v_k = K_p e_k + I_k$ be the requested command, and u_k its value clipped to the actuator range. An unconstrained integrator adds $hK_i e_k$ to I_k at every sample. While an unreachable command holds a heater at its upper limit, positive error keeps accumulating; when the command becomes reachable

again the stored integral may keep the heater fully on. This delayed recovery is called windup. Logging only actual heater power hides the growth of the internal demand. Log both signals and the integral state.

A back-calculation update feeds the clipping discrepancy back to the integrator:

$$I_{k+1} = I_k + h[K_i e_k + (u_k - v_k)/T_t].$$

When the request exceeds the upper stop, the second term is negative. At fixed error and fixed clipping level, this update attracts the requested command toward a value consistent with saturation, provided its discrete tracking dynamics are stable. For example, with $h = 0.1$ s, $K_i = 1$, $e = 2$, $u = 5$, $v = 8$, and $T_t = 0.5$ s, the integral increment is -0.4, compared with $+0.2$ without correction. Tuning T_t must account for the sample interval and actuator dynamics. Anti-windup cannot make an infeasible steady torque available.

Use a numerical result to distinguish two errors

Return to the OneSimulation tracking experiment. It models a continuous, unsaturated proportional loop and reports the error at a finite time. For a unit step from rest with loop gain $g = KK_p$, the error after N closed-loop time constants is $e_N = (1 + ge^{-N})/(1 + g)$. With $g = 8$ and $N = 6$, predict 0.11331. Its difference from the equilibrium $1/9$ is transient motion. Halving the plant speed changes the physical time needed to reach the same dimensionless sample, while this sampled error stays fixed.

Test the interpretation by recording at three and six time constants and taking the ratio of the two deviations from equilibrium. Predict e^{-3} for the later deviation divided by the earlier. A persistent discrepancy warrants checking the recorded timestamps, the implementation of the step and whether a limit was active. To examine sampling or clipping, explicitly add these mechanisms to the model before comparing with the equations above. The logged linear-loop runs answer the finite-record question directly.

Derive an exact sampled stability boundary with a delayed command

A digital controller may acquire a sample at the start of an interval but only deliver its computed command at the following update. The one-sample computation delay then introduces an additional dynamic state. Consider a first-order plant $\tau\dot{y} + y = Ku$ with an ideal zero-order hold and sampling period T. Its exact sample-to-sample relation is

$$y_{k+1} = ay_k + bu_k,$$

with $a = e^{-T/\tau}$ and $b = K(1 - a)$. Assume proportional negative feedback about a zero reference and express gains with dimensions chosen so that bg is dimensionless.

With negligible computation delay, $u_k = -gy_k$ yields one pole $a - bg$, so stability requires $|a - bg| < 1$. With a one-sample delay the applied command is $u_k = -gy_{k-1}$, giving $y_{k+1} = ay_k - bgy_{k-1}$. Substitution of a trial sequence z^k gives the characteristic polynomial $z^2 - az + bg$. For a real monic

quadratic $z^2 + c_1 z + c_0$, the unit-disc conditions are $1 + c_1 + c_0 > 0,\ 1 - c_1 + c_0 > 0,\ 1 - c_0 > 0$. For $0 < a < 1,\ b > 0,\ g > 0$ only the last condition limits increasing gain: $bg < 1$.

Once the delayed poles form a complex pair, their product equals their squared modulus. That makes the bound physically transparent: as bg reaches one the oscillation no longer decays between samples. This mechanism is invisible if a continuously implemented controller is analysed and merely evaluated at plotting intervals.

Compare two software schedules at the same sampling rate

Take $K = 1$,

$$\tau = 0.50\,\mathrm{s},$$
$$T = 0.10\,\mathrm{s}.$$

The exact coefficients are

$$a = 0.8187,\ b = 0.1813.$$

Without delay the positive gain limit is

$$(1 + a)/b = 10.03.$$

With a delayed command it is $1/b = 5.52$. A gain of 7.0 gives a stable immediate pole $0.8187 - 7(0.1813) = -0.450$. The negative sign predicts an alternating error whose sample values decrease in magnitude.

Apply the same gain through the delayed schedule. The characteristic polynomial becomes $z^2 - 0.8187z + 1.2689$. Its roots are complex with modulus $\sqrt{1.2689} = 1.126$, implying about 12.6% growth per sample in the envelope. An input limit may eventually bound the motion or produce a limit cycle, but it does not restore local linear stability around the unsaturated equilibrium. A controller test that records sensor samples and commanded values with separate timestamps can distinguish this delay from a sign error or an incorrectly discretised plant. Reduce the delayed gain to 4.0 and the predicted modulus becomes 0.852; the same clock can operate stably with a different response speed.

To recognise the delayed oscillation in a recorded trace, compute its angle as well as its modulus. At gain seven each real part is

$$a/2 = 0.4094.$$

Dividing by the modulus 1.126 gives $\cos\theta = 0.364$, so $\theta = 1.20\,\mathrm{rad/sample}$. The apparent oscillation period is about $2\pi(0.10)/1.20 = 0.52$ s. By comparison the immediate negative real pole produces a two-sample alternating pattern with period 0.20 s. A controller log showing roughly five samples per growing cycle accords with the delayed implementation, whereas alternating every sample points towards the immediate model operating too aggressively. This interpretation uses the synchronised

physical actuator output; a software variable computed today for application tomorrow may have a different timestamp from the hardware.

Why a fast vibration can appear as a slow feedback error

An analogue position sensor may detect a genuine low-frequency displacement and a much faster frame vibration. Sampling can make the two contributions indistinguishable. For a sampling interval $T_s = 1/f_s$, evaluate a complex sinusoid at sample n:

$$e^{i2\pi(f+kf_s)nT_s} = e^{i2\pi fnT_s}e^{i2\pi kn} = e^{i2\pi fnT_s}, \qquad k \in \mathbb{Z}.$$

A real cosine is also unchanged by reversal of frequency, so its apparent frequency can be folded into $0 \le f_a \le f_s/2$. At $f_s = 100\,\mathrm{Hz}$, a 92 Hz cosine gives precisely the samples of an 8 Hz cosine with reversed phase angle. If both are present, their sampled amplitudes combine according to phase. Removing 8 Hz in software would remove the true 8 Hz motion as well. Filter the continuous sensor signal before sampling if the higher component must be attenuated independently.

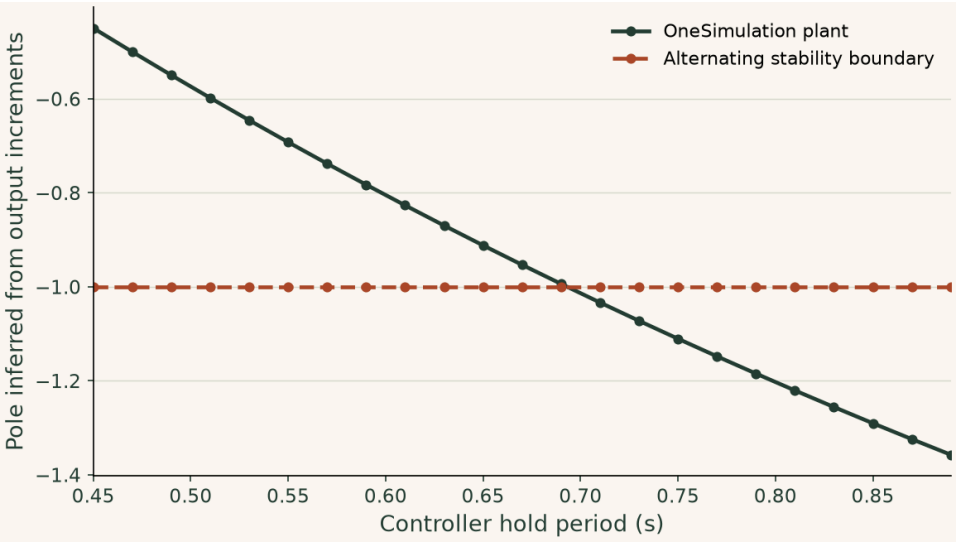
Suppose an encoder sees an unwanted sinusoidal displacement of amplitude $0.50\,\mathrm{mm}$ at 92 Hz, while feedback must track motion through 8 Hz. A fourth-order analogue Butterworth low-pass filter with cutoff 20 Hz has magnitude $|H(f)| = [1 + (f/20\,\mathrm{Hz})^8]^{-1/2}$. At 92 Hz the ratio is 0.00223, or approximately $-53.0\,\mathrm{dB}$: the folded displacement amplitude is only $1.12\,\mu\mathrm{m}$. At 8 Hz the ratio is 0.99967, so a $0.10\,\mathrm{mm}$ desired component retains about $99.97\,\mu\mathrm{m}$. These are separate amplitude tests on the analogue filter. If both vibrations persist, the sampled 8 Hz component can vary with their relative phase.

A passband magnitude close to one does not bound the filter's phase lag. A controller designed with substantial gain near 8 Hz must also include the filter's complex frequency response in its phase-margin calculation. Aliasing, sample-and-hold delay and computational delay are three different contributions to the observed loop dynamics.

Practise on sheet 3 · Choose an analogue filter order against a folding vibration. Further reading · University of Pennsylvania · ESE 1500 Digital Audio Basics · Lecture 5, Anti-Aliasing, Parts 1–2.

Find the stability boundary from three measured outputs

Sample the unit first-order plant three times with gain 3 and a constant reference. A one-pole response gives $\Delta y_{k+1} = z\Delta y_k$, so the ratio of increments identifies the discrete pole without knowing the final value. Here the plant uses forty RK4 substeps per controller interval.



The measured pole crosses -1 between 0.69 and 0.71 seconds. From the exact hold map $z = 4\exp(-h) - 3$, the critical period is $\ln 2$ seconds. A continuous plant integrated more accurately at the same hold period still sees the same delay in the controller update.

▶ **Read values from the plot**

Download the OneSimulation model (Python)

Interactive calculation · change the inputs of this example

Change controller gain and plant time constant, then find the sample interval where the discrete pole crosses minus one.

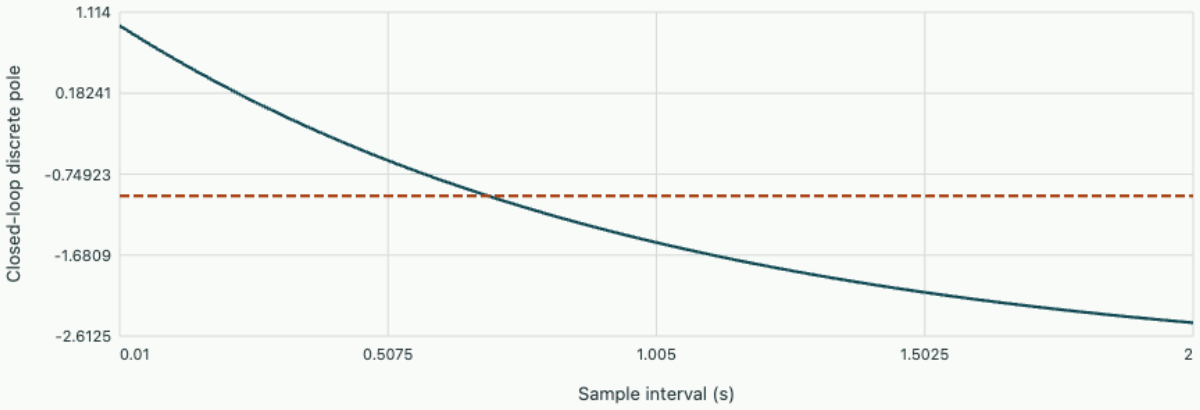
The curves below are recalculated from this section's equations in your browser. The fixed reference plot above remains available for comparison.

Open this live calculation

Try the calculation

Proportional gain 3

Plant time constant (s) 1



Exact zero-order hold Lower stability boundary

Stability requires sample interval below 0.693 s. More accurate plant integration does not remove the held-feedback limit.

Problem sheet 3 · What a controller can learn and deliver

Use rank tests, an exact sampled model and a finite-time numerical observation to diagnose three different limits.

1. A pair of decaying modes

Let $A = \operatorname{diag}(-1, -2),\ B = (1, 1)^{\mathsf{T}}$. (a) Test reachability. (b) Test observability for $C = (1, 1)$ and then for $C = (1, 0)$. (c) For zero input, recover the initial state from the first sensor when $y(0) = 3, \dot{y}(0) = -5$.

▼ **Use both direction and time history**

The reachability columns are $(1, 1)$ and $(-1, -2)$; their determinant is -1. The first observability matrix has rows $(1, 1), (-1, -2)$ and also has full rank. The second has rows $(1, 0), (-1, 0)$, hence never sees the second mode. Solving $x_1 + x_2 = 3,\ -x_1 - 2x_2 = -5$ gives

$$x_1 = 1,\ x_2 = 2.$$

In practice several noisy time samples are more reliable than a single numerical derivative.

---

2. A sampled proportional loop

A held-input plant has $K = 2,\ \tau = 1$ s. Use $K_p = 2$. (a) Find the largest sample interval for asymptotic stability. (b) Pick an interval giving a zero discrete pole. (c) For a unit step from rest, find the corresponding output at the first update and its equilibrium error.

▼ **Solve both bounds using the exponential**

The pole is $z = 5e^{-h} - 4$. For a positive sample interval it stays below 1; the lower bound requires $e^{-h} > 3/5$, giving $h < \ln(5/3) = 0.511$ s. Equality gives an undamped alternating mode. For $z = 0$, take $h = \ln(5/4) = 0.223$ s. The initial held input is 2, so the first output is

$$2(1 - 0.8)2 = 0.8.$$

It stays there on subsequent samples, leaving error 0.2. A fast transient and accurate steady tracking are independent specifications.

---

3. Explain a numerical tracking record

Use the feedback experiment. (a) Predict the output errors at three and six closed-loop time constants for $K = 2, K_p = 4$. (b) Predict which values change if the plant time constant doubles. (c) A drive limited to 2 command units shows a different early trace after a unit step. Identify a signal that could confirm the cause before changing gains.

▼ **Compare a finite observation with the steady balance**

Here $e_N = (1 + 8e^{-N})/9$. The two predictions are 0.15537 and 0.11331; the steady error is 0.11111. Doubling τ doubles the times in seconds while preserving those two dimensionless observations. At switch-on the controller asks for $K_p e = 4$, twice the limit. Plot requested and delivered command on the same axes and identify the interval when they differ. The unsaturated model still supplies a valid comparison after checking the actual initial condition for recovery.

Choose an analogue filter order against a folding vibration

A sampled displacement loop runs at 100 Hz. Genuine motion below 10 Hz must retain at least 99% of its amplitude. A 78 Hz frame oscillation must be attenuated to at most 2% before sampling. With a 20 Hz cutoff, decide whether a second- or third-order Butterworth filter meets both amplitude conditions. Determine where the residual vibration appears after sampling.

▼ **Show the worked solution**

For order n, use $G(f) = [1 + (f/20)^{2n}]^{-1/2}$. At 10 Hz the ratios are 0.9701 and 0.9923 for orders two and three. At 78 Hz they are 0.0656 and 0.0169. Thus the third-order candidate meets both requirements. The second-order candidate misses both. The vibration folds to $100 - 78 = 22\,\mathrm{Hz}$. Even though it lies outside the specified tracking band after sampling, it can still enter sampled diagnostics or excite a controller with appreciable gain there.

Revisit chapter 6 · Why a fast vibration can appear as a slow feedback error.

Explore the governing equations after solving the sheet: States, reachability and estimation; Sampling, saturation and commissioning. Compare the assumptions and units before transferring your chosen parameters; the controls may cover a different numerical range from the written exercise.

Notebook practice · Why one sample of latency can change the stability margin

Chapter 7 · Following poles as a controller gain changes

Use the geometry of the characteristic equation to discover which transient specifications one adjustable gain can meet.

FROM FIRST PRINCIPLES

Derive the angle and gain conditions for a root locus

For a linear negative-feedback system with open-loop transfer $KG(s)$, the free closed-loop motion has a nonzero amplitude only when the characteristic factor vanishes. Treat K as a real nonnegative adjustable parameter. A candidate pole s is on the root locus precisely when the complex number $KG(s)$ equals the real number -1.

$$1 + KG(s) = 0,$$
$$|KG(s)| = 1,$$
$$\arg G(s) = (2j+1)\pi.$$

Factor the fixed transfer into real positive leading gain g, zeros z_i and poles p_j. Multiplication adds complex angles and division subtracts them. The resulting geometric test and gain value are

$$\sum_i \arg(s - z_i) - \sum_j \arg(s - p_j) = (2\ell + 1)\pi,$$
$$K = \frac{\prod_j |s - p_j|}{g \prod_i |s - z_i|}.$$

A point can have an attractive decay rate or oscillation frequency yet fail the angle condition for every positive gain. Once that condition holds, the distance products give the unique gain at an ordinary, finite point of the locus. At zero gain the branches start at the open-loop poles by continuity.

DERIVATION AND APPLICATIONS

Predict where a root-locus branch approaches its asymptote

A pole sketch can show the direction in which a gain change sends fast modes. Consider real-coefficient negative feedback with characteristic equation $D(s) + KN(s) = 0$. Take monic polynomials with n open-loop poles p_j, m finite zeros z_j, $n > m$, and positive adjustable gain K. Include repeated roots with their multiplicities and absorb any constant numerator factor into the gain. For large $|s|$, expand the product of root factors and divide the leading terms.

$$\frac{D(s)}{N(s)} = s^{n-m}\left[1 - \frac{\sum_j p_j - \sum_j z_j}{s} + O(s^{-2})\right].$$

Write $q = n - m$. A shifted power $(s - \sigma)^q$ has the same first two large-s coefficients if $q\sigma$ equals the difference of root sums. The remaining characteristic equation for distant branches then reduces to a power equal to negative real gain.

$$\sigma = \frac{\sum_j p_j - \sum_j z_j}{q}, \qquad (s - \sigma)^q \simeq -K.$$

Taking the q roots gives the asymptote directions. Complex poles and zeros come in conjugate pairs, so the centroid σ lies on the real axis. These are asymptotic guides; a branch at finite gain generally curves towards its line.

$$s \simeq \sigma + K^{1/q} \exp\left[i\frac{(2j+1)\pi}{q}\right],$$
$$j = 0, \ldots, q-1.$$

WORKED EXAMPLE

Identify a high-gain difficulty in a three-lag loop

For poles at $-1, -2, -6$ s⁻¹ and no finite zeros, $q = 3$ and $\sigma = -3$ s⁻¹. The three angles are 60°, 180° and 300°. Two branches ultimately enter the right half-plane. As a check the exact polynomial is $s^3 + 9s^2 + 20s + 12 + K$; cubic stability demands $9(20) > 12 + K$, or $K < 168$. The sketch predicted a finite gain limit and the polynomial supplies its exact value.

Change the inputs and inspect this calculation below

SECOND APPLICATION

Recover a missing zero from measured asymptotes

A separate system has poles at $0, -4, -8$ s⁻¹ and one real zero. Its two high-gain branches approach vertical lines with real coordinate -4.50 s⁻¹. The centroid equation gives $2(-4.50) = -12 - z$, hence $z = -3.00$ s⁻¹. The remaining finite branch tends to that zero as gain grows. Vertical asymptotes alone identify the pole excess; their real intercept supplies the missing location.

Interactive calculation · Change the inputs of this example

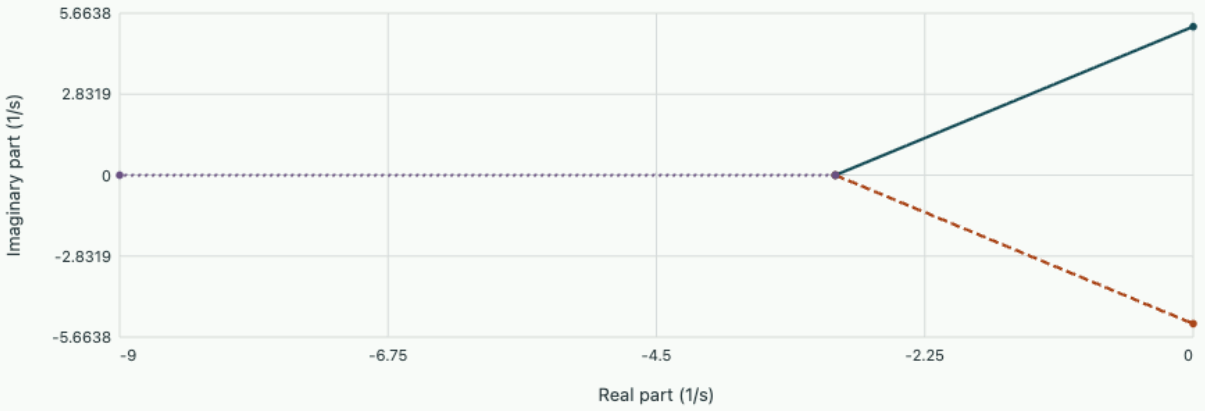
Use the adjustable quantities to test the two worked calculations above. The plot recomputes this derivation's balance or constitutive relation; its numerical readout reports the quantities needed for the forward calculation and the inverse or design question. Keep the fixed assumptions stated in the readout when comparing with the example values.

Open this live calculation

Try the calculation

Third open-loop pole rate (1/s) **6**

High-gain radial distance **6**



Asymptote centroid -3.000 /s. Exact cubic stability limit K<168.000 /s³ for poles −1,−2,−6. Straight rays describe high gain, not the finite-gain exact locus.

Start from this chapter. Follow the single proportional-feedback pole as gain varies. Use the sampled step to distinguish its settling rate from the remaining steady tracking error.

OneSimulation laboratory · Predict the settling time and offset of a feedback loop

A family of closed-loop systems

The earlier pole calculations chose a gain and found the poles. Reverse that question: if an engineer turns one gain dial continuously, which parts of the complex plane can the poles ever visit? For unity negative feedback write the open-loop transfer as $KN(s)/D(s)$, with N, D coprime and $K \geq 0$. The closed-loop characteristic equation is

$$D(s) + KN(s) = 0.$$

At a trial complex point away from an open-loop zero, the required gain is $-D(s)/N(s)$. The point is attainable only if this number is real and nonnegative. Its magnitude gives the gain; its phase supplies a separate feasibility test.

At $K = 0$ branches begin at open-loop poles. At very large gain, finite roots must approach zeros of N. If the denominator has greater degree, the remaining roots go to infinity. Since a real-coefficient polynomial has conjugate roots, the locus is symmetric about the real axis. These observations put strong constraints on a sketch even before computing roots.

Derive a breakaway and read the transient

Take a positioning plant $P = 1/[s(s+4)]$ with proportional gain K. The closed-loop poles are $s = -2 \pm \sqrt{4 - K}$. They start at 0 and –4, approach each other along the real axis and meet at –2 when $K = 4$. Equivalently the real-axis gain $K(s) = -s(s+4)$ has a stationary point at –2. Above the breakaway the complex pair moves vertically, so raising gain increases oscillation frequency while leaving the exponential envelope e^{-2t} unchanged.

The constant term gives $\omega_n = \sqrt{K}$ and the linear coefficient gives

$$\zeta = 2/\sqrt{K}.$$

A specification of decay rate 3 s⁻¹ and damped frequency 4 rad/s would require poles $-3 \pm 4i$. Since every complex pole here has real part –2, no choice of proportional gain meets both requirements. Failure to find such a gain numerically is a statement about controller structure.

Add a second degree of freedom

Feed back measured velocity as well as position:

$$u = K_p(r - y) - K_v\dot{y}.$$

The closed-loop denominator becomes $s^2 + (4 + K_v)s + K_p$; the reference numerator is just K_p. For desired poles $-3 \pm 4i$ the polynomial is $s^2 + 6s + 25$, hence

$$K_v = 2,\ K_p = 25.$$

The derivative acts only on measured output, so a discontinuous reference does not directly request an impulse. Differentiating a noisy measurement can nevertheless demand excessive actuator effort. A separate velocity sensor or a filtered estimate changes the high-frequency dynamics and must be included in the design check.

Include the numerator when interpreting a pole plot

Moving a controller zero also reshapes the locus because it changes the angle of N/D. A zero in the reference-to-output map changes step-response residues even if the poles coincide with a previous design. In particular a positive real zero imposes a severe tracking constraint. For the stable unit-DC transfer $G = (1 - s/z)/(1 + s\tau)^2$ with $z > 0$, the unit-step initial slope is $\lim_{s\to\infty} sG(s) = -1/(z\tau^2)$. Yet the final output is positive. The initial motion in the wrong direction follows directly from the zero. Increasing gain to make the initial transient faster cannot remove that geometry.

Place a controller zero from a specified root

A proportional controller moves poles along the root locus of its plant. A chosen design point off that locus can sometimes be made reachable with a controller zero. For a unity-feedback plant $P(s) = 1/[s(s + 2)]$, take an ideal proportional–derivative compensator $C(s) = K(s + z)$ with real $K, z > 0$. The locus condition $1 + C(s)P(s) = 0$ becomes

$$K = -s(s + 2)/(s + z).$$

For a desired nonreal pole s_d, this ratio must be real and positive. This single complex constraint is equivalent to one phase and one magnitude condition.

Choose $s_d = -3 + 4i\,\mathrm{s}^{-1}$, so the target characteristic polynomial is $s^2 + 6s + 25$. Coefficient matching is particularly transparent for this second-order plant: the characteristic polynomial is $s^2 + (2 + K)s + Kz$. Equating coefficients gives $K = 4$ and $z = 6.25\,\mathrm{s}^{-1}$. Geometrically the zero at −6.25 supplies precisely the missing phase at the specified point. Substituting s_d in $-s(s + 2)/(s + 6.25)$ recovers positive real gain 4, an independent check that this is the positive-gain branch.

The target pair has $\omega_n = 5\,\mathrm{rad/s}$ and $\zeta = 0.60$. The ideal controller contributes derivative gain 4 and proportional gain 25 in consistent normalised units. The closed-loop reference numerator contains the zero, so predicting exact overshoot requires the complete response even though the poles now match the design point.

Determine the consequence of giving differentiation a finite bandwidth

The ideal derivative gain grows without bound at high frequency. A real implementation can use $C_f = 25 + 4s/(1 + s/20)$, placing a filter pole at −20 s⁻¹ on the derivative path. Combining terms gives

$$C_f = (25 + 5.25s)/(1 + s/20),$$

which shifts the controller zero to $-25/5.25 = -4.76\,\mathrm{s}^{-1}$. The characteristic equation of the actual loop is

$$s(s+2)(1+s/20)+25+5.25s=0,$$

$$s^3+22s^2+145s+500=0.$$

The target point no longer satisfies it: inserting $-3+4i$ gives $28+96i$, a nonzero residual in the monic characteristic polynomial. The implementation remains stable by the cubic Routh test because all coefficients are positive and $22(145)>500$, but its exact transient differs. The high-frequency controller gain approaches $25+4(20)=105$, now finite and directly relevant when estimating actuator activity from encoder noise. Placing the filter much farther left reduces the difference near the designed poles while increasing this asymptotic noise gain. A design specification with both pole placement and an allowable actuator-noise level must therefore include the filter during synthesis rather than adding it only in software.

A reference ramp distinguishes this controller from a high static-gain proportional loop that happens to yield similar dominant poles. With unity feedback and plant integrator, the ramp error to $r(t)=vt$ tends to v/K_v provided the closed loop is stable, where $K_v=\lim_{s\to 0} sC(s)P(s)$. Both ideal and filtered designs have $C(0)=25$, giving $K_v=25/2=12.5\,\mathrm{s}^{-1}$. A 5 mm/s position ramp therefore leaves a 0.40 mm asymptotic tracking error. Moving the derivative filter changes fast behaviour while leaving this low-frequency comparison unchanged. If a much smaller ramp error is required, speed feedforward or additional low-frequency loop gain must be assessed separately; the selected complex poles by themselves do not establish that tracking specification.

Chapter 8 · Uncertainty envelopes and robust stability

Turn a bounded discrepancy between plant and model into a frequency-by-frequency test of the feedback design.

FROM FIRST PRINCIPLES

Derive a feedback bound from a fractional plant error

A frequency-dependent model error matters through the feedback path in which it enters. Suppose an actual plant can be represented as $P = P_0(1 + W\Delta)$, with nominal plant P_0, a known stable weight W and stable uncertainty satisfying $|\Delta(i\omega)| \le 1$. Use a fixed controller C and assume the nominal interconnection is internally stable. Define $L_0 = CP_0$ and expand the perturbed characteristic factor.

$$1 + CP = 1 + L_0 + L_0 W\Delta = (1 + L_0)(1 + T_0 W\Delta),$$
$$T_0 = \frac{L_0}{1 + L_0}.$$

The second factor describes how the unknown part feeds back around the nominal closed loop. If the magnitude of its variable term stays strictly below one at every frequency, it cannot reach the critical value -1. The small-gain condition for stable interconnected factors is

$$\sup_{\omega} |T_0(i\omega)W(i\omega)| < 1.$$

This gives a sufficient certificate for every uncertainty in the specified class. The bound permits less modelling error where the complementary sensitivity is large. A failed inequality identifies a missing certificate; determining actual instability requires the particular plant error or a tighter description of it.

Background · Sensitivity

$$S(s) = 1/[1 + L(s)]$$

For a scalar unity negative-feedback loop this gives one closed-loop transfer. The point where a disturbance enters determines which transfer the observer measures.

DERIVATION AND APPLICATIONS

Represent a transport-delay error as multiplicative uncertainty

A frequency sweep of a conveyor or thermal transport path may fit the gain well while leaving an uncertain delay. Suppose the actual stable plant is $P(s) = P_0(s)e^{-s\Delta t}$, with a nominal model containing the accurately known part of the delay. Let the extra delay lie between zero and d. At

sinusoidal frequency ω, define the fractional model error as $E = P/P_0 - 1$. Factoring the difference of the two unit phasors gives

$$E(i\omega) = e^{-i\omega\Delta t} - 1 = -2i\, e^{-i\omega\Delta t/2} \sin(\omega\Delta t/2).$$

This is a chord on the unit circle. Its magnitude grows with delay until the phasors point in opposite directions. Maximising over the stated delay interval therefore gives a monotone envelope which stays at two after the interval first contains a half-cycle delay.

$$W_d(\omega) = \max_{0 \le \Delta t \le d} |E| = \begin{cases} 2\sin(\omega d/2), & \omega d \le \pi, \\ 2, & \omega d \ge \pi. \end{cases}$$

For nominal loop $L_0 = CP_0$ the perturbation multiplies the complementary sensitivity $T_0 = L_0/(1 + L_0)$ in the characteristic equation. If the nominal closed loop is internally stable and the bounded error is stable, keeping the product below unity at all frequencies gives a sufficient robust-stability test.

$$1 + CP = (1 + L_0)(1 + T_0 E), \qquad |T_0(i\omega)| W_d(\omega) < 1.$$

At low frequencies the envelope approaches ωd. This dependence on frequency explains how an almost exact steady gain can coexist with a substantial feedback discrepancy near crossover.

WORKED EXAMPLE

Audit a transport path at one frequency

A thermal experiment bounds extra delay by 20.0 ms. At $\omega = 30.0$ rad/s the worst relative error is $2\sin(0.300) = 0.591$. If $|T_0| = 1.40$, the pointwise product is 0.827, within the bound at that frequency. Repeating at every relevant frequency is necessary for the full small-gain conclusion; this single value establishes only the local allowance.

Change the inputs and inspect this calculation below

SECOND APPLICATION

Derive a timing requirement from a measured peak

At 40.0 rad/s another loop has $|T_0| = 2.00$. The allowable envelope must be below 0.500. On the rising branch the maximum extra delay is therefore

$$d < 2\arcsin(0.250)/40.0 = 12.6\,\text{ms}.$$

A measured 15.0 ms spread would give $4\sin(0.300) = 1.18$ for the product, so this sufficient certificate would no longer settle stability there.

Interactive calculation · Change the inputs of this example

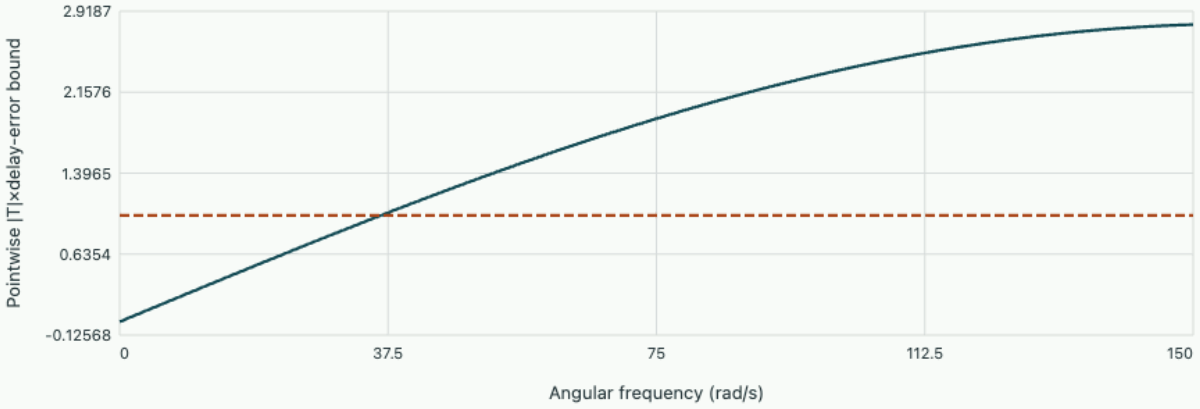
Use the adjustable quantities to test the two worked calculations above. The plot recomputes this derivation's balance or constitutive relation; its numerical readout reports the quantities needed for the forward calculation and the inverse or design question. Keep the fixed assumptions stated in the readout when comparing with the example values.

Open this live calculation

Try the calculation

Maximum extra delay (ms) **20**

Measured complementary gain **1.4**



Specified magnitude times delay envelope Small-gain threshold

At30 rad/s the pointwise product is 0.827. Holding |T| constant here is a local-envelope experiment; a real robustness test needs its measured frequency dependence.

What do several plant tests tell us?

The gain of a drive changes as it warms up, and a flexible mounting introduces poorly measured phase lag at high frequency. Let P_0 be a nominal model. One useful description is multiplicative uncertainty, $P = P_0(1 + W\Delta)$, with stable dimensionless Δ of gain at most one and a specified frequency weight W. The magnitude of W bounds relative error. Estimate it from $|P/P_0 - 1|$ across an ensemble of measurements, retaining phase in the complex ratio. A 20% amplitude difference and a 20° phase difference generally produce a larger relative error than either alone.

Where P_0 is close to zero this relative description can be poorly conditioned. An additive uncertainty $P = P_0 + W_a\Delta$ can describe an absolute error floor there. A bound fitted only at sampled frequencies requires a reason to hold between samples, especially near flexible resonances. A robustness calculation is conditional on this modelling work.

Factor the perturbed characteristic equation

For the nominal controller C, set $L_0 = P_0 C$ and $T_0 = L_0/(1 + L_0)$. Under multiplicative perturbation,

$$1 + PC = (1 + L_0)(1 + W T_0 \Delta).$$

Assume internal stability of the nominal interconnection and stable, proper uncertainty factors with no hidden unstable cancellation. If $\sup_\omega |W(i\omega)T_0(i\omega)| < 1$, the perturbation loop has gain less than one and cannot destabilise the feedback. One way to see the mechanism is the convergent inverse series $1 - G + G^2 - \cdots$ for $G = WT_0\Delta$ in the stable-system norm. The pointwise bound protects against all allowed dynamic phase choices. Failure of a sufficient test alone does not prove the particular measured plant unstable.

Work a frequency-domain budget

At one frequency a nominal loop has $L_0 = -0.8 + 0.4i$. Then $|1 + L_0| = \sqrt{0.2}$, $|L_0| = \sqrt{0.8}$, and $|T_0| = 2$. Relative uncertainty below 50% at that frequency keeps $|WT_0| < 1$ there. A measured 60% bound crosses this threshold. Geometrically the perturbed loop samples a disk centred at L_0 of radius $|L_0 W|$. At 50% its rim touches -1. Other frequencies still require examination, and a physically narrower family such as a single real gain may occupy far less than the full disk.

Add a performance requirement

A stable loop can still amplify an output disturbance unacceptably. For the uncertain plant the sensitivity is $S = S_0/(1 + WT_0\Delta)$. The triangle inequality gives $|S| \le |S_0|/(1 - |WT_0|)$ where the denominator is positive. If $|S_0| = 0.10$ and $|WT_0| \le 0.25$, disturbances can be amplified by at most 0.133 in this bound. Very near the robustness limit that bound becomes large. Shaping the nominal loop for a small uncertainty product near troublesome resonances can therefore protect both stability and useful performance. Record the sensor configuration, loading and frequency band over which the underlying plant measurements were made.

Distinguish absolute from relative uncertainty

Several tested plants may differ by an approximately fixed absolute transfer at high frequency although their nominal output gain is nearly zero there. Dividing each difference by a near-zero nominal transfer gives an enormous relative error and may make a multiplicative envelope awkward. An additive description writes $P_a = P_0 + W_a\Delta$, where $|\Delta(i\omega)| \le 1$, the stable filter W_a has the same units as the plant and its magnitude bounds absolute discrepancies. The controller C and nominal plant first form an internally stable feedback system. Assume the uncertainty interconnection is well posed and the normalised uncertainty is stable.

Factor the uncertain return difference using the nominal sensitivity $S_0 = 1/(1 + CP_0)$:

$$1 + CP_a = (1 + CP_0)[1 + W_a CS_0 \Delta].$$

For unstructured complex uncertainty the small-gain bound $\|W_a CS_0\|_\infty < 1$ is sufficient for robust stability. The relevant nominal transfer is from measurement error to control effort, CS_0, rather than complementary sensitivity alone. At high frequency S_0 often approaches one, so a derivative action that increases high-frequency controller gain can stress an additive uncertainty bound even where nominal loop gain has already fallen.

The inequality uses the worst permitted complex phase at each frequency. A point violating the envelope test indicates loss of this guarantee. It does not by itself establish that a particular measured plant is unstable. Conversely a few passing frequency samples do not certify the continuous-frequency supremum.

Budget uncertainty for a lightly coupled actuator resonance

Suppose a calibration expresses position in millimetres and actuator input in volts. At 40 rad/s the nominal gain is only $0.020\,\mathrm{mm/V}$, while an unmodelled flexible path can add up to $0.030\,\mathrm{mm/V}$ in any phase. A measured nominal loop has $|CS_0| = 18\,\mathrm{V/mm}$ there, giving an uncertainty product $0.030(18) = 0.54$. The corresponding allowable absolute uncertainty before the unit threshold is

$$1/18 = 0.0556\,\mathrm{mm/V}.$$

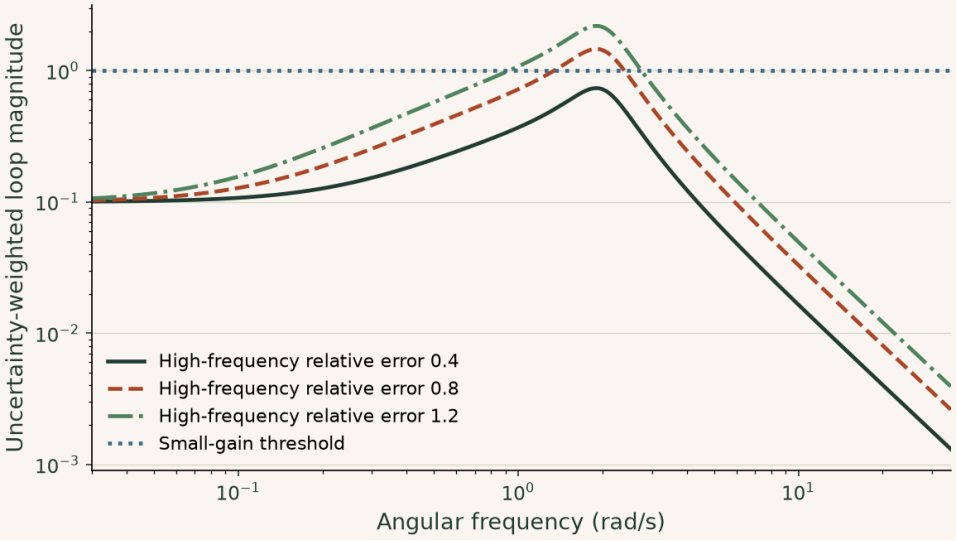
Stating only the 150% relative plant discrepancy would conceal the controller-dependent margin at that frequency.

At 70 rad/s suppose the absolute uncertainty bound increases to $0.045\,\mathrm{mm/V}$ and $|CS_0| = 25\,\mathrm{V/mm}$. Their product is 1.125. To reserve a factor of 1.25 on the uncertainty envelope, impose a target product of at most 0.80. Then $|CS_0|$ must be at most $0.80/0.045 = 17.8\,\mathrm{V/mm}$ at that frequency. A revised compensator with measured $|CS_0| = 16$ gives 0.72 at the suspect point. Check the complete response around any new filter resonance and recompute nominal tracking. Reducing this single frequency peak cannot establish whether the remainder of the frequency band also satisfies the reserve.

Suppose the absolute uncertainty envelope comes from three nominally identical machines. At each frequency calculate the complex difference of each measured response from the nominal complex response before taking magnitudes. At 70 rad/s, two transfer measurements of equal magnitude but opposite phase can differ by twice that magnitude. Subtracting magnitudes alone would assign zero discrepancy. Also retain independent test uncertainty such as force calibration and motion-sensor noise, which may be especially important where nominal motion is small. The largest observed discrepancy only covers those three machines and operating conditions. A justified production bound requires considering the frequency regions between measurement points and whether later mounting or payload changes can introduce another flexible path.

Test a frequency-dependent uncertainty budget

Use the same nominal loop and model multiplicative uncertainty by the weight $W(s) = (0.1 + as)/(1 + s)$ times a stable unit-gain perturbation. Inspect the product $|WT|$ rather than the two factors separately. A uniform bound below one supplies a sufficient robust-stability test.



The largest products occur near the nominal resonance, not at the largest plotted frequency where |T| is already small. A curve crossing unity means this envelope test no longer certifies the whole allowed family; a particular tested plant may still be stable.

▶ **Read values from the plot**

Interactive calculation · change the inputs of this example

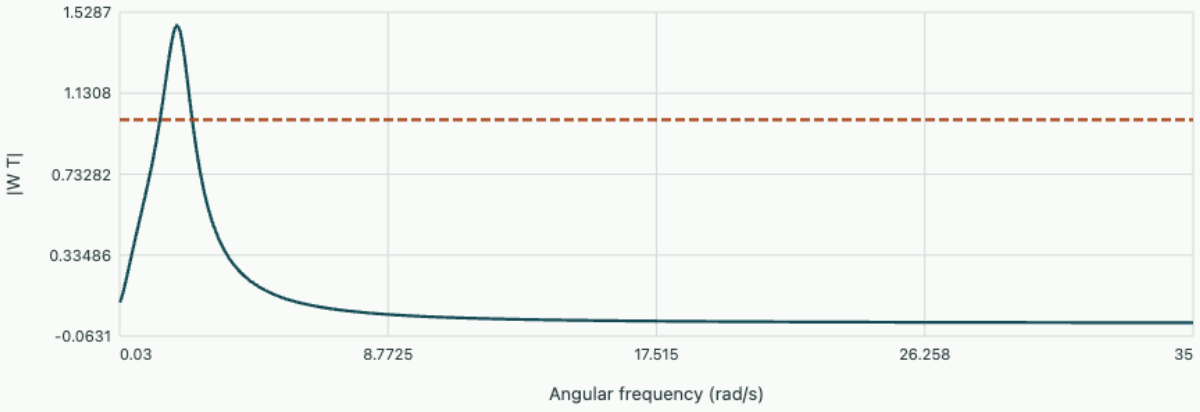
Increase the high-frequency uncertainty budget and locate the maximum weighted response over the plotted band.

The curves below are recalculated from this section's equations in your browser. The fixed reference plot above remains available for comparison.

Open this live calculation

Try the calculation

Loop gain	**4**
High-frequency relative error	**0.8**
Low-frequency relative error	**0.1**



Maximum over the sampled band 1.463. Crossing one removes this sufficient envelope certificate; it does not classify every particular plant.

Study the full derivation

Sensitivity, bandwidth and the integral price of disturbance rejection · System dynamics, feedback and estimation

A regulator attenuates a slowly varying output load yet amplifies another disturbance at a higher frequency. Separate the paths by which reference, measurement noise and physical disturbance reach the recorded variable, then derive an integral identity that explains why a stable loop can redistribute sensitivity across frequency.

Problem sheet 4 · Pole geometry and plant uncertainty

Find which poles are achievable, then test what happens when the plant deviates from the nominal design.

1. A position loop with one gain

Take $P = 1/[s(s+6)]$ and $u = K(r-y)$. (a) Find the breakaway gain and pole location. (b) Choose the gain that gives damped frequency 4 rad/s. (c) Add velocity feedback and select both gains for poles at $-4 \pm 3i$.

▼ **Compare the desired monic polynomial with the plant**

The roots are $-3 \pm \sqrt{9-K}$; they meet at -3 for $K = 9$. A damped frequency of 4 requires $K = 25$, yielding poles $-3 \pm 4i$. The second desired pair has polynomial $s^2 + 8s + 25$. With $u = K_p(r-y) - K_v\dot{y}$ its coefficients give

$$K_v = 2,\ K_p = 25.$$

Both designs have the same natural frequency but different envelopes and peak times.

---

2. An initial movement in the wrong direction

Let $G = (1 - s/2)/(1+s)^2$. (a) Determine the final response and initial slope for a unit step. (b) Obtain the full step response. (c) Find the time and value of its minimum.

▼ **Resolve the step into a repeated-pole response**

The DC value is 1 and $\lim_{s\to\infty} sG = -1/2$, giving the initial slope. Partial fractions of G/s yield

$$y = 1 - e^{-t} - 1.5te^{-t}.$$

Its derivative is $(-0.5 + 1.5t)e^{-t}$; the minimum occurs at $t = 1/3$ s and equals

$$1 - 1.5e^{-1/3} \simeq -0.0748.$$

It is a zero-induced inverse response in a stable system.

---

3. How much complex model error fits?

At one frequency $L_0 = -0.6 + 0.3i$. (a) Compute $|S_0|, |T_0|$. (b) Test a multiplicative uncertainty radius of 40%. (c) Estimate a worst-case upper bound on uncertain sensitivity there. Explain what remains to establish robust stability.

▼ Measure the distance to the critical point

The distance $|1 + L_0| = 0.5$ gives $|S_0| = 2$ and $|T_0| = \sqrt{0.45}/0.5 = 1.342$. The uncertainty product is about 0.537, below one here. The triangle bound yields

$$|S| \leq 2/(1 - 0.537) = 4.32.$$

Check the full frequency band, nominal internal stability and the properties of the uncertainty class before drawing a loop-wide conclusion.

Explore the governing equations after solving the sheet: Following poles as a controller gain changes; Uncertainty envelopes and robust stability. Compare the assumptions and units before transferring your chosen parameters; the controls may cover a different numerical range from the written exercise.

Chapter 9 · Optimal state feedback and the Riccati equation

Choose how heavily to penalise tracking error and actuator effort, then derive the feedback gain that minimises that stated cost.

FROM FIRST PRINCIPLES

Derive a feedback gain by minimising future cost

A control objective must say what counts as expensive. Consider a known continuous linear system $\dot{x} = Ax + Bu$ with state penalty $x^{\mathsf{T}}Qx$ and effort penalty $u^{\mathsf{T}}Ru$, where Q is positive semidefinite and R positive definite. Assume the usual stabilisability and detectability conditions for the infinite-horizon problem. If $V(x)$ is the least remaining accumulated cost, minimising over one short interval and then using the optimal continuation gives

$$0 = \min_{u}\{x^{\mathsf{T}}Qx + u^{\mathsf{T}}Ru + \nabla V^{\mathsf{T}}(Ax + Bu)\}.$$

Take the quadratic value form $V = x^{\mathsf{T}}Px$ with symmetric P. Differentiate the expression inside braces with respect to u. The positive effort penalty makes its stationary value a minimum.

$$2Ru + 2B^{\mathsf{T}}Px = 0,$$

$$u_{*} = -R^{-1}B^{\mathsf{T}}Px.$$

Substitute this control back into the optimality condition and require the quadratic state coefficient to vanish for every state.

$$A^{\mathsf{T}}P + PA - PBR^{-1}B^{\mathsf{T}}P + Q = 0.$$

The stabilising solution gives the feedback gain and the predicted least cost from the initial state. Hard limits on actuator effort require a further model constraint.

DERIVATION AND APPLICATIONS

Relate an optimal feedback value function to energy spent

The Riccati solution predicts more than a feedback gain. Consider a scalar integrator $\dot{x} = bu$ with $b > 0$ and a cost accumulated over an unlimited horizon. Use positive weights q, r and charge both state deviation and control amplitude squared. Assume the actuator has no amplitude or rate limit and the chosen stabilising control returns the state to zero. A quadratic candidate value $V = Px^2$ changes at rate $2Pbxu$. Add this derivative to the running cost and complete the square in the control.

$$qx^2 + ru^2 + \dot{V} = r\left(u + \frac{bP}{r}x\right)^2 + \left(q - \frac{b^2P^2}{r}\right)x^2.$$

Choose P so that the residual state term vanishes. The positive solution makes the candidate value nonnegative and gives a stabilising feedback. Integrating in time then writes the entire cost as its initial value plus a nonnegative penalty for departures from that control.

$$P = \frac{\sqrt{qr}}{b},$$

$$J = Px_0^2 + \int_0^\infty r\left(u + \frac{bP}{r}x\right)^2 dt.$$

Equality holds when $u = -Kx$ with $K = \sqrt{q/r}$. Substitution into the state balance gives an exponential return with time constant $1/(bK)$. The optimum assigns equal instantaneous penalties to state and effort in this specific integrator model.

$$x = x_0 e^{-bKt},$$

$$J_x = J_u = \frac{qx_0^2}{2bK},$$

$$J_{\min} = Px_0^2.$$

WORKED EXAMPLE

Price a change in the initial deviation

A normalised stage has $b = 2.00$, $q = 9.00$ and $r = 4.00$. Its optimal $K = 1.50$ gives a decay rate of 3.00 s⁻¹. For $x_0 = 0.400$, the minimum cost is $(\sqrt{36}/2)(0.400)^2 = 0.480$ weighted units. Half comes from control effort. Doubling the starting deviation quadruples total cost to 1.92 and doubles the magnitude of the initial command.

Change the inputs and inspect this calculation below

SECOND APPLICATION

Test a proposed control weight against an actuator limit

A second integrator has $b = 1.00$, $q = 4.00$, initial deviation 0.300 and an available command magnitude of 0.200. Since peak command is $K|x_0|$, the unconstrained solution fits this experiment only when $K \le 2/3$, requiring $r \ge 9.00$. At the boundary the time constant is 1.50 s and total optimal cost is

$$\sqrt{4 \times 9}(0.300)^2 = 0.540.$$

A larger starting deviation would need its own constraint check.

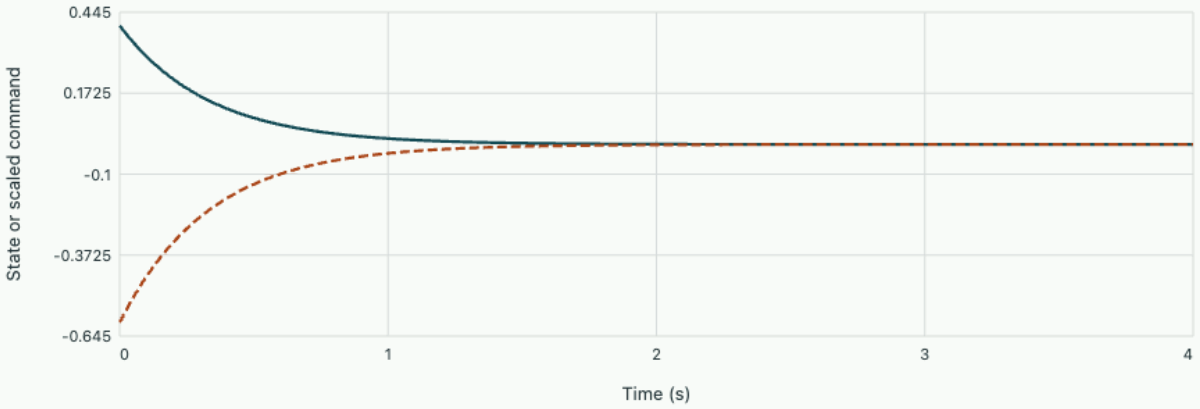
Interactive calculation · Change the inputs of this example

Use the adjustable quantities to test the two worked calculations above. The plot recomputes this derivation's balance or constitutive relation; its numerical readout reports the quantities needed for the forward calculation and the inverse or design question. Keep the fixed assumptions stated in the readout when comparing with the example values.

Open this live calculation

Try the calculation

State cost q	9
Control cost r	4
Initial deviation	0.4



Optimal feedback K=1.500; initial command -0.600; minimum cost 0.480. State and control contributions are each half of this total for the scalar integrator.

Make the design objective explicit

Pole placement prescribes decay rates but says nothing about how much actuator energy a manoeuvre consumes. For a continuous linear system $\dot{x} = Ax + Bu$ consider the regulator objective $J = \int_0^\infty (x^\mathsf{T} Q x + u^\mathsf{T} R u)\, dt$. Assume Q positive semidefinite and R positive definite. A large displacement penalty or a small control penalty generally encourages stronger corrective action. The ratio matters: multiplying both matrices by the same positive scalar leaves the best control unchanged. Scale states and inputs before comparing entries; a displacement in millimetres and a velocity in metres per second cannot share an arbitrary numerical weight without a unit convention.

Derive the continuous algebraic Riccati equation

For an infinite-horizon stationary problem propose a quadratic remaining cost $V(x) = x^{\mathsf{T}} P x$. During a short interval dt, optimality requires the minimum of instantaneous cost plus the change in remaining cost to vanish. In the limit this gives

$$0 = \min_{u} [x^{\mathsf{T}} Q x + u^{\mathsf{T}} R u + 2 x^{\mathsf{T}} P (A x + B u)].$$

Completing the square in u yields

$$u^{*} = -R^{-1} B^{\mathsf{T}} P x,$$

$$A^{\mathsf{T}} P + P A - P B R^{-1} B^{\mathsf{T}} P + Q = 0.$$

The desired solution is the stabilising symmetric solution. Under standard stabilisability and detectability conditions it exists and yields finite nonnegative cost. Other algebraic roots need not make sense as stored future cost. Once a candidate is found, check the eigenvalues of $A - B R^{-1} B^{\mathsf{T}} P$ and the residual of the matrix equation.

Solve the scalar unstable case

Let $\dot{x} = x + u$ with $Q = R = 1$. The scalar equation is $2P - P^2 + 1 = 0$, giving $P = 1 \pm \sqrt{2}$. Only $1 + \sqrt{2}$ is positive and makes the closed-loop pole $1 - P = -\sqrt{2}$. From initial state $x_0 = 2$, the theoretical minimum cost is $4(1 + \sqrt{2}) \simeq 9.657$. Verify independently: insert $x = 2e^{-\sqrt{2}t}$, $u = -Px$ into the cost integral to obtain $4(1 + P^2)/(2\sqrt{2})$, the same value. By contrast a stabilising gain of 2 gives state pole –1 and cost 10. Stabilising the state is necessary but still leaves a meaningful optimisation problem.

Distinguish control penalties from hard actuator limits

The optimal initial input for the scalar example has magnitude $2P = 4.83$. If the actuator is limited to 3, clipping the command invalidates the quadratic unconstrained solution during the clipped interval. Increasing R makes large inputs more expensive but generally supplies no absolute guarantee for arbitrary initial conditions. A concrete certificate is possible inside a stated region. In one dimension $|u| \le u_{\max}$ is guaranteed by $|x| \le u_{\max}/P$ under this gain, and the resulting stable exponential stays in that interval. Later we will impose hard constraints directly in a finite-horizon design.

The MIT notes on linear quadratic regulators extend this derivation to time-varying and finite-horizon problems.

Solve the Riccati equation for a position-and-velocity plant

A carriage moving on nearly frictionless guides is locally a double integrator. Take its position error x_1 and velocity x_2 as states and acceleration command u as input. After selecting consistent physical scales, use $A = \left(\begin{smallmatrix} 0 & 1 \\ 0 & 0 \end{smallmatrix}\right)$, $B = (0, 1)^T$ and diagonal cost $Q = \operatorname{diag}(q_x, q_v)$, $R = r$. Assume $q_x > 0,\ q_v \geq 0,\ r > 0$. Both states are reachable because acceleration changes velocity immediately and position through integration. Choosing a positive position penalty also makes a stationary position error visible to the cost.

Set $P = \left(\begin{smallmatrix} p_{11} & p_{12} \\ p_{12} & p_{22} \end{smallmatrix}\right)$ in $A^T P + PA - PBr^{-1}B^T P + Q = 0$. Matching the upper-left, lower-right and off-diagonal entries gives three scalar equations:

$$p_{12}^2 / r = q_x,$$

$$p_{22}^2 / r = 2p_{12} + q_v,$$

$$p_{11} = p_{12} p_{22} / r.$$

Choose the positive roots giving stabilising feedback. The gains in $u = -k_x x_1 - k_v x_2$ become

$$k_x = \sqrt{q_x / r},$$

$$k_v = \sqrt{2\sqrt{q_x / r} + q_v / r}.$$

Substitution into the state equations gives

$$s^2 + k_v s + k_x = 0.$$

The effort penalty and position penalty enter the squared natural frequency through a square root; the natural frequency itself varies as a fourth root. This explains why a moderate change of settling speed may require a surprisingly large change in a numerical cost weight.

Set a settling speed and calculate its actuator cost

In chosen metre, second and acceleration units let $q_x = 16,\ q_v = 0,\ r = 1$, with weights understood to include the needed unit conversions. Then

$$k_x = 4\,\mathrm{s}^{-2},\ k_v = \sqrt{8} = 2.828\,\mathrm{s}^{-1}.$$

The poles are $-1.414 \pm 1.414i\,\mathrm{s}^{-1}$, with natural frequency 2 rad/s and damping ratio $1/\sqrt{2}$. From an initial positive position error of 0.20 m and zero speed, commanded acceleration is −0.80 m/s². The optimum quadratic cost is $V = x^T P x$. Here $p_{12} = 4$, $p_{22} = 2.828$, $p_{11} = 11.314$, so the initial cost value is $11.314(0.2)^2 = 0.453$ in the chosen cost units.

To double natural frequency while keeping $q_v = 0,\ r = 1$, increase q_x by sixteen to 256. The new position gain is 16 s⁻², velocity gain is $\sqrt{32} = 5.657\,\mathrm{s}^{-1}$, and the same initial condition demands −3.2

m/s². An acceleration-limited drive rated for ±1.5 m/s² cannot execute this ideal motion from that initial error. For the larger gain, the zero-velocity initial error must be no more than $1.5/16 = 0.0938\,\mathrm{m}$ for the first command to lie within its limit. That calculation addresses the initial state only; subsequent velocity feedback must also be checked. Comparing physical demand rather than weight values makes the practical tradeoff visible.

The chosen gains can be tested without explicitly solving the Riccati equation again. Along an optimal unconstrained trajectory $V = x^T P x$ decreases according to $\dot{V} = -(x^T Q x + u^T R u)$. At the original initial state the instantaneous cost is $16(0.20)^2 + (-0.80)^2 = 1.28$ cost units per second. Directly differentiating V there gives $2x^T P \dot{x} = 2(0.20)(4)(-0.80) = -1.28$, since initial speed is zero and acceleration equals the command. The sign and magnitude agree. This check simultaneously tests the gain sign, state ordering and cost convention. Introducing an input clip breaks the square-completion identity for the proposed unconstrained controller, so the same matrix cannot simply be read as the realised accumulated cost of the clipped manoeuvre.

Chapter 10 · Estimating hidden states from noisy measurements

Propagate both the state estimate and its uncertainty, and derive how much of an unexpected sensor reading to trust.

FROM FIRST PRINCIPLES

Choose a measurement gain by minimising estimation variance

Suppose a scalar prediction $\hat{x}^-$ has zero-mean error of variance P^-, while an independent unbiased sensor gives $y = x + v$ with noise variance R. Correct the prediction by a fraction K of its disagreement with the sensor. The posterior estimation error combines the original error and sensor noise with different weights.

$$\hat{x}^+ = \hat{x}^- + K(y - \hat{x}^-),$$
$$x - \hat{x}^+ = (1 - K)(x - \hat{x}^-) - Kv.$$

Independence eliminates the cross term when the squared error is averaged. Differentiate the resulting variance with respect to the gain and solve for its minimum.

$$P^+ = (1 - K)^2 P^- + K^2 R,$$
$$\frac{dP^+}{dK} = -2(1 - K)P^- + 2KR = 0.$$

The optimum gain grows when prediction uncertainty rises relative to sensor uncertainty. Substituting it into the variance gives

$$K = \frac{P^-}{P^- + R},$$
$$P^+ = \frac{P^- R}{P^- + R}.$$

The innovation $y - \hat{x}^-$ has variance $P^- + R$ under the same assumptions. Monitoring that disagreement in actual samples checks whether the stated uncertainty model is consistent with the observed data.

DERIVATION AND APPLICATIONS

Find the steady uncertainty of a repeatedly sampled sensor

A sequence of scalar Kalman updates can settle to a fixed confidence even if the physical state continues wandering. Model the state as $x_{k+1} = x_k + w_k$ with independent zero-mean increments of variance Q. A sensor gives $y_k = x_k + v_k$ with independent noise variance $R > 0$. Use the same

sampling interval each time and assume known, constant variances. If P denotes the posterior variance immediately after one reading, propagation adds the process variance before the next reading.

$$P^{-} = P + Q,$$

$$K = \frac{P^{-}}{P^{-} + R},$$

$$P_{\text{next}} = (1 - K)P^{-} = \frac{R(P + Q)}{P + Q + R}.$$

At steady operation set $P_{\text{next}} = P$ and multiply by the denominator. The product RP cancels from both sides, leaving a quadratic. Variance must be nonnegative, so only one root is admissible for $Q > 0$.

$$P^2 + QP - RQ = 0,$$

$$P_{\infty} = \frac{\sqrt{Q^2 + 4QR} - Q}{2}.$$

Since the posterior update can also be written $P_{\text{next}} = RK$, its steady value immediately fixes the long-run gain. Conversely, measuring a gain and knowing sensor variance supplies an estimate of the assumed process increment variance.

$$K_{\infty} = P_{\infty}/R,$$

$$\frac{Q}{R} = \frac{K_{\infty}^2}{1 - K_{\infty}}.$$

For large Q/R, the state changes substantially between readings and the gain approaches one. With vanishing process noise repeated independent readings continue reducing uncertainty; the positive fixed point then tends to zero.

WORKED EXAMPLE

Predict a level sensor's residual spread

A drifting liquid level has $Q = 1.00$ mm² per sample and sensor variance $R = 4.00$ mm². The posterior variance tends to $(\sqrt{17} - 1)/2 = 1.56$ mm², giving a posterior standard deviation of 1.25 mm. The steady gain is 0.390. A 3.00 mm innovation consequently moves the estimate 1.17 mm. The innovation variance includes both prediction and sensor uncertainty, giving $1.56 + 1 + 4 = 6.56$ mm².

Change the inputs and inspect this calculation below

SECOND APPLICATION

Infer the assumed drift from a filter log

Another estimator settles at gain 0.200 with independently measured $R = 9.00$ K². Then $Q = 9(0.200)^2/(1 - 0.200) = 0.450$ K² per sample and posterior variance is $RK = 1.80$ K². If the

physical drift is a random walk with variance proportional to elapsed time, doubling the sample interval would double Q to 0.900 K². The predicted new steady gain becomes

$$(\sqrt{0.900^2 + 4(0.900)(9)} - 0.900)/18 = 0.270.$$

Interactive calculation · Change the inputs of this example

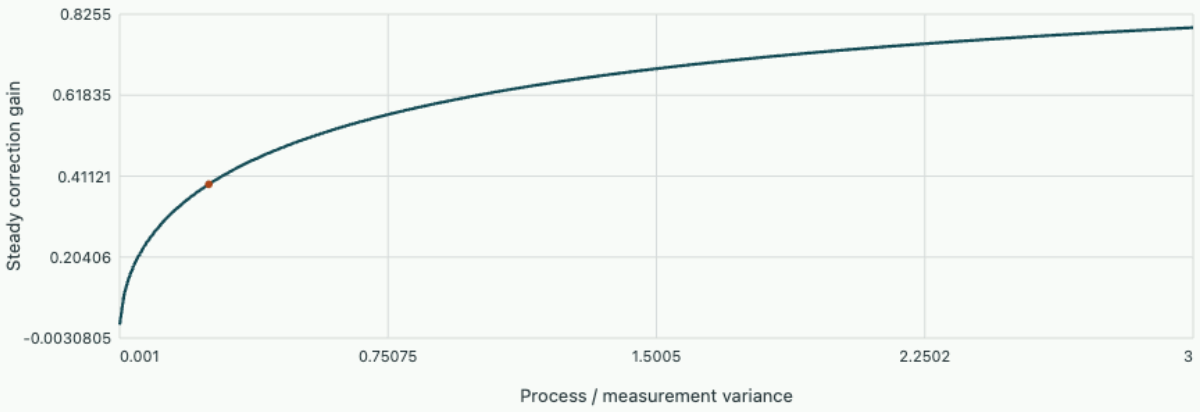
Use the adjustable quantities to test the two worked calculations above. The plot recomputes this derivation's balance or constitutive relation; its numerical readout reports the quantities needed for the forward calculation and the inverse or design question. Keep the fixed assumptions stated in the readout when comparing with the example values.

Open this live calculation

Try the calculation

Process variance Q (mm²) 1

Measurement variance R (mm²) 4



Steady random-walk filter | Selected variance pair

Posterior standard deviation 1.250 mm; gain 0.390. A3 mm innovation shifts the estimate by 1.171 mm; innovation variance=6.562 mm².

A prediction and an observation carry different information

Consider a sampled model $x_{k+1} = Ax_k + Bu_k + w_k$, $y_k = Cx_k + v_k$. Model disturbances w_k describe unmeasured forcing or model error, while v_k describes measurement error. Suppose they are independent, zero mean and white, with known covariances W, V, and independent of the initial estimation error. Start each update with a predicted mean $\hat{x}^-$ and covariance P^-. The innovation

$y - C\hat{x}^-$ is the part of the reading that the model did not predict. Its covariance is $S = CP^-C^\mathsf{T} + V$. Units are those of the sensor squared; a large innovation in volts is meaningful only relative to this scale.

Minimise posterior variance

Update the estimate by $\hat{x}^+ = \hat{x}^- + K(y - C\hat{x}^-)$. In the scalar direct-measurement case the squared error has expected value $P^+ = (1-K)^2P^- + K^2V$. Differentiating with respect to K yields

$$K = P^-/(P^- + V).$$

A precise sensor raises the gain; a precise prior prediction lowers it. The same covariance minimisation gives the matrix update

$$K = P^-C^\mathsf{T}S^{-1},$$
$$P^+ = (I - KC)P^-(I - KC)^\mathsf{T} + KVK^\mathsf{T}.$$

The displayed covariance formula retains positive-semidefinite structure in computation better than a raw subtraction of nearly equal matrices. For the next prediction use $\hat{x}^-_{k+1} = A\hat{x}^+_k + Bu_k$ and $P^-_{k+1} = AP^+_kA^\mathsf{T} + W$. The control input affects the mean; assumed known exactly, it adds no direct covariance term. With Gaussian assumptions the recursion gives the conditional mean and covariance, and with just second moments it gives the optimal linear estimate under the stated model.

Recover velocity from a position reading

Suppose the prior covariance for position and velocity is $\left(\begin{smallmatrix} 4 & 1.5 \\ 1.5 & 1 \end{smallmatrix}\right)$ in compatible metre and second units. A position sensor has variance 1. Then $S = 5$,

$$K = (0.8, 0.3)^\mathsf{T}.$$

If measured position exceeds prediction by 2 m, the updated position increases by 1.6 m and velocity by 0.6 m/s. The sensor did not measure velocity directly: the nonzero cross-covariance makes a position surprise informative about velocity. The posterior covariance is $\left(\begin{smallmatrix} 0.8 & 0.3 \\ 0.3 & 0.55 \end{smallmatrix}\right)$. A robust implementation would obtain the gain by solving with S instead of constructing a matrix inverse.

Test innovations during commissioning

Under a correct linear Gaussian model the standardised innovations should have mean zero, appropriate variance and little serial correlation. A persistent offset points toward a biased sensor or missing force. A succession of innovations of the same sign during accelerations suggests unmodelled dynamics or too little process uncertainty. Simply enlarging V to suppress the innovations in the correction makes the filter trust the model more; it does not remove a systematic error. Compare held-out records that include

manoeuvres and quiet intervals. When observer and regulator are combined, estimation and control share the quality of the underlying dynamic model.

Fuse two measurements with a shared error source

Two temperature sensors mounted on one test head can share calibration drift or electromagnetic interference. Treating their errors as independent exaggerates the information gained by averaging. Suppose they measure the same scalar state at a single instant, $y_1 = x + v_1,\ y_2 = x + v_2$, with zero-mean errors, variances r_1, r_2 and covariance c. Seek an unbiased linear estimate $\hat{x} = w y_1 + (1 - w) y_2$, initially without a separate prior. Its error variance is

$$J(w) = w^2 r_1 + (1 - w)^2 r_2 + 2w(1 - w)c.$$

Known covariance makes this a scalar optimisation rather than a heuristic choice of weights.

Differentiating and setting $J'(w) = 0$ gives

$$w = \frac{r_2 - c}{r_1 + r_2 - 2c},$$

$$R_{\text{fused}} = \frac{r_1 r_2 - c^2}{r_1 + r_2 - 2c}.$$

The denominator is exactly the variance of the sensor difference. If their errors are identical, this difference contains no information for estimating a weighting and the duplicate measurement cannot improve precision. Positive semidefiniteness of the covariance requires

$$c^2 \le r_1 r_2.$$

Negative weights can appear when one error contains a predictable correlated component that can be subtracted. Such weights require a well-supported covariance model, especially when the denominator is small.

Update one state without counting a common disturbance twice

Take sensor standard deviations 2 K and 3 K and covariance $c = 3\,\mathrm{K}^2$, corresponding to correlation 0.50. The first sensor's optimal weight is $(9 - 3)/(4 + 9 - 6) = 6/7$, with $1/7$ on the second. Readings of 305 and 308 K give a fused reading of 305.429 K and variance $(36 - 9)/7 = 3.857\,\mathrm{K}^2$. The independent-sensor formula would instead claim $36/13 = 2.769\,\mathrm{K}^2$, materially overstating the improvement. Each sensor may appear repeatable by itself while both shift in the same direction between experiments.

Suppose a prior estimate from a thermal model is 302 K with variance 4 K², independent of both current measurement errors. One scalar Kalman update using the fused measurement has gain

$4/(4+3.857) = 0.5091$. Its estimate is $302 + 0.5091(305.429 - 302) = 303.745\,\mathrm{K}$ and its posterior variance is

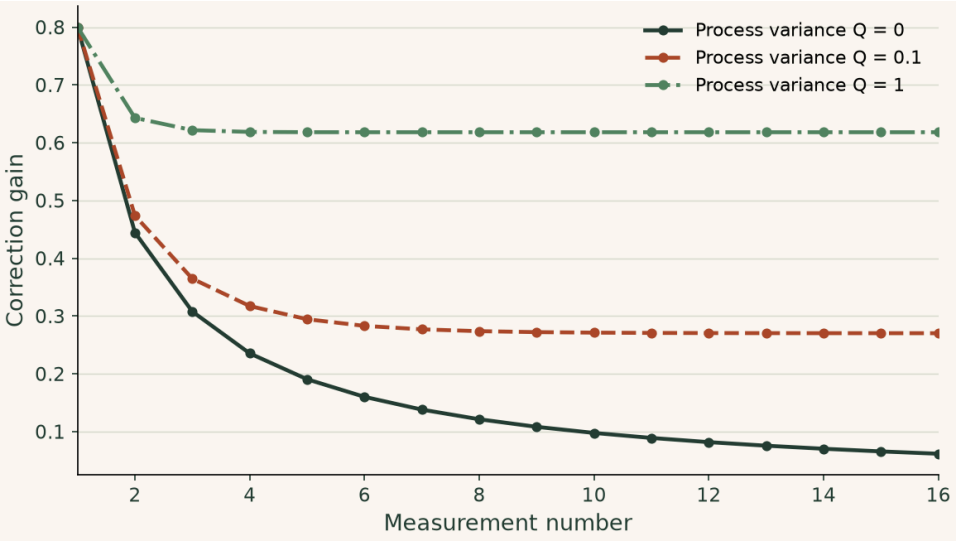
$$(1 - 0.5091)4 = 1.964\,\mathrm{K}^2.$$

The same answer follows from a two-row measurement matrix and the full covariance matrix with its off-diagonal terms. If the model has already incorporated sensor 1 at this same instant, reusing the fused reading as an independent update counts some evidence twice. Synchronisation and calibration history therefore determine which covariance belongs in the estimator, as well as the arithmetic of its gain.

Sensor disagreement provides a second diagnostic that does not require knowing the true temperature. Because the two sensors observe the same instantaneous state, their difference $y_1 - y_2$ should have variance $r_1 + r_2 - 2c = 7\,\mathrm{K}^2$. The observed difference of −3 K corresponds to about $-3/\sqrt{7} = -1.13$ standard deviations. A long sequence of such differences tests the assumed relative-error covariance without a calibrated thermal reference. It cannot identify a common offset, since that cancels from every difference. If both instruments warm their local mounting head or use the same drifting reference voltage they may agree closely while being wrong together. Periodic comparison against an independent standard is needed before interpreting small disagreement as strong evidence of absolute accuracy.

Watch repeated observations change the Kalman gain

A scalar random walk has initial predicted variance 4 and independent measurement variance 1. At each reading use $K = P^{-}/(P^{-} + 1)$, then update $P^{+} = (1 - K)P^{-}$ and predict $P^{-}_{\text{next}} = P^{+} + Q$.



All three filters start with gain 0.8. Without process uncertainty the gain steadily shrinks, while Q = 1 approaches about 0.618. A stable nonzero gain means that uncertainty is replenished between readings.

▶ **Read values from the plot**

Interactive calculation · change the inputs of this example

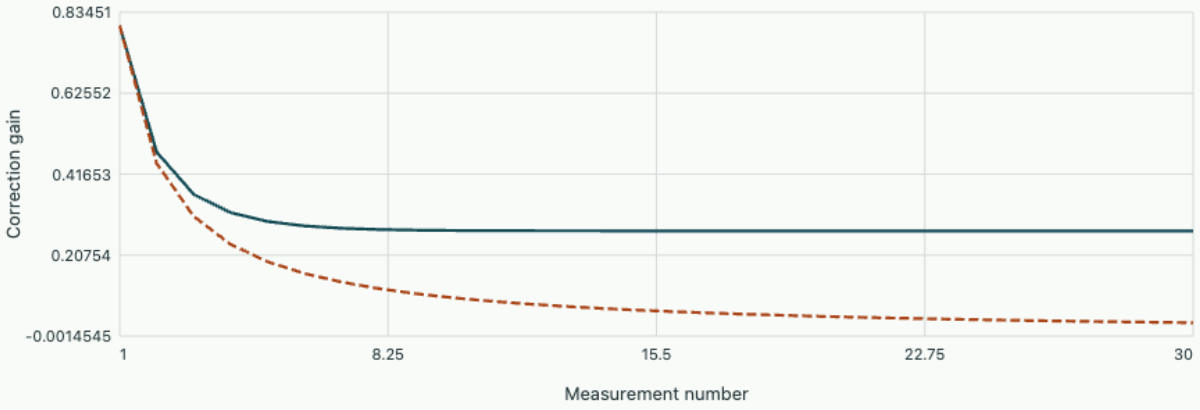
Separate the effect of initial uncertainty from uncertainty added at every step. Measurement variance sets how much confidence to put in each observation.

The curves below are recalculated from this section's equations in your browser. The fixed reference plot above remains available for comparison.

Open this live calculation

Try the calculation

Process variance Q	0.1
Measurement variance R	1
Initial predicted variance	4



Selected process uncertainty No process uncertainty

First gain 0.800; gain after30 observations 0.270. Independent process and observation noise are assumed.

Study the full derivation

Estimating velocity from position with a model-based observer · System dynamics, feedback and estimation

An optical encoder returns position and the actuator drive supplies a commanded acceleration. Reconstruct an unmeasured velocity without differentiating successive noisy samples. Derive the observer, explain what a pair of chosen poles guarantees, and predict the tradeoff visible when the encoder carries periodic interference.

Problem sheet 5 · Regulate a state and estimate it from data

Calculate an optimal feedback policy and decide when filtering or sensor placement limits its implementation.

1. Energy and effort on an integrator

For $\dot{x} = u$, minimise $\int_0^\infty (9x^2 + 4u^2)\,dt$. (a) Find the stabilising Riccati solution and feedback gain. (b) For $x_0 = 2$ calculate the optimal cost directly from the exponential trajectory. (c) If $|u| \le 2$, identify a region where the unconstrained policy is feasible.

▼ **Check the cost with a direct integral**

The algebraic equation $9 - P^2/4 = 0$ gives $P = 6$; $u = -1.5x$. Then $x = 2e^{-1.5t}$ and the integrand is $72e^{-3t}$. Its integral 24 agrees with Px_0^2. Feasibility is guaranteed for $|x_0| \le 4/3$ since the magnitude then decreases monotonically. The initial state 2 is outside this region.

---

2. A scalar filter through a missed measurement

A random-walk model has initial predicted mean 8, variance 4, and measurement variance 1. A reading of 10 arrives. (a) Calculate the corrected mean and variance. (b) Predict one step with process variance 0.5. (c) One reading is then lost. Find the gain at the following available reading.

▼ **Follow the prediction and correction chronology**

The first gain is $4/5$, so the corrected mean is 9.6 and variance 0.8. The next prior mean is 9.6 and variance 1.3. With no reading there is no correction; another prediction raises variance to 1.8. The next available reading obtains gain $1.8/2.8 = 0.643$. Setting a missed reading equal to its prediction and carrying out a normal update would falsely shrink the variance.

---

3. An apparently quiet but unreliable sensor

A scalar innovation sequence is $2.0, 2.0, 1.8, 2.2, 2.0$ in sensor units, with predicted innovation standard deviation 0.5. (a) Compute the standardised mean and range. (b) Compare two mechanisms that could generate the trace. (c) Propose a test that separates the mechanisms.

▼ **Use sign and time structure as well as magnitude**

The standardised values are $4, 4, 3.6, 4.4, 4$, with mean 4 and range 3.6 to 4.4. A biased sensor and a missing nearly constant load can both produce a sustained signed discrepancy. During a known calibration hold compare the device with an independent sensor. Then vary the load while holding

sensor conditions fixed; innovations that follow the load implicate the model. Simply increasing the assumed noise variance could hide both effects.

Explore the governing equations after solving the sheet: Optimal state feedback and the Riccati equation; Estimating hidden states from noisy measurements. Compare the assumptions and units before transferring your chosen parameters; the controls may cover a different numerical range from the written exercise.

Chapter 11 · Interacting actuators, pairing and decoupling

Predict how closing one loop changes the apparent gain of another, and identify a physically expensive actuator combination.

FROM FIRST PRINCIPLES

Eliminate the second actuator to find a constrained gain

With several actuators and outputs, a gain depends on what the other controller holds fixed. Consider small steady deviations described by a square two-by-two gain matrix G. First hold u_2 fixed at zero. The direct response of y_1 to u_1 has slope g_{11}. Now assume an ideal second controller adjusts u_2 slowly enough to maintain $y_2 = 0$ in steady state.

$$y_1 = g_{11}u_1 + g_{12}u_2,$$
$$0 = y_2 = g_{21}u_1 + g_{22}u_2.$$

For $g_{22} \neq 0$, the second actuator must move by $-g_{21}u_1/g_{22}$. Substitution gives the constrained gain seen by the first loop and the ratio of the direct to constrained gains.

$$\left.\frac{y_1}{u_1}\right|_{y_2=0} = \frac{\det G}{g_{22}},$$
$$\lambda_{11} = \frac{g_{11}g_{22}}{\det G}.$$

The ratio requires a nonsingular matrix. A value near unity means closing the second loop changes this steady gain little. A negative value signals a reversal of the gain sign under the constraint. Dynamic phase and actuator limits still affect whether the pairing is usable.

DERIVATION AND APPLICATIONS

Determine how actuator limits reshape the reachable demands

A two-actuator mechanism may have abundant force in either output separately but be unable to deliver the two requested forces simultaneously. Work with a static local model and identical actuator units. Choose outputs as the sum y_1 and difference y_2 of two actuator forces u_1, u_2. Assume each is bidirectional and bounded by magnitude U, and the static gain holds over the region of interest. The forward and inverse maps are

$$
\begin{aligned}
y_1 &= u_1 + u_2, \\
y_2 &= u_1 - u_2, \\
u_1 &= \frac{y_1 + y_2}{2}, \\
u_2 &= \frac{y_1 - y_2}{2}.
\end{aligned}
$$

Substitution of the inverse map into the two component limits gives a pair of strips in output space. Their intersection is the reachable region. The maximum of the absolute sum and absolute difference of two real numbers is their absolute-value sum, which simplifies the description to a diamond.

$$
|y_1 + y_2| \le 2U, \qquad |y_1 - y_2| \le 2U \quad \Longleftrightarrow \quad |y_1| + |y_2| \le 2U.
$$

A supervisor can retain the relative amounts of the two requested outputs by scaling a nonzero desired vector d by one factor. The largest such factor up to unity follows by placing the scaled request on the reachable boundary whenever the original request exceeds capacity.

$$
\begin{aligned}
y &= \alpha d, \\
\alpha &= \min\left(1, \frac{2U}{|d_1| + |d_2|}\right).
\end{aligned}
$$

This allocation preserves the requested direction. If one output has priority, it may be preferable to maintain that output and use only the residual capacity for the other. The physical feasible set is the same for both decisions.

WORKED EXAMPLE

Preserve the direction of a combined request

Two linear actuators each permit 50.0 N. A requested combination $(80, 60)$ N would need $(70, 10)$ N. Scaling by $100/140 = 0.7143$ yields outputs $(57.1, 42.9)$ N and actuator loads $(50.0, 7.14)$ N. Checking the inverse transformation shows exactly which actuator has reached its limit while the other retains substantial headroom.

Change the inputs and inspect this calculation below

SECOND APPLICATION

Preserve a preload while alternating the differential force

A separate rig with $U = 30.0$ N requires total preload $y_1 = 44.0$ N throughout a test. The reachable inequality leaves $|y_2| \le 60.0 - 44.0 = 16.0$ N. At the positive extreme the actuator forces are $(30.0, 14.0)$ N; at the negative extreme they exchange roles. A demanded 20.0 N alternating differential amplitude would require reducing the fixed preload to at most 40.0 N.

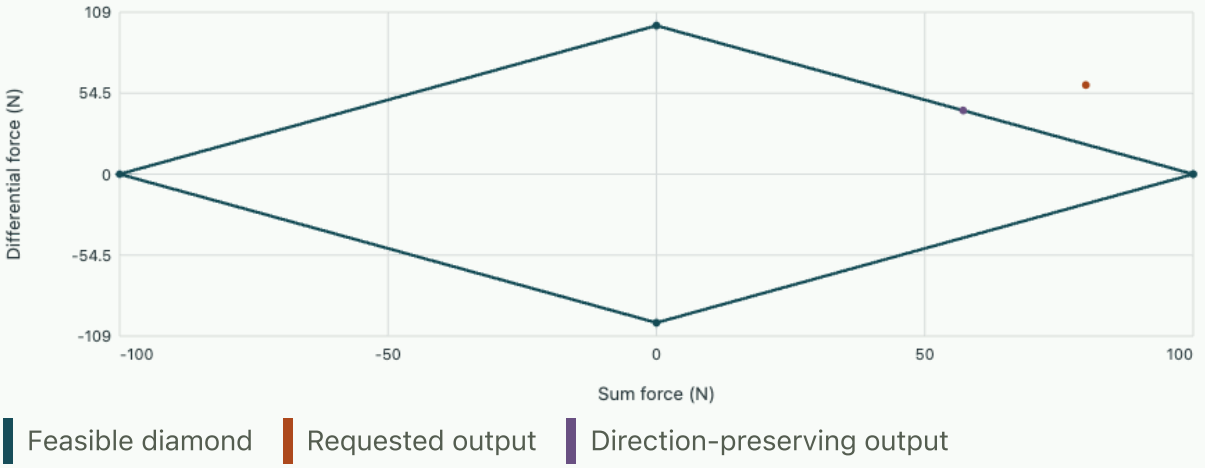
Interactive calculation · Change the inputs of this example

Use the adjustable quantities to test the two worked calculations above. The plot recomputes this derivation's balance or constitutive relation; its numerical readout reports the quantities needed for the forward calculation and the inverse or design question. Keep the fixed assumptions stated in the readout when comparing with the example values.

Open this live calculation

Try the calculation

Actuator force limit (N)	50
Requested sum force (N)	80
Requested differential force (N)	60



Scale 0.714: actuator forces 50.000 and 7.143 N. Feasible outputs depend on coupled actuator limits.

A second loop changes the experiment

Imagine two thermal zones with separate heaters and two temperature sensors. At steady state let $y = Gu$, with $G = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$ K per unit power. Increasing heater 1 with heater 2 fixed raises both temperatures. Increasing heater 1 while actively holding temperature 2 fixed is a different experiment: heater 2 must back off. The effective gain in the first channel changes. This difference matters when a controller is tuned with its neighbour disconnected and then commissioned with both loops active.

For a general nonsingular two-by-two matrix hold $y_2 = 0$. Solving $g_{21}u_1 + g_{22}u_2 = 0$ gives

$$y_1/u_1 = g_{11} - g_{12}g_{21}/g_{22}.$$

The ratio of the fixed-input gain to this fixed-output gain is $g_{11}g_{22}/\det G$. More generally the relative gain array has entries $\Lambda_{ij} = g_{ij}(G^{-1})_{ji}$. Its rows and columns each sum to one. Large magnitudes suggest cancellation. A negative relative gain indicates that the paired steady response changes sign when the other output is perfectly held, an important warning for independent-loop tuning.

Compare candidate pairings

For the two-zone matrix above the diagonal relative gains are $4/3$ and the off-diagonal gains are $-1/3$. The direct pairing has fixed-input gain 2 and fixed-other-output gain 1.5. The cross-pairing changes sign. These steady results do not include transport delays or transient coupling. If the direct heater path is very slow but the cross path changes quickly, a frequency-dependent model is needed around the intended control bandwidth.

Diagonalise a symmetric plant by physical combinations

Commanding both heaters together excites the common direction $(1, 1)$ with gain 3; commanding them oppositely excites $(1, -1)$ with gain 1. For the desired temperature increment $(1, -1)$, exactly equal and opposite input increments $(1, -1)$ are required. Negative input here is a reduction from a positive bias; a heater already at zero cannot supply it. Inspecting a transformed system without retaining the original actuator limits can make an impossible target seem easy.

Now bring the gains closer together:

$$G = \begin{pmatrix} 1 & 0.98 \\ 0.98 & 1 \end{pmatrix}.$$

The common gain is 1.98, while the differential gain is just 0.02. A unit differential target requires differential input of magnitude 50 in each actuator. Static inversion $u = G^{-1}v$ algebraically decouples the two commands but amplifies measurement, calibration and modelling errors along that weak direction. If both temperature sensors move together during every available test, the weak direction is also particularly hard to identify.

Choose a practical validation test

First perturb each actuator separately about the operating bias and measure both outputs. Repeat at different loads and frequencies. Next excite the predicted weakest input combination within safe operating limits and check whether the differential output emerges above the noise. A design can prioritise the common temperature and relax allowable imbalance when the plant is inherently hard to split. A constrained controller can express this trade directly by penalising both output errors while respecting individual actuator bounds.

Find difficult demand directions with a singular-value decomposition

Relative gains help pair individual inputs and outputs, but some machines are controlled through combinations of actuators. A steady mapping $y = Gu$ with adequate static rank can still require a large actuator change in a particular direction. Scale inputs by allowed physical excursions and outputs by meaningful error or command scales. After scaling, the Euclidean norm is interpretable and singular values quantify directional gains. Without scaling, changing position units from metres to millimetres could dominate the result.

For a real square nonsingular G, the symmetric positive definite matrix $G^T G$ has orthonormal eigenvectors v_j and positive eigenvalues σ_j^2. A unit actuator pattern v_j yields an output of length σ_j in direction $u_j = Gv_j/\sigma_j$. Writing $G = U\Sigma V^T$ then gives a transparent inverse. A desired output y_r requires actuator components $(u_j^T y_r)/\sigma_j$ along v_j. The smallest singular value therefore controls the largest gain of the inverse, and the ratio $\sigma_{\max}/\sigma_{\min}$ measures conditioning.

This geometry describes steady input effort, not necessarily dynamic stability. Two channels with different lags can have different difficult directions at higher frequencies. The singular values of the complex frequency-response matrix can reveal that change after consistent scaling at every frequency.

Allocate two actuators with a tight current budget

Consider a scaled steady plant $G = \left(\begin{smallmatrix} 1 & 0.9 \\ 0.9 & 1 \end{smallmatrix}\right)$. Its symmetry gives unit common and differential directions $(1, 1)/\sqrt{2}$ and $(1, -1)/\sqrt{2}$, with singular values 1.9 and 0.1. A common output target $(0.2, 0.2)$ requires only $u = (0.1053, 0.1053)$. A differential target $(0.2, -0.2)$ needs $u = (2, -2)$, a factor nineteen more effort for the same output norm. If each drive is limited to magnitude one, the first target is feasible and the second is not.

One possible allocation trades exact tracking against effort by minimising $\|Gu - y_r\|^2 + \lambda\|u\|^2$. The stationary normal equations are

$$(G^T G + \lambda I)u = G^T y_r;$$

each inverse singular value is replaced by $\sigma_j/(\sigma_j^2 + \lambda)$. For $\lambda = 0.01$ and the differential target, that factor is

$$0.1/(0.01 + 0.01) = 5.$$

The resulting input is $(1, -1)$ and output $(0.1, -0.1)$. The allocation respects these particular bounds while openly accepting half the requested differential change. A penalty does not enforce an input limit for every possible target. If exact limit enforcement is required, solve the bounded allocation problem or constrain the admissible command directions explicitly.

A simple perturbation test makes the weak direction visible in hardware. Command $u = (0.1, 0.1)$ in scaled units and predict $y = (0.19, 0.19)$. Reversing only the second command gives $u = (0.1, -0.1)$

and outputs only $(0.01, -0.01)$. If each sensor has an uncertainty of ±0.01 output units, the first test has a comfortable signal level but the second barely resolves its predicted value. A poorly estimated differential gain may therefore vary considerably between identification records even though the common gain is repeatable. Increasing the differential test amplitude within operating bounds or averaging repeated experiments improves that estimate. A static inverse based on a single small pulse at each actuator may otherwise amplify calibration error precisely in the least observable output combination.

Study the full derivation

Relative gains, loop pairing and nearly singular process measurements · System dynamics, feedback and estimation

Two heaters act on two temperature sensors. With the second heater fixed, a test of the first gives a well-determined positive response. Closing a loop around the other temperature can change both the size and sign of that apparent gain. Derive the relative gain array from the experiment one actually performs.

Chapter 12 · Finite-horizon control with real actuator constraints

Build a small predictive controller from a sampled model and determine when the requested target is physically reachable.

FROM FIRST PRINCIPLES

Predict a finite trajectory before choosing the first command

Consider a scalar sampled plant $x_{k+1} = ax_k + bu_k$, with a measured initial state and known coefficients. A predictive controller uses the model repeatedly to compare candidate future command sequences. For two steps, substitute the first state update into the second; each input affects a different interval and the first input also affects the second predicted state through propagation.

$$x_1 = ax_0 + bu_0,$$
$$x_2 = a^2x_0 + abu_0 + bu_1.$$

Even a one-step horizon shows how a quadratic objective selects a command. Let r be the desired next state, $q > 0$ a tracking penalty and $\rho > 0$ an effort penalty. Differentiate the objective after substituting its predicted state.

$$J = q(ax_0 + bu_0 - r)^2 + \rho u_0^2,$$
$$u_0^* = \frac{qb(r - ax_0)}{qb^2 + \rho}.$$

For this single scalar step, an interval limit on u_0 clips the unconstrained minimiser to the nearest feasible endpoint. Over longer horizons the state and input constraints couple the commands. Only the first chosen command is applied before the state is measured again and the prediction repeated.

DERIVATION AND APPLICATIONS

Predict the stopping distance of a sampled axis

A predictive controller can satisfy every instantaneous position bound and still enter a state from which the next position constraint must be violated. To see the role of velocity, consider a frictionless axis moving towards a positive position limit. The command is constant over sample interval h and bounded in acceleration by $a > 0$. Suppose the current speed v is positive, the axis must stop without reversing, and a final fractional braking command is allowed. Constant acceleration integrates exactly over each step.

$$v_{k+1} = v_k + hu_k,$$
$$x_{k+1} = x_k + hv_k + \tfrac{1}{2}h^2u_k, \qquad |u_k| \le a.$$

Apply maximum braking for $n = \lfloor v/(ah) \rfloor$ full samples. The remaining speed $w = v - nah$ is in $[0, ah)$. During these n samples the speed decreases linearly, so the travelled distance is their duration times the average of initial and final speed.

$$d_n = nh\,\frac{v + w}{2} = \frac{v^2 - w^2}{2a}.$$

If $w > 0$, select $u = -w/h$ for the last interval. The last step travels $hw/2$ and ends at rest exactly on a sampling instant. Combining the two contributions yields the minimum stopping distance with this no-reversal and held-command requirement.

$$d_{\text{stop}} = \frac{v^2}{2a} + \frac{w(ah - w)}{2a}.$$

Maximum early braking minimises future speeds and hence the travelled distance. The extra term compared with continuous switching is nonnegative and vanishes for an integer number of full braking intervals. A position-only terminal constraint can overlook this extra travel.

WORKED EXAMPLE

Check if a carriage can stop before its limit

Let $v = 0.550$ m/s, $a = 2.00$ m/s² and $h = 0.100$ s. Two full braking steps leave $w = 0.150$ m/s. They consume $0.200(0.550 + 0.150)/2 = 0.0700$ m. The final step needs $u = -1.50$ m/s² and travels 7.50 mm, giving 77.5 mm total. A remaining clearance of 76.0 mm exceeds the continuous stopping distance of 75.6 mm but is insufficient for the sampled axis.

Change the inputs and inspect this calculation below

SECOND APPLICATION

Bound the error without knowing the initial speed

For $a = 1.20$ m/s² and $h = 0.200$ s the correction is largest at $w = ah/2$. Its maximum is

$$ah^2/8 = 6.00\,\text{mm}.$$

At $v = 0.360$ m/s a full step leaves exactly 0.120 m/s, so the continuous estimate of 54.0 mm increases to 60.0 mm. Reducing the sample interval by half reduces this worst-case excess distance by a factor of four.

Interactive calculation · Change the inputs of this example

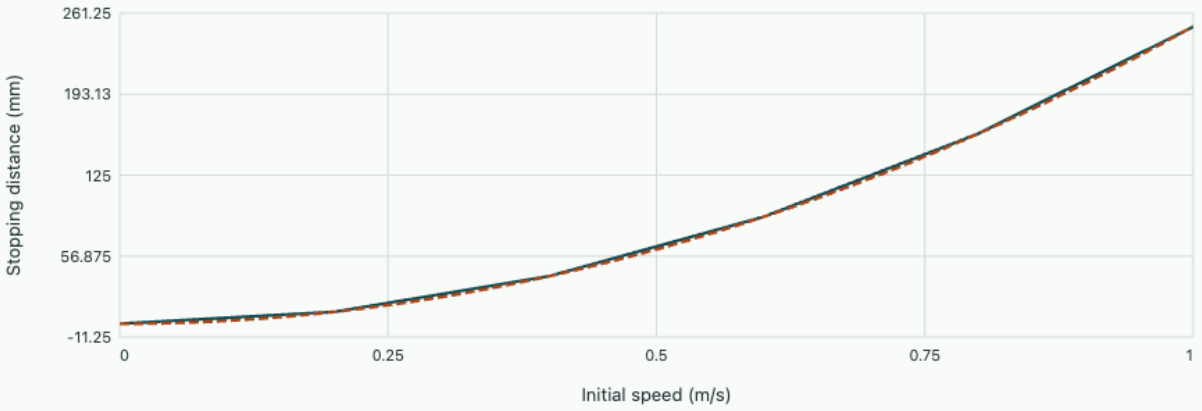
Use the adjustable quantities to test the two worked calculations above. The plot recomputes this derivation's balance or constitutive relation; its numerical readout reports the quantities needed for the forward calculation and the inverse or design question. Keep the fixed assumptions stated in the readout when comparing with the example values.

Open this live calculation

Try the calculation

Braking acceleration (m/s²) **2**

Sample interval (s) **0.1**



Sampled braking with final partial command Continuous ideal braking

At0.55 m/s stopping needs 77.50 mm. Maximum extra distance over the continuous formula is 2.500 mm.

Predict a sequence before applying the first action

For $x_{k+1} = Ax_k + Bu_k$ with measured or estimated x_k, successive substitutions give

$$x_{k+2} = A^2x_k + ABu_k + Bu_{k+1}.$$

The initial state sets the free response; the planned input sequence contributes linearly. With quadratic tracking and effort penalties, linear dynamics and linear bounds on states or inputs, planning becomes a convex quadratic programme. At each update apply only the first planned command, acquire a new measurement and solve again. Receding the horizon provides feedback when disturbances make actual motion depart from the earlier prediction.

Solve a two-step constrained problem

Take the dimensionless integrator $x_{k+1} = x_k + u_k$, $x_k = 2$, and target zero. Minimise $x_{k+1}^2 + x_{k+2}^2 + u_k^2 + u_{k+1}^2$. With $u = u_k, v = u_{k+1}$, the objective is $(2+u)^2 + (2+u+v)^2 + u^2 + v^2$. Its stationary equations are

$$3u + v = -4,\ u + 2v = -2.$$

The unconstrained plan $u = -1.2,\ v = -0.4$ reaches states 0.8 and 0.4. If both input magnitudes are limited to one, the first command is infeasible. Fixing $u = -1$ and reoptimising the second gives $v = -0.5$, reaching 1 and 0.5. The derivative with respect to u points toward smaller u at this boundary, confirming the active limit.

After applying –1, suppose the new state is exactly 1. The next two-step unconstrained plan scales the original one by one half, giving first action –0.6. The earlier second action was –0.5. This change is consistent even without disturbances, since the new optimisation includes a further future step. A terminal cost approximating the remaining cost of control can reduce this artificial dependence on the planning horizon.

Feasible commands and feasible trajectories differ

A command bound always has feasible choices if zero lies within it. Adding a hard state constraint, such as a position stop, may make all predicted trajectories infeasible from the current state. Before asking for a time-optimal move, compute elementary reachability bounds. With $|u| \le 1$, a scalar integrator can move at most N units in N samples. A terminal constraint demanding zero from state 3 in two samples is impossible. Allowing slack in a comfort-related temperature band may be defensible; a physical actuator limit must still be enforced. Report which constraints were active and what performance was sacrificed.

Close the engineering loop

For a practical comparison, first establish a verified baseline with the OneSimulation proportional-loop record. Reconstruct its pole and finite-record bias from the model parameters. This record witnesses how an unconstrained linear loop behaves; it does not itself exercise a predictive optimiser. Now design a separate held-input or constrained test whose command reaches a known bound, record measured timestamps and actuator signals, and compare with a by-hand prediction of the first active limit. The useful performance claims differ: accurate steady tracking, stable sampled dynamics, adequate estimator uncertainty and feasibility under limits each require an observation capable of testing them.

Derive a terminal region that can be maintained indefinitely

A finite-horizon planner can reach a state at its last predicted sample that looks inexpensive but has no feasible continuation. A terminal set provides a constructive remedy. For $x_{k+1} = Ax_k + Bu_k$, choose

local feedback $u = Kx$. A set X_f is positively invariant under this feedback if every $x \in X_f$ satisfies the state and input bounds and $(A + BK)x \in X_f$. Requiring the last predicted state to belong to this set gives the next optimisation a candidate, provided the model is exact and a feasible plan already exists.

To see this, suppose the accepted plan contains controls $u_0, \dots, u_{N-1}$ and states $x_0, \dots, x_N \in X$, with $x_N \in X_f$. Apply u_0. At the next time, shift the remaining controls one step and append Kx_N. Its last state is $(A + BK)x_N$, still in X_f. Every intermediate state and command came from the old feasible plan. This argument establishes recursive feasibility for the ideal deterministic system; the cost can then be chosen to establish a separate performance or stability claim.

For an unstable scalar plant $x^+ = 1.2x + u$, bounds $|u| \le 0.5,\ |x| \le 2$, choose $u = -0.7x$. The resulting factor is 0.5. Choosing $|x| \le 0.5/0.7 = 5/7$ protects the local input limit and the next-state magnitude contracts. This interval is one admissible terminal set; it need not be the largest invariant region obtainable with all possible controls.

Construct the two-step feasible region by working backwards

For the same plant, suppose a terminal interval is $[-b, b]$. A current state can reach it in one step if the reachable interval $[1.2x - 0.5, 1.2x + 0.5]$ intersects $[-b, b]$. Solving the two endpoint inequalities gives $|x| \le (b + 0.5)/1.2$, additionally restricted by $|x| \le 2$. Starting from $b_0 = 5/7 = 0.7143$, the one-step interval has half-width $b_1 = 1.0119$, and the two-step interval has $b_2 = 1.2599$. A current error $x = 1.20$ is therefore feasible for two steps but cannot reach the terminal set in one.

A feasible two-step plan applies $u_0 = u_1 = -0.5$. It generates $x_1 = 0.94,\ x_2 = 0.628$, inside the terminal interval. The appended local command is $-0.7(0.628) = -0.440$, generating $x_3 = 0.314$. From $x = 1.40$, however, even two maximal negative commands yield

$$x_1 = 1.18,\ x_2 = 0.916 > 0.7143.$$

A solver reporting infeasibility there has identified a physical horizon limitation. Lengthening the horizon can enlarge the feasible region. External disturbances and estimator error can spoil the exact shifted-plan construction, motivating reserve inside the hard bounds or an explicitly disturbance-aware terminal set.

Observe the limiting sequence of backwards-reachable interval widths. Ignoring the physical state cap for a moment, the recursion has fixed point $b = (b + 0.5)/1.2 = 2.5$. At $x = 2.5$ the strongest negative input merely holds position, and beyond it even that input allows outward divergence. With the stated state limit the physically relevant bound is tighter at 2.0. From $x = 2.0$ the strongest input moves to 1.9, so that boundary can be brought inward over enough samples. This examination distinguishes insufficient horizon from insufficient actuator authority. A solver change cannot rescue a state above the unrestricted 2.5 threshold with the given input, whereas increasing horizon can rescue a feasible physical state that lies just beyond the current backwards-reachable interval.

Now suppose implementation records establish an additive one-step disturbance bound $|w| \le 0.05$ in $x^+ = 1.2x + u + w$. At local feedback $u = -0.7x$, guaranteeing that $[-b, b]$ maps to itself under

every such disturbance requires

$$0.5b + 0.05 \le b.$$

The terminal interval above satisfies this because $b = 0.7143 > 0.10$, and the feedback still respects the input limit. To guarantee that a chosen predecessor reaches this interval in one step, the nominal next state needs to lie within $[-b + 0.05, b - 0.05]$. The robust predecessor half-width is thus $(b - 0.05 + 0.5)/1.2 = 0.970$, smaller than the nominal 1.012.

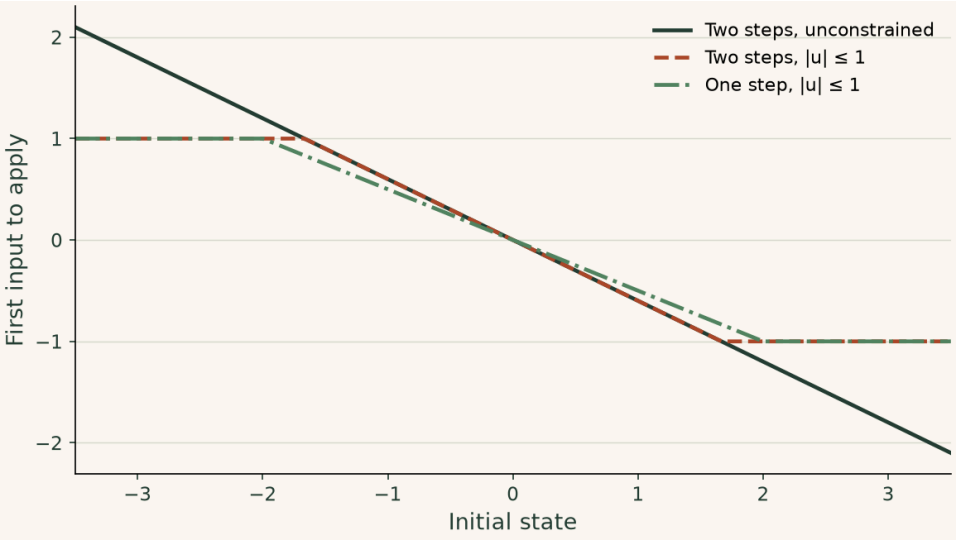
Applying the backward rule twice gives the robust two-step half-width $(0.970 - 0.05 + 0.5)/1.2 = 1.183$ when an updated action can be selected at the intermediate state. The previously feasible initial state 1.20 is outside that guarantee. Indeed maximal negative actuation with two consecutive positive disturbances yields

$$x_1 = 0.99,\ x_2 = 0.738 > 0.7143.$$

A nominally feasible trajectory and an invariant terminal set alone therefore do not make the whole finite-horizon plan robust. The intermediate constraints and the information available when choosing the next input must also be included in the guarantee.

Locate the first active constraint in a predictive policy

For $x^+ = x + u$, penalise the sum of squared future states and inputs. In one step the unconstrained action is $-x/2$. Eliminating the second action in a two-step plan gives first action $-3x/5$. Solve again on the active face once the input bound is reached.



The two-step policy reaches the limit at $|x| = 5/3$, earlier than the one-step policy at $|x| = 2$. In this scalar case the optimal first action agrees with clipping the linear feedback; in a coupled multivariable system componentwise clipping need not optimise the remaining plan.

▶ **Read values from the plot**

Interactive calculation · change the inputs of this example

Raise the actuator penalty and then reduce its magnitude limit. Compare when the one-step and two-step policies first saturate.

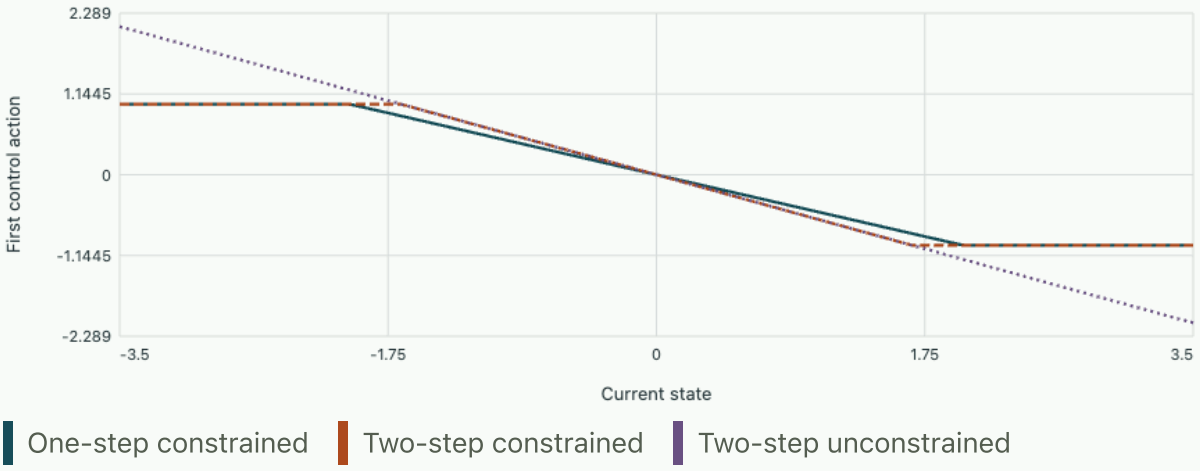
The curves below are recalculated from this section's equations in your browser. The fixed reference plot above remains available for comparison.

Open this live calculation

Try the calculation

Input cost / state cost 1

Actuator magnitude limit 1



Two-step action reaches the actuator limit at |x| = 1.667. Clipping is optimal for this scalar integrator example, not for an arbitrary multivariable plant.

Study the full derivation

Predicting a delayed process and testing a Smith controller · System dynamics, feedback and estimation

A heater is upstream of a temperature sensor. Its internal thermal response begins immediately while a material parcel reaches the sensor several seconds later. Build a feedback signal from two model outputs, derive which transfer function loses its delay, and then deliberately introduce a delay mismatch to see why a plausible prediction may fail.

Problem sheet 6 · Coupled plants and constrained predictions

End the course by testing an actuator pairing, a finite-horizon plan and a specification that initially looks feasible.

1. Unequal heat paths

For $G = \left(\begin{smallmatrix} 3 & 1 \\ 2 & 2 \end{smallmatrix}\right)$, (a) compute its relative gain array. (b) Calculate both actuator changes needed to raise output 1 by one unit while holding output 2 fixed. (c) Are these changes available if both actuators already sit at their lower bounds?

▼ **Solve the outputs before choosing a controller**

The determinant is 4, so the array has diagonal entries $6/4 = 1.5$ and cross entries $-2/4 = -0.5$. The desired change $G^{-1}(1, 0)^{\mathsf{T}} = (0.5, -0.5)^{\mathsf{T}}$ needs the second heater to reduce its command. At its lower bound this is infeasible. Reassigning loop pairings does not remove this physical constraint.

---

2. Active limits in a predictive move

Use the chapter's two-step integrator problem with initial state 3 and $|u|, |v| \le 1$. (a) Find the unconstrained plan. (b) Optimise the second action when the first is at its lower bound and compute the objective. (c) Can a requirement $x_{k+2} = 0$ be added?

▼ **Check the unconstrained optimum, then the active face**

Scaling the unconstrained plan gives $(-1.8, -0.6)$. Fix $u = -1$; the remaining derivative is $2(2 + v) + 2v$, giving $v = -1$, which is feasible. The input gradients at the lower bounds are nonnegative (the second is zero), establishing optimality for the convex objective. The states are 2 and 1, so the cost is $4 + 1 + 1 + 1 = 7$. The largest two-step reduction is 2, so the terminal equality is infeasible.

---

3. A requirement beyond actuator authority

A speed loop has steady plant $y = 2u - d$, with $|u| \le 2$, opposing disturbance $d = 1$ and desired speed 4 in normalised units. (a) Test whether any controller can track that command indefinitely. (b) Find the best possible steady output and error. (c) Give the measurements needed to distinguish saturation from an incorrect model gain in a logged experiment.

▼ **Apply the balance before retuning the loop**

Steady tracking requires $u = (4 + 1)/2 = 2.5$, beyond the upper bound. The maximum steady speed is 3, leaving error 1. Log the controller's requested action, the delivered actuator signal, the load and the

measured response. Verify the gain using smaller feasible steady commands; if delivered and requested signals diverge at the limit, the balance predicts the observed loss of tracking. Increasing integral gain cannot change the feasible steady range.

Commission a sampled loop with a limited actuator

A first-order plant has unit gain and time constant 1 s. A proportional controller of gain 3 holds its command for h seconds. With a unit reference and zero initial output, find the steady error, the sampling limit and the first three sampled outputs for h = 0.50 s. Then limit the command to [0, 1]. Predict the first sample at which the upper limit releases.

▼ **Worked solution**

Integration during one hold gives $y_{k+1} = ay_k+(1–a)u_k$ with $a = \exp(–h)$. Before clipping, $u_k = 3(1–y_k)$, the pole is $4a–3$, and the equilibrium is 0.75. The error 0.25 is independent of h for a stable loop. Requiring the pole to exceed –1 gives h less than ln 2 seconds. At h = 0.50 the pole is –0.5739; the other stability inequality holds for positive h.

The first three outputs are $0.75[1–(4a–3)^k]$ for k = 1, 2 and 3, approximately 1.180, 0.503 and 0.892. The first unconstrained command is 3. After clipping the actuator stays at one until a sample reaches $y \geq 2/3$, when the requested input no longer exceeds one.

While the limit is active the output ticks follow $y_k = 1–\exp(–0.5k)$. At k = 2 the output is 0.632 and the controller still requests 1.104. At k = 3 the output is 0.777 and the request is 0.669. The upper limit releases at 1.5 s. In the course experiment compare the unlimited trajectory first. Applying the scalar pole relation across a saturated interval would identify a different dynamics.

Explore the governing equations after solving the sheet: Interacting actuators, pairing and decoupling; Finite-horizon control with real actuator constraints. Compare the assumptions and units before transferring your chosen parameters; the controls may cover a different numerical range from the written exercise.

Chapter 13 · Proving a region of attraction with the energy of a pendulum

Pole locations describe a small neighbourhood of a nonlinear equilibrium. An energy argument can prove that a much larger set of initial conditions converges to the same resting point. Work out the region explicitly and test individual trajectories against its boundary.

Extract a storage function from a torque balance

Measure angle θ from the downward vertical, let $\omega = \dot{\theta}$, and use pivot inertia J and gravitational torque coefficient $K = mgl$. Add viscous damping b and a control torque $u = -k_d\omega$ that opposes motion. The closed dynamics are $J\dot{\omega} = -K\sin\theta - (b + k_d)\omega$. Define a scalar function with units of energy by adding kinetic energy to the integral of the restoring torque from zero angle:

$$V(\theta, \omega) = \frac{1}{2}J\omega^2 + K(1 - \cos\theta),$$

$$\dot{V} = J\omega\dot{\omega} + K\sin\theta\,\omega = -(b + k_d)\omega^2.$$

A positive damping coefficient ensures energy cannot increase. On the branch $-\pi < \theta < \pi$, V is positive away from the hanging rest state and zero there. Its derivative is zero whenever velocity vanishes, including ordinary turning points. To justify eventual convergence we need to establish whether any motion can remain indefinitely in this zero-derivative set.

Distinguish a turning point from an invariant state

At $\omega = 0$ a nonzero $\sin\theta$ produces angular acceleration, taking the trajectory immediately away from zero velocity. Inside the well the only state remaining forever at zero velocity is $(0, 0)$. The upright position sits at potential energy $2K$. Each closed sublevel $V \le c < 2K$ containing the downward rest state is compact and cannot leave the well, because reaching $\theta = \pm\pi$ would cost at least $2K$. LaSalle's invariance principle then establishes asymptotic convergence on that component.

$$|\omega_0| < \sqrt{\frac{2K}{J}(1 + \cos\theta_0)}$$

This expresses the strict sublevel as a speed allowance at each angle. A state outside it can still settle in the same well because damping removes energy before it reaches the top. The inequality is a sufficient certificate that can be tested from an initial measurement. Exactly at the upright rest state no damping torque acts and the system can remain there.

Compare two releases at the same angle

Let $J = 0.20$ kg m², $K = 2.0$ N m and $b + k_d = 0.30$ N m s/rad. At 60° the potential energy is 1.0 J and the certificate allows speeds below $\sqrt{30} = 5.48$ rad/s. At 4 rad/s the total is 2.6 J, safely below the 4 J

barrier. The turning angle cannot exceed $\arccos(1 - 2.6/2) = 107.5^\circ$ in magnitude. Positive damping reduces the actual maximum further.

At 6 rad/s the initial energy is 4.6 J, outside that guarantee. The instantaneous dissipation is $0.30(6)^2 = 10.8$ W, so dissipation before any top crossing may be substantial. Integrate the equation with the same initial angle and compare the true energy trace with the barrier. Moving the damping slider changes the rate of energy loss, while the sublevel calculated from J and K remains fixed.

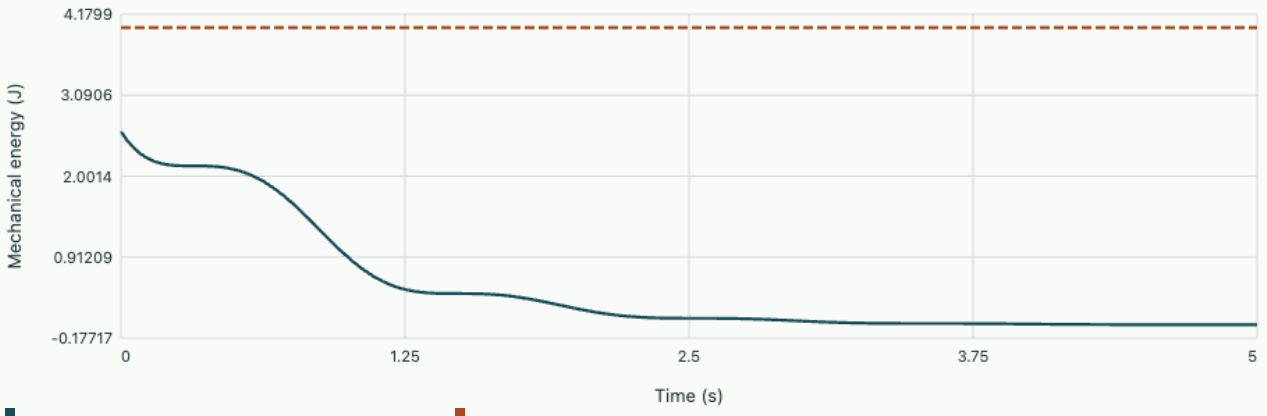
Raise the starting speed until the initial energy just exceeds the barrier. Compare a damped run with a zero-damping run. Use the two traces to distinguish a guaranteed sublevel from energy actually removed during a particular trajectory.

Open this experiment online

Try the calculation

Initial angular speed (rad/s) **4**

Total viscous damping (N m s/rad) **0.3**



The 60° release starts at 2.600 J and has 0.003 J after 5 s. Its starting sublevel stays inside the initial well.

Tutorial · design an admissible release speed

At 120° with the same parameters, choose a speed ensuring the trajectory remains below 90% of the barrier. What changes in the proof if the velocity gain cancels the physical damping exactly?

▼ **Show the energy budget**

The potential energy is

$$2(1 - \cos 120^\circ) = 3.0\,\mathrm{J}.$$

The target maximum total energy is 3.6 J, leaving 0.6 J of kinetic energy. Thus $|\omega_0| \le \sqrt{2(0.6)/0.20} = 2.45$ rad/s to the quoted precision. When $b + k_d = 0$, energy is constant. The sublevel remains invariant, but whole undamped orbits now lie in the zero-derivative set, so convergence to rest no longer follows.

Further reading · Åström and Murray, Feedback Systems · chapter 5, Lyapunov stability

Chapter 14 · Why integral action rejects an unknown constant load

A motor regulates speed while an unknown torque is applied to its shaft. A proportional controller allows a steady speed offset. Add a state that remembers the offset and derive the complete load transient as well as its eventual rejection.

Find the equilibrium before tuning poles

Use $J\dot{y} + by = u - d$, where y is angular speed, u is motor torque, and d is an opposing load. For $u = K_p(r - y)$ at constant reference r and load d the steady speed is $(K_p r - d)/(b + K_p)$. With r=0 and a positive load the speed settles below zero in these deviation variables. A constant motor torque is needed to cancel the load, yet pure proportional control at zero error produces zero torque.

Introduce $z' = r - y$ and $u = K_p(r - y) + K_i z$. At any equilibrium the condition $z' = 0$ requires $y = r$; the plant equation then requires

$$K_i z = br + d.$$

The controller state can store the torque needed at zero error. This reasoning assumes available actuator torque and closed-loop stability. If the actuator cannot supply $br + d$, no such regulated equilibrium exists.

Place the poles and solve a disturbance step

Set r=0 and transform from initial equilibrium with zero deviations. Eliminating z gives

$$\frac{Y(s)}{D(s)} = \frac{-s}{Js^2 + (b + K_p)s + K_i},$$

$$K_i = J\omega_n^2,$$

$$K_p = 2J\zeta\omega_n - b.$$

The numerator zero at s=0 removes constant disturbances in steady state. It follows from the integrator pole in the open loop. For a load step $D = d_0/s$, the initial rate is $-d_0/J$ regardless of gains because the feedback states have not had time to change. For a repeated real pole at $-a$, inversion yields

$$y(t) = -\frac{d_0}{J}\, te^{-at},$$

$$t_{\text{peak}} = 1/a,$$

$$|y_{\text{peak}}| = \frac{d_0}{Jea}.$$

A peak and recovery are expected even with exact final rejection. Increasing a improves that peak in the ideal model, at the price of higher control gains and possible sensitivity to sensor noise or actuator limits.

Quantify how a speed loop recovers

Let $J = 0.20$ kg m², $b = 0.50$ N m s/rad and choose $a = 4$ s⁻¹. The critical gains are $K_p = 1.10$ N m s/rad and $K_i = 3.20$ N m/rad. A step of 2 N m gives a peak deviation of −0.920 rad/s after 0.25 s. At 1 s it has recovered to −0.183 rad/s and ultimately reaches zero. The stored state supplies exactly 2 N m once speed has recovered.

With the same K_p but no integral action, the step gives $y(t) = -2[1 - e^{-(1.6/0.2)t}]/1.6$, tending to −1.25 rad/s. Its initial slope is also −10 rad/s². Use the two traces to identify an interval where a short measurement could make the controllers look similar, and compare their final speed offsets. For a disturbance of the opposite sign the curves reverse by linearity.

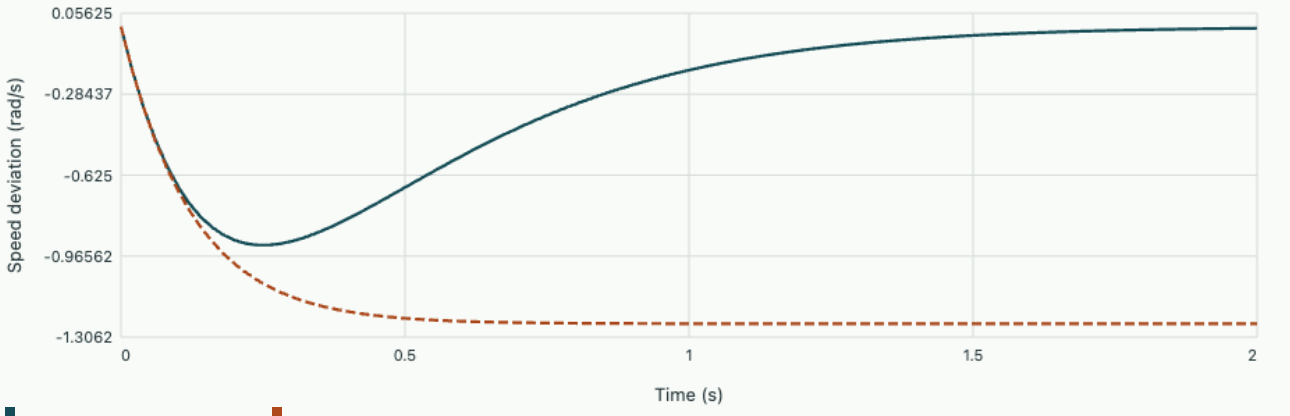
Double the pole rate and predict the new peak time and peak magnitude. Compare the very first slopes of PI and proportional-only responses. Which of the measurements could reveal increased bandwidth and which depends only on inertia?

Open this experiment online

Try the calculation

Critical pole rate (1/s) **4**

Opposing load step (N m) **2**



The PI deviation reaches -0.920 rad/s after 0.250 s and tends to zero. The proportional-only offset is -1.250 rad/s.

A related OneSimulation calculation

Use the course first-order feedback study to measure how proportional gain changes the transient. Compare the trends with the proportional-only curve above before introducing the extra integral state.

Run the OneSimulation feedback comparison

Tutorial · pick poles from a peak-error specification

A 1.5 N m load step must produce at most 0.50 rad/s peak error with $J = 0.20$ kg m² and the same b. Choose the smallest critical pole rate meeting this condition and calculate the gains. What torque must remain available after the transient?

▼ Show the bound and gains

Using the exact peak gives $a \geq 1.5/[0.20e(0.50)] = 5.52$ s⁻¹. Thus $K_i \simeq 6.09$ and $K_p \simeq 1.71$ in the units above. The motor needs at least 1.5 N m of extra steady torque to reject the new load at the original speed. Pole placement alone cannot establish the available transient or steady torque margin.

Further reading · Åström and Murray, Feedback Systems · chapter 7, integral action

Chapter 15 · Separating reference tracking from disturbance rejection

A positioning or speed loop can reject a load well but follow a demanded trajectory poorly. Add a feedforward signal computed from the requested motion, and derive precisely how mismatch and command smoothness affect tracking.

Use a physical balance to predict the required input

Take a first-order plant $J\dot{y} + by = u - d$ with positive inertia and drag. Suppose the commanded speed $r(t)$ and its derivative are available from a trajectory generator. From the plant balance the model input that would exactly sustain $y = r$ at zero disturbance is $\widehat{J}\dot{r} + \hat{b}r$, where hats indicate estimated parameters. Add ordinary proportional correction on the tracking error $e = r - y$:

$$u = \widehat{J}\dot{r} + \hat{b}r + K_p e,$$

$$J\dot{e} + (b + K_p)e = (J - \widehat{J})\dot{r} + (b - \hat{b})r + d.$$

This error equation separates three causes of deviation. Feedback changes the decay rate and attenuates each forcing term. With exact parameters, zero disturbance and correct initial conditions, arbitrary differentiable trajectories are tracked exactly in this ideal model. A discontinuous step calls for an impulse through the derivative term; a finite actuator cannot follow it. A generated smooth trajectory supplies a realizable alternative.

> **Background · Complementary sensitivity**
>
> $$T(s) = L(s)/[1 + L(s)]$$
>
> For the same feedback configuration S plus T is one. Measurement noise and plant-input disturbances reach the output on different paths.

Design a command with finite endpoint slopes

Move from zero to target R during an interval T with $r = R(3s^2 - 2s^3)$, $s = t/T$. It has zero velocity derivative at both ends and peak $\dot{r} = 1.5R/T$ at the midpoint. Doubling T halves the inertial part of the nominal torque without changing the final drag torque. An acceleration or torque budget can therefore determine trajectory duration before feedback gains are selected. A sensor derivative of a noisy measured speed is unnecessary here because $\dot{r}$ is computed from the commanded polynomial.

For a constant-slope segment $r = vt$, suppose drag is estimated exactly and the load is zero. The mismatch forcing is $(J - \widehat{J})v$. Away from segment transitions the settled tracking lag is $(J - \widehat{J})v/(b + K_p)$. At a hold after the segment the mismatch forcing disappears, and this lag decays with time constant $J/(b + K_p)$.

A tracking error that reveals the wrong inertia

Let $J = 0.20$, $b = 0.50$ and $K_p = 1.5$ in consistent SI units. The tracking time constant is 0.10 s. With $\widehat{J} = 0.15$, correct drag and a 4 rad/s² ramp, the lag approaches $0.05(4)/2 = 0.10$ rad/s. After 0.30 s it reaches

$$0.10(1 - e^{-3}) = 0.0950\,\mathrm{rad/s}.$$

A constant opposing load of 0.40 N m adds a steady 0.20 rad/s lag regardless of this inertia mismatch. Correcting feedforward inertia can improve tracking without changing load rejection.

For a cubic transition to 4 rad/s in 1 s, the midpoint has $r = 2$ rad/s and $\dot{r} = 6$ rad/s². The nominal motor torque at that instant is $0.20(6) + 0.50(2) = 2.20$ N m. Repeating the move in 2 s reduces the midpoint request to 1.60 N m; the final torque remains 2.0 N m for both durations. The plot integrates the error equation, so it also captures lag while the slope changes.

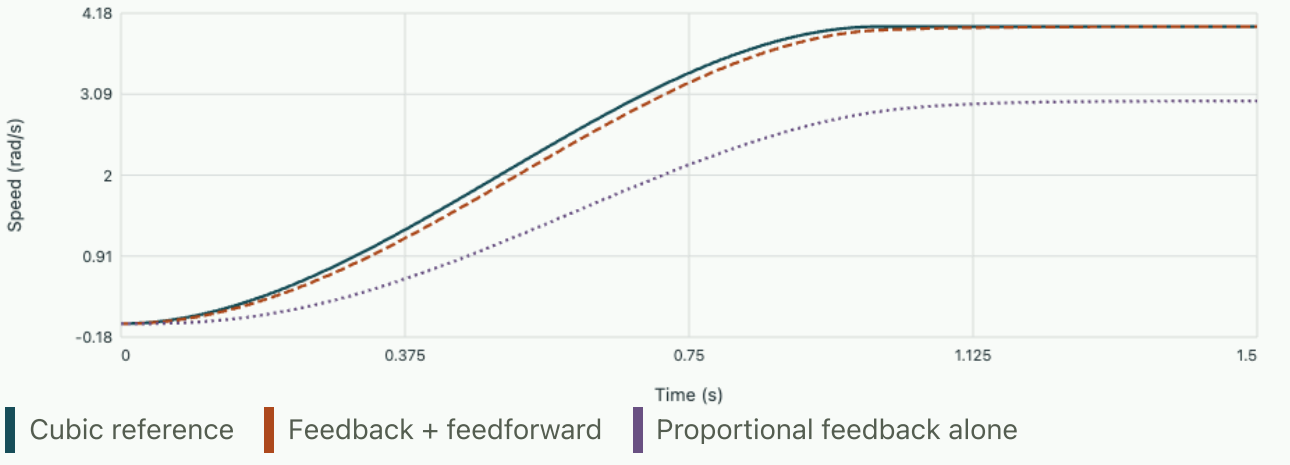
Set estimated inertia equal to the actual 0.20 kg m² before changing move duration. Then halve the estimate and compare lag during acceleration and during the hold. Explain how shortening move duration tests the available torque separately from feedback damping.

Open this experiment online

Try the calculation

Estimated inertia (kg m²) **0.15**

Move duration (s) **1**



The midpoint inertial-plus-drag feedforward command is 1.90 N m. The completed hold requires 2.00 N m at the target speed.

Tutorial · distinguish drag bias from inertia bias

With no external load, a long constant-speed hold at 4 rad/s leaves a 0.10 rad/s lag. Use the same true b and tuned K_p to estimate $\hat{b}$. Can a measurement during this hold identify the feedforward inertia as

well?

▼ **Show the identification argument**

At constant reference $\dot{r} = 0$ and at equilibrium $\dot{e} = 0$. Thus $0.10 = (0.50 - \hat{b})(4)/2$, giving $\hat{b} = 0.45$ N m s/rad. The inertia term vanishes identically on this data segment. A segment with nonzero $\dot{r}$, ideally repeated at opposite slopes at the same speed, supplies an independent observation of inertial mismatch.

Further reading · Åström and Murray, Feedback Systems · chapter 12, feedforward design

Chapter 16 · How much transport delay a feedback loop can tolerate

A communication or transport delay changes phase at every frequency while leaving gain untouched. Use that property to derive the first stability boundary for a specific loop and to interpret the phase margin in units of seconds.

Locate the critical imaginary-axis crossing

Use negative unity feedback with $L(s) = Ke^{-s\tau}/(1 + sT)$, where K is dimensionless, T is a plant time constant and τ is a pure delay. With $\tau = 0$ and positive K the single closed-loop pole is $-(1 + K)/T$. At a purely imaginary characteristic root $s = j\omega$, the condition $1 + L(j\omega) = 0$ requires both unit magnitude and an odd multiple of $-\pi$ phase. Since the delay has magnitude one, the unique positive gain-crossover frequency for $K > 1$ obeys

$$\frac{K}{\sqrt{1 + (\omega_c T)^2}} = 1,$$

$$\omega_c = \frac{\sqrt{K^2 - 1}}{T}.$$

The undelayed plant contributes $-\arctan(\omega_c T)$. The delay contributes an unwrapped phase $-\omega_c\tau$ in radians. The first critical delay therefore satisfies

$$\tau_* = \frac{\pi - \arctan(\omega_c T)}{\omega_c},$$

$$\mathrm{PM}(\tau) = \pi - \arctan(\omega_c T) - \omega_c\tau.$$

Further crossings at the same frequency would require another 2π of delay phase. The first one governs the delay margin because the zero-delay system is stable. For $0 < K < 1$ no positive frequency reaches unit gain in this particular first-order model. At $K = 1$ the only unit magnitude occurs at zero frequency where the phase is zero.

Background · Loop delay

$$L_{\mathrm{delayed}}(i\omega) = L(i\omega)e^{-i\omega\tau}$$

A pure delay leaves the magnitude unchanged and adds frequency-dependent phase. Use angular frequency when converting a margin to seconds.

Check that the first crossing enters instability

The characteristic equation can be written $1 + sT + Ke^{-s\tau} = 0$. Implicit differentiation and substitution of $Ke^{-s\tau} = -(1 + sT)$ give

$$ds/d\tau = -s(1+sT)/[T+\tau(1+sT)].$$

To test the direction it is simpler to invert this derivative:

$$\left(\frac{ds}{d\tau}\right)^{-1} = -\frac{T}{s(1+sT)} - \frac{\tau}{s},$$

$$\Re\left[\left(\frac{ds}{d\tau}\right)^{-1}\right]_{s=j\omega_c} = \frac{T^2}{1+\omega_c^2 T^2} > 0.$$

A complex number and its reciprocal have real parts of the same sign. Thus the conjugate pair enters the right half-plane when the delay increases past the first critical value. The oscillation period there is $2\pi/\omega_c$.

Latency budget at two loop gains

With $T = 0.50$ s and $K = 2$, $\omega_c = 3.464$ rad/s. The undelayed phase at crossover is −60°, leaving 120° or 2.094 rad. The critical delay is 0.605 s and the crossing period is 1.81 s. An installation with 0.30 s of delay retains about 60.5° of phase margin.

Raise K to 4 with the same T. Then $\omega_c = 7.746$ rad/s, the undelayed phase is −75.5°, and the delay margin falls to 0.235 s. The same 0.30 s installation would cross the stability boundary. Reading a phase plot wrapped modulo 360° can conceal such a crossing. Use the unwrapped curves and locate the frequency where gain equals one before judging stability.

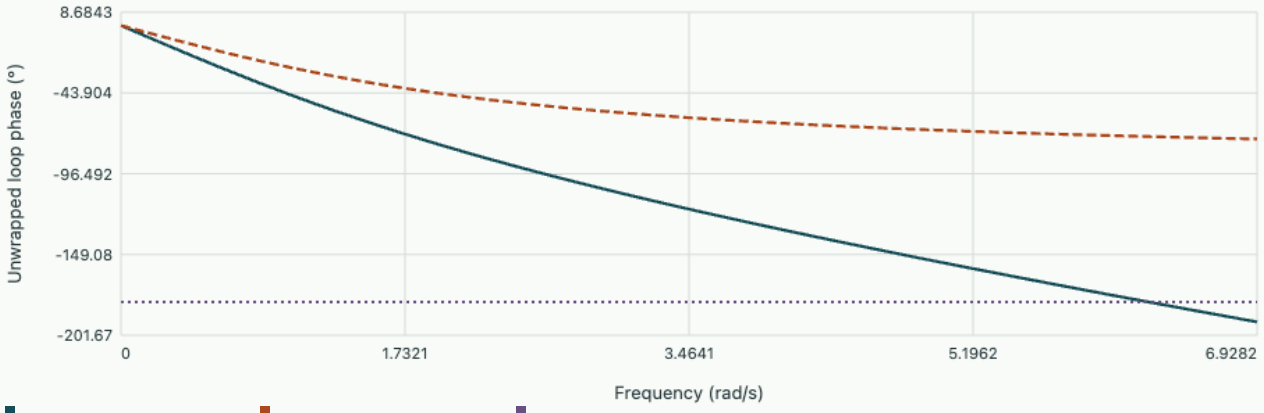
At fixed loop gain increase delay until the reported margin vanishes. Use the frequency of that point to estimate the oscillation period. Repeat after increasing gain and check whether the same physical time delay spends the same phase budget.

Open this experiment online

Try the calculation

Loop gain **2**

Delay (s) **0.3**



Loop with delay · Undelayed loop · Critical phase

Gain crossover is 3.464 rad/s. The remaining phase margin is 60.5° and the first critical total delay is 0.605 s.

A related OneSimulation calculation

Measure how a gain change affects the undelayed first-order closed loop in the course laboratory. Then use its increased bandwidth to explain the reduced delay budget calculated here.

Open the OneSimulation first-order feedback laboratory

Tutorial · allocate a reserve in degrees

With $K = 2$, $T = 0.50$ s require at least 45° phase margin. Calculate the largest total delay and the spare time when fixed transport already takes 0.20 s. Explain the effect of using degrees directly in $\omega\tau$.

▼ **Show the phase budget**

Subtract the required 45° from the undelayed 120°, leaving 75° for delay. Converting to radians gives $1.309/3.464 = 0.378$ s total. The remaining budget is about 0.178 s. The expression $\omega\tau$ is in radians for ω in rad/s. Dividing the number 75 directly by 3.464 would overestimate the budget by $180/\pi$.

Further reading · Åström and Murray, Feedback Systems · chapter 10, Nyquist criterion and stability margins

Study the full derivation

Predicting a delayed process and testing a Smith controller · System dynamics, feedback and estimation

A heater is upstream of a temperature sensor. Its internal thermal response begins immediately while a material parcel reaches the sensor several seconds later. Build a feedback signal from two model outputs, derive which transfer function loses its delay, and then deliberately introduce a delay mismatch to see why a plausible prediction may fail.

Chapter 17 · Why a stable plant can initially move in the wrong direction

Two processes may act on a measured output at different speeds and with opposite signs. Their combined output can first fall after a positive command and later rise. Derive that transient and see why trying to cancel its zero creates an unstable internal state.

Build a right-half-plane zero from two stable responses

Let a unit input drive two stable first-order processes in parallel. A slow path has gain $a > 0$ and time constant T_s; a fast path is subtracted, with gain $b > 0$ and time constant $T_f < T_s$. The measured output transfer function is

$$G(s) = \frac{a}{1 + sT_s} - \frac{b}{1 + sT_f} = \frac{(a - b) + (aT_f - bT_s)s}{(1 + sT_s)(1 + sT_f)}.$$

Choose $a > b$ so the final step response is positive. Its initial slope is $a/T_s - b/T_f$. When this slope is negative, $bT_s > aT_f$, so the numerator has a real positive zero $z = (a - b)/(bT_s - aT_f)$. Each storage process has a stable pole. The right-half-plane zero expresses the way their contributions are combined in the output.

Locate the undershoot without a numerical search

For a unit step apply each path's time-domain solution and subtract. Setting the time derivative to zero and taking logarithms gives the time of minimum.

$$y(t) = a(1 - e^{-t/T_s}) - b(1 - e^{-t/T_f}),$$

$$t_{\min} = \frac{\ln[bT_s/(aT_f)]}{1/T_f - 1/T_s}.$$

For the inverse-response conditions this is positive. The steady value diagnosis alone would see only $a - b$; a short test would reveal mostly the fast negative path. Identifying both requires measurements on both time scales. When the fast and slow slopes exactly match, the numerator loses its s term and the finite zero moves to infinity.

Read a time-domain constraint directly from the zero

For a bounded causal unit-step response $Y(s) = G(s)/s$. At the zero $Y(z) = 0$, hence

$$\int_0^\infty e^{-zt} y(t)\, dt = 0.$$

An output ultimately settling to a positive constant cannot stay positive at every time while obeying this weighted integral. A controller implementing the exact inverse $1/G$ would have a pole at z and an

exponentially growing response to small errors in its internal state. Shaping the command over a longer interval can reduce an undershoot without introducing that unstable inversion.

Two experiments with the same final reading

Take $a = 2,\ b = 1,\ T_s = 4$ s and $T_f = 0.50$ s. The zero is at $1/3$ s^{-1}. The initial slope is $0.5 - 2 = -1.5$ s^{-1} and the minimum occurs at 0.792 s. Evaluating the response gives about −0.435 before it ultimately reaches +1. This initial dip can look like a sensor wired with the wrong sign until the slow path has had time to act.

Lengthen the fast path to 1.5 s. The same final value +1 now accompanies a less negative initial slope of −0.167 s^{-1}. The zero moves to $1/(4 - 3) = 1$ s^{-1} and the minimum occurs at 0.690 s with an undershoot about −0.052. Try moving the fast time constant through 2 s. Predict the initial slope and disappearance of the undershoot before examining the curves.

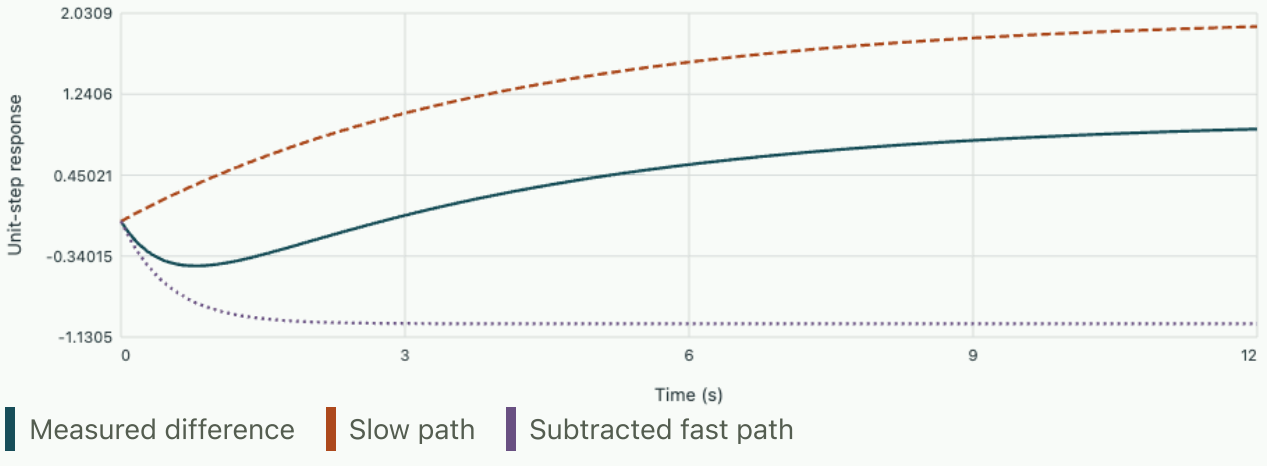
Increase the fast time constant through the equal-slope value of 2 s at the default gains. Predict the disappearance of the negative excursion. Reduce the subtracted gain and derive the new equal-slope condition before repeating the test.

Open this experiment online

Try the calculation

Fast-path time constant (s) **0.5**

Subtracted gain **1**



Initial slope is -1.500 per second and final gain is 1.00. The minimum is -0.436 at 0.792 s.

Tutorial · infer a hidden fast time constant

A fitted slow path has $a = 2,\ T_s = 4$ s. The steady step response is 1 and the measured initial slope is −0.75 s^{-1}. Infer b, T_f, the zero position and the time of minimum.

▼ **Show the parameter recovery**

The final reading gives

$$b = a - 1 = 1.$$

The initial slope gives $1/T_f = 2/4 + 0.75 = 1.25$ s⁻¹, so $T_f = 0.80$ s. The zero is $1/(4 - 1.6) = 0.417$ s⁻¹ and $t_{\min} = \ln(4/1.6)/(1.25 - 0.25) = 0.916$ s. Combining a static observation with a dynamic one makes both path parameters identifiable in this ideal two-path model.

Further reading · Åström and Murray, Feedback Systems · chapter 14 on fundamental limits

Chapter 18 · Designing a lead network at a specified gain crossover

A loop is too lightly damped at the bandwidth needed for tracking. A lead network can supply positive phase over a frequency band. Locate the band mathematically, set its gain crossover deliberately and measure the accompanying high-frequency gain increase.

Build a lead element from one pole and one zero

Take $C(s) = K(1 + sT)/(1 + sT/\alpha)$ with $\alpha > 1$. The zero occurs at $1/T$ and the pole at the higher frequency α/T. Its phase contribution is $\phi = \arctan(\omega T) - \arctan(\omega T/\alpha)$. At very low or very high frequency that phase tends to zero. Differentiate with respect to $z = \omega T$ to locate the interior maximum:

$$\frac{1}{1+z^2} - \frac{1/\alpha}{1+z^2/\alpha^2} = 0,$$

$$z = \sqrt{\alpha},$$

$$\sin\phi_{\max} = \frac{\alpha - 1}{\alpha + 1}.$$

The maximum occurs at the geometric mean of the zero and pole frequencies. At this same frequency the magnitude of the dynamic part is $\sqrt{\alpha}$, obtained by substituting $z^2 = \alpha$ into $\sqrt{(1+z^2)/(1+z^2/\alpha^2)}$. Thus positive phase comes with a magnitude change. Adding lead without resetting scalar K can move crossover to a frequency where the anticipated plant phase and lead are different.

Set the phase and magnitude at a common design point

For a stable minimum-phase design example with an open-loop integrator, take $G(s) = 1/[s(1 + s)]$ and select $\omega_c = 2$ rad/s. Plant phase there is

$$-90^\circ - \arctan 2 = -153.43^\circ.$$

Choose 30° extra lead, giving nominal phase margin 56.57°. Invert the peak-phase relation and put the peak at the target frequency:

$$\alpha = \frac{1 + \sin\phi_{\max}}{1 - \sin\phi_{\max}},$$

$$T = \frac{\sqrt{\alpha}}{\omega_c},$$

$$K = \frac{1}{|G(i\omega_c)|\sqrt{\alpha}}.$$

A scalar-gain controller chosen for the same crossover would have $K_P = 1/|G(i\omega_c)|$ and margin only 26.57°. This comparison holds crossover fixed, so the phase advantage can be attributed to the

compensator. The return ratio magnitude must still be checked away from crossover, especially for a plant with additional modes or delay.

Calculate the network and identify a robustness cost

For the 30° design $\alpha = 3,\ T = 0.866$ s, giving zero and pole frequencies of 1.155 and 3.464 rad/s. Plant magnitude at 2 rad/s is $1/(2\sqrt{5}) = 0.2236$, so K=2.582. The high-frequency control gain tends to $K\alpha = 7.746$, while its low-frequency gain is 2.582. High-frequency measurement noise passed through a feedback controller can therefore induce larger actuator activity than a low-frequency calculation suggests. The actual noise transfer to the input must include the loop sensitivity.

Require a more aggressive 45° boost at the same crossover. Then $\alpha = 5.828,\ T = 1.207$ s and $K = 1.852$. Margin becomes 71.57° and the high-frequency gain reaches 10.80. The larger boost also lowers the DC controller gain, which changes low-frequency disturbance attenuation in this plant. Comparing the two phase traces alongside their designed crossover avoids selecting a controller purely from its maximum phase number.

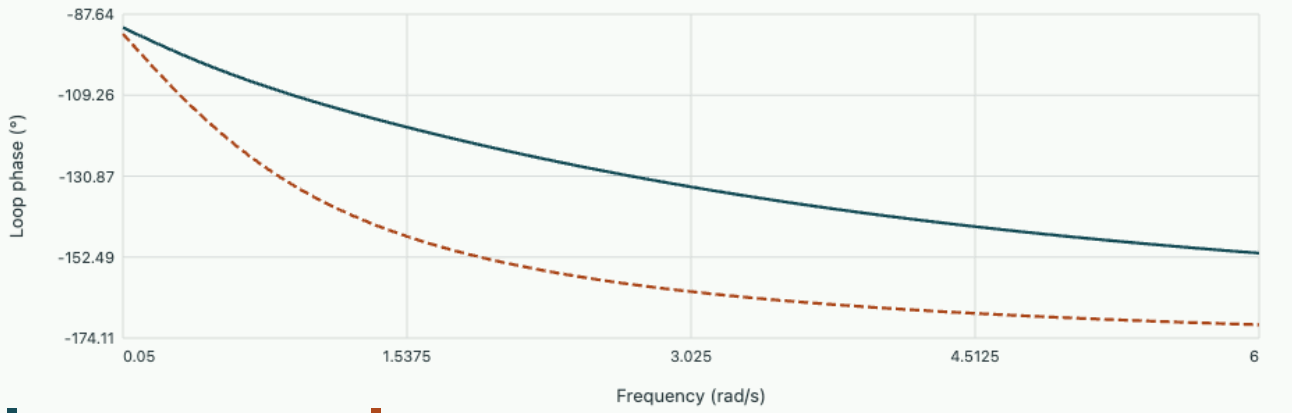
Keep crossover at 2 rad/s and increase the required phase from 30 to 45 degrees. Compare the high-frequency controller gains. Now fix lead and move crossover upward to inspect how plant phase can consume the added margin.

Open this experiment online

Try the calculation

Lead at crossover (°) **30**

Gain crossover (rad/s) **2**



At crossover, nominal margin is 56.57°. The required scalar gain is 2.582 and the high-frequency controller gain is 7.746.

Tutorial · allow for a known delay

Return to the 30° design and insert a 0.10 s pure delay. Compute the remaining phase margin at 2 rad/s. What maximum additional delay is compatible with a required 40° margin? Why is there no gain crossover shift for this change?

▼ **Spend the phase reserve in radians**

At crossover 0.10 s adds $2(0.10) = 0.20$ rad or 11.46° of phase lag. Margin falls to 45.11°. To retain 40° permits a further $(45.11 - 40)\pi/[180(2)] = 0.0446$ s. Pure delay has unit magnitude at every frequency, so the designed crossover stays fixed. In contrast, adding a physical low-pass filter would change both gain and phase and require a fresh crossover calculation.

Further reading · Åström and Murray, Feedback Systems · chapter 12, feedback design by loop shaping

Study the full derivation

Sensitivity, bandwidth and the integral price of disturbance rejection · System dynamics, feedback and estimation

A regulator attenuates a slowly varying output load yet amplifies another disturbance at a higher frequency. Separate the paths by which reference, measurement noise and physical disturbance reach the recorded variable, then derive an integral identity that explains why a stable loop can redistribute sensitivity across frequency.

Chapter 19 · The period and duty cycle of a thermostat with hysteresis

A two-position thermostat can have a stable, repeatable temperature oscillation even though neither operating state has an equilibrium inside the chosen control band. Use event times to calculate the cycle and test its heat balance.

Couple a thermal balance to a discrete memory state

Let a uniform thermal mass have heat capacity C, conductance G to ambient at T_a, and heater power P when the switch state $s = 1$. With heater off $s = 0$. The continuous state obeys $C\dot{T} = -G(T - T_a) + Ps$. The relay turns on when cooling reaches T_-, and turns off when heating reaches $T_+ > T_-$. Between thresholds it retains the most recent switch state. Thus specifying a temperature inside the band without the binary state does not fully specify the future.

$$\tau = C/G,$$

$$T_{\text{on}} = T_a + P/G,$$

$$T_{\text{off}} = T_a,$$

$$T(t) = T_{\text{eq}} + [T(0) - T_{\text{eq}}]e^{-t/\tau}.$$

Both thresholds must be reachable in their respective modes for periodic cycling: $T_a < T_- < T_+ < T_{\text{on}}$. If $T_{\text{on}} = T_+$ the threshold is approached only asymptotically in this ideal model. A temperature outside the band initially can lead to a transient before the repeating orbit begins.

Solve the crossing conditions exactly

Start at T_- with the heater on. Solve the exponential trajectory for the time to T_+. Switch state, set that endpoint as the cooling initial condition and solve for the return time. Each logarithm compares distances from the relevant equilibrium:

$$t_{\text{on}} = \tau \ln \frac{T_{\text{on}} - T_-}{T_{\text{on}} - T_+},$$

$$t_{\text{off}} = \tau \ln \frac{T_+ - T_a}{T_- - T_a},$$

$$D = \frac{t_{\text{on}}}{t_{\text{on}} + t_{\text{off}}}.$$

Integrate the energy equation over one complete period. Net stored energy change is zero, so $PD = G(\bar{T} - T_a)$. This gives the actual time-average temperature directly from the duty ratio. The arithmetic midpoint of the thresholds need not be the time average, because heating and cooling speeds vary across the swing.

Choose band width and estimate switching frequency

Let $C = 2000$ J/K, $G = 10$ W/K, $P = 200$ W and $T_a = 15°$C. The time constant is 200 s and the heating equilibrium is 35 °C. For thresholds 20 and 22 °C the on-time is $200 \ln(15/13) = 28.62$ s and the off-time $200 \ln(7/5) = 67.29$ s. Period is 95.9 s and duty 29.84%. Mean power is 59.7 W and the implied mean temperature 20.97 °C. The shorter heating interval reflects the larger temperature distance from the heated equilibrium.

Narrow the band to 20.5 and 21.5 °C at the same midpoint. The on-time falls to 14.29 s and off-time to 33.41 s. Switching frequency roughly doubles while duty remains close to 30%. In a real relay, sensor lag can permit overshoot beyond thresholds. A narrow hysteresis band then ceases to predict the observed extrema. Increasing numerical resolution in a continuous integrator only helps if its logic also detects the threshold crossing at the proper time.

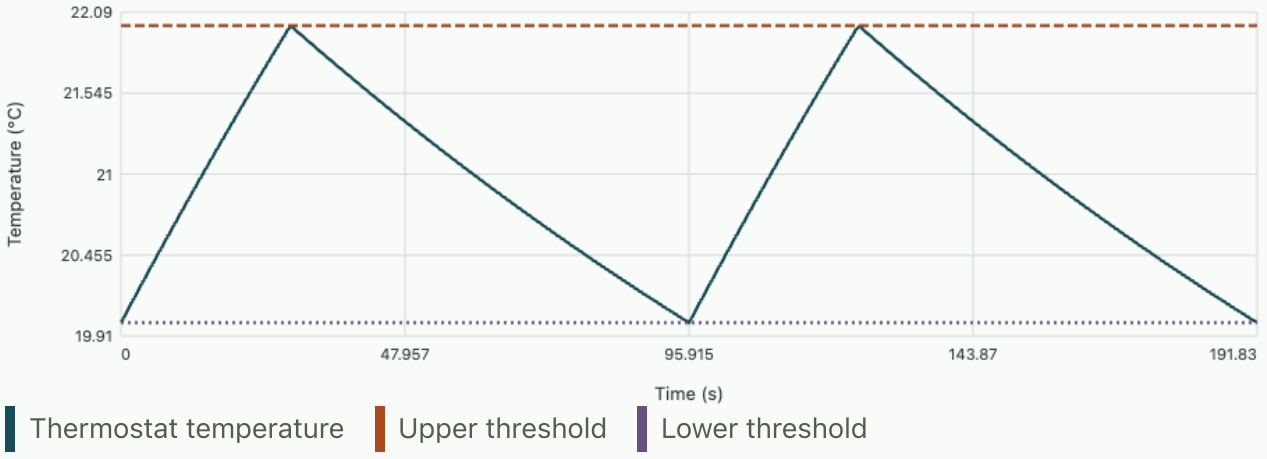
Narrow the switching band by half while keeping its midpoint fixed. Predict the switching frequency and mean heating power, then read both event times. Raise heater power and identify which branch of the thermal trajectory should remain unchanged.

Open this experiment online

Try the calculation

Band half-width about 21 °C (K) **1**

Heater power (W) **200**



Heating takes 28.62 s and cooling takes 67.29 s. Duty is 29.84% and energy balance gives a mean temperature of 20.97 °C.

Tutorial · infer the thermal capacitance from one cycle

Suppose the wider band gives a measured cooling time of 80.0 s with ambient at 15 °C and the same conductance. Infer C and predict heating time if heater power remains 200 W. If the observed heating time disagrees but the cooling curve fits, identify one additional independent measurement.

▼ **Infer the time constant then check the energized mode**

The cooling expression gives

$$\tau = 80 / \ln(7/5) = 237.8\,\mathrm{s},$$
$$C = G\tau = 2378\,\mathrm{J/K}.$$

Predicted heating time is

$$237.8 \ln(15/13) = 34.0\,\mathrm{s}.$$

Measuring electrical heating power or the long-time temperature reached under a continuous on command tests P/G. The cooling branch alone identifies the ratio C/G and does not determine that energized equilibrium.

Further reading · UC Berkeley Ptolemy Project · thermostat as a continuous plant controlled by a finite-state machine

Chapter 20 · Learn a repeatable positioning move from its previous error

An inspection stage follows the same position command on hundreds of nominally identical passes. A small force disturbance returns at the same part of each move. The complete tracking record from the previous pass is available before the next move begins. Derive a learning update that uses the stage dynamics to turn that record into a new feedforward command. Its convergence can be predicted from singular values and tested with a few lines of code.

Separate time within the move from the number of the move

Let $k = 0, \dots, N-1$ index input samples during a single pass, and let j count completed learning updates between passes. Before every pass return the stage to the same initial state. The reference trajectory, sampling period and number of samples must match from one pass to the next. If the fixture slips in time or the first sample corresponds to a slightly different position, an otherwise repeatable disturbance appears to move through the record. Alignment of trials is thus a physical part of implementing the mathematical assumptions.

Take the simple sampled model $x_{k+1} = a x_k + (1-a) u_k$, with $0 < a < 1$ and $x_0 = 0$. It can represent a position servo whose fast inner control loop has already been stabilised. The outer experiment changes its feedforward command between repetitions. A unit command produces a downstream effect starting at the *next* sample. Stack the commands as $(u_0, \dots, u_{N-1})$ and the readings as $(y_1, \dots, y_N)$. This ordering keeps the direct response of each command on the diagonal of a square matrix. Stacking y_0 instead introduces a row carrying no information about any command.

For the first three readings, direct substitution gives $x_1 = (1-a)u_0$,

$$x_2 = (1-a)(a u_0 + u_1),$$

$$x_3 = (1-a)(a^2 u_0 + a u_1 + u_2).$$

Repeated substitution produces a lower triangular matrix G. In zero-based row and column indexing $G_{kn} = (1-a)a^{k-n}$ for $k \geq n$, and zero otherwise. Its columns are shifted, truncated impulse responses. Assemble the repeatable output disturbance into a vector d. For the jth pass the measured output is $\mathbf{y}_j = G\mathbf{u}_j + \mathbf{d}$.

Use virtual changes of the command to obtain the update direction

Let $\mathbf{r}$ be the desired trajectory and define error $\mathbf{e}_j = \mathbf{r} - \mathbf{y}_j$. Consider the scalar tracking objective $J(\mathbf{u}) = \|\mathbf{r} - G\mathbf{u} - \mathbf{d}\|^2 / 2$. A perturbation $\delta\mathbf{u}$ changes the measured output by $G\delta\mathbf{u}$. Multiplying out the new squared error and retaining the linear term gives

$$\delta J = -\mathbf{e}^T G \delta\mathbf{u}.$$

The gradient in input coordinates is therefore $-G^T\mathbf{e}$. Taking a step in the opposite direction gives the update

$$\mathbf{u}_{j+1} = \mathbf{u}_j + \gamma G^T \mathbf{e}_j,$$

with a positive learning step γ. Each input sample receives a weighted sum of subsequent output errors from the preceding trial. This apparent use of the future is ordinary causality when the two time indices are kept separate. At the moment the new command file is written, every sample from the previous pass is already known. No sensor can deliver the future errors of the pass currently being executed.

For a three-sample example let $a = 1/2$ and take the error record $(0, 0, 1)$. The last row of G is $(1/8, 1/4, 1/2)$. Consequently

$$G^T \mathbf{e} = (1/8, 1/4, 1/2).$$

An error in the final output calls for some correction to each earlier input because the plant remembers each command. Assigning the entire correction to the input with the same array index would neglect two of the available physical paths. An imperfect model can give poorly scaled weights, which is one reason to examine actual error after each update.

Derive a bound for every repeatable error direction

Substitute the new command into the output equation and subtract from the reference. The unknown disturbance cancels because it is the same vector in both trials. The result is $\mathbf{e}_{j+1} = (I - \gamma G G^T)\mathbf{e}_j$. This gives a closed dynamical equation *across trials*. Its eigenvalues describe learning convergence; they are not the within-trial pole a of the servo.

Take a singular value decomposition $G = U \Sigma V^T$. Expressing the output error in the orthogonal basis U shows that its ith component is multiplied by $1 - \gamma \sigma_i^2$. Thus all attainable directions contract monotonically in the Euclidean norm provided

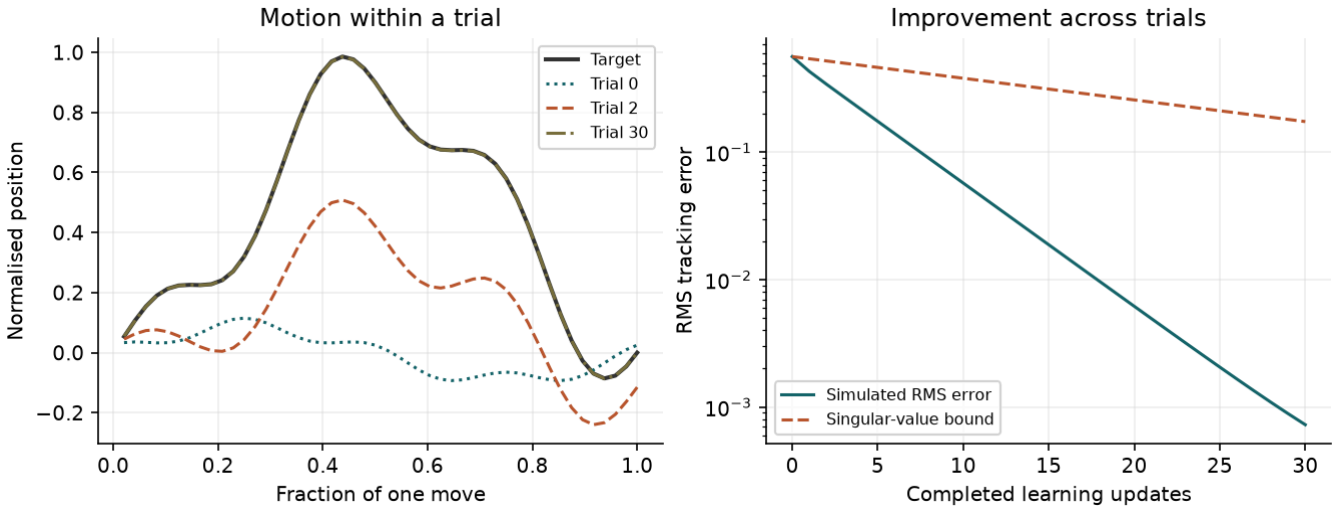
$$0 < \gamma < \frac{2}{\sigma_{\max}^2}, \qquad \|\mathbf{e}_{j+1}\|_2 \le q \|\mathbf{e}_j\|_2, \qquad q = \max_i |1 - \gamma \sigma_i^2|.$$

Here G is nonsingular because its diagonal is nonzero. A more general stage with unobservable samples, missing actuators or pure delay longer than the chosen alignment can have zero singular values. Components of error outside its reachable output space then persist, however many trials are available. Inspecting the norm alone cannot reveal whether a plateau comes from this structural limit, measurement noise or a deliberately imposed actuator bound.

The slow directions have small singular values: a substantial command correction produces only a small output change. Inverting G directly would eliminate the deterministic error in one idealised iteration, but the same inversion magnifies sensor noise and can demand impractically oscillatory commands. Gradient learning spends many more trials on these directions while retaining a tunable update size. For a positive full-rank spectrum, balancing the two endpoint contraction magnitudes gives the ideal minimax step $2/(\sigma_{\max}^2 + \sigma_{\min}^2)$. Its usefulness still depends on how accurate the assumed plant is.

Open editable notebook Download .ipynb

Read the stage trajectory and the trial history together



Left, each curve is a complete motion in time. Right, each point is a statistic calculated from an entire completed motion. Movement along the two horizontal axes represents different experiments.

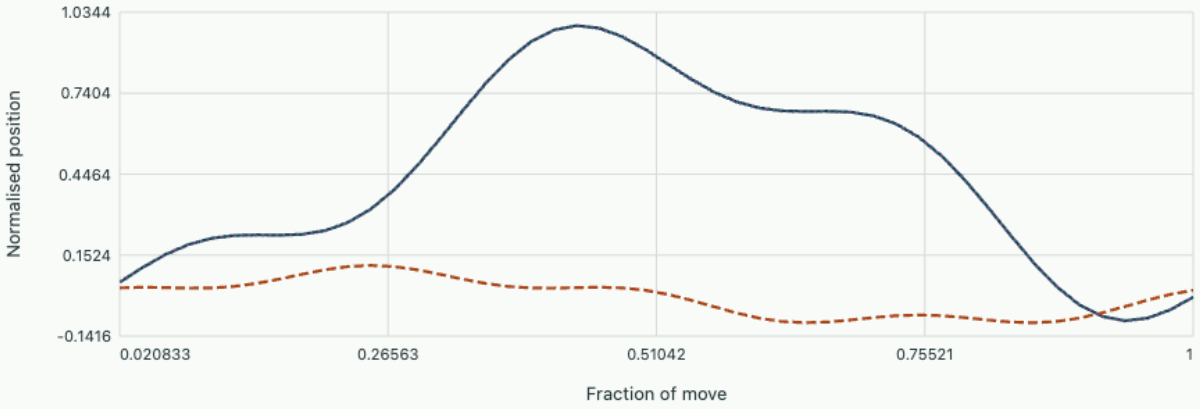
Interactive calculation · Learning motion

Change the physical and numerical inputs of the notebook model. Compare the updated curves and balance checks with the recorded notebook figures. Each plot uses the same equations and assumptions documented in that notebook; its readout names any analytical or reduced-model reference.

Open this live calculation

Try the calculation

Sampled plant memory	0.75
Fraction of limiting learning step	0.9
Learning updates	30



Learning gain 1.879, RMS error 0.56264 → 0.00073. Largest squared singular value is evaluated by100 power iterations. Values above the limiting-step fraction1 can destabilise an error component.

Download this example's editable notebook

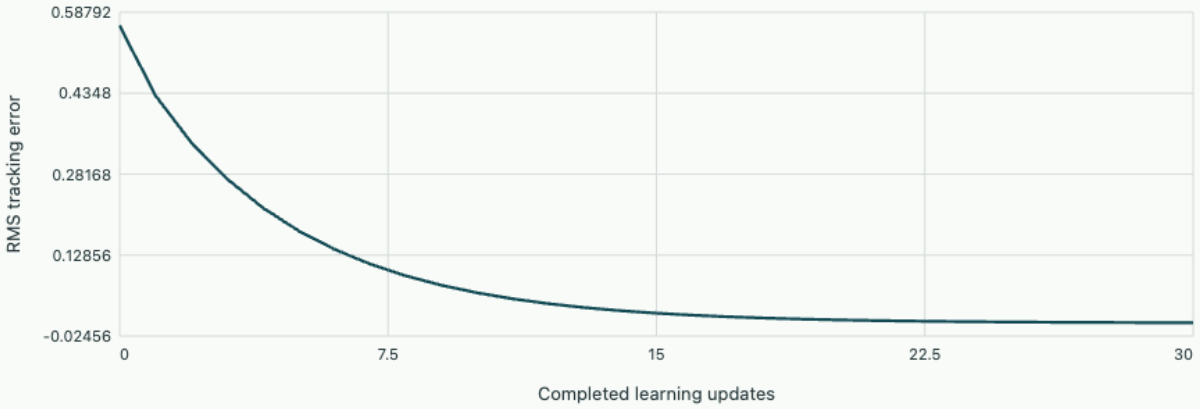
Interactive calculation · Learning errors

Change the physical and numerical inputs of the notebook model. Compare the updated curves and balance checks with the recorded notebook figures. Each plot uses the same equations and assumptions documented in that notebook; its readout names any analytical or reduced-model reference.

Open this live calculation

Try the calculation

Sampled plant memory	0.75
Fraction of limiting learning step	0.9
Learning updates	30



Error after each update

Learning gain 1.879, RMS error 0.56264 → 0.00073. Largest squared singular value is evaluated by100 power iterations. Values above the limiting-step fraction1 can destabilise an error component.

Download this example’s editable notebook

Use 48 commands, an inner-loop pole of 0.75 and a smooth reference combined with a smaller oscillation. Add a repeatable sinusoidal displacement to the output as a model of the fixture. The largest singular value is 0.9788 and the smallest is 0.1429. Choosing 90% of the upper admissible step gives $\gamma = 1.879$ and $q = 0.9616$. In the notebook the RMS tracking error falls from 0.56264 to about 0.00073 after thirty updates. The norm bound predicts a slower worst-case decline because the chosen reference has limited content along some slowly learned directions.

A change of step can produce alternating signs in a well-observed error direction while its magnitude still shrinks. For the strongest direction at the default step the multiplier is

$$1 - 1.8 = -0.8.$$

Comparing signed errors at one time instant may look oscillatory even though the total norm decreases. Conversely, a deliberately excessive step can produce several initially improving records if the unstable direction is weakly represented in the original error. Evaluate a proposed step with the singular spectrum and with independent repeated trials.

Account for trial variation before interpreting a plateau

Let the disturbance differ between passes. Redoing the same subtraction now adds $\mathbf{d}_j - \mathbf{d}_{j+1}$ to the nominal error update. A slowly changing load introduces drift in trial number. Independent sensor noise produces a new random input to every update. Learning too aggressively can write the noise of one trial into the command of the next. To distinguish this mechanism from an unlearnable repeatable error, run several identical command files without updating and compare the within-sample scatter across the repeats.

Other practical departures have different signatures. An amplitude-limited actuator clips the proposed command and breaks the linear update law. A stage with backlash may respond differently around reversal, even when two command records are close. A safely implemented laboratory algorithm can stop updating after the measured repeats reach their uncertainty floor and can retain a copy of the previous feasible input. These choices follow from the measured limitations of the particular motion and do not change the algebraic result for the unconstrained model.

Identify a lifted column from two physical trials

Suppose the stage dynamics are imperfectly known. Start with one baseline run and save its entire output record. On a second run change the third input sample by a small positive amount δ, leaving the other inputs fixed. If the reset and disturbance repeat, subtracting the baseline record gives δ times the third column of G. The first response increment appears at the following reading. For the assumed pole $a = 0.75$, a unit perturbation gives the sequence $0.25, 0.1875, 0.140625$ at the first three affected readings. Their successive ratio estimates the pole without requiring the disturbance to be known. The absolute size of the first increment identifies the input gain. A departure from a constant ratio is evidence to inspect the one-pole model or the measurement repeatability.

When measurement resolution is poor, a perturbation can be too small to distinguish from baseline variability. Try perturbations of the same size and opposite signs on otherwise identical passes. Divide the difference in recorded outputs by 2δ. A common offset cancels and an even quadratic dependence on input amplitude cancels as well. A large difference between the columns estimated at two perturbation sizes signals curvature or saturation. Repeat the calibration at an earlier and a later command index. Their overlapping samples should agree after an appropriate time shift if the closed-loop stage is time invariant during the move. This experiment tests the actual matrix used in the update, including timing and reset assumptions.

A final notebook cell forms the error-propagation matrix explicitly and asks OneSimulation for its symmetric eigenvalues when available locally. Their largest absolute value must agree with the worst

error gain computed from singular values. The two computations expose a misplaced transpose or an incorrectly scaled learning step.

Exercises on what can be learned

1. With a memoryless sampled plant $G = gI$, what step removes every deterministic output error in one update?

▼ **Work through the solution**

The multiplier in each direction is $1 - \gamma g^2$. For $g \neq 0$, the choice $\gamma = 1/g^2$ sets it to zero. If g is overestimated, this choice takes too small a step and more trials are needed. If its magnitude is sufficiently underestimated, the learning gain can exceed the contraction bound.

---

2. Suppose the control engineer includes the initial reading y_0 and drops the final one. Which output sample can no feedforward input in that move change?

▼ **Work through the solution**

The initial reading precedes every available command. Its row in the lifted matrix is zero; the corresponding error direction has singular value zero. Choosing a larger gain leaves that component untouched. Align the output record with the first response sample or state explicitly that the initial mismatch belongs to the reset procedure.

---

3. Why can reducing the measured error in one pass make the next pass look worse when an external disturbance is freshly drawn?

▼ **Work through the solution**

The update uses the previous measured error, which includes a random component. Its correction can point in a direction unrelated to the newly drawn disturbance. The term $\mathbf{d}_j - \mathbf{d}_{j+1}$ survives the subtraction and the deterministic norm bound fails. Repeat the unmodified input several times to estimate a repeatable mean and the variability around it before changing the learning rate.

Compare the lifted calculation with experimental learning control

Bristow, Tharayil and Alleyne's survey of iterative learning control relates time-domain learning laws to frequency-domain and robustness considerations. Identify the role of a stable within-trial feedback loop before applying a between-trial update to an actual positioning system.

Problem sheet 7 · A speed regulator with an unknown load and a measured delay

A motor must reject sudden torque changes and track a smooth speed request over a delayed feedback connection. Select nominal integral gains from a time-domain requirement, calculate the delay margin for the actual PI loop and separate reference feedforward from feedback stability.

1 · Recover from a load with a bounded speed excursion

The motor obeys $J\dot{y} + by = u - d$ with $J = 0.20$ kg m² and $b = 0.50$ N m s/rad. Use $u = K_p(r - y) + K_i z,\ \dot{z} = r - y$. A 2 N m step load should cause a nominal peak error below 1.0 rad/s. Place both undelayed closed-loop poles at −4 s⁻¹. Find the gains and peak.

▼ **Check the pole choice against the performance requirement**

The characteristic polynomial is $Js^2 + (b + K_p)s + K_i = J(s + 4)^2$, so $K_p = 1.10,\ K_i = 3.20$ in SI units. At constant reference the step load gives $y = -10te^{-4t}$ in deviation coordinates. Its largest magnitude is $10/(4e) = 0.920$ rad/s at 0.25 s, meeting the undelayed specification. The controller must retain 2 N m extra torque after recovery. Saturation can invalidate this nominal transient.

Recover the peak excursion and compare proportional-only regulation

2 · Calculate the margin of this actual loop

The entire feedback command is delayed by τ. Before adding delay, the loop transfer function is $L_0(s) = (K_p s + K_i)/[s(Js + b)]$. Obtain its positive gain-crossover frequency by equating squared magnitudes. Calculate the first critical delay and the largest delay retaining at least 45° phase margin.

▼ **Use the PI zero, integrator and motor pole together**

Write $z = \omega^2$. Unit gain requires

$$J^2 z^2 + (b^2 - K_p^2)z - K_i^2 = 0.$$

With the chosen values the positive root is $z = 32$, giving $\omega_c = 5.657$ rad/s. Undelayed phase is $\arctan(K_p\omega_c/K_i) - \pi/2 - \arctan(J\omega_c/b)$, giving margin 86.63°. The first critical delay is therefore

$$1.512/5.657 = 0.267\,\text{s}.$$

A 45° reserve permits about 0.128 s. The separate delay experiment uses a first-order proportional loop. Check there that pure delay changes phase by $-\omega\tau$, then apply that relation here at the PI loop's own crossover.

Test the linear relation between phase loss and delay

3 · Predict feedforward torque during a smooth transition

Command 0 to 4 rad/s in 1 s with $r = 4(3s^2 - 2s^3)$, where $s = t/(1\ \mathrm{s})$. Add $u_{\mathrm{ff}} = \widehat{J}\dot{r} + \hat{b}r$. Determine nominal feedforward torque at midpoint and hold. What mismatch appears at midpoint if the estimate is $\widehat{J} = 0.15$ while drag is correct?

▼ **Calculate required and modeled input at the same time**

Midpoint speed is 2 rad/s and acceleration is 6 rad/s². The nominal torque is $0.20(6) + 0.50(2) = 2.20$ N m. During the final hold it is 2.00 N m. With the low inertia estimate the midpoint command is 1.90 N m, leaving a 0.30 N m shortfall for the feedback loop. The pure reference path changes tracking forcing while leaving the nominal feedback characteristic equation unchanged when signals sum linearly and actuators stay unsaturated.

Set the estimated inertia to 0.15 and 0.20 kg m² during the same move

4 · Interpret a measured commissioning delay

Timing measurements give a repeatable delay of 0.100 s. Calculate the remaining nominal phase margin and the spare latency compatible with the 45° reserve. Can the undelayed 0.920 rad/s load peak now be reported as the delayed result?

▼ **Separate the frequency and transient checks**

The delay costs $5.657(0.100) = 0.566$ rad or 32.4°, leaving about 54.2°. The spare latency is roughly 0.028 s. The delayed system has a transcendental characteristic equation, so the repeated real poles and analytic load transient of the zero-delay model no longer describe its exact step. A disturbance step experiment or an integration retaining the delay is needed to check the peak requirement at this operating point.

Problem sheet 8 · A saturated PI loop remembers an impossible set point

A first-order laboratory speed loop has a bounded actuator. A high command exceeds its available steady speed and is later replaced by a reachable command. Predict how long the output remains stuck near its limit, derive a controller that tracks the actual actuation and use a reproducible time sweep to distinguish actuator saturation from slow mechanical response.

Start with the linear closed-loop model

Normalise steady actuator gain and plant time constant to one, so $\dot{y} = -y + u$. Take $e = r - y$, proposed command $v = 2e + I$, ordinary integral update $\dot{I} = e$ and actual actuation $u = \min(1, \max(0, v))$. Begin at $y = I = 0$. Derive the characteristic polynomial without saturation and explain why its stable poles cannot answer the high-command experiment.

▼ **Close the loop within the valid linear region**

With $u = v$, Laplace transformation under zero initial conditions gives a PI controller $2 + 1/s$ in series with plant $1/(s+1)$. The reference-to-output transfer function is $(2s+1)/(s^2+3s+1)$. Its poles are approximately minus 0.382 and minus 2.618 per second, both stable. The integral action makes the steady error zero for a constant attainable reference. At $r = 2$, however, a steady plant output of two requires $u = 2$, which the actuator cannot deliver. The clipping creates a different dynamical regime immediately because the proportional term alone requests four units.

No linear pole of this original model describes the growth of the integral memory during the unreachable hold. Saturation also differs from a time delay: a delayed but sufficiently powerful actuator could eventually attain two, whereas the bounded plant cannot. A reference step within the range is a useful control experiment, although a rapid step may still transiently saturate due to the proportional kick.

Compute the integral state before the reference drops

Hold $r = 2$ for six seconds. Solve for plant output and integral state while the actuator is saturated. At $t = 6$ replace the reference by 0.4. Estimate how long the plain PI controller keeps the actuator at its upper limit after this switch.

▼ **Integrate the actual plant input rather than the requested command**

During the high reference $u = 1$, hence $y = 1 - e^{-t}$. Error is $1 + e^{-t}$, so $I = t + 1 - e^{-t}$. At six seconds $y = 0.9975$ and $I = 6.9975$. Immediately after the drop, $v = 2(0.4 - 0.9975) + 6.9975 \simeq 5.80$, so the actuator still clips at one even though the error is now negative. The plant remains near one while the integrator slowly unwinds.

Let s denote time after switching during this second saturated interval. Then $y = 1 - e^{-6-s}$, and $I = I(6) - 0.6s + e^{-6}(1 - e^{-s})$. Substituting into the command gives

$$v = 5.8 - 0.6s + e^{-6-s}.$$

Setting $v = 1$ gives approximately 8.00 further seconds at the upper limit. A first-order plant with a one-second time constant can therefore appear unresponsive much longer than its own transient. Inspecting the actual and requested commands makes the cause identifiable.

Derive a back-calculation update and test its consequences

Replace the integrator by $\dot{I} = e + (u - v)/T_t$. Explain the sign of the correction and derive its approximate long-hold value for $r = 2$, $u = y = 1$ and tracking time $T_t = 0.5$ s. Does the modified controller change the small-signal dynamics when no limit is active?

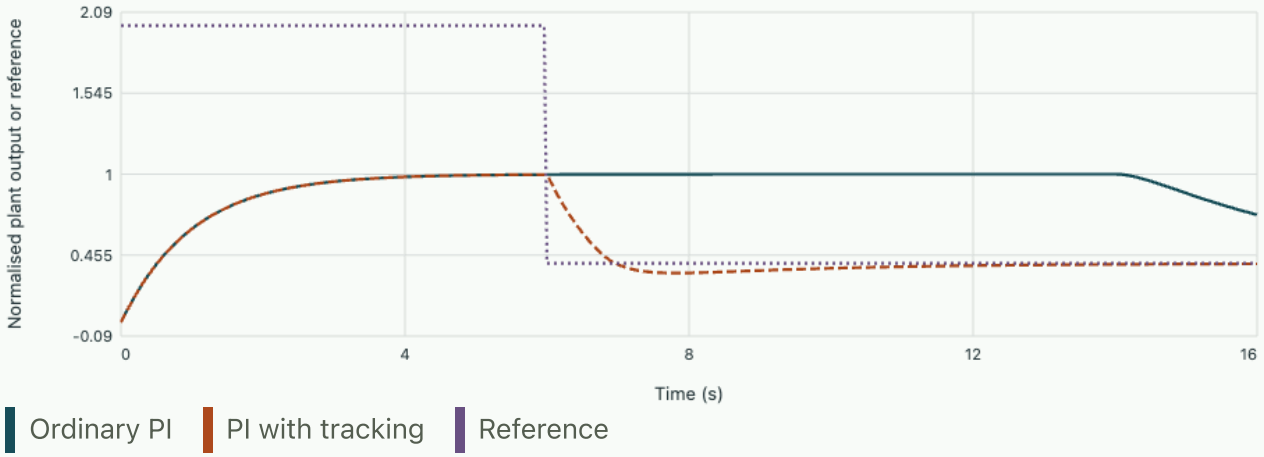
▼ **Track the attainable input and preserve the unforced integral law**

At upper saturation the requested command exceeds the actual one and $u - v$ is negative. The correction cancels continued integral growth rather than adding to it. At the long-hold steady state $0 = 1 + (1 - v)/T_t$, giving $v = 1 + T_t = 1.5$ and $I = v - 2e = -0.5$. When reference falls to 0.4 the raw command immediately becomes approximately minus 1.7. The lower actuator limit then becomes active; the plant can finally decay. On leaving both limits $u = v$ and the modification disappears exactly from the linear controller.

Open this experiment online

Try the calculation

High reference duration (s)	6
Tracking time (s)	0.5
Actuator maximum	1



After the reference drops, time still at the upper limit while output exceeds 0.6 is 8.00 s for ordinary PI and 0.00 s with tracking. The plant time constant is 1 s.

Compare the two time traces at equal plant, gains and limits. At the default hold the unmodified controller spends about eight additional seconds pinned at the upper limit while its output exceeds 0.6. The tracking controller spends essentially none there. Reducing the tracking time does not give an unlimited physical speedup: plant dynamics and the lower input bound remain. An extremely small tracking time can also make a sampled implementation numerically stiff if its discretisation does not resolve the tracking correction.

Separate time-step error, sensor clipping and true actuator saturation

The graph uses a held actuator on short time intervals, an exact first-order plant update over each interval and an explicit integral-state update. Describe how to verify a digital implementation with a longer sample interval and how an experiment could distinguish an output sensor limit from genuine actuator saturation.

▼ **Compare independent diagnostics before changing gains**

Halve the integration step at fixed switch time and compare the times at which the requested command first returns within its limits. A good trace should reproduce the analytic eight-second release time for ordinary PI. In the tracking controller a bounded integrating step must resolve both the one-second plant time and T_t. If actuator dynamics, rate limiting or a downstream safety cap changes actual actuation, feed back the measured or appropriately modelled applied value. Tracking only the software clamp in front of an additional physical limit can leave substantial windup.

Log reference, raw command, applied actuator signal, integral state and an independent output measurement. An output sensor clipped at one can conceal a real plant response above one, while an actually bounded actuator will show the applied input pinned even as its request grows. Reverse the operating sequence and use several hold durations. An additional release delay growing with unreachable hold time is characteristic of integral memory; an unchanged time constant after every switch suggests a different plant or measurement limitation.

Åström and Murray · Feedback Systems, integral windup and tracking in PID control

Notebook practice · Learn a repeatable positioning move from its previous error

System dynamics, feedback and estimation

Start from a physical plant and its measured signals. Derive a feedback or estimation model and test which dynamical behaviour can be controlled or inferred.

1. ## Estimating velocity from position with a model-based observer

   An optical encoder returns position and the actuator drive supplies a commanded acceleration. Reconstruct an unmeasured velocity without differentiating successive noisy samples. Derive the observer, explain what a pair of chosen poles guarantees, and predict the tradeoff visible when the encoder carries periodic interference.

2. ## Relative gains, loop pairing and nearly singular process measurements

   Two heaters act on two temperature sensors. With the second heater fixed, a test of the first gives a well-determined positive response. Closing a loop around the other temperature can change both the size and sign of that apparent gain. Derive the relative gain array from the experiment one actually performs.

3. ## Predicting a delayed process and testing a Smith controller

   A heater is upstream of a temperature sensor. Its internal thermal response begins immediately while a material parcel reaches the sensor several seconds later. Build a feedback signal from two model outputs, derive which transfer function loses its delay, and then deliberately introduce a delay mismatch to see why a plausible prediction may fail.

4. ## Sensitivity, bandwidth and the integral price of disturbance rejection

   A regulator attenuates a slowly varying output load yet amplifies another disturbance at a higher frequency. Separate the paths by which reference, measurement noise and physical disturbance reach the recorded variable, then derive an integral identity that explains why a stable loop can redistribute sensitivity across frequency.

5. ## Which sampling times reveal the velocity of an oscillator?

   A sensor measures displacement while a controller needs velocity as well. Use the exact free vibration between two readings to recover the missing state, then quantify how a poorly chosen sample interval amplifies measurement noise.

6. ## When does a stable operating point disappear?

   A shaft can hold a static torque against gravity up to a finite limit. Trace the equilibrium branches and distinguish local stability, loss of an equilibrium and the possibility of oscillation.

7. ## Can any controller meet this bandwidth and peak-gain specification?

Delay uses up phase at the crossover even when it changes no gain. An unstable pole near a right-half-plane zero imposes a separate peak-gain limit. Derive each restriction before tuning a controller.

8. Designing current, speed and position loops for a drive

An encoder measures shaft motion, while the power stage acts on voltage. Decide which disturbances each measurement can reject and how fast each loop can move before the unseen inner dynamics erode the position margin.

System dynamics, feedback and estimation · 1. Estimating velocity from position with a model-based observer

An optical encoder returns position and the actuator drive supplies a commanded acceleration. Reconstruct an unmeasured velocity without differentiating successive noisy samples. Derive the observer, explain what a pair of chosen poles guarantees, and predict the tradeoff visible when the encoder carries periodic interference.

Find out which initial conditions are distinguishable

Model a translating stage over its usable travel by $\dot{x}_1 = x_2,\ \dot{x}_2 = u$, with measured output $y = x_1 + n$. Here the input is an acceleration in the chosen position and time units. With zero noise and known input, two different initial positions give different initial readings. Two different initial velocities with the same position yield traces whose initial slopes differ. Formally the observability matrix stacks $C = (1, 0)$ above $CA = (0, 1)$, giving rank two. If only velocity were measured, a constant offset of initial position would remain invisible, whatever the command. Observability belongs to the chosen model and output together.

Run a copy of the plant and drive it with both the known command and a correction proportional to innovation $y - \hat{x}_1$:

$$\dot{\hat{x}}_1 = \hat{x}_2 + l_1(y - \hat{x}_1),$$

$$\dot{\hat{x}}_2 = u + l_2(y - \hat{x}_1).$$

Define error as true minus estimated state. For a perfect model and noise-free measurement the error matrix has rows $(-l_1, 1)$ and $(-l_2, 0)$. Expanding the determinant of $sI - (A - LC)$ gives $s^2 + l_1 s + l_2$. Matching the standard oscillator denominator sets $l_1 = 2\zeta\omega_o$ and $l_2 = \omega_o^2$. The units are inverse seconds and inverse seconds squared. A constant position innovation can therefore request corrections to both estimated velocity and acceleration. Stable poles guarantee decay of initial error for the ideal model. They do not specify how large the measurement error becomes when injected continually.

Follow one sensor tone all the way to the velocity estimate

Set the true state and command to zero and solve the two linear equations in the Laplace domain with zero initial condition. Eliminating estimated position gives the transfer of position noise to estimated velocity:

$$\frac{\hat{V}(s)}{N(s)} = \frac{l_2 s}{s^2 + l_1 s + l_2}.$$

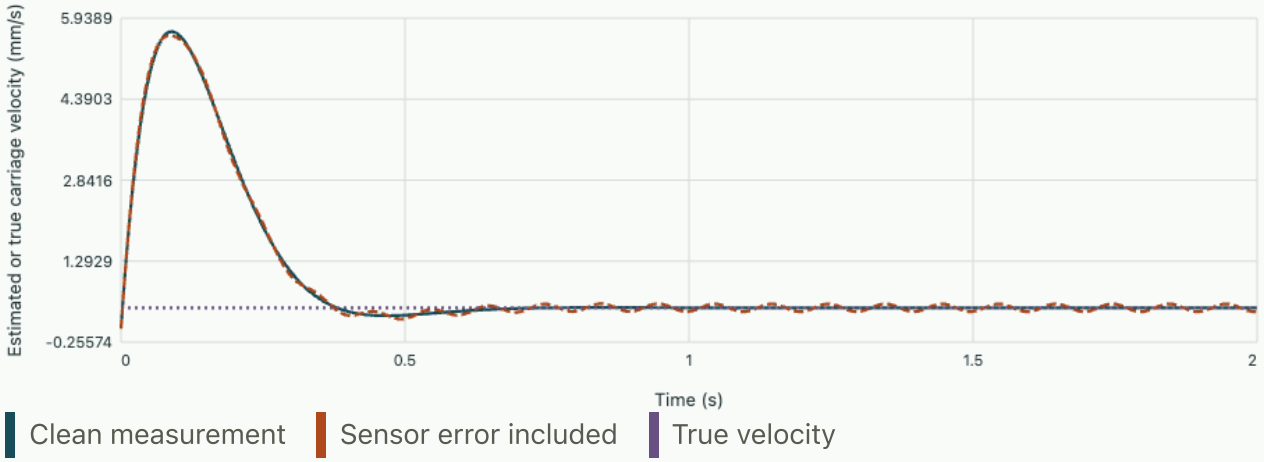
For slow variation it behaves like a differentiator, so a constant position offset ultimately produces zero spurious velocity. At high frequency the gain falls as l_2/ω, whereas differentiating raw data multiplies its amplitude by frequency. Noise near the observer bandwidth can nevertheless be strongly visible. With observer frequency 2 Hz, damping ratio 0.75 and a 10 Hz sensor oscillation of amplitude 0.03 mm, the

predicted velocity ripple is about 0.075 mm/s. The stage moves at 0.40 mm/s in the example. Compare the late-time difference between clean and noisy traces with the predicted peak ripple. The initial position estimate is deliberately displaced to reveal transient convergence.

Open the live experiment

Try the calculation

Observer natural frequency (Hz)	2
Sinusoidal sensor error (mm)	0.03
Sensor error frequency (Hz)	10



The predicted steady velocity ripple amplitude from the sensor error is 0.075 mm/s. Initial true position is 1 mm; both estimated initial states are zero.

Background · What if successive position errors are correlated?

Estimate velocity as the difference between consecutive positions divided by the sample interval h. Suppose each position error has variance σ^2 and the correlation between successive errors is ρ. The variance of the difference contains the covariance twice with a negative sign.

$$\operatorname{Var}\left(\frac{e_k - e_{k-1}}{h}\right) = \frac{2\sigma^2(1-\rho)}{h^2}.$$

A constant sensor bias cancels exactly from this difference, although the reconstructed position can remain wrong. Independent 0.10 mm position errors sampled 0.01 s apart give a 14.1 mm/s velocity standard deviation. With the same marginal variance and correlation 0.9 it falls to 4.47 mm/s. A sensor data sheet reporting only RMS position noise leaves this prediction underdetermined.

Use the innovation as a diagnostic signal

If the true acceleration exceeds the commanded value by constant d, the steady error equations yield

$$e_1 = d/l_2,$$

$$e_2 = l_1 d/l_2.$$

Raising the observer frequency reduces these model-error terms while passing more encoder noise at a fixed sufficiently high frequency. Adding a slowly evolving bias state is another modelling choice; its observability and process-noise assumptions need to be checked again. A near-zero innovation in a high-gain observer alone is weak evidence of correct velocity, since the algorithm was designed to force estimated position toward the sensor.

On the bench, estimate the encoder noise spectrum with the stage held stationary and the illumination and sampling electronics operating normally. During motion, compare estimated speed with an independently timed travel between two reference marks. An encoder fitted only by its own residual offers no independent absolute scale. If a motion command changes numerically from metres to millimetres, transform acceleration, position and initial estimates together. Pole locations have units of inverse time and do not depend on this uniform length conversion.

Check whether an apparent velocity ripple aliases from above the sampling band

Suppose the encoder is sampled at 100 Hz and its electronics pick up a 90 Hz sinusoid. At those sampling instants the measured sequence is indistinguishable from a ten-hertz sinusoid with suitable phase. A downstream digital observer cannot determine which analogue frequency generated the samples. Repeating with a different acquisition clock moves an alias while leaving a genuine ten-hertz process feature at its physical frequency. An analogue anti-alias filter can attenuate the high-frequency signal before sampling.

Compare this experiment with the frequency knob in the displayed model. The browser directly synthesises a known tone at the integration rate and therefore tests the observer frequency response in isolation. A measured disagreement on a real sampled encoder can reveal an upstream acquisition effect even when the observer discretisation is accurate.

Tutorial · Identify a constant offset and an acceleration mismatch

An encoder reads 0.20 mm high at rest. Separately, an unmodelled 0.20 mm/s squared acceleration acts on an otherwise perfect stage. Find steady errors for the two cases at the default observer setting.

▼ Solve the steady error equations for each input

A sensor offset eventually places estimated position 0.20 mm above truth while the estimated velocity returns to zero. For the acceleration mismatch, $l_2 = 157.91$ and $l_1 = 18.85$ in the appropriate second-

based units. The true position exceeds the estimate by 0.00127 mm and true velocity exceeds its estimate by 0.0239 mm/s. The persistent residual in the second case is tiny despite a clear velocity bias. Repeat the test while changing bandwidth to distinguish a bias in the sensor from a mismatch in acceleration.

Tutorial · Compare with numerical differentiation

With 1 ms sample separation, independent position errors have standard deviation 0.01 mm per reading. What standard deviation does a backward difference produce for velocity? Explain how this differs from the deterministic 10 Hz test.

▼ **Propagate the variance of the difference**

Subtracting independent zero-mean errors adds their variances. Divide the resulting standard deviation by the time step to obtain

$$\sqrt{2}(0.01)/0.001 = 14.14\,\mathrm{mm/s}.$$

A 10 Hz interference tone has correlated neighbouring samples and a fixed phase evolution. Its output is predicted from sinusoidal frequency response, not from independence. Averaging a longer interval reduces random position-difference noise but measures a longer-time mean speed and responds more slowly to acceleration.

Tutorial · Check discrete implementation before increasing gain

A software observer updates by explicit Euler every 10 ms. If the continuous error poles are both minus 300 per second, what happens to an initial error? How could the implementation preserve the intended decay?

▼ **Map a continuous pole through the actual update**

The scalar Euler factor is

$$1 + h\lambda = 1 - 3 = -2;$$

its magnitude exceeds one, so errors grow while changing sign. Exact discretisation would map the pole to $e^{-3} = 0.0498$. A suitably faster sample interval, a correctly discretised observer, or a discrete observer designed directly for the sampled measurement can restore stability. These choices still depend on the actual measurement rate. Interpolating old encoder samples at a faster software clock does not create independent new position information.

Caltech · Åström and Murray, Output Feedback, observability and state estimation

System dynamics, feedback and estimation · 2. Relative gains, loop pairing and nearly singular process measurements

Two heaters act on two temperature sensors. With the second heater fixed, a test of the first gives a well-determined positive response. Closing a loop around the other temperature can change both the size and sign of that apparent gain. Derive the relative gain array from the experiment one actually performs.

Compare two physically different partial derivatives

Linearise a two-input, two-output process about an equilibrium, using deviations from the original input and output settings. Write its steady gain matrix as K, with entry $k_{ij} = \partial y_i / \partial u_j$ with the other input held fixed. If instead the second output is held fixed by feedback, the other input must move too. Eliminate it from the second row:

$$0 = k_{21}\, du_1 + k_{22}\, du_2,$$

$$\left.\frac{dy_1}{du_1}\right|_{y_2} = k_{11} - \frac{k_{12}k_{21}}{k_{22}} = \frac{\det K}{k_{22}}.$$

The relative gain for this pairing is the open-other-input gain divided by the gain with the other output held constant. Inverting the gain matrix gives the general expression $\Lambda_{ij} = k_{ij}(K^{-1})_{ji}$, an element-by-element product with the transposed inverse. It requires a nonsingular square matrix; the alternative ratio interpretation requires that the selected constrained experiment also be well defined. Each row and column sums to one. A quick check uses the diagonal of $KK^{-1} = I$. A negative element signals a gain reversal between the two experiments. A very large element indicates considerable cancellation in the constrained response.

For matrices $K = \left(\begin{smallmatrix} 2 & 1 \\ 1 & 2 \end{smallmatrix}\right)$, the diagonal relative gains are 4/3 and the off-diagonal entries are minus one third. A unit increase of the first input holding the second output fixed requires reducing the second input by one half. The first output then rises by 1.5, compared with 2 when the other input is held fixed. Off-diagonal pairing would confront a reversal when the other loop closes. A static array can guide selection, but faster dynamics, delays, actuator bounds and output priorities still matter.

Show which numerical conclusions depend on chosen units

Let individual output units be rescaled by a diagonal matrix D_y and input units by D_u. The new gain matrix is $D_y K D_u^{-1}$ and its inverse transforms in the opposite order. The scale factors cancel separately in each product defining a relative gain. Reporting a temperature deviation in millikelvin instead of kelvin does not change the array. Singular values, by contrast, do depend on units; meaningful comparisons require physical normalisation such as allowed input ranges and tolerated output deviations.

For $K = \left(\begin{smallmatrix} 1 & 0.9 \\ 1.1 & 1 \end{smallmatrix}\right)$ the determinant is 0.01. The diagonal entries of the array reach 100 and the off-diagonal entries minus 99. Nominally holding the second output constant requires $du_2 = -1.1 du_1$, leaving only 0.01 output units per unit of first input. A requested unit output therefore asks for input changes of 100 and minus 110 in this local model. Saturation or departure from the linear neighbourhood

would usually intervene first. A nearly singular response indicates that the two actuators push in almost the same output direction.

A test programme distinguishes cancellation from sensor drift

Perturb each input on both sides of the reference point with the other held fixed, allow slow transients to settle, and fit slopes including an offset for drift. Repeat with both inputs coordinated to hold one output near its initial value. The small difference of large responses may have uncertainty as large as its point estimate. Record confidence intervals for the matrix elements and propagate them through the determinant before prescribing a controller pairing. Independent tests on a second operating point reveal whether the conclusion survives gain changes. A temperature sensor substituted into both output channels can accidentally make an otherwise invertible plant look rank deficient.

Compare the time-domain inverse-response study with a steady relative gain. A right-half-plane zero can constrain a pairing that looks attractive at zero frequency.

Make the nearly invisible constrained response visible

The first curve holds the second input constant; the second holds the second output constant. Both pass through the reference point because the model uses deviations about a common equilibrium. With the nominal gains the latter slope is only one hundredth of the former. Change the lower-right coefficient from one to 0.99 and then to 0.98. The constrained slope passes through zero and reverses sign. At the singular value a finite relative gain is undefined, while the physical experiment remains perfectly interpretable: the coordinated input direction makes no output change.

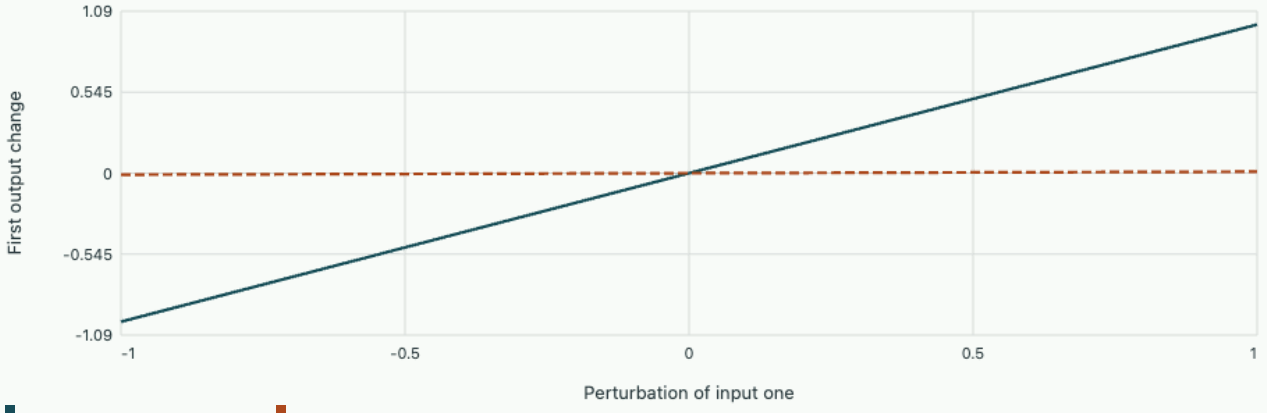
Open the live experiment

Try the calculation

Cross gain k_{12} **0.9**

Cross gain k_{21} **1.1**

Direct gain k_{22} **1**



The constrained gain is 0.0100. The diagonal relative gain is 100.00. For a unit first input the second actuator must move -1.100 units to keep its output fixed.

A change of less than the uncertainty of a single gain measurement can thus reverse the preferred feedback direction. If the second actuator reaches a limit during the constrained test, the measured slope shifts back toward the first curve because the supposedly fixed second output can no longer be maintained. Record both actuator trajectories and the residual at the constrained output before claiming to have measured a relative gain.

Tutorial · Infer the hidden actuator motion

A two-zone stage has gain matrix with rows (3, 1) and (2, 2), in consistent scaled units. Increase the first actuator by one unit while regulating the second output exactly. Find the first output and relative gain.

▼ Write the constrained row first

The second row requires $du_2 = -1$. The first output changes by $3 - 1 = 2$. With the second input fixed it would have changed by three, so the first relative gain is 1.5. The complementary entries are minus one half, since row and column sums must equal one. For a slowly regulating second loop the early transient can resemble the fixed-input experiment while the settled change follows the constrained one. Compare time scales before using the settled gain in a fast-loop tuning rule.

Tutorial · Quantify a near-singular inference

Suppose only the lower-right coefficient of the nearly singular example is uncertain by plus or minus 0.02. Can the sign of the diagonal relative gain be inferred from the nominal values?

▼ **Bound the determinant before taking the ratio**

At the endpoints the determinant becomes minus 0.01 and plus 0.03. At 0.99 for that coefficient the matrix is exactly singular, so the interval crosses a pole of the calculation. The sign is not identified. A covariance estimate or repeated constrained tests may narrow the uncertainty. Taking an arithmetic average of very large positive and negative sampled gains around a singularity is especially misleading; the two sides correspond to qualitatively different controller directions.

Tutorial · Test a proposed decoupling command

For the first symmetric matrix, request changes (1, minus 1) in the two temperatures. Compare the required inputs with those for (1, 1). Why is the independent-channel approximation misleading?

▼ **Invert the matrix along its two physical directions**

The equal direction has gain three and needs input changes (one third, one third). The differential direction has gain one and needs (1, minus 1). Treating both diagonal gains as two would predict changes of magnitude one half for either case. A common-mode validation therefore overstates ability to control a gradient. Test both common and differential directions at suitable amplitudes before reporting usable output ranges, especially with bounded actuators.

ETH Zürich · Control Systems II, multivariable systems and control design

ETH Zürich · MIMO directionality, singular values and relative gain

System dynamics, feedback and estimation · 3. Predicting a delayed process and testing a Smith controller

A heater is upstream of a temperature sensor. Its internal thermal response begins immediately while a material parcel reaches the sensor several seconds later. Build a feedback signal from two model outputs, derive which transfer function loses its delay, and then deliberately introduce a delay mismatch to see why a plausible prediction may fail.

Keep the measured and predicted signals distinct

Represent a stable process as a rational delay-free part $G_0(s)$ followed by transport delay e^{-Ls}. Let the controller see a prediction $y_p = y + \hat{G}_0 u - \hat{G}_0 e^{-\hat{L}s} u$. The two model terms are the currently predicted response and a stored, delayed prediction. On a perfect model, the delayed part cancels the measured response to the command, leaving an undelayed estimate in the feedback path. The measured output remains at the physical sensor downstream. Nothing in the arrangement advances an unknown disturbance in time.

With controller $u = C(r - y_p)$ and actual command-to-output plant P, direct substitution gives

$$\frac{Y}{R} = \frac{PC}{1 + C[\hat{G}_0 + P - \hat{G}_0 e^{-\hat{L}s}]},$$

$$\left.\frac{Y}{R}\right|_{\text{exact}} = e^{-Ls}\frac{G_0 C}{1 + G_0 C}.$$

The characteristic denominator is delay-free in the exact model. The reference-to-measurement signal still includes the full delay. Internal stability requires sensible models and implementation. In particular, cancelling an unstable physical mode in a transfer expression can hide a divergent internal signal. This elementary design assumes a stable plant and meaningful initial conditions for the stored model trajectory.

Calculate the entire nominal step before adjusting a tuning knob

Take $G_0 = 1/(4s + 1)$, proportional gain two and a delay of three seconds. The measured unit step response is zero until three seconds. Afterwards it is $(2/3)[1 - \exp(-3(t - 3)/4)]$. Its final offset is one third because a proportional controller is being used. It attains 63.2 percent of its own final change at 4.333 seconds after the reference step and about 95 percent at seven seconds. A plotted prediction rising before three seconds is perfectly consistent with a real sensor trace that remains flat. Labelling a model trajectory as a sensor measurement would conceal a missing or jammed sensor.

Now suppose the true transport delay is four seconds while the model still stores three. A sinusoid of angular frequency ω gives a mismatch between the two unit-magnitude delay factors equal to $2|\sin[\omega(L - \hat{L})/2]|$. At one radian per second and one second timing error this is 0.959, even though the steady mismatch vanishes. Multiply by the plant gain at that frequency before interpreting its effect. Time stamps, sample-and-hold, filtering and passage distance each contribute phase. Fitting a single dead time over a narrow frequency band cannot tell them apart.

Use the ordinary delayed-loop phase-margin study to contrast physical phase lag with the model-corrected feedback path.

Track an unknown load applied at the measured output

Let $y = Pu + d$, where a load disturbance is added after the physical process. In the exact predictor the feedback observation becomes $G_0 u + d$. With reference zero it follows that $u = -Cd/(1 + CG_0)$, so the actual output responds according to $1 - PC/(1 + CG_0)$. An abrupt output disturbance appears immediately at the sensor, but the compensating physical response arrives only after transport. Input disturbances occurring upstream have a different transfer path. Sketching the injection point avoids confusing these two bench tests.

Measure the transport delay by recording a known actuator perturbation and the two temperatures on a common clock. A feature appearing at an electronic sensor downstream may include sensor lag as well as travel. Move the sensor a known distance without changing flow and repeat. A shift proportional to added distance is evidence for transport; a constant extra lag suggests sensor or actuator dynamics. Repeat at another flow rate and revise the model if the material speed has changed.

Distinguish the two delayed model buffers in a simulation

The experiment stores the output history of the first-order lag. One buffer index represents the actual transit time and another supplies the delayed copy subtracted by the predictor. When the two delay sliders match, the difference between the physical measured curve and its ideal reference is limited to the small time-discretisation error. Change only the actual delay to test mismatch. The correction begins when one delayed signal has arrived and the other has not. Watch the predicted feedback signal separately from the physical output across that interval.

Open the live experiment

Try the calculation

Actual transport delay (s) 3

Model transport delay (s) 3

Proportional gain 2



The largest measured deviation from an exact-delay benchmark during 30 s is 0.001. The driver updates every 0.01 s. No physical response is possible before 3.0 s.

Raise proportional gain at the mismatched delay. The physical transit time stays fixed but the model-based correction becomes more sensitive to cancellation errors. A measurement which shows oscillatory excursions in this demonstration can still settle to the nominal final offset, since a pure delay has unit gain at zero frequency. Agreement at the final value cannot by itself verify the timing model.

Tutorial · Distinguish two time constants on a response plot

Two labelled traces reach 0.42 output units at 1.33 s and 4.33 s respectively after a unit step. A technician calls the second plant three times slower. Explain both traces in the nominal example and design a second check.

▼ **Align the response origins**

The 63.2 percent level of the 0.667 final value is about 0.421. An undelayed prediction reaches it at the closed-loop time constant of 1.33 s. The measured trace reaches it one transport interval later at 4.33 s. Shift the measured trace three seconds earlier and compare its entire shape, amplitude and late drift. Agreement at one level alone would also occur for many models combining a different lag and delay.

Tutorial · Quantify a frequency at which a delay error matters

For an error of 0.2 seconds, at what first positive angular frequency does the difference of unit delay factors reach magnitude one? Is that automatically a stability boundary?

▼ **Solve the phasor geometry and inspect the remaining loop**

Set $2\sin(0.1\omega) = 1$, giving $\omega = \pi/(0.6) = 5.24$ rad/s. The physical plant magnitude there is about 0.048 for the stated first-order lag. Controller gain and phase, additional modes and the nominal denominator determine the closed-loop consequence. A unit difference of raw delay phasors is an interpretable timing-error measure and does not by itself place a closed-loop pole on the imaginary axis.

Tutorial · Explain a failed transport assumption

A predictor tuned at constant flow develops oscillatory residuals when flow periodically changes. Name observations that could distinguish model-delay variation from an oscillating load.

▼ **Use independent transport and command records**

Compute the passage delay from an independent flow measurement and known pipe volume, accounting for the actual parcel history when speed varies appreciably during transit. Compare residual phase against the command and repeat with a small deliberate input perturbation at a different frequency. A response phase tracking flow supports a varying delay. An output oscillation continuing at fixed flow with no corresponding command suggests an external load or another mode. Preserve a baseline with the predictor disabled to identify plant changes independently of the feedback algorithm.

ETH Zürich · Control Systems II, dead-time compensation and the Smith predictor

System dynamics, feedback and estimation · 4. Sensitivity, bandwidth and the integral price of disturbance rejection

A regulator attenuates a slowly varying output load yet amplifies another disturbance at a higher frequency. Separate the paths by which reference, measurement noise and physical disturbance reach the recorded variable, then derive an integral identity that explains why a stable loop can redistribute sensitivity across frequency.

Solve the loop once for three independent inputs

Let the actual output be $y = Pu + d$, the measured output be $y + n$, and the command be $u = C(r - y - n)$. Here d is an additive load at the output and n is sensor error. Division by $1 + PC$ gives

$$S = \frac{1}{1 + L},$$

$$T = \frac{L}{1 + L},$$

$$L = PC,$$

$$y = Tr + Sd - Tn,$$

$$u = CS(r - d - n).$$

Thus $S + T = 1$ as a complex identity. Magnitudes generally do not add to one. Strong rejection of a slow physical load requires small sensitivity at that frequency; the measured noise can then appear almost unchanged, with opposite sign, in the actual output. A readout of the same sensor used by the controller can hide this motion. A second independent sensor can reveal it. An input load inserted before the plant would instead have transfer PS, so always identify the disturbance injection point in a reported test.

For $L = 4/[(s + 1)(s + 2)]$, both the open and closed loops are stable. The zero-frequency sensitivity is one third. At three radians per second the numerator is $-7 + 9i$ and denominator is $-3 + 9i$, giving magnitude $\sqrt{130/90} = 1.202$. A 0.10-unit output disturbance at that frequency produces a 0.120-unit actual output oscillation. A time plot of reference following alone would miss this amplification. The sensitivity peak also measures proximity of the Nyquist curve to minus one, since $1/|S| = |1 + L|$.

Derive the stable rational case from its pole and zero distances

For a closed-loop stable rational sensitivity tending to one at high frequency, write its numerator and denominator as monic products of roots. For a real negative root at minus a, a factor contributes $\frac{1}{2}\log(\omega^2 + a^2)$. The difference for two such factors integrates by differentiating under the integral in parameter a. The derivative $a/(\omega^2 + a^2)$ has integral $\pi/2$ over the positive frequency axis. Integration back gives $\pi(a - b)/2$. For complex conjugate roots the same result follows by pairing both signs of frequency and shifting the full-axis integration variable by the imaginary part. The area contribution depends on the absolute real part.

The sensitivity zeros are the open-loop poles and its poles are the closed-loop poles in a minimal description. For relative degree at least two the coefficient of the next-to-highest power in the open- and closed-loop characteristic polynomials is unchanged; their signed root sums therefore coincide. If every open-loop pole lies in the left half-plane, the absolute-real-part sums also coincide and the integral vanishes. An open-loop pole in the right half-plane reverses its absolute-real-part contribution and leaves the general result

$$\int_0^\infty \log|S(i\omega)|\,d\omega = \pi \sum_{p\in\mathrm{RHP}} \operatorname{Re} p.$$

This statement assumes internal closed-loop stability, suitable high-frequency roll-off and no imaginary-axis poles in the elementary derivation. It is an area on a linear angular-frequency axis with natural logarithm. Estimating it as signed visual area on a conventional log-frequency Bode plot weights the frequencies differently. A stable low-pass system can reduce sensitivity over a wide low-frequency interval by allowing higher sensitivity elsewhere. The identity does not fix an individual peak without bandwidth or shape constraints.

Connect the complementary sensitivity to model error

For multiplicative plant uncertainty $P_{\text{actual}} = P(1 + W\Delta)$, factor the return difference as $(1 + L)(1 + TW\Delta)$. If nominal interconnection is internally stable, the uncertainty and weighting are stable, and $\|TW\|_\infty < 1$, the small-gain theorem ensures stability for all such perturbations with $\|\Delta\|_\infty \le 1$. A failure of this sufficient check invites closer analysis of the particular physical uncertainty. A poorly known resonance above bandwidth can matter even after a nominal step trace has settled smoothly. State the frequency range on which input-output identification actually constrains the model.

Change a lead network and compare crossover with measured sensitivity near a resonance.

Locate the sensitivity peak and test vector addition

Sweep the gain numerator at fixed pole locations. At zero gain the load sees essentially no correction. Increasing gain initially improves low-frequency load rejection but can create a distinct sensitivity maximum between the pole scales and the high-frequency range. Find the frequency where the sensitivity curve crosses one. At that point the distance of the loop phasor from minus one equals one. The complementary-sensitivity magnitude there need not vanish; the sum identity concerns complex values with their phases retained.

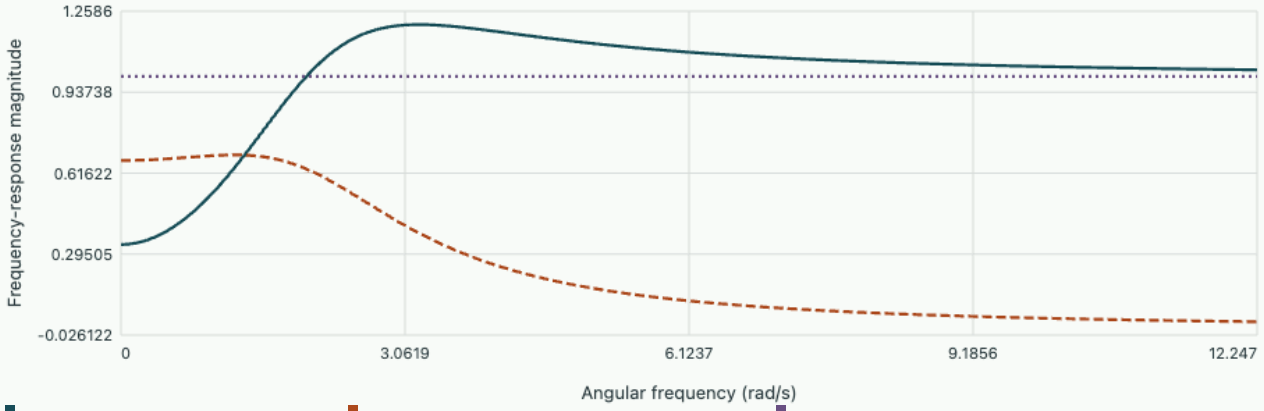
Open the live experiment

Try the calculation

Loop numerator K 4

First pole frequency (rad/s) 1

Second pole frequency (rad/s) 2



Output disturbance |S| Reference or sensor noise |T| Unit magnitude

Static sensitivity is 0.333. The sampled peak is 1.206 near 3.20 rad/s. At 3 rad/s |S| = 1.202.

To check the curves on a real apparatus, inject small reference tones to estimate complementary sensitivity and separate, calibrated output loads to estimate sensitivity. Record coherent amplitude and phase at matching frequencies with the same time origin. Sum the two complex estimates and compare against unity. Delay in the reference measurement chain alone can spoil that check, even if both measured magnitudes seem plausible. The test is most informative where both responses have appreciable amplitude.

Tutorial · Test an apparent counterexample

For $L = K/(s + a)$ with positive K, a, sensitivity is below one at all finite frequencies. Evaluate the logarithmic area and locate the failed assumption.

▼ **Integrate the log-distance difference directly**

The sensitivity equals $(s + a)/(s + a + K)$. The real-root identity gives

$$\pi[a - (a + K)]/2 = -\pi K/2.$$

This loop has relative degree one. Its numerator and denominator root sums differ by K, invalidating the high-frequency cancellation used for the zero-area result. Adding a finite fast actuator pole restores relative degree two and can move the compensating sensitivity region far beyond the frequencies first plotted. Check both the analytical degree and measurement bandwidth before diagnosing a contradiction.

Tutorial · Use two independent sensors

At a low frequency the loop has real positive gain 19. A calibrated sinusoid of amplitude 0.20 mm is electronically added to the feedback sensor with no physical load. Predict the actual motion and its indicated motion at that sensor.

▼ **Follow the noise to two separate outputs**

Sensitivity is 0.05 and complementary sensitivity 0.95. The actual stage moves with amplitude 0.19 mm opposite the injected signal. Its indicated position at the corrupted sensor has amplitude 0.01 mm in phase with the injection. An independent reference encoder sees the much larger actual motion. Increasing gain further would make the feedback readout look quieter while causing more physical motion in this particular test.

Tutorial · Bound uncertainty at a measured spectral line

At one frequency the nominal measured complementary sensitivity is 1.25 and the magnitude bound on multiplicative uncertainty is 0.60. At a second frequency the values are 1.05 and 1.10. Interpret both products.

▼ **Apply the pointwise ingredient of the small-gain check**

The products are 0.75 and 1.155. The first frequency meets the sufficient magnitude bound there. The second exceeds it, so a claim based on that global criterion is unavailable. A tighter phase-aware model or structured uncertainty can refine the conclusion. Checking two isolated frequencies cannot establish a supremum bound, since resonant features may occur between samples. Estimate frequency resolution from the narrowest expected damping bandwidth and verify both sides of each measured peak.

Åström and Murray · Feedback Systems, chapters on sensitivity functions and Bode's integral formula

System dynamics, feedback and estimation · 5. Which sampling times reveal the velocity of an oscillator?

A sensor measures displacement while a controller needs velocity as well. Use the exact free vibration between two readings to recover the missing state, then quantify how a poorly chosen sample interval amplifies measurement noise.

Observe a rotation through one coordinate

For an undamped, unforced oscillator $\ddot{x} + \omega^2 x = 0$, use the state $z = [x, \dot{x}/\omega]^T$. Both entries then have units of length. A delay Δt advances phase by $\theta = \omega \Delta t$. Integrating the equation exactly gives

$$x(t + \Delta t) = x(t)\cos\theta + [\dot{x}(t)/\omega]\sin\theta.$$

The velocity equation follows by differentiating with respect to Δt. Together they rotate the state by a matrix with first row $(\cos\theta, \sin\theta)$ and second row $(-\sin\theta, \cos\theta)$.

The position sensor has observation row $(1, 0)$. At consecutive samples its noiseless readings obey

$$\begin{bmatrix} y_0 \\ y_1 \end{bmatrix} = \underbrace{\begin{bmatrix} 1 & 0 \\ \cos\theta & \sin\theta \end{bmatrix}}_{O} \begin{bmatrix} x_0 \\ v_0/\omega \end{bmatrix}.$$

The determinant is $\sin\theta$. At integer half-period intervals the sensor measures the same direction in the state plane twice, allowing opposite velocities to produce exactly the same pair of readings. Away from those intervals,

$$v_0 = \omega(y_1 - y_0\cos\theta)/\sin\theta.$$

The determinant diagnoses whether an exact reconstruction exists. The singular values describe whether that reconstruction will tolerate small errors. Because OO^T has diagonal entries one and off-diagonal entries $\cos\theta$, its eigenvalues are $1 \pm |\cos\theta|$. Their square roots are the singular values of O in the chosen state scaling.

Propagate the uncertainty in two readings

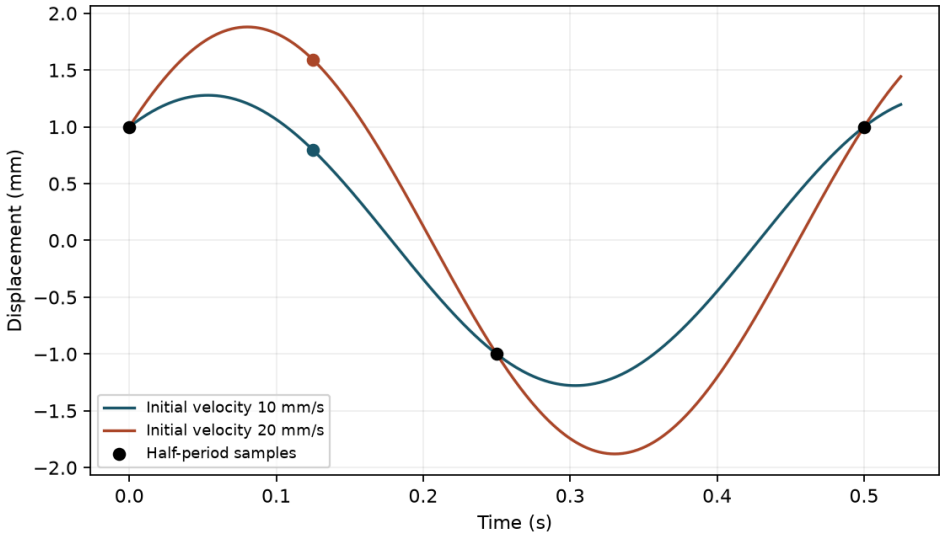
Suppose each recorded position has independent zero-mean noise of standard deviation σ_x. A linear combination $y_1 - y_0\cos\theta$ has variance $\sigma_x^2(1 + \cos^2\theta)$. Multiply by the square of the velocity gain to obtain

$$\sigma_v = \omega\sigma_x\sqrt{1 + \cos^2\theta}/|\sin\theta|.$$

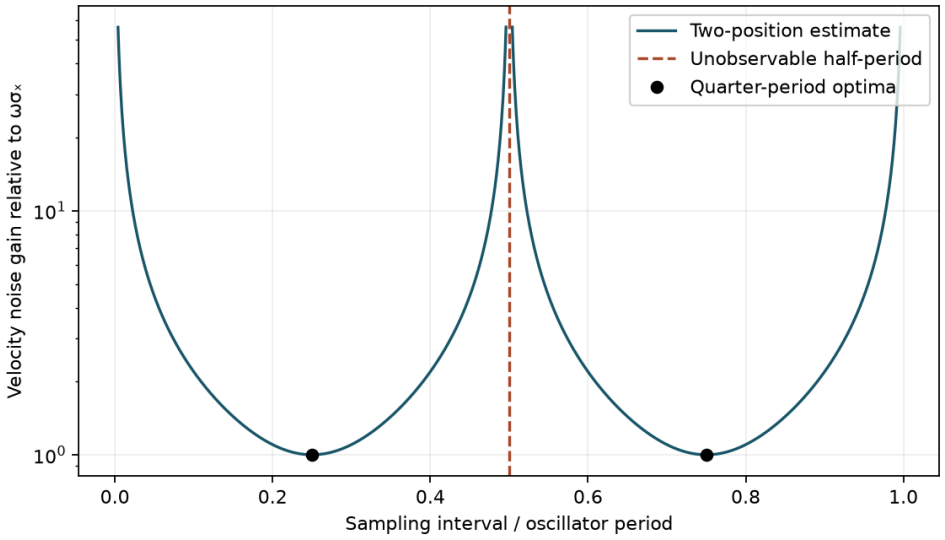
An exact knowledge of the sample times and natural frequency is assumed in this expression.

A 2 Hz oscillator has period 0.5 s and angular frequency 4π rad/s. With a 0.125 s interval the second reading directly measures v_0/ω. For 0.10 mm position noise the recovered velocity has standard

deviation 1.257 mm/s. Increasing the delay to 0.240 s moves the second observation close to the direction opposite the first. The same position uncertainty now produces 14.12 mm/s of velocity uncertainty, although the mathematical reconstruction still exists. The notebook varies the sampling interval continuously and checks the formula against noise propagation through the inverse matrix.



Compare the equal half-period readings with the selected off-grid observation. Continuous motion between sample points can carry information that the readings discard.



The smallest uncertainty occurs at the quarter-period points where the second position reading depends only on the initial velocity.

Problem 1 · Take more readings at exactly one half-period

A colleague proposes collecting 100 equally spaced positions to recover velocity at $\Delta t = T/2$. Does the larger data set recover the missing state for this exact model?

▼ Work through the solution

After k steps the observation row is $(\cos k\pi, \sin k\pi) = ((-1)^k, 0)$. Every reading measures only the initial position, up to its alternating sign. Averaging can reduce the noise in x_0, while the initial velocity

remains arbitrary. One extra measured position at a known quarter-period after the first would add row $(0, 1)$ and restore the missing independent equation.

Problem 2 · Relate observability to a finite difference

For $\omega \Delta t \ll 1$, derive the leading noise in the reconstructed velocity. Explain why decreasing the interval can increase the velocity error here.

▼ Work through the solution

Use $\sin\theta \simeq \theta = \omega \Delta t$ and $\cos\theta \simeq 1$. The estimate becomes $(y_1 - y_0)/\Delta t$ to leading order and $\sigma_v \simeq \sqrt{2}\sigma_x / \Delta t$. The actual displacement accumulated over a short interval shrinks, while the two independent noise samples have a difference of fixed standard deviation. This result applies to the two-reading estimate; fitting many accurately timed readings over a fixed observation duration uses different information.

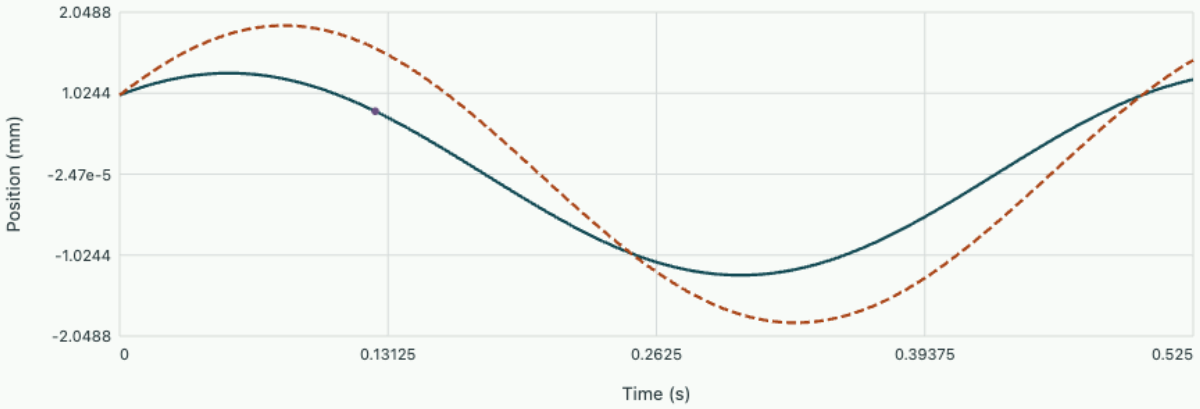
Interactive calculation · Oscillator sampling

Change the physical and numerical inputs of the notebook model. Compare the updated curves and balance checks with the recorded notebook figures. Each plot uses the same equations and assumptions documented in that notebook; its readout names any analytical or reduced-model reference.

Open this live calculation

Try the calculation

Natural frequency (Hz)	**2**
Initial velocity (mm/s)	**10**
Sampling / period	**0.25**



For independent0.10 mm position noise the inferred velocity standard deviation is 1.257 mm/s.

Download this example’s editable notebook

Interactive calculation · Sampling noise

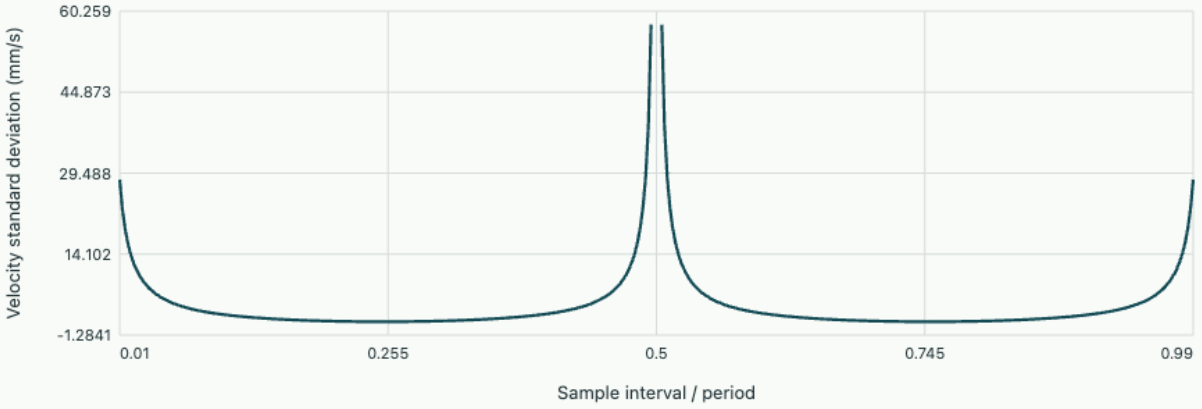
Change the physical and numerical inputs of the notebook model. Compare the updated curves and balance checks with the recorded notebook figures. Each plot uses the same equations and assumptions documented in that notebook; its readout names any analytical or reduced-model reference.

Open this live calculation

Try the calculation

Natural frequency (Hz) **2**

Position noise standard deviation (mm) **0.1**



Two-position reconstruction

Quarter-period minimum standard deviation 1.257 mm/s. The divergence near half-period and full-period sampling is a loss of observability, not floating-point noise.

Download this example's editable notebook

Open editable notebook Download .ipynb

System dynamics, feedback and estimation · 6. When does a stable operating point disappear?

A shaft can hold a static torque against gravity up to a finite limit. Trace the equilibrium branches and distinguish local stability, loss of an equilibrium and the possibility of oscillation.

Construct the operating diagram from a torque balance

A pendulum driven by a constant torque u obeys $J\ddot{\theta} + d\dot{\theta} + k\sin\theta = u$, where $k = mg\ell$, $d > 0$ and $J > 0$. With $v = \dot{\theta}$ the state arrows are $(v, (u - dv - k\sin\theta)/J)$. The first nullcline is $v = 0$. The second is $v = (u - k\sin\theta)/d$. Plotting both gives the local arrow directions before any time integration. Their intersections are precisely static operating points.

For $|u| < k$ there are two intersections per revolution: $\theta_s = \arcsin(u/k)$ and $\theta_u = \pi - \theta_s$, modulo 2π. Linearize about an intersection to obtain

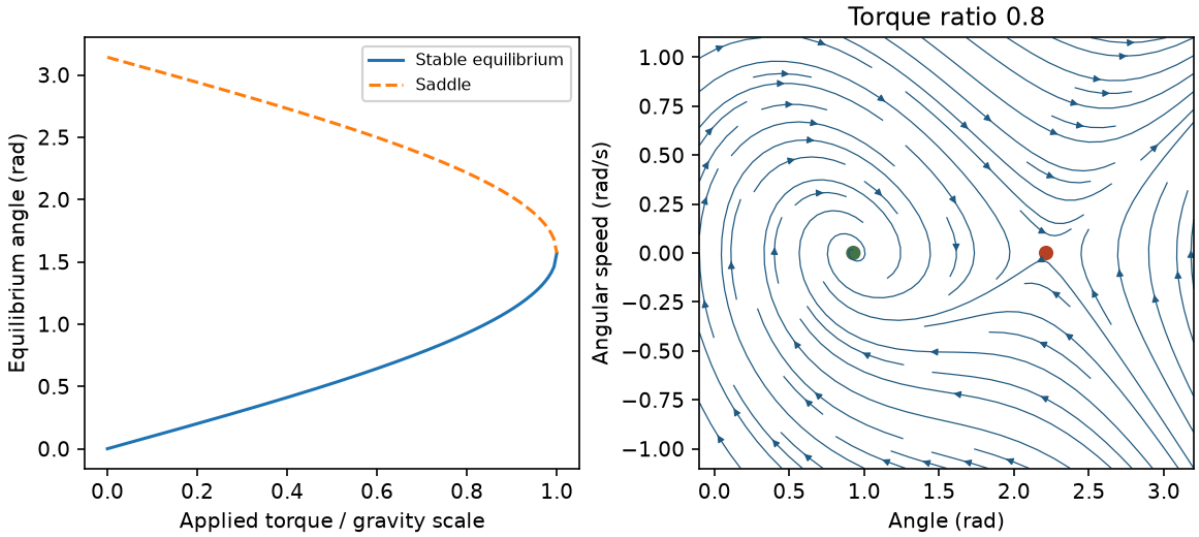
$$J\lambda^2 + d\lambda + k\cos\theta_e = 0.$$

Positive incremental restoring stiffness gives stable poles for $d > 0$. Negative stiffness gives roots of opposite signs and therefore a saddle. At $u = k$ the branches meet and one pole is zero. For $u > k$ there is no static balance. A gain tuned from a linearization on the former branch cannot hold a nonexistent open-loop equilibrium at the same constant torque.

Recover the saddle-node scaling without fitting data

Set $\theta = \pi/2 + x$ and $u = k + \mu$. Since $\sin(\pi/2 + x) = 1 - x^2/2 + O(x^4)$, the slow balance near the merging points is $d\dot{x} \simeq \mu + kx^2/2$. The omitted acceleration changes the fast transient; at the merger one eigenvalue remains $-d/J$ while the other vanishes. For $\mu < 0$ the displacement from the merger scales as $|x| \simeq \sqrt{-2\mu/k}$. The restoring torque gradient tends to zero and perturbations take longer to decay. This is a fold or saddle-node bifurcation.

Use $J = 1$ kg m², $d = 0.4$ N m s/rad, $k = 1$ N m and $u = 0.8$ N m. The equilibria are 0.9273 and 2.2143 rad. Their stiffnesses are +0.6 and −0.6 N m/rad. The stable poles are $-0.2 \pm 0.748i$ s⁻¹. The saddle poles are $+0.6$ and -1.0 s⁻¹. A simulated trajectory that sits exactly at the saddle can stay there indefinitely in exact arithmetic. Perturb it along either eigenvector to identify its phase-plane structure.



The stable and saddle branches meet at the load limit. The phase portrait at 80% load shows trajectories attracted to the lower-angle equilibrium and deflected from the saddle.

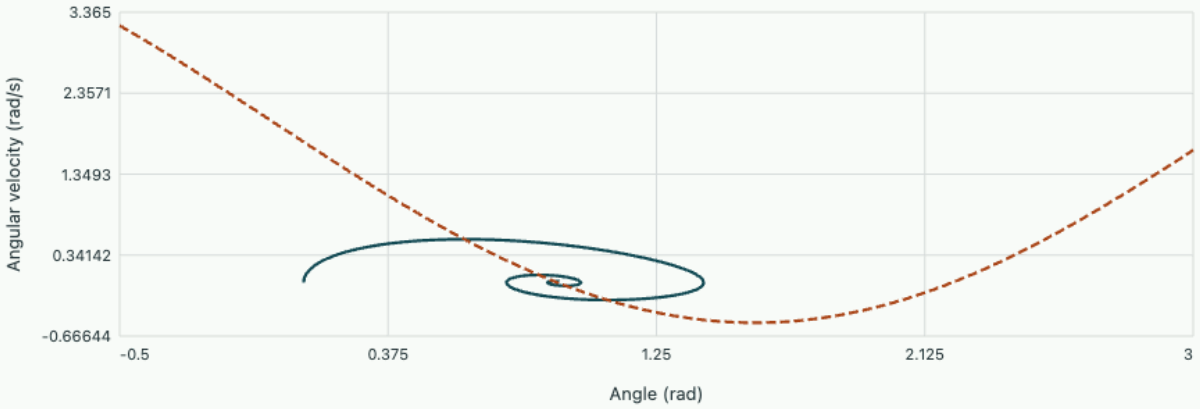
Interactive calculation · Move through the loss of a static operating point

Increase the applied torque to and beyond the gravity limit. Change the starting angle to approach either side of the saddle. The browser integrates the actual pendulum equation with RK4, using inertia1 kg m² and gravity-torque scale1 N m.

Open this live calculation

Try the calculation

Applied / maximum static torque	**0.8**
Damping (N m s/rad)	**0.4**
Initial angle (rad)	**0.1**



Trajectory from rest Acceleration nullcline

Stable branch angle 0.927 rad; saddle branch 2.214 rad. At unity load the branches meet. Inertia and gravity torque scale are both one.

Compare a fold with the onset of a periodic motion

A local stability change need not destroy the equilibrium. In the illustrative planar normal form $\dot{x} = \mu x - y - (x^2 + y^2)x$, $\dot{y} = x + \mu y - (x^2 + y^2)y$, put $r^2 = x^2 + y^2$. Differentiation gives $\dot{r} = \mu r - r^3$ and $\dot{\varphi} = 1$. The origin persists on both sides. For $\mu > 0$ it loses stability and a stable periodic orbit of radius $\sqrt{\mu}$ appears. This supercritical Hopf example has a fundamentally different phase portrait from the torque fold.

What a sweep should record

A bifurcation diagram needs an observable appropriate to each branch. Record equilibrium angles after transients for static branches and maxima and minima for cycles. Direct forward simulation usually misses unstable branches. Solving the equilibrium equations and inspecting their Jacobian recovers them.

Exercises on stability evidence

At $u/k = 0.99$, use the fold expansion to estimate the angular separation of the two equilibria.

▼ **Worked solution**

The near-merger displacement is $\sqrt{2(0.01)} = 0.1414$ rad on either side. The separation is 0.2828 rad, close to the exact $2\arccos(0.99) = 0.2831$ rad.

At $\mu = 0.04$ in the planar normal form, what does a radial perturbation to the periodic orbit do?

▼ **Worked solution**

The orbit radius is 0.2. The radial derivative $\mu - 3r^2 = -0.08$ gives local decay with time constant 12.5 model time units. A phase shift along the orbit persists.

Further reading · Åström and Murray, Feedback Systems, 2nd ed., Chapter 5.

System dynamics, feedback and estimation · 7. Can any controller meet this bandwidth and peak-gain specification?

Delay uses up phase at the crossover even when it changes no gain. An unstable pole near a right-half-plane zero imposes a separate peak-gain limit. Derive each restriction before tuning a controller.

Spend a phase budget at the gain crossover

In a scalar negative-feedback loop let $L = PC$, $S = 1/(1+L)$ and $T = L/(1+L)$. At $|L(i\omega_c)| = 1$, a required margin ϕ_m imposes $\arg L(i\omega_c) \geq -\pi + \phi_m$ on the continuously unwrapped phase at the crossing relevant to stability. Factor a right-half-plane zero at $z > 0$ into a minimum-phase zero and the unit-gain all-pass factor $(z-s)/(z+s)$. Its phase is $-2\arctan(\omega/z)$. Pure delay contributes $-\omega\tau$. Magnitude data alone cannot reveal these phase costs.

For a nominal minimum-phase design contributing -90° at crossover, the available phase budget gives the following exact condition under that assumption. An attainable gain crossover with multiple crossings still requires the full Nyquist stability test.

$$2\arctan(\omega_c/z) + \omega_c\tau \;\leq\; \pi/2 - \phi_m.$$

If $z = 40$ rad/s, $\tau = 0$ and $\phi_m = 45^\circ$, then $\omega_c \leq 40\tan(\pi/8) = 16.57$ rad/s. A proposed 25 rad/s crossover would leave only $90^\circ - 2\arctan(25/40) = 25.99^\circ$. A changed controller may buy lead with noise amplification or a steeper high-frequency roll-off. Reevaluate the complete minimum-phase contribution rather than treating this particular 16.57 as universal.

Use the maximum-modulus principle for a controller-independent test

Now suppose the plant has a simple real unstable pole $p > 0$ and a right-half-plane zero $z > 0$ with $p \neq z$. Assume rational proper internally stabilizing feedback without unstable cancellations. S and T are bounded and analytic in the open right half plane. Internal stability of PS requires $S(p) = 0$, hence $T(p) = 1$. At the plant zero $T(z) = 0$. Divide T by the stable all-pass factor $B_z(s) = (s-z)/(s+z)$. The apparent singularity at z is removable. On the imaginary axis $|B_z| = 1$, so $F = T/B_z$ and T have identical boundary magnitudes.

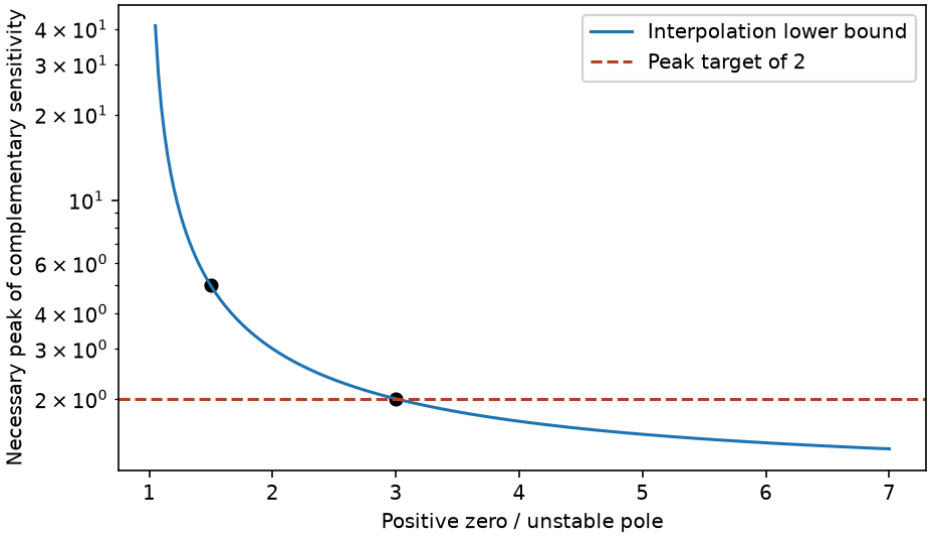
The maximum-modulus principle bounds the interior value $|F(p)|$ by the largest boundary magnitude (including the limit at infinity). Substituting the interpolation values gives

$$\|T\|_\infty \geq |F(p)| = \left|\frac{p+z}{p-z}\right|.$$

For $p = 2$ and $z = 3$ s⁻¹ the peak is at least 5, approximately 14.0 dB. A claimed peak below 2 is infeasible for every controller obeying the stated assumptions, even if the measured range missed the peak. When the pole and zero approach each other the bound diverges. Trying to cancel them inside the controller sacrifices internal stability.

Why the sensitivity integral and this test say different things

The Bode integral constrains the area of log sensitivity across frequency. The maximum-modulus argument uses exact internal interpolation at two specific plant singularities to bound a peak of complementary sensitivity. Which result is sharper depends on the requirement and measured plant.



The necessary peak diverges as the unstable pole and positive zero approach each other. A zero-to-pole ratio of three is the first compatible with this necessary bound and a peak target of two.

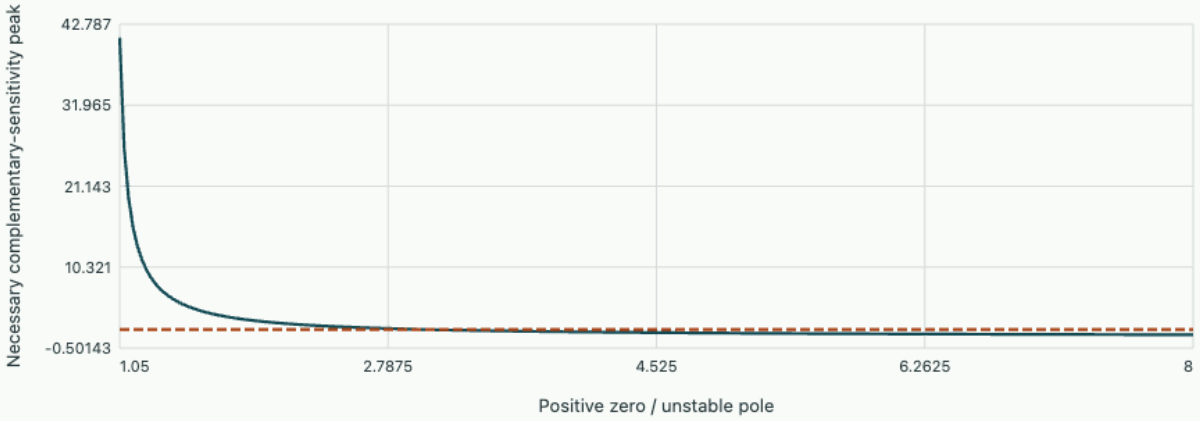
Interactive calculation · Move a positive zero toward an unstable pole

Compare the necessary peak-gain bound with the specified target. The delay affects the separate45° phase budget, not the interpolation inequality.

Open this live calculation

Try the calculation

Unstable pole p (1/s)	1
Allowed peak gain	2
Transport delay (s)	0.05



The positive zero must satisfy z/p ≥ 3.000. At z=3p, a 45° phase budget gives crossover at most 1.143 rad/s. Both conditions are necessary, not a controller construction.

Exercises on impossible requirements

A measured minimum-phase loop has –110° phase at its proposed crossover of 30 rad/s. How much additional pure delay permits a 45° margin?

▼ **Worked solution**

The spare phase is $180 - 110 - 45 = 25^\circ = 0.4363$ rad. At the stated unchanged crossover this permits at most $0.4363/30 = 14.5$ ms.

How far must a positive zero $z > p$ lie above an unstable pole $p = 2$ to avoid ruling out a peak target of 2 by the interpolation bound alone?

▼ **Worked solution**

Require $(z+2)/(z-2) \leq 2$, giving $z \geq 6$ s⁻¹. This removes one necessary-condition violation. It does not construct a controller satisfying every constraint.

Further reading · Åström and Murray, Feedback Systems, 2nd ed., Chapters 13–14, especially sections 14.3–14.4.

System dynamics, feedback and estimation · 8. Designing current, speed and position loops for a drive

An encoder measures shaft motion, while the power stage acts on voltage. Decide which disturbances each measurement can reject and how fast each loop can move before the unseen inner dynamics erode the position margin.

Place a fast loop where the disturbance enters

Use the small-signal motor equations $L\dot{i} + Ri = v - k_e\omega$, $J\dot{\omega} + B\omega = k_t i - \tau_L$, and $\dot{\theta} = \omega$. Measure current at the converter and motion at the shaft. A speed loop alone leaves rapid electrical disturbances to propagate before they appear in shaft motion. Close a fast local current loop and send it a current reference from the mechanical loop. Treat back EMF as a disturbance or add measured-speed feedforward.

Choose $C_i(s) = k_{pi} + k_{ii}/s$ with $k_{ii}/k_{pi} = R/L$. For the nominal $1/(Ls + R)$ current plant the open loop reduces to $k_{pi}/(Ls)$ and the closed transfer is $\omega_i/(s + \omega_i)$, with $k_{pi} = L\omega_i$ and $k_{ii} = R\omega_i$. At $L = 0.010$ H, $R = 1$ ohm and $\omega_i = 1000$ rad/s the gains are 10 V/A and 1000 V/(A s). The approximation presupposes adequate measurement, switching and computation bandwidth.

Choose the middle loop against an inner-loop error budget

If current follows its command over the speed-loop band, the torque-to-speed plant is $k_t/(Js + B)$. A speed PI controller with $k_{p\omega} = J\omega_v/k_t$ and $k_{i\omega} = B\omega_v/k_t$ gives an approximate closed speed response $\omega_v/(s + \omega_v)$. With $J = 0.020$, $B = 0.020$, $k_t = 0.40$ in SI and $\omega_v = 100$ rad/s, the gains are 5 in the corresponding current-per-speed and integral units. The neglected current lag contributes $\arctan(100/1000) = 5.71^\circ$ at that speed-loop crossover. Reserve this phase when validating the full model. A torque disturbance enters the mechanical balance directly and is primarily rejected by this middle loop.

Convert a position requirement into a realizable crossover

Use position proportional feedback $\omega_{ref} = K(\theta_{ref} - \theta)$, the achieved speed model $\omega_v/(s + \omega_v)$ and total outer-path delay τ. In this reduced model the exact position-loop frequency response is $L(s) = Ke^{-s\tau}/[s(1 + s/\omega_v)]$. Choosing K for a desired amplitude crossover separates magnitude tuning from the residual phase budget:

$$K = \omega_c\sqrt{1 + (\omega_c/\omega_v)^2},$$

$$\phi_m = \pi/2 - \arctan(\omega_c/\omega_v) - \omega_c\tau.$$

At $\tau = 6$ ms, $\omega_v = 100$ rad/s and required margin 45°, solving the monotone equation gives $\omega_c \le 51.56$ rad/s. The 50 rad/s design has only 46.25° before any additional sensor or drive lag. At 20 rad/s the nominal margin is 71.81°. The notebook separately evaluates the delay-free closed-loop poles with OneSimulation. Their decay rate and the delayed loop's phase margin measure different properties.

Interactive experiment · Spend the phase budget with bandwidth and delay sliders

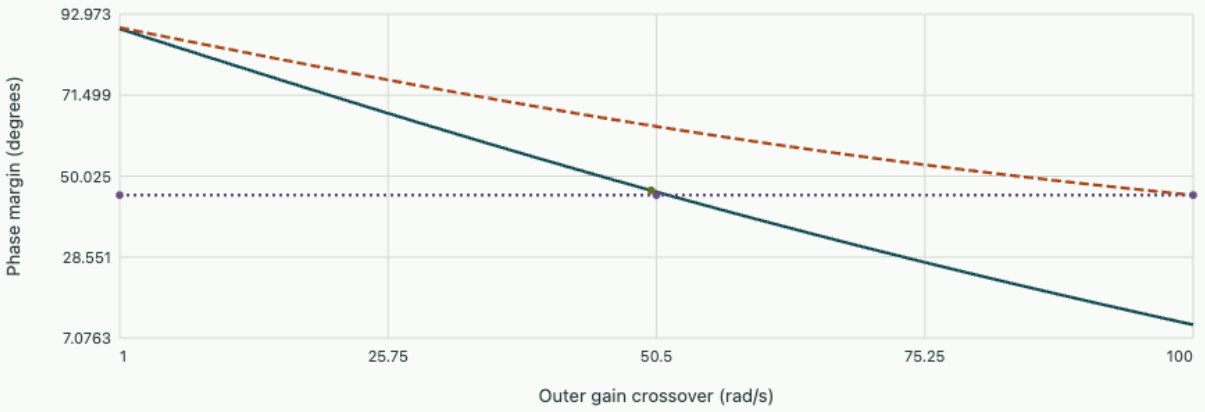
Choose the achieved inner-loop bandwidth, the outer-path delay and the intended outer crossover. The browser adjusts the proportional gain to place the amplitude crossover exactly at that frequency, then evaluates the remaining phase margin. The 45° line is a design target; it is not the boundary of stability itself.

Begin at an inner bandwidth of 100 rad/s, delay 6 ms and crossover 50 rad/s. Double the delay and observe the lost 17.19°. Then lower the requested crossover until the target is recovered. The zero-delay pole margin in the readout is computed from the quadratic closed-loop characteristic polynomial. Compare it with the OneSimulation stability diagnostic in the notebook; it does not certify the delayed system.

Open this experiment online

Try the calculation

Achieved inner bandwidth (rad/s)	100
Outer-path delay (ms)	6
Chosen outer crossover (rad/s)	50



Chosen gain K = 55.902 /s gives 46.25° phase margin. The 45° boundary is 51.56 rad/s. The delay-free closed-loop pole margin is 50.000 /s; it is a separate diagnostic from the delayed phase margin.

Continue in the OneSimulation notebook · Download its Python model

OneSimulation experiment · An achieved inner speed bandwidth and transport delay bound the outer-loop crossover

Run the notebook with OneSimulation 0.2.0, NumPy, SciPy and Matplotlib. Change the input data and compare your run with the recorded result.

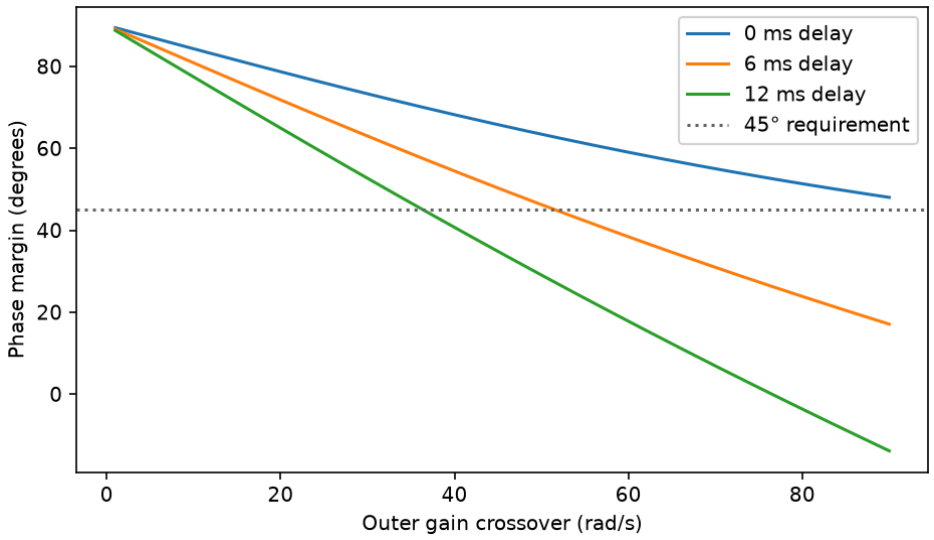
```
"""Evaluate inner-lag and transport-delay limits on an outer position loop."""
import numpy as np
import matplotlib.pyplot as plt
from onesim.physics.systems.control import stability_margin

inner_pole = 100.0  # rad/s, achieved bandwidth of the closed speed loop
def outer_gain(crossover):
    return crossover*np.sqrt(1+(crossover/inner_pole)**2)
def phase_margin(crossover, delay):
    return 90-np.rad2deg(np.arctan(crossover/inner_pole)+crossover*delay)

frequencies = np.linspace(1,90,250)
fig, ax = plt.subplots(figsize=(6.4,3.7), layout="constrained")
for delay in [0.0,0.006,0.012]:
    ax.plot(frequencies, phase_margin(frequencies,delay),label=f"{1000*delay:.0f} ms delay")
for crossover in [20.0, 50.0]:
    gain = outer_gain(crossover)
    poles = np.roots([1.0,inner_pole,inner_pole*gain])
    print(f"crossover {crossover:.0f} rad/s: delay-free pole margin = "
          f"{stability_margin(poles,discrete=False):.3f} /s, "
          f"6 ms phase margin = {phase_margin(crossover,0.006):.2f} deg")
ax.axhline(45,color="0.4",ls=":",label="45° requirement")
ax.set(xlabel="Outer gain crossover (rad/s)",ylabel="Phase margin (degrees)")
ax.legend()
plt.show()
```

Reference run

```
crossover 20 rad/s: delay-free pole margin = 28.543 /s, 6 ms phase margin = 71.81 deg
crossover 50 rad/s: delay-free pole margin = 50.000 /s, 6 ms phase margin = 46.25 deg
```



An achieved inner speed bandwidth and transport delay bound the outer-loop crossover

Download editable notebook · Download Python script

Inspect failure signatures from the inside out

Measure the current-reference response with the mechanical loop open in the model. Then close speed feedback and compare measured torque-step recovery with the linear prediction. Finally apply a position command and log reference and achieved current, shaft speed and position together. Under voltage clipping the actual current cannot track its command at speed. A speed integrator that keeps charging can dominate the later position excursion. Use actuator feedback for anti-windup and check the actual limits at the operating speed.

Which loop receives the position sample delay?

A rotor-position timestamp used to derive speed can delay both mechanical loops. A transport delay introduced only between the supervisory controller and the drive delays the outer loop. Draw the measurement paths before inserting a single lumped delay or counting it twice.

Exercises on a changed architecture

Keep crossover at 50 rad/s while total delay rises from 6 to 12 ms. What does this do to the reduced-model margin?

▼ **Worked solution**

The gain is unchanged by pure delay. The added lag is $50(0.006) = 0.30$ rad, or 17.19°. Margin falls from 46.25° to 29.06°, below the target.

At high speed the current sensor follows its local model but digital speed estimates have 4 ms of additional filtering delay. Where should the new uncertainty enter?

▼ **Worked solution**

Place it in the speed measurement path. Recompute the middle-loop closed dynamics before using them as the position plant. Adding it only to the outer-loop delay would miss loss of speed-loop robustness.

Further reading · Åström and Murray, Feedback Systems, 2nd ed., Chapters 10–11 and 15.

From a moving plant to identified dynamics

Derive a plant from force and energy, then compare its modes with measured shifts and decay. Examine the uncertainty that remains when two parameters produce similar responses.

1. Free vibration, energy loss and ring-down data
2. Harmonic forcing, phase and dissipated power
3. Coupled oscillators and normal coordinates
4. Energy coordinates and Lagrange’s equation
5. Finite-amplitude oscillations and model validity
6. A moving-coil galvanometer as an electrical and mechanical system
7. Coupled optical and mechanical resonances
8. Predictions, parameter profiles and nonlinear uncertainty
9. Learning a drifting temperature from a sequence of readings
10. Predicting a new measurement when the fitted parameters are correlated
11. Measure gravity with a pendulum of finite amplitude
12. Measure inertia and gravity with two suspension axes

Free vibration, energy loss and ring-down data

Machine dynamics

Infer stiffness and damping from a released machine mount and check the implications of a viscous damping model.

FROM FIRST PRINCIPLES

Determine the frequency and decay of a released mount

Displace a mass m from its static equilibrium on a linear spring of stiffness k and release it. A viscous damper exerts force opposite to velocity with coefficient c. Measuring displacement x relative to static equilibrium removes any constant force already balanced by the spring. Newton's second law gives the free-motion equation.

$$m\ddot{x} = -kx - c\dot{x}.$$

An exponential trial e^{st} turns differentiation into multiplication by s, so nonzero motion requires a root of the characteristic polynomial. Its two roots depend on the undamped natural frequency ω_n and damping ratio ζ.

$$ms^2 + cs + k = 0,$$

$$\omega_n = \sqrt{k/m},$$

$$\zeta = \frac{c}{2\sqrt{km}}.$$

For $0 < \zeta < 1$ the roots are a conjugate pair. Their oscillatory contributions combine to give a real damped sinusoid with envelope $e^{-\zeta\omega_n t}$.

$$s = -\zeta\omega_n \pm i\omega_n\sqrt{1-\zeta^2}.$$

The initial position and velocity determine amplitude and phase, while the parameters set decay and period. Same-sign peaks one period apart therefore measure the exponential envelope without requiring an accurately known release time.

Background · Damping ratio

$$\zeta = c/(2\sqrt{km})$$

This compares a viscous damper with critical damping for a single linear oscillator. It cannot represent Coulomb friction at every amplitude.

DERIVATION AND APPLICATIONS

Infer damping from the loss of mechanical energy per cycle

A lightly damped mount sheds a measurable fraction of stored energy on each oscillation. Consider a single translating mass m with linear stiffness k, viscous damping c, and no forcing after release. Measure displacement from static equilibrium. Multiply the equation of motion by velocity to convert inertia and spring terms into total time derivatives.

$$m\ddot{x} + c\dot{x} + kx = 0,$$

$$\frac{d}{dt}\left(\tfrac{1}{2}m\dot{x}^2 + \tfrac{1}{2}kx^2\right) = -c\dot{x}^2.$$

For weak damping, amplitude A changes little during one period and $x \simeq A\cos\omega_n t$, with $\omega_n = \sqrt{k/m}$. Integrating viscous power over the cycle then gives the leading loss. Divide by the stored maximum spring energy evaluated at the same amplitude.

$$\Delta E_{\text{loss}} \simeq cA^2\omega_n^2\frac{\pi}{\omega_n} = \pi c\omega_n A^2, \qquad \frac{\Delta E_{\text{loss}}}{E} \simeq 4\pi\zeta,$$

$$\zeta = \frac{c}{2m\omega_n}.$$

For appreciable damping use two same-sign turning points directly. At each peak the kinetic energy vanishes, so the exact energy ratio equals the square of the amplitude ratio. The measured logarithmic decrement relates to viscous damping for an underdamped oscillator.

$$\frac{E_{j+1}}{E_j} = \left(\frac{A_{j+1}}{A_j}\right)^2 = e^{-2\delta},$$

$$\zeta = \frac{\delta}{\sqrt{4\pi^2 + \delta^2}},$$

$$\delta = \ln\frac{A_j}{A_{j+1}}.$$

The energy fraction is independent of starting amplitude for linear viscous damping. Amplitude dependence in repeated runs is therefore a direct reason to examine extra friction or nonlinear damping.

WORKED EXAMPLE

Determine heat release from two peak readings

For $k = 800$ N/m, consecutive positive peaks are 10.0 and 9.00 mm. The spring energies are 40.0 mJ and 32.4 mJ, so 7.60 mJ is dissipated. The decrement is $\ln(10/9) = 0.1054$, giving $\zeta = 0.0168$. If the first amplitude were doubled and the same linear damper retained, the energy lost in that cycle would quadruple.

Change the inputs and inspect this calculation below

SECOND APPLICATION

Test a damper from its measured cycle work

A separate mount has $k = 2000$ N/m, $\omega_n = 50.0$ rad/s and small oscillations of amplitude 2.00 mm. A damper dynamometer measures 0.250 mJ of work per cycle. The leading estimate gives $c = \Delta E/(\pi\omega_n A^2) = 0.398$ N s/m. From $m = k/\omega_n^2 = 0.800$ kg, $\zeta = 0.00497$. The estimated loss fraction $4\pi\zeta = 0.0625$ supports the nearly constant-amplitude approximation for a single cycle.

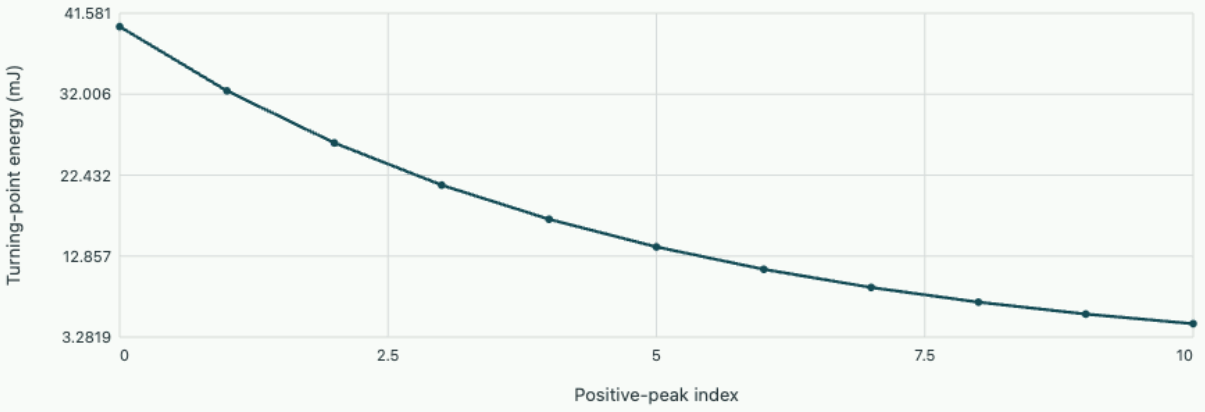
Interactive calculation · Change the inputs of this example

Use the adjustable quantities to test the two worked calculations above. The plot recomputes this derivation's balance or constitutive relation; its numerical readout reports the quantities needed for the forward calculation and the inverse or design question. Keep the fixed assumptions stated in the readout when comparing with the example values.

Open this live calculation

Try the calculation

Peak displacement (mm)	**10**
Next/previous peak	**0.9**
Stiffness (N/m)	**800**



Measured geometric peak decay

First-cycle loss 7.600 mJ; inferred ζ=0.01677.

Displace a machine from its static position

A pump on a spring mount settles under its weight. Pull it slightly farther down and release it. The weight still acts, but the extra motion is driven by the change in spring force. Define x as displacement from the loaded equilibrium, take constant local spring stiffness k, and represent dissipation by a force opposing velocity of magnitude $c|\dot{x}|$. Newton's law becomes

$$m\ddot{x} + c\dot{x} + kx = 0.$$

This model describes one isolated mode over an amplitude range where inertia, stiffness and damping are approximately constant.

Multiply the equation by $\dot{x}$. The mass term differentiates kinetic energy, and the spring term differentiates stored elastic energy. Consequently $d(m\dot{x}^2/2 + kx^2/2)/dt = -c\dot{x}^2$. Energy cannot grow for positive damping. Displacement can increase temporarily as kinetic energy converts back into spring energy, so the instantaneous displacement alone is not an energy meter.

Obtain the decay and oscillation rates

Trying $x = e^{st}$ gives

$$ms^2 + cs + k = 0.$$

Define $\omega_n = \sqrt{k/m}$ and $\zeta = c/(2m\omega_n)$. The roots are

$$s = -\zeta\omega_n \pm \omega_n\sqrt{\zeta^2 - 1}.$$

For $0 < \zeta < 1$ their imaginary part gives the damped angular frequency $\omega_d = \omega_n\sqrt{1-\zeta^2}$. The real part gives a decaying exponential. A general underdamped trace is

$$x(t) = e^{-\zeta\omega_n t}\left[A\cos(\omega_d t) + B\sin(\omega_d t)\right].$$

At $\zeta = 1$ the repeated root gives $(A + Bt)e^{-\omega_n t}$. Above critical damping both roots are negative and real. A lack of visible oscillation therefore does not by itself determine the stiffness. For an underdamped trace, use two peaks of the same sign: their separation is $T_d = 2\pi/\omega_d$, and their amplitude ratio depends only on the envelope decay.

Background · Which energy is represented by a sequence of displacement peaks?

At a turning point the velocity is zero. If displacement is measured from static equilibrium, the energy of the linear oscillator there is entirely elastic. Two peaks of the same sign one damped period apart give an exact energy ratio even though the instantaneous dissipation varies within that period.

$$\frac{E_{n+1}}{E_n} = \left(\frac{x_{n+1}}{x_n}\right)^2,$$

$$\frac{\Delta E_{\text{lost}}}{E_n} = 1 - e^{-2\delta},$$

$$\delta = \ln\frac{|x_n|}{|x_{n+1}|}.$$

If successive positive peaks fall from 10 to 8 mm, 36% of the initial peak energy has been lost in one cycle. The amplitude has fallen by only 20%. Near critical damping it may be difficult to resolve a full period above sensor noise. The definition then remains valid for a measurable underdamped cycle, but the inferred ratio becomes sensitive to peak errors.

Infer parameters from four cycles

A 20.0 kg mounted component rings down from a positive peak of 8.0 mm to a positive peak of 2.0 mm four cycles later, after 1.60 s. The period is $T_d = 0.400$ s and $\omega_d = 15.71$ rad/s. The log decrement per cycle is

$$\delta = \ln(8/2)/4 = 0.3466.$$

Since $x_j/x_{j+1} = \exp(\zeta\omega_n T_d)$, the envelope decay rate is

$$\alpha = \delta/T_d = 0.8664\,\text{s}^{-1}.$$

Now $\omega_n = \sqrt{\omega_d^2 + \alpha^2} = 15.73$ rad/s and $\zeta = \alpha/\omega_n = 0.0551$. The inferred stiffness is $m\omega_n^2 = 4.95$ kN/m, while $c = 2m\alpha = 34.7$ N·s/m. After eight cycles the prediction is 0.50 mm. Using peaks separated by several cycles reduces sensitivity to noise in any one amplitude or timestamp.

Challenge the damping assumption

Subtract the equilibrium sensor reading before taking amplitude ratios. A fixed offset changes the ratio of small late peaks more strongly than large early peaks. Plot $\ln|x_j|$ against time for peaks whose signal clearly exceeds the noise floor. A straight trend supports the viscous envelope. Approximately constant

loss of amplitude per cycle suggests a different mechanism such as dry friction. Two nearby natural frequencies may produce a beat envelope that grows and shrinks even though total energy decays.

A second experiment can separate stiffness from added inertia. If a small known mass is bolted to the same moving assembly and the support stiffness is unchanged, $1/\omega_n^2$ should increase in proportion to effective mass. This check assumes that the added mass participates fully in the chosen coordinate. A flexible appendage that moves out of phase changes the system being identified.

Correct a biased displacement signal

Suppose a displacement sensor reads 0.50 mm high relative to the true equilibrium. The first and fifth positive peaks then appear as 8.50 and 2.50 mm. Their raw log ratio over four cycles gives $0.25 \ln(8.5/2.5) = 0.306$, underestimating the true 0.347 decrement. The error grows as the decaying signal approaches the offset. One remedy is to measure equilibrium independently when excitation has stopped. Another is to use the difference between consecutive positive and negative extrema and compare such differences one whole period apart. The constant offset cancels while, for a single underdamped mode, successive peak-to-trough ranges retain the same exponential ratio. Late peaks below the sensor resolution still carry little reliable damping information.

Derive the repeated-root response and its crossing condition

The oscillatory ring-down described above assumes $\zeta < 1$. For a return mechanism holding a fragile component, the question can instead be whether it reaches equilibrium without crossing it. At critical damping, $c = 2\sqrt{mk}$, the characteristic equation is

$$(s + \omega_n)^2 = 0.$$

The repeated root gives two independent free solutions $e^{-\omega_n t}$ and $te^{-\omega_n t}$. Imposing $x(0) = x_0, \dot{x}(0) = v_0$ yields

$$x(t) = [x_0 + (v_0 + \omega_n x_0)t]e^{-\omega_n t}.$$

For release from rest at positive displacement, the bracket stays positive and $\dot{x} = -\omega_n^2 x_0 t e^{-\omega_n t}$ stays negative for $t > 0$. If a mechanism reaches the same x_0 with negative incoming velocity, the bracket can cross zero once even at critical damping. In particular $v_0 < -\omega_n x_0$ gives crossing time $-x_0/(v_0 + \omega_n x_0)$. Critical damping rules out repeated free oscillations, but the initial velocity still determines whether the target is overshot.

For overdamping the distinct real decay rates are $\omega_n[\zeta \pm \sqrt{\zeta^2 - 1}]$. Their product is ω_n^2. Increasing viscous resistance makes the fast root decay more rapidly but the slow root approach zero. Simply adding damping can consequently prolong the final approach to a sensing threshold.

Select a damper for a spring-return slide

A slide of mass 3.0 kg has a return spring of 300 N/m, so $\omega_n = 10\,\mathrm{rad/s}$ and $c_{\mathrm{cr}} = 60\,\mathrm{N\,s/m}$. Released from rest at $x_0 = 50$ mm with critical damping, its normalised position is $(1 + 10t)e^{-10t}$. To enter a 5% band, solve $(1 + s)e^{-s} = 0.05$ with $s = 10t$. Testing 4.7 and 4.8 gives 0.0518 and 0.0477; the crossing is about 4.74, or $t = 0.474$ s.

The same slide at the same 50 mm location might arrive with $v_0 = -0.80\,\mathrm{m/s}$ after an actuator disconnects. The bracket is now $0.050 - 0.30t$, which crosses at 0.167 s. Differentiating gives its negative peak at $t = 0.267\,\mathrm{s}$, with an overshoot magnitude about 2.08 mm. An optical stop positioned only 1 mm beyond equilibrium would register a crossing even though the damper is critical. Compare that with a slow overdamped return from rest: for $\zeta = 2$ its dominant decay rate falls to $10(2 - \sqrt{3}) = 2.68\,\mathrm{s^{-1}}$. An assessment based simply on whether the motion oscillates would overlook both the clearance requirement and this much slower return.

The return time depends on which event the instrumentation calls completion. Some switches trigger when the position first enters a band and remain latched even if it later leaves. For the critical release with negative entering speed, first crossing the equilibrium plainly does not mean motion has stopped. A position-plus-velocity criterion is stricter. At the first zero crossing, the bracket in x vanishes, so the velocity is its constant derivative multiplied by the exponential, $-0.30e^{-10(0.167)} = -0.0567\,\mathrm{m/s}$. A slide with a maximum docking speed of 0.020 m/s would fail that condition despite its position being exactly correct at that instant. The appropriate response test must state both allowed path excursion and the speed at any mechanical contact.

Checking only the first displacement peak cannot identify this velocity-limited docking event.

Distinguish sliding-friction decay from a viscous ring-down

Suppose a slide moves on a dry horizontal guide under a spring force $-kx$. While the slide moves left, kinetic friction F_c acts right, and $m\ddot{x} + kx = +F_c$. The solution is harmonic about the shifted equilibrium $+F_c/k$. During rightward sliding the equilibrium switches to $-F_c/k$. A complete undisturbed sliding half-cycle still lasts $\pi\sqrt{m/k}$, but friction dissipates energy in proportion to distance travelled. At zero speed, static friction can take any value up to F_s, so checking the spring force before starting the next half-cycle is essential.

Let a leftward excursion start at $x = A > 0$ and reach a turning point at $x = -B$, with $A, B > 0$. The spring energy drop equals friction work,

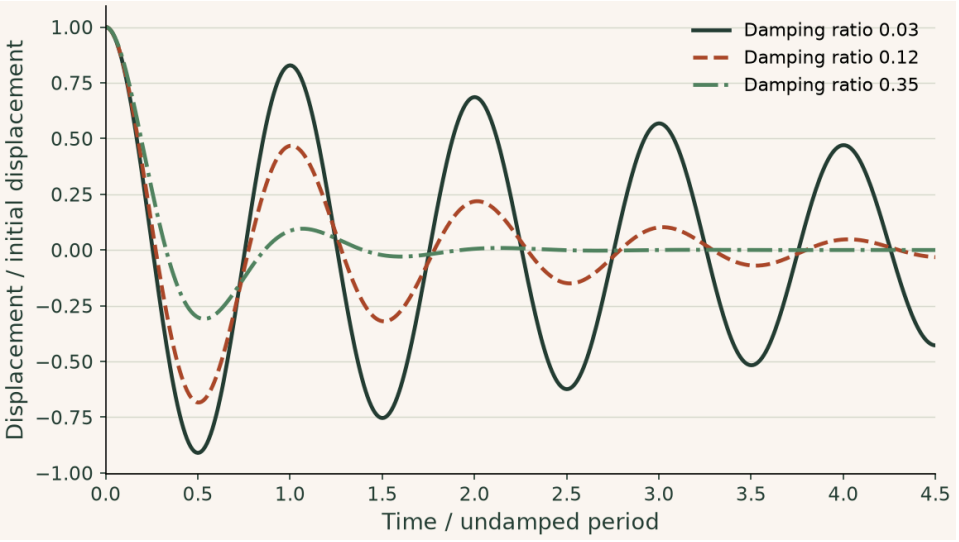
$$k(A^2 - B^2)/2 = F_c(A + B).$$

Cancellation leaves $B = A - 2F_c/k$. Same-side peak magnitudes fall by $4F_c/k$ each complete sliding cycle. This is a constant absolute decrement. A viscous exponential gives a constant ratio instead. If a late half-cycle does not cross zero, use its signed turning-point coordinate $x_{\mathrm{next}} = 2F_c/k - x_{\mathrm{start}}$ for leftward motion and apply the static criterion there.

A 2 kg slide has $k = 800\,\mathrm{N/m}$, $F_c = 2.0$ N and $F_s = 2.4$ N. Release it from 18 mm. Its initial spring force is 14.4 N, so it moves. Each half-cycle reduces the alternating peak magnitude by 5.0 mm, giving signed positions $+18, -13, +8, -3$ mm. At −3 mm the spring pull is 2.4 N, equal to the static threshold. Under an ideal static-friction law it can remain there. It took $3\pi\sqrt{2/800} = 0.471$ s to reach that position. The final displacement need not be zero. Interpreting its first two positive peaks with a logarithmic decrement would give a damping estimate that fails to predict the finite stop.

Estimate damping from repeated peaks

Release an initially displaced mass with zero initial velocity. The underdamped waveform contains both cosine and sine terms; the latter enforces zero initial velocity. Compare successive same-sign peaks using a logarithmic decrement.



At damping ratio 0.12 each positive peak retains $\exp(-2\pi\zeta/\sqrt{(1-\zeta^2)}) \approx 0.468$ of the previous one. Taking consecutive positive and negative extrema instead halves the interval and changes the inferred decrement.

▶ **Read values from the plot**

Interactive calculation · change the inputs of this example

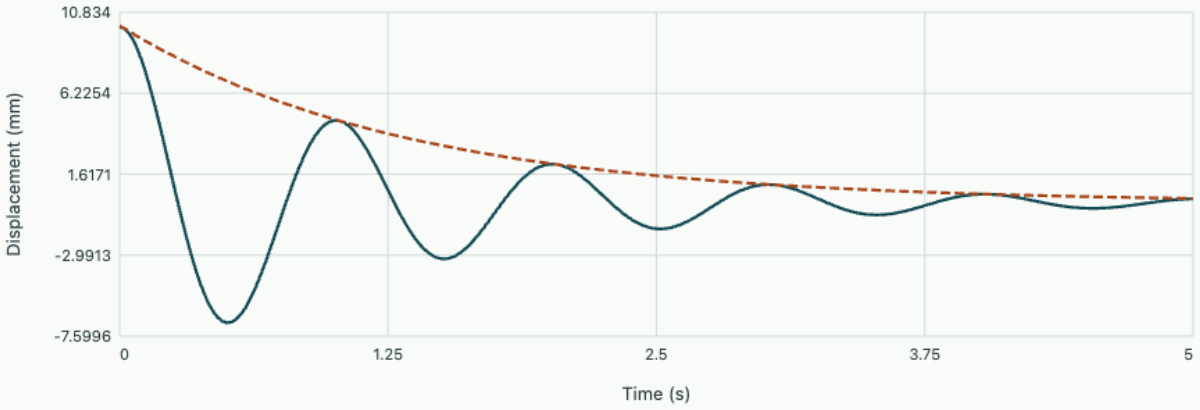
Change damping before changing natural frequency. Separate the decay per cycle from the number of cycles per second.

The curves below are recalculated from this section's equations in your browser. The fixed reference plot above remains available for comparison.

Open this live calculation

Try the calculation

Damping ratio	**0.12**
Natural frequency (Hz)	**1**
Initial displacement (mm)	**10**



Release from rest Exponential envelope

Successive same-sign peak ratio 0.468. Damped period 1.007 s.

Harmonic forcing, phase and dissipated power

Machine dynamics

Predict the steady vibration under a rotating force and relate the phase lag to the power absorbed by damping.

FROM FIRST PRINCIPLES

Obtain forced vibration amplitude from dynamic equilibrium

A machine on a linear spring and viscous damper is driven by force $F_0 \cos \omega t$ on its mass m. Newton's second law sums inertia, damping and stiffness forces. Assume a stable mount and wait until the transient part has decayed. Seek a sinusoidal response at the forcing frequency, written $x(t) = \mathrm{Re}[X e^{i\omega t}]$ to retain both magnitude and phase.

$$m\ddot{x} + c\dot{x} + kx = F_0 \cos \omega t, \qquad (k - m\omega^2 + ic\omega)X = F_0.$$

The real part of the dynamic stiffness combines spring and inertia, while the damping contribution is in quadrature. Taking its modulus gives displacement amplitude; taking its argument gives the lag of displacement behind the applied force.

$$|X| = \frac{F_0}{\sqrt{(k - m\omega^2)^2 + (c\omega)^2}},$$

$$\phi = \operatorname{atan2}(c\omega, k - m\omega^2).$$

At low frequency the displacement approaches static deflection F_0/k. Near $\sqrt{k/m}$ the spring and inertia terms cancel, leaving damping to limit the steady response. High-frequency response falls approximately as $F_0/(m\omega^2)$ for a fixed force amplitude.

Background · Damped frequency

$$\omega_d = \omega_n \sqrt{1 - \zeta^2}$$

This free-decay oscillation frequency exists in the underdamped regime. A resonance peak in a driven response uses a different frequency condition.

DERIVATION AND APPLICATIONS

Separate displacement resonance from maximum absorbed power

The frequency of largest displacement can differ from the frequency of largest mean heating in a damper. Drive a single mass m on a linear spring k and viscous damper c with a fixed peak force F at angular frequency ω. Assume sinusoidal steady state, a rigid base and coefficients independent of amplitude. With complex displacement amplitude X, mean energy dissipated per unit time is c times mean squared velocity. Force equilibrium gives

$$X = \frac{F}{k - m\omega^2 + ic\omega},$$
$$\bar{P} = \frac{1}{2} c\omega^2 |X|^2.$$

Rewrite the power by dividing numerator and denominator by ω^2. The resulting mechanical impedance has a real dissipative part c and a reactive part $m\omega - k/\omega$. The power peak occurs where this latter term vanishes.

$$\bar{P} = \frac{cF^2/2}{c^2 + (m\omega - k/\omega)^2},$$
$$\omega_{\text{peak}} = \sqrt{k/m},$$
$$\bar{P}_{\max} = \frac{F^2}{2c}.$$

At half maximum the magnitude of the reactive term equals c. The positive solutions of the two resulting quadratic equations give lower and upper half-power frequencies. Their difference is exact for this force-driven, velocity-power measurement.

$$\omega_{\pm} = \frac{\sqrt{c^2 + 4mk} \pm c}{2m},$$
$$\omega_+ - \omega_- = \frac{c}{m},$$
$$\omega_+ \omega_- = \frac{k}{m}.$$

By comparison, the displacement peak contains the additional $1/\omega$ in converting velocity to displacement, so applying these exact bandwidth relations to a displacement trace requires care.

WORKED EXAMPLE

Predict the heating bandwidth

Use $m = 2.00$ kg, $k = 800$ N/m, $c = 8.00$ N s/m and $F = 4.00$ N. The peak is at 20.0 rad/s and maximum mean dissipation is

$$16/16 = 1.00\,\mathrm{W}.$$

The half-power frequencies are 18.10 and 22.10 rad/s. At the peak, velocity is in phase with force and has peak amplitude $F/c = 0.500$ m/s. The corresponding displacement peak amplitude at this frequency is 25.0 mm.

Change the inputs and inspect this calculation below

SECOND APPLICATION

Identify spring and damper from two half-power points

Another oscillator of mass 1.50 kg has measured half-power frequencies of 12.0 and 18.0 rad/s. Their difference gives $c = 1.50(6.00) = 9.00$ N s/m and their product gives

$$k = 1.50(12)(18) = 324\,\mathrm{N/m}.$$

The predicted power peak is at the geometric mean, $\sqrt{216} = 14.7$ rad/s. A direct power sweep can test this prediction without a calibrated displacement amplitude.

Interactive calculation · Change the inputs of this example

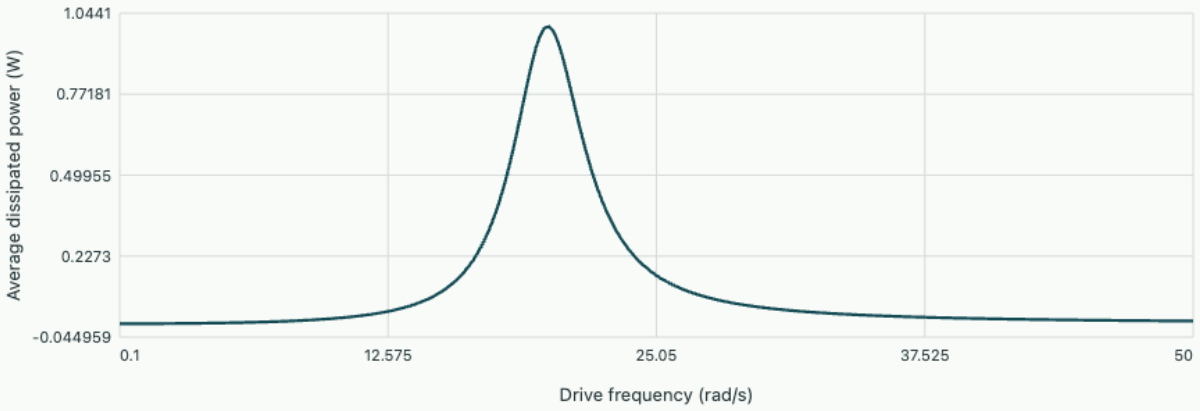
Use the adjustable quantities to test the two worked calculations above. The plot recomputes this derivation's balance or constitutive relation; its numerical readout reports the quantities needed for the forward calculation and the inverse or design question. Keep the fixed assumptions stated in the readout when comparing with the example values.

Open this live calculation

Try the calculation

Damper (N s/m) 8

Force amplitude (N) 4



Damping heat

At natural frequency20 rad/s, maximum power=1.000 W. Half-power bandwidth c/m=4.000 rad/s for mass2 kg.

Extract one sinusoidal component

A slightly eccentric rotating part exerts a periodic force on its housing. Measure the housing displacement from static equilibrium and first represent a single transverse force component by $F_0 \cos \omega t$. Linear response then satisfies

$$m\ddot{x} + c\dot{x} + kx = F_0 \cos \omega t.$$

Its total solution contains the free transient plus a steady periodic part. Wait several decay times before comparing measured peak amplitudes with a harmonic prediction.

Write the steady part as $\mathrm{Re}\{Xe^{i\omega t}\}$, taking the real force phasor F_0 as phase reference. Each differentiation contributes $i\omega$. The displacement phasor therefore satisfies

$$(k - m\omega^2 + ic\omega)X = F_0.$$

The factor in parentheses is dynamic stiffness in N/m. At low speed the phasor is close to F_0/k. At high speed the inertia term dominates and displacement is almost opposite the applied force. A phase measurement can distinguish those regimes even if their displacement amplitudes happen to be equal.

Read magnitude and phase separately

Introduce $r = \omega/\omega_n$, $\omega_n = \sqrt{k/m}$ and $\zeta = c/(2m\omega_n)$. Take the magnitude and argument of the dynamic stiffness to obtain

$$|X|/(F_0/k) = [(1 - r^2)^2 + (2\zeta r)^2]^{-1/2}.$$

The displacement lags by $\phi = \operatorname{atan2}(2\zeta r, 1 - r^2)$. At $r = 1$ elastic and inertia forces cancel in phase and the remaining quadrature force is viscous. The displacement lags by 90 degrees and has amplitude $F_0/(c\omega_n)$.

For constant force amplitude, maximum displacement occurs at $r = \sqrt{1 - 2\zeta^2}$ if $\zeta < 1/\sqrt{2}$. This follows by differentiating the squared denominator with respect to r. Thus the frequency of greatest displacement need not coincide exactly with the undamped natural frequency. A rotating imbalance also changes force amplitude with speed, so its peak cannot be inferred from a constant-force plot alone.

Check a resonant measurement by energy

Let $m = 10$ kg, $k = 4000$ N/m, $c = 40$ N·s/m and $F_0 = 8.0$ N. Then $\omega_n = 20$ rad/s and $\zeta = 0.10$. At resonance the predicted displacement amplitude is

$$8/[40(20)] = 10.0\,\mathrm{mm}.$$

The velocity amplitude is 0.200 m/s. The time-average input power of a sinusoidal force and velocity is half the product of their peak magnitudes times the cosine of their phase difference. Since velocity is in phase with force at this frequency, the average is 0.800 W.

Independently, the damper dissipates $\overline{P} = c\,\overline{\dot{x}^2} = c\omega^2|X|^2/2 = 0.800$ W. The spring returns the energy it stores each cycle and the mass has no net change in kinetic energy. If force and displacement peaks are simply multiplied and reported as watts, the dimensions and the missing phase both reveal the error.

Apply the result to a rotor

A rotating eccentric mass m_e displaced a radius e generates force amplitude $m_e e\omega^2$ when its circular motion is prescribed. With $m_e e = 0.020\,\mathrm{kg\,m}$ and $\omega = 20$ rad/s, this is the 8.0 N used above. At 40

rad/s the applied force grows to 32 N. The dynamic stiffness magnitude becomes $\sqrt{(4000-10(40)^2)^2+(40(40))^2}=12.1$ kN/m, giving about 2.64 mm displacement. The force grew fourfold while the response fell because the system moved away from resonance. A speed sweep therefore needs both the changing excitation and the changing dynamic stiffness.

Estimate damping from a frequency sweep

A useful second identification uses the vibration velocity at constant force amplitude. At resonance the velocity amplitude is F_0/c. At the two frequencies where mean dissipated power has fallen to half its maximum, the velocity amplitude is $F_0/(\sqrt{2}c)$. Inverting the velocity response gives

$$|k/\omega - m\omega| = c.$$

Solving the two sign cases and subtracting gives the bandwidth $\omega_2-\omega_1=c/m$ for this viscous single-mode system. In the numerical example the difference is 4 rad/s; the predicted limits are about 18.10 and 22.10 rad/s. This result applies to the power or velocity response at constant force amplitude. A sweep driven by a growing rotor imbalance would also require correcting the measured amplitudes for their changing excitation.

Decompose a nonsinusoidal periodic machine force

A cam, pump or reciprocating actuator can generate a nearly rectangular force over each cycle. For a linear system in periodic steady operation, expand this force into Fourier components and apply the dynamic stiffness separately at each harmonic. The construction assumes linear coefficients over the motion range and that the free transient has decayed. A periodic input does not imply that the fundamental component dominates the measured displacement; a higher harmonic near a natural frequency can be amplified strongly.

For a symmetric square force $F(t)=F_s\ \mathrm{sgn}(\sin\Omega t)$, the mean vanishes and half-wave symmetry cancels all even harmonics. Its sine coefficient over one period is $b_n=(2/\pi)\int_0^{\pi}F_s\sin(n\theta)d\theta$ for the half-wave-symmetric terms, giving $4F_s/(n\pi)$ for odd n. Each odd harmonic has its own displacement amplitude and phase:

$$X_n=\frac{4F_s/(n\pi)}{\sqrt{[k-m(n\Omega)^2]^2+(cn\Omega)^2}},$$

$$\phi_n=\operatorname{atan2}(cn\Omega,\,k-m(n\Omega)^2).$$

Reconstruct the time history by adding $X_n\sin(n\Omega t-\phi_n)$. Adding the magnitudes gives a peak upper bound, not generally the actual peak because phase shifts differ. Orthogonality gives an exact mean-square contribution $X_n^2/2$ for each distinct harmonic, providing a convenient comparison of vibration levels without finding coincident maxima.

Identify the dominant vibration in an indexing machine

Take $m = 5.0\,\mathrm{kg}$, $k = 4500\,\mathrm{N/m}$, and $c = 9.0\,\mathrm{N\,s/m}$, so $\omega_n = 30\,\mathrm{rad/s}$ and $\zeta = 0.03$. The force alternates between ±12 N at fundamental $\Omega = 10\,\mathrm{rad/s}$. The first-harmonic force is

$$48/\pi = 15.28\,\mathrm{N}.$$

Its dynamic stiffness magnitude is $\sqrt{4000^2 + 90^2} = 4001\,\mathrm{N/m}$, giving $X_1 = 3.82\,\mathrm{mm}$ with a small lag. The third harmonic has force 5.09 N but falls exactly at the natural frequency, giving $X_3 = 5.09/(9 \times 30) = 18.9\,\mathrm{mm}$ at 90° lag.

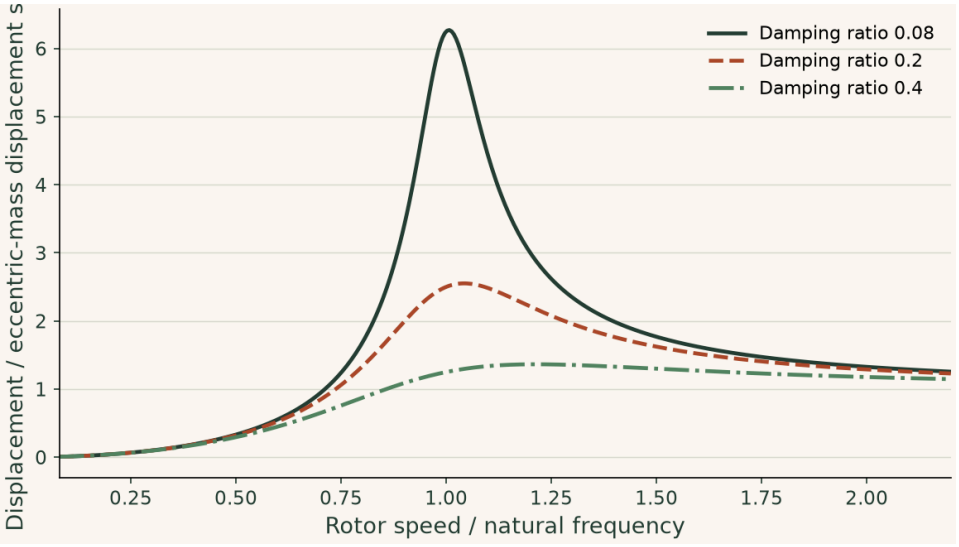
At 50 rad/s the fifth component has force 3.06 N and stiffness magnitude $\sqrt{8000^2 + 450^2}$, giving only 0.381 mm. The first three nonzero harmonics give RMS displacement $\sqrt{(3.82^2 + 18.9^2 + 0.381^2)/2} = 13.6\,\mathrm{mm}$. The largest measured spectral peak would appear at three times the operating frequency. It could arise from the drive's waveform rather than a separate three-times-speed rotating fault. Altering the rise time of the periodic force can reduce high-order Fourier coefficients, while adding damping around this resonance reduces the mechanical amplification. Which intervention addresses the observed peak can be tested by measuring the force spectrum as well as the motion.

The energy entering the damping element gives a further spectral check because high frequencies carry more velocity for the same displacement. The mean dissipated power of harmonic n is $c(n\Omega X_n)^2/2$. In the example the first harmonic dissipates about $9(10 \times 0.00382)^2/2 = 0.0066$ W, while the third dissipates $9(30 \times 0.0189)^2/2 = 1.45$ W. The third harmonic therefore dominates heating even more strongly than it dominates displacement. A sensor attached near a nodal point could suppress a local displacement reading, but the overall energy dissipation still has to be provided by the drive. Synchronous force and velocity measurements over a whole period can independently check this average input power.

A centred force–velocity recording also excludes a false average power component introduced by a sensor's DC offset.

Which speed excites an unbalanced rotor most strongly?

A rotating unbalance excites a single lateral mode with force $m_e e \omega^2$. Divide the harmonic displacement by $m_e e/m$. The resulting factor has r^2 in its numerator because the exciting force grows with speed.



The response tends toward one at high speed and peaks just above the undamped natural frequency for these damping ratios. Treating unbalance as a constant-amplitude shaker force predicts the wrong high-speed asymptote.

▶ **Read values from the plot**

Interactive calculation · change the inputs of this example

Increase damping and compare the resonance neighbourhood with the high-speed asymptote.

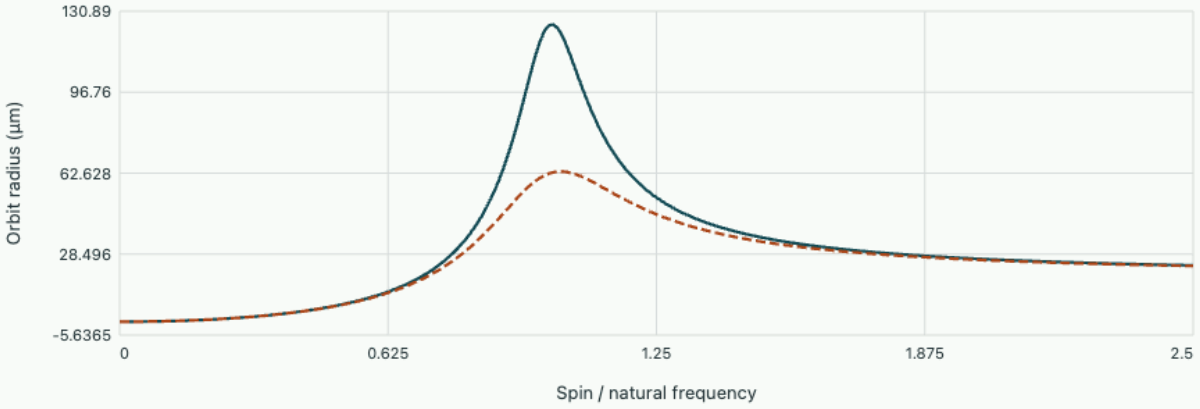
The curves below are recalculated from this section's equations in your browser. The fixed reference plot above remains available for comparison.

Open this live calculation

Try the calculation

Damping ratio **0.08**

Unbalance / rotor mass (µm) **20**



Selected damping Twice damping

At resonance the amplitude is 125.000 µm; its high-speed limit is 20.000 µm because the excitation grows with speed squared.

Coupled oscillators and normal coordinates

Machine dynamics

Derive and interpret the two modes of a coupled machine, then decide which mode an applied load can excite.

FROM FIRST PRINCIPLES

Find natural modes from the free force balance

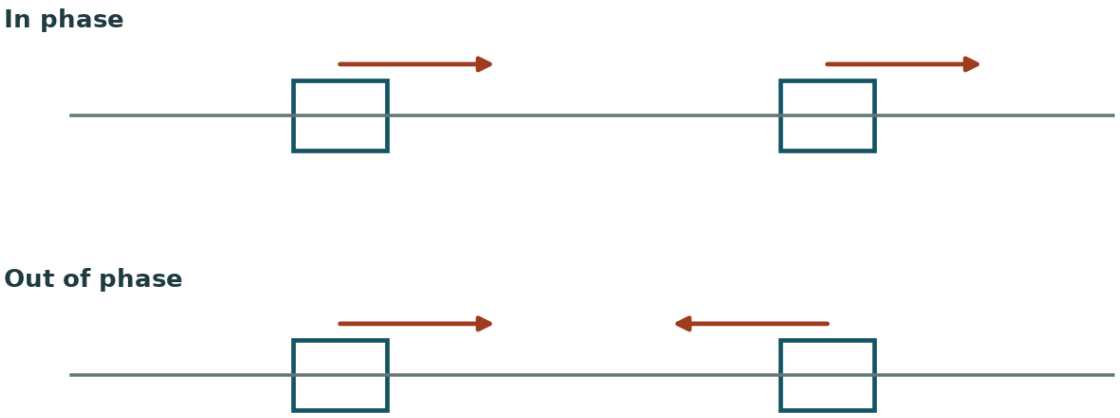
For small motions of a conservative system with several coordinates, linear inertia forces are represented by a positive-definite mass matrix M and linear restoring forces by a symmetric stiffness matrix K. Collect the free balance and try a motion in which the coordinate amplitudes stay in one fixed ratio as each oscillates at frequency ω.

$$M\ddot{\mathbf{x}} + K\mathbf{x} = 0,$$
$$\mathbf{x} = \boldsymbol{\phi}\cos(\omega t) \implies (K - \omega^2 M)\boldsymbol{\phi} = 0.$$

A nonzero shape $\boldsymbol{\phi}$ can solve these homogeneous equations only when the coefficient matrix is singular. The determinant condition gives candidate squared frequencies, and the nullspace at each frequency gives a mode shape up to scale.

$$\det(K - \omega^2 M) = 0,$$
$$K\boldsymbol{\phi}_j = \omega_j^2 M\boldsymbol{\phi}_j.$$

For identical masses linked by a spring, a common translation leaves that link unstretched and an opposite motion stretches it from both ends. This observation identifies common and differential shapes even before evaluating the determinant. Initial displacement and velocity select their amplitudes; the eigenproblem identifies the motions the structure can support.



Which spring deforms in each mode?
Assume equal end springs k, equal masses m and a central coupling spring k^c. A common displacement leaves the coupling spring at its original length. In the lower drawing each mass moves a distance x towards the other, compressing the central spring by 2x. Half a cycle later the same spring extends by 2x. The ratios of restoring force to displacement are k and k + $2k^c$, giving squared frequencies k/m and (k + $2k^c$)/m.

DERIVATION AND APPLICATIONS

Quantify the splitting of two weakly coupled modes

When two machines have the same uncoupled natural frequency, adding a weak connecting spring splits that frequency into two. Consider identical sliders of mass m each attached to ground by stiffness k and linked to each other by stiffness k_c. Neglect damping and choose displacements in the same direction from equilibrium. The coupling spring responds only to the displacement difference. Adding and subtracting the two force balances therefore gives separate equations for $s = (x_1 + x_2)/2$ and $d = (x_1 - x_2)/2$.

$$m\ddot{s} + ks = 0,$$

$$m\ddot{d} + (k + 2k_c)d = 0.$$

Release slider one at A and slider two at zero, both from rest. The two modal coordinates then both start at $A/2$. Write $\omega_s = \sqrt{k/m}$ and $\omega_d = \sqrt{(k + 2k_c)/m}$ and combine their solutions.

$$x_1 = \tfrac{A}{2}(\cos\omega_s t + \cos\omega_d t),$$

$$x_2 = \tfrac{A}{2}(\cos\omega_s t - \cos\omega_d t).$$

Introduce $\bar{\omega} = (\omega_s + \omega_d)/2$ and $\Delta\omega = \omega_d - \omega_s$. The sum-to-product identities separate the fast motion from slow modulation. When the split is small, the modulation resembles a repeated transfer of motion between the sliders.

$$x_1 = A\cos(\bar{\omega}t)\cos(\Delta\omega t/2),$$
$$x_2 = A\sin(\bar{\omega}t)\sin(\Delta\omega t/2).$$

The first minimum of the slider-one displacement envelope occurs at $t = \pi/\Delta\omega$. Its instantaneous velocity there need not vanish because the envelope itself is changing. The envelope timing remains identifiable from repeated peaks at weak coupling.

WORKED EXAMPLE

Time the exchange of motion

For $m = 1.00$ kg, $k = 100$ N/m and $k_c = 10.5$ N/m, the two frequencies are 10.0 and 11.0 rad/s. The first envelope minimum appears at 3.14 s and the first full recovery of its magnitude at 6.28 s. Releasing the first slider by 8.00 mm sets both modal initial amplitudes to 4.00 mm. Releasing both sliders together instead excites only the slower common motion.

Change the inputs and inspect this calculation below

SECOND APPLICATION

Estimate coupling from a beat measurement

A second pair has mass 2.00 kg per slider and measured lower frequency 25.0 rad/s. Consecutive magnitude maxima of one slider's envelope occur 5.00 s apart. Thus

$$\Delta\omega = 2\pi/5.00 = 1.257\,\mathrm{rad/s},$$
$$k_c = m[(25 + 1.257)^2 - 25^2]/2 = 64.4\,\mathrm{N/m}.$$

The independently inferred grounding stiffness is $m\omega_s^2 = 1250$ N/m, making the connector about 5.2 percent as stiff.

Interactive calculation · Change the inputs of this example

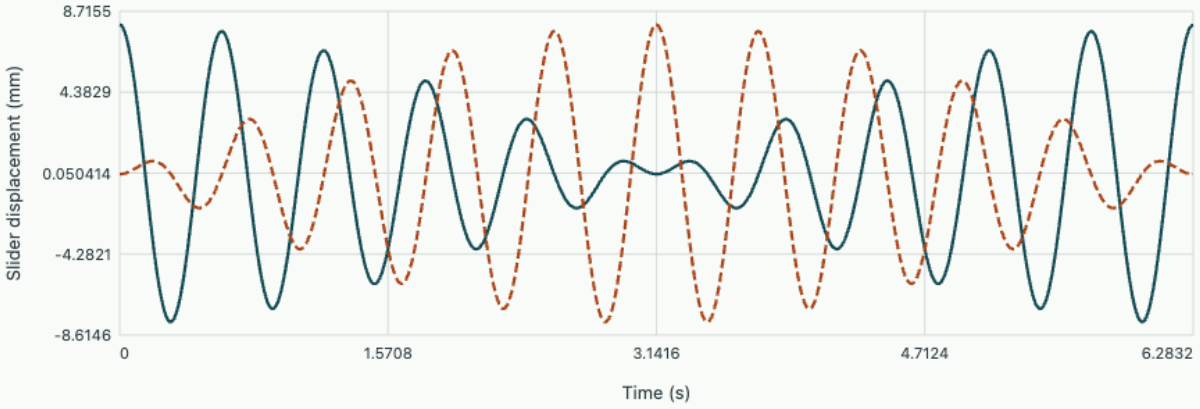
Use the adjustable quantities to test the two worked calculations above. The plot recomputes this derivation's balance or constitutive relation; its numerical readout reports the quantities needed for the forward calculation and the inverse or design question. Keep the fixed assumptions stated in the readout when comparing with the example values.

Open this live calculation

Try the calculation

Coupling stiffness (N/m) **10.5**

Initial release (mm) **8**



Frequencies 10.000 and 11.000 rad/s; first envelope minimum at 3.142 s. Equal masses1 kg; grounded stiffness100 N/m.

Two masses allow two independent shapes

Two identical instruments sit on separate spring mounts and are connected by a flexible link. Pull both by the same amount and the link keeps its natural length. Pull them in opposite directions and the link stretches twice as much as either displacement. The two tests reveal distinct restoring mechanisms. A model with one equivalent spring and the sum of both masses would discard one of them.

Let x_1, x_2 be horizontal displacements from equilibrium. Each mass is m, each ground spring has stiffness k, and the coupling spring has stiffness k_c. With light springs, small displacements and no damping, the kinetic energy is $m(\dot{x}_1^2 + \dot{x}_2^2)/2$ and the elastic energy is $k(x_1^2 + x_2^2)/2 + k_c(x_1 - x_2)^2/2$. The coupling energy contains the difference of displacements because it is the extension between its ends.

Separate space from time

Taking the derivative of elastic energy with respect to each coordinate gives $m\ddot{x}_1 + (k + k_c)x_1 - k_c x_2 = 0$ and the symmetric second equation. Look for $\mathbf{x} = \boldsymbol{\phi}\cos\omega t$. Nonzero modal shapes solve $(\mathbf{K} - \omega^2\mathbf{M})\boldsymbol{\phi} = 0$. For this pair the determinant equation is

$$(k + k_c - m\omega^2)^2 - k_c^2 = 0.$$

Taking both signs of the square root gives

$$\omega_1^2 = k/m,$$
$$\omega_2^2 = (k + 2k_c)/m.$$

Substitution back yields shapes proportional to $(1, 1)$ and $(1, -1)$. For the first, the connector transmits no incremental force. For the second, the connector acts strongly. Reversing both signs of a plotted eigenvector or multiplying it by a nonzero number describes the same physical mode.

Predict a release test at both instruments

Take $m = 5.0$ kg, $k = 2000$ N/m and $k_c = 1500$ N/m. The two angular frequencies are 20.0 and 31.62 rad/s, or 3.18 and 5.03 Hz. Pull only the first instrument 4.0 mm and release from rest. The initial displacement vector equals $2(1, 1) + 2(1, -1)$ mm. Each modal amplitude starts at its corresponding value with zero initial velocity. Hence

$$x_1(t) = 2\cos(20t) + 2\cos(31.62t),$$
$$x_2(t) = 2\cos(20t) - 2\cos(31.62t)$$

in millimetres. At zero time the second displacement vanishes while the connector exerts 6.0 N on its mass. Its initial acceleration is 1.2 m/s² toward the first instrument’s displacement. This apparent transfer of motion needs neither forcing nor damping: modal displacements add and cancel differently at the two observation points.

Distinguish available modes from visible modes

Define average and relative coordinates $q_s = (x_1 + x_2)/2$ and $q_a = (x_1 - x_2)/2$. Adding or subtracting the original equations decouples them. Equal external forces excite only q_s; equal and opposite forces excite only q_a. Taking the signal $(x_1 + x_2)/2$ likewise hides the antisymmetric mode. A missing spectral peak can therefore reveal the symmetry of a load or sensor arrangement, even though the corresponding eigenmode still exists.

If the two masses differ, the simple sum and difference cease to diagonalise the equations. The general route is to form the physical mass and stiffness matrices and solve the symmetric generalised eigenproblem. Distinct modes of an undamped conservative system are orthogonal with respect to

mass, $\boldsymbol{\phi}_i^T \mathbf{M} \boldsymbol{\phi}_j = 0$. This weighted orthogonality is what permits a measured motion or distributed load to be projected into separate modal coordinates.

Predict which mode responds to adding rotary inertia

For a unit mode shape ϕ, the quotient $\omega^2 = \phi^T K \phi / (\phi^T M \phi)$ compares stored elastic energy with inertia. Increasing inertia at a coordinate matters most when that coordinate has appreciable motion in the mode. A frame with three tip coordinates offers a way to check this reasoning beyond the symmetric two-mass illustration.

Compare two changes to the same frame: double only the tip mass from 200 to 400 kg, then restore it and double only the tip rotary inertia from 60 to 120 kg m². Predict whether either change can raise a natural frequency for the same stiffness and compute the fractional shift of the third mode in each case. Use the negligible-member-mass row throughout. Check the frame frequency measurements.

Derive bounce and pitch modes of a supported machine

A long machine supported on two mounts can translate and rotate as a nearly rigid body even if there is negligible frame bending. Let the mass centre move vertically by z and rotate by small angle θ, with positive rotation raising a point to the right. Vertical spring deflections relative to ground are $z - a\theta$ on the left and $z + b\theta$ on the right. Measure displacements from static equilibrium. The kinetic energy is $(m\dot{z}^2 + I_G\dot{\theta}^2)/2$, and the spring energy is $[k_L(z - a\theta)^2 + k_R(z + b\theta)^2]/2$.

Differentiating potential energy gives stiffness terms $K_{zz} = k_L + k_R,\ K_{\theta\theta} = k_L a^2 + k_R b^2,\ K_{z\theta} = -k_L a + k_R b$. The mass matrix is diagonal because translation was measured at the centre of mass. A vertical translation produces a moment whenever the weighted mount locations fail to balance, and a pure rotation then also produces a vertical force by reciprocity. If $k_L a = k_R b$ the modes are pure translation and pure pitch. Equal spring rates alone do not guarantee decoupling when the lever arms differ.

For free motion solve $(K - \omega^2 M)\phi = 0$. In mixed coordinates, metres and radians label different displacement entries, so a numerical component ratio needs units or a stated length scale. Measured vertical motions at two actual sensor locations provide an unambiguous observable mode shape.

Separate a heaving mode from a rocking mode

Let $m = 200$ kg, $I_G = 40\,\mathrm{kg\,m^2}$, $a = b = 0.50$ m and mount rates $k_L = 30000,\ k_R = 50000\,\mathrm{N/m}$. The stiffness entries are 80000 N/m, 20000 N m/rad and 10000 N in the mixed term. Setting the determinant to zero gives $(80000 - 200\lambda)(20000 - 40\lambda) - 10000^2 = 0$ with $\lambda = \omega^2$. Dividing through by 8000 yields

$$\lambda^2 - 900\lambda + 187500 = 0.$$

The squared frequencies are 327.5 and 572.5 s⁻², giving 2.88 and 3.81 Hz.

For the lower mode the first row implies

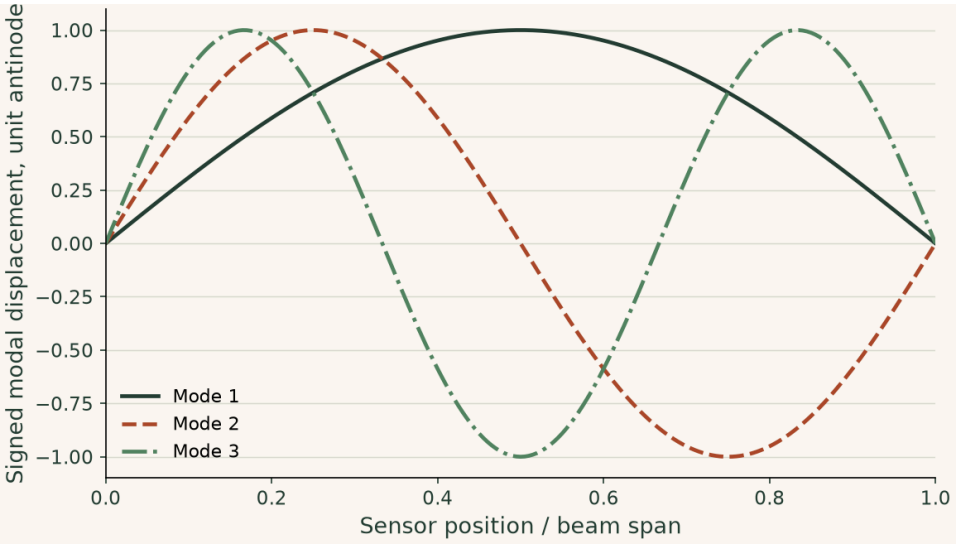
$$\theta/z = -(80000 - 200 \times 327.5)/10000 = -1.45\,\mathrm{rad/m}.$$

If mass-centre motion has 1 mm amplitude, the mount-point displacements are 1.725 and 0.275 mm, in phase but strongly unequal. The higher mode gives $\theta/z = +3.45\,\mathrm{rad/m}$, so the left point now moves oppositely to the right. An accelerometer placed close to a mode's nodal point can scarcely detect it although that mode remains physically present. For the low mode the vertical node lies about $1/1.45 = 0.69$ m to the right of the centre, beyond the right mount. Two well-separated measurement points make the rotational contribution identifiable.

A static load experiment can identify the stiffness coupling before a modal test. Apply a known vertical force F at the centre and no external moment. The pitch-equilibrium equation gives $\theta = -K_{z\theta}z/K_{\theta\theta} = -0.50z$ rad per metre for these mounts. Substitution into the vertical balance leaves effective static stiffness $80000 - 10000^2/20000 = 75000\,\mathrm{N/m}$. Under 150 N centre load, predicted motion is 2.00 mm and pitch −0.0010 rad, giving left and right mount readings of 2.50 and 1.50 mm. A centre gauge alone would only indicate an apparent softening, while the unequal pair exposes the rotational coupling that later appears in the two mode shapes.

Find a sensor position that can see a mode

Use the first three bending modes of an ideal simply supported uniform beam as a continuous analogue of coupled oscillators. A local displacement transducer reports a modal coordinate multiplied by the signed shape at the measurement location.



A midspan sensor misses the second mode even if excitation and damping would produce a large second modal coordinate. At one third span the third mode is invisible. Several sensor locations help distinguish a missing mode from a measurement node.

▶ **Read values from the plot**

Interactive calculation · change the inputs of this example

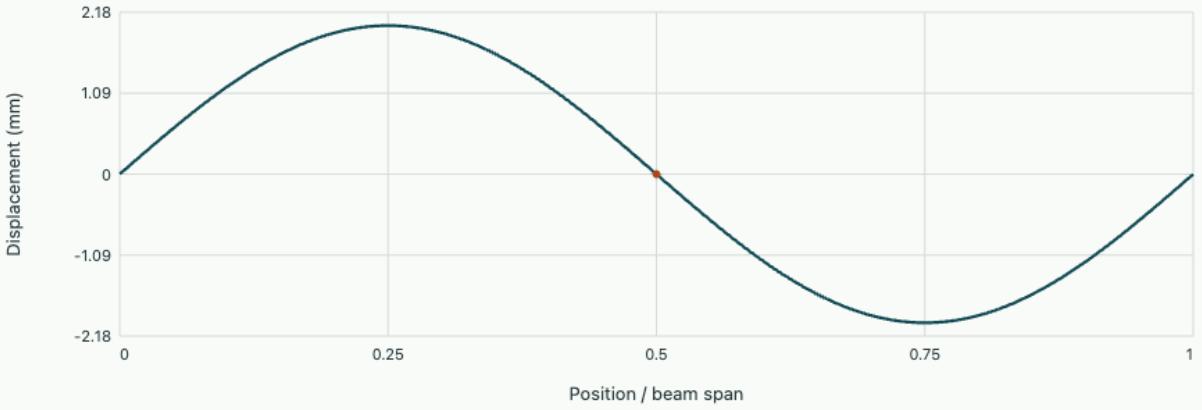
Move the sensor to a node and then change mode number. A zero local reading does not imply a stationary beam.

The curves below are recalculated from this section's equations in your browser. The fixed reference plot above remains available for comparison.

Open this live calculation

Try the calculation

Mode number	**2**
Sensor position / span	**0.5**
Modal amplitude (mm)	**2**



Selected mode Sensor reading

The sensor reads 0.000 mm. Mode 2 has 1 internal displacement nodes. A zero reading can conceal nonzero modal energy.

Energy coordinates and Lagrange's equation

Machine dynamics

Derive a coupled translation–rotation equation with a rolling constraint and identify which reaction disappears from the energy balance.

FROM FIRST PRINCIPLES

Derive the equation of motion using one independent coordinate

An ideal geometric constraint ties several motions together. Choose an independent coordinate q that automatically satisfies it and express kinetic energy $T(q, \dot{q})$ and potential energy $V(q)$ in that coordinate. In the principle of virtual work, ideal constraint reactions do no work along admissible small motions. The balance of inertial, conservative and applied virtual work gives Lagrange's equation

$$\frac{d}{dt}\frac{\partial T}{\partial \dot{q}} - \frac{\partial T}{\partial q} + \frac{\partial V}{\partial q} = Q_{\mathrm{nc}},$$

$$\delta W_{\mathrm{nc}} = Q_{\mathrm{nc}}\,\delta q.$$

For a rigid cylinder rolling on a stationary surface without slip, take centre displacement $q = x$ and angular speed $\dot{x}/R$. Its translation and rotation contribute to kinetic energy; a horizontal spring attached at the centre supplies the potential.

$$T = \tfrac{1}{2}(m + I_G/R^2)\dot{x}^2,$$

$$V = \tfrac{1}{2}kx^2, \qquad (m + I_G/R^2)\ddot{x} + kx = Q_{\mathrm{nc}}.$$

Static friction supplies the contact reaction required for rolling. The contact point has zero instantaneous velocity, so that reaction contributes zero power here. The solution must still demand no more friction than the surface can supply.

DERIVATION AND APPLICATIONS

Derive the effective stiffness of a pendulum on an accelerating cart

A continuously accelerating support changes both the equilibrium direction and the small-oscillation frequency of a hanging load. Model a point mass m on a massless rigid link of length ℓ with an ideal hinge. Let the support move horizontally with imposed coordinate $X(t)$ and measure angle θ from the downward vertical towards positive horizontal displacement. Ignore drag. With mass coordinates $x = X + \ell\sin\theta$ and $y = -\ell\cos\theta$, the kinetic and potential energies are

$$T = \tfrac{1}{2}m[\dot{X}^2 + 2\dot{X}\ell\dot{\theta}\cos\theta + \ell^2\dot{\theta}^2],$$

$$V = -mg\ell\cos\theta.$$

Differentiate $\partial T/\partial\dot\theta$ with respect to time and subtract $\partial T/\partial\theta$. The velocity cross terms involving $\dot X\dot\theta$ cancel, leaving a forcing term set by support acceleration, independent of support speed.

$$\ell\ddot\theta + g\sin\theta + \ddot X\cos\theta = 0.$$

For constant support acceleration a, a stationary angle obeys $\tan\theta_0 = -a/g$. Choose the hanging equilibrium. Expanding the restoring term about θ_0 gives the effective gravitational acceleration along this new direction and the frequency of small oscillations.

$$\theta = \theta_0 + \eta,$$

$$\ddot\eta + \frac{\sqrt{g^2+a^2}}{\ell}\eta = 0,$$

$$\omega_n = \sqrt{\frac{\sqrt{g^2+a^2}}{\ell}}.$$

A positive cart acceleration tips the pendulum backwards. The first correction to angle is linear in a/g, while the first relative correction to frequency is quadratic. The exact equilibrium result holds at finite a; the frequency expression requires small oscillations about it.

WORKED EXAMPLE

Predict a suspended sensor's direction

A cart accelerates at 3.00 m/s² carrying a pendulum of length 0.600 m. Its steady angle is

$$-\arctan(3.00/9.81) = -17.0^\circ.$$

The effective acceleration is 10.26 m/s², giving $\omega_n = 4.13$ rad/s and period 1.52 s. A sensor that treats the pendulum direction as vertical during this manoeuvre would carry the full angular bias.

Change the inputs and inspect this calculation below

SECOND APPLICATION

Infer acceleration from a steady deflection

In another rig a 0.400 m pendulum settles 8.00° backwards after weak damping removes its initial swing. Constant acceleration is

$$9.81\tan 8.00^\circ = 1.38\,\mathrm{m/s^2}.$$

The effective gravity is $9.81/\cos 8.00^\circ = 9.91$ m/s², predicting a small-oscillation period $2\pi\sqrt{0.400/9.91} = 1.26$ s for weak damping. The angle is much more sensitive to modest acceleration than the period.

Interactive calculation · Change the inputs of this example

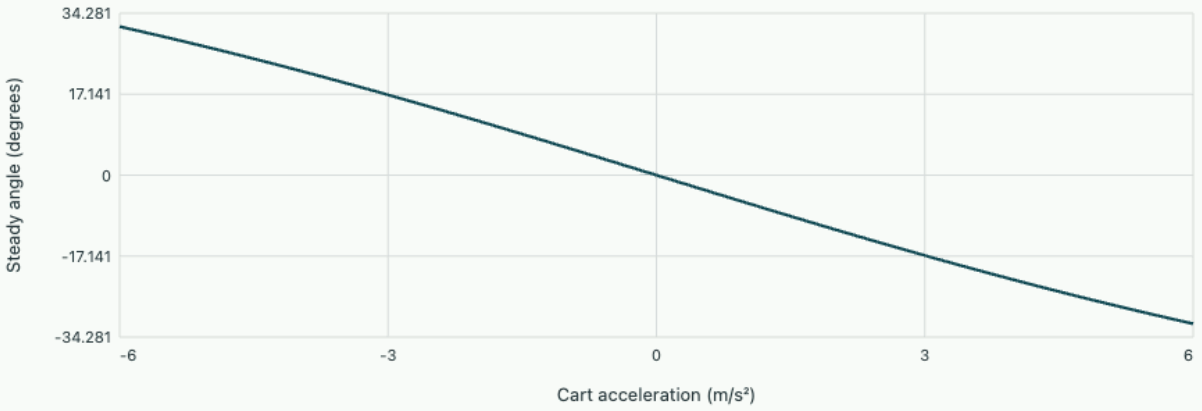
Use the adjustable quantities to test the two worked calculations above. The plot recomputes this derivation's balance or constitutive relation; its numerical readout reports the quantities needed for the forward calculation and the inverse or design question. Keep the fixed assumptions stated in the readout when comparing with the example values.

Open this live calculation

Try the calculation

Cart acceleration (m/s²) **3**

Pendulum length (m) **0.6**



Effective-gravity alignment

Selected equilibrium angle -17.004°; small oscillation frequency 4.135 rad/s.

Let geometry remove a reaction

A cylinder rolls along a horizontal rail while a spring attached at its centre pulls it towards equilibrium. The rail friction is essential for rolling, but an ideal contact point is instantaneously at rest and the friction force does no work. One can write separate translational and rotational equations with an unknown friction force. Alternatively choose a coordinate that satisfies rolling from the beginning. The reduced energy then contains the rotational inertia without introducing a contact reaction.

Take the rightward displacement x of the cylinder centre as independent coordinate and clockwise angle as positive. On the fixed surface the no-slip condition gives $x = R\theta$, up to a reference offset. A body of total mass m and centre inertia I_G has kinetic energy from both translation and rotation. A centre spring with stiffness k has elastic potential $kx^2/2$.

$$T = \tfrac{1}{2}(m + I_G/R^2)\dot{x}^2,$$
$$V = \tfrac{1}{2}kx^2.$$

Connect energy to a dynamical equation

The virtual work of a horizontal force F applied at the centre is $F\delta x$. In this coordinate its generalised force is simply F. A torque M acting clockwise would do $M\delta\theta = (M/R)\delta x$ and would enter as a force M/R. With a centre damper c, the dissipative force is $-c\dot{x}$. Lagrange's equation for this system becomes

$$\frac{d}{dt}\frac{\partial T}{\partial \dot{x}} - \frac{\partial T}{\partial x} + \frac{\partial V}{\partial x} = F - c\dot{x}, \qquad (m + I_G/R^2)\ddot{x} + c\dot{x} + kx = F.$$

For coordinates whose inertia depends on position, retain both derivatives involving T. The differentiation of $\partial T/\partial \dot{x}$ then generates a term proportional to $\dot{x}^2$, part of which cancels $\partial T/\partial x$. Simply multiplying a configuration-dependent effective mass by acceleration would miss the remaining term.

Predict a frequency and check rolling

A solid cylinder has $m = 8.0\,\mathrm{kg}$, $R = 0.10\,\mathrm{m}$, and $I_G = mR^2/2 = 0.040\,\mathrm{kg\,m^2}$. The effective mass seen by the spring is

$$8 + 0.040/0.10^2 = 12.0\,\mathrm{kg}.$$

With $k = 480\,\mathrm{N/m}$ and negligible damping the natural frequency is

$$\sqrt{480/12}/(2\pi) = 1.007\,\mathrm{Hz}.$$

Release it from rest at $x = +0.060\,\mathrm{m}$. Its initial acceleration is

$$-480(0.060)/12 = -2.40\,\mathrm{m/s^2}.$$

Writing only the centre's translation and ignoring rotation would overpredict the acceleration magnitude.

Reconstruct the previously eliminated contact force from translation. The spring pulls leftwards with 28.8 N and the centre accelerates leftwards with $m\ddot{x} = -19.2\,\mathrm{N}$. Friction must therefore point rightwards with 9.6 N to supply the appropriate angular acceleration. The coefficient of static friction must be at least $9.6/(8 \times 9.81) = 0.122$ at release. For a smaller coefficient, the assumed rolling constraint fails and the model needs separate rotational and translational coordinates together with a slipping contact law.

Compare two cylinders with identical mass

A thin-walled hoop of the same mass and rolling radius has $I_G = mR^2$. With the original spring its effective mass doubles the translational mass, giving natural frequency $\sqrt{480/16}/(2\pi) = 0.872\,\mathrm{Hz}$. A no-slip ring-down can therefore distinguish two bodies whose masses and outer dimensions are identical but whose internal mass distributions differ.

Use a cyclic coordinate to reduce coupled motion

Mount a pendulum on a cart that can roll horizontally on a level frictionless guide. Let cart displacement be x, bob angle θ from downward vertical, length l, bob mass m and remaining cart mass M. The bob position is $(x + l\sin\theta, -l\cos\theta)$. Differentiating and squaring its speed gives a cross term $2l\dot{x}\dot{\theta}\cos\theta$ because cart and pendulum velocities need not be perpendicular. The full energy terms are

$$T = \tfrac{1}{2}(M+m)\dot{x}^2 + ml\cos\theta\,\dot{x}\dot{\theta} + \tfrac{1}{2}ml^2\dot{\theta}^2,$$

$$V = mgl(1-\cos\theta).$$

Horizontal position itself does not occur in either expression. It is a cyclic coordinate, so $\partial T/\partial\dot{x} = (M+m)\dot{x} + ml\cos\theta\,\dot{\theta}$ is conserved if no external horizontal force acts. This is horizontal momentum written in generalised coordinates. The pendulum equation from Lagrange's equation is

$$l\ddot{\theta} + \ddot{x}\cos\theta + g\sin\theta = 0.$$

For small swings from rest, differentiate the momentum relation and linearise to obtain

$$\ddot{x} = -ml\ddot{\theta}/(M+m).$$

Substituting gives

$$[lM/(M+m)]\ddot{\theta} + g\theta = 0.$$

Free horizontal motion makes the small-swing frequency larger than for a clamped cart, by $\sqrt{(M+m)/M}$. This result reflects motion about the system's horizontally constrained centre of mass. The extra degree of freedom also permits uniform horizontal translation at zero frequency. Its presence does not imply an additional oscillatory mode.

Predict opposing cart travel in a release test

Let $M = 3.0$ kg, $m = 1.0$ kg and $l = 0.60$ m. Clamp the cart: the oscillation period would be $2\pi\sqrt{0.60/9.81} = 1.55\,\mathrm{s}$. Release its horizontal guide and the period becomes

$$2\pi\sqrt{0.60(3)/(9.81\times 4)} = 1.35\,\mathrm{s}.$$

At release the bob is at $\theta_0 = 10^\circ$ and both bodies are at rest. Because horizontal momentum remains zero, integrate the exact relation to obtain

$$(M+m)x + ml\sin\theta = \text{constant}.$$

As the bob first crosses the vertical, the cart has shifted $ml\sin 10^\circ/(M+m) = 0.0260\,\text{m}$ to the right from its starting point.

At the vertical the potential-energy decrease is

$$mgl(1-\cos 10^\circ) = 0.0894\,\text{J}.$$

Substituting $\dot{x} = -ml\dot{\theta}/(M+m)$ into the kinetic energy at this instant gives effective angular inertia $ml^2M/(M+m) = 0.270\,\text{kg}\,\text{m}^2$. Thus the angular speed magnitude is $\sqrt{2(0.0894)/0.270} = 0.814\,\text{rad/s}$, and the cart speed magnitude is $0.15(0.814) = 0.122\,\text{m/s}$. The frequency estimate used a small-angle approximation; the centre-of-mass and energy results at the vertical used the actual release angle. A horizontal cable pulling on the cart or measurable wheel friction would change the conserved momentum and should be measured before attributing deviations solely to pendulum nonlinearity.

The two-sensor experiment can measure conservation without estimating acceleration. Track the cart position and bob angle from video, then plot $x + [ml/(M+m)]\sin\theta$. For the chosen masses its value should stay constant at the initial offset despite the cart and pendulum both moving. A steadily sloping line indicates nonzero initial horizontal momentum; a curved departure suggests an external force or tracking bias. This separation is useful because rolling resistance can be weak enough to leave the first measured period nearly unchanged while producing visible drift over several cycles. The conserved combination is specific to the inertial laboratory coordinate, so a camera fixed on the moving cart needs a separate record of cart displacement.

Finite-amplitude oscillations and model validity

Machine dynamics

Use conservation of energy and a frequency approximation to quantify when a small-angle pendulum model loses timing accuracy.

FROM FIRST PRINCIPLES

Derive the pendulum period and identify the small-angle assumption

A point mass m suspended on a massless rigid link of length ℓ feels a gravitational torque $-mg\ell\sin\theta$ about the hinge. Assume a stationary pivot and negligible drag. Angular momentum balance gives an exact nonlinear equation. Replacing $\sin\theta$ by θ only for small angles in radians gives a linear oscillator and its predicted period.

$$m\ell^2\ddot{\theta} = -mg\ell\sin\theta,$$

$$T_{\text{small}} = 2\pi\sqrt{\ell/g}.$$

To track finite amplitude, release from rest at angle $A < \pi$ and use mechanical energy conservation. At any intermediate angle, loss of potential energy equals gained rotational kinetic energy.

$$\tfrac{1}{2}m\ell^2\dot{\theta}^2 = mg\ell(\cos\theta - \cos A),$$

$$T = 4\sqrt{\frac{\ell}{2g}}\int_0^A \frac{d\theta}{\sqrt{\cos\theta - \cos A}}.$$

The factor four uses the symmetry of the four quarter cycles. The mass cancels even at large amplitude. The exact restoring torque grows more slowly with deflection than its small-angle approximation, so finite-amplitude swings take longer. Measured variation of period with release angle directly tests the linearisation.

DERIVATION AND APPLICATIONS

Derive the period of an oscillator with purely cubic stiffness

A spring with almost no linear stiffness near equilibrium can have a period set primarily by release amplitude. Consider mass m on a horizontal guide with restoring force $-\beta x^3$ for $\beta > 0$. Ignore damping and assume the fitted force law holds between $-A$ and $+A$. Release from rest at $x = A$. Integrating the force law gives a quartic potential, and conservation of energy provides the speed anywhere on the inward leg.

$$\tfrac{1}{2}m\dot{x}^2 + \tfrac{1}{4}\beta x^4 = \tfrac{1}{4}\beta A^4,$$

$$|\dot{x}| = \sqrt{\frac{\beta}{2m}}\,\sqrt{A^4 - x^4}.$$

By symmetry the time from an extreme position to the centre is one quarter of a full cycle. Write $dt = |dx|/|\dot{x}|$ and substitute $u = x/A$ in the integral. This isolates the amplitude dependence without guessing a sinusoidal shape.

$$T = \frac{4}{A}\sqrt{\frac{2m}{\beta}}\int_0^1 \frac{du}{\sqrt{1-u^4}}.$$

The dimensionless integral evaluates numerically to 1.31103. Its endpoint singularity is integrable because the speed near the turning point vanishes as the square root of the distance remaining. The resulting period and angular frequency are

$$T = \frac{7.4163}{A}\sqrt{\frac{m}{\beta}},$$

$$\omega = \frac{2\pi}{T} = 0.8472\,A\sqrt{\frac{\beta}{m}}.$$

The positive cubic restores the mass for displacements of either sign, but its tangent stiffness at zero is zero. The full nonlinear dynamics therefore predicts a period diverging with decreasing amplitude. Even a weak residual linear spring would eventually dominate in that limit.

WORKED EXAMPLE

Predict the timing of two releases

Take $m = 0.500$ kg, $\beta = 2.00 \times 10^5$ N/m³ and $A = 20.0$ mm. The predicted period is

$$7.4163\sqrt{0.500/(2.00\times10^5)}/0.0200 = 0.586\,\mathrm{s}.$$

Reducing the release to 10.0 mm doubles the predicted period to 1.17 s. At the larger amplitude the centre-crossing speed is

$$A^2\sqrt{\beta/(2m)} = 0.179\,\mathrm{m/s}.$$

Change the inputs and inspect this calculation below

SECOND APPLICATION

Infer stiffness from a measured period

A second fixture of mass 0.800 kg completes a cycle in 0.750 s after release at 15.0 mm. The implied coefficient is $\beta = m[7.4163/(AT)]^2 = 3.48 \times 10^5$ N/m³. Its predicted restoring-force magnitude at release is

$$\beta A^3 = 1.17\,\mathrm{N}.$$

An independent static force measurement there tests the dynamic estimate without assuming a sinusoidal motion.

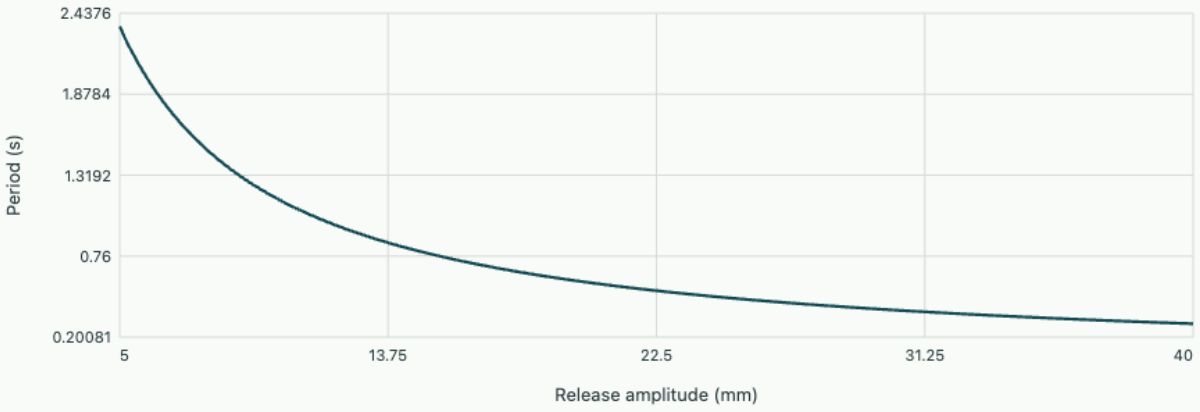
Interactive calculation · Change the inputs of this example

Use the adjustable quantities to test the two worked calculations above. The plot recomputes this derivation's balance or constitutive relation; its numerical readout reports the quantities needed for the forward calculation and the inverse or design question. Keep the fixed assumptions stated in the readout when comparing with the example values.

Open this live calculation

Try the calculation

Mass (kg)	0.5
Cubic force coefficient (kN/m³)	200



Energy-integral period

At20 mm amplitude, period=0.586 s. This pure cubic spring has no nonzero small-amplitude linear frequency.

Retain the actual restoring moment

A pendulum gives a practical test of a linear model because its length and release angle are easy to set while its period can be measured over many cycles. Approximate the moving body first by a point mass m at distance l from a frictionless pivot. With θ measured from vertically downwards, gravity produces torque $-mgl\sin\theta$. Dividing the rotational balance by ml^2 leaves $\ddot{\theta}+(g/l)\sin\theta=0$. The linear small-angle frequency $\omega_0=\sqrt{g/l}$ follows only after replacing $\sin\theta$ by θ in radians.

The leading correction has a definite sign:

$$\sin\theta \simeq \theta-\theta^3/6.$$

At a large positive angle the restoring force magnitude is smaller than the linear extrapolation. The pendulum should take longer to complete its swing. A sign test like this is useful before calculating a period correction.

Derive an integral valid at finite amplitude

For release from rest at amplitude $A<\pi$, conservation of mechanical energy gives

$$\tfrac{1}{2}ml^2\dot{\theta}^2+mgl(1-\cos\theta)=mgl(1-\cos A).$$

During one quarter of the orbit the angle travels between zero and A. Substitute $\sin(\theta/2)=\sin(A/2)\sin\psi$ into the time integral. Cancelling common square roots gives

$$T=4\sqrt{l/g}\int_0^{\pi/2}\frac{d\psi}{\sqrt{1-k^2\sin^2\psi}},$$

$$k=\sin(A/2).$$

For modest amplitudes expand the integrand as $1+\tfrac{1}{2}k^2\sin^2\psi+\cdots$. The average of $\sin^2\psi$ over the quarter interval is one half. Therefore $T/T_0\simeq 1+k^2/4\simeq 1+A^2/16$ through second order in amplitude. The oscillation period depends on energy, while in the linear spring model all amplitudes share the same period.

Estimate a measurable timing error

Take $l=0.80\,\mathrm{m}$, $g=9.81\,\mathrm{m/s^2}$ and release at $A=30^\circ=0.5236\,\mathrm{rad}$. The small-angle period is

$$T_0=2\pi\sqrt{0.80/9.81}=1.794\,\mathrm{s}.$$

The leading correction is 1.71%, giving about 1.825 s. Over ten full oscillations the predicted delay relative to the small-angle model is roughly 0.31 s. Reducing the release angle to 15° reduces the

leading relative correction by a factor of four.

A timing series from a damped experimental pendulum may show the period drifting towards T_0 as its amplitude decays. Measuring only the first release angle while averaging many later periods mixes different amplitudes. To test the finite-amplitude law, use consecutive same-direction crossings to measure each local period and estimate its associated amplitude from neighbouring turning points. The correction above is an expansion around zero amplitude. As A approaches the inverted equilibrium, the exact integral grows without bound and a low-order polynomial is no longer an adequate approximation.

Check the scale of the neglected term

Retaining one further term in the integrand gives $1 + \frac{1}{2}k^2 \sin^2\psi + \frac{3}{8}k^4 \sin^4\psi$. The quarter-interval average of the fourth power of sine is $3/8$, so $T/T_0 \simeq 1 + k^2/4 + 9k^4/64$. At 30° this gives 1.01738, while the leading expansion in amplitude gave 1.01713. The small difference is measurable only if the uncertainty in pendulum length, release angle and crossing times supports the comparison.

Derive an amplitude-dependent resonance relation

A pendulum softens as its amplitude grows because the sine restoring law rises more slowly than its linear tangent. A mechanical flexure can instead harden through geometry. A simple modal model is $m\ddot{x} + c\dot{x} + kx + \beta x^3 = F\cos\omega t$. For positive β the restoring stiffness increases with excursion. Assume one dominant harmonic and take $x \approx A\cos(\omega t - \phi)$. Expanding the cubic gives

$$x^3 = A^3[3\cos(\omega t - \phi) + \cos(3\omega t - 3\phi)]/4.$$

Keeping only the fundamental turns the cubic term into an amplitude-dependent in-phase force.

Balance in-phase and quadrature components against the input, square and add to eliminate ϕ. The first-harmonic amplitude estimate satisfies

$$F^2 = A^2\left[\left(k + \tfrac{3}{4}\beta A^2 - m\omega^2\right)^2 + (c\omega)^2\right].$$

In a lightly damped free motion the same approximation gives a backbone frequency $\omega_b^2 \approx [k + (3/4)\beta A^2]/m$. This is an amplitude-dependent estimate of conservative oscillation frequency, not a new eigenvalue of the linearised equilibrium. At very small amplitude it recovers $\sqrt{k/m}$, while a negative cubic coefficient tilts the relation towards lower frequencies. If the cubic approximation is only a local expansion, do not extrapolate it until it predicts an unphysical loss of global stiffness.

Identify a hardening spring from two release amplitudes

A test component has moving mass 1.0 kg and small-amplitude stiffness 400 N/m. Low-level ring-down gives $\omega_0 = 20\,\mathrm{rad/s}$. At measured oscillation amplitude 20 mm the dominant free frequency is approximately 22 rad/s. Under the weakly nonlinear one-harmonic model, $\beta \simeq 4m(22^2 - 20^2)/[3(0.020)^2] = 2.80 \times 10^5\,\mathrm{N/m^3}$. At 10 mm the inferred backbone is $\sqrt{400 + 0.75(2.8 \times 10^5)(0.010)^2} = 20.52\,\mathrm{rad/s}$, a direct third reading against which to test the fit.

Now drive near 22 rad/s with viscous coefficient $c = 0.80\,\mathrm{N\,s/m}$. At the 20 mm backbone amplitude the spring and inertia fundamentals cancel, so maintaining this motion takes forcing $F = c\omega A = 0.352\,\mathrm{N}$ in quadrature with displacement. The neglected cubic third-harmonic force amplitude is

$$\beta A^3/4 = 0.56\,\mathrm{N};$$

its displacement effect can nevertheless be small because the oscillator is driven far from its fundamental resonance at 66 rad/s. A separate mode near that frequency would invalidate this simplification. In a forced frequency sweep the amplitude equation can admit several positive roots near resonance. Its algebra does not establish their stability; slow upward and downward sweeps, each with sufficient settling time, can reveal different stable responses and jump transitions.

An initial static measurement tests the sign of the fitted coefficient independently. At a displacement of 20 mm the linear restoring force is

$$400(0.020) = 8.0\,\mathrm{N};$$

the cubic contribution is

$$2.8 \times 10^5 (0.020)^3 = 2.24\,\mathrm{N}.$$

The predicted static force is therefore 10.24 N, and the tangent stiffness is

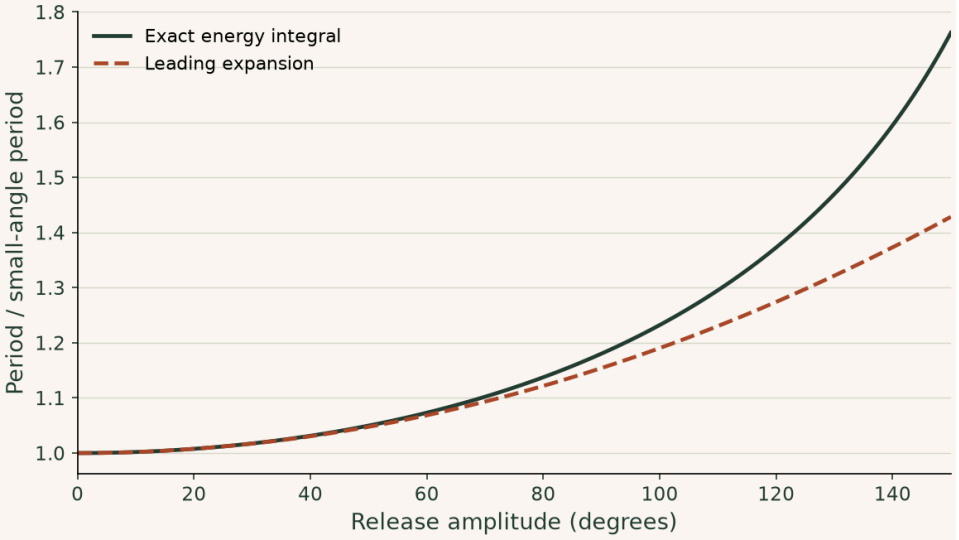
$$400 + 3(2.8 \times 10^5)(0.020)^2 = 736\,\mathrm{N/m}.$$

This tangent stiffness applies to small vibrations around a held biased equilibrium. Using it directly for an oscillation centred at zero would overestimate the measured amplitude-dependent frequency because the instantaneous stiffness changes during each cycle. Static bias and free oscillation amplitude are physically different tests of the same nonlinear restoring law.

Repeating a decay test around a deliberately shifted equilibrium distinguishes the two experimentally.

Set an amplitude limit for a pendulum-based timing model

For a rigid pendulum released from rest the exact period ratio is $2K(\sin^2(\theta_0/2))/\pi$, where the complete elliptic integral here uses its parameter rather than its modulus. Compare it with the first amplitude correction $1 + \theta_0^2/16$, with angle in radians.



At 90° the exact period is about 1.180 times the linear prediction; the leading expansion gives 1.154. The series is useful at modest amplitudes, but it cannot reproduce the divergence on approach to an upright release.

▶ **Read values from the plot**

Interactive calculation · change the inputs of this example

Increase the maximum release angle and compare the energy integral with the first series correction. Changing length rescales both periods.

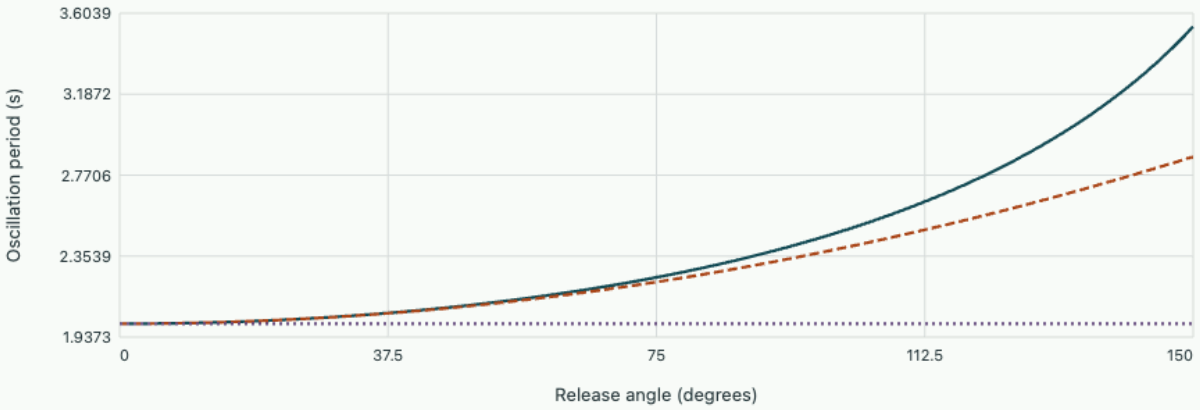
The curves below are recalculated from this section's equations in your browser. The fixed reference plot above remains available for comparison.

Open this live calculation

Try the calculation

Maximum release angle (degrees) **150**

Pendulum length (m) **1**



Energy integral via AGM · First amplitude correction · Small-angle limit

At 150° the exact period is 3.535 s, 76.22% above the linear estimate.

A moving-coil galvanometer as an electrical and mechanical system

Electromechanics, optics and measurement

A scale division records an angle, yet the quantity marked on the dial is electric current. A spring, a magnetic field and the resistance of the external measuring circuit connect the two. Derive the calibration, predict how a pointer approaches its reading, and find which experiment measures the charge of a short pulse.

Convert the force on two straight sides into a torque

A light rectangular coil has N turns, active length l, width w, and area per turn $A = lw$. It pivots about an axis parallel to its long sides. Shaped pole pieces confine an approximately radial magnetic field of magnitude B to an air gap. Over the working angular range the magnetic field is perpendicular to the active wire and lies approximately in the plane of the coil. An elementary length carrying conventional current i experiences $d\mathbf{F} = i\,d\mathbf{l} \times \mathbf{B}$. The two long sides carry current in opposite directions and experience forces of equal magnitude $NilB$. Acting with lever arms $w/2$, the forces make a couple whose torque is $NiAB$. Their net force is approximately zero. The short ends supply negligible torque about the pivot under the ideal geometry.

Write $K = NAB$, and choose the positive sense of current to give positive torque. Its units are $\mathrm{N\,m/A}$. In a uniform Cartesian field an ordinary rigid loop instead gives torque $NiAB\sin\alpha$, where α is the angle between its normal and the field. The radial gap maintains the factor equivalent to $\sin\alpha$ close to unity as the pointer turns. This is a design choice that produces a nearly linear scale. If a measured calibration curve flattens at a large angle, field nonuniformity and a nonlinear torsion spring are candidate causes that can be tested separately.

For a soft torsion spring with stiffness κ, the return torque is $-\kappa\theta$ at small deflections from the zero-current setting. At static equilibrium $\theta = Ki/\kappa$. With $N = 200$, $A = 1.00\,\mathrm{cm^2}$, $B = 0.250\,\mathrm{T}$ and $\kappa = 8.0 \times 10^{-4}\,\mathrm{N\,m/rad}$, the torque constant is $0.0050\,\mathrm{N\,m/A}$ and the current sensitivity is $6.25\,\mathrm{rad/A}$. A $5.0\,\mathrm{mA}$ current turns the pointer through $0.03125\,\mathrm{rad}$, about 1.79°. A mirror mounted on the same axle deflects a reflected laser through twice this angle. At a screen distance of $0.50\,\mathrm{m}$ the spot shifts $0.50\tan(2\theta) = 31.3\,\mathrm{mm}$, making the small mechanical rotation easy to resolve.

Recover the reciprocal voltage from work

Now rotate the coil through a small positive angle while it carries positive current. The electromagnetic torque does mechanical work $Ki\,d\theta$. The voltage associated with the motion must therefore draw electrical work $ie_{\mathrm{back}}\,dt$ of equal magnitude when the motor torque assists that motion. This gives the back-emf relation $e_{\mathrm{back}} = K\dot\theta$, measured as a positive drop in the motor sign convention. Its coefficient has units $\mathrm{V\,s/rad}$, numerically the same SI coefficient as the torque constant. One can also derive it from the motional emf on the two active sides. Each moves at speed $(w/2)\dot\theta$, crosses field B, and generates a voltage $Bl(w/2)\dot\theta$ along its wire. The two contributions add over every turn.

Let J denote the rotational inertia, b the mechanical viscous damping coefficient, L the coil inductance and R the sum of winding plus external series resistance seen by a voltage source $v(t)$. The electromechanical equations are

$$J\ddot{\theta} + b\dot{\theta} + \kappa\theta = Ki,$$

$$L\dot{i} + Ri + K\dot{\theta} = v(t).$$

Multiplying the first equation by $\dot{\theta}$ and the second by i eliminates the coupling power $Ki\dot{\theta}$. The result says that source power vi equals the rate of increase of $J\dot{\theta}^2/2 + \kappa\theta^2/2 + Li^2/2$, plus $b\dot{\theta}^2 + Ri^2$. This identity fixes the relative signs of the torque and back emf. If a sign mistake turns a passive closed coil into a self-accelerating pointer, the same mistake produces a residual coupling term in this power balance. Mechanical friction and winding heat represent distinct sinks even though both can damp the observed displacement.

Predict overshoot and find a critical external resistance

If the electrical relaxation time L/R is short compared with the oscillation time $1/\sqrt{\kappa/J}$, set $i \simeq (v - K\dot{\theta})/R$. Substituting into the mechanical equation increases its effective damping to $b_{\mathrm{eff}} = b + K^2/R$. Set $v = 0$ when discussing free relaxation with the electrical terminals connected through a resistor. A moving shorted coil generates a current that opposes its motion. Opening the circuit instead suppresses that current and leaves the mechanical damping. An ideal current source holding specified current through all pointer motions also changes the back-emf response because its output voltage adjusts to enforce that current.

The undamped angular frequency is $\omega_0 = \sqrt{\kappa/J}$ and the damping ratio is

$$\zeta = b_{\mathrm{eff}}/(2\sqrt{\kappa J}).$$

Take $J = 2.0 \times 10^{-6}\,\mathrm{kg\,m^2}$, $b = 1.6 \times 10^{-5}\,\mathrm{N\,m\,s/rad}$, and the preceding K, κ. Then $\omega_0 = 20.0\,\mathrm{rad/s}$, $2\sqrt{\kappa J} = 8.0 \times 10^{-5}\,\mathrm{N\,m\,s/rad}$, and an open circuit gives $\zeta = 0.20$. With $R = 1.00\,\Omega$ the electric contribution to damping is $2.50 \times 10^{-5}\,\mathrm{N\,m\,s/rad}$, giving $\zeta = 0.5125$. For a voltage step the ideal fractional overshoot in the reduced model is

$$\exp[-\pi\zeta/\sqrt{1-\zeta^2}] = 0.153.$$

The first peak occurs after $\pi/(\omega_0\sqrt{1-\zeta^2}) = 0.183\,\mathrm{s}$.

Critical damping occurs when $b + K^2/R = 2\sqrt{\kappa J}$, which requires a total resistance of about $0.391\,\Omega$. A winding resistance of $0.300\,\Omega$ therefore leaves $0.091\,\Omega$ for external source and leads. If the winding alone exceeds the target resistance, a simple passive external series resistor cannot reach this damping in the reduced model. A finite L must also be checked: with $L = 5.0\,\mathrm{mH}$ and $R = 1.0\,\Omega$, $\omega_0 L/R = 0.10$, so the approximation provides a useful starting estimate. Reducing R indefinitely eventually invalidates the elimination of current. The full pair of equations then predicts the damping and frequency dependence.

Distinguish a charge impulse from a slow current reading

Suppose a brief imposed current pulse ends long before the pointer has turned appreciably. Integrating the torque equation over that interval gives

$$J\Delta\dot{\theta} \simeq K\int i\,dt = Kq.$$

The initial signal is now angular velocity proportional to total transmitted charge. For weak subsequent damping, the first excursion is approximately $Kq/(J\omega_0)$. A dose of $20\,\mu\mathrm{C}$ gives $2.50\,\mathrm{mrad}$ with the instrument above, corresponding to about $2.50\,\mathrm{mm}$ on the mirror scale. Two current pulses of identical charge but different shapes give nearly the same reading if both are short on the mechanical time scale. Their peak currents can be entirely different.

For appreciable damping, use the impulse solution $\theta(t) = Kq\,e^{-\zeta\omega_0 t}\sin(\omega_d t)/(J\omega_d)$, with $\omega_d = \omega_0\sqrt{1-\zeta^2}$ for an underdamped pointer. Its maximum obeys $\tan(\omega_d t_p) = \omega_d/(\zeta\omega_0)$ in the first quarter cycle. Substitution corrects the ballistic calibration instead of treating the damped peak as an undamped amplitude. The state of the external circuit after the pulse matters because it determines ζ during the throw. A calibration pulse should leave the measuring circuit connected in the same way as the unknown pulse.

Measure the torque constant without assuming the magnetic field

The preceding design calculation multiplies turn count, active area and a nominal air-gap induction. An independent calibration can measure the same constant without knowing the field distribution along every wire. Move the pointer through a small controlled angle while measuring its open-circuit voltage with a high-impedance instrument. In the same working interval the voltage is $K(\theta)\dot{\theta}$ up to a polarity choice. Integrating the voltage over the sweep gives $\int_{\theta_1}^{\theta_2} K(\theta)\,d\theta$. If K is constant this is simply $K\Delta\theta$, independent of sweep speed. It is often easier to integrate a recorded voltage pulse accurately than to differentiate the angle trace to obtain instantaneous speed.

For a sweep through $0.0800\,\mathrm{rad}$, suppose the measured voltage integral is $0.401\,\mathrm{mV\,s}$. The generator constant is $5.01\times10^{-3}\,\mathrm{V\,s/rad}$. Independently add a known small torque to the spring and measure its static angle. A torque $32.0\,\mu\mathrm{N\,m}$ rotates the pointer by $0.0400\,\mathrm{rad}$, giving $\kappa = 8.00\times10^{-4}\,\mathrm{N\,m/rad}$. Combining the two measurements predicts a current sensitivity of $6.27\,\mathrm{rad/A}$. Splitting the sweep into smaller equal angular intervals tests whether the measured K depends on angle. Sweeping backwards should reverse the integrated voltage. A voltage offset instead contributes its offset times duration, so repeating at two sweep speeds helps separate it.

A dynamic cross-check uses free decay under two load conditions. Measure the exponential decay rate α_{open} with a high-impedance voltage probe and α_R with a known total resistance in the closed circuit. In the short-electrical-time limit, subtraction of the two damping laws yields

$$K^2 = 2JR(\alpha_R - \alpha_{\mathrm{open}}).$$

With $J = 2.0 \times 10^{-6}\,\mathrm{kg\,m^2}$, $R = 1.00\,\Omega$, and decay rates 4.0 and $10.25\,\mathrm{s^{-1}}$, this gives

$$K = 0.0050\,\mathrm{N\,m/A}.$$

The method needs two identifiable decay rates and independent inertia. If the loaded response is overdamped, the envelope of two oscillation peaks is no longer available; fit the full two-real-pole response to the measured displacement.

Keep the third dynamical variable in a frequency-response experiment

For a sinusoidal voltage at angular frequency ω the exact linear transduction law follows by eliminating the complex current after transforming both differential equations. The mechanical impedance in angular-displacement form is $\kappa - J\omega^2 + i\omega b$; the winding impedance is $R + i\omega L$. Using phasors with time dependence $e^{i\omega t}$ gives

$$\frac{\Theta}{V} = \frac{K}{(R + i\omega L)(\kappa - J\omega^2 + i\omega b) + i\omega K^2}.$$

At zero frequency the gain is $K/(R\kappa)$ and the phase tends to zero, as the static calibration requires. At a sufficiently high frequency both the electrical and inertial impedances grow. For $L > 0$ the magnitude approaches $K/(LJ\omega^3)$. The reduced model with the current eliminated algebraically falls only as $1/\omega^2$. The change of the asymptotic slope provides an experimental way to distinguish finite inductance from an incorrectly adjusted viscous damping constant. Beyond the working band of a real instrument, elastic modes of the pointer or parasitic capacitance can enter before that asymptote is resolved.

At the nominal undamped frequency $20\,\mathrm{rad/s}$, with $R = 1.0\,\Omega$ and $L = 5.0\,\mathrm{mH}$, the stiffness and inertia terms cancel. The denominator becomes $-\omega^2 L b + i\omega(Rb + K^2)$, or approximately $-0.000032 + i\,0.000820$ in consistent SI units. Its magnitude gives about $6.09\,\mathrm{rad/V}$. A sinusoid of $1.0\,\mathrm{mV}$ peak produces about $6.09\,\mathrm{mrad}$ peak, whereas the static response to $1.0\,\mathrm{mV}$ is $6.25\,\mathrm{mrad}$. The phases differ radically despite their similar amplitudes: near this frequency the pointer lags by about 92.2°. A gain-only calibration at a single frequency would miss the phase lag.

Change the electrical load and compare two step models

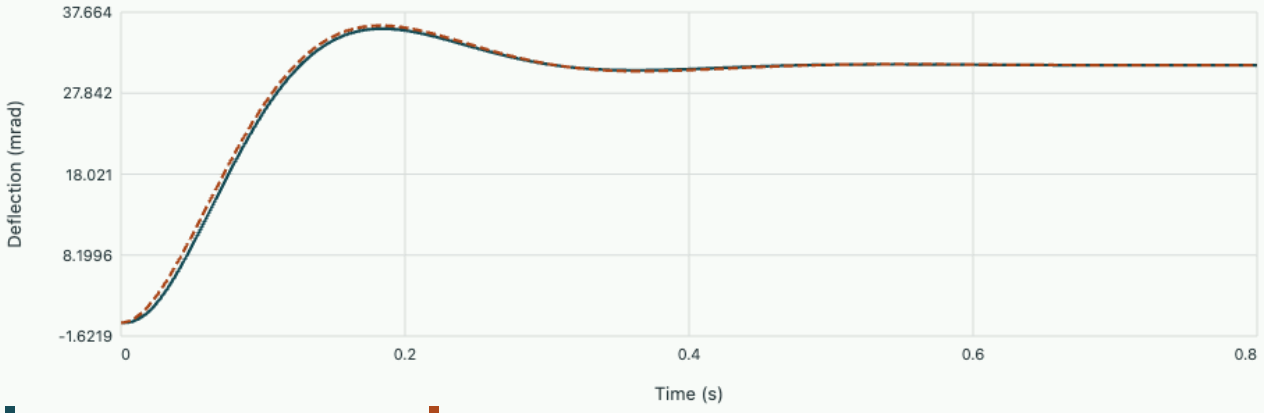
In the live graph the source voltage is chosen anew at each resistance to give a final current of $5.0\,\mathrm{mA}$ after the pointer settles. That convention keeps the final angle common to all curves and isolates the transient response of the load. The full model retains $i, \theta, \dot\theta$; the second curve sets $L\dot i = 0$. Predict first which trace should start flatter. With finite inductance the current initially vanishes and increases at a finite rate. The initial torque and angular acceleration therefore vanish, while the reduced model applies a finite torque immediately. Increasing the inductance makes that distinction easier to see, especially when the electrical response time becomes comparable to the period of mechanical oscillation.

Open the live experiment

Try the calculation

Total resistance (Ω) 1

Winding inductance (mH) 5



Coupled winding and pointer | Short electrical time approximation

The reduced damping ratio is 0.512. The largest deflection in the first 0.80 s is 35.67 mrad. The common static limit is 31.25 mrad.

How could a student change the static range of the instrument without changing its suspension or its torsional stiffness?

▼ **Work through the solution**

For a dc ammeter, connect a calibrated shunt resistance in parallel with the coil, allowing only a known fraction of total current into it. With a coil winding resistance $0.300\,\Omega$ and a $0.0333\,\Omega$ shunt, the coil carries about ten percent of a steady terminal current and the indicated full-scale terminal current increases about tenfold. The total circuit seen by a motional voltage during a pointer oscillation depends on the shunt and source impedance, so the static range multiplier does not by itself give the new damping. For a voltmeter, a series resistance instead sets the current produced by an applied terminal voltage. In both cases record the resistance network whenever reporting a dynamic response.

Estimate the back action on the signal being measured

A galvanometer becomes part of the circuit it observes. Suppose a source with open-circuit voltage $10.0\,\mathrm{mV}$ and internal resistance $2.00\,\Omega$ drives a coil of resistance $0.300\,\Omega$. After settling, the current is $4.35\,\mathrm{mA}$ and the pointer turns by $27.2\,\mathrm{mrad}$. The undisturbed short-circuit current of the source would have been $5.00\,\mathrm{mA}$. Interpreting the indicated value as the current that would flow with a perfect zero-resistance meter underestimates that quantity by about thirteen percent. Measuring the open-circuit voltage instead requires a sufficiently large input resistance, another electrical boundary condition.

During pointer motion the disturbance includes the back emf as well as winding resistance. If the same coil is briefly turning at $0.40\,\mathrm{rad/s}$ it develops a $2.0\,\mathrm{mV}$ back emf. Neglecting inductance for that estimate, the resulting current is

$$(10.0 - 2.0)\,\mathrm{mV}/2.30\,\Omega = 3.48\,\mathrm{mA}.$$

The deflection therefore contains information about both the source and the pointer's earlier motion. A static calibration alone cannot recover a rapid source waveform. A known source resistance and the two coupled dynamic equations provide a way to estimate the waveform from synchronised angle and current measurements.

Tutorial on calibration and transient identification

A current source produces a static mirror shift of $12.5\,\mathrm{mm}$ on a screen $0.50\,\mathrm{m}$ away. Estimate the current and identify a reversible test for an angular zero offset.

▼ **Work through the solution**

Use $\theta = \tfrac{1}{2}\arctan(0.0125/0.50) = 0.01250\,\mathrm{rad}$ and $i = \kappa\theta/K = 2.00\,\mathrm{mA}$. Repeat with reversed current of equal magnitude. Half the difference of the two angles cancels a fixed angular zero offset and isolates the response odd in current. A changing temperature during reversal can alter the spring or winding resistance, so alternate the order and record the zero afterwards if the required uncertainty is small.

For an open-circuit free decay the ratio of two successive same-side positive peak angles is 3.60. Their time separation is $0.321\,\mathrm{s}$. Recover the mechanical parameters that can be inferred if J is known.

▼ **Work through the solution**

The logarithmic decrement is $\ln 3.60 = 1.281$. The decay rate is $\alpha = 1.281/0.321 = 3.99\,\mathrm{s}^{-1}$, and the damped angular frequency is

$$2\pi/0.321 = 19.57\,\mathrm{rad/s}.$$

Then $\omega_0 = \sqrt{\omega_d^2 + \alpha^2} = 19.97\,\mathrm{rad/s}$,

$$b = 2J\alpha = 1.60 \times 10^{-5}\,\mathrm{N\,m\,s/rad},$$

$$\kappa = J\omega_0^2 = 7.98 \times 10^{-4}\,\mathrm{N\,m/rad}.$$

This identification assumes viscous damping and a sufficiently high impedance to make the electrical damping negligible.

Explore how motion generates voltage in the induction chapter.

Coupled optical and mechanical resonances

Electromechanics, optics and measurement

A vibrating mirror can imprint its position on an optical field, while the stored light pushes back on the mirror. An apparent pair of peaks can indicate two hybrid modes, interference at a readout port or two unrelated resonances. Starting from energy in a moving cavity, derive the coupled response and learn which measurements establish a coupling strength and its uncertainty.

From a moving boundary to a force per photon

Take a nearly empty Fabry–Pérot cavity of optical length L with one moving end. A longitudinal resonance has $\omega_c = m\pi c/L$ for integer m, neglecting frequency-dependent mirror phase. For a small extension x, $\delta\omega_c = (\partial\omega_c/\partial x)x$ with signed slope $G_x = -\omega_c/L$. At fixed photon number n the optical energy is $n\hbar\omega_c(x)$. Its derivative gives the radiation-pressure force on the coordinate. For this mirror $F = -n\hbar G_x > 0$ pushes outward. The derivative must be at fixed photon number. Differentiating a driven steady-state photon number as though it were a closed-system energy would incorrectly include work exchanged with the pump.

For a 10 mm cavity resonant near 1064 nm the frequency slope magnitude in ordinary frequency is about 28.2 MHz per nanometre. A static shift of 10 pm therefore changes the optical resonance by approximately 282 kHz. One photon exerts a mean force $\hbar\omega_c/L \simeq 1.87 \times 10^{-17}$ N. An equivalent classical derivation counts a momentum change twice per normal reflection for each relevant round trip. The two descriptions yield the same mean force. The mirror displacement is the properly normalised mode coordinate; in a flexural device the optically sampled position and the total physical mass are generally different from centre-of-mass displacement and mass.

Introduce a single mechanical eigenmode with effective mass m_{eff}, angular frequency Ω_m and displacement scale $x_{\mathrm{zpf}} = \sqrt{\hbar/(2m_{\mathrm{eff}}\Omega_m)}$. The single-photon coupling magnitude is $g_0 = |G_x|x_{\mathrm{zpf}}$. This normalisation is convenient even when the observed vibrations are entirely classical. For $m_{\mathrm{eff}} = 10$ ng and $\Omega_m/(2\pi) = 1.0$ MHz, $x_{\mathrm{zpf}} = 9.16 \times 10^{-16}$ m and $g_0/(2\pi) = 25.8$ Hz for the stated optical geometry. The small displacement scale explains why a coherent pump is normally used to make the interaction measurable.

How a laser couples frequencies that are otherwise far apart

An optical carrier oscillates at hundreds of terahertz, whereas the mirror may vibrate at only megahertz. Direct comparison of these two bare frequencies would never suggest a resonant exchange. A strong coherent pump supplies a phase reference and converts a mechanical displacement into optical sidebands. Move into a frame rotating at the pump frequency and write the cavity amplitude as its large steady part plus a small perturbation. The leading frequency-shift interaction is proportional to photon number times displacement. Expanding the intensity around the large pump keeps terms linear in both the probe field and displacement. The resulting coupling is $G = g_0\sqrt{n_c}$, where n_c is mean intracavity pump photon number.

Tune the pump about one mechanical frequency below the optical cavity resonance. In the resolved-sideband and weak-fluctuation limit a near-resonant process moves one excitation between the

mechanical mode and an upper optical sideband. The rapidly rotating opposite-sideband terms average out. Now the two coupled frequencies are an optical sideband offset and the mechanical frequency, measured relative to a common rotating reference. The large carrier frequency has already been accounted for by the pump. The pump phase can be absorbed into the mode definitions, so G may be chosen positive for this isolated pair.

If instead the pump is placed near the upper sideband, the surviving interaction creates or annihilates excitations in pairs. The passive exchange equations below cease to apply and strong pumping can produce amplification or instability. A bare two-oscillator machine driven mechanically near a shared natural frequency often has analogous exchange equations without a carrier pump. Identical matrices can describe different energy supplies; identify the drive and rotating reference before assigning a literal photon-to-phonon energy conversion.

Derive complex poles and the effect of unequal damping

Let a, b be small energy-normalised optical and mechanical complex amplitudes. In a frame where the uncoupled optical sideband has zero frequency, let the mechanical detuning be δ. Let κ, Γ denote energy linewidths in angular frequency, so amplitude decay rates are half these values. With an optical probe s at offset ω, the linear system is

$$\begin{pmatrix} \kappa/2 - i\omega & iG \\ iG & \Gamma/2 - i(\omega - \delta) \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} s \\ 0 \end{pmatrix}.$$

Zeros of the determinant give autonomous complex mode frequencies. In the lossless case the eigenvalues are $\delta/2 \pm \sqrt{G^2 + (\delta/2)^2}$. Far off resonance each branch approaches a bare mode. At zero detuning the two eigenvectors have equal optical and mechanical weights and the splitting is $2G$. The eigenvectors rotate continuously through the avoided crossing, an important discriminator from two uncoupled lines simply passing through one another.

At zero detuning with losses, completing the square in the determinant gives

$$\omega_{\pm} = -i(\kappa + \Gamma)/4 \; \pm \sqrt{G^2 - (\kappa - \Gamma)^2/16}.$$

The real parts separate once $G > |\kappa - \Gamma|/4$. The shared imaginary part above that threshold corresponds to a hybrid energy linewidth $(\kappa + \Gamma)/2$. Resolving two intensity features on a given instrument is a stronger, readout-dependent condition. The dimensionless cooperativity $C = 4G^2/(\kappa\Gamma)$ is yet another measure, comparing interaction against the product of dissipation scales. A value greater than one does not automatically prove two resolved normal-mode poles in a device with very unequal linewidths.

Use frequency numbers divided by 2π consistently from here for examples. With optical linewidth 240 kHz, mechanical linewidth 40 kHz and coupling 160 kHz, the real pole separation is

$$2\sqrt{160^2 - 50^2} = 304\,\mathrm{kHz}.$$

Both hybrid energy linewidths are 140 kHz, and cooperativity is 10.7. If coupling falls to 45 kHz, cooperativity is still about 0.844 and no real zero-detuning pole separation appears. At 50 kHz the square root reaches zero. Near that threshold even a small linewidth or detuning calibration error has a large effect on a coupling estimate inferred from a single splitting.

Compute what an optical or mechanical receiver actually sees

Invert the matrix. With $D = (\kappa/2 - i\omega)[\Gamma/2 - i(\omega - \delta)] + G^2$, the small amplitudes are $a = s[\Gamma/2 - i(\omega - \delta)]/D$ and $b = -iGs/D$. The poles are shared. Their numerators are different, so the two channels can show different apparent maxima. A port in reflection also has a directly reflected field that interferes with the leaking intracavity amplitude. Its intensity can display a narrow dip or asymmetric feature without a matching resonance-height pattern inside the cavity. A legitimate forward model must describe the measured port.

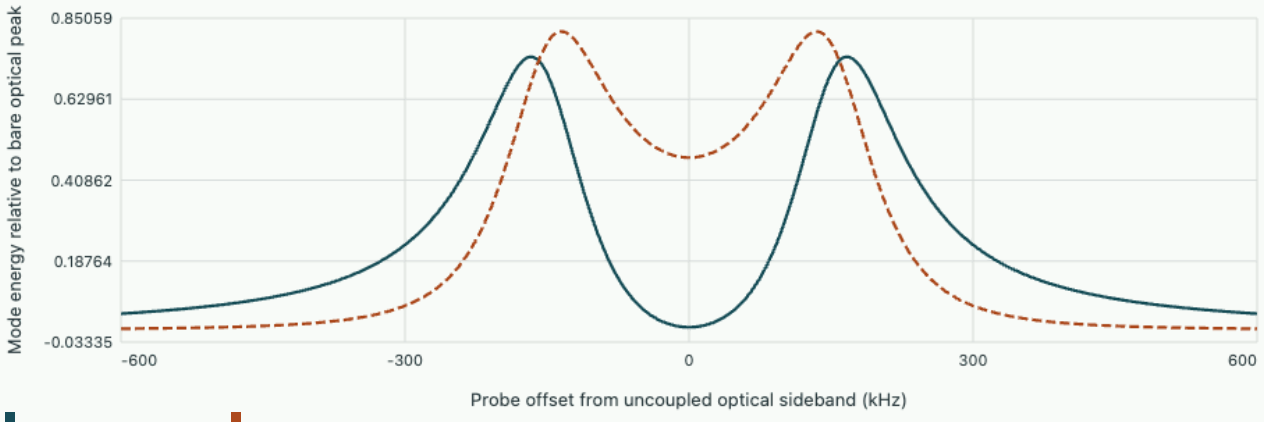
Open the live experiment

Try the calculation

Coupling G/(2π) (kHz) **160**

Optical linewidth κ/(2π) (kHz) **240**

Mechanical detuning (kHz) **0**



At zero bare detuning the real pole separation is 304.0 kHz. For the selected coupling and linewidths the cooperativity is 10.67. The mechanical linewidth is fixed at 40 kHz.

The graph plots the internal squared amplitudes relative to the bare on-resonance optical probe energy. Set the mechanical detuning to zero and reduce the coupling through 50 kHz. Compare the optical dip with the predicted real pole separation. Then set a substantial positive detuning and follow the transfer of mechanical response onto one branch. As a control set the coupling exactly to zero. The optical curve recovers the bare cavity Lorentzian and the mechanical probe-induced response vanishes. A thermally excited mechanical background, absent from this driven calculation, could remain in a real detector.

At zero detuning and zero probe offset,

$$a/a_{\mathrm{bare}} = 1/(1 + C).$$

For the default parameters the internal optical probe intensity at the centre is only $1/(1 + 10.67)^2 = 0.00735$ of the bare peak. This deep feature alone still does not establish the pole spacing. An incoherent background, a frequency-dependent detector transfer or coherent feedthrough can alter peak and dip positions. Fit complex quadratures across a detuning series when available, and verify bare linewidths by independent ringdown or far-detuned spectra.

Check how energy linewidth and amplitude ringdown are related for one isolated resonance

Watch energy exchange in time and identify the linewidth convention

In an ideal equal-linewidth system starting with optical perturbation $a(0) = a_0$, $b(0) = 0$ at zero detuning, change to symmetric and antisymmetric modes. Recombination gives

$$a(t) = a_0 e^{-\kappa t/2} \cos Gt,$$

$$b(t) = -i a_0 e^{-\kappa t/2} \sin Gt.$$

Their normalised total energy decays as $e^{-\kappa t}$, while energy moves back and forth between optical sideband and vibration. The first full transfer occurs at $t = \pi/(2G)$ if the rotating-wave approximation holds. For $G/(2\pi) = 100$ kHz, this is 2.50 microseconds. With a shared energy linewidth of 20 kHz in ordinary frequency the surviving energy fraction then is

$$e^{-2\pi(20\ \mathrm{kHz})(2.50\ \mu\mathrm{s})} = 0.730.$$

A gated excitation shorter than the exchange period but long enough to avoid unwanted modes can prepare an approximately optical initial perturbation. Independent detection of an optical revival and a delayed mechanical response tests coherent exchange more strongly than the presence of two peaks alone. Comparing times of maxima directly with one half the reciprocal measured optical linewidth misidentifies the exchange physics. Check the decay of the sum of appropriately calibrated energies and the relative phase of the exchange. Electrical or optical feedthrough can produce a prompt apparent response with no mechanical build-up.

Infer a coupling with propagated measurement errors

When the two hybrid pole frequencies have been obtained from a justified fit and zero bare detuning is established, the coupling follows from the measured real separation S:

$$G = \sqrt{S^2/4 + (\kappa - \Gamma)^2/16}.$$

All quantities here can be expressed in the same ordinary-frequency units. If $S = 300 \pm 8$ kHz, $\kappa = 240 \pm 10$ kHz and $\Gamma = 40 \pm 4$ kHz, the estimate is 158.1 kHz. Independent small errors give

$$\partial G/\partial S = S/(4G) = 0.474,$$

$$\partial G/\partial \kappa = -\partial G/\partial \Gamma = (\kappa - \Gamma)/(16G) = 0.0791.$$

Combining in quadrature gives about 3.9 kHz uncertainty. A common frequency calibration can correlate all three inputs, and a fitted detuning and splitting from one scan may also be correlated. Use their covariance rather than multiplying independent error bars blindly.

A second inference varies intracavity pump power. Since $G^2 = g_0^2 n_c$, plot the squared fitted coupling against independently calibrated n_c. A nonzero intercept could indicate an additional direct coupling, baseline fitting bias or a mode incorrectly assigned at low power. The linewidths themselves can vary with pump power because of heating or backaction, so fit or measure them across the scan. Incident optical power is not identical to stored photon number; cavity coupling, detuning and loss determine their relation.

A measured frequency shift and linewidth change reveal optical feedback

The same instrument can be operated below the resolved splitting regime and read out a dispersive shift of the mechanical response. Algebraically eliminate the optical amplitude in the two-mode equations while driving the mechanical mode weakly. Its inverse susceptibility acquires a self-energy contribution proportional to the optical susceptibility. Let x be probe frequency measured from the bare optical sideband and use the same linewidth convention as above. In the mechanical denominator the additional term is

$$\Sigma(x) = \frac{G^2}{\kappa/2 - i x} = \frac{G^2(\kappa/2 + i x)}{(\kappa/2)^2 + x^2}.$$

Its real part adds to the amplitude damping, while its imaginary part shifts the resonant condition. At very weak coupling one can evaluate it at the bare mechanical frequency $x = \delta$, because variation across the narrow mechanical peak is small. Near an avoided crossing or when the mechanical peak samples a large change in optical susceptibility, solve the coupled poles directly; simply adding a linewidth and shift then neglects the very hybridisation being tested.

In ordinary-frequency units choose $G = 10$ kHz, optical linewidth 200 kHz, mechanical linewidth 2 kHz and detuning $+100$ kHz. At that offset the optical amplitude half-width is 100 kHz, giving $\mathrm{Re}\,\Sigma = 0.50$ kHz and $\mathrm{Im}\,\Sigma = 0.50$ kHz. The full mechanical linewidth increases from 2.0 to approximately 3.0 kHz. Its centre moves upwards by about 0.50 kHz with the adopted matrix signs. Moving to minus 100 kHz reverses the shift and preserves the added damping in this symmetric model. At zero detuning the predicted shift vanishes but the added linewidth is largest.

A scan on both sides is more discriminating than measuring broadening alone. A background rise in mechanical temperature changes integrated thermal noise and may change intrinsic loss, but need not

reproduce this odd shift and even broadening pattern. Measure intrinsic linewidth far from the optical response and subtract that baseline. An unknown constant frequency offset can be removed with the two symmetric detunings, but an optical power dependence of cavity frequency can move the actual symmetry centre. Monitor the optical resonance during the scan.

The effective susceptibility describes classical feedback through the delayed optical field. It does not by itself specify the final mechanical temperature. A bath occupation or noise spectrum is also required to convert a damping change into an energy distribution. A driven response peak that broadens and lowers in amplitude is evidence for a changed susceptibility; thermal cooling additionally requires a calibrated measurement of the fluctuating mechanical displacement and its optical transduction.

Tutorial · A high cooperativity without resolved modes

A device has ordinary optical linewidth 1.00 MHz, mechanical linewidth 1.0 kHz and pump-enhanced coupling 30 kHz. Calculate cooperativity, compare the coupling with the real-splitting threshold and estimate the relative intracavity probe intensity at the central frequency. What observation would distinguish a narrow interference window from two separated poles?

▼ **Evaluate the three criteria independently**

In kHz units the cooperativity is

$$4(30)^2/(1000 \cdot 1) = 3.6.$$

The real-splitting threshold is $999/4 = 249.75$ kHz, far above the actual coupling. Yet the zero-detuning centre intensity is $1/(1+3.6)^2 = 0.0473$ relative to the bare optical peak. A narrow, deep structured feature is possible with no separated real pole frequencies in this model. The linewidth contrast permits mechanical response to interfere strongly within a broad optical susceptibility.

Acquire both amplitude and phase while scanning the absolute mechanical detuning and compare a coupled complex-response model with independently measured bare parameters. A pair of well-resolved poles should trace two consistent trajectories in both observables and can generate underdamped exchange in the time domain. A single optical intensity scan containing two flanks around a dip is insufficient to make that inference.

Tutorial · Diagnose an erroneous estimate of the single-photon coupling

An independent linewidth analysis gives pump-enhanced couplings 80 and 160 kHz at photon numbers 4.0×10^6 and 1.6×10^7. Estimate $g_0/(2\pi)$ and predict the coupling at 9.0×10^6 photons. A calibration revision doubles all inferred photon numbers while leaving the measured spectra unchanged. Which conclusions move?

▼ Use coherent amplitude scaling and separate calibration from spectroscopy

Divide the respective couplings in ordinary frequency by the square roots of their photon numbers. Both give 40 Hz and the intermediate prediction is 120 kHz. Doubling all inferred photon numbers lowers the inferred single-photon coupling by $1/\sqrt{2}$, to 28.3 Hz. The observed enhanced couplings, pole locations and cooperativities computed from measured linewidths stay the same. A prediction labelled by fixed physical photon number changes; a prediction at a repeatable laser setting can still be calibrated empirically.

To test the absolute calibration, compare the inferred single-photon value with the independent cavity frequency slope times the mechanical zero-point displacement. This requires an effective mass appropriate to the measured displacement and the actual cavity optical overlap. An incorrect choice of mechanical mode normalisation can mimic an error in stored photon number. Agreement across a power ratio tests the square-root scaling but cannot by itself determine the absolute optical or mechanical calibration.

Aspelmeyer, Kippenberg and Marquardt · cavity optomechanics, linearisation and strong coupling

Predictions, parameter profiles and nonlinear uncertainty

Inverse problems and least squares

Calculate intervals for a quantity of interest, then find where a linear covariance approximation loses the geometry of a nonlinear fit.

FROM FIRST PRINCIPLES

Propagate fitted uncertainty to a reported quantity

Suppose a fit gives a parameter estimate m with covariance C, and the quantity to be reported is $z = h(p)$. Near m, expand a differentiable h to first order. Writing $g = \nabla h(m)$, each small parameter change contributes through its directional sensitivity,

$$h(m + \delta p) - h(m) \simeq g^T \delta p,$$
$$\operatorname{var}(z) \simeq g^T C g.$$

Correlations can either reduce or enlarge the propagated variance. For a genuinely nonlinear quantity or a large parameter region, a local line may miss the full shape. One alternative holds the target value fixed and finds the best fit compatible with it. With an objective Φ, define

$$\Phi_{\text{prof}}(z) = \min_{p:\, h(p)=z} \Phi(p).$$

Comparing this constrained objective with its unconstrained minimum tracks compensating changes in the nuisance parameters. A confidence level for its increase requires an explicit statistical model; tracing an objective curve alone supplies no probability calibration.

DERIVATION AND APPLICATIONS

Derive a profile interval using the covariance of a linear fit

Suppose a Gaussian linear experiment gives unconstrained estimate m and strictly positive covariance C, with correctly known noise scale. Its negative log likelihood relative to the optimum is the quadratic distance $\frac{1}{2}(p - m)^T C^{-1}(p - m)$. To profile a reported scalar $z = c^T p$, require the candidate parameters to satisfy that equality. Introduce a multiplier and differentiate with respect to the vector,

$$C^{-1}(p - m) + \lambda c = 0,$$
$$p = m - \lambda C c.$$

The constraint determines $\lambda = (c^T m - z)/(c^T C c)$. Substitution gives both the parameter combination selected at each target value and the profile height,

$$p_z = m + Cc\frac{z - c^T m}{c^T C c},$$

$$2[\Phi_{\text{prof}}(z) - \Phi(m)] = \frac{(z - c^T m)^2}{c^T C c}.$$

This reproduces the usual propagated Gaussian variance while showing exactly how the other fitted coordinates move along the profile. For a monotone function $q = g(z)$, imposing q fixes the corresponding $z = g^{-1}(q)$. The profile then transforms without approximation,

$$2\Delta\Phi(q) = \frac{[g^{-1}(q) - c^T m]^2}{c^T C c}.$$

An interval selected by a likelihood-ratio threshold maps by applying g to its two endpoints. It can be asymmetric even when the underlying likelihood in p is exactly quadratic. If g is not one-to-one, every feasible branch must be considered and the smallest objective chosen.

WORKED EXAMPLE

Track the nuisance displacement along a profile

Take $m = (2, 4)^T$, $C = \left(\begin{smallmatrix} 1 & 0.5 \\ 0.5 & 1 \end{smallmatrix}\right)$, and target $z = p_1 + p_2$. The target mean is 6 and variance 3. At $z = 9$, $Cc = (1.5, 1.5)^T$, giving constrained state $(3.5, 5.5)^T$. The twice-log-likelihood increase is $3^2/3 = 3$. Holding the second parameter fixed would impose a different and more restrictive path.

SECOND APPLICATION

Transform an interval for a positive output

A second model estimates $z = \ln(q/q_0)$ with mean 0.4 and deviation 0.3; let $q_0 = 10$. At threshold $2\Delta\Phi = 4$, the interval in z is $[-0.2, 1.0]$. Mapping its endpoints gives $q \in [8.19, 27.18]$, around the maximum-likelihood value $10e^{0.4} = 14.92$. A symmetric interval about that value would lose the exact profile geometry.

Push uncertainty through the prediction

Often a user needs tomorrow's temperature or a time to threshold, rather than all fitted coefficients. For a linear target $h(x) = c^{\mathsf{T}} x$ and Gaussian posterior, h is Gaussian with mean $c^{\mathsf{T}} m$ and variance $c^{\mathsf{T}} C c$. Include covariance terms: positively correlated errors raise uncertainty in a sum and reduce it in a difference. For a new noisy measurement add its own independent measurement variance. Reporting this predictive interval as the uncertainty of the noise-free underlying response would mix two questions.

For a smooth nonlinear h, expand about the mean m. The first-order variance is $\nabla h^{\mathsf{T}} C \nabla h$; the leading correction to the mean is $\frac{1}{2}\mathrm{tr}[H_h(m)C]$. If the gradient vanishes, first-order variance propagation can return zero despite substantial uncertainty. For example, if x is centred normal with variance s^2 and

$h = x^2$, then $\mathbb{E}h = s^2$ and $\operatorname{Var} h = 2s^4$, using the Gaussian fourth moment. The linear approximation at zero would miss both.

A positive rate and a nonlinear lifetime

Suppose $\theta = \ln(k/k_{\text{unit}})$ is normal with mean m and standard deviation s. A decay lifetime is

$$t_* = 1/k = e^{-\theta}/k_{\text{unit}}.$$

Its median is e^{-m}/k_{unit}; its mean is larger by $e^{s^2/2}$. A central 95% interval is $[e^{-m-1.96s}, e^{-m+1.96s}]/k_{\text{unit}}$. These endpoints need not be equally far from the median. For median 10 s and $s = 0.5$, the interval is approximately [3.75, 26.6] s and mean 11.33 s. Transform samples or quantiles if strong nonlinearity makes symmetric error bars misleading.

A profile follows compensating changes

Let the fit have parameters (p, q). To profile p, fix it temporarily at each trial value and minimise the objective over q. The resulting curve reflects how much the other parameter can compensate. Holding q fixed instead gives a slice that is generally steeper and gives deceptively narrow intervals. For the sum-of-squares surface $F = (p + q - 2)^2 + 0.04(q - 1)^2$, minimising gives

$$q = (2.04 - p)/1.04.$$

The resulting profile is $0.04(p - 1)^2/1.04$. A slice at the joint optimum's $q = 1$ is just $(p - 1)^2$. Both have minimum at one but radically different curvature.

Read the level as well as the curve

A profile likelihood is a maximisation over nuisance parameters. A Bayesian marginal instead integrates their probability. Even when their modes agree the two objects can differ in shape in nonlinear models. A likelihood-ratio threshold yields approximate confidence sets under regularity assumptions and enough informative data. Boundaries, multiple modes and poorly identified nuisance parameters may invalidate a familiar chi-square cutoff. Report the objective convention and how the threshold was calibrated, e.g. by simulated repetitions of the full experiment. A plot of the raw profile and representative predicted curves along it remains interpretable even when an asymptotic confidence level is questionable.

Invert a test to put an interval on a ratio

A scientifically relevant quantity may be $q = \mu_y/\mu_x$, for example the response per unit excitation. Estimates $\hat{x}, \hat{y}$ can each have nearly Gaussian uncertainty even when their ratio is highly asymmetric. Suppose for now that they are independent normals with known variances s_x^2, s_y^2, and that the true

denominator is nonzero. To test a proposed ratio q, consider the contrast $\hat{y} - q\hat{x}$. If the proposed ratio is correct, that contrast has mean zero and variance $s_y^2 + q^2 s_x^2$. The source of difficulty, the uncertain denominator, now enters the reference variance explicitly.

Collect all values whose absolute standardised contrast is at most a normal quantile z. Squaring and expanding gives the quadratic inequality

$$(\hat{x}^2 - z^2 s_x^2) q^2 - 2\hat{x}\hat{y} q + (\hat{y}^2 - z^2 s_y^2) \le 0.$$

This inversion gives a confidence set because it contains the true ratio exactly when the correctly standardised contrast falls in the specified acceptance region. Its shape need not resemble a symmetric error bar. If the leading coefficient changes sign, the accepted set may be unbounded. A covariance between $\hat{x}$ and $\hat{y}$ changes the variance to $s_y^2 + q^2 s_x^2 - 2q\,\mathrm{Cov}(\hat{x}, \hat{y})$, so the preceding coefficients require modification.

A precise reading divided by an uncertain excitation

Use $z = 2$, corresponding to about 95.45 percent coverage for the stated normal model. Set $\hat{x} = 2, \hat{y} = 4,\ s_x = 0.5,\ s_y = 1$. The quadratic coefficients are 3, –16 and 12. Its roots are $(16 \pm \sqrt{112})/6$, about 0.903 and 4.431. The accepted set is the closed interval between them because the polynomial opens upward. The point estimate is two. The interval extends considerably farther upward than downward because a smaller plausible excitation can produce a much larger ratio. A symmetric linearised interval loses this asymmetry.

Now keep the observations and numerator uncertainty fixed but raise s_x to 1.5. The resulting inequality is $-5q^2 - 16q + 12 \le 0$, whose roots are $(-16 \pm \sqrt{496})/10$, about –3.827 and 0.627. The accepted set consists of the two exterior rays. It includes very large positive and negative values. For such values the standardised contrast approaches $|\hat{x}|/s_x = 4/3$, inside the threshold two, precisely because the data cannot rule out an excitation arbitrarily close to zero.

An investigator with independent physical information that the excitation is strictly positive may intersect the set with the resulting permitted parameter space. A numerical routine that simply discards negative draws or clips ratios at a convenient maximum is instead changing the uncertainty analysis. The unbounded set is a measurable statement about what an experiment failed to localise, and identifies the denominator measurement as an appropriate target for redesign.

When $\hat{x}^2 = z^2 s_x^2$ exactly, the quadratic term disappears. The confidence set is then ordinarily a half-line determined by a linear inequality. This is the transition between bounded and exterior solutions. A robust implementation must handle it directly instead of dividing by a leading coefficient that may be zero or extremely small.

Change parameters · Explore the worked calculation

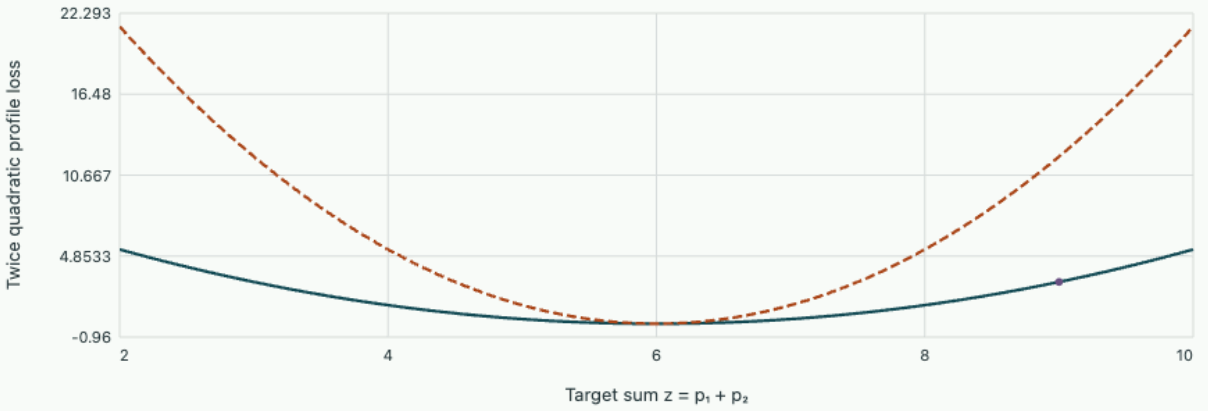
Change one input at a time and compare the recalculated curves with the argument above. The browser evaluates this small model locally; the accompanying numerical values state its assumptions and reference checks. Reset restores the reference inputs, and the data can be downloaded as CSV.

Open this calculation with adjustable inputs

Try the calculation

Prior parameter correlation **0.5**

Selected sum target z **9**



Both parameters refitted | Second parameter held at its mean | Selected profiled target

Selected profiled state (3.5, 5.5). Sum variance 3. Profiling moves nuisance parameters; fixing them gives a different curvature.

Learning a drifting temperature from a sequence of readings

Inverse problems and least squares

Derive the scalar Kalman update by completing the square and see how process uncertainty controls memory of the past.

Propagate a distribution through time

A drifting offset follows $x_k = x_{k-1} + w_k$. A thermometer reports $z_k = x_k + v_k$. Assume independent, zero-mean Gaussian process increments of variance Q and sensor noise of variance R, independent at each time and of the initial state. Before the next reading, the mean stays at $m^- = m_{k-1}$ and the variance adds, $P^- = P_{k-1} + Q$.

The negative log posterior, up to constants, is $(x - m^-)^2/(2P^-) + (z_k - x)^2/(2R)$. Collect the coefficient of x^2 and its linear term. The result can be written in two equivalent forms:

$$P_k = (1/P^- + 1/R)^{-1},$$

$$m_k = m^- + K(z_k - m^-),$$

$$K = P^-/(P^- + R).$$

Thus the fraction of the innovation used is a model prediction. When sensors are noisy the gain falls. When the model permits rapid drift, uncertainty grows between readings, giving the new measurement a larger role. These are variances, so a stated standard deviation must be squared first.

A reading with units and a repeated reading

A prior says 20.0 ± 2.0 °C (one standard deviation), with negligible intervening drift. A sensor of standard deviation 1 °C reads 23 °C. The gain is $4/(4+1) = 0.8$, giving 22.4 °C and variance 0.8 °C². A second independent reading of 23 °C gives gain $0.8/1.8$ and mean 22.6667 °C. Treating a duplicate record as independent would falsely claim the same information gain.

In a long random walk with $Q = 0.25$ and $R = 1$ in squared units, a steady posterior variance satisfies

$$P = (P + Q)R/(P + Q + R).$$

Rearrangement gives $P^2 + QP - QR = 0$, so the physically positive root is about 0.3904. New uncertainty replenishes the information squeezed out by measurement.

Follow a change in the underlying temperature

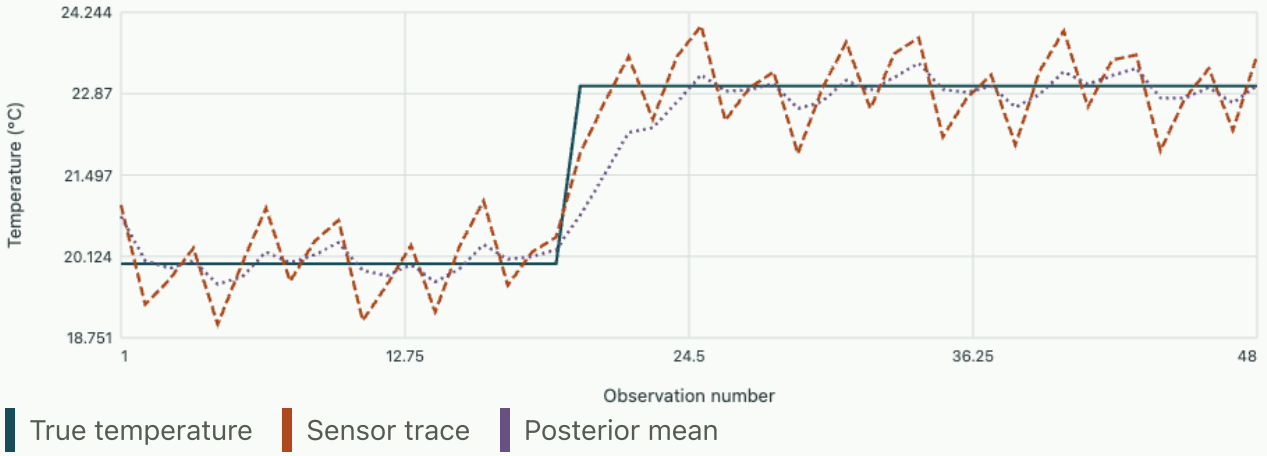
For a reproducible illustration, a fixed synthetic thermometer trace contains a smooth periodic disturbance and a step in the hidden temperature. It is deliberately held fixed while the filter's assumed Q and R change. Compare the lag just after the step with the wiggle once the temperature settles.

Open this experiment online

Try the calculation

Assumed process variance Q (°C²) **0.2**

Assumed sensor variance R (°C²) **1**



Last posterior mean 23.005 °C, standard deviation 0.59855 °C. Gain 0.35826.

Tutorial on noisy data over time

If $Q = 0$, show how n independent readings with the same variance affect precision.

▼ **Worked solution**

Each update adds $1/R$ to precision $1/P$. Induction gives

$$1/P_n = 1/P_0 + n/R.$$

This is precisely the precision of a batch Gaussian update.

What pattern in the innovations would call the random-walk assumptions into question?

▼ **Worked solution**

A long run of positive innovations after a systematic ramp suggests bias in the prediction. Serial correlation suggests missing dynamics or correlated observation errors. Check against predicted innovation variance $P^- + R$ and examine signs and lags as well as their sizes.

Read the university treatment

Extend the scalar precision update to several state variables and identify the matrix made invertible by a positive observation variance. Stephen Boyd · Stanford EE363, estimation and the Kalman filter

Predicting a new measurement when the fitted parameters are correlated

Inverse problems and least squares

Propagate the full posterior covariance to a quantity of interest, distinguish latent and measured uncertainty and select the observation that reduces the uncertainty of the actual prediction.

One model supplies many possible targets

For parameters θ with posterior mean m and covariance C, a linear target $s = h^T\theta$ has mean $h^T m$. Subtract the mean and expand the quadratic product to obtain

$$\operatorname{Var}(s) = h^T C h.$$

An independent sensor with variance R changes the variance of a future measured value to $h^T C h + R$. Repeated fits of noisy datasets and repeated physical measurements are distinct sources of variation.

A candidate new observation $z = a^T\theta + \epsilon$ with variance R_a has covariance $h^T C a$ with the target. For a Gaussian posterior the conditional covariance update is $C_+ = C - Caa^T C/(a^T C a + R_a)$, found either by completing the square or projecting out the optimal multiple of the innovation. The target variance decreases by

$$\Delta \operatorname{Var}(s) = \frac{(h^T C a)^2}{a^T C a + R_a}.$$

Background · Posterior predictive

$$p(y_*|y) = \int p(y_*|m)p(m|y)\,dm$$

A distribution over parameters induces uncertainty in a new observation. Measurement noise and uncertainty of the prediction mean are distinct contributions.

A precise sum and an uncertain difference

Let C have diagonal entries 1 and correlation -0.9. A target $h = (1, 1)$ has variance $2 - 1.8 = 0.2$. The difference $h = (1, -1)$ has variance 3.8. Ignoring correlations would predict 2 for both. With sensor variance 0.25, a newly measured sum has variance 0.45 although the latent sum variance remains 0.2.

For the target sum, measuring the first coordinate with variance 0.25 reduces variance by $(1 - 0.9)^2/(1 + 0.25) = 0.008$. Measuring the sum itself at the same variance reduces it by $0.2^2/(0.2 + 0.25) = 0.0889$. These two sensors have differently normalised sensitivity vectors. Their stated noise levels refer to their own output units and must be calibrated before a cost comparison.

Choose the measurement direction

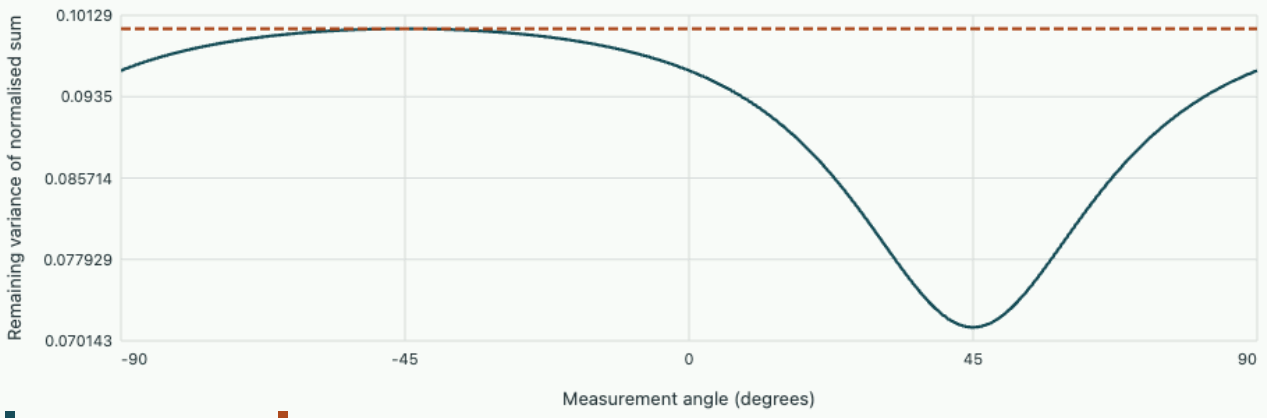
For a fairer geometric comparison, restrict both measurement and prediction directions to unit length. Fix the target along the normalised sum. Sweep candidate observation angles and read the remaining prediction variance. Change the parameter correlation: when does a coordinate sensor improve the target appreciably?

Open this experiment online

Try the calculation

Parameter correlation **-0.9**

Candidate sensor variance **0.25**



Target variance 0.1 initially, 0.096 with a first-coordinate sensor, 0.071429 with a sensor along the target.

Tutorial on prediction and parameter recovery

If $h^T C a = 0$, can a noise-free measurement of $a^T \theta$ improve the linear prediction under these Gaussian assumptions?

▼ **Worked solution**

The covariance reduction is exactly zero. The measurement may determine another direction precisely while carrying no linear information about the target.

The posterior on θ is strongly non-Gaussian. Which conclusions still hold exactly?

▼ **Worked solution**

Mean and variance of a fixed linear prediction still follow from its first two moments, and independent sensor variance still adds. The Gaussian conditional update and an interval based only on two standard deviations need additional assumptions.

Read the university treatment

Compare the projection of a vector measurement with the covariance of a scalar target. Check the direction of each matrix product. Stephen Boyd · Stanford EE363, estimation

Measure gravity with a pendulum of finite amplitude

Mechanics in the laboratory

A metre-long pendulum is an appealing clock because its small-angle period depends only on length and gravity. A precise clock can nevertheless give a precisely wrong value of gravity. Derive the full period, trace the systematic error to an identifiable approximation and design a measurement that distinguishes a length error from a release-angle error.

Define the measured length and time

Use a point bob on a massless, inextensible string attached to a fixed pivot. Let L be the distance from pivot to the bob's centre of mass, and θ the signed departure from the downward vertical. For a rigid extended bob, rotational inertia about its own centre adds to mL^2; a length measured to the bottom of the bob is neither of these dynamical lengths. Measure the elapsed time between successive crossings of a photogate in the same direction. Crossings in opposite directions give half of the period in a symmetric oscillation. A stationary turning point is a poor timing marker because the position error there produces a large time error.

At rest at the release angle θ_0, the energy is $mgL(1-\cos\theta_0)$. At a later angle the speed is $L|\dot\theta|$. Equating kinetic plus potential energy to the release value gives

$$\dot\theta^2 = (2g/L)(\cos\theta - \cos\theta_0).$$

Only a quarter cycle is needed. Inverting the angular speed and integrating from zero to θ_0 produces a square-root singularity at release. The period is finite since integrating the inverse square root over a small interval gives a square root.

Resolve the turning point before approximating

Write $k = \sin(\theta_0/2)$ and substitute $\sin(\theta/2) = k\sin u$. The vanishing square root cancels a factor $\cos u$ from the differential. One obtains the complete elliptic integral of the first kind with modulus k.

$$T = 4\sqrt{\frac{L}{g}}\,K(k),$$

$$K(k) = \int_0^{\pi/2} \frac{du}{\sqrt{1-k^2\sin^2 u}}.$$

Expanding its integrand yields

$$K(k) = (\pi/2)(1 + k^2/4 + 9k^4/64 + \cdots).$$

Since $k \simeq \theta_0/2$ at small amplitude, the first fractional period correction is $\theta_0^2/16$ with angle in radians. It is positive. At larger amplitude the tangential component of gravity rises more slowly with angle than the

linear restoring law predicts. The pendulum spends longer near the turns. Passing degrees directly into the quadratic correction would give nonsense even if the timer and recorded angle were perfect.

Infer the bias in gravity

Suppose the true acceleration is $9.81\,\mathrm{m\,s^{-2}}$, the length is 1.000 m and the release angle is 30 degrees. The small-angle period is about 2.0061 s and the finite-amplitude period about 2.0410 s. Interpreting the longer measured period by the linear formula would give about $9.48\,\mathrm{m\,s^{-2}}$. The fractional gravity error is about twice the fractional period error with opposite sign because gravity is inferred from T^{-2}. Timing a hundred periods reduces random timing error but leaves this systematic shift.

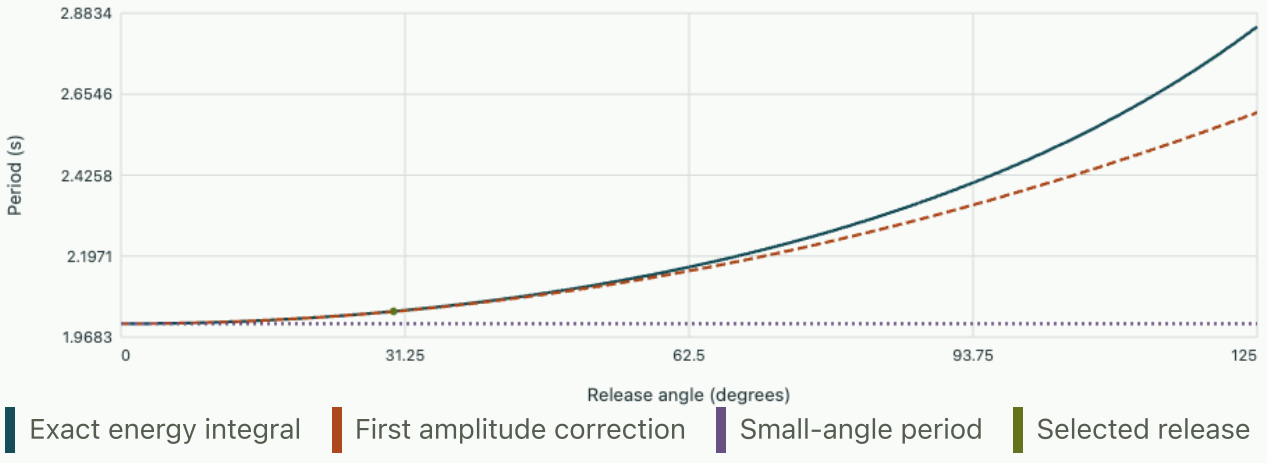
The numerical curve evaluates the complete elliptic integral. It uses the arithmetic–geometric mean identity $K(k) = \pi/[2\,\mathrm{AGM}(1, \sqrt{1-k^2})]$, so no sampled point sits on the turning point. The single-correction curve truncates the amplitude series after its first nonconstant term. Vary angle and length, comparing both the predicted period and the result a linear analysis would infer. Which change rescales both period curves together? Which changes their relative discrepancy?

Open this experiment online

Try the calculation

Release angle (degrees) 30

Pendulum length (m) 1



Period 2.041 s. Small-angle analysis would infer g = 9.4772 m/s².

Design a two-release check

For the same pendulum the ratio $T(\theta_a)/T(\theta_b) = K(k_a)/K(k_b)$ cancels the unknown length and local gravity. Measure at two amplitudes before deciding whether to recalibrate the length. If the ratio agrees but both absolute times differ from prediction, check the effective length, clock rate and model of the

bob. If the ratio changes during a decaying swing, record initial and final amplitudes for the timed sequence. The period associated with a long sequence is an average over its slowly changing amplitude. Substituting the initial angle throughout would overcorrect the measurement.

Tutorial

At small amplitude, what release angle gives a fractional period bias of approximately one part in a thousand?

▼ **Work through the solution**

Set $\theta_0^2/16 = 0.001$. This gives 0.1265 rad, about 7.25 degrees. The gravity bias from an uncorrected measurement would be approximately two parts in a thousand. A tolerance on one quantity is not automatically the same tolerance on the other.

A photogate records successive interruptions every 1.02 seconds for the metre pendulum. A calculation gives nearly four times local gravity. Identify one possible error in the observable.

▼ **Work through the solution**

The gate may have counted both directions and returned a half period. Doubling that interval gives about 2.04 s. Because the inferred acceleration is inversely proportional to the square of the period, the missing factor of two gives a factor of four.

Why does reducing the grid spacing in an integration of the linear equation fail to recover the finite-amplitude curve?

▼ **Work through the solution**

Integration error vanishes as the step is refined, but the restoring torque in that equation is $mgL\theta$. The true torque is $mgL\sin\theta$. A converged numerical solution can therefore provide independent evidence of an inadequate physical approximation.

For the equation of motion, phase portrait and dissipation, see the MIT treatment of the simple pendulum.

Measure inertia and gravity with two suspension axes

Mechanics in the laboratory

A rigid pendulum can oscillate from either of two knife edges. If the periods agree at distinct distances from the centre of mass, the spacing of the edges determines gravity. Derive the cancellation, identify a second way equal periods can arise and find why a poorly chosen geometry magnifies timing errors.

Torque about an offset pivot

Take a rigid body of mass m with parallel suspension axes perpendicular to its plane of oscillation. Its moment of inertia about the centre of mass is $I_C = m\rho^2$; ρ is the radius of gyration. For a pivot a distance $a > 0$ from the centre, the parallel-axis theorem gives

$$I_a = m(\rho^2 + a^2).$$

Gravity exerts torque $-mga\sin\theta$. In the small-angle limit,

$$T_a^2 = \frac{4\pi^2}{g}\,\frac{\rho^2 + a^2}{a},$$

$$L_a = a + \frac{\rho^2}{a}.$$

Here L_a is the length of a point pendulum with the same period. It can lie beyond the physical centre of mass. A tap applied at this distance from a pivot gives the centre of percussion when the ideal instantaneous pivot reaction is to vanish. The same combination of mass distribution and geometry appears in an impulsive and in an oscillatory measurement because both depend on angular momentum about that pivot.

Match the periods and keep both algebraic branches

Put the second knife edge on the opposite side of the centre at distance b. Equal periods require

$$a + \rho^2/a = b + \rho^2/b.$$

Rearrange before dividing by $a - b$:

$$(a - b)\left(1 - \frac{\rho^2}{ab}\right) = 0.$$

Two equal distances always give equal periods. That branch provides no information about ρ. If the distances are distinct, equality implies

$$\rho^2 = ab.$$

Both equivalent lengths then equal $a + b$, the directly measurable edge separation. For $a = 0.200\,\mathrm{m},\ b = 0.450\,\mathrm{m},\ \rho = 0.300\,\mathrm{m}$, the equivalent length is 0.650 m and both periods at standard laboratory gravity are about 1.62 s. A small value of $a - b$ is an experimental warning sign, even if the observed equality is exceptionally clean.

Background · Can unequal release amplitudes imitate a mismatch between pivots?

A rigid pendulum about either knife edge obeys the same sine nonlinearity as a simple pendulum. To leading order in its release angle θ in radians, the measured period exceeds the small-amplitude period by the fraction $\theta^2/16$. At matched small-amplitude periods, measurements with amplitudes θ_a, θ_b therefore give

$$\frac{T_a - T_b}{T_0} \simeq \frac{\theta_a^2 - \theta_b^2}{16}.$$

Releasing at 6° about the first edge and 2° about the other creates a fractional mismatch of about 6.1×10^{-4}. For a two-second period this is 1.2 ms. Record the two amplitudes along with the timing data and reduce both periods to the same limit before moving the trim mass. Otherwise the trim can compensate a preparation error.

Extract gravity when the two measured periods differ

Multiply the first period equation by a and the second by b. Subtract to eliminate ρ^2. The exact small-angle result for unequal distances is

$$g = 4\pi^2\,\frac{a^2 - b^2}{aT_a^2 - bT_b^2}.$$

At matched periods this reduces to the simple expression in terms of the edge spacing. The general formula involves a difference in its denominator. If a and b nearly coincide, the denominator is small and two independent timing errors subtract with poor relative precision. This is a property of the inverse measurement, even if computing each individual period is perfectly well conditioned.

Once g is known, use $\rho^2 = gaT_a^2/(4\pi^2) - a^2$ to recover the body inertia per unit mass. A negative value may indicate a wrong distance measured to the centre or a timing observable that is actually a half period. Verify the result using the second pivot and recover a separately predicted period after shifting a suspension point. A single successful cancellation is weaker than reproducing an additional measurement.

Move one knife edge in the experiment

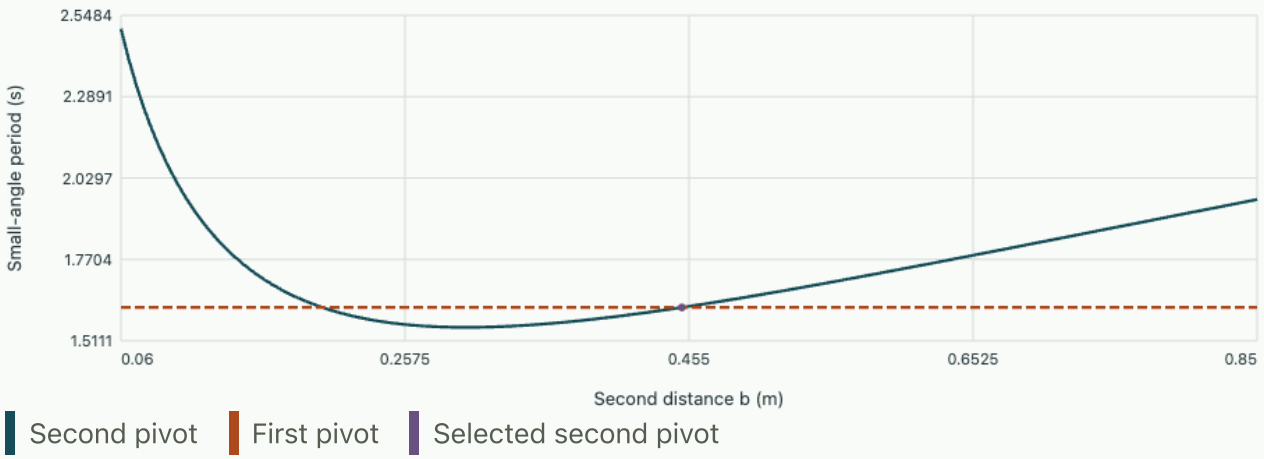
Hold $a = 0.200\,\mathrm{m}$ fixed and vary both the radius of gyration and the second distance. The graph plots the period against the possible second distance. Before inspecting intersections, predict where the second match should lie as ρ grows. The dependence on distance has a minimum at $b = \rho$ because effective length comprises a growing and a decreasing term.

Open this experiment online

Try the calculation

Radius of gyration (m) **0.3**

Second pivot distance (m) **0.45**



First period 1.6173 s; second 1.6173 s. The other matching distance is 0.45 m.

With a finite oscillation amplitude, the physical pendulum has the same elliptic correction as a point pendulum of length L_a, provided the body is rigid and the frictionless pivot stays fixed. Matching different amplitudes can bias the cancellation. Measuring both at the same small angle removes the leading difference, and extrapolating both to zero amplitude provides a check. Rotational compliance of the knife edge and a shifting centre of mass require a different model.

Tutorial

For $a = 0.240$ m and $\rho = 0.300$ m, where should the second distinct match occur and what length should be used in $g = 4\pi^2 L/T^2$?

▼ **Work through the solution**

The second branch gives

$$b = 0.0900/0.240 = 0.375\,\mathrm{m}.$$

Because the pivots straddle the centre of mass the equivalent length is 0.615 m.

Both pivots lie on the same side of the centre, at the same distances a, b as in the previous question. What changes in the cancellation?

▼ Work through the solution

The periods depend on the magnitudes of the offsets, so equal periods still imply $\rho^2 = ab$ on the distinct branch. The sum $a + b$ still gives the equivalent length, but the physical ruler distance between pivots is $|a - b|$. The direct measurement shortcut relies on the chosen geometry.

Why does setting both axes exactly at the same distance make even noise-free measurements insufficient to determine both g and ρ?

▼ Work through the solution

The two equations are then identical. For every positive choice of inertia radius the first period determines a corresponding value of gravity. Improving timing precision cannot provide the missing independent equation.

Connect the offset-pivot calculation to the rigid-body inertia chapter.

Time stepping and robust computation

Investigate state growth in apparently stable systems, propagate derivatives and judge whether numerical error can account for a switching event.

1. Absolute stability and stiffness
2. Adaptive steps and event locations
3. Tangent equations and adjoint outputs
4. How two decaying modes produce a growing signal
5. Locating a grazing event between time steps
6. Eigenvalues, modes and spectral gaps
7. Singular values and minimum-norm solutions
8. Deciding how far to trust a linearised residual
9. Master equations, detailed balance and probability currents

Absolute stability and stiffness

Numerical ordinary differential equations

Determine the step sizes that damp a decaying mode and explain why fast, physically uninteresting modes can dominate the cost of explicit integration.

FROM FIRST PRINCIPLES

Find a stability limit directly from repeated multiplication

A small perturbation around an equilibrium often satisfies a linear equation $y' = \lambda y$. Begin with real $\lambda < 0$, for which the continuous value is multiplied by $e^{h\lambda}$, strictly between zero and one. Apply Euler's left- and right-endpoint approximations to the integral of this same rate. Let $z = h\lambda$:

$$y_{n+1} = y_n + z y_n,$$
$$y_{n+1} = y_n + z y_{n+1}.$$

Solve each update for its new value and iterate n steps. Persistence of roundoff or initial error is governed by the same powers:

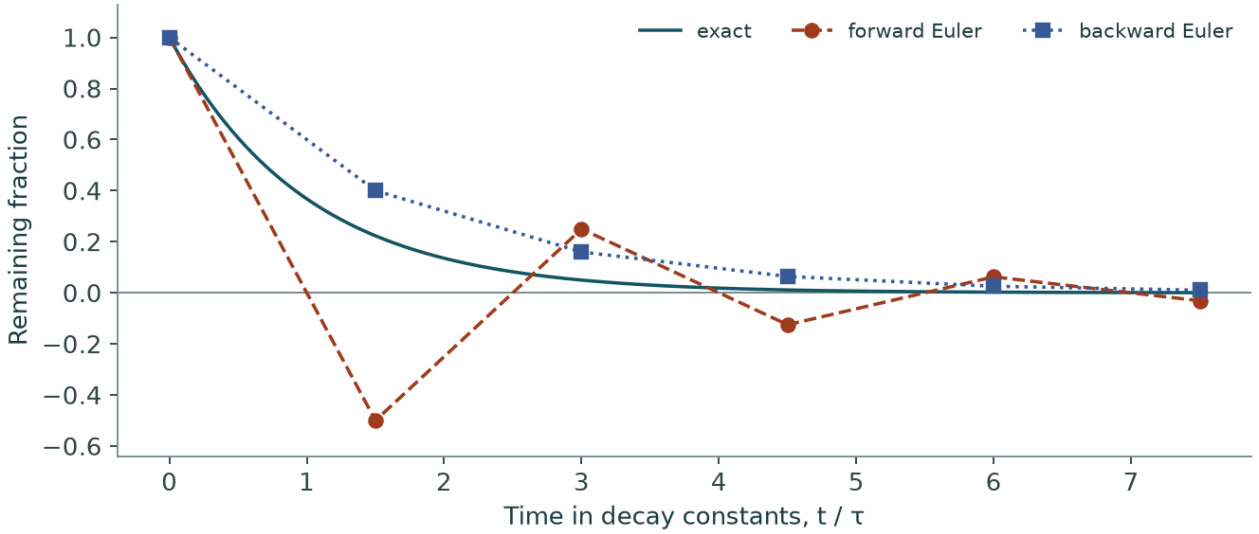
$$y_n = R(z)^n y_0,$$
$$R_{\text{forward}} = 1 + z,$$
$$R_{\text{backward}} = (1 - z)^{-1}.$$

For strict decay, $|R| < 1$ is necessary. Forward Euler on this real mode requires $-2 < z < 0$, and avoiding sign alternation further requires $z \geq -1$. Backward Euler damps every negative real z, even when the step is too large to approximate a transient accurately. With modes at -1 and -1000 s⁻¹, the fastest mode restricts a forward step to less than 0.002 s even after that transient has become tiny. This mismatch between stability and the resolution needed for the slow observable is the source of stiffness in this example.

Background · Stability function

$$y_{n+1} = R(h\lambda) y_n$$

On the scalar linear test equation, absolute stability depends on the point h times lambda. It gives only part of the story for a nonnormal matrix or nonlinear problem.



A stable update can predict an impossible concentration

Take a decay time τ and a time step h = 1.5τ. The exact one-step factor is exp(−1.5) ≈ 0.223. Forward Euler multiplies by −0.5 and backward Euler by 0.4. Both numerical magnitudes decay, yet the first update changes sign. Refinement tests accuracy even when positivity and stability hold.

DERIVATION AND APPLICATIONS

Separate unconditional stability from damping of an unresolved mode

A stiff dissipative mode may have decayed almost completely during a single time step. Merely keeping its computed amplitude bounded is then a weak requirement. Examine the family of one-step formulas that weights the old and new slopes with $1-\theta$ and θ. On the scalar test equation $y' = -\lambda y$, $\lambda > 0$, define the dimensionless decay over a step as $r = h\lambda$. Collecting the new value on the left gives

$$y_{n+1} = y_n - r[(1-\theta)y_n + \theta y_{n+1}],$$

$$G(r) = \frac{y_{n+1}}{y_n} = \frac{1-(1-\theta)r}{1+\theta r}.$$

The denominator is positive for $\theta \geq 0$. The amplification never exceeds one from above; enforcing $G \geq -1$ supplies the other half of the real-axis stability condition:

$$|G| \leq 1 \quad \Longleftrightarrow \quad 2 + (2\theta - 1)r \geq 0.$$

Thus all positive step sizes are stable along the negative real eigenvalue axis for $\theta \geq 1/2$. Now investigate what actually appears in the stored solution when the mode is unresolved:

$$\lim_{r\to\infty} G(r) = -\frac{1-\theta}{\theta} \quad (\theta > 0),$$

$$\lim_{r\to\infty} e^{-r} = 0.$$

The trapezoidal choice $\theta = 1/2$ approaches a sign-reversing factor of unit magnitude. Backward Euler $\theta = 1$ suppresses the fast mode. These properties govern the handling of transients; the order of approximation for a smooth, resolved mode is a separate question.

WORKED EXAMPLE

Check a relaxation after one coarse step

Take $\lambda = 1000$ s⁻¹, $h = 0.100$ s and an initial perturbation of 1.00 K. The exact remnant is e^{-100} K, negligible. Trapezoidal stepping leaves $(1-50)/(1+50) = -0.961$ K, while backward Euler leaves $1/101 = 0.00990$ K. After two trapezoidal steps the perturbation becomes positive again at 0.923 K. That oscillatory persistence is numerical, even though the amplification satisfies the stability condition.

SECOND APPLICATION

Choose a step for monotonically decreasing transients

Suppose a solver uses $\theta = 0.750$ and should preserve positivity for a spectrum of decay rates up to 400 s⁻¹. The numerator of G requires $1 - 0.250h\lambda \geq 0$, hence $h \leq 0.0100$ s. A 0.0200 s step would remain stable on every mode but gives $G = (1-2)/(1+6) = -1/7$ for the fastest one. A positivity requirement can impose a finite time step even when boundedness does not.

Examine a single amplification factor

For the real decay problem $y' = -\lambda y$, $\lambda > 0$, the physical solution multiplies by $e^{-h\lambda}$ over a step. A linear one-step method replaces this by $R(z)$, with $z = -h\lambda$. Absolute stability at z means repeated multiplication does not amplify magnitude, conventionally $|R(z)| \leq 1$. A truly decaying mode requires a strict inequality for numerical damping. Positivity and phase fidelity impose additional requirements beyond this magnitude test.

Forward Euler has $R = 1 + z$. In the complex plane the stability region is the disk of radius one centred at -1. On the real negative axis, decay requires $0 < h\lambda < 2$; nonnegative decay requires $h\lambda \leq 1$. Backward Euler has $R = 1/(1-z)$, and the trapezoidal method has $R = (1+z/2)/(1-z/2)$. Both contain the entire left half-plane in their stability regions. As z becomes large and negative, backward Euler tends to zero while trapezoidal tends to minus one. The former strongly damps unresolved fast components; the latter can keep them as alternating numerical transients.

Follow a slow and a fast component together

Let $u' = -u,\ v' = -1000v$, with $u(0) = v(0) = 1$. At $t = 0.1$ the exact slow state is 0.90484, while the fast state is e^{-100}, effectively zero on a plot at unit scale. Suppose the slow state is the desired observable. One forward Euler step of 0.1 s gives $u = 0.9$ but $v = -99$. The method becomes unstable even though the desired slow response is well resolved. To damp the fast mode explicitly, h must stay below 0.002 s whenever that mode can be excited.

Backward Euler with $h = 0.1$ gives

$$u = 1/1.1 = 0.90909,$$

$$v = 1/101 = 0.00990.$$

Trapezoidal gives

$$u = 0.95/1.05 = 0.90476,$$

$$v = -49/51 = -0.96078.$$

The second-order rule is especially accurate on the slow mode but retains a conspicuous oscillatory fast component. Neither the maximum order nor a broad stability region alone identifies the right integrator. The observable, acceptable transient error and stiffness all enter the decision.

Trace the cost back to the model

A diffusion equation discretised on a spatial mesh produces modes with rates up to order D/h_x^2. Halving spatial spacing roughly quadruples the magnitude of the fastest eigenvalue. An explicit time step chosen for a coarse grid can therefore fail after spatial refinement, even if the initial physical signal looks smooth. Small high-frequency components from boundary changes, rounding or nonlinear interactions may excite modes absent from an idealised initial state.

Implicit stability shifts work into linear or nonlinear systems inside each time step. An oversized implicit step can miss a forcing pulse altogether, or produce a poor slow-phase estimate despite decaying every homogeneous mode. Compare the output on two time grids, distinguish the time required to resolve a source from the time required merely for stability, and inspect at least one fast component when interpreting unexpected oscillations. The distinction is visible in a reaction where exact species totals coexist with substantial errors in the individual states. Compare reaction histories under Euler and RK4.

Explain why a tiny initial fast component can still matter

Take the same two-rate model but start with $v_0 = 10^{-12}$ and apply forward Euler with $h = 0.01$. The physical fast mode should disappear almost immediately. The numerical factor is $1 - 10 = -9$, so after twelve steps its magnitude is $9^{12}10^{-12} \simeq 0.282$. An initial disturbance that was invisible in the output has become comparable with the desired slow signal in little physical time.

If the fast variable started exactly at zero in a perfectly decoupled model, exact arithmetic would keep it zero. That special trajectory would give a misleading sense of stability. Roundoff, forcing or a weak coupling can generate a nonzero fast component at later stages. A robust experiment deliberately perturbs fast directions and examines whether numerical damping agrees with the model under those perturbations. This is especially relevant when a mesh refinement adds high-frequency modes that were absent from the coarser spectrum.

Separate damping of a perturbation from tracking a moving state

A useful stiff test has an exact slow path $g(t)$ embedded in $y' = \lambda(y - g(t)) + g'(t)$, with $\lambda < 0$. Subtracting g gives

$$(y - g)' = \lambda(y - g).$$

An initial condition on the path stays on it, whereas an off-path disturbance decays exponentially. A numerical method must therefore address two different tasks: damp erroneous components perpendicular to the slow path and follow the path as it moves. A homogeneous stability test measures only the first.

Apply backward Euler and define $e_n = y_n - g(t_n)$. Subtract $g(t_{n+1})$ after the implicit update. The error satisfies

$$(1 - h\lambda)e_{n+1} = e_n + hg'(t_{n+1}) - [g(t_{n+1}) - g(t_n)].$$

The last two terms are the error in approximating the change of g using its new endpoint derivative. Even with $e_n = 0$, this forcing can create error at the next time. The factor $1/(1 - h\lambda)$ damps both old error and newly created defect. Increasing stiffness changes their balance, while changing the time step alters both the source and its propagation.

Track a quadratic path with a fast relaxation

Take $g = t^2,\ \lambda = -100$ per second, $h = 0.1$ s, $e_0 = 0$. Since $h(2t_{n+1}) - (t_{n+1}^2 - t_n^2) = h^2$, the recurrence is

$$e_{n+1} = (e_n + 0.01)/11.$$

The first error is 0.00090909, so the first computed state is 0.01090909 when the exact value is 0.01. The second error is approximately 0.00099174. Solving the scalar recurrence shows that the error tends to 0.001, the fixed point of $11e = e + 0.01$. Stability does not force it to vanish because each step creates a new tracking defect.

For comparison, trapezoidal integration of the slow-path derivative integrates the linear function $g' = 2t$ exactly over each interval. Substituting the exact nodal values $g(t_n)$ therefore satisfies its full discrete equations, so an exactly initialized trajectory stays on this particular path. However, a small off-path

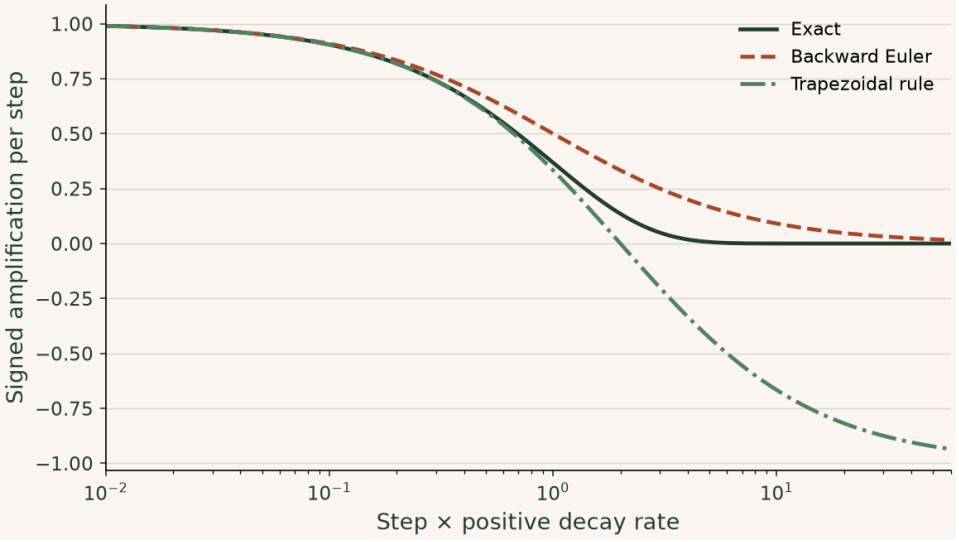
perturbation is multiplied per step by $(1-5)/(1+5) = -2/3$, whereas backward Euler multiplies it by $1/11$. Thus one method is exact on the specified moving path and the other damps unrelated fast perturbations more rapidly.

A numerical test that initialises precisely on g and uses only this quadratic path could make the trapezoidal method appear flawless. Add a known small perturbation and measure its sign and amplitude separately. Conversely a test with $g = 0$ detects damping but cannot expose a tracking defect. Both controlled initialisations are needed to tell whether an observed error arises from resolving the moving forcing or dissipating the artificially fast component.

A final scaling experiment holds the stiffness fixed and halves the step. The long-time tracking error in this quadratic case is $h/|\lambda|$, obtained by cancelling the fixed-point terms in the recurrence. With $h = 0.05$ and the same relaxation rate it becomes 0.0005. A one-step error alone, evaluated from the exact state on each grid, would not reveal the same accumulation mechanism. Report whether an observed order concerns a local defect, a settled tracking error or an initial fast transient.

Separate boundedness from physical decay in a stiff mode

For $y' = -ky$ with positive initial state, an unresolved fast mode should disappear on the reporting timescale. Compare the actual signed factors. A magnitude-only graph would conceal a spurious sign change.



At $hk = 10$ the exact factor is 4.54×10^{-5}, backward Euler gives $1/11$, and trapezoidal gives $-2/3$. The trapezoidal solution remains bounded but an alternating fast component can contaminate physically nonnegative output.

▶ **Read values from the plot**

Change parameters · Separate boundedness from physical decay in a stiff mode

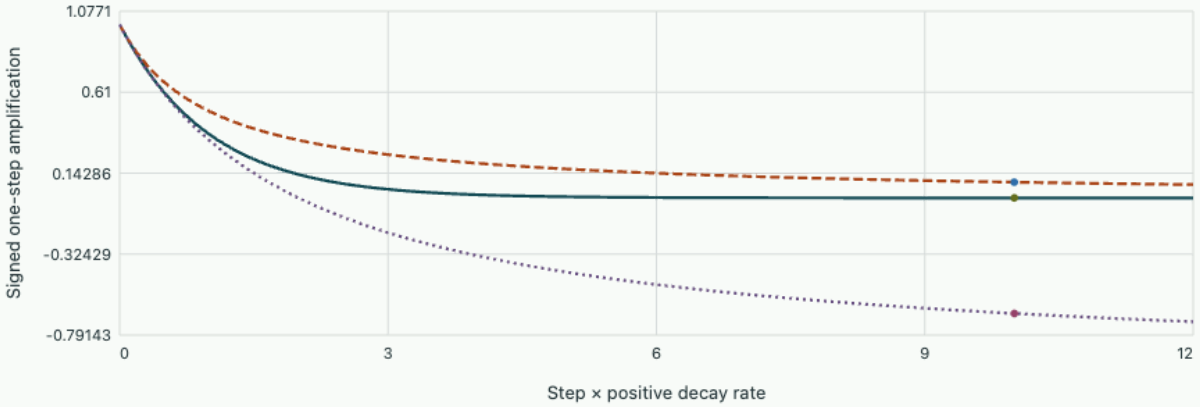
Change one input at a time and compare the recalculated curves with the argument above. The browser evaluates this small model locally; the accompanying numerical values state its assumptions and reference checks. Reset restores the reference inputs, and the data can be downloaded as CSV.

Open this calculation with adjustable inputs

Try the calculation

Maximum step × rate **12**

Selected step × rate **10**



Exact exponential · Backward Euler · Trapezoidal · Selected exact amplification · Selected backward Euler · Selected trapezoidal

At selected hk = 10: exact 0.0000454, backward Euler 0.090909, trapezoidal -0.66667. Negative amplification creates an alternating unresolved mode.

Adaptive steps and event locations

Numerical ordinary differential equations

Estimate local integration error with a low–high order pair, choose a dimensionally consistent tolerance and bracket a zero of an event function in time.

FROM FIRST PRINCIPLES

Obtain a step controller from the measured error of two approximations

Take a proposed step h twice from the same current state, once using an order-p formula and once using a more accurate formula whose leading defect is one order smaller. For smooth data and sufficiently short steps the two values have a common exact target y_*. Their expansion gives

$$\begin{aligned} y_{\text{low}} &= y_* + Ch^{p+1} + O(h^{p+2}), \\ y_{\text{high}} &= y_* + O(h^{p+2}), \\ d &= y_{\text{low}} - y_{\text{high}}. \end{aligned}$$

The difference estimates the leading local error of the lower-order formula. It has the units of the state. Scale each component by an absolute tolerance with those units plus a dimensionless relative tolerance times the current magnitude:

$$r = \max_i \frac{|d_i|}{\text{atol}_i + \text{rtol}\,\max(|y_{n,i}|, |y_{\text{high},i}|)}.$$

If $r \le 1$ the proposed step passes this error estimate. Provided its leading coefficient varies slowly, a new step should change the estimated error by the ratio of step sizes raised to $p+1$. Targeting unit scaled error gives

$$h_{\text{new}} = s\,h\,r^{-1/(p+1)}, \qquad 0 < s < 1.$$

The safety factor leaves room for variation in the neglected terms. A real implementation bounds growth and handles $r = 0$ separately. A discontinuity or an event between endpoints can violate the smooth-error assumption even when the endpoints appear accurate.

Background · Embedded error estimate

$$e = y_{n+1}^{[p]} - y_{n+1}^{[p-1]}$$

Two formulas share stages and estimate a local error. The controller also needs absolute and relative scales for each component.

DERIVATION AND APPLICATIONS

How component tolerances become a single step decision

A time step acceptable for one component can be inaccurate for another whose magnitude or units differ. Consider an embedded pair of methods sharing their internal stages and producing two endpoint estimates y_i^H, y_i^L. Let $d_i = y_i^H - y_i^L$. For a smooth right-hand side and a pair whose lower formula has order p, this difference is normally dominated by an h^{p+1} local term. Assume all components use the same step but have physically appropriate absolute scales $a_i > 0$ and relative tolerances r_i.

Convert each difference to a dimensionless ratio using the larger recent endpoint magnitude. An infinity norm requires every component to meet its own allowance:

$$s_i = a_i + r_i \max(|y_{i,n}|, |y_i^H|),$$
$$E = \max_i |d_i|/s_i.$$

With approximately constant leading error coefficient and scale over the proposed retry, the new norm is predicted from the power law. Solving for a target near one and including a safety factor $\nu < 1$ gives

$$E_{\text{new}} \simeq E(h_{\text{new}}/h)^{p+1},$$
$$h_{\text{new}} = \nu h E^{-1/(p+1)}.$$

Accept and advance with the higher-order estimate only when $E \le 1$ under this policy. If a component is converted to another unit by a factor c, both its measured difference and absolute tolerance must scale:

$$|cd_i|/(|c|a_i + r_i|c| \max |y_i|) = |d_i|/s_i.$$

The decision is therefore invariant under consistent unit conversion. A bounded growth factor handles zero estimated differences and prevents a single unusually small defect from triggering a huge next step. This controller targets local agreement; its acceptance is not a proof of a global error bound.

WORKED EXAMPLE

Decide a two-component retry

Let an order-two embedded lower estimate at $h = 0.100$ s produce differences 0.000800 m and 0.00600 m/s. The chosen component scales are 0.00100 m and 0.00200 m/s. Their ratios are 0.800 and 3.00. Reject the step and take $\nu = 0.9$, giving $h_{\text{new}} = 0.9(0.100)3^{-1/3} = 0.0624$ s. The leading-error model predicts a new norm $0.9^3 = 0.729$.

SECOND APPLICATION

Compare a componentwise limit with an averaged one

Suppose a system has 100 components, with 99 normalized errors zero and one equal to 3.00. Its root-mean-square norm is $\sqrt{9/100} = 0.300$, so an RMS policy would accept. The infinity policy rejects because one variable missed its tolerance threefold. An RMS policy can be appropriate when errors are distributed, while a strict component requirement demands the maximum or a separate check for a critical variable.

Ask different parts of a trajectory for different resolutions

A cooling transient changes rapidly at first and slowly near equilibrium. A uniform step small enough for the beginning can waste effort later. An adaptive controller proposes a step, estimates its error and accepts or rejects it. To estimate the error without the exact solution, two formulas can share function evaluations. With Euler and Heun, compute $k_1 = f(t_n, y_n)$, predict $y_E = y_n + hk_1$, then evaluate $k_2 = f(t_n + h, y_E)$ and $y_H = y_n + h(k_1 + k_2)/2$.

Their difference $y_H - y_E = h(k_2 - k_1)/2$ estimates the leading local error of the lower-order Euler step. It normally scales as h^2 in smooth regions. An embedded difference does not equal the local error of the higher-order accepted value; the higher-order value generally has a smaller error starting from accurate data. Nor does a local tolerance directly bound the error at the end of the whole run. Previous accepted steps already carry accumulated trajectory error.

Scale the estimate and choose a step

For a scalar use $w = \text{atol} + \text{rtol}\max(|y_n|, |y_H|)$ and the dimensionless estimate $E = |y_H - y_E|/w$. Accept when $E \le 1$. If the next difference scales as h^2, a proposed change is $h_{new} = 0.9hE^{-1/2}$. The factor 0.9 leaves a margin for changes in curvature. Practical implementations limit the ratio between successive steps and handle a zero difference by bounded growth. A rejected step is retried from the same original state and time.

For $y' = -2y,\ y_n = 1,\ h = 0.1$, Euler gives 0.8 and Heun gives 0.82, a difference 0.02. With absolute tolerance 0.005 and zero relative tolerance for this demonstration, $E = 4$. The step is rejected and a revised step is

$$0.9(0.1)/2 = 0.045.$$

Re-evaluation from the original state gives

$$y_E = 0.91,\ y_H = 0.91405,$$

with difference 0.00405 and $E = 0.81$. The accepted state corresponds to time 0.045, not 0.145. Advancing time on rejection would skip the unresolved portion.

Background · What tolerance should a solver apply near zero concentration?

Two state components can be reported in very different units. An error of 10^{-6} metres may be harmless for a position and an error of 10^{-6} moles may dominate a trace species. A common error test divides each component by its own physical absolute tolerance plus a relative contribution, for instance

$$e_i^{\text{scaled}} = \frac{e_i}{a_i + r \max(|y_{n,i}|, |y_{n+1,i}|)}.$$

At a zero crossing, $a_i > 0$ keeps the requested absolute accuracy finite. If concentration units change from mol/L to mmol/L, multiply the matching absolute tolerance by 1000 along with the numerical concentrations. The relative tolerance is dimensionless. Leaving the old absolute tolerance in place silently imposes a thousandfold tighter accuracy requirement near zero.

Estimate error using one full step and two half steps

A solver without an embedded pair can estimate its local error by applying the same order-p stepper twice from a common starting state: take one full step for Y_c, and separately two successive half steps for Y_f. Suppose the local error for one smooth step, numerical minus exact, is $Ch^{p+1} + O(h^{p+2})$. The two half-step errors accumulate to $2C(h/2)^{p+1} + O(h^{p+2})$ because propagating the first over the short second half changes its leading coefficient only at higher order. Thus Y_f carries approximately 2^{-p} of the full-step error.

Subtracting the two approximations gives an estimate of the correction to Y_f: $(Y_f - Y_c)/(2^p - 1)$. Adding it gives a Richardson-extrapolated value. The exponent in a local step controller remains $1/(p+1)$ because the correction estimate scales as h^{p+1}. The denominator $2^p - 1$ comes from taking *two* shorter steps to the same endpoint. Confusing that denominator with a single half-step local error factor introduces a systematic error in the estimate.

Take Heun's second-order method on $y' = -y$, starting from one with $h = 0.4$. The full-step amplification is

$$1 - 0.4 + 0.4^2/2 = 0.68.$$

Each half step has amplification 0.82, so $Y_f = 0.6724$. The fine correction is

$$(0.6724 - 0.68)/3 \approx -0.0025333.$$

The extrapolated value is 0.6698667. The exact endpoint $e^{-0.4} \approx 0.6703200$ requires correction -0.0020800 to Y_f. The estimated sign and scale are correct while higher-order terms are still visible. Discontinuous forcing, a skipped event or inconsistent internal solver tolerances can spoil this comparison even when the two output times agree.

Locate events between stored states

An event is a zero of $g(t, y(t))$, such as crossing a chosen temperature threshold. When its endpoint values have opposite signs, continuity brackets at least one event. Use a dense approximation to the trajectory or refine the time interval to locate the crossing. A linear interpolation between endpoints may be adequate for a smooth signal away from a turning point. If a value changes by a small amount near a flat crossing, a small field error can produce a much larger error in event time.

For $y = e^{-t}$, readings at 0.6 and 0.8 are 0.54881 and 0.44933. A straight interpolation to 0.5 gives $t \simeq 0.69813$, while the exact time is $\ln 2 = 0.69315$. The discrepancy arises from the curved trajectory even if the endpoints are exact. A tangential event such as $(t - 1)^2 = 0$ never changes sign, and two crossings in one step can also be missed. Declare the crossing direction, the required timing accuracy and the physical effect of reaching it. After an event that changes the model, start the next segment using the event state and preserve the time at which the transition actually occurred.

Test the event interpolant with a known direction

At a crossing where a scalar signal has nonzero derivative, an approximate amplitude error δy produces a first-order timing error $\delta t \simeq -\delta y / y'$. For $e^{-t} = 0.5$, the derivative is -0.5 per second. A signal that is 0.002 too high near the crossing therefore appears to reach the threshold about 0.004 s late. This local estimate converts a solver tolerance on a state into a physically interpretable timing requirement.

A source that switches precisely at the threshold makes the event time part of the subsequent model. A timing error then affects all later states even if ordinary integration error within each regime is small. When comparing grids, compare both the located event times and the post-event states at shared physical times. Two traces that agree before a switch can diverge later purely because one entered the new law sooner.

Relate state accuracy to event-time accuracy

Let an event be defined by $G(t, y(t)) = 0$ and write the scalar event signal as $g(t)$. Its derivative along a smooth trajectory is $g' = G_t + G_y f$. At a simple crossing $g'(t_\star) \neq 0$, an error δg in evaluating the signal displaces the reported time by approximately $-\delta g / g'(t_\star)$. More rigorously, if the true signal derivative has magnitude at least m throughout the final bracket, the mean value theorem bounds time error by the event-signal error divided by m, in addition to any error left by the scalar root search.

Thus an event locator can solve its interpolated scalar equation to a tiny tolerance while returning a poor physical time. Errors in the interpolated state, uncertain event threshold and flat slope all propagate into

the time estimate. At a tangential contact the derivative vanishes and the first-order estimate ceases to apply. Two simple events can merge into one contact or disappear altogether under a perturbation that would merely shift a steep isolated crossing by a small amount.

Time two crossings around the top of a trajectory

Take vertical position $y(t) = 10t - 5t^2$ metres, with time in seconds. Its maximum is 5 m at one second. A detector at 4.8 m records two crossings, since $5 - 5(t-1)^2 = 4.8$ gives $t = 1 \pm 0.2$, namely 0.8 and 1.2 s. The respective vertical velocities are $+2$ and -2 m/s. A small threshold increase of 0.01 m therefore moves the rising crossing about 0.005 s later and the falling crossing about 0.005 s earlier.

Check the exact effect at 4.81 m. The half-separation becomes

$$\sqrt{0.19/5} = \sqrt{0.038} \approx 0.194936\,\mathrm{s}.$$

The actual time shifts have magnitude approximately 0.005064 s, so the local slope estimate is already informative. At a threshold of 4.999 m the speed is instead only $10\sqrt{0.001/5} \approx 0.14142$ m/s. Another increase of 0.01 m moves the threshold above the maximum and removes both crossings. Applying the previous linear timing estimate at this near-tangent setting would predict a displacement for an event that no longer exists.

A direction flag distinguishes the rising and falling events if each is bracketed, but it cannot recover two crossings hidden inside one large step with endpoint signals of the same sign. In this model the velocity event $y' = 0$ identifies the turning point and splits the time interval into monotone pieces. On each piece check the threshold against its attainable position range before starting a time root search. This strategy uses knowledge of the dynamics to ensure the event actually exists.

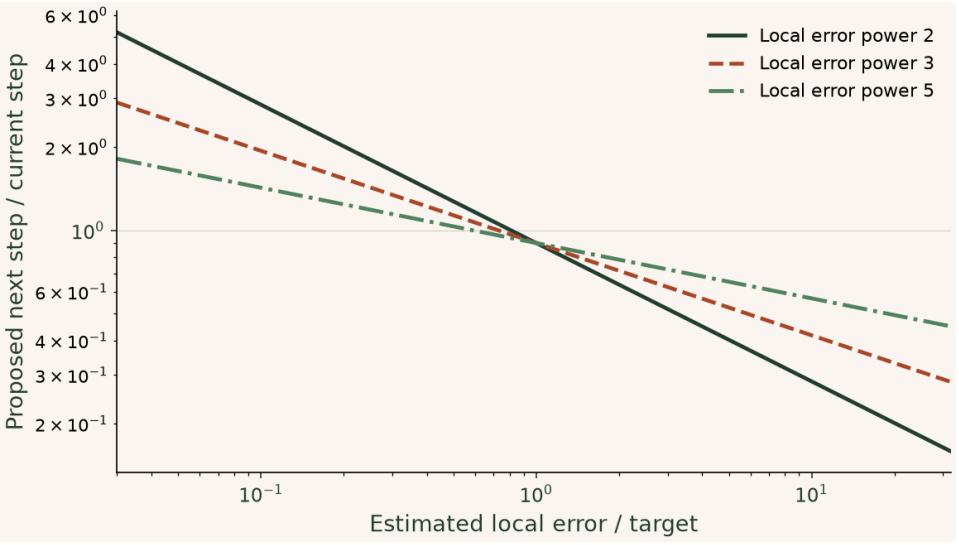
For example, a single step from 0.7 to 1.3 s gives position 4.55 m at both endpoints and therefore no sign difference relative to the 4.8 m detector. Inserting the peak at 1 s gives a positive signal there and immediately produces two valid brackets. Decreasing only the scalar root tolerance would have found neither event on the original unsplit step.

Choose the next step when a local error estimate misses its target

Suppose the local estimator behaves as $E \simeq Ch^p$. Matching the next error to tolerance by rescaling h, then applying a safety factor 0.9, predicts

$$h_{\mathrm{next}}/h = 0.9(\mathrm{tol}/E)^{1/p}.$$

Treat p as the power of this estimator.



With E eight times the target, a cubic estimator requests a new step 0.45 times the old one. The formula proposes a length; an acceptance test must separately decide whether to retry the current step. Sudden events can invalidate the assumed scaling.

▶ **Read values from the plot**

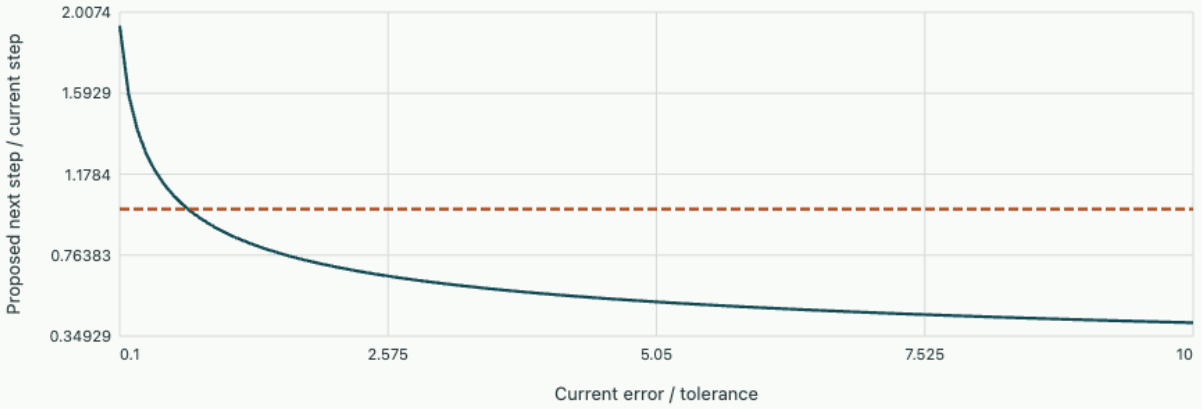
Change parameters · Choose the next step when a local error estimate misses its target

Change one input at a time and compare the recalculated curves with the argument above. The browser evaluates this small model locally; the accompanying numerical values state its assumptions and reference checks. Reset restores the reference inputs, and the data can be downloaded as CSV.

Open this calculation with adjustable inputs

Try the calculation

Estimator power p	3
Safety factor	0.9
Selected error / tolerance	8



Power-law step proposal No step change

Selected ratio requests hnext/h = 0.45. Accept/reject the current step separately; this rule assumes the estimator scales as h^3.

Tangent equations and adjoint outputs

Numerical ordinary differential equations

Compute how an observed final state changes when an initial state or a model parameter is perturbed.

FROM FIRST PRINCIPLES

Differentiate the trajectory that your instrument measures

Let a scalar parameter θ enter both a smooth rate law $y' = f(t, y, \theta)$ and the initial data $y(0) = g(\theta)$. Compare two solutions at the same physical time. Their difference divided by a small parameter increment tends to the sensitivity $S = \partial y / \partial \theta$. Differentiating the rate law at fixed time supplies a term from the moving state and a direct parameter term:

$$\dot{S} = f_y S + f_\theta,$$
$$S(0) = g'(\theta).$$

For a scalar terminal measurement $J = \psi(y(T), \theta)$ the chain rule gives

$$J' = \psi_y S(T) + \psi_\theta.$$

When many parameters enter a single output, a backward equation can reuse their common output weighting. Choose $\lambda(T) = \psi_y^T$ and $\dot{\lambda} = -f_y^T \lambda$. Then the terms containing the unknown sensitivity cancel in the derivative of their scalar product:

$$\frac{d}{dt}(\lambda^T S) = \lambda^T f_\theta,$$
$$J' = \psi_\theta + \lambda(0)^T g'(\theta) + \int_0^T \lambda^T f_\theta \, dt.$$

Evaluate all derivatives on the original trajectory. For $y' = -\theta y$ with fixed $y(0)$, direct differentiation gives $S(T) = -T y(T)$, the sign expected when increasing the decay rate. A parameter-dependent observation time would add an endpoint term.

DERIVATION AND APPLICATIONS

Differentiate the numerical update that produced the observed output

A gradient used in fitting a parameter should agree with the discretized objective actually evaluated. Consider explicit Euler on a scalar equation $y' = f(y, \theta, t)$ with fixed, parameter-independent time steps and initial state. A single final observation defines $J = \Phi(y_N, \theta)$. Treat the computed states as a chain of differentiable maps rather than replacing their dependence by that of the exact differential equation.

For the step $F_n(y, \theta) = y + hf(y, \theta, t_n)$ a small perturbation satisfies

$$\begin{aligned} \delta y_{n+1} &= A_n \delta y_n + B_n \delta\theta, \\ A_n &= 1 + h f_y, \\ B_n &= h f_\theta \end{aligned}$$

with derivatives evaluated on the stored numerical trajectory. Start from $\delta J = \Phi_y \delta y_N + \Phi_\theta \delta\theta$ and substitute the last step. Repeating this substitution backward defines a weight on each earlier state:

$$\begin{aligned} \lambda_N &= \Phi_y, \\ \lambda_n &= A_n \lambda_{n+1}. \end{aligned}$$

Each stage contributes its local parameter derivative weighted by the future effect of that stage. For a fixed initial condition the total derivative is

$$dJ/d\theta = \Phi_\theta + \sum_{n=0}^{N-1} \lambda_{n+1} B_n.$$

With vector states the reverse update uses the transpose Jacobian. An initial condition depending on θ adds $\lambda_0 \, \partial_\theta y_0$. The resulting gradient is exact for the discrete map up to arithmetic precision, though it can differ from the gradient of the continuous solution. Adaptive control decisions need special care if differentiating through changes in the time grid.

WORKED EXAMPLE

Differentiate a two-step fitted decay

Let $y' = -\theta y$, $y_0 = 2.00$, $h = 0.250$, $\theta = 1.00$, and $J = (y_2 - 1)^2/2$. The Euler states are 1.50 and 1.125, giving $\lambda_2 = 0.125$, $\lambda_1 = 0.09375$. The local parameter derivatives are -0.500 and -0.375. Thus

$$J' = 0.09375(-0.500) + 0.125(-0.375) = -0.09375.$$

Increasing the decay rate locally reduces the positive final residual.

SECOND APPLICATION

Compare with the continuous sensitivity at the same time

At $t = 0.500$ the exact value is $2e^{-0.5} = 1.21306$ and its derivative is $-0.500(1.21306) = -0.60653$. The continuous objective therefore has derivative -0.1292. That difference is expected at this coarse step. A centred parameter perturbation of the two-step numerical objective should approach -0.09375, providing a direct implementation check without assuming that the time integrator is already accurate.

Differentiate the evolution

Let $y' = f(t, y, \alpha)$ with smooth f and $y(0) = y_0(\alpha)$. Assume that the solution exists and depends differentiably on α throughout the observation window. Differentiating the equation along the solution gives

$$S' = f_y S + f_\alpha,$$
$$S(0) = y_0'(\alpha),$$

where $S = \partial y / \partial \alpha$. The first term propagates state changes through the dynamics and the second injects the direct parameter effect. Losing the initial derivative gives the wrong sensitivity whenever a parameter also changes the initial preparation.

For an output $Q = \phi(y(T), \alpha)$, the chain rule gives

$$Q_\alpha = \phi_y S(T) + \phi_\alpha.$$

Each parameter requires its own tangent column. Many observations can then reuse the propagated columns. Scaling matters because a derivative with respect to milliseconds is numerically different from the derivative with respect to seconds. Multiplying by a stated parameter perturbation converts the derivative back into a predicted observation change that can be checked directly.

A rate parameter and a final measurement

Take $y' = -\alpha y,\ y(0) = 2$. Then

$$S' = -\alpha S - y,\ S(0) = 0.$$

Multiplying by $e^{\alpha t}$ leaves $(e^{\alpha t} S)' = -2$, hence $S = -2te^{-\alpha t}$. At $\alpha = 0.5$ per second and $T = 2$ s the sensitivity of final amplitude is $-4/e \approx -1.47152$ amplitude units per inverse second. Raising the decay rate by 0.01 per second predicts a decrease of 0.014715. The exact change is $2e^{-1}(e^{-0.02} - 1) \approx -0.01457$; the discrepancy is quadratic in the parameter perturbation.

Testing a tangent code by differences requires controlling both the ODE integration error and subtraction error. Use several perturbation sizes. With a forward difference the truncation error normally decreases linearly at first, until numerical noise in subtracting two nearly equal outputs dominates.

Run one dual equation for a scalar output

Define a column vector z backward in time by $-z' = f_y^T z,\ z(T) = \nabla_y \phi$. Differentiating $z^T S$ cancels the two terms containing f_y. Integrating what remains gives

$$Q_\alpha = \phi_\alpha + z(0)^T y_0'(\alpha) + \int_0^T z^T f_\alpha \, dt.$$

For the preceding final-amplitude observation $z = e^{-\alpha(T-t)}$. Its product with $f_\alpha = -y$ is the constant $-2e^{-\alpha T}$, whose integral reproduces the tangent result. One adjoint trajectory can provide gradients with respect to many parameters for this single output. Its coefficients require the forward trajectory at earlier times, so storing or reconstructing that trajectory is part of the numerical cost. If the desired derivative is of the discrete numerical update itself, differentiating its specific steps and interpolation choices can matter at finite time resolution.

A cancellation at a particular observation time

Compare observing the final amplitude with observing the lost amplitude $Q = y(0) - y(T)$. If the rate is the only parameter, changing the observation reverses the sign of the computed rate sensitivity. If instead the unknown is the common initial amplitude, the direct derivative of $y(0)$ is one and the propagated derivative is $e^{-\alpha T}$. The final derivative of the loss is $1 - e^{-\alpha T}$, tending to zero as the observation window shrinks. Omitting the direct term would predict a large negative initial sensitivity for a loss that is identically zero at zero time.

Differentiate the time at which a model switches

If a trajectory switches dynamics at a threshold, the event time itself depends on the parameter. Let a pre-event solution $y^-(t, p)$ meet $G(t, y, p) = 0$ at $\tau(p)$. Denote by S^- its parameter derivative at fixed time, evaluated on the nominal event. Differentiating the event equation gives

$$\tau_p = -\frac{G_y S^- + G_p}{G_t + G_y f^-}.$$

The denominator is the rate at which the trajectory crosses the threshold. This formula requires a transversal crossing and a locally unchanged event sequence. At a grazing contact the derivative may diverge or cease to exist, and a perturbed trajectory can encounter a different number of events.

Assume the state itself is continuous while the right-hand side changes from f^- to f^+. The total parameter change of the event state is $S^- + f^- \tau_p$. But the post-event tangent at a fixed clock time must remove the amount of post-event evolution gained or lost by shifting the start time, giving $S^+ = S^- + (f^- - f^+)\tau_p$. This jump in a derivative of a continuous state is sometimes surprising. Comparing trajectories taken at the same physical clock time is different from comparing their states at their respective event times.

Estimate the final position after a velocity switch

Let a cart start at zero and travel at constant positive speed p until it reaches position one. It then travels at fixed speed 0.5. Observe its position at $T = 2$, and examine $p = 1$, so the event occurs at $\tau = 1$. The pre-event position is pt and its tangent at fixed time is

$$S^- = t = 1.$$

Using $G = y - 1$ gives

$$\tau_p = -1/p^2 = -1.$$

Since $f^- = 1$ and $f^+ = 0.5$, the tangent immediately after the switch becomes

$$1 + (1 - 0.5)(-1) = 0.5.$$

Subsequent evolution at the fixed speed adds no further parameter source.

Check directly: as long as the event precedes the observation, $y(T, p) = 1 + 0.5(2 - 1/p)$. Its derivative at one is indeed 0.5. At $p = 1.01$ the exact final position is approximately 1.5049505 instead of 1.5. The first-order prediction is 1.505. Carrying the pre-event tangent unchanged through the switch would predict 1.51, nearly doubling the response. Resetting it to zero because the event position is fixed would miss the change entirely.

At $p = 0.5$ the switch is exactly at the observation endpoint. Parameters below this value give $y(T) = 2p$, whereas parameters above it use the switched expression. Both one-sided derivatives happen to equal two here because the two velocities match at this specific threshold value. Their second derivatives differ. Changing the post-event speed to 0.25 makes even the first derivatives disagree at the boundary. An apparent smooth result in one carefully chosen switching case cannot justify ignoring event timing throughout a parameter study.

A finite-difference sensitivity test should consequently check which events occurred on both perturbed trajectories. Agreement in total simulated duration and integrator tolerance is insufficient if the jump sequence changed. Recording event times together with the reported output makes it possible to distinguish an error in the reset law from a derivative that has ceased to exist across the selected parameter interval.

Propagate a reset depending on the event location

If a switching event resets the state by $y^+ = R(t, y^-, p)$, differentiate the reset along the moving event before returning to a fixed clock time. The result is

$$S^+ = R_y S^- + R_p + (R_t + R_y f^- - f^+)\tau_p.$$

This reduces to the continuous-state formula when R is the identity. The occurrence time must still come from the pre-event trajectory, since the post-event state generally does not satisfy the original trigger condition.

Modify the cart example so the trigger position is $H(p) = 1 + 0.2(p - 1)$, and the indicated position is reset to $R = 0.3y^- + 0.2$ before it increases at speed 0.5. At $p = 1$, $\tau = H/p = 1$ and $\tau_p = 0.2 - 1 = -0.8$. The pre-event fixed-time tangent is one. The reset gives

$$S^+ = 0.3 + (0.3 - 0.5)(-0.8) = 0.46.$$

The indicated final position at time two is $0.3H + 0.2 + 0.5(2 - H/p)$, exactly one at the nominal parameter. Differentiating this expression gives $0.3(0.2) - 0.5(0.2 - 1) = 0.46$, independently verifying the sign and the event-time correction. Simply differentiating the reset map at the nominal crossing position would give 0.3 and lose the varying length of the post-event trajectory.

Change parameters · Explore the worked calculation

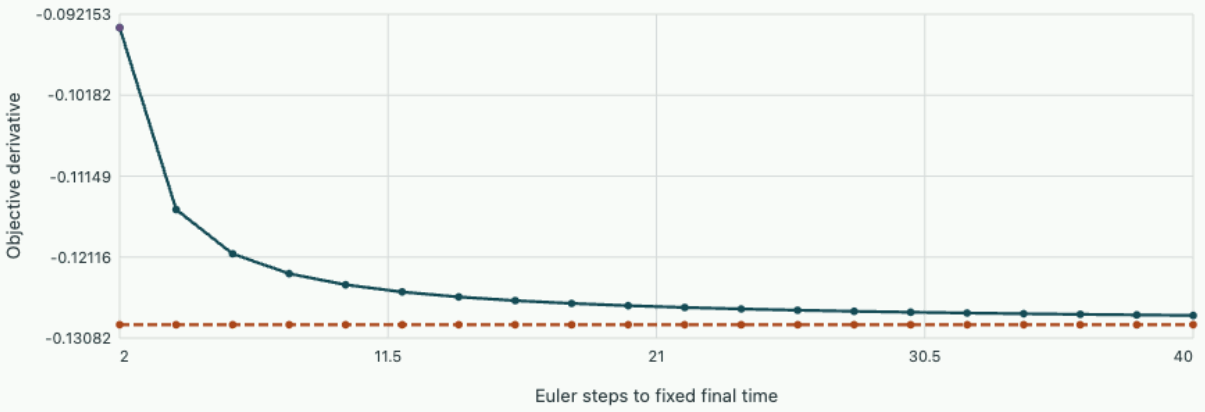
Change one input at a time and compare the recalculated curves with the argument above. The browser evaluates this small model locally; the accompanying numerical values state its assumptions and reference checks. Reset restores the reference inputs, and the data can be downloaded as CSV.

Open this calculation with adjustable inputs

Try the calculation

Decay parameter θ 1

Euler steps to t = 0.5 2



Exact derivative of discrete objective | Continuous objective derivative | Chosen grid

At 2 Euler steps the derivative is -0.09375 versus continuum -0.12923. A correct discrete adjoint differentiates the numerical trajectory actually computed.

How two decaying modes produce a growing signal

Numerical ordinary differential equations

Solve a triangular system exactly, derive the energy growth criterion and separate a real transient from numerical instability.

A coupling hidden by the eigenvalues

Let $y_1' = -y_1 + K y_2,\ y_2' = -2y_2$, with time in scaled units. The eigenvalues are –1 and –2 for every K. Their negative real parts guarantee asymptotic decay of this system. They do not say that $\|y(t)\|$ must decay monotonically. The second component acts as a temporary source in the first component before being damped away.

With initial state $(0, 1)$, $y_2 = e^{-2t}$. Multiply the first equation by e^t and integrate from zero to t. The result is

$$y_1 = K(e^{-t} - e^{-2t}).$$

The largest absolute excursion of the first component occurs when $2e^{-2t} = e^{-t}$, giving $t = \log 2$ and $y_1 = K/4$. At $K = 12$, a unit initial state generates a measured first component of 3.

Which initial directions grow at the initial instant?

Write $E = \|y\|_2^2/2$. Differentiation gives

$$E' = y^T S y,$$

with $S = (A + A^T)/2$. Here S has diagonal $(-1, -2)$ and off-diagonal $K/2$. A positive eigenvalue appears when the determinant $2 - K^2/4$ becomes negative. Thus some direction has positive instantaneous energy growth for $|K| > \sqrt{8}$. The specific initial vector above nevertheless has $E'(0) = -2$. Existence of a growing direction and immediate growth of every chosen initial condition are separate statements.

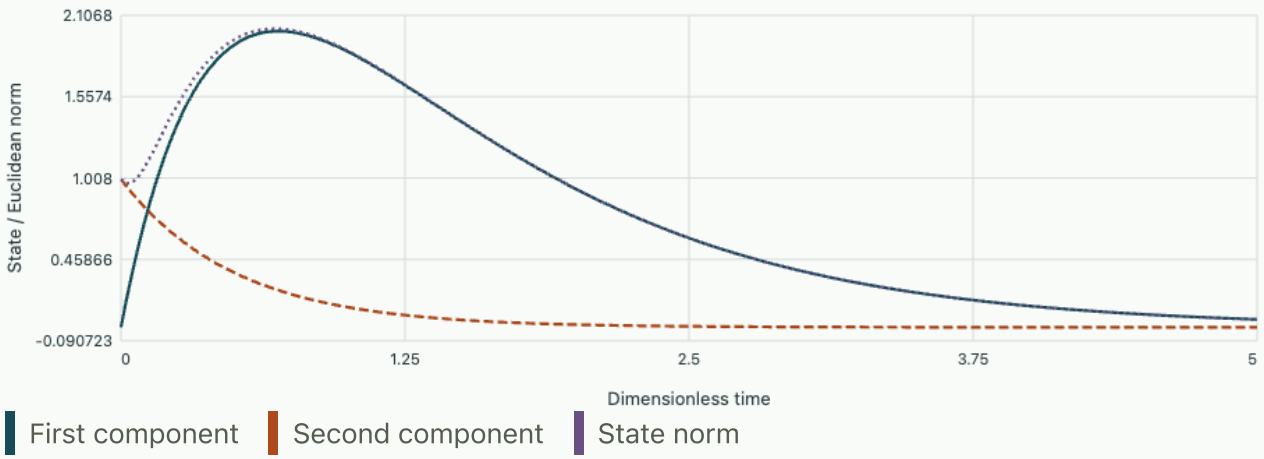
A numerical check with a false lead

Set $K = 8$. At $t = \log 2$ the exact components are $2, 1/4$. Forward Euler with a sufficiently small step will show an initial dip followed by a rise in the norm. Refining time steps changes the approximation to the peak but does not eliminate the physical growth. By contrast, using a coarse Euler step $h > 1$ puts the scalar mode –2 outside the forward-Euler stability interval and introduces genuine numerical growth.

Open this experiment online

Try the calculation

Coupling K 8



At t = ln 2 the components are (2, 0.25). Largest eigenvalue of the energy-rate matrix 2.5311.

The experiment plots the analytic components and norm. Use the signed coupling to reverse the first component. Explain why the norm is invariant under that sign change, and why a successful eigenvalue stability test alone cannot certify a bound on a short-time output.

Tutorial on transient amplification

Can a normal matrix with only negative real eigenvalues produce this Euclidean-norm amplification?

▼ **Worked solution**

A normal matrix is unitarily diagonalizable. Its evolution is a unitary change of independent decaying factors, giving a norm at most one for $t \geq 0$. Nonorthogonality of the damped directions matters here.

Change the initial state to $(1, 1)$ at $K = 6$. What is the initial energy slope?

▼ **Worked solution**

$$E' = -y_1^2 - 2y_2^2 + Ky_1y_2 = -1 - 2 + 6 = 3.$$

The instrumental first-component peak and the maximum total Euclidean norm generally occur at different times.

Locating a grazing event between time steps

Numerical ordinary differential equations

Derive a sensitivity formula for event time, then measure how a harmless error in a trajectory can create or remove two threshold crossings.

A time is an implicitly defined output

Suppose an integrator advances $y' = f(t, y)$ and an event occurs at $g(t, y(t)) = 0$. Along the solution the derivative of the event function is $\dot{g} = g_t + \nabla_y g \cdot f$. A perturbation δg in the computed event function shifts a simple event time by $\delta t \simeq -\delta g / \dot{g}$. This estimate separates trajectory accuracy from event-time accuracy. At a tangency the denominator vanishes and the formula cannot be used.

Background · Event transversality

$$\dot{g} = \partial_t g + \nabla g \cdot f$$

For a simple event this directional derivative is nonzero. Near grazing, a small state error can create a much larger error in the event time.

Resolve two crossings analytically

Take $g(t) = (t - t_0)^2 - d$, a local model for a turning point relative to a sensor threshold. For $d > 0$ the events are $t_0 \pm \sqrt{d}$. Their separation is $2\sqrt{d}$, and their slopes are $\pm 2\sqrt{d}$. For $d = 0$ there is only a tangent touch. If $d < 0$, the sensor is never triggered. An additive constant perturbation ϵ changes the effective parameter from d to $d - \epsilon$.

At $t_0 = 1$, $d = 0.01$ and $\epsilon = 0.001$, the first event moves from 0.9 to about 0.90513. The linear prediction is

$$0.001 / 0.2 = 0.005.$$

At $d = 0.0004$ the same perturbation removes both crossings entirely. Reporting more digits from a polished scalar rootfinder would not recover them.

What two endpoint samples fail to tell you

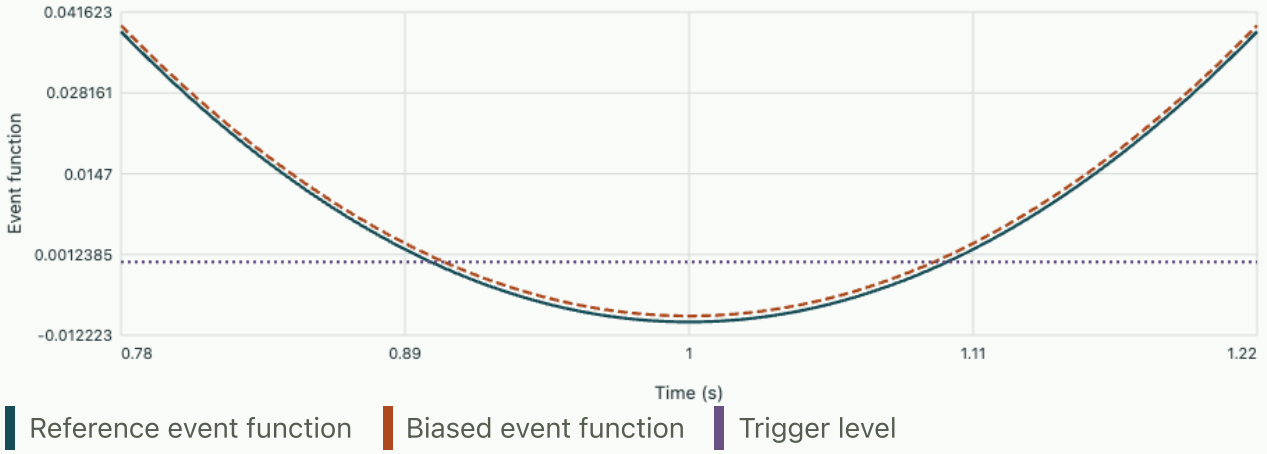
With $d = 0.0025$, a step from 0.9 to 1.1 has equal positive endpoint values of 0.0075 but crosses the threshold twice. Dense interpolation followed by an extremum check can detect this pair if the interpolant resolves the turning point. A sign-change test at accepted endpoints alone cannot. An interpolant is itself an approximation, so event feasibility should be tested against an error scale at the extremum.

Open this experiment online

Try the calculation

Threshold offset d **0.01**

Trajectory bias ε **0.001**



Reference: 0.9 s and 1.1 s. Perturbed: 0.90513 s and 1.0949 s.

Tutorial on event tolerances

For a simple crossing with slope magnitude 0.04 units/s, what error in the event function is consistent with a 1 ms time tolerance to first order?

▼ **Worked solution**

Multiply the slope by the desired time uncertainty. The event function should be accurate to about 4×10^{-5} units near the crossing.

A contact model needs the first event, even if it only touches the threshold and returns. How should the experiment be reported near grazing?

▼ **Worked solution**

Report the minimum event-function value and an estimated error on it, in addition to any time estimates. If zero lies within that uncertainty one cannot distinguish no crossing, tangency and a nearby pair from these data alone.

Eigenvalues, modes and spectral gaps

Numerical linear algebra

Estimate a dominant vibration or diffusion mode, explain the role of eigenvalue separation and use a residual to check a computed eigenpair.

FROM FIRST PRINCIPLES

Watch a power iteration select a mode

Start with a real symmetric matrix A and express a trial vector v_0 in its orthonormal eigenvectors u_j. Order the eigenvalues so $|\lambda_1| > |\lambda_2| \geq \cdots$ and suppose the starting component c_1 is nonzero. Applying A repeatedly multiplies each component by its own eigenvalue,

$$v_0 = \sum_j c_j u_j,$$

$$A^k v_0 = \lambda_1^k \left[c_1 u_1 + \sum_{j>1} c_j (\lambda_j/\lambda_1)^k u_j \right].$$

Normalizing after each application keeps vector lengths manageable without changing the relative size of the components. The largest competing component therefore shrinks relative to the first at a geometric rate. A scalar estimate of the eigenvalue comes from projecting the action back onto the trial vector,

$$\rho(v) = \frac{v^T A v}{v^T v}.$$

For an exact eigenvector the quotient equals its eigenvalue. Equal leading magnitudes can prevent selection of a single direction. A negative leading eigenvalue can make the displayed vector alternate sign even while its one-dimensional direction converges.

DERIVATION AND APPLICATIONS

Derive the eigenvalue shift produced by weak coupling

Two individually measured modes can move apart in frequency or stiffness when they are coupled. Model a real symmetric operator in an orthonormal basis by diagonal values α, β and off-diagonal coupling g. All three numbers must have the same units. A nonzero eigenvector exists only when the determinant of the shifted matrix vanishes,

$$(\alpha - \lambda)(\beta - \lambda) - g^2 = 0.$$

Set $m = (\alpha + \beta)/2$ and $d = (\alpha - \beta)/2$. The quadratic becomes $(\lambda - m)^2 = d^2 + g^2$, so the two eigenvalues and their separation are

$$\lambda_{\pm} = m \pm \sqrt{d^2 + g^2},$$
$$\lambda_{+} - \lambda_{-} = 2\sqrt{d^2 + g^2}.$$

Even when the uncoupled values coincide, a nonzero coupling leaves a gap $2|g|$. For well-separated diagonal entries, assume $\alpha > \beta$ and expand the square root for $|g| \ll \alpha - \beta$. The largest eigenvalue has the shift

$$\lambda_{+} - \alpha = \frac{g^2}{\sqrt{d^2 + g^2} + d} \simeq \frac{g^2}{\alpha - \beta}.$$

The rationalized form gives an accurate small shift without subtracting almost equal eigenvalues. The eigenvalue sum remains exactly $\alpha + \beta$ because the trace is unchanged. A mass-normalized vibration model typically returns squared angular frequencies, so any measured ordinary frequencies must first be converted to the variables of the operator.

WORKED EXAMPLE

Predict a shift away from a crossing

With $\alpha = 10$, $\beta = 6$ and $g = 0.5$, the upper shift is $0.25/(\sqrt{4.25} + 2) = 0.0616$; the leading approximation is 0.0625. The lower eigenvalue is 5.9384 and the upper is 10.0616. The opposite signs of the shifts preserve the trace of 16.

SECOND APPLICATION

Recover coupling and detuning from a measured gap

A second two-mode system has eigenvalues 19 and 25, giving a mean of 22 and half-gap of 3. An independent uncoupled calibration gives $|\alpha - \beta| = 4$, so $d^2 = 4$. The inferred coupling magnitude is

$$\sqrt{3^2 - 2^2} = \sqrt{5} = 2.24.$$

The eigenvalues do not determine the sign of g, which also changes when one basis vector is reversed.

What an eigenproblem measures

In an undamped mechanical model $M\ddot{u} + Ku = 0$, insert $u = v\cos(\omega t)$. Nonzero amplitudes require

$$Kv = \omega^2 M v.$$

The normal frequencies are square roots of generalised eigenvalues and the mode shapes are determined only up to scale and sign. For symmetric positive definite M, substituting $w = M^{1/2}v$ gives the ordinary symmetric operator $M^{-1/2}KM^{-1/2}$. This step explains why mass weighting matters when comparing shapes from two different meshes.

For an ordinary symmetric matrix A with orthonormal eigenvectors, a normalised trial vector $v = \sum c_i u_i$ has Rayleigh quotient $\rho = v^T A v = \sum_i \lambda_i c_i^2$. Hence it lies between the smallest and largest eigenvalues. Near an isolated mode its value can be much more accurate than the vector direction: shifting a small amount of amplitude into another orthogonal mode changes the quotient by a term proportional to the square of that amplitude.

Select a mode by repeated application

The power iteration repeatedly forms Av_k and normalises it. Before normalisation,

$$A^k v_0 = \sum c_i \lambda_i^k u_i.$$

If one eigenvalue has strictly largest magnitude and the initial vector overlaps its eigenspace, the ratio of any competing component to the leading one gains a factor $(\lambda_i/\lambda_1)^k$. Normalisation avoids overflow while preserving those ratios. A negative leading eigenvalue can cause sign changes between successive normalised vectors. Comparing their raw difference then misrepresents convergence.

For $A = \operatorname{diag}(4, 2)$ and a starting vector proportional to $(1, 1)^T$, the ratio of the second component to the first is $1/2$ after one multiplication and $1/4$ after two. A vector proportional to $(4, 1)^T$ has Rayleigh quotient $(4 \cdot 16 + 2)/17 = 66/17 = 3.88235$. After four steps the ratio is $1/16$, and the quotient is

$$1026/257 = 3.99222.$$

Check that the quotient error decreases roughly fourfold per additional step, although the directional error decreases only twofold.

Use shifts to find another part of the spectrum

Power iteration selects magnitude. To find a mode near a target σ, apply the inverse of $A - \sigma I$ implicitly by solving a linear system each iteration. The eigenvectors are the same and the eigenvalues are $1/(\lambda_i - \sigma)$. For eigenvalues 2 and 4 with $\sigma = 2.1$, the transformed magnitudes are 10 and about 0.526. The eigenvalue near 2 dominates for a suitable starter. A fixed shift permits factor reuse, but choosing it exactly equal to an eigenvalue makes the shifted solve singular.

Test a candidate pair by forming $r = Av - \rho v$ with $\|v\|_2 = 1$. In the orthonormal basis,

$$\|r\|^2 = \sum_i (\lambda_i - \rho)^2 c_i^2.$$

At least one eigenvalue must lie within $\|r\|$ of ρ. This is a numerical certificate for a symmetric problem. To certify a particular eigenvector, the gap to neighbouring eigenvalues also matters. At a repeated eigenvalue every normalised linear combination of its modes is valid. Comparing subspaces or projecting one span onto the other is then more meaningful than matching individual component signs.

Follow the whole spectrum by QR iteration

To find several eigenvalues of a small dense operator, factor $A_k - \mu I = QR$ and form $A_{k+1} = RQ + \mu I$. Orthogonality gives $A_{k+1} = Q^T A_k Q$, an exact similarity. Hence every step preserves eigenvalues, trace and determinant while changing the coordinates used to represent the operator. This use of QR is different from solving a triangular least-squares system. Repeated unshifted steps amount to an orthogonalised power iteration on multiple directions. A shift near a separated eigenvalue can make the last coordinate nearly invariant and accelerate its separation.

For a symmetric matrix a preliminary sequence of orthogonal similarities reduces it to tridiagonal form. When the coupling between the last two coordinates is negligible relative to the desired accuracy, setting that coupling and its transpose to zero separates a scalar eigenproblem. This removal is called deflation. If the off-diagonal magnitude is e, setting the pair to zero is a symmetric perturbation of spectral norm $|e|$, giving a direct measure for that decision. For a general real nonsymmetric matrix, the converged result can also contain real two-by-two blocks representing complex eigenvalue pairs.

A shifted step with a check on the trailing mode

Take $A = \left(\begin{smallmatrix} 4 & 1 \\ 1 & 2 \end{smallmatrix}\right)$ and $\mu = 2$. Normalising the first shifted column gives $(2, 1)^T/\sqrt{5}$. Complete the orthogonal basis with $(-1, 2)^T/\sqrt{5}$. Multiplying $Q^T A Q$ gives diagonal $(22/5, 8/5)$ and off-diagonal $-1/5$. The trace remains six and the determinant remains seven. The last-coordinate Rayleigh estimate, initially two, is now 1.6 compared with the exact lower eigenvalue $3 - \sqrt{2} \approx 1.586$. The residual of this coordinate has fallen from one to 0.2. A subsequent shift can use the improved estimate. Preserving eigenvalues alone would never certify that a coordinate is ready for deflation.

Which boundary modes are physical?

An elastic assembly suspended in free space admits rigid translations, and in three dimensions also rigid rotations when its geometry allows them. In a correctly assembled model these motions cost zero strain energy, so the stiffness matrix has a nullspace and some generalised eigenvalues are zero. Attempting unshifted inverse iteration on that stiffness matrix cannot locate a lowest positive vibration by simply inverting the operator. One can instead constrain rigid motions consistently or project into the mass-orthogonal complement of their span.

For a two-mass system with unit masses and a unit spring connecting them,

$$K = \left(\begin{smallmatrix} 1 & -1 \\ -1 & 1 \end{smallmatrix}\right).$$

Translation $(1,1)^T$ has zero squared frequency; opposite displacement $(1,-1)^T$ has squared frequency two. Grounding the first mass eliminates translation but changes the second mode to an oscillator with frequency one. Projection leaves the original free-system vibration at $\sqrt{2}$. These are physically different comparisons; the treatment of a nullspace determines which frequency an eigenvalue calculation is actually reporting.

Convert an eigenpair residual into an angle bound

Given a real symmetric matrix A, a unit-vector approximation v and a scalar θ, write its residual $r = Av - \theta v$. Expand $v = \sum_j c_j q_j$ in an orthonormal eigenbasis. The squared residual is $\sum_j (\lambda_j - \theta)^2 |c_j|^2$. This identity always implies that at least one eigenvalue lies within $\|r\|$ of θ. It does not by itself show whether the reported vector represents one particular physical mode. Two almost coincident eigenvalues can produce a tiny residual for many different mixtures of their eigenvectors.

To bound leakage out of a target eigenspace, let γ be a lower bound on the distance from θ to every eigenvalue outside that space. The squared residual is at least $\gamma^2 \sum_{\text{outside}} |c_j|^2$. Consequently the sine of the angle between v and the target space is at most $\|r\|/\gamma$. This result assumes a symmetric or Hermitian operator under the chosen inner product. For a mass-weighted mechanical problem, transform to mass-normal coordinates or perform the argument in the proper mass inner product. An ordinary Euclidean angle between unscaled coordinates may have little relevance to vibrational energy.

Resolve a cluster more reliably than either member

Take $A = \operatorname{diag}(10, 10.02, 30)$, and $v = (1,1,0)^T/\sqrt{2}$. Its Rayleigh quotient is 10.01. Multiplication gives $r = (-0.01, 0.01, 0)^T/\sqrt{2}$, of length 0.01. Relative to the cluster near ten this residual is small, yet the angle from the first coordinate eigenvector is 45 degrees. Using only the other low eigenvalue as a separation gives $\gamma = 0.01$, so the bound on that individual angle has sine at most one and provides no useful orientation guarantee.

Now take the target as the entire first two-dimensional subspace. Its gap from the reported quotient to the outside eigenvalue is

$$30 - 10.01 = 19.99.$$

The same residual implies leakage no greater than $0.01/19.99 \approx 0.000500$. In fact the true leakage of this particular vector is exactly zero. If the measured eigenvalues are squared angular frequencies, the two low modes have nearly the same frequency but can mix strongly under a small asymmetric manufacturing perturbation. Their combined subspace may remain very stable and may therefore be the more sensible object to compare with measurements.

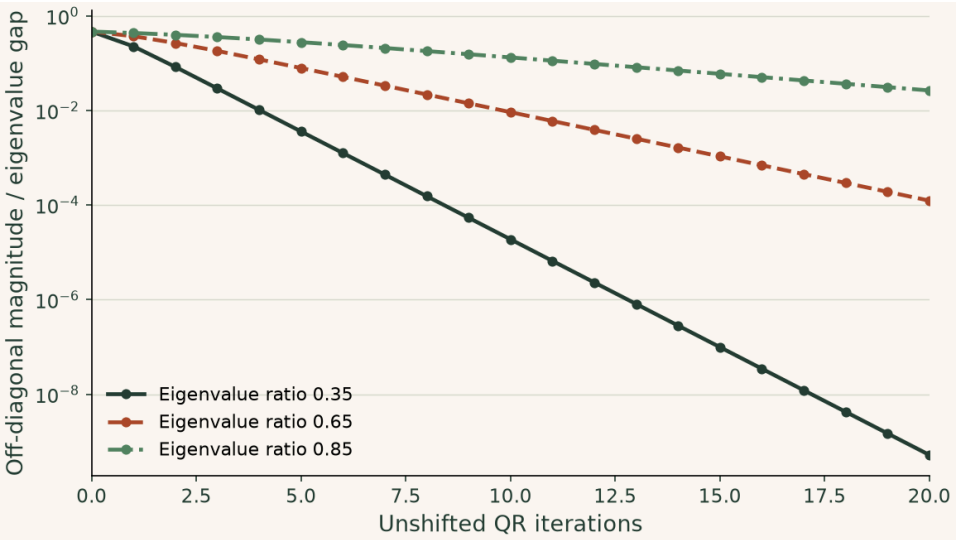
A laboratory comparison can project independent excitation patterns onto the computed cluster and compare the total energy in that cluster. Such a result is invariant under rotations of its basis, which a simple table of pairwise eigenvector differences is not. If the gap to the rest of the spectrum closes as

well, the subspace estimate becomes weak. Measuring the residual alone cannot repair that missing separation information.

For example the sum of squared projections of the excitation $(1, 0, 0)^T$ onto the first two computed orthonormal modes is one for any orthonormal basis of the exact low cluster. Individual projections can range between zero and one as the basis rotates. A test based on cluster projection assesses the reproducible quantity without requiring a uniquely labelled first mode.

Watch the QR iteration resolve two close eigenvalues

Rotate each diagonal matrix with eigenvalues $1, \rho$ by 0.6 radians, then factor $A_k = Q_k R_k$ and form $A_{k+1} = R_k Q_k$. Each step is an orthogonal similarity. Dividing the off-diagonal by the gap lets all three matrices start with the same normalized mixing.



The tails are straight on a logarithmic vertical axis. Once the dominant direction is nearly aligned, one step multiplies the remaining mixing by approximately the eigenvalue ratio. At a ratio near one, an unchanged spectrum can coexist with slow convergence to a diagonal matrix.

▶ **Read values from the plot**

Change parameters · Watch the QR iteration resolve two close eigenvalues

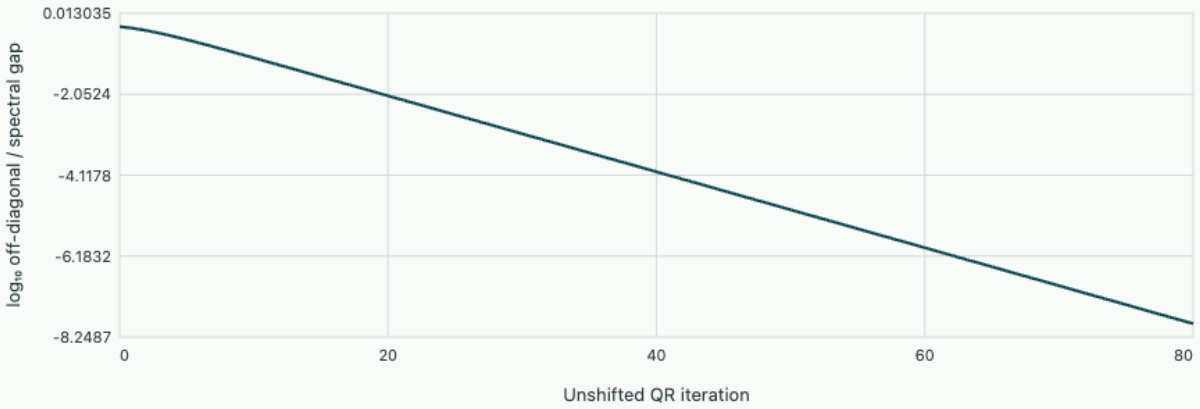
Change one input at a time and compare the recalculated curves with the argument above. The browser evaluates this small model locally; the accompanying numerical values state its assumptions and reference checks. Reset restores the reference inputs, and the data can be downloaded as CSV.

Open this calculation with adjustable inputs

Try the calculation

Smaller eigenvalue ratio ρ **0.8**

Initial rotation (degrees) **35**



Orthogonal-similarity mixing

Asymptotic contraction factor 0.8. The eigenvalues remain 1 and 0.8 throughout the iteration.

Singular values and minimum-norm solutions

Numerical linear algebra

Resolve a linear measurement into independent directions, compute its pseudoinverse and predict the effect of truncating a weakly observed component.

FROM FIRST PRINCIPLES

Count what a rectangular measurement can recover

For a real matrix A, the squared output length in a unit input direction v is $v^T A^T A v$. Since $A^T A$ is symmetric and nonnegative, it has orthonormal eigenvectors v_j with eigenvalues $\sigma_j^2 \geq 0$. Each positive eigenvalue defines a unit output direction $u_j = A v_j / \sigma_j$. Distinct orthogonal input eigenvectors then have orthogonal output directions because

$$u_i^T u_j = \frac{v_i^T A^T A v_j}{\sigma_i \sigma_j} = \delta_{ij}.$$

Resolve an input into these vectors and the null space. A component along v_j is amplified by σ_j, while every null component disappears. For measured data b, matching the observed output components gives

$$x_{\min} = \sum_{\sigma_j > 0} \frac{u_j^T b}{\sigma_j} v_j.$$

Adding any null vector produces the same prediction but increases the squared norm by its own squared length. This proves the minimum-norm choice. Data perpendicular to all the u_j remain as a least-squares residual, and noise along a small positive singular value can be amplified substantially.

Background · Singular value

$$A v_i = \sigma_i u_i$$

Input and output directions can live in different spaces. A small positive singular value can cause large inverse amplification.

DERIVATION AND APPLICATIONS

Find the smallest reconstruction in physically scaled coordinates

The minimum-norm solution of an underdetermined fit depends on which units and physical scale define the norm. Consider exact independent constraints $Ax = b$ with fewer rows than columns, and let W be a symmetric positive definite matrix of penalties. Minimize $\frac{1}{2}x^T W x$ among all feasible states. A reference value could first be subtracted if zero represents an inappropriate baseline. Factor $W = L^T L$ and express the optimization in the scaled variable $y = Lx$,

$$Ax = b \quad \Longrightarrow \quad (AL^{-1})y = b,$$
$$x^T W x = \|y\|_2^2.$$

Thus the usual minimum-norm singular-value solution applies to the scaled matrix. An alternative derivation introduces a multiplier λ. Differentiating $\frac{1}{2}x^T W x - \lambda^T(Ax - b)$ gives

$$Wx = A^T \lambda.$$

Substituting in the constraints yields

$$(AW^{-1}A^T)\lambda = b,$$
$$x = W^{-1}A^T(AW^{-1}A^T)^{-1}b.$$

The reduced matrix is positive definite for independent constraint rows. For any feasible perturbation h with $Ah = 0$, the cross term in the objective vanishes:

$$(x + h)^T W(x + h) - x^T W x = 2\lambda^T Ah + h^T W h = h^T W h.$$

This establishes uniqueness and explains the role of the weighting. Directions that data leave unresolved are assigned by a stated metric. Changing units while transforming both A and the metric consistently preserves the physical result.

WORKED EXAMPLE

Distribute a measured total with unequal motion costs

A balance gives $x_1 + x_2 = 10$ and the squared cost is $x_1^2 + 4x_2^2$. Stationarity requires $x_1 = 4x_2$, yielding $(8, 2)$ at cost 80. The unweighted minimum-norm pair $(5, 5)$ satisfies the balance but has weighted cost 125. There is no change in agreement with the measurement.

SECOND APPLICATION

Check invariance under a change of coordinate units

Write a second balance $u + v = 6$ and equal physical cost $u^2 + v^2$. If $w = 100v$, the new row is $(1, 0.01)$ and the cost is $u^2 + 0.0001w^2$. The weighted solution is $u = 3, w = 300$, recovering $v = 3$. Applying an unweighted norm to the rescaled coordinates instead gives $u = 6/(1.0001) \simeq 5.9994$, exposing a unit-dependent extra assumption.

Find directions that a rectangular map measures separately

For real A, the matrix A^TA is symmetric positive semidefinite since $v^TA^TAv = \|Av\|^2$. Choose orthonormal eigenvectors v_i with positive eigenvalues σ_i^2 and define $u_i = Av_i/\sigma_i$. Dotting two such vectors shows that the u_i are orthonormal. Completing both bases gives the singular value decomposition $A = U\Sigma V^T$. Each σ_i is a gain from one parameter direction into a corresponding output direction. Zero singular values indicate directions that no observation in this model can distinguish.

To minimise $\|Ax - b\|_2$, express both vectors in these orthogonal bases. For each positive singular value the residual component vanishes by taking $v_i^Tx = u_i^Tb/\sigma_i$. Output components outside the range remain unavoidable, while input components in the nullspace can take any value without changing the fit. Setting those free components to zero gives the minimum-norm least-squares answer. This derivation identifies the pseudoinverse without assuming either a square matrix or a unique exact solution.

Distinguish an unobserved coordinate from an error

Suppose two instruments read $b_1 = x_1 + x_2$ and $b_2 = 0.02x_3$, giving readings 4 and 0.06. One singular value is $\sqrt{2}$, corresponding to the mean direction of x_1, x_2; another is 0.02 along x_3. The difference $x_1 - x_2$ is unobserved. The minimum-norm exact fit is $(2, 2, 3)^T$, and every $(2 + s, 2 - s, 3)^T$ has identical data. More precision on the mean measurement cannot determine the hidden difference.

A noise change of +0.01 in the second reading raises the reconstructed third parameter by 0.5. If the admissible parameter scale is one, this is a substantial perturbation. Truncating the 0.02 singular direction instead sets $x_3 = 0$. Then the residual grows from zero to the entire second reading, but its noise no longer controls the fitted third coordinate. The decision trades bias against uncertainty; it does not prove that the true third parameter is zero.

Assess a reduced measurement model

Keeping the largest k singular components produces $A_k = \sum_{i \le k} \sigma_i u_i v_i^T$. In the spectral norm its error is σ_{k+1}: expansion in the right singular basis gives this upper bound, and applying the difference to v_{k+1} attains it. Any rank-k alternative must annihilate some unit vector in the span of the first $k + 1$ right singular vectors, on which the original map has norm at least σ_{k+1}. Hence no lower-rank replacement improves the worst-case absolute output error.

That best-approximation result concerns forward predictions in a specified norm. Inverting small singular values magnifies noise, so a good low-rank approximation for forward use does not automatically determine the appropriate inverse for parameter recovery. Work with physically scaled parameters and noise-weighted observations before deciding which singular components are meaningful.

Choose an output before discarding a direction

Suppose an instrument observes x_1 accurately and $0.001x_2$ with uncertainty 0.01. The second variable alone has uncertainty of order ten. The output x_1, however, has no projection onto the poorly observed direction. Its uncertainty receives no contribution from that amplification. Conversely the output $x_1 + x_2$ inherits the full uncertainty. This distinction can change whether a rank decision is acceptable. The spectrum describes the data acquisition matrix, while the projection of the chosen observable onto right singular vectors determines which losses of information affect the reported result.

Prove that truncation is optimal in two matrix norms

When compressing a dense linear map A, decide which discrepancy is being minimised over all approximations of rank at most k. Its spectral norm measures the largest output error over unit inputs. For any rank-k map B, consider the span of the first $k+1$ right singular vectors of A. Its dimension exceeds the rank of the restriction of B, so there exists a unit vector w in this span satisfying $Bw = 0$. Therefore $\|(A-B)w\| = \|Aw\| \geq \sigma_{k+1}$. Retaining the first k singular pairs attains this lower bound, as its discarded part has largest singular value exactly σ_{k+1}.

For the Frobenius norm, project each column of A into the column space of a candidate B. Orthogonal projection is the closest approximation with columns in that space. If its projector is P, the captured squared norm is $\sum_j \sigma_j^2 \|Pu_j\|^2$. The weights $\|Pu_j\|^2$ lie between zero and one and sum to at most k. The largest possible weighted sum assigns weight one to the k largest squared singular values. Hence the optimal discarded Frobenius norm is $(\sum_{j>k} \sigma_j^2)^{1/2}$. Both arguments require the norm and coordinate scaling just stated. Entrywise maximum error or application-specific weighted errors generally have different optimisers.

Decide how many responses a reduced map must reproduce

Suppose the four singular values of a unit-scaled instrument model are 8, 3, 2 and 1. A rank-two approximation has best possible worst-case gain error two and Frobenius error $\sqrt{2^2 + 1^2} = \sqrt{5}$. Its relative squared Frobenius loss is $5/(64+9+4+1) = 5/78$, approximately 6.41 percent. Describing this as a loss of just 6.41 percent can be misleading if the experiment deliberately excites the third singular direction. For that unit input, the exact output has length two and the compressed output is zero, giving complete loss of the signal of interest.

Suppose instead inputs have equal independent unit variance in all four right singular coordinates. Then the expected squared error in the output is the sum of the discarded squared gains, exactly five. If the

third coordinate is excited with variance one hundred while the others retain variance one, the error from truncating that coordinate alone becomes 400. Under that different usage distribution the meaningful objective is $\|(A-B)C^{1/2}\|_F$, where C is the input covariance.

In these diagonal coordinates the weighted gains become $8, 3, 20, 1$. The best rank-two design for expected output error now retains directions three and one, and the discarded expected squared error becomes

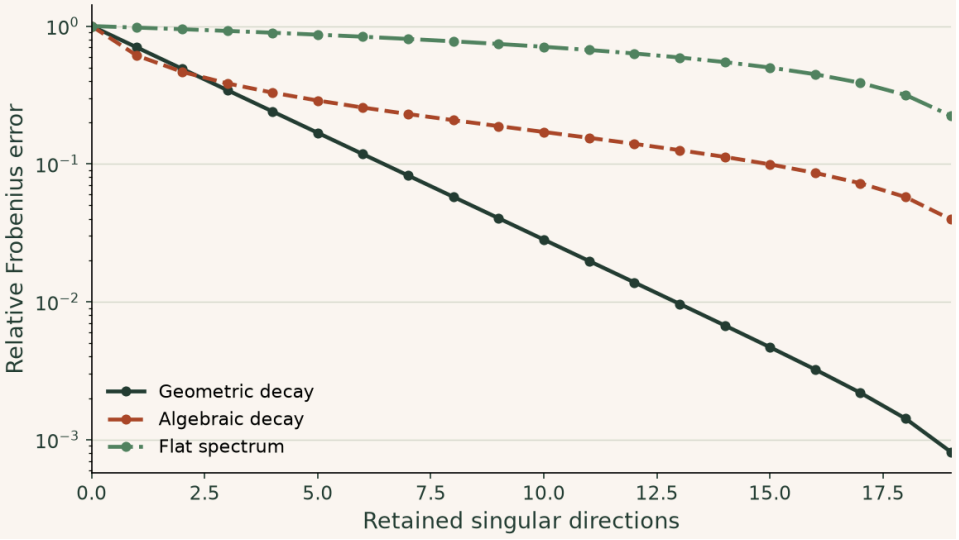
$$3^2 + 1^2 = 10.$$

The ranking changed because the physical question changed. It did not change the optimality theorem for the unweighted matrix; it changed which matrix correctly represents the intended error measure.

When an input covariance is singular, some directions are never excited under the stated design. The weighted objective can be indifferent to large errors in just those directions. A later change of operating regime may activate them, so a reduced model qualified for one input ensemble should retain the conditions under which its ranking was derived.

Decide whether a low-rank approximation is worthwhile

The best rank-k approximation discards the tail of the singular spectrum. Its relative Frobenius error is the square root of discarded squared singular values divided by total squared singular values, with the square root applied to the complete ratio.



For a flat spectrum, retaining ten of twenty directions still leaves an error of $1/\sqrt{2}$. Geometric decay makes the same rank far more effective. The spectral-norm error instead depends only on the largest discarded singular value.

▶ **Read values from the plot**

Change parameters · Decide whether a low-rank approximation is worthwhile

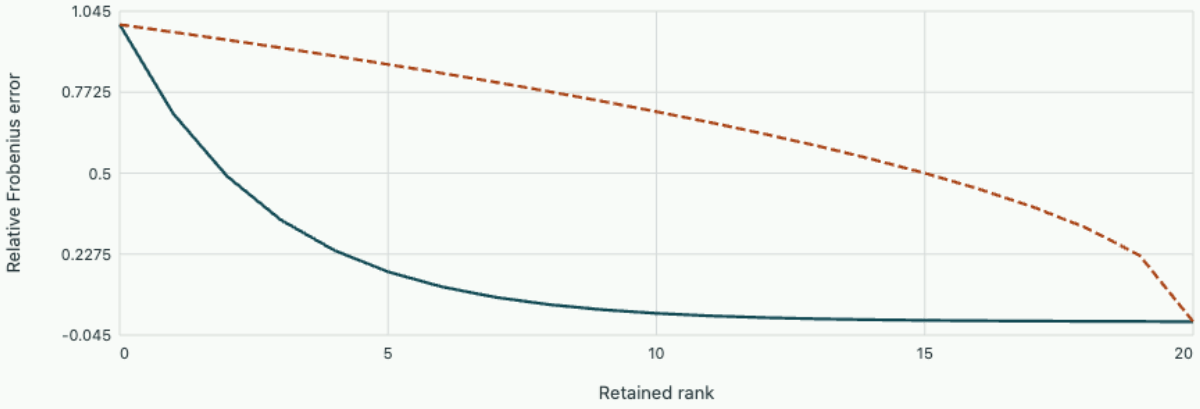
Change one input at a time and compare the recalculated curves with the argument above. The browser evaluates this small model locally; the accompanying numerical values state its assumptions and reference checks. Reset restores the reference inputs, and the data can be downloaded as CSV.

Open this calculation with adjustable inputs

Try the calculation

Singular directions **20**

Geometric decay ratio **0.7**



Geometric spectrum Flat spectrum

The best spectral-norm error after rank k is σk+1; this plot instead measures the complete discarded energy.

Deciding how far to trust a linearised residual

Nonlinear equations

Derive an acceptance test from predicted versus actual decrease and see how it rescues a Newton direction that crosses the root by a large distance.

Choosing a step from a local model

For a scalar residual f, finding a zero also minimises $\phi = f^2/2$ if a zero is attainable. Near x, approximate $f(x+s)$ by $f(x) + f'(x)s$, yielding a quadratic model $m(s) = (f + f's)^2/2$. Minimise this model subject to $|s| \le \Delta$. In one dimension this clips the ordinary Newton step to the interval; when $f' = 0$, the model contains no preferred direction.

Measure both predicted and actual decrease. Provided the predicted decrease is positive, define

$$\rho = \frac{\phi(x) - \phi(x+s)}{m(0) - m(s)}.$$

A ratio near one supports the local model. A negative ratio means that a predicted gain produced an actual loss. Shrink the radius and repeat from the same point. A successful boundary step can justify enlarging the next radius. These decisions require an extra residual evaluation and do not require a globally accurate Taylor approximation.

Background · Trust radius

$$\|s\| \le \Delta$$

The local model is minimised on this region. Agreement between predicted and achieved reduction guides the next radius.

Two proposals for the same equation

For $f(x) = x^2 - 2$ at $x = 0.2$, $f = -1.96$, $f' = 0.4$. The full Newton proposal $s = 4.9$ lands at 5.1, where the residual is 24.01. The linear model predicted zero. The merit function rises catastrophically.

With radius one, $s = 1$ gives $x = 1.2$ and $f = -0.56$. Predicted merit decrease is

$$(1.96^2 - 1.56^2)/2 = 0.704.$$

Actual decrease is

$$(1.96^2 - 0.56^2)/2 = 1.764.$$

The ratio exceeds one because the curvature accelerated the improvement. A large ratio is possible and has a direct interpretation.

Look along the step before accepting it

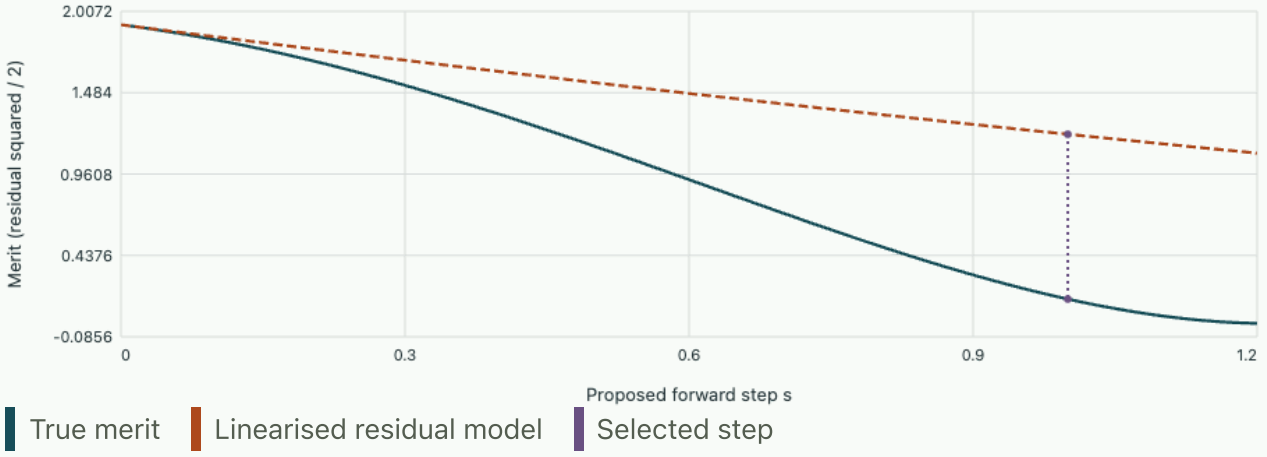
Compare the true merit function and the local model along positive steps from a starting value below the root. Move the trust radius and read the ratio for the constrained minimiser. Locate a region in which predicted and observed changes have opposite signs.

Open this experiment online

Try the calculation

Starting value 0.2

Trust radius 1



Proposed step 1. Actual / predicted decrease 2.5057.

Tutorial on choosing a merit function

Start at zero for the same equation. Can reducing the radius find a direction from this particular model?

▼ **Worked solution**

Here $f' = 0$ while $f = -2$. The linearised residual is constant, hence every step predicts zero decrease. The ratio is undefined. A curvature model, perturbation, or another starting point is needed.

Is a small gradient of $f^2/2$ sufficient to certify a root?

▼ **Worked solution**

Its gradient is ff'. It can vanish either at a root or at a stationary residual. Report the original residual and test feasibility to distinguish these outcomes.

Read the university treatment

Follow the nonlinear least-squares section to compare damping a proposed step with solving a constrained local subproblem. Driscoll and Braun · Nonlinear equations

Master equations, detailed balance and probability currents

Statistical mechanics from microscopic counting to material response

A two-state detector can switch at random while displaying reproducible occupation fractions and relaxation times. A three-state detector can maintain exactly constant occupancies while carrying a persistent directed current. Derive the probability dynamics, infer transition rates from timed records and decide what a stationary histogram can and cannot reveal about equilibrium.

From short-time transition probabilities to a generator

Take discrete states i resolved by an instrument. In a short time dt, conditional on current state i, let the probability of one transition $i \to j$ be $w_{ij}dt + o(dt)$. The probability of remaining in i is one minus the sum of departures to this order. Assume time-independent rates and the Markov property, meaning the conditional law of future transitions is determined by the current state. Sum gains and losses to derive

$$\dot{p}_i = \sum_{j \neq i} (p_j w_{ji} - p_i w_{ij}),$$

$$\sum_i \dot{p}_i = 0.$$

Each gain in one state is a loss in another, so total probability is conserved. In a column-vector convention $\dot{\mathbf{p}} = L\mathbf{p}$, off-diagonal entry $L_{ij} = w_{ji}$ and columns sum to zero. In a row-vector convention transpose the matrix. This is not cosmetic in a numerical calculation: using a matrix assembled for column probabilities while evolving rows can lose normalisation or give an incorrect stationary distribution.

For a given state the exit rate is

$$r_i = \sum_{j \neq i} w_{ij}.$$

Its survival probability obeys $\dot{S}_i = -r_i S_i$, so $S_i(t) = e^{-r_i t}$. Conditional on exiting, the destination has probability w_{ij}/r_i. This permits direct event-driven simulation: draw an exponential waiting time, choose a destination with these probabilities, advance the clock and repeat. Sampling one transition per arbitrary fixed time step generally distorts waiting-time statistics unless a controlled approximation or exact finite-time propagator is used.

Determine two rates from one occupancy and one time

For a two-state system name the rates $a = w_{01}$, $b = w_{10}$. Eliminating $p_0 = 1 - p_1$ gives

$$\dot{p}_1 = a - (a + b)p_1.$$

Its stationary value $p_* = a/(a + b)$ is independent of the common rate scale, and deviations decay with time constant $\tau = 1/(a + b)$. Thus one stationary fraction gives a ratio, one relaxation time gives a sum, and the two together identify both rates.

A location sensor records an upper-state fraction 0.30 and relaxation time 0.20 seconds. Then $a = 1.50$ and $b = 3.50$ per second. Mean waiting times are $1/a = 0.667$ s in the lower state and $1/b = 0.286$ s in the upper one. Neither equals the population relaxation time of 0.20 s. A transition rate inferred by taking the reciprocal of the average of these two waiting times estimates a different quantity.

Define the indicator $X = 1$ in state one and zero otherwise. Conditional probabilities from the same linear solution give

$$\langle [X(t) - p_*][X(0) - p_*] \rangle = p_*(1 - p_*) e^{-|t|/\tau}.$$

Consequently a normalised equilibrium autocorrelation determines τ without needing a deliberate preparation away from stationarity. A bandlimited detector may miss short visits, reducing counts and distorting waiting-time distributions even if the mean occupation looks plausible. Compare time sampling with both exit times as well as the relaxation time, or model the measured signal with an explicit observation process.

Detailed balance is stronger than stationarity

Let π_i denote a stationary distribution. The net probability current on an oriented edge is $J_{ij} = p_i w_{ij} - p_j w_{ji}$. Stationarity requires only that currents sum to zero at each vertex. Detailed balance requires $\pi_i w_{ij} = \pi_j w_{ji}$ on every edge individually. For a system in equilibrium with one heat bath, with states even under time reversal and energies E_i, thermal rates satisfy $w_{ij}/w_{ji} = e^{-\beta(E_j - E_i)}$ for nondegenerate states. If states collect multiple microstates, degeneracies or coarse-grained free energies enter their equilibrium weights.

For the previous two-state process at equilibrium with equal state degeneracies, $p_*/(1 - p_*) = 3/7$ implies

$$(E_1 - E_0)/(k_B T) = \log(7/3) = 0.847.$$

Multiplying both rates by ten preserves the energy inference and speeds relaxation tenfold. Thermodynamic weights therefore do not determine the kinetic barrier. A physical sample may change its rate scale with solvent viscosity or transition geometry without changing the equilibrium ratio.

A convenient test on a connected graph with reciprocal edges is the product of rate ratios around each closed cycle. With detailed balance all stationary factors telescope to one. Conversely, if the product is one on every independent loop, consistent state weights can be reconstructed from any reference state by integrating ratios along a path. Path independence makes these weights a detailed-balance stationary distribution after normalisation. A missing reverse edge or unobserved channel requires additional care; a detector that cannot resolve two routes can hide directed currents.

A steady nonequilibrium system with the same state histogram

Consider three equally named sites arranged in a ring, with clockwise rate k_+ and anticlockwise rate k_- on every link. Symmetry gives stationary occupation $1/3$ at each site for all positive rates. Each oriented clockwise edge carries current $J = (k_+ - k_-)/3$. The cycle affinity in dimensionless units is $\mathcal{A} = \log[(k_+/k_-)^3]$. For $k_+ = 4$, $k_- = 1$ per second, each edge transports net probability one per second, and $\mathcal{A} = 3\log 4 = 4.159$. A perfectly flat histogram conceals the directed operation.

For Markov jumps with resolved reversible edges and their conventional time reversals, the total stochastic entropy production rate is

$$\dot{S}_{\mathrm{tot}} = \frac{k_B}{2}\sum_{ij}(p_i w_{ij} - p_j w_{ji}) \log \frac{p_i w_{ij}}{p_j w_{ji}} \geq 0.$$

Each term is nonnegative under the product $(x - y)\log(x/y)$; the factor of one half prevents counting both directions separately. For the stationary ring this gives $k_B J \mathcal{A} = 4.159 k_B$ per second. An actual heat dissipation rate needs a physical local detailed-balance model identifying environments, work and entropy flows. A collection of counted transitions without that calibration can still yield evidence for broken time-reversal symmetry at its level of resolution.

Stationary Shannon entropy $-k_B \sum p_i \log p_i$ is constant in this ring. The positive total entropy production is compatible with that constancy because an environment sustains the directed current and carries compensating entropy. Equating production with just the derivative of stationary state entropy would predict zero for every driven stationary process.

Why a distribution relaxes toward a stationary law

For a finite irreducible time-independent chain let $\pi_i > 0$ be its unique stationary distribution. The relative entropy $D(p\|\pi) = \sum p_i \log(p_i/\pi_i)$ is nonnegative. To see its contraction over a finite observation interval, write the propagated probability as a stochastic mixture of earlier probabilities and apply the log-sum inequality to each output state. Summing gives

$$D(Pp\|P\pi) \leq D(p\|\pi).$$

Since $P\pi = \pi$, the relative entropy to stationarity cannot grow.

For a detailed-balance chain, substituting the master equation into this derivative gives $-k_B dD/dt = \dot{S}_{\mathrm{tot}}$ in the simple closed heat-bath model. For a driven chain the system may remain at π with constant relative entropy while producing positive steady entropy. Relaxation toward stationarity and dissipation required to maintain that stationary state are distinct measurable quantities.

Compare relaxation inferred from a time trace with the thermodynamic irreversibility experiment

Infer generator entries from residence times

An ideal continuously observed Markov record contains successive states and dwell lengths. Let T_i be accumulated time in state i and N_{ij} the number of resolved $i \to j$ jumps. Its log likelihood contains $\sum_{ij} N_{ij} \log w_{ij} - \sum_i T_i \sum_{j \neq i} w_{ij}$. Differentiation gives

$$\widehat{w}_{ij} = N_{ij}/T_i.$$

Dividing by total experiment time instead estimates the gross stationary flux $\pi_i w_{ij}$, not a conditional transition rate.

For a nearly stationary two-state record with $T_0 = 70$ s, $T_1 = 30$ s and 105 transitions in each direction the estimates are 1.50 and 3.50 per second. Counting uncertainty for each reasonably large Poisson-like count is roughly the inverse square root of the count in relative terms; precise interval estimates condition correctly on exposures and the observation process. Equality of opposite transition counts in a long two-state record is essentially forced by alternation. It cannot by itself establish a thermal energy gap or exclude multiple hidden driven pathways.

Turn a barrier diagram into two independently measurable rates

A thermally activated transition between two metastable configurations often has a forward rate $a = \nu_A e^{-(G^\ddagger - G_A)/(k_B T)}$ and reverse rate $b = \nu_B e^{-(G^\ddagger - G_B)/(k_B T)}$, where $G^\ddagger$ denotes a dividing-surface barrier under a chosen model. If the same transition-state construction applies in both directions with consistent partition factors, their ratio reproduces the equilibrium free-energy difference. The overall barrier controls how rapidly equilibrium is approached. Changing a barrier while leaving both wells unchanged can alter both kinetics by many orders of magnitude without changing stationary occupancy.

As an illustration let the higher-energy state cost $2k_B T$, and put a transition configuration $10k_B T$ above the lower state. With a common attempt rate of 10^6 per second, the upward rate is 45.4 per second and the downward rate 335 per second. The stationary upper fraction is approximately 0.119 and the relaxation time 2.63 ms. Lower the barrier by $2k_B T$ without changing the two state energies. Both rates increase by $e^2 = 7.39$, while occupancy is unchanged. A detector recording only the equilibrium level histogram could miss the intervention entirely.

For a temperature scan, a straight line in the log rate against inverse temperature is evidence for an approximately constant activation energy and prefactor over that interval. A free-energy barrier includes entropic as well as enthalpic contributions; its numerical value can vary with temperature even if the microscopic potential does not. A change of dominant pathway can bend the apparent Arrhenius plot. Measure forward and reverse rates independently, then compare their ratio with an independently determined equilibrium population. A barrier fit explaining one rate but violating the equilibrium ratio usually indicates inconsistent state definitions, missed short visits or an unmodelled pathway.

A useful control is a weak static bias that lowers one well relative to the other. The equilibrium log-odds shift measures the difference between the state couplings to the bias. The individual log-rate shifts additionally report the related dividing-surface response. If the state-occupation ratio moves as expected but only one rate changes appreciably, the effective transition configuration may be much

closer in bias response to the other state. Such statements depend on the activated single-path interpretation; the response slopes themselves are more directly observed than an inferred barrier shape.

Use uniformisation to propagate a generator without negative probabilities

The formal finite-time propagator of a time-independent jump process is $P(t) = e^{Lt}$. A forward Euler update $p(t + \Delta t) = (I + L\Delta t)p(t)$ conserves total probability if the generator is assembled correctly, but a large step can yield negative diagonal probabilities. An exact algebraic representation starts with a rate q at least as large as every total exit rate. Then $K = I + L/q$ is a stochastic matrix in the same column convention. Write $L = q(K - I)$; since K and the identity commute,

$$e^{Lt} = e^{-qt} \sum_{n=0}^{\infty} \frac{(qt)^n}{n!} K^n.$$

This can be read as drawing a Poisson number of opportunities at rate q, followed by n transitions of K. At a lightly connected state some opportunities become self transitions. The construction leaves real waiting-time statistics correct while eliminating negative transition probabilities. Its truncation error in total probability is bounded by the omitted Poisson tail because every matrix power is stochastic.

For a two-state model with rates two and six per second choose $q = 6$ per second. Starting in the lower state, the first opportunity after some waiting time moves upward with probability one third and otherwise does nothing. In the upper state the next opportunity moves downward with probability one. At 0.25 s the Poisson mean is 1.5. Retaining terms through order eight leaves a tail below 3×10^{-5}. Compare the resulting upper probability against $(2/8)[1 - e^{-8(0.25)}] = 0.2162$. A small discrepancy due solely to truncation should shrink with the computable tail probability.

The representation also suggests a measured-data sanity check. A fitted generator with negative off-diagonal rates cannot yield a valid K at any positive q. A snapshot transition matrix can itself be nonnegative and stochastic but need not arise from a time-homogeneous continuous-time generator. In the two-state case its nontrivial eigenvalue must be $e^{-(a+b)t} > 0$. A measured negative eigenvalue rules out that simple generator model under perfect sampling, though finite counts and hidden periodic driving can distort empirical estimates. Taking a matrix logarithm blindly may return complex or negative rates rather than a physical kinetic explanation.

Population relaxation and a first-arrival experiment answer different questions

A first-passage measurement stops its clock when a target is first reached, even if that target would subsequently be left many times. To calculate its mean, treat the target as absorbing. For any other state i, wait on average $1/r_i$, then add the mean remaining time from the next state weighted by its conditional transition probability. The backward equations are

$$-1 = \sum_j w_{ij}(t_j - t_i),$$

with $t_{\mathrm{target}} = 0$. Their solution depends on possible returns before the first successful passage. It generally differs from an inverse equilibrium occupation or from a slowest relaxation time.

Consider a three-site linear chain $0 \leftrightarrow 1 \leftrightarrow 2$ with equal rate k on each adjacent directed edge, starting at zero and stopping at two. From zero one must reach one after mean $1/k$. From one the mean wait is $1/(2k)$, then there is probability one half of returning to zero and one half of finishing. Solving $t_0 = 1/k + t_1$, $t_1 = 1/(2k) + t_0/2$ gives

$$t_0 = 3/k.$$

Ignoring returns would yield $1.5/k$, exactly half the correct first-passage mean.

The unmodified closed chain has equal stationary weights, but that fact alone says nothing about the numerical value of $3/k$. Speeding every link by ten leaves occupancies unchanged and reduces passage times tenfold. A detector that registers only visits longer than a threshold makes a brief first arrival invisible and can report even longer apparent passage times. Calibrate detector dead time and distinguish entrance to the target from persistence within it.

For a larger network, solve the sparse backward linear system and compare against many event-driven paths. Check boundary initial conditions by starting simulations exactly at the absorbing target: the first-passage time must then be zero under that convention. Some experiments instead ask for a first return after leaving, which requires a different initial condition. A zero-lag event in one definition is a physically different measurement in the other.

Estimate a cycle affinity with finite counting statistics

A clockwise ring trajectory contains gross counts in both directions on each edge. Estimating each conditional rate as count divided by origin-state exposure, the product of clockwise-to-anticlockwise rates around a complete closed loop cancels the exposures. The estimated cycle affinity is the sum of log forward counts minus log reverse counts, assuming every required count is nonzero and rates were stationary during the record. This is a useful cross-check independent of how the total time was partitioned among states.

If all three edges register roughly 120 clockwise and 80 anticlockwise events, the estimated affinity is

$$3\log(1.5) = 1.216.$$

For large independent Poisson-like gross counts the approximate variance of this log product is $3(1/120 + 1/80) = 0.0625$, giving standard error 0.25 as a first estimate. Actual path constraints couple counts; uncertainty can be assessed from independent records or a complete stochastic likelihood. The point estimate alone does not tell whether an observed directionality is statistically distinguishable from zero.

One missing reverse event makes the unregularised logarithm infinite, but a finite record cannot establish an infinite physical affinity. Report a confidence bound based on the exposure and upper limit on the unseen reverse rate. Extend the acquisition or change sampling speed to exclude missed short-lived reverse states. Numerical addition of an arbitrary count can help regularise a plotting routine but is not a substitute for a specified prior or measurement model when claiming physical entropy production.

A driven occupation loop can carry charge without changing concentrations

Place three sites in a cycle and drive their transition rates by two controls, for example energies and barriers. Even if each instantaneous frozen generator satisfies detailed balance with a well-defined equilibrium distribution, the physical occupations need not equal that moving equilibrium distribution at each moment. A slow lag can produce a small edge current at every instant and a finite integrated current over a cycle. The current through an edge is time-integrated transfer; it can be nonzero while all occupations return to their initial values after a complete drive period.

A simple piecewise illustration avoids subtleties of adiabatic geometric pumping. In stage one, connect only sites zero and one while strongly favouring one. In stage two connect only one and two while favouring two. In stage three connect only two and zero while favouring zero. Begin with nearly all probability at zero and wait long enough for each enabled pair to relax. Almost the entire population travels once around the loop per cycle. No site accumulates probability over repeated complete cycles, but each edge registers about one forward crossing net per cycle. External control changes the energy and barrier landscape and supplies the required work.

If each programmed transfer succeeds with probability 0.95, it is tempting to announce a current of 0.95^3 per cycle. This is the probability of one fully successful trajectory starting on the prescribed site in a simplified no-return preparation, not generally the long-run edge current of a periodically driven ensemble. Failed transfers leave probability on other sites in later stages. The correct periodic distribution is the unit-eigenvalue vector of the product of the three finite-time propagators. Compute edge currents during each stage using its evolving occupations and integrate, then compare a complete cycle after transients have decayed.

This construction supplies a statistical check on experimental counting. Over a complete cycle the integrated net inflow at a site equals its change in occupancy probability. In a periodically stationary state that change is zero, so all three consistently oriented integrated edge currents must agree. A disagreement between measured edges can reveal a missed transition channel, a start-up transient or incorrect gating. Agreement does not prove absence of drive or detailed balance. A torque-free stationary ring and a three-stage external pump can share nearly the same time-averaged site histogram yet differ clearly in time-resolved edge currents.

Finally compare the work and heat bookkeeping with the master-equation entropy-production expression. During an external change of a site's energy without a transition, the state's energy changes by work. During a transition at frozen controls, assigning heat requires the energy difference and the relevant bath. Lumping an entire driven cycle into a single time-averaged generator can conceal those distinctions and underestimate dissipation.

The ordered protocol can also be run backwards. If the three stages are reversed while preparations are equilibrated in each corresponding periodic state, the preferential current direction should reverse in the ideal transfer picture. A fixed instrumental preference for counting neighbouring sites in one direction would not generally reverse. This distinguishes controlled transport from a direction-dependent detection efficiency. Do not confuse this observed current reversal with exact microscopic time reversal of each individual stochastic trajectory: the reversed experiment has its own ensemble of random jumps. Comparing path probabilities, rather than pairing individual paths as though their noise histories were reproducible, is the appropriate starting point for entropy-production and fluctuation-relation tests. The distinction becomes particularly visible when a stage transfers almost all the probability in one direction while a few rare trajectories travel against the applied protocol.

Distinguish stationarity from detailed balance

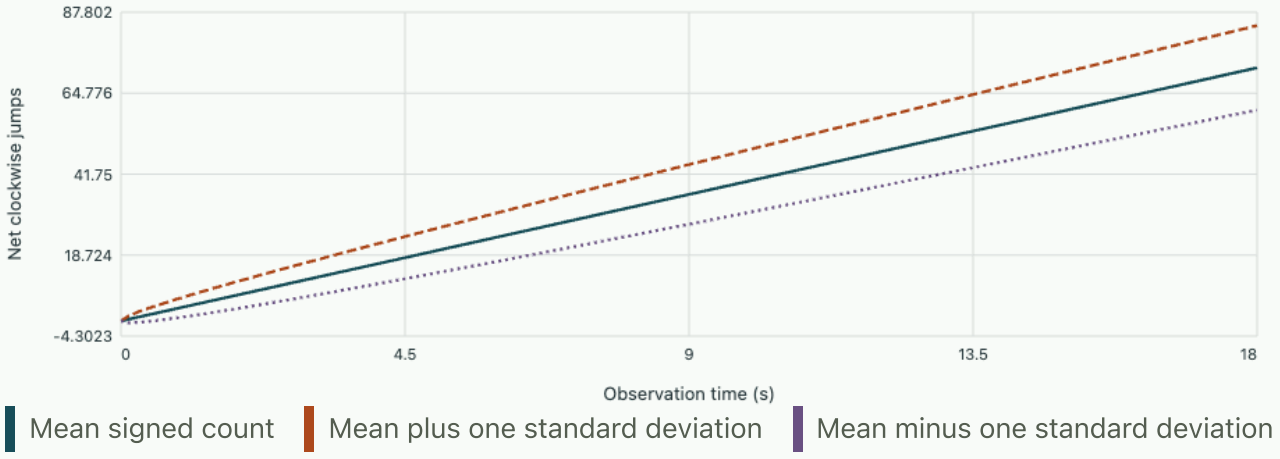
Keep the three populations identical while varying the directed transition rates. Identify the observable that reveals time-reversal asymmetry.

Open this experiment online

Try the calculation

Clockwise bias **0.5**

Total jump rate (s⁻¹) **8**



Every stationary population is 1/3. Net clockwise edge current 1.3333 s⁻¹. Entropy production 4.3944 kB/s.

Tutorial · Recover a generator from snapshots rather than full events

At sampling interval 0.10 s, a two-state record yields $P_{01} = 0.12$ and $P_{10} = 0.28$. Infer the continuous-time rates, and compare with directly dividing each transition probability by 0.10 s.

▼ **Use the exact finite-interval propagator**

The exact transition probabilities are $P_{01} = p_*[1 - e^{-(a+b)\Delta t}]$ and $P_{10} = (1 - p_*)[1 - e^{-(a+b)\Delta t}]$. Their sum is 0.40, so $a + b = -\log(0.60)/0.10 = 5.108$ per second. The stationary upper fraction is

$$0.12/0.40 = 0.30.$$

Thus $a = 1.532$ and $b = 3.576$ per second. Dividing by the interval gives only 1.2 and 2.8 because return transitions inside a sampling interval can erase evidence of the intermediate visit.

Tutorial · Can measured rates arise from one equilibrium potential?

For a three-state triangle suppose clockwise to counterclockwise conditional rate ratios are 2, 3 and 0.25 around the three distinct edges. Could these be equilibrium weights in a common bath? If a single uncertain ratio were corrected, what would it have to be on the third link?

▼ **Apply the closed-loop consistency condition**

The product is $2 \cdot 3 \cdot 0.25 = 1.5$, giving nonzero loop affinity $\log 1.5 = 0.405$. No consistent detailed-balance set of three state weights exists for these exact ratios. The third ratio would have to be $1/6$, giving a telescoping unit product. If counts are noisy, propagate uncertainty in the log ratios or compare a constrained detailed-balance likelihood against an unconstrained model. A value barely inconsistent with unity within one sparse record is not enough to establish driving.

Tutorial · Detect a hidden state with the same observed level

A detector reports two visible levels, but its lower level combines two unresolved substates. The measured lower-state survival probability is $0.5e^{-t} + 0.5e^{-9t}$ with time in seconds. Calculate the initial hazard and its long-time limit. Can a two-state Markov fit reproduce both?

▼ **Differentiate survival before interpreting the tail**

The hazard is minus the derivative of $\log S$. At zero time it is the initial weighted mean of rates, 5 per second. At long times the rapidly departing component has almost disappeared and the hazard tends to 1 per second. A single Markov exit rate would give a constant hazard for all dwell ages. The observed model cannot reproduce both limits with one rate.

A changing apparent hazard may also result from limited detector temporal resolution or variable preparation of the substates. Repeat with shorter acquisition intervals and condition the dwell distribution on the preceding visible transition or an independent preparation marker. A fitted sum of

exponentials identifies a need for a richer observation or state model; it does not uniquely identify microscopic structures absent in the measurements.

Cambridge Kinetic Theory · stochastic processes and approach to equilibrium

Applied study

Design feedback for a tracking requirement and examine how loop gain changes speed, offset and stability.

Turn a machine balance into a plant model

A speed controller drives a motor against viscous drag. The shaft has inertia J, drag coefficient b, angular speed ω and commanded torque u. With a constant opposing load torque d, the balance is

$$J\dot{\omega} = u - b\omega - d.$$

Without feedback a step in load changes the steady speed by $-d/b$. An engineer who can measure speed can command more torque as soon as the speed falls.

Divide by b. With output $y = \omega$, time constant $\tau = J/b$ and static gain $K = 1/b$, the unloaded plant becomes

$$\tau\dot{y} + y = Ku.$$

A heating element with finite heat capacity and heat loss has the same first-order form under constant material properties. Transfer functions or step-response measurements can therefore guide controller design before the full machine is available.

Choose the loop gain from a tracking requirement

A proportional controller applies $u = K_p(r - y)$ for a reference r. Substitution gives

$$\tau\dot{y} + (1 + KK_p)y = KK_p r.$$

If $1 + KK_p > 0$, the closed loop settles with time constant $\tau_{\mathrm{cl}} = \tau/(1 + KK_p)$. Following a step from rest,

$$y(t) = \frac{KK_p}{1 + KK_p} r\,(1 - e^{-t/\tau_{\mathrm{cl}}}),$$

$$e_\infty = \frac{r}{1 + KK_p}.$$

The steady error supplies the command that balances the drag. For a normalised model with $K = 2$, $\tau = 20$ s, $r = 1$ and $K_p = 4$, the loop gain is 8. The output tends to $8/9 = 0.8889$ with time constant 2.222 s. Keeping the steady error within 5% of the setpoint requires $1/(1 + 2K_p) \le 0.05$, hence $K_p \ge 9.5$. At the instant of a unit step this higher gain asks for 9.5 units of actuator effort. A drive limited to 5 units initially saturates, so the linear prediction for the rise time must be revised.

Reject a load disturbance with integral action

For the motor with a constant load, proportional control settles at $y = (K_p r - d)/(b + K_p)$. Adding $K_i \int (r - y)\, dt$ to the command lets the controller supply constant torque even at zero error. If the loop settles and the actuator has sufficient reserve, a bounded steady integrator state requires $r - y = 0$.

The extra state changes stability and damping. For the normalised plant and a PI controller, the closed-loop characteristic polynomial is $\tau s^2 + (1 + KK_p)s + KK_i$. At $K = 2$, $\tau = 20$ s, $K_p = 4$ and $K_i = 1$ per second, its roots are $-0.225 \pm 0.222i$ per second. Both decay, but a disturbance can now excite an oscillation. A gain that removes steady error can change the transient response substantially.

Problem class

A motor picks up a load

Take $b = 0.50$ N·m·s/rad, $K_p = 2.0$ N·m·s/rad and reference speed 10 rad/s. A 1 N·m load is added after the motor has settled. Find the two steady speeds and the torque an integral controller would need to restore 10 rad/s.

▼ Balance torque before and after the load step

Before the load, $y = 20/2.5 = 8.0$ rad/s. After it, $y = (20 - 1)/2.5 = 7.6$ rad/s. At the desired speed the drag is 5 N·m, so the restored steady state needs a total of 6 N·m from the motor. At zero error the integral term supplies that torque.

Design for critical damping

Keep the normalised plant and $K_p = 4$. Find the integral gain for a repeated pole. Predict the effect of doubling this integral gain on the pole locations.

▼ Find the poles from the characteristic polynomial

Critical damping requires $(1 + KK_p)^2 - 4\tau K K_i = 0$, giving $K_i = 81/160 = 0.50625$ per second. The pole is $s = -9/40 = -0.225$ per second. Doubling K_i leaves the real part unchanged and gives an imaginary part of magnitude 0.225 per second. The poles now form a complex pair.

ONESIMULATION LABORATORY

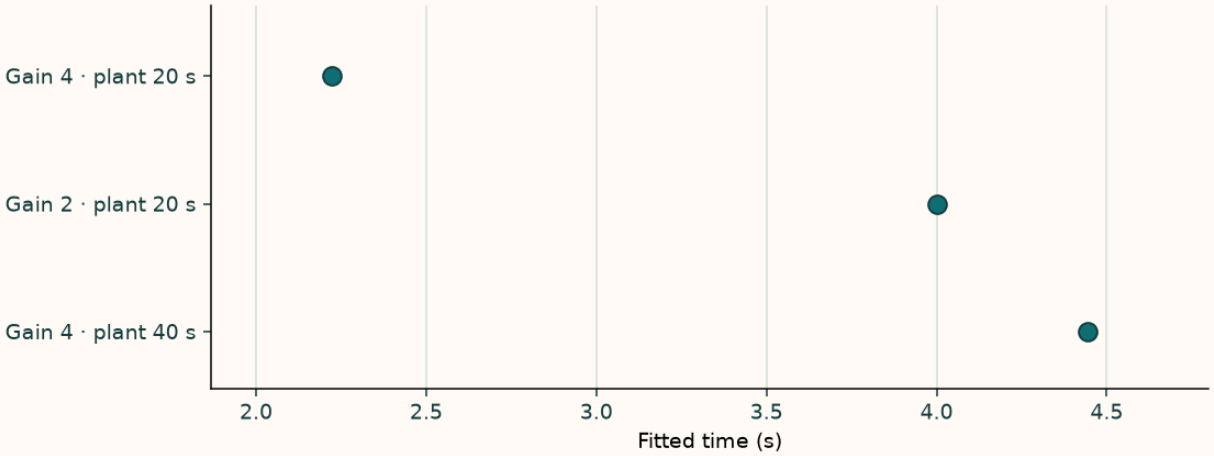
Predict the settling time and offset of a feedback loop

Compare recorded OneSimulation runs on this page. To test your own inputs, download the notebook and model files and run the model in an activated OneSimulation environment.

The OneSimulation signal graph connects a setpoint, error junction, controller and first-order plant. Start with `plant_gain = 2.0` and `setpoint = 1.0`. Change `proportional_gain` or `time_constant`, keeping the other parameters fixed. Before comparing, decide which change affects the asymptotic error.

Numerically fitted time constant and tracking error sampled after six closed-loop time constants. The unit step starts from rest.

Controller gain and plant time constant	Fitted time (s)	Sampled error
Gain 4 · plant 20 s	2.22222	0.11331
Gain 2 · plant 20 s	4.00000	0.20198
Gain 4 · plant 40 s	4.44444	0.11331



Reducing proportional gain slows the loop. Doubling the plant time constant with gain fixed doubles the fitted closed-loop time while leaving the error at six closed-loop time constants almost unchanged.

Try $K_p = 9.5$ in the download. Compare the recorded final error against $e_\infty = 0.05$. To explain the difference, evaluate the exponential at the actual last sample time. Inspect the signal graph and its time histories.

Download experiment files (ZIP)

Extract the files together and open `first_order_plant_p_control.ipynb` in a local OneSimulation environment.

Interactive reference for the OneSimulation laboratory

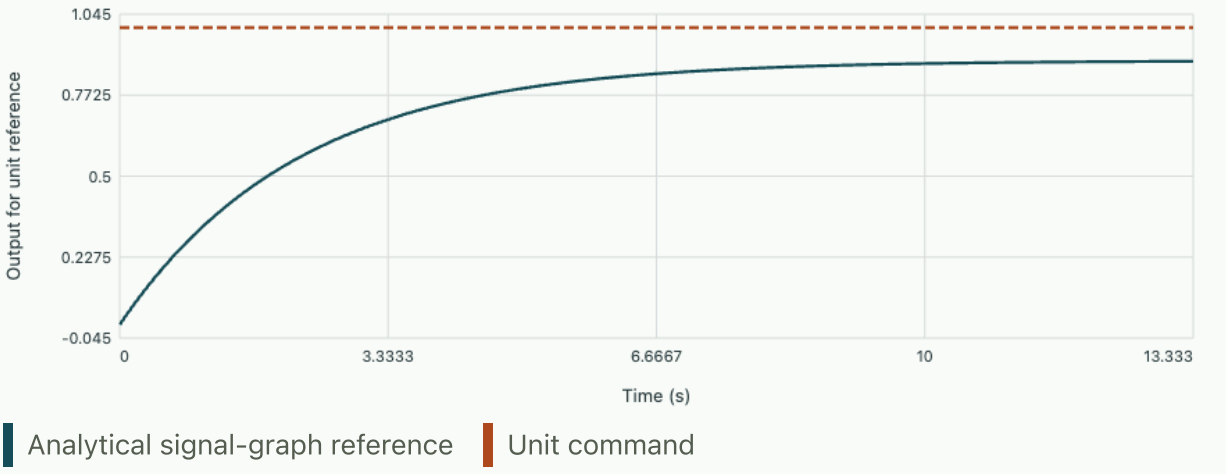
Change gain and plant lag independently. The browser evaluates the unsaturated analytical reference for the linked OneSimulation signal graph.

Open the live laboratory reference

Try the calculation

Proportional gain **4**

Plant time constant (s) **20**



Closed-loop time constant 2.22222 s; asymptotic error 0.11111; error after6 time constants 0.11331. This is the unsaturated reference for the linked OneSimulation signal model.

Problem sheet 2 · Identify a controller from a finite-time observation

Notation and assumptions

Read each defining equation with its assumptions and units, then follow the link to the argument that gives the equation its meaning.

Complementary sensitivity

$$T(s) = L(s)/[1 + L(s)]$$

For the same feedback configuration S plus T is one. Measurement noise and plant-input disturbances reach the output on different paths.

Derivation · Separating reference tracking from disturbance rejection

Discrete pole

$$z = e^{sT_s}$$

This maps a continuous modal exponent into a sampled multiplier over interval Ts. Different continuous frequencies can give the same sampled phase.

Derivation · Sampling, saturation and commissioning

Loop delay

$$L_{\text{delayed}}(i\omega) = L(i\omega)e^{-i\omega\tau}$$

A pure delay leaves the magnitude unchanged and adds frequency-dependent phase. Use angular frequency when converting a margin to seconds.

Derivation · How much transport delay a feedback loop can tolerate

Observability

$$\mathcal{O} = [C^{\mathsf{T}}, (CA)^{\mathsf{T}}, \ldots]^{\mathsf{T}}$$

Full column rank in the standard finite-dimensional linear model allows the initial state to be inferred from ideal output history. Noise still controls achievable precision.

Derivation · States, reachability and estimation

Sensitivity

$$S(s) = 1/[1 + L(s)]$$

For a scalar unity negative-feedback loop this gives one closed-loop transfer. The point where a disturbance enters determines which transfer the observer measures.

Derivation · Uncertainty envelopes and robust stability

Transfer function

$$G(s) = Y(s)/U(s)$$

State the input, output and zero initial condition. Interconnection signs and unmodelled dynamics affect what a measured closed-loop response represents.

Derivation · Transfer functions and connections between systems

Further reading

- Compare time-domain pole placement with loop-shaping predictions. Trace an input disturbance and measurement noise separately through the same feedback loop. Åström and Murray · Feedback Systems
- Use the cost-to-go derivation to relate the Riccati matrix to the geometry of a controlled trajectory and its terminal cost. Russ Tedrake · Linear Quadratic Regulators

Compare specific topics with the university sources