Inverse problems and least squares

Identify observable parameter combinations, quantify noise amplification, regularise unstable fits, quantify uncertainty and design measurements that test a reconstruction.

Read the subject in depth

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

- ### Inferring parameters and fields from observations

  Follow information from a forward model back to its unknowns. Evaluate identifiability, numerical sensitivity and how an additional observation changes the inference.

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. ### Which field can be recovered from boundary data?

   Relate an observation to its spatial sensitivity. Use reciprocity, smoothing with distance and changes of electrode configuration to identify recoverable features.

2. ### A measured curve and competing mechanisms

   Fit material response, decay and resonant splitting. Design at least one new observation that can distinguish explanations left unresolved by the first experiment.

How the course fits together

1. **Chapters 1–4**

   A forward map predicts observations from parameters. Whiten noise before measuring a misfit. Singular directions reveal poorly observed parameter combinations, and regularisation controls their influence.

2. **Chapters 5–8**

   Linearise the forward map near a trial parameter and choose measurements that distinguish competing fits. A prior yields a posterior distribution. Profiles ask which physical predictions remain supported as other parameters readjust.

3. **Chapters 9–12**

   Evaluate gradients with an adjoint, stop iterations to limit noise amplification, and impose feasible sets or edge-preserving penalties. Resolution analysis asks which features survive the full measurement and reconstruction process.

Contents

1. Chapter 1 · Forward maps, information and identifiability
2. Chapter 2 · Weighted least squares and residual diagnostics
3. Problem sheet 1 · Which combinations do the data determine?
4. Chapter 3 · Singular systems and the amplification of noise
5. Chapter 4 · Regularisation, resolution and the price of stability
6. Problem sheet 2 · Spectral filtering and recoverable detail
7. Chapter 5 · Nonlinear fitting and local models
8. Chapter 6 · Design an informative experiment and test its predictions
9. Problem sheet 3 · Nonlinear estimates and experimental design
10. Chapter 7 · Bayesian inversion and the information supplied by a prior
11. Chapter 8 · Predictions, parameter profiles and nonlinear uncertainty
12. Problem sheet 4 · Which uncertainty came from the experiment?
13. Chapter 9 · Adjoint gradients through a forward solver
14. Chapter 10 · Iteration as a regulariser
15. Problem sheet 5 · Derivatives and early stopping
16. Chapter 11 · Constraints, sparse signals and edge-preserving penalties
17. Chapter 12 · A reconstruction audit from observations to a defensible claim
18. Problem sheet 6 · Edges, decision thresholds and validation
19. Chapter 13 · Choosing regularisation when the true answer is hidden
20. Chapter 14 · Learning a drifting temperature from a sequence of readings
21. Chapter 15 · When a single uncertainty interval hides two explanations
22. Chapter 16 · How model error masquerades as a fitted physical parameter
23. Chapter 17 · Predicting a new measurement when the fitted parameters are correlated
24. Chapter 18 · Calibrating two instruments when both axes fluctuate
25. Problem sheet 7 · Project · Which thermometer should the laboratory buy?
26. Problem sheet 8 · Estimate optical thickness when both measurements count photons
27. Inferring parameters and fields from observations
28. Which field can be recovered from boundary data?
29. A measured curve and competing mechanisms
30. Applied study
31. Notation and assumptions
32. Further reading

Chapter 1 · Forward maps, information and identifiability

Decide what a measurement can tell you before attempting to fit it. Separate nonuniqueness, noise sensitivity and model error.

FROM FIRST PRINCIPLES

Decide what a measurement uniquely determines

A forward model predicts readings from a state. In a linear discretization write $y = Ax$; row i lists the sensitivity of the ith reading to each state component. If two states give the same data, subtract their equations. Their difference must be invisible to every measurement,

$$Ax_1 = Ax_2 \quad \Longleftrightarrow \quad A(x_1 - x_2) = 0.$$

Thus, if x_0 is one solution, all exact solutions are $x_0 + h$ for h in the null space of A. A reported quantity $q = c^T x$ can still be unique even if the full state is ambiguous: its value must be unchanged by every invisible displacement. To test this, find a weight vector w that expresses c in terms of the measurement rows. Substituting it into the quantity of interest gives

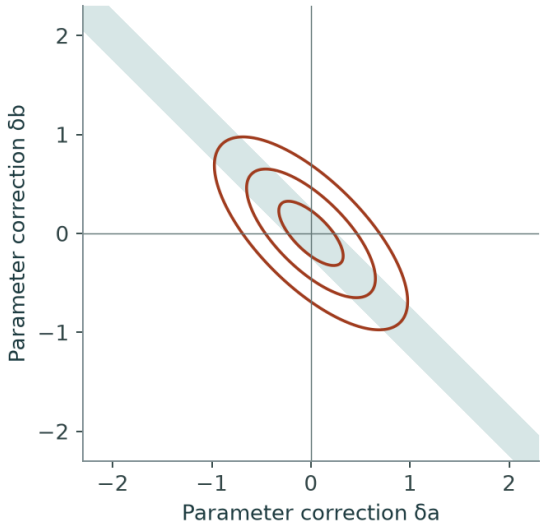
$$\begin{aligned} A^T w &= c, \\ q &= c^T x = (A^T w)^T x \\ &= w^T (Ax) = w^T y. \end{aligned}$$

This equality identifies the quantity directly from the measurements. Small but nonzero sensitivities create a different difficulty: the state can be unique in exact arithmetic while realistic measurement errors leave its numerical value uncertain.

Background · Forward operator

$$y = G(m) + \eta$$

The operator maps parameters to predicted observations. The noise and discrepancy model must use the same physical units as the readout.



When does a second sensor resolve an ambiguity?
The first sensor measures a + b with standard deviation 0.25. Changes along a = −b remain invisible. A second sensor observes a − b with standard deviation 0.60. For independent Gaussian errors, squared residuals in units of their respective variances add. The ellipses show sums of 1, 4 and 9, each constant along its contour. Taking a second independent reading with the first sensor alone would narrow its band by a factor of $\sqrt{2}$ while preserving its invisible direction. Copying the original reading supplies no new information.

DERIVATION AND APPLICATIONS

Separate amplitude and decay rate using two observation times

A single positive reading of $y(t) = Ae^{-kt}$ admits many pairs of amplitude A and rate k. Two measurements at different times can remove that ambiguity, but the spacing matters when readings are imperfect. Assume a single exponential, constant positive amplitude and positive noiseless signal, with independently calibrated times $t_1 < t_2$. Taking the ratio of the two exact data eliminates the unknown amplitude,

$$\frac{y_1}{y_2} = e^{k(t_2 - t_1)},$$

$$k = \frac{\ln y_1 - \ln y_2}{t_2 - t_1},$$

$$A = y_1 e^{kt_1}.$$

This inversion is unique even though an individual observation cannot distinguish the two parameters. The local sensitivity of the inferred rate is obtained by differentiating both logarithms. For small errors and independently known times,

$$\delta k \simeq \frac{\delta y_1 / y_1 - \delta y_2 / y_2}{t_2 - t_1}.$$

If the measurements have independent relative standard deviations s_1, s_2, squaring the coefficients and adding their variances gives

$$\operatorname{sd}(\widehat{k}) \simeq \frac{\sqrt{s_1^2 + s_2^2}}{t_2 - t_1}.$$

A shared multiplicative calibration factor cancels from the ratio exactly. A shared additive background generally does not cancel, particularly once the signal becomes small. Choosing measurements far apart helps at fixed fractional precision; under fixed absolute noise the later signal may lose sufficient precision to reverse that benefit. Two data can fit two parameters exactly and consequently do not check whether a single exponential describes intermediate times.

WORKED EXAMPLE

Recover a decay and test an unused observation

At $t = 1$ and $4\,\mathrm{s}$, a sensor reads 12 and 3 in the same units. The inferred rate is

$$\ln 4/3 = 0.4621\,\mathrm{s}^{-1},$$
$$A = 12e^{0.4621} = 19.05.$$

At $2.5\,\mathrm{s}$, halfway between the observations, the model predicts the geometric mean $\sqrt{12 \times 3} = 6$. A third reading of 7 reveals a deviation that the exact two-point fit could not show.

SECOND APPLICATION

Specify the timing separation required for precision

For a second instrument the fractional standard deviation is 0.030 in both channels. To achieve rate standard deviation $0.010\,\mathrm{s}^{-1}$, the separation must be at least $\sqrt{2}(0.030)/0.010 = 4.24\,\mathrm{s}$ under these noise assumptions. Observations separated by only $0.50\,\mathrm{s}$ give $0.0849\,\mathrm{s}^{-1}$, even though they still determine a unique mathematical solution.

Turn a measurement protocol into an operator

Imagine reconstructing the temperatures of two adjoining cells, x_1, x_2, from sensor readings. A sensor averaging both cells gives

$$y = (x_1 + x_2)/2.$$

The vector x describes physical properties; the forward map predicts what this particular instrument will report. A second instrument may measure the same properties through a different map. Include spatial averaging, response time and calibration in that map if they are significant. Treating a finite-width sensor as an exact point sample can make an accurate experiment look inconsistent.

For a linear experiment write $y = Ax + \epsilon$, separating the ideal response from data error. With the averaging sensor, $x = (20, 24)$ and $(21, 23)$ both give 22. The difference is a multiple of $(1, -1)$, which A maps to zero. The data specify an affine family $x = x_0 + z$, $z \in \ker A$. More decimal places for the very same measurement cannot select one member. A unique reported answer may therefore reflect a chosen convention such as minimum norm or smoothness instead of new physical information.

Build a ray matrix from attenuation measurements

How does a detector reading become a row of a matrix? In a simple transmission scanner, assume a narrow monochromatic beam, negligible scattering into the detector and a known incident intensity I_0. Over a short path segment ds, the fractional intensity loss is $\mu(s)\,ds$ to first order. Dividing the resulting differential equation $dI/ds = -\mu I$ by I, then integrating along the ray γ_i, gives

$$y_i = -\log(I_i/I_0) = \int_{\gamma_i} \mu(s)\,ds = \sum_j \ell_{ij}\mu_j.$$

The last equality assumes the attenuation is constant within each model cell. The coefficient ℓ_{ij} is the length of the path through cell j. An unvisited cell contributes zero. Model refinement changes these lengths and the number of columns. With lengths in centimetres and attenuation in inverse centimetres, each log reading is dimensionless. This construction models individual rays. A finite detector that pools their intensities requires an additional averaging step in the forward model.

Background · Why does a logarithm change the statistical weight of a ray?

Photon transmission with a known incident mean I_0 produces a mean outgoing count $\mu = I_0 e^{-a^T x}$. Geometry becomes linear in optical depth after taking a logarithm. The resulting noise is no longer constant across rays. If the detected count N is Poisson and reasonably large, its variance is μ. Linearising $y = -\log(N/I_0)$ therefore gives

$$\operatorname{Var}(y) \simeq \left(\frac{-1}{\mu}\right)^2 \mu = \frac{1}{\mu}.$$

A ray with 100 expected counts has an optical-depth standard deviation near 0.10, compared with roughly 0.01 for 10,000 counts. Equal weights in the log-transformed fit give these readings equal influence despite their different precision. At very low intensities, zero counts make the logarithm undefined; fitting a count likelihood preserves those observations.

Which extra direction resolves a four-cell ambiguity?

Arrange four square cells of side 1 cm in two rows and list them in row-major order. Two horizontal and two vertical centre rays have log readings $0.4, 0.6, 0.5, 0.5$. Each traversed cell contributes one centimetre. The sum of the two horizontal readings equals the sum of the vertical readings, leaving three independent equations. If the top-left attenuation is t, successive substitutions give the family

$$(\mu_1, \mu_2, \mu_3, \mu_4) = (t,\ 0.4 - t,\ 0.5 - t,\ 0.1 + t) \quad \mathrm{cm}^{-1}.$$

Even positivity admits $0 \le t \le 0.4$. The vector $(1, -1, -1, 1)$ describes a checkerboard change invisible to the original views. Add a diagonal ray joining the outer corners through cells 1 and 4. Each diagonal chord is $\sqrt{2}$ cm. If the reading is $0.3\sqrt{2}$, division by $\sqrt{2}$ gives $2t + 0.1 = 0.3$, so $t = 0.1$. The field is $(0.1, 0.3, 0.4, 0.2)$ in inverse centimetres. The extra row acts on the previous null vector by $2\sqrt{2}$. Its orientation supplies the missing information. Repeating any original ray could improve its measurement precision but would preserve the checkerboard ambiguity.

Practise on sheet 1 · A detector that pools two unresolved rays. Further reading · Technical University of Denmark · Scientific Computing for X-Ray Computed Tomography · Discretization of CT Problems, slides 7, 9–10 and 16, Intersection of Line and Pixel, The Line Model and The System Matrix.

Find the observable combinations

Every row of A is a measured linear combination. Every vector in its null space is orthogonal to all those rows. Thus a derived linear quantity $c^{\mathsf{T}}x$ is uniquely determined by ideal data exactly when c lies in the row space. If $c = A^{\mathsf{T}}w$, compute it as $w^{\mathsf{T}}y$. Conversely, any component of c along a null direction changes the quantity while leaving the data unchanged. This test can rescue a useful question from an underdetermined experiment. Both temperatures may be ambiguous while their mean is measured precisely.

Take a second reading $(x_1 + x_2)/2 + \eta(x_1 - x_2)/2$. Subtract the first reading to isolate the contrast,

$$x_1 - x_2 = 2(y_2 - y_1)/\eta.$$

With $\eta \neq 0$ both temperatures are identifiable in exact arithmetic. At $\eta = 0.01$, a difference error of just 0.02 K changes the recovered contrast by 4 K. Algebraic rank has improved, but the measurement has barely separated the two states. Changing the second sensor's spatial weighting can be much more effective than repeating its reading.

Translate a nonlinear ambiguity into an experiment

For a decay $y(t) = Ae^{-kt}$, an observation at $t = 0$ identifies A and says nothing about k. A single later observation allows the continuum $A = y(t)e^{kt}$. Two exact positive observations at distinct times yield

$$k = \ln[y(t_1)/y(t_2)]/(t_2 - t_1).$$

This identifies both parameters under the assumed model. When the times are almost equal the denominator is small, which makes the slope sensitive to measurement noise. Late measurements can also be difficult when responses approach the instrument's noise floor.

Always state the admissible parameter set. A power reading $y = p^2$ determines $|p|$; if physics requires $p \ge 0$, the sign ambiguity disappears, though sensitivity near zero still needs attention. In later chapters

we will assess local sensitivity with derivatives. A full-rank local derivative distinguishes sufficiently close parameters under suitable regularity; widely separated alternatives can remain.

Establish a baseline for a reconstruction

Write down the unknowns with units, the controlled inputs, the measurements and their precision. Predict what changes when one unknown increases with all others fixed. Look for symmetries and parameter combinations that move predictions together. Finally name a second measurement able to separate them. This sequence gives a reason for the later numerical choices: least squares reconciles inconsistent data, while regularisation controls amplification or deliberately selects among weakly distinguished solutions.

Characterise an identifiable output without reconstructing the state

Suppose an experiment provides exact linear observations $Ax = y$ and a scientist wants a scalar $Q = c^T x$. Any two compatible states differ by a vector v satisfying $Av = 0$. Their outputs differ by $c^T v$. Consequently the output is identifiable for every compatible data set precisely when c is orthogonal to the nullspace. This condition can be checked entirely from the design, before taking a measurement. If it fails, arbitrarily precise measurements with the same operator cannot answer the scientific question unless other assumptions restrict the states.

There is a constructive version. Every row of A is orthogonal to every null vector. The row space has dimension equal to the rank, whereas the nullspace dimension is the number of columns minus the rank. These dimension counts show that all vectors orthogonal to the nullspace lie in the row space. Hence the test is equivalent to solving $A^T w = c$ for observation weights w. When the test succeeds, $Q = w^T A x = w^T y$. The weights specify exactly which combination of instrument readings answers the question without requiring a particular choice of unobserved coordinates.

Recover one observable from overlapping measurements

Three masses sit on a support. One instrument reports $x_1 + x_2 = 7$ kg and another reports $x_2 + x_3 = 9$ kg. Subtracting gives $x_3 - x_1 = 2$ kg. In matrix language the target $c = (-1, 0, 1)^T$ equals the second measurement row minus the first, so $w = (-1, 1)^T$. Independently parameterise all solutions by $x = (7 - t, t, 9 - t)^T$. Every admissible value of t produces the same difference. In contrast the total mass is $16 - t$, so even requiring all masses to be nonnegative leaves totals ranging from 9 to 16 kg. A convenient equal-mass assumption would select a value by adding external information.

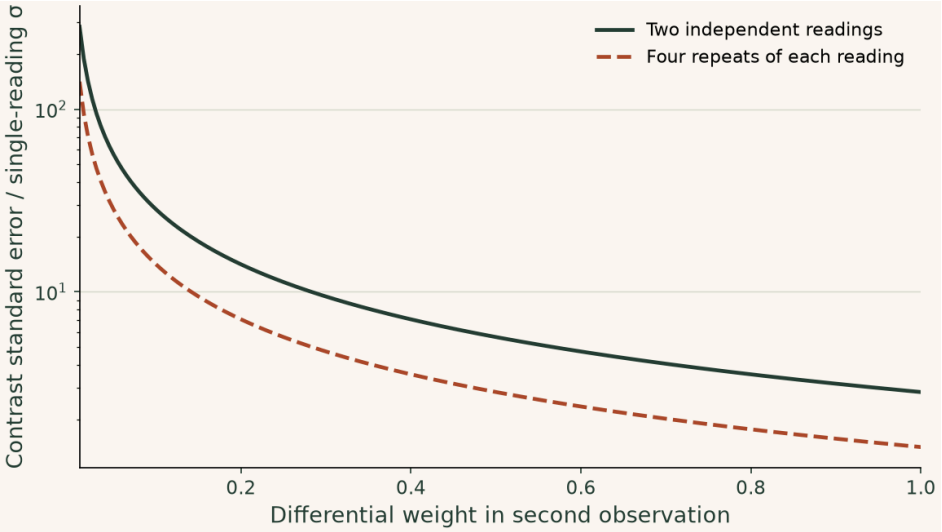
Now give the two readings independent measurement errors of standard deviation 0.20 and 0.30 kg. The same weighted difference has variance $0.20^2 + 0.30^2 = 0.13$ kg² and standard deviation about 0.361 kg. It remains an unbiased estimate of the mass difference. A shared additive zero-offset error cancels in this difference, whereas it would affect an output using the sum. This is a second design question: identifiability determines whether suitable weights exist, while the error covariance determines whether applying those weights yields useful precision.

A third instrument measuring x_2 alone would remove the remaining ambiguity. Measuring the first overlapping pair again would tighten one known sum but retain the null direction $(-1, 1, -1)^T$. When choosing the next measurement, evaluate its dot product with that unresolved direction. A nonzero product measures some of the missing information; a very small product makes the newly identifiable quantity fragile under noise.

For example a reading of $x_1 + 2x_2 + x_3$ merely adds the two original readings. It has nonzero sensitivity to every individual mass yet zero dot product with the missing direction. The instrument can improve precision on a known combination while leaving exactly the same continuum of compatible total masses.

How much information does a second sensor actually add?

The first measurement is the mean of two cells. The second is the same mean plus $a(x_1 - x_2)/2$. Differencing cancels the common mean and allows the contrast to be recovered. Assume independent, identically noisy readings.



At a = 0.1 the contrast error is about 28 single-reading standard deviations. Repeating each condition four times cuts uncertainty in half; changing the second sensor's weight from 0.1 to 0.4 cuts it by four.

▶ **Read values from the plot**

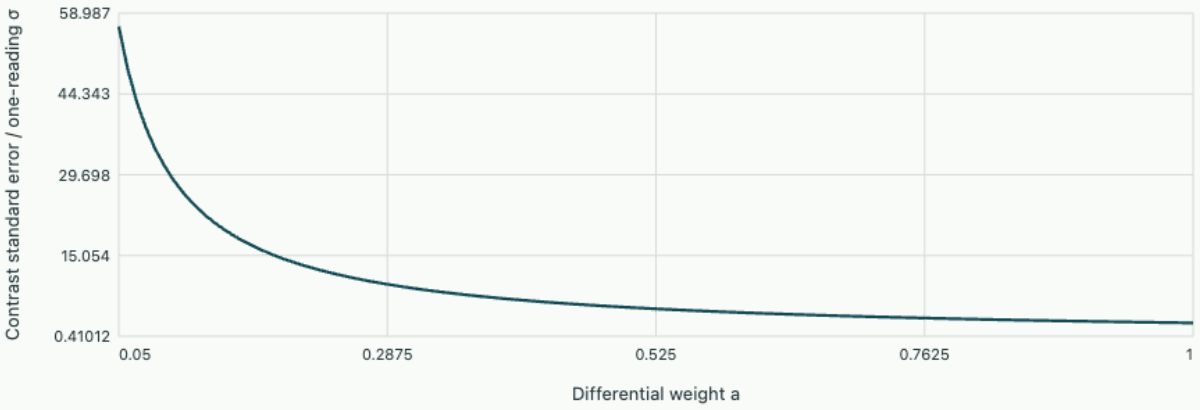
Change parameters · How much information does a second sensor actually add?

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

Independent repetitions	1
Selected differential weight	0.1



Two-sensor contrast

Selected contrast standard error 28.284 times one-reading σ. Repetition improves as 1/√N; changing sensor geometry improves as 1/a.

Study the full derivation

Eliminate linear nuisance parameters before fitting a decay · Inferring parameters and fields from observations

A measured relaxation has an unknown rate, amplitude and background. Derive a one-dimensional profile fit without fixing the nuisance parameters, explain why short observation windows can hide the rate and test an actual numerical dataset.

Chapter 2 · Weighted least squares and residual diagnostics

Derive the fitting criterion from measurement error and separate the uncertainty of an estimated mean from the scatter of a new observation.

FROM FIRST PRINCIPLES

Derive the least-squares distance from the instrument noise

Suppose $y = Ax + \epsilon$, with Gaussian measurement noise of zero mean and known positive definite covariance Σ. Independent errors of unequal sizes should not count equally in raw units, and correlated errors should not be counted twice. Factor $\Sigma = LL^T$ and multiply the model by L^{-1}. The transformed noise has unit covariance,

$$b = L^{-1}y,$$
$$B = L^{-1}A,$$
$$\mathrm{Cov}(L^{-1}\epsilon) = I.$$

For a candidate state, the negative log likelihood differs by a constant from half the squared norm of the transformed residual. This follows from the Gaussian density,

$$\tfrac{1}{2}(y - Ax)^T\Sigma^{-1}(y - Ax) = \tfrac{1}{2}\|b - Bx\|_2^2.$$

The likeliest state therefore minimizes ordinary squared residuals in these whitened coordinates. Solves with the triangular factor L preserve this interpretation without forming its inverse. If the covariance is estimated or state dependent, the stated probability model and resulting uncertainty estimates need to include that extra information.

Background · Whitened residual

$$r = \Sigma^{-1/2}(y - G(m))$$

For a known positive covariance this expresses mismatch in noise units and removes the modelled correlations between measurements.

DERIVATION AND APPLICATIONS

Construct a model check that requires no fitted parameters

A residual can be compared with noise most cleanly when the particular combination of readings was forced to vanish by the model itself. Consider $y = Ax + \epsilon$ with zero-mean noise of known positive

definite covariance Σ. Choose a nonzero contrast c satisfying $A^T c = 0$. No value of the unknown x can change the predicted contrast because its mean is

$$\mathrm{E}(c^T y) = c^T A x = 0,$$
$$\mathrm{var}(c^T y) = c^T \Sigma c.$$

Under a Gaussian noise model its standardized value is a unit normal random variable. Squaring it gives a one-degree-of-freedom discrepancy,

$$Z = \frac{c^T y}{\sqrt{c^T \Sigma c}}, \qquad Z^2 \sim \chi^2_1.$$

The corresponding fit interpretation follows by whitening $b = L^{-1} y,\ B = L^{-1} A$. The vector $u = L^T c / \sqrt{c^T \Sigma c}$ has unit norm and is perpendicular to every column of B. Thus every fitted residual has projection Z along the same direction,

$$u^T (b - B\widehat{x}) = u^T b = Z, \qquad \|b - B\widehat{x}\|^2 \geq Z^2.$$

When there is exactly one independent residual direction, equality holds for the least-squares fit. The contrast is a model check only to the extent that the covariance and zero-mean error assumption are justified. It identifies a combination of inconsistent readings, without uniquely locating a faulty sensor.

WORKED EXAMPLE

Check an affine trend from three measurements

At times $0, 1, 2$ a linear trend predicts zero second difference. Observations $2, 3, 6$, each with independent standard deviation 0.20, give contrast $2 - 2(3) + 6 = 2$. Its standard deviation is $0.20\sqrt{6} = 0.490$, hence $Z = 4.08$ and $Z^2 = 16.7$. Fitting the intercept and slope cannot remove that discrepancy.

SECOND APPLICATION

Include a shared calibration mode in a conservation test

Three independent channels should satisfy a balance with contrast $c = (1, 1, -1)^T$. Suppose their independent errors each have deviation 0.10 and a common additive offset has independent deviation 0.20. Since $c^T (1, 1, 1)^T = 1$, the offset contributes 0.20^2 to contrast variance. For readings $1.0, 2.0, 2.6$ the discrepancy is $0.4 / \sqrt{0.03 + 0.04} = 1.51$ standard deviations. Dropping the shared offset would inflate it to 2.31.

Start from this chapter. Increase the noise scale while keeping temperatures and sampling times fixed. Predict how the covariance of the activation energy changes before refitting the same model.

> OneSimulation laboratory · Find out what a temperature series identifies

A distance in data space depends on the instrument

Let $y = Ax + \epsilon$ with zero-mean error of known positive-definite covariance Σ. If the errors are jointly Gaussian, the negative log likelihood differs from $(y - Ax)^{\mathsf{T}}\Sigma^{-1}(y - Ax)/2$ only by constants independent of x. This expression still defines a weighted least-squares criterion without normality. Correlated readings can vary together substantially while their difference is very precise; individually dividing them by standard deviations fails to capture that information.

Factor $\Sigma = LL^{\mathsf{T}}$ and solve $Lb = y,\ LB = A$ by triangular substitution. Then $b = Bx + \eta$ with $\operatorname{Cov}\eta = I$; $\|b - Bx\|^2$ is the original weighted distance. This is whitening. It also cures a units problem: residuals measured in different units are compared after division by their uncertainty scales. For truly independent readings of standard deviations s_i, the transformation simply divides each row and response by s_i.

Derive the projection and compute it without losing precision

Differentiating the squared norm gives

$$B^{\mathsf{T}}(b - B\hat{x}) = 0.$$

Thus $B\hat{x}$ is the orthogonal projection of b onto the column space. For full column rank write a thin QR factorisation $B = QR$ with $Q^{\mathsf{T}}Q = I$. The minimiser solves $R\hat{x} = Q^{\mathsf{T}}b$. Forming the normal matrix $B^{\mathsf{T}}B$ instead squares its two-norm condition number. QR avoids this extra source of precision loss, though it cannot create information absent from the measurements.

Suppose two instruments measure one constant as 10 and 14 with independent standard deviations 1 and 2. The weighted mean is $(10 + 14/4)/(1 + 1/4) = 10.8$, with standard deviation $(1 + 1/4)^{-1/2} = 0.894$. The less precise instrument still contributes, but with one quarter of the weight. Giving it half the weight would confuse standard deviation with variance. If both instruments share an unknown calibration bias, the stated independence model overstates what averaging can achieve.

Background · Why can two whitening transformations give the same fitted curve?

A symmetric inverse square root and a triangular Cholesky solve generally give different coordinates for the whitened residual. Let both transformations W_1, W_2 whiten the same positive definite covariance Σ. Then the change of coordinates $Q = W_2 W_1^{-1}$ obeys

$$QQ^T = W_2 \Sigma W_2^T = I.$$

Whitened residual vectors are therefore related by an orthogonal transformation. Their squared lengths and hence their least-squares minimisers agree. Two individual whitened components can nevertheless have different magnitudes. If a diagnostic needs to identify which physical sensor was faulty, interpret its residual in the original measurement coordinates alongside the correlation model. A large component obtained after rotation cannot by itself identify a physical channel.

Work with residuals after fitting

The fitted whitened response is Hb with $H = QQ^{\mathsf{T}}$, and residual $r = (I - H)b$ has covariance $I - H$. Consequently raw whitened residuals after fitting do not have equal independent unit variance. Observation i has leverage H_{ii}; a high-leverage point can pull the fit toward itself and show a deceptively small residual. Under the stated model, $\mathbb{E}\|r\|^2 = m - p$ for m observations and p independent fitted coefficients. With Gaussian errors the sum has a chi-square distribution with those degrees of freedom. Plot signed residuals against input, time and sensor identity to look for omitted curvature, drift or serial correlation.

Uncertainty in a fitted trend and a future measurement

In the full-rank linear model $\operatorname{Cov} \hat{x} = (B^{\mathsf{T}} B)^{-1}$. For the model mean $c^{\mathsf{T}} \hat{x}$ the variance is $c^{\mathsf{T}} \operatorname{Cov}(\hat{x}) c$; solve $R^{\mathsf{T}} v = c$ to obtain it as $\|v\|^2$. Predicting a future *individual reading* requires adding its independent measurement variance. This distinction matters when validating fitted response curves. A single point can legitimately lie far outside a narrow interval for the mean. When only relative weights are known, estimate a common variance scale from the residual degrees of freedom and state the additional assumption.

Separate common and independent errors before estimating a mean

Measurements made after one calibration often share a random offset. Write $y_i = \mu + b + e_i$, where the common offset has mean zero and variance τ^2, and the individual errors are independent of one another and of b, with variances s_i^2. The covariance matrix is the diagonal matrix of individual variances plus the rank-one matrix $\tau^2 \mathbf{1}\mathbf{1}^T$. Consider any linear estimate $w^T y$ that reproduces an arbitrary constant, so that $\sum_i w_i = 1$. Its variance separates into $\tau^2 + \sum_i w_i^2 s_i^2$. The first term survives every admissible choice of weights because the shared error is indistinguishable from the unknown mean within this set of readings.

Minimising the second term with a multiplier for the sum constraint gives

$$2 s_i^2 w_i = \lambda.$$

Normalisation yields weights proportional to the individual precisions, with variance $\tau^2 + 1/\sum_i s_i^{-2}$. In particular, estimating the weights from the marginal variances $s_i^2 + \tau^2$ as if every reading were independent gives both an incorrect covariance and generally incorrect relative weights. This derivation uses only second moments. Gaussian assumptions enter if the standard deviation is subsequently turned into a normal-theory probability interval.

Decide whether to repeat a reading or recalibrate

A balance has a random calibration uncertainty of 0.30 g. The first reading is 12.40 g with individual standard deviation 0.10 g. The second, acquired during a shorter averaging window, is 11.90 g with individual standard deviation 0.20 g. The individual precisions are 100 and 25 per g², so the relative weights are 0.8 and 0.2. The estimate is 12.30 g. Its variance is $0.09 + 1/125 = 0.098$ g², giving standard deviation 0.313 g. A long sequence of additional readings using this same calibration can only push that standard deviation toward 0.30 g.

The difference between these two readings removes both μ and the shared offset. Its standard deviation is

$$\sqrt{0.10^2 + 0.20^2} = 0.224\,\mathrm{g}.$$

The observed difference of 0.50 g is about 2.24 such standard deviations. That discrepancy is relevant to the model for individual readings and averaging windows. It supplies no direct evidence for or against a common calibration offset of a particular size, because that offset has cancelled algebraically.

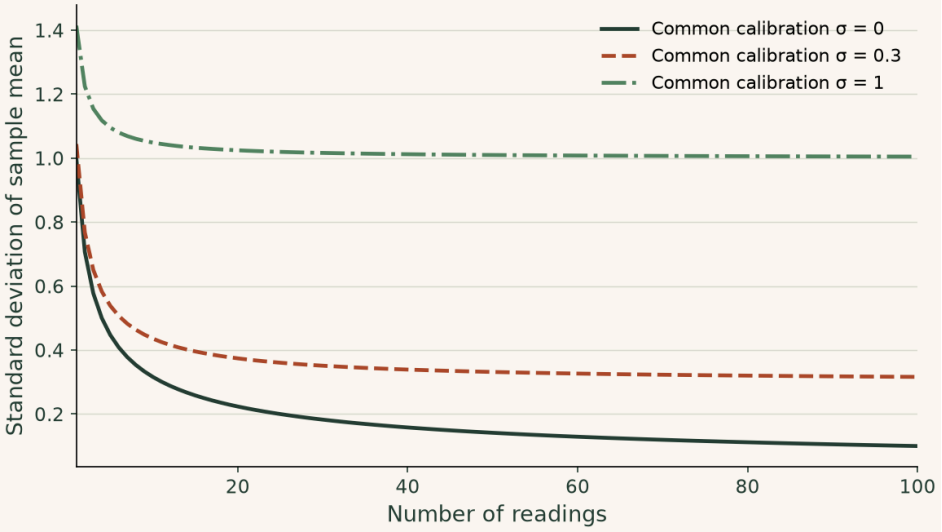
What if the entire two-reading procedure is repeated after an independent calibration? Averaging the two independently calibrated group estimates gives variance $0.098/2 = 0.049$ g² and standard deviation 0.221 g. Taking the same number of readings with one calibration instead gives variance $0.09 + 1/250 = 0.094$ g². The choice of sampling protocol changes the uncertainty even when the balance, total reading count and individual variances are identical. If calibrations share an uncertain reference standard, independence between groups must itself be revised.

A separate reference mass weighed during each calibration provides another way to estimate a shared offset. Subtracting that reference reading removes the offset but also introduces the reference measurement error into every corrected value. The corrected readings are therefore correlated again, now through the shared reference. Their covariance should be recomputed from this modified measurement protocol.

Find a measurement floor that averaging cannot remove

Each reading has independent noise of variance one and the same additional zero-mean calibration offset. Repeating the measurement divides the independent variance by the number of readings. The

calibration error is shared by every term in their average.



With common standard deviation 0.3, averaging one hundred readings with independent sensor noise would still leave total standard deviation about 0.316. Fitting a diagonal covariance and extrapolating a $1/\sqrt{N}$ law misses this limiting uncertainty.

▶ **Read values from the plot**

Change parameters · Find a measurement floor that averaging cannot remove

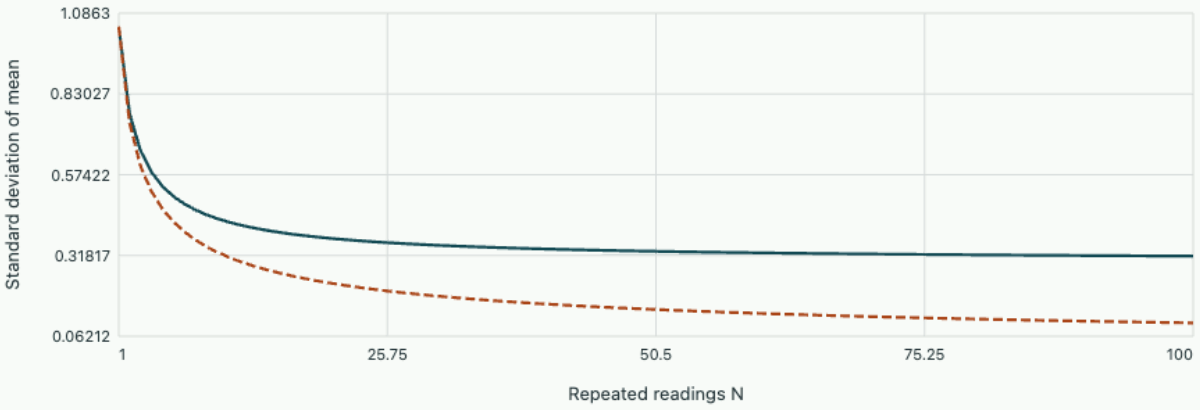
Change one input at a time and compare the recalculated curves with the argument above. The browser evaluates this small model locally; the accompanying numerical values state its assumptions and reference checks. Reset restores the reference inputs, and the data can be downloaded as CSV.

Open this calculation with adjustable inputs

Try the calculation

Shared calibration SD **0.3**

Independent reading SD **1**



Correct common-offset covariance | Incorrect independent-offset assumption

The limiting standard deviation is the shared calibration SD 0.3. Repeating the same instrument does not remove its common unknown offset.

Study the full derivation

Choose a regularisation scale by predicting omitted data · Inferring parameters and fields from observations

Regularised least squares can fit each observation partly with itself. Derive leave-one-out residuals from a single factorisation, approximate them by a trace and identify when a generalised cross-validation score misses a highly influential sensor.

Problem sheet 1 · Which combinations do the data determine?

Use a null-space argument, a covariance model and a residual calculation to assess three proposed measurements.

1. Two instruments on three cells

Let $y_1 = x_1 + x_2 = 5,\ y_2 = x_2 + x_3 = 7$. (a) Write all exact solutions. (b) Determine whether $x_3 - x_1$ and $x_1 + x_2 + x_3$ are identifiable. (c) Name one additional reading that makes the full vector identifiable and prove it.

▼ **Test each quantity along the null direction**

With $x_2 = t$ all solutions are $(5 - t, t, 7 - t)$. Their directional change is $(-1, 1, -1)$. The contrast $x_3 - x_1 = 2 = y_2 - y_1$ stays fixed, while the sum $12 - t$ varies. A reading of the total gives a unique t, since its dot product with the null direction is -1. Repeating either original reading cannot remove this null direction.

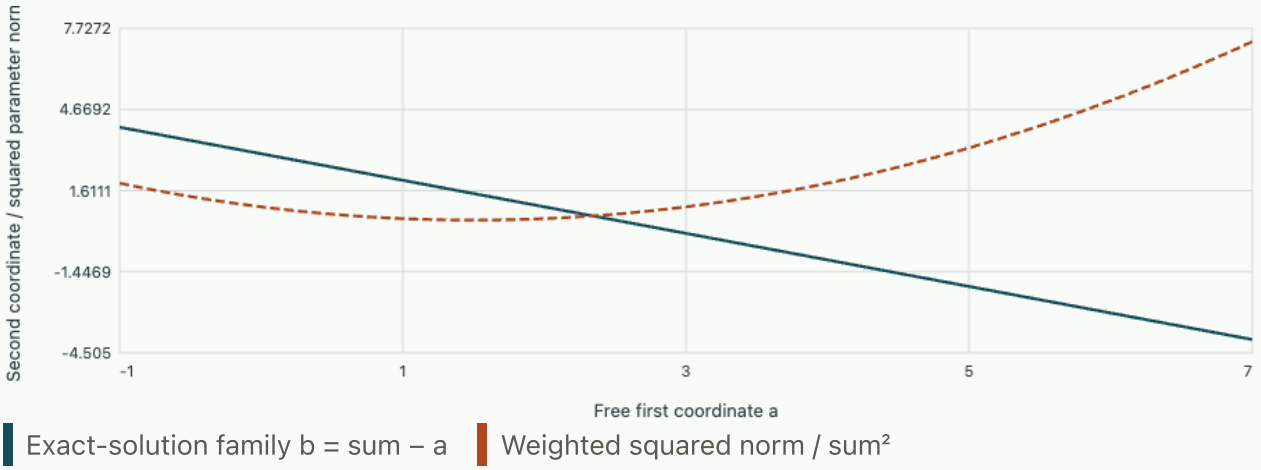
Try this calculation. Explore the continuum of parameter pairs with the same observed sum, and see how a chosen norm selects one pair.

Open this adjustable calculation

Try the calculation

Identified sum a+b **3**

Quadratic weight on second parameter **1**



The minimum weighted norm occurs at (1.5, 1.5). The data identify the sum; the choice within its nullspace comes from the stated parameter metric. The norm curve is normalised by sum².

2. A shared calibration error

Two sensors read the same constant plus independent random errors of variance 1 and a shared random offset of variance 4. All three errors have zero mean and are mutually independent before the shared offset is added to both readings. (a) Write their covariance matrix. (b) Find the variance of their mean and difference. (c) Extend the mean variance to N independent sensor-specific errors with the same shared offset.

▼ **Track the common component once**

Each reading has variance 5 and their covariance is 4. Thus the mean variance is $(5 + 2 \cdot 4 + 5)/4 = 4.5$, while difference variance is

$$5 - 2 \cdot 4 + 5 = 2.$$

For N readings it is $4 + 1/N$ for the mean. Treating all readings as independent would instead predict $5/N$ and would hide the calibration floor. An external calibration can address that floor.

Explore the related calculation with adjustable inputs. Change the parameters in the corresponding calculation and compare its independently recomputed values with the equation derived in this solution.

---

3. A line with room to test its fit

At $x = -1, 0, 1$ observe $y = 0, 2, 1$, with independent unit-variance errors. Fit $y = b + kx$. (a) Find the coefficients and signed residuals (observed minus fitted). (b) Compute the residual sum of squares and its degrees of freedom. (c) Compare variance of the fitted mean and variance of a future observation at $x = 2$.

▼ **Use the orthogonal design columns**

The columns have squared lengths 3 and 2 and zero inner product. The intercept is the mean, 1, and the slope is $1/2$. The fitted responses are 0.5, 1 and 1.5, leaving residuals $(-0.5, 1, -0.5)$ of squared norm 1.5 and one residual degree of freedom. The coefficient covariance is diagonal with entries $1/3, 1/2$. The fitted mean at $x = 2$ has variance $1/3 + 4/2 = 7/3$; an independent future reading there has variance $10/3$.

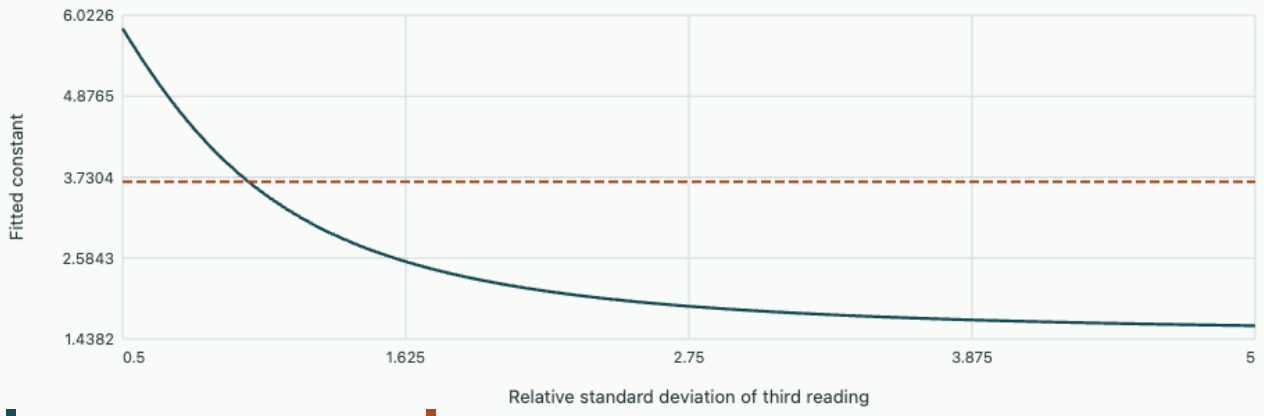
Try this calculation. Change one observation's uncertainty and recompute the weighted projection. More than one fitted coefficient requires the additional basis columns specified in this exercise.

Open this adjustable calculation

Try the calculation

Third reading 8

Third-reading SD / other SD 2



Readings are (1,2,8). At the selected uncertainty ratio the fit is 2.2222. The third weight scales as inverse variance, not inverse standard deviation.

A detector that pools two unresolved rays

A detector averages the transmitted intensities of two equally illuminated rays before reporting one reading. Each traverses 1 cm of material, with attenuation 0.2 and 0.8 cm^{-1}, respectively. Compute the apparent homogeneous attenuation. Compare it with the equal-weight arithmetic mean and explain whether one pooled reading identifies the two attenuations.

▼ **Worked solution**

The normalized transmitted intensities are

$$e^{-0.2} = 0.818731,$$

$$e^{-0.8} = 0.449329.$$

Their mean is 0.634030, so the apparent homogeneous attenuation is

$$-\log(0.634030)/(1\ \text{cm}) = 0.455659\ \text{cm}^{-1}.$$

Averaging the two separately measured log readings would have given $0.5\ \text{cm}^{-1}$.

The pooled result lies $0.044341\ \text{cm}^{-1}$ below the arithmetic mean. Convexity of e^{-x} explains the sign: the more transparent ray contributes disproportionately to the transmitted intensity. An infinite family of

attenuation pairs has this same summed transmission. Two equal attenuations of $0.455659\ \mathrm{cm}^{-1}$ provide another example.

Explore the related calculation with adjustable inputs. Rotate an omitted physical effect relative to the measured directions; treat the nonlinear pooling calculation here as an explicit source of model discrepancy.

Revisit chapter 1 · Build a ray matrix from attenuation measurements.

Chapter 3 · Singular systems and the amplification of noise

Identify which patterns survive a forward measurement and which become dominated by uncertainty when inverted.

FROM FIRST PRINCIPLES

Identify the measured pattern hidden in each singular direction

A forward measurement $y = Ax + \epsilon$ is easiest to interpret in orthonormal directions where A acts by simple scaling. For a singular triplet with positive σ_j, $Av_j = \sigma_j u_j$ and $A^T u_j = \sigma_j v_j$. Project the complete data equation onto its output direction,

$$u_j^T y = \sigma_j v_j^T x + u_j^T \epsilon.$$

Recovering the coefficient in the state divides the measured projection by σ_j. Thus the reconstruction error in that direction is

$$v_j^T(\widehat{x} - x) = \frac{u_j^T \epsilon}{\sigma_j}.$$

When the data are whitened with unit noise covariance, the projected noise has standard deviation one and the recovered coefficient has noise standard deviation $1/\sigma_j$. A zero singular value hides the corresponding state component completely. The numerical rank of a matrix and the set of coefficients that can be inferred at useful precision can therefore differ even for an exactly known forward model.

DERIVATION AND APPLICATIONS

Derive the singular spectrum of a local averaging instrument

An instrument that averages neighbouring locations transmits broad patterns and attenuates rapid changes. Consider a periodic one-dimensional array of N equally spaced cells. At every cell the detector returns a symmetric three-point average, with response coefficient $0 \le a \le 1/2$. With wraparound indices the noiseless map is

$$(Ax)_j = a x_{j-1} + (1 - 2a) x_j + a x_{j+1}.$$

Translation symmetry suggests using normalized Fourier patterns $v_k(j) = e^{2\pi i k j / N} / \sqrt{N}$. Neighbouring samples of one mode differ only by a phase. Substitution yields its eigenvalue,

$$Av_k = \lambda_k v_k,$$

$$\lambda_k = 1 - 2a + 2a\cos(2\pi k/N) = 1 - 4a\sin^2(\pi k/N).$$

The averaging matrix is symmetric and the Fourier modes are orthonormal, so the singular value of each mode is $|\lambda_k|$. A negative eigenvalue reverses the measured phase of a real sinusoidal pattern but has the same noise amplification as a positive one of equal magnitude. A zero destroys the mode. For independent equal-variance data errors of variance s^2, an inverse on a nonzero mode gives

$$\mathrm{E}\,|\delta x_k|^2 = \frac{s^2}{|\lambda_k|^2}.$$

The periodic boundary assumption supplies the Fourier eigenvectors; a finite array with distinct edge rules generally has a different spectrum. Preservation of the constant mode, $\lambda_0 = 1$, checks that the normalized averaging kernel conserves total signal.

WORKED EXAMPLE

Identify an invisible checkerboard pattern

For $N = 8$ and $a = 1/4$, the alternating vector corresponds to $k = 4$ and has $\lambda_4 = 0$. No improvement in readout precision can infer its coefficient from these data. The quarter-cycle mode $k = 2$ instead has $\lambda_2 = 1/2$; noise of standard deviation 0.020 in its measured coefficient becomes 0.040 in the inferred amplitude.

SECOND APPLICATION

Change the optical mixing and forecast a contrast limit

A second array has $a = 0.20$ and even N, giving alternating-mode gain $1 - 4(0.20) = 0.20$. A true normalized contrast coefficient 0.30 appears as 0.060 in the data. At coefficient noise deviation 0.030 that is only a two-standard-deviation response. To infer the same state coefficient with deviation 0.050, the measurement deviation must fall to $0.20(0.050) = 0.010$.

Start from this chapter. Near a single temperature, changing the prefactor can mimic an activation-energy change. Narrow the experimental temperature range and check how the reported standard error changes.

OneSimulation laboratory · Find out what a temperature series identifies

Rotate the coordinates to expose the experiment

For a whitened matrix $B = U\Sigma V^{\mathsf{T}}$, the unit vector v_j is mapped to $\sigma_j u_j$. Parameter patterns v_j and measurement patterns u_j are orthonormal in their respective coordinates. A small singular value means a full unit change in a particular parameter combination makes only a small change in the data. This conclusion depends on scaling: metres or millimetres for one coordinate change the geometry. Scale each parameter by a physically meaningful range before interpreting a singular vector as a weak process.

For nonzero singular values the least-squares minimum-norm inverse takes the form

$$\hat{x} = \sum_{\sigma_j > 0} \frac{u_j^{\mathsf{T}} b}{\sigma_j} v_j.$$

Left-singular directions orthogonal to the range contribute to the residual and cannot be fitted. Right-null directions contribute nothing to the data and are assigned zero by the minimum-norm convention. For a full-column-rank matrix the largest response of the inverse to a unit data perturbation has norm $1/\sigma_{\min}$. A relative worst-case bound involves $\kappa_2 = \sigma_{\max}/\sigma_{\min}$; the actual perturbation direction can be far less damaging.

A blur that almost removes the contrast

Let a sensor mix two unknown levels with $A = \left(\begin{smallmatrix} 0.505 & 0.495 \\ 0.495 & 0.505 \end{smallmatrix}\right)$. It passes the average direction unchanged and attenuates the difference direction by 0.01. For $x = (2, 0)$ it predicts $y = (1.01, 0.99)$. Add the modest perturbation $(0.01, -0.01)$. Inverting the perturbed data exactly gives $(3, -1)$: a 0.01 perturbation per sensor produced a unit change in each reconstructed level. In contrast, perturbation $(0.01, 0.01)$ just raises both fitted levels by 0.01. Both inversions have zero residual. Compare the recovered patterns with independently known precision to decide which interpretation is supportable.

Why finer discretisation can make an inverse problem harder

An integrating or blurring instrument smooths rapidly oscillating parameter fields. Discretising more finely introduces additional oscillatory patterns, each with very small image under the forward map. For a convolution on a periodic grid the Fourier modes themselves diagonalise the operation: the response to each frequency is the Fourier transform of the kernel there. In a Gaussian blur this multiplier falls rapidly with frequency. Formally dividing by it at every resolved frequency amplifies high-frequency noise. Finer output sampling cannot automatically recover what the original acquisition smoothed away.

Plot σ_j, $|u_j^{\mathsf{T}} b|$, and their ratio against index. For recoverable modes the projected signal can be distinguished from the noise floor. When the singular values keep decreasing but the data coefficients fluctuate at roughly constant noise magnitude, division begins to generate structure from noise. The transition can be gradual and projections are random, especially in a short record. Use noise estimates or replicated observations before choosing a cutoff.

Distinguish numerical rank from usable resolution

A singular value large enough to represent in floating point may still give a parameter uncertainty too large for the scientific question. Conversely several poorly measured parameters may admit a well-measured sum. In whitened coordinates the uncertainty of a linear functional follows from how it projects onto v_j. A functional orthogonal to small-singular-value directions can remain stable without the full reconstruction being stable. The next chapter makes the decision to suppress these directions explicit and quantifies the bias it creates.

Ask whether a continuum solution has finite energy

A finite matrix has only finitely many singular values, so a formally inverted vector can always be printed after zero modes have been removed. For a compact smoothing operator on a space of square-integrable functions, there can be infinitely many positive singular values tending to zero. Write the data coefficients as b_j in the left singular basis, with no data component orthogonal to the closure of the range. Solving mode by mode forces the state coefficients to be $x_j = b_j/\sigma_j$. Orthogonality gives

$$\|x\|^2 = \sum_j |b_j/\sigma_j|^2.$$

A square-integrable solution exists exactly when that last series converges; an arbitrary additional nullspace component can then be added if the operator is not injective.

This range condition explains why diminishing data amplitudes are insufficient evidence of a reconstructible state. They must diminish sufficiently fast relative to the strength with which the instrument transmits each mode. The converse is constructive: if the squared quotient series converges then the partial reconstructed states converge in norm, and continuity of the forward operator takes their limit to the original compatible data. A finite truncation conceals failure of this condition because every partial sum is finite.

Follow two spectra as the reconstruction grid grows

Take singular values $\sigma_j = 1/j$, and compare data sets with coefficients j^{-2} and $j^{-3/2}$, respectively. Both data sets have finite squared norm. In the first, the reconstructed coefficients are $1/j$ and the squared state norm is $\sum j^{-2}$, which converges. Beyond a cutoff N, its tail is bounded by $\int_N^\infty x^{-2}dx = 1/N$. Keeping 100 modes therefore leaves state error at most 0.10 in this particular normalisation. A corresponding forward residual can be much smaller because applying the operator divides each tail term by an additional j.

In the second data set, the state coefficients are $1/\sqrt{j}$. Their squared norm through mode N is the harmonic sum. From 100 to 200 retained modes it increases by about $\log 2$, and the same approximate increase recurs on every doubling. Meanwhile the squared data discrepancy after truncation is $\sum_{j>N} j^{-3}$, bounded by $1/(2N^2)$. At 100 modes the data residual is at most 0.0071. The residual can become arbitrarily small although no finite-energy exact state exists for this idealised data set.

In a measured spectrum, only finitely many coefficients are observed and their far tail includes instrument noise. Exact membership in an infinite-dimensional range cannot be proven by drawing a straight line through a short log plot. The continuum calculation instead suggests two independent checks during refinement: inspect whether reconstruction norms stabilise, and compare data coefficients to an independently estimated noise floor before dividing them by small singular values.

For the first sequence even an excellent forward residual is a relatively optimistic diagnostic: the squared forward tail behaves like N^{-3}, while the squared state tail behaves only like N^{-1}. Comparing norms without specifying whether the operator has already smoothed the discrepancy makes the two convergence rates appear contradictory.

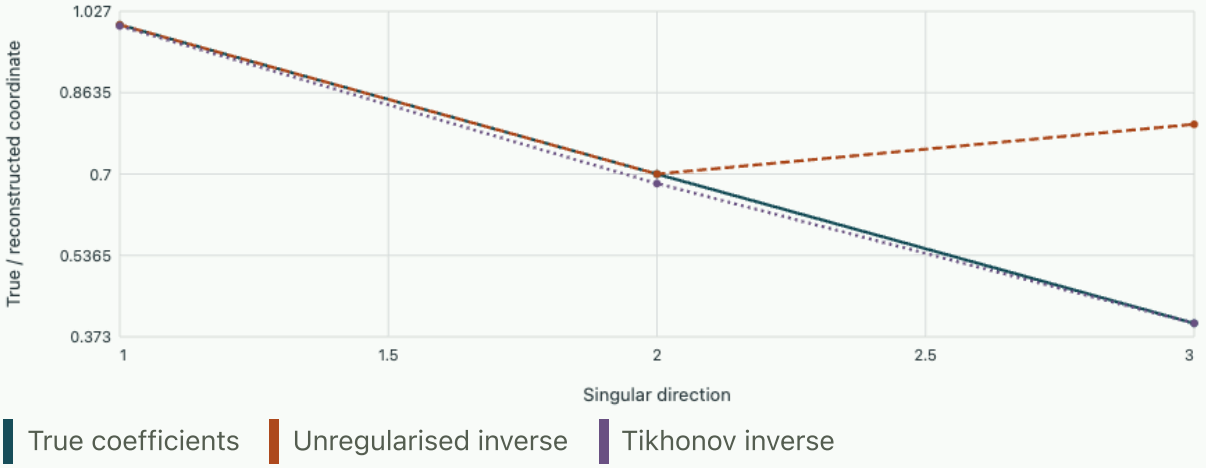
Change parameters · Singular systems and the amplification of noise

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

Weak singular value	0.05
Noise in weak data coordinate	0.02
Regularisation μ	0.05



Weak-mode noise is amplified by 20 before regularisation. The regularised weak-mode resolution gain is 0.5.

Study the full derivation

Recovering sparse coefficients without hiding amplitude bias · Inferring parameters and fields from observations

A sensor records a linear mixture of a few active components. Derive the soft-threshold map, use it in an iterative reconstruction and decide when correlated response columns make an apparently sharp sparse answer ambiguous.

Chapter 4 · Regularisation, resolution and the price of stability

Choose which features to recover, derive the associated filter and quantify the bias introduced by that choice.

FROM FIRST PRINCIPLES

Derive how a quadratic penalty filters small singular values

Suppose whitened data b are modeled by Ax, and large reconstructed coefficients are implausible. Penalize their squared norm with strength $\lambda^2 > 0$. The minimizer balances the measured residual and penalty,

$$\min_x \|Ax - b\|^2 + \lambda^2 \|x\|^2.$$

In singular-vector coordinates let $z_j = v_j^T x$ and $\beta_j = u_j^T b$, where $Av_j = \sigma_j u_j$. Orthogonality makes the objective a sum of scalar terms $(\sigma_j z_j - \beta_j)^2 + \lambda^2 z_j^2$, plus a constant from unmodeled data directions. Setting the derivative of each term to zero gives

$$z_j = \frac{\sigma_j}{\sigma_j^2 + \lambda^2} \beta_j.$$

At $\sigma_j \gg \lambda$ this is close to the ordinary inverse. At a tiny singular value it avoids dividing the noise by that tiny number, at the price of attenuating the true coefficient. A null component is set to zero by this penalty. A different physical reference state or a derivative penalty imposes different information and changes what the reconstruction preserves.

Background · Tikhonov estimate

$$\arg\min_m \{\|Gm - y\|^2 + \alpha \|Lm\|^2\}$$

The regularisation operator selects which departures are costly. Compare values of alpha only after fixing how both terms are scaled.

DERIVATION AND APPLICATIONS

Predict a withheld observation without repeating every regularized fit

Suppose a linear model has independent equal-variance readings y_i, design rows a_i^T, and a fixed positive quadratic penalty $\lambda^2\|x\|^2$. One way to compare penalty strengths is prediction on withheld readings. Define $H = A^TA + \lambda^2 I$, fit all rows to obtain $\widehat{x}$, and let $\widehat{x}_{-i}$ be the fit with row i removed while retaining the same penalty. Adding and subtracting the omitted row in the smaller normal equations yields

$$H\widehat{x} = A^Ty,$$
$$H\widehat{x}_{-i} = A^Ty - a_i(y_i - a_i^T\widehat{x}_{-i}).$$

Subtract the second identity from the first. Multiplying by $a_i^TH^{-1}$ shows that the shift in the fitted prediction is proportional to the withheld residual. The constant of proportionality h_i is its regularized leverage,

$$a_i^T(\widehat{x} - \widehat{x}_{-i}) = h_i(y_i - a_i^T\widehat{x}_{-i}),$$
$$h_i = a_i^TH^{-1}a_i.$$

Writing $r_i = y_i - a_i^T\widehat{x}$ and rearranging gives the leave-one-out residual and corresponding validation score,

$$y_i - a_i^T\widehat{x}_{-i} = \frac{r_i}{1-h_i},$$
$$\mathrm{CV}(\lambda) = \frac{1}{m}\sum_{i=1}^{m}\left(\frac{r_i}{1-h_i}\right)^2.$$

A point with large leverage helps set its own fitted value, so its ordinary residual can be misleadingly small. The formula assumes rowwise omission makes sense for the data. If nearby observations share noise or the intended prediction is a whole new experiment, withholding complete independent groups answers a different question.

WORKED EXAMPLE

Check the shortcut against two explicit refits

Fit a constant to $y = (1, 3)^T$ with $\lambda^2 = 1$. The all-data estimate is $4/3$, and both leverages are $1/3$. Dividing the ordinary residuals by $2/3$ gives withheld errors $-0.5, 2.5$, hence $\mathrm{CV} = 3.25$. Fitting only 3 yields $3/2$; fitting only 1 yields $1/2$, confirming both predictions.

SECOND APPLICATION

Detect a high-leverage reading

For another scalar fit take sensitivities $1, 3$, data $2, 4$ and $\lambda^2 = 1$. The estimate is $14/11$, and the second reading has ordinary residual $2/11$ and leverage $9/11$. Its withheld error is

$$(2/11)/(2/11) = 1.$$

Directly using only the first reading gives $x = 1$, which predicts 3 at the second station rather than the observed 4.

Two ways to control the unstable inverse

A truncated SVD keeps only the first k singular vectors ordered by decreasing singular value. All retained coefficients are inverted exactly; the others are set to zero. Its noiseless reconstruction is a projection onto the retained right-singular subspace. The choice is transparent but a hard cutoff can place two comparably informative directions on opposite sides of the threshold. In some spatial reconstructions sharp spectral cutoffs also introduce oscillations.

A smoother choice is zero-order Tikhonov regularisation. In scaled parameters and whitened measurements minimise $\|Bx - b\|^2 + \lambda^2 \|x\|^2$ for $\lambda > 0$. Differentiation gives

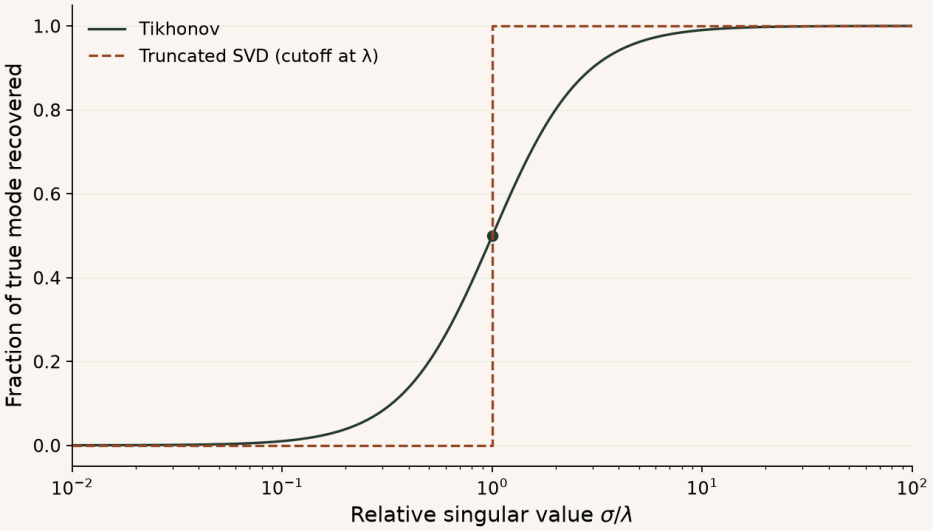
$$(B^{\mathsf{T}} B + \lambda^2 I) x_\lambda = B^{\mathsf{T}} b.$$

Solve with an augmented QR or with the SVD. Each coefficient now carries a filter:

$$x_\lambda = \sum_j \frac{\sigma_j}{\sigma_j^2 + \lambda^2} (u_j^{\mathsf{T}} b) v_j,$$

$$f_j = \frac{\sigma_j^2}{\sigma_j^2 + \lambda^2}.$$

If $b = Bx$ exactly, the estimate of $v_j^{\mathsf{T}} x$ is the true coefficient times f_j. Where $\sigma_j = \lambda$ the noiseless amplitude is halved. For a very small singular value, the noise gain now approaches zero; the price is that a real fine-scale feature in the same direction is also suppressed.



Noiseless resolution of an individual singular mode. A value of one half means that the inversion retains half of the true amplitude, even with perfect data.

Calculate the bias and variance for one component

Suppose a scalar instrument reads $b = 0.1x + \epsilon$, true $x = 2$, with noise standard deviation 0.1. Direct inversion gives an unbiased estimate with standard deviation 1. At $\lambda = 0.1$ the filter is $1/2$; the estimated mean is 1 and the standard deviation is 0.5. Mean squared error is bias squared plus variance: 1.25 here, compared with 1 for direct inversion. For true $x = 0$ the same regularisation would reduce mean squared error from 1 to 0.25. Stabilising an estimate does not guarantee a smaller error for every possible signal. The importance assigned to plausible signals is part of the modelling decision.

Turn a source condition into a noise-dependent accuracy bound

The bias–variance calculation describes a particular signal and a noise distribution. Suppose instead we know only a deterministic data-error bound $\|b^\delta - Bx^\dagger\|_2 \le \delta$. To derive a uniform reconstruction guarantee we must also specify which signals qualify. One useful assumption is the source condition $x^\dagger = B^T w,\ \|w\|_2 \le R$, in scaled coordinates. In singular coordinates a coefficient of $x^\dagger$ is then σ_j times a coefficient of w. Thus arbitrarily large features cannot hide in arbitrarily weak directions for the given R. This assumption also excludes a component in the null space.

Write $\alpha = \lambda^2$ for the penalty coefficient. Split the reconstruction error into the data perturbation passed through the regularised inverse and the loss produced by regularising exact data. In a mode of singular value σ, the noise multiplier is $\sigma/(\sigma^2 + \alpha)$. The noiseless bias is $-\alpha/(\sigma^2 + \alpha)$ times the true coefficient, or $-\alpha\sigma/(\sigma^2 + \alpha)$ times the source coefficient. Because $(\sigma - \sqrt{\alpha})^2 \ge 0$, the first multiplier is bounded by $1/(2\sqrt{\alpha})$. The same inequality bounds the magnitude of the second by $\sqrt{\alpha}/2$. Orthogonality and the triangle inequality now give

$$\|x_\alpha^\delta - x^\dagger\|_2 \le \frac{\delta}{2\sqrt{\alpha}} + \frac{R\sqrt{\alpha}}{2}.$$

If $R, \delta > 0$, the two terms balance at $\alpha = \delta / R$, giving the bound $\sqrt{R\delta}$. This is an a priori parameter choice based on stated uncertainty. For a family of increasingly resolved operators, keeping R controlled matters. If the model class admits large coefficients in directions of vanishing sensitivity, the same claim has no uniform basis.

When bias and data error reinforce one another

Take $B = \operatorname{diag}(1, 0.1)$ and $w = (0, 1)^T$, so $R = 1$, the true signal is $(0, 0.1)^T$, and exact data are $(0, 0.01)^T$. Suppose the second measurement is perturbed downwards by 0.01. Then

$$\delta = 0.01, \ \alpha = 0.01.$$

With exact data, reconstruction would retain half the weak coefficient, producing 0.05 and bias -0.05. The data perturbation is multiplied by $0.1/(0.01 + 0.01) = 5$, adding another -0.05. The reported signal is zero, and the error 0.1 attains the theoretical bound. Here the vulnerable direction has precisely $\sigma = \sqrt{\alpha}$. On another spectrum the achievable error may be much smaller.

Practise on sheet 2 · A source ball and a different weak direction. Further reading · University of Cambridge · Inverse Problems, Mathematical Tripos Part III · §3.2.3 Tikhonov regularisation and §3.2.4 Source-conditions, pp. 37–39.

Choose what the penalty expresses

For an expected background x_0 and a derivative operator D, a suitable criterion may use $\|D(x - x_0)\|^2$. A first-difference penalty discourages adjacent jumps while leaving constants unpenalised; a second-difference penalty instead leaves affine trends unpenalised on a simple uniform open grid. The minimiser is unique if $\ker B \cap \ker D = \{0\}$ for $\lambda > 0$. Check this condition before interpreting a smooth result. Penalising a field whose true behaviour has a sharp interface spreads the interface; inspect the response to a synthetic sharp feature before claiming a physical transition width.

Choose a strength from measurements, then report resolution

If a bound δ for the whitened data error is known, the discrepancy principle seeks a fit whose residual norm is comparable to δ. Smaller λ usually fits data more closely; larger values raise residual and reduce the penalised norm. A noise level estimated in the original physical units must be transformed consistently. If model error alone exceeds the assumed discrepancy, no amount of careful parameter tuning makes the noise model correct. Other choices include validation on genuinely independent observations and examining the residual-versus-penalty curve across several decades.

Report the chosen operator, units, strength and selection criterion. Alongside a reconstruction, show how its resolution changes under a modest change of regularisation. A feature that vanishes under a noise-consistent alternative requires more evidence. A predicted quantity can remain stable despite substantial uncertainty in the underlying image; testing that quantity directly can be the scientifically relevant validation.

Per Christian Hansen develops the singular-value picture and parameter choice further in Discrete Inverse Problems.

When a specified discrepancy determines a unique strength

Consider the zero-background quadratic penalty and express the data in left singular coordinates b_j. Substituting the filtered solution back into the forward model gives residual coefficient $\lambda^2 b_j/(\sigma_j^2+\lambda^2)$, with the sign chosen as data minus prediction. The data component $b_\perp$ orthogonal to the range of the matrix survives unchanged for every strength. Orthogonality therefore expresses the squared discrepancy as $\|b_\perp\|^2+\sum_j |b_j|^2[\lambda^2/(\sigma_j^2+\lambda^2)]^2$, with the sum taken over positive singular values. This equation allows a statement more definite than the heuristic that smoothing worsens the fit.

For each positive singular value the quantity in brackets increases strictly from zero to one as λ grows from zero to infinity. Thus the residual norm increases continuously from the irreducible distance $\|b_\perp\|$ to the full data norm. There is a unique positive finite strength for every target strictly between these endpoints provided at least one measured singular component is nonzero. A target below the first endpoint is infeasible. A target near the upper endpoint can permit a nearly zero reconstruction. The exact same reasoning for general penalties needs care when some physical modes receive no penalty.

Allocate a discrepancy after accounting for an unmodelled reading

A scaled instrument produces $b=(3,1)^T$ for a single unknown x, with predictions $Bx=(2x,0)^T$. The second channel, which should be zero according to the stated model, already consumes one unit of residual norm. Set a total target of $\sqrt{1.36}$ and use the zero-background criterion. Squaring the norm leaves $1.36-1=0.36$ for the squared discrepancy in channel one. The required nonnegative first-coordinate discrepancy is therefore 0.6. Its prediction is $3-0.6=2.4$, so the reconstructed coefficient must be 1.2.

Check by solving the actual penalised problem. The normal equation is

$$(4+\lambda^2)x=6.$$

Substituting $x=1.2$ gives $4+\lambda^2=5$, hence $\lambda=1$. The noiseless resolution factor for that coefficient is

$$4/(4+1)=0.8.$$

This calculation identifies the amplitude tradeoff associated with the selected discrepancy. It also shows why setting the target to 0.9 could never work: the second channel alone contributes one unit for every value of x.

Suppose a second independent reading of the missing channel also repeatedly lies near one, despite a nominal uncertainty of only 0.05 units. The persistent signed defect would be evidence to inspect the forward model or reference subtraction. Treating that behaviour as an excuse to enlarge the noise budget can produce an attainable discrepancy while systematically oversmoothing the channel that does

measure x. Inspect residuals by physical observation as well as their aggregate norm before attributing a discrepancy-selected strength to the uncertainty of the instrument.

In this example the formula for arbitrary strength is $x = 6/(4 + \lambda^2)$. Differentiating it shows that near $\lambda = 1$ a strength increase of 0.1 initially lowers the coefficient by about 0.048. This local sensitivity is another interpretable number to publish if the noise target is itself estimated from a short calibration series.

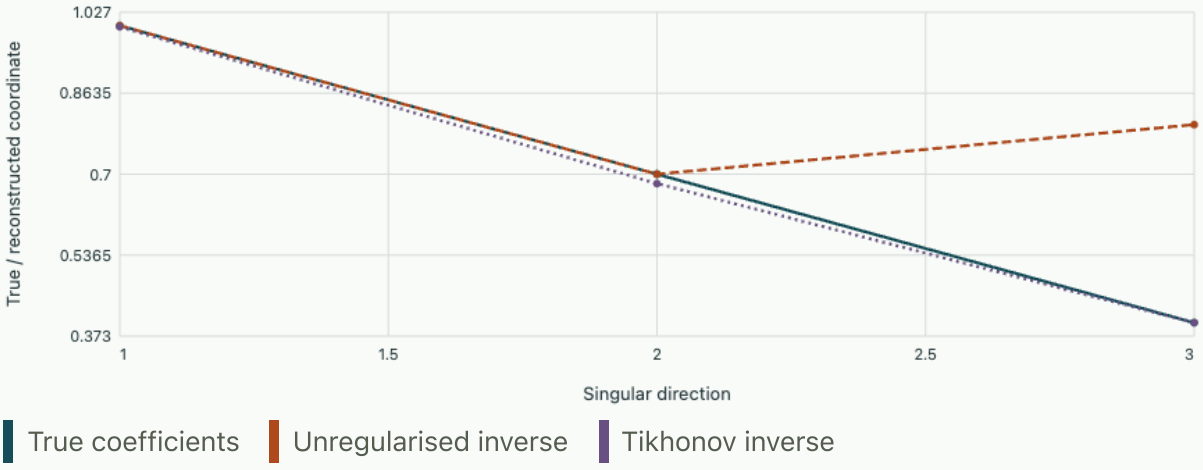
Change parameters · Regularisation, resolution and the price of stability

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

Weak singular value	**0.05**
Noise in weak data coordinate	**0.02**
Regularisation μ	**0.05**



Weak-mode noise is amplified by 20 before regularisation. The regularised weak-mode resolution gain is 0.5.

Study the full derivation

Choose a regularisation scale by predicting omitted data · Inferring parameters and fields from observations

Regularised least squares can fit each observation partly with itself. Derive leave-one-out residuals from a single factorisation, approximate them by a trace and identify when a generalised cross-validation score misses a highly influential sensor.

Problem sheet 2 · Spectral filtering and recoverable detail

Calculate what a reconstruction can borrow from the data and what it assumes when it suppresses a noisy direction.

1. A small residual after a large error

In singular coordinates let $B = \mathrm{diag}(1, 0.02)$, true $x = (1, 1)$, and data perturbation $(0, 0.01)$. (a) Solve the perturbed inverse exactly. (b) Reconstruct with the second mode truncated and compare residuals and errors. (c) Give an example of a stable target functional.

▼ **Compare in parameter and measurement space**

The observed data are $(1, 0.03)$. Exact inversion gives $(1, 1.5)$ with zero residual and parameter error 0.5. Truncation gives $(1, 0)$ with residual norm 0.03 and parameter error 1. In this particular realisation truncation loses a real feature and has greater error. The target x_1 is stable under either estimator; its gain from data noise in this example is one.

Explore the related calculation with adjustable inputs. Change the parameters in the corresponding calculation and compare its independently recomputed values with the equation derived in this solution.

---

2. Choose a discrepancy level

A scalar system has $B = 0.20,\ b = 0.30$. (a) Express the magnitude of the residual under zero-order Tikhonov as a function of λ. (b) Choose λ to give residual 0.06 and compute the estimate. (c) Quantify the resulting noiseless resolution factor.

▼ **Solve for the strength in measurement units**

The fitted response is $b\,0.04/(0.04 + \lambda^2)$, leaving magnitude $0.30\lambda^2/(0.04 + \lambda^2)$. For residual 0.06 solve $\lambda^2 = 0.01$, so $\lambda = 0.10$ and $x_\lambda = 0.20(0.30)/0.05 = 1.20$. The retained noiseless amplitude is

$$0.04/0.05 = 0.8.$$

Interpreting 0.06 as a sensible discrepancy requires an external noise or model-error estimate.

Explore the related calculation with adjustable inputs. Change the parameters in the corresponding calculation and compare its independently recomputed values with the equation derived in this solution.

---

3. Penalising jumps without penalising the mean

For three cell values use $Dx = (x_2 - x_1, x_3 - x_2)$. (a) Find its null space. (b) Show whether measuring $x_3 - x_1$ alone yields a unique regularised solution. (c) Repeat with the mean as the sole measurement, and state what the resulting solution assumes.

▼ **Intersect the null spaces**

The penalty vanishes on all constant vectors. A measurement of $x_3 - x_1$ also vanishes on constants, so adding a constant to every candidate preserves both terms in the objective. A mean measurement sees constants, and the common null space becomes trivial. With observed mean b the unique solution for positive λ is (b, b, b) since both residual and differences then vanish. The flat profile reflects the smoothness assumption; the measurement established only its mean.

Explore the related calculation with adjustable inputs. Inspect soft-thresholding and quadratic shrinkage. For two-cell total variation, apply the threshold to the half-difference and retain the mean as derived here.

A source ball and a different weak direction

Let $B = \mathrm{diag}(1, 0.2)$, $w = (0, 1)^T$, and measured data $b^\delta = (0, 0.0375)^T$, with certified error $\delta = 0.0025$. Compute the balanced α, the bound and the observed error. What happens to the guaranteed accuracy if the noise certificate is quartered while R stays fixed?

▼ **Worked solution**

The signal $B^T w = (0, 0.2)^T$ produces exact weak data 0.04. Their measured perturbation is -0.0025. With $R = 1$, the balanced $\alpha = 0.0025$ guarantees an error at most $\sqrt{0.0025} = 0.05$. The measured weak reconstruction is $0.2(0.0375)/(0.04 + 0.0025) = 0.1764706$, an error of 0.0235294.

The weak singular value 0.2 differs from $\sqrt{\alpha} = 0.05$. Thus the sharp per-mode multiplier in the proof need not be reached. For a new measurement satisfying a quarter-sized certificate, choosing $\alpha = 0.000625$ gives guaranteed error 0.025, only half the former bound.

Explore the related calculation with adjustable inputs. Change the parameters in the corresponding calculation and compare its independently recomputed values with the equation derived in this solution.

Revisit chapter 4 · Turn a source condition into a noise-dependent accuracy bound.

Chapter 5 · Nonlinear fitting and local models

Derive a Gauss–Newton step, recognise misleading local curvature and use step acceptance to protect a nonlinear fit.

FROM FIRST PRINCIPLES

Build a local inverse from sensitivities

Let a differentiable forward model $f(p)$ predict several measurements from parameters p. Whiten its prediction errors by the measured noise covariance, and denote the resulting residual by $r(p)$. To infer a parameter change s from the current mismatch, expand about the current state,

$$r(p+s) = r(p) + J(p)s + O(\|s\|^2).$$

Here column j of J records the response to changing parameter j, with the other parameters fixed at the current values. Minimizing the squared norm of the linearized prediction supplies the Gauss–Newton step,

$$s = \operatorname*{argmin}_d \|r + Jd\|_2^2,$$

$$J^T J s = -J^T r.$$

Independent columns make the local step unique. Nearly parallel columns indicate parameter combinations that can compensate for each other in the data. Because J describes only the present neighbourhood, compare the actual residual at the proposed new state with the linear prediction. A new trial point must also respect any physical parameter bounds assumed by the forward model.

DERIVATION AND APPLICATIONS

Compare finite updates in a parameter and its logarithm

A positive coefficient such as a diffusivity or decay rate can be fitted directly or via a logarithm that enforces positivity automatically. Consider $p > 0$ and use a reference p_0 of the same units so that $q = \ln(p/p_0)$ is dimensionless. Let $r(p)$ be a whitened residual vector and J_p its derivative column at the current parameter. The chain rule and a linearized solve give

$$J_q = pJ_p,$$

$$\delta q = -\frac{J_q^T r}{J_q^T J_q} = \frac{\delta p}{p},$$

$$\delta p = -\frac{J_p^T r}{J_p^T J_p}.$$

Thus direct and logarithmic coordinates describe the same first-order change when the column is nonzero. Their undamped finite proposals differ. The direct step adds the change; the logarithmic step

exponentiates it. Expanding the exponential gives the difference explicitly,

$$p_{\text{direct}} = p + \delta p,$$

$$p_{\log} = p\, e^{\delta p / p} = p + \delta p + \frac{\delta p^2}{2p} + O(\delta p^3 / p^2).$$

The local variance from the same linear model transforms consistently, assuming unit whitened noise and no added prior:

$$\operatorname{var}(\hat{q}) \simeq \frac{1}{p^2 J_p^T J_p},$$

$$p^2 \operatorname{var}(\hat{q}) \simeq \operatorname{var}(\hat{p}).$$

The positivity of a proposal does not ensure it reduces the nonlinear residual. A line search or other acceptance test can still be needed. Likewise, a prior density specified in one coordinate must be transformed with its Jacobian when a Bayesian analysis changes coordinates.

WORKED EXAMPLE

Compare two proposals against an exact one-channel fit

A direct sensor predicts $y = p$, with reading 1 and initial $p = 4$. The direct correction is -3, reaching the exact fit immediately. The log-coordinate correction is $-3/4$, giving $4e^{-0.75} = 1.889$. Both reduce the original error of 3, but the two linearizations have different finite remainders. The logarithmic model is nonlinear even for this linear physical sensor.

SECOND APPLICATION

Translate a local interval into physical units

A second inversion gives $p = 2.00\,\mathrm{mm}^2/\mathrm{s}$ and local standard deviation $s_q = 0.10$ in log coordinates. First-order propagation gives $s_p = 0.20\,\mathrm{mm}^2/\mathrm{s}$. Mapping the symmetric log interval $q \pm 0.20$ exactly gives $p \in [2e^{-0.2}, 2e^{0.2}] = [1.637, 2.443]\,\mathrm{mm}^2/\mathrm{s}$. The asymmetric physical interval retains the bounds specified in logarithmic units.

> **Start from this chapter.** Fit simulated concentration histories and compare the fitted activation energy with the value that generated the data. Repeating with different seeds separates noise scatter from a systematic discrepancy.
>
> OneSimulation laboratory · Find out what a temperature series identifies

Expand the residual around a trial parameter

Let $r_i(p)$ be predicted minus observed data after whitening. Near p_k,

$$r(p_k + d) = r(p_k) + Jd + O(\|d\|^2),$$

where $J_{ij} = \partial r_i / \partial p_j$. Minimise the squared norm of this linear expression to obtain the Gauss–Newton step. The exact Hessian of $\frac{1}{2}\|r\|^2$ is $J^\mathsf{T}J + \sum_i r_i \nabla^2 r_i$. Gauss–Newton omits the second term. It is especially accurate near a good fit with small residuals. With substantial residuals or severe model curvature it can predict the wrong improvement even if the Jacobian is computed perfectly.

Compute the step from a QR or SVD solve $Jd \simeq -r$. Verify a derivative independently using a directional difference $[r(p + hv) - r(p - hv)]/(2h)$ and compare with Jv across a few h. Errors that shrink and then grow as h decreases are consistent with truncation followed by roundoff. A disagreement stable across several step sizes often signals inconsistent units, residual signs or parameter ordering.

Work one update by hand

For $g(t; p) = e^{-pt}$, observe $y(1) = 0.50,\ y(2) = 0.25$ with equal standard deviations. Start at $p = 0$. The residuals are 0.50 and 0.75 and the Jacobian column is $(-1, -2)^\mathsf{T}$. The least-squares step is

$$d = -J^\mathsf{T}r / (J^\mathsf{T}J) = 0.4.$$

At $p = 0.4$ the two predictions are about 0.670 and 0.449. The residual shrinks, but a single linear step falls short of the exact parameter $\ln 2$. Recompute the derivative at the new point before the next update. Fitting the logarithms instead would change the implied measurement-noise model unless the weights are transformed appropriately.

Limit poor steps and enforce physical scales

A damped step minimises $\|r + Jd\|^2 + \mu\|Dd\|^2$. The diagonal scale D makes a step of one natural parameter unit comparable across components. Large μ restricts the step strongly; small μ approaches the ordinary Gauss–Newton solve. Evaluate the true objective at the trial point and compare its actual decrease with the decrease predicted by the linear residual model. A step with negative actual decrease should be rejected or shortened. Repeatedly tiny steps with a large gradient are evidence of a stalled solve, not convergence.

For a strictly positive rate, one can write $k = e^\theta$; then $\partial g/\partial\theta = k\,\partial g/\partial k$. This transformation enforces positivity and represents relative changes naturally. It also changes what a penalty on parameter distance means. Distinguish damping of an individual iteration from regularisation in the final objective: damping may decrease toward zero as a local method converges even when no penalty on the final parameters is intended.

The limits of a local uncertainty ellipse

At a well-determined full-rank fit with correctly scaled independent errors, $(J^\mathsf{T}J)^{-1}$ approximates parameter covariance in a local linear model. For $g(p) = p^2$ and observed positive y, either root has

nonzero slope and a narrow local interval can arise around each. That interval misses the other root entirely. Bounds, multiple starts, profile fits and independent predictions give different checks on global uniqueness and nonlinearity. Distinct parameter values with similar residuals are a result to report; selecting the one reached by the first start can hide a scientifically important ambiguity.

Eliminate linear coefficients from a nonlinear fit

A decay amplitude may be unknown alongside its time constant. For fixed positive rate k the vector of predictions is $av(k)$, with $v_i(k) = e^{-kt_i}$ and an unknown scalar amplitude a. Assume the data have already been whitened, applying the same transformation also to v. Minimising the squared residual at fixed k is an ordinary one-dimensional projection. Differentiating with respect to a gives $v^T(av - y) = 0$, so $a(k) = v^T y/(v^T v)$ whenever the column is nonzero. Substitution leaves a scalar objective $F(k) = \|a(k)v(k) - y\|^2/2$.

To compute its derivative, differentiate the complete residual $r = av - y$. The chain rule gives

$$F' = r^T[a'v + av'].$$

But the optimal amplitude has already made $r^T v = 0$, so the first term vanishes exactly. Thus the profile derivative is $ar^T v'$, even though $a(k)$ changes as the rate changes. This cancellation requires an accurately solved inner projection. If an approximate amplitude leaves a visible residual along the column, ignoring its derivative can yield the wrong outer step. Several linear columns can be handled by the same projection argument while their rank remains fixed.

Diagnose the effect on three observations

Take times 0, 1 and 2 seconds, observed amplitudes $y = (4, 2, 1.5)^T$, and trial rate $k = \log 2$ per second. Then $v = (1, 1/2, 1/4)^T$, $v^T v = 21/16$, and $v^T y = 43/8$. Hence $a = 86/21$. Subtract the observations from the predictions to obtain

$$r = (2, 1, -10)^T/21.$$

Its product with the design column is proportional to $2 + 1/2 - 10/4 = 0$, checking that no amplitude adjustment alone improves the fit. The value of the objective is

$$105/(2 \cdot 21^2) = 5/42.$$

Now $v' = (0, -1/2, -1/2)^T$ in seconds, so

$$r^T v' = 3/14.$$

Multiplying by the fitted amplitude gives $F' = 43/49$ per unit rate in the chosen amplitude scaling. The positive derivative says that slightly reducing k lowers the local discrepancy. This sign has a physical

explanation: the late observation lies above the fitted curve and a smaller decay rate raises the relative late amplitude. The amplitude will readjust simultaneously to keep the projection optimal.

What if all three acquisition times are the same? Every $v(k)$ is a nonzero multiple of the same constant column. Changing k rescales that column and the best amplitude compensates perfectly. The profiled objective is then flat although the inner projection is perfectly well defined. Eliminating a linear nuisance coefficient improves the structure of a search but cannot cure missing information about decay rates when the timing protocol transmits only one measurement direction.

One can diagnose that limiting design before taking data by projecting the derivative column v' perpendicular to v. Only this perpendicular component can change the fitted curve after amplitude compensation. At identical times it is identically zero because $v' = -tv$. With distinct times it usually survives. Under approximately constant additive noise, inspecting its magnitude across candidate sampling schedules gives a quantitative way to choose the times that distinguish a changed shape from a changed amplitude.

The projected derivative has units of amplitude times time, so compare candidate schedules using the same time and amplitude scales.

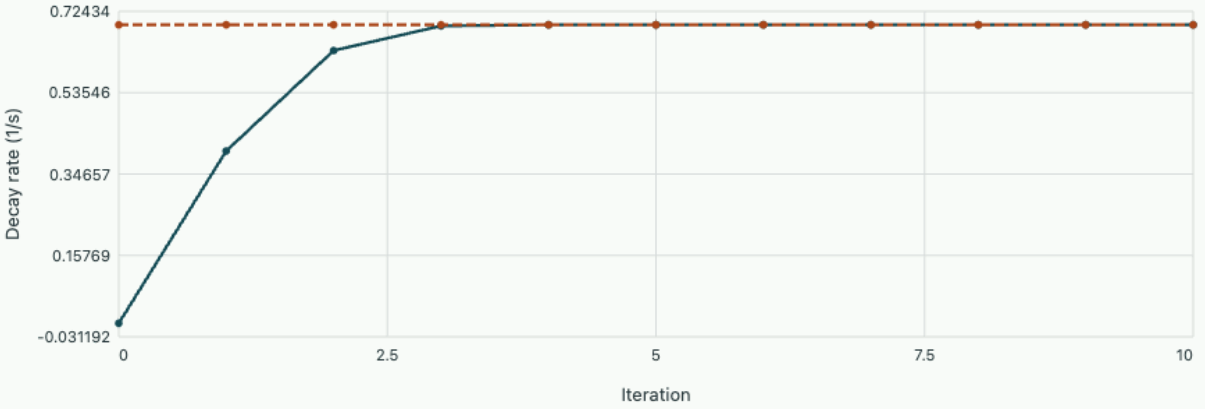
Change parameters · Nonlinear fitting and local models

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

Initial rate (1/s)	0
Second observation offset	0
Damping μ	0



Final rate 0.69315. Changing the second observation makes the two data inconsistent with a single exact decay; the solver then seeks a least-squares compromise.

Study the full derivation

Eliminate linear nuisance parameters before fitting a decay · Inferring parameters and fields from observations

A measured relaxation has an unknown rate, amplitude and background. Derive a one-dimensional profile fit without fixing the nuisance parameters, explain why short observation windows can hide the rate and test an actual numerical dataset.

Chapter 6 · Design an informative experiment and test its predictions

Choose observations that separate parameters, then check uncertainty against the behaviour of new data.

FROM FIRST PRINCIPLES

Derive the information supplied by an observation

Near a reference parameter, suppose a reading of known standard deviation s_i changes by $a_i^T \delta p$. Under independent, zero-mean Gaussian errors, the local squared mismatch for a parameter perturbation is the sum of the squared responses in noise units. Collecting terms gives the information matrix,

$$\sum_i \frac{(a_i^T \delta p)^2}{s_i^2} = \delta p^T H \delta p,$$

$$H = \sum_i \frac{a_i a_i^T}{s_i^2}.$$

If some nonzero direction is perpendicular to every sensitivity row, the experiment has no local information in that direction. A new independent reading with sensitivity a and noise deviation s adds a positive semidefinite term,

$$H_{\text{new}} = H + \frac{a a^T}{s^2}.$$

The new reading is most informative for combinations to which it responds strongly relative to its noise. In the full-rank linear Gaussian model H^{-1} is the parameter covariance. For a nonlinear experiment the rows depend on the reference parameters, so a prospective design should be checked across the range in which the experiment may operate.

DERIVATION AND APPLICATIONS

Value a new observation for a specified prediction

Suppose a calibrated inverse experiment already supplies a full-rank linear estimate with parameter covariance C. An independent additional reading would measure $a^T x$ with noise variance $s^2 > 0$. The information matrices add, giving $C_{\text{new}}^{-1} = C^{-1} + a a^T / s^2$. The decision concerns a prediction $z = c^T x$, so compare its variance before and after the proposed observation. Multiplying out the following candidate inverse verifies the rank-one update,

$$C_{\text{new}} = C - \frac{C a \, a^T C}{s^2 + a^T C a}.$$

The denominator is the total predicted variance of the new reading. Its numerator connects the uncertain target and the proposed noise-free measurement through the existing covariance. Applying c^T and c to either side gives

$$\mathrm{var}_{\mathrm{old}}(z) - \mathrm{var}_{\mathrm{new}}(z) = \frac{(c^T C a)^2}{s^2 + a^T C a}.$$

A candidate with excellent precision can have zero value for this target if the covariance-weighted cross term vanishes. Normalize by the initial prediction variance to express fractional gain as a squared correlation with the actual noisy reading,

$$\frac{\Delta\,\mathrm{var}(z)}{c^T C c} = \frac{(c^T C a)^2}{(c^T C c)(s^2 + a^T C a)}.$$

The fraction is between zero and one by the covariance inequality. These formulas presume independently acquired data and an unchanged linear model. Reusing a correlated reading as if independent overstates the gain and can distort a sensor-placement decision.

WORKED EXAMPLE

Choose between an indirect and a direct reading

Let $C = \left(\begin{smallmatrix} 4 & 1 \\ 1 & 1 \end{smallmatrix}\right)$ and target $z = x_2$. Measuring x_1 with $s = 0.5$ removes $1^2/(4 + 0.25) = 0.235$ from target variance 1. Measuring x_2 with the less precise $s = 1$ removes $1/(1 + 1) = 0.500$. The second location gives the more precise final target despite its larger detector noise.

SECOND APPLICATION

Identify a target-blind but accurate sensor

A separate model has $C = \mathrm{diag}(9, 1)$, and the target is $x_1 + x_2$ with variance 10. A candidate measures $x_1 - 9x_2$ with $s = 0.1$. The cross term is $9 - 9 = 0$, so target variance stays 10. A noisy direct measurement of the target with $s = 3$ instead reduces variance by $100/(10 + 9) = 5.26$, leaving 4.74.

> **Start from this chapter.** Keep five experiments and the same sensor noise. Change their temperature spacing to alter the sensitivity directions, then compare the resulting uncertainty in activation energy.
>
> OneSimulation laboratory · Find out what a temperature series identifies

Express design as a change in sensitivity

For independent whitened observations a candidate measurement with row derivative j adds $j^{\mathsf{T}} j$ to the local information matrix. A repeat of the same measurement increases information along its existing direction, while a row with a new orientation can raise rank. After defining physically meaningful

parameter scales, one can maximise the smallest eigenvalue to protect the weakest direction, or maximise the determinant to shrink a local uncertainty volume. If the scientific target is a single prediction $h(p)$, minimising $\nabla h^{\mathsf{T}}(J^{\mathsf{T}}J)^{-1}\nabla h$ at full rank may select a different design. Choose a criterion for the quantity actually needed.

Derive temperature and time sensitivities

Consider $c = c_0 e^{-kt}$ with $k = k_r \exp[-(E/R)(1/T - 1/T_r)]$. Estimate $\theta = \ln k_r$ and E, taking c_0 as known. At fixed T the derivatives are

$$\partial_\theta c = -ktc,$$
$$\partial_E c = (ktc/R)(1/T - 1/T_r).$$

Within one temperature series all sensitivity rows are parallel in this two-parameter space, however many sampling times are used. Exactly at T_r the second component vanishes. Temperatures on opposite sides of the reference rotate the rows in opposite directions and can reduce parameter correlation. For equal absolute concentration uncertainty, the factor kte^{-kt} peaks at $kt = 1$. A wider temperature range changes the rate as well as the orientation, so a time schedule informative at one temperature can miss the decay at another.

OneSimulation · When should the second thermometer reading be taken?

A vessel cools toward a known ambient temperature. The temperature excess is Ae^{-kt}, with nominal amplitude 4 K and nominal rate 0.4 per hour. Two independent readings, each with standard deviation 0.08 K, are available. One is taken at time zero. If the amplitude were known exactly, the time sensitivity to the rate would peak in magnitude at $t = 1/k$. But the first reading estimates the unknown amplitude and introduces uncertainty into the rate estimate.

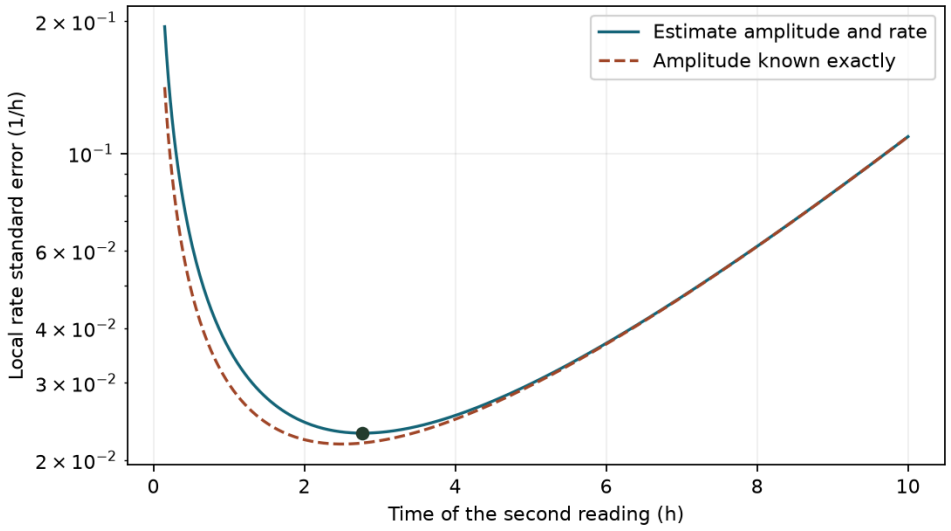
The OneSimulation solve computes the local two-parameter covariance from the sensitivity matrix. Another derivation comes from taking the log ratio of the readings and propagating their errors to first order. The variance of the estimated rate is

$$\operatorname{Var}(\hat{k}) \simeq \frac{\sigma^2}{A^2 t^2}\left(e^{2kt} + 1\right).$$

Writing $z = kt$, the minimum satisfies $z = 1 + e^{-2z}$. Thus the preferred time is about 2.77 h, slightly later than the 2.50 h optimum with a known amplitude. Two close readings hardly discriminate a rate; a very late reading is nearly lost in absolute instrument noise. The curve is a local design calculation at the nominal rate, before observing any temperatures.

```
"""Choose a sampling time for a cooling rate when its amplitude is also unknown."""
import matplotlib.pyplot as plt
import numpy as np
from onesim.numerics.linear.matrices.dense_linear_algebra import solve_dense

amplitude = 4.0  # K above a known ambient temperature
rate = 0.4  # per hour
noise = 0.08  # K, standard deviation of each independent measurement
```

```
times = np.linspace(0.15, 10.0, 400)  # first reading at 0 h; vary the second
standard_errors = []
for t in times:
    decay = np.exp(-rate*t)
    sensitivity = np.array([[1., 0.], [decay, -amplitude*t*decay]])
    covariance = solve_dense(sensitivity.T @ sensitivity / noise**2, np.eye(2))
    standard_errors.append(np.sqrt(covariance[1, 1]))
known_amplitude = noise * np.exp(rate*times) / (amplitude*times)
analytic = noise * np.sqrt(1 + np.exp(2*rate*times)) / (amplitude*times)
assert np.allclose(standard_errors, analytic)
best = np.argmin(standard_errors)
print(f"Optimal second reading on grid: {times[best]:.2f} h; rate SE {standard_errors[best]:.4f} /h")
fig, ax = plt.subplots(figsize=(6.8, 3.8), constrained_layout=True)
ax.semilogy(times, standard_errors, color="#176879", label="Estimate amplitude and rate")
ax.semilogy(times, known_amplitude, "--", color="#a4492b", label="Amplitude known exactly")
ax.scatter(times[best], standard_errors[best], color="#273d30", zorder=3)
ax.set(xlabel="Time of the second reading (h)", ylabel="Local rate standard error (1/h)")
ax.legend()
ax.grid(alpha=0.2)
fig.savefig("inverse-problems.png", dpi=160)
```



The solid curve allows for estimating the initial temperature. Early sampling is particularly expensive once that additional unknown is included.

Download the Python example

Compare an inference with three numerical experiments

The OneSimulation Arrhenius experiment supplies three fits to noisy synthetic histories. First compare two runs at the wider temperature range: doubling the noise scale changes the reported standard error of E from 0.3636 to 0.7228 kJ/mol. Next restore the original noise and narrow the spacing between temperatures from 5 K to 1 K. The standard error reaches 1.4300 kJ/mol. The second change alters the geometry of the sensitivity rows, even though the number of observations stays the same. The exact factor also depends on the sampling times and nonlinear fit.

The known generating value is 115 kJ/mol. A single fitted standard error describes expected spread under local assumptions. It is not a bound on any particular realised error. Vary the seed and record both the

fitted estimate and its reported uncertainty. Compare the empirical spread with the average nominal error and inspect the distribution of standardised errors. A systematic shift under repeated seeds suggests estimator bias or a mismatch in the analysis; one unusual seed alone cannot settle that question.

Validate the model where it will be used

Hold out an entire temperature curve to test the predicted temperature dependence. Randomly omitting isolated time points from every curve mostly tests interpolation within already represented conditions; it also risks leakage if neighbouring errors are correlated. Fit only the training data, make the held-out prediction with its uncertainty, and plot signed prediction errors against time. If different temperatures systematically show opposite curvature errors, revisit the rate model before reporting more digits in E.

A reproducible inference records the forward equations, units, noise law, parameter scales, data split, residual plots and the specific prediction the fit supports. Parameter identifiability is tied to the observations available. Adding a second measurement condition can change the answer more than switching between two competent optimisation algorithms.

Choose a design for the prediction that will be reported

Consider a linear calibration $y = a + bx + \epsilon$ on an instrument with two available settings $x = -1, +1$. Assume independent errors of common known variance σ^2. Any number of repetitions at the same setting estimates its mean response. Denote the two setting means by m_- and m_+. The calibration then predicts the value at any target t by linear interpolation or extrapolation,

$$Q(t) = (1 - t)m_-/2 + (1 + t)m_+/2.$$

This expression follows by equating the coefficients of the constant and linear functions, both of which must be reproduced exactly.

With n_- and n_+ readings, the variance is $\sigma^2[(1 - t)^2/(4n_-) + (1 + t)^2/(4n_+)]$. For a fixed budget $N = n_- + n_+$, minimise over positive group sizes. Taking the derivative with respect to n_-, or applying Cauchy–Schwarz, gives

$$n_- : n_+ = |1 - t| : |1 + t|.$$

The absolute values matter outside the calibration interval, where one interpolation weight becomes negative while its contribution to variance remains positive. This result assumes the slope law is trusted at the target and both settings cost the same.

Allocate twelve readings for a target near one endpoint

At $t = 1/2$ the target weights are $1/4$ and $3/4$. The optimal division of twelve readings is therefore three at the negative setting and nine at the positive setting. For unit measurement variance, the resulting prediction variance is

$$1/(16 \cdot 3) + 9/(16 \cdot 9) = 1/12.$$

Splitting the readings equally instead gives

$$(1 + 9)/(16 \cdot 6) = 5/48.$$

The first design reduces this particular variance by one fifth. Both designs can reconstruct the full line because both endpoints have been observed.

If the target were a tight joint estimate of intercept and slope, the answer would change. The information matrix in these two parameters has diagonal entries N and off-diagonal entries $n_+ - n_-$ for unit variance. Its determinant is $N^2 - (n_+ - n_-)^2 = 4n_-n_+$, maximised by a six–six design. For the three–nine design the determinant is 108 versus 144 when balanced. Giving up joint precision here buys precision specifically for the proposed positive-side prediction. A design decision cannot be assessed without knowing which quantity is intended for publication.

As a further check, allow the physical calibration to contain an unmodelled quadratic term dx^2. At both design points that contribution equals the same constant d, so it is absorbed into the fitted intercept. At $t = 1/2$ its actual contribution is only $d/4$. Every design using the two endpoints alone therefore incurs prediction bias $3d/4$, however small its nominal variance. An additional interior validation setting can reveal this defect in a way that more repeated endpoint measurements cannot.

A further constraint may come from instrument startup time. If changing between settings is expensive while taking repeated readings is cheap, run the readings in groups but randomise or balance their order across independent sessions. Otherwise a gradual drift can become indistinguishable from the slope. The independent-error variance derivation offers no protection against an unmodelled time effect confounded with the chosen temperature or excitation.

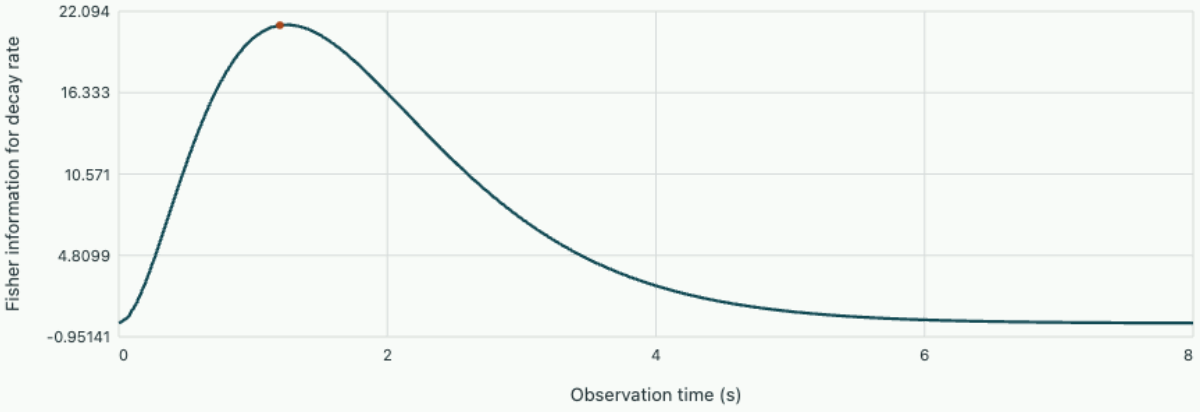
Change parameters · Design an informative experiment and test its predictions

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 rate (1/s)	**0.8**
Observation noise SD	**0.1**
Selected observation time (s)	**1.2**



One known-amplitude decay reading Selected observation time

Maximum information occurs at t = 1/k = 1.25 s. At the selected time the local rate SD is 0.21764 1/s. Unknown amplitude or correlated noise changes this design problem.

Study the full derivation

Choose a regularisation scale by predicting omitted data · Inferring parameters and fields from observations

Regularised least squares can fit each observation partly with itself. Derive leave-one-out residuals from a single factorisation, approximate them by a trace and identify when a generalised cross-validation score misses a highly influential sensor.

Problem sheet 3 · Nonlinear estimates and experimental design

Distinguish a successful optimisation step from a reliable parameter estimate and a prediction that generalises.

1. One step, two possible answers

For $r(p) = p^2 - 4$, (a) compute the Gauss–Newton steps from $p = 1$ and $p = -1$. (b) Compare the actual and linear predicted residuals after the positive-start step. (c) Explain what happens from $p = 0$, and why one local standard error cannot describe all possibilities.

▼ Inspect the slope at each starting point

The step is $-r/(2p)$ for $p \neq 0$. It is $+1.5$ at 1 and -1.5 at –1. From 1 the linear model predicts residual zero, whereas $r(2.5) = 2.25$. Its magnitude still decreased from 3. At zero the derivative vanishes, and the linear model predicts no improvement in any direction. The two separated exact solutions $+2, -2$ produce distinct local uncertainty regions unless physics fixes the sign.

Explore the related calculation with adjustable inputs. Explore how initial values and the residual slope affect Newton and secant convergence on the quadratic reference problem; retain the stated residual when checking the present hand calculation.

---

2. Which two measurement conditions?

In dimensionless whitened coordinates each design may add two rows. Compare $J_A = \left(\begin{smallmatrix} 1 & 0 \\ 1 & 0 \end{smallmatrix}\right)$ and $J_B = \left(\begin{smallmatrix} 1 & -0.5 \\ 1 & 0.5 \end{smallmatrix}\right)$. (a) Determine each information matrix and rank. (b) Find the uncertainty of the second parameter under design B and after repeating the whole design four independent times. (c) Relate the calculation to measuring decay only at a reference temperature.

▼ Compute the information added by each row

Design A gives $\mathrm{diag}(2, 0)$, rank one. Design B gives $\mathrm{diag}(2, 0.5)$, rank two. In the local linear unit-noise model the second parameter has variance 2, hence standard deviation $\sqrt{2}$. Four independent copies multiply information by four and halve that standard deviation. At the reference temperature the Arrhenius activation derivative vanishes. More sampling at that one temperature cannot recover the missing direction.

Explore the related calculation with adjustable inputs. Change the parameters in the corresponding calculation and compare its independently recomputed values with the equation derived in this solution.

---

3. Read a fit as a prediction

Use the temperature-design experiment. (a) Compute the narrow-to-wide standard-error ratio for E at fixed noise. (b) Compare the wide-range fit error with its own standard error using the known 115 kJ/mol truth. (c) Design a validation that tests a new temperature and one that tests noise calibration.

▼ **Change a single condition per comparison**

The ratio is

$$1.4300/0.3636 = 3.93.$$

For the wide, lower-noise fit the signed error is $114.0367 - 115 = -0.9633$ kJ/mol, about -2.65 reported standard errors. This is one draw; it does not estimate the long-run coverage fraction. To assess transfer, hold out a complete temperature series, fit the others and predict that entire curve. To assess noise calibration, repeat the same design with many independent seeds and compare scatter or interval coverage against the local uncertainty calculation.

Choose between repeating and spreading measurements

At fixed noise the wide five-temperature design gives a local standard error of 0.3636 kJ/mol and the narrow design gives 1.4300 kJ/mol. Suppose independent complete replicas of the narrow design are possible. Derive how many are needed to equal or beat the wide design, using the linearised least-squares covariance at a fixed nominal parameter.

▼ **Work through the solution**

Whiten the observations and call the design Jacobian J. Stacking n identical independent designs adds n copies of $J^T J$. The covariance is its inverse, so the standard errors shrink by $1/\sqrt{n}$. Require $1.4300/\sqrt{n} \leq 0.3636$, giving $n \geq 15.47$, hence 16 complete sets, or 80 temperature experiments. This calculation uses the reported errors as local estimates. Shared calibration offsets or a changed nominal fit can invalidate the independent-replica scaling.

Use the OneSimulation laboratory to check your answer

Chapter 7 · Bayesian inversion and the information supplied by a prior

Combine an explicit distribution for the parameters with a noise model. Calculate exactly which uncertainties an observation reduces.

FROM FIRST PRINCIPLES

Derive a posterior from a prior and a noisy reading

Suppose the scalar parameter x has Gaussian prior mean m_0 and variance $v_0 > 0$, and a sensor reads $y = x + \epsilon$, with independent Gaussian error of variance $s^2 > 0$. Bayes' rule multiplies the prior density by the probability of obtaining the reading at each trial x. Its negative log, apart from constants, is

$$-\log p(x \mid y) = \frac{(x - m_0)^2}{2v_0} + \frac{(y - x)^2}{2s^2} + \text{constant}.$$

The coefficient of $x^2/2$ gives the inverse posterior variance. Matching the linear coefficient after completing the square determines the posterior mean,

$$v_1 = \left(\frac{1}{v_0} + \frac{1}{s^2}\right)^{-1},$$

$$m_1 = v_1\left(\frac{m_0}{v_0} + \frac{y}{s^2}\right).$$

Greater precision means larger weight. An extremely uncertain reading leaves the prior nearly unchanged, while an extremely precise measurement concentrates the posterior near the reading. These statements describe the specified prior and instrument model. They depend on the reliability of both inputs.

DERIVATION AND APPLICATIONS

Predict a future reading using the full posterior uncertainty

A posterior can describe both an unknown state and measurements still to come. Suppose current data give $x \mid y \sim N(m, C)$. A new independent instrument will observe $y_* = a^T x + \epsilon_*$, with known zero-mean Gaussian noise variance s_*^2. Conditional on a fixed x, the only variability is instrumental. Averaging over the posterior makes the expected new reading

$$\mathrm{E}[y_* \mid y] = \mathrm{E}_{x|y}[\mathrm{E}(y_* \mid x)] = a^T m.$$

Apply the law of total variance. The expected conditional variance is the new instrument variance, and the variance of its conditional mean is the uncertainty in the latent signal,

$$\mathrm{var}(y_* \mid y) = \mathrm{E}_{x|y}[s_*^2] + \mathrm{var}_{x|y}(a^T x) = s_*^2 + a^T C a.$$

The new reading is Gaussian because it is a sum of independent Gaussian contributions. Two future readings along the same vector a share the latent uncertainty while retaining independent measurement errors. For n such readings of equal precision, their predicted mean has

$$\operatorname{var}(\bar{y}_* \mid y) = a^T C a + \frac{s_*^2}{n}.$$

Increasing the number of future repeats reduces only the noise of their average in this prediction. Once actually observed, those data can be used to update the state further. A predictive check is most informative when the new observations were excluded from the original inference and the instrument model applies under the future conditions.

WORKED EXAMPLE

Compare a latent interval and a future reading

A posterior for one temperature has mean $20.0\,^{\circ}\mathrm{C}$, variance $0.16\,\mathrm{K}^2$. A new thermometer has deviation $0.30\,\mathrm{K}$. Its predicted reading deviation is

$$\sqrt{0.16 + 0.09} = 0.50\,\mathrm{K}.$$

A new reading of $21.0\,^{\circ}\mathrm{C}$ lies two predictive standard deviations above the mean. Dividing its discrepancy by the latent deviation $0.40\,\mathrm{K}$ would assess a different quantity.

SECOND APPLICATION

Forecast the precision of an averaged observation

A separate model predicts a linear contrast with $a^T C a = 0.09$, and a single future reading has noise variance 1.44. Averaging 16 readings gives predictive deviation $\sqrt{0.09 + 1.44/16} = 0.424$. Even infinitely many future repeats would leave a prior-to-measurement predictive deviation 0.300. To reduce it to 0.350 would require $n \geq 1.44/(0.350^2 - 0.09) = 44.31$, hence at least 45 readings.

From one answer to a distribution

In an experiment with $y = Ax + \epsilon$, let the unknown have a Gaussian prior with mean m_0 and covariance C_0, and let measurement error be independent Gaussian with mean zero and covariance Γ. The prior describes the state of knowledge before seeing these data. Bayes' rule multiplies the likelihood by that prior and normalises over parameters. For positive-definite covariances the negative log posterior, up to constants, is

$$\tfrac{1}{2}(y - Ax)^{\mathsf{T}}\Gamma^{-1}(y - Ax) + \tfrac{1}{2}(x - m_0)^{\mathsf{T}}C_0^{-1}(x - m_0).$$

The minimiser is the maximum a posteriori estimate. This objective has the same algebra as Tikhonov regularisation centred on a background. The probabilistic interpretation adds assumptions: if the distributions are taken seriously their scales and correlations imply uncertainty, not just a convenient

stabiliser for a solver. A guessed smoothness penalty cannot be given calibrated probability intervals without justifying its scale.

Complete the square

Expanding the quadratic in x gives precision $C_{\text{post}}^{-1} = C_0^{-1} + A^{\mathsf{T}}\Gamma^{-1}A$ and mean $m_{\text{post}} = C_{\text{post}}(C_0^{-1}m_0 + A^{\mathsf{T}}\Gamma^{-1}y)$. Since the exponent is quadratic and the precision positive definite, the posterior is Gaussian and its mean is also its mode. For implementation, factor and solve the precision system. When data dimension is much smaller than parameter dimension, the equivalent data-space update can be cheaper:

$$m_{\text{post}} = m_0 + C_0A^{\mathsf{T}}(AC_0A^{\mathsf{T}} + \Gamma)^{-1}(y - Am_0).$$

The same matrix multiplying the data discrepancy determines how far knowledge moves from the prior. A reading much more precise than the prior prediction has high influence in its observed direction; an observation compatible with a broad noise distribution does less. In this linear Gaussian setting posterior covariance depends on design and noise, while the posterior mean also depends on the realised data values.

Recover a sum and ask about the difference

Take two independent prior components with means zero and variances 1. Observe their sum as 3 with noise variance 1. The predicted observation variance is $1 + 1 + 1 = 3$, so the update gives $m_{\text{post}} = (1, 1)$. The posterior covariance is $I - \frac{1}{3}\left(\begin{smallmatrix}1 & 1\\ 1 & 1\end{smallmatrix}\right)$. Thus the sum has posterior variance $2/3$, down from 2, while the difference still has variance 2. A unique best estimate of the difference exists (zero), but here the measurement supplied no information about it. Reporting only the point estimate would conceal that distinction.

Interpret prior correlations and check their consequences

If nearby temperatures have positively correlated priors, observing one cell can reduce uncertainty in another even when its forward sensitivity is zero. That reduction is conditional on the correlation model. Change the correlation length, rerun the update and examine the target prediction. Before training, draw plausible prior fields or use representative quantiles and pass them through the forward operator. If almost every predicted data set violates basic instrument ranges or physical constraints, revisit the prior or parameterisation. This prior-predictive check uses no fitted residual and can uncover a consequential modelling error before a numerical optimiser converges.

Condition a second update on what the first already revealed

Suppose two observation errors share a known source, so processing each reading as an independent likelihood risks counting some information twice. For a scalar parameter and common response

coefficient one, write $y_1 = x + e_1,\ y_2 = x + e_2$. Let the errors be jointly Gaussian with variances v_1, v_2, covariance c, and independence from the prior parameter. Subtract $\alpha = c/v_1$ times the first reading from the second. The transformed second error $e_2 - \alpha e_1$ has zero covariance with e_1. Gaussianity strengthens that fact to independence. Its variance is $v_2 - c^2/v_1$ and its response coefficient is $1 - \alpha$.

The data can now be processed sequentially with the same result as a full joint update. If the new transformed coefficient is small, its effect is small even if the reading itself appears precise. The assumption that the original covariance is positive definite guarantees a positive residual variance. Perfectly correlated errors require a separate treatment of any exact difference equation. With non-Gaussian errors, decorrelation alone does not establish independent likelihoods, even when a linear unbiased estimator remains available.

Compare an independent replicate with a correlated one

Start with prior $x \sim N(0, 4)$. Two readings, 2 and 3, have unit marginal error variances and covariance $1/2$. After the first observation, the precision rises from $1/4$ to $5/4$ and the precision-weighted mean rises from zero to 2. Thus the intermediate posterior has variance $4/5$ and mean $8/5$. The adjusted second datum is $3 - (1/2)2 = 2$, with observation coefficient $1/2$ and noise variance $1 - 1/4 = 3/4$. Its contribution to precision is $(1/2)^2/(3/4) = 1/3$, much smaller than the unit contribution of an independent replicate.

The final precision is

$$5/4 + 1/3 = 19/12.$$

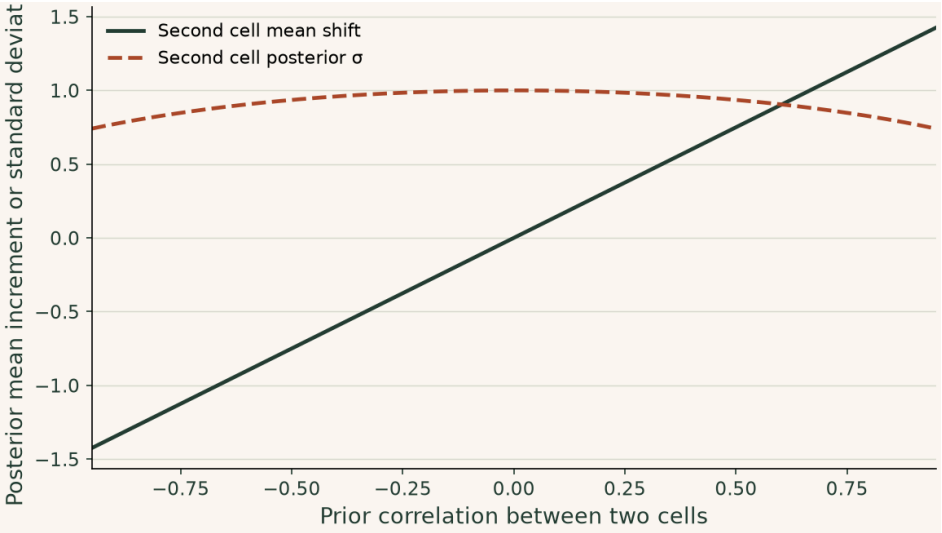
The adjusted datum adds $(1/2)2/(3/4) = 4/3$ to the weighted mean, giving posterior variance $12/19$ and mean $(2 + 4/3)(12/19) = 40/19$. The mean is about 2.11 and the standard deviation about 0.795. Incorrectly treating both original readings as independent would give variance $4/9$, mean $20/9$, and standard deviation $2/3$. The same pair of readings has therefore yielded a more confident answer solely by suppressing part of the noise model.

An independent check uses the average and difference of the two observations. Their respective signal coefficients are one and zero; equal marginal variances make these two errors uncorrelated and hence independent. The average is 2.5 with variance $(1 + 1 + 2 \cdot 0.5)/4 = 0.75$ and therefore adds precision $4/3$ to the prior. The difference remains available for checking the joint error model, even though it supplies no information about x. Both routes give the same posterior when each part of the evidence is used exactly once.

For the numbers used here the difference $3 - 2 = 1$ has variance $1 + 1 - 2(1/2) = 1$. It is one standard deviation from the expected zero difference. Thus there is no substantial conflict between these two readings under their correlation model, even though treating them as independent would exaggerate the information available about their shared underlying value.

Update an unobserved parameter through prior correlation

Two cell values have prior means zero, unit variances and correlation ρ. Only the first cell is observed, yielding 3 in measurement noise of variance one. The data-space update transfers information to the other cell through cross-covariance.



At $\rho = 0$ the second cell is unaffected. At $\rho = \pm 0.8$ its mean moves by ± 1.2 while both posterior standard deviations of that second cell equal $\sqrt{0.68}$. The magnitude of the uncertainty reduction hides the sign of the assumed physical relation.

▶ **Read values from the plot**

Change parameters · Update an unobserved parameter through prior correlation

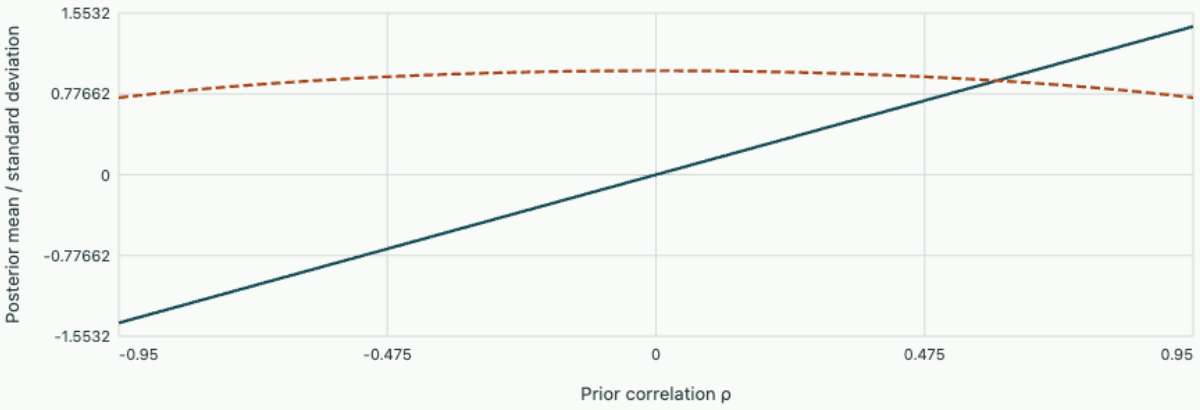
Change one input at a time and compare the recalculated curves with the argument above. The browser evaluates this small model locally; the accompanying numerical values state its assumptions and reference checks. Reset restores the reference inputs, and the data can be downloaded as CSV.

Open this calculation with adjustable inputs

Try the calculation

Measured first coordinate 3

Measurement noise variance 1



The observed-coordinate mean is 1.5 and its SD is 0.70711. The second coordinate is updated only through the specified prior covariance.

Study the full derivation

What each new sensor reading changes in a Gaussian inverse problem · Inferring parameters and fields from observations

Start with an uncertain vector, condition on one noisy linear observation, and reconstruct the posterior mean and covariance without repeating a batch inversion. Check shared calibration errors before treating repeated measurements as independent evidence.

Chapter 8 · Predictions, parameter profiles and nonlinear uncertainty

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 $\mathrm{Var}\, h = 2s^4$, using the Gaussian fourth moment. The linear approximation at zero would miss both.

A positive rate and a nonlinear lifetime

Suppose $\theta = \ln(k/k_{\mathrm{unit}})$ is normal with mean m and standard deviation s. A decay lifetime is

$$t_* = 1/k = e^{-\theta}/k_{\mathrm{unit}}.$$

Its median is e^{-m}/k_{unit}; its mean is larger by $e^{s^2/2}$. A central 95% interval is $[e^{-m-1.96s}, e^{-m+1.96s}]/k_{\mathrm{unit}}$. 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

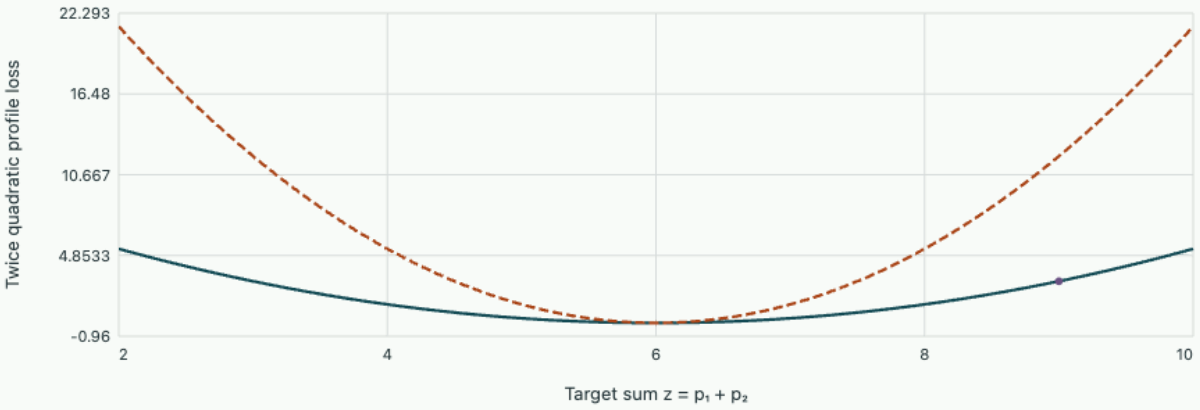
Change one input at a time and compare the recalculated curves with the argument above. The browser evaluates this small model locally; the accompanying numerical values state its assumptions and reference checks. Reset restores the reference inputs, and the data can be downloaded as CSV.

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.

Problem sheet 4 · Which uncertainty came from the experiment?

Compare a posterior update, a nonlinear prediction and a likelihood profile on problems with exact solutions.

1. Learning about an unmeasured cell

Two temperatures have prior mean $(20, 20)$ and covariance $\left(\begin{smallmatrix} 4 & 2 \\ 2 & 4 \end{smallmatrix}\right)$ K². The first cell is measured as 23 °C with independent noise variance 1 K². (a) Compute the updated mean and covariance. (b) Repeat the update of the second cell using zero prior correlation and the same marginal variances. (c) Interpret the difference.

▼ **Follow the cross-covariance into the unmeasured cell**

The prior observation variance including noise is 5. Its cross-covariance with the state is $(4, 2)^{\mathsf{T}}$, giving updated mean $(22.4, 21.2)$. Subtract $(4, 2)^{\mathsf{T}}(4, 2)/5$ from the prior covariance to obtain $\left(\begin{smallmatrix} 0.8 & 0.4 \\ 0.4 & 3.2 \end{smallmatrix}\right)$ K². With zero correlation the second mean and variance stay at 20 and 4. The indirect update relies entirely on the prior correlation assumption and the new first-cell measurement together.

Explore the related calculation with adjustable inputs. Change the parameters in the corresponding calculation and compare its independently recomputed values with the equation derived in this solution.

---

2. A nonlinear prediction at a stationary point

Let $x \sim N(0, 0.25)$, and take a detector intensity proportional to $I = x^2$. (a) Calculate the mean and variance of I. (b) Compare with the first-order delta method. (c) Find the upper endpoint of a one-sided 95% predictive interval for the ideal intensity using $P(|Z| \le 1.96) \simeq 0.95$.

▼ **Transform the Gaussian amplitude before reporting intensity**

The mean is 0.25 and variance is

$$2(0.25)^2 = 0.125.$$

At the mean amplitude the derivative $2x$ vanishes, so the first-order method predicts zero variance. The ideal intensity is nonnegative and $P(I \le (1.96 \times 0.5)^2) \simeq 0.95$, giving upper endpoint 0.9604. A new noisy intensity reading would also require an observation model.

Explore the related calculation with adjustable inputs. Change the parameters in the corresponding calculation and compare its independently recomputed values with the equation derived in this solution.

---

3. Slicing a valley or profiling it

Given $F = (p + q - 3)^2 + 0.25(q - 1)^2$, (a) locate its joint minimum. (b) Derive the profile in p and the slice with q fixed at its optimum. (c) Compare the widths of the two sets defined by $F \le 1$.

▼ **Allow the nuisance parameter to compensate**

The joint minimum has $q = 1, p = 2$. For fixed p, differentiation gives

$$q = (3.25 - p)/1.25;$$

substitution yields

$$F_{\text{prof}} = 0.20(p - 2)^2.$$

The slice is $(p - 2)^2$. The half-widths are $\sqrt{5}$ and 1 respectively. Their ratio is independent of the chosen positive contour level for this quadratic surface. Attaching a confidence level still requires a likelihood scale and sampling assumptions.

Explore the related calculation with adjustable inputs. Change the parameters in the corresponding calculation and compare its independently recomputed values with the equation derived in this solution.

Chapter 9 · Adjoint gradients through a forward solver

Differentiate a scalar objective with respect to many parameters without recomputing a separate perturbed solution for each one.

FROM FIRST PRINCIPLES

Differentiate a scalar objective through a linear state equation

Suppose the state u solves $K(p)u = f(p)$ for a given parameter vector, with nonsingular K, and a scalar objective $J(u, p)$ measures mismatch. A small parameter perturbation changes the state through both the load and the operator. Differentiating the balance gives

$$K\,du = df - (dK)u,$$
$$dJ = J_u^T du + J_p^T dp.$$

To obtain a scalar derivative without computing every state sensitivity, solve the transpose system $K^T\lambda = J_u$ once. Substitution eliminates du,

$$dJ = J_p^T dp + \lambda^T[df - (dK)u].$$

Each parameter then contributes through its derivatives of f, K, J. The transpose is required even if the forward solver is more familiar; it moves the scalar objective's weights backward through the linear response. A directional finite difference of the complete objective is an independent check on a computed adjoint gradient.

Background · Adjoint sensitivity

$$J^{\mathsf{T}} r$$

This operation maps a data-space vector back to parameter space. For a discretised physical model the chosen inner products determine the correct adjoint.

DERIVATION AND APPLICATIONS

Propagate a terminal measurement backward through continuous time

Consider a scalar state satisfying $\dot{u} = a(t)u + p\,b(t)$ over $0 \le t \le T$, with known initial state independent of p. A sensor observes $u(T)$, giving loss $J = \frac{1}{2}[u(T) - y]^2$. Assume continuous coefficients and exact evolution for this calculation. The forward derivative $w = \partial u/\partial p$ satisfies an inhomogeneous initial-value problem, and the objective derivative depends on its endpoint,

$$\dot{w} = a(t)w + b(t),$$
$$w(0) = 0,$$
$$\frac{dJ}{dp} = [u(T) - y]w(T).$$

Instead of solving for every possible forcing sensitivity separately, introduce a terminal weight $\lambda(T) = u(T) - y$. Choose its evolution to cancel the term proportional to w in the derivative of λw. The product rule gives

$$\dot{\lambda} = -a(t)\lambda,$$
$$\frac{d}{dt}(\lambda w) = \lambda b(t).$$

Integrate over the interval. The initial product vanishes because $w(0) = 0$, and the terminal product equals the desired gradient. Thus

$$\frac{dJ}{dp} = \int_0^T \lambda(t)b(t)\,dt.$$

One backward weight suffices for many scalar parameters entering through different forcing shapes. A parameter-dependent initial condition would add $\lambda(0)\partial_p u(0)$. A discrete time integrator approximating the continuous equations also introduces integration error; differentiating the actual discrete steps gives the derivative of the computed loss at that finite step size.

WORKED EXAMPLE

Reconstruct a sensitivity in a decaying reservoir

For $a = -1,\ b = 1,\ T = \ln 2,\ u(0) = 0,\ p = 4$, the terminal state is

$$4(1 - e^{-T}) = 2.$$

With observation $y = 1.5$, the terminal adjoint is 0.5. Earlier weights are $\lambda(t) = 0.5e^{-(T-t)}$, so

$$dJ/dp = 0.5(1 - e^{-T}) = 0.25.$$

Direct differentiation of $\frac{1}{2}(p/2 - 1.5)^2$ at four confirms the result.

SECOND APPLICATION

Compare two controls observed only at the endpoint

For $\dot{u} = -u$ between two instantaneous unit-gain input pulses, consider their amplitudes as independent parameters. A pulse at time t_i contributes $e^{-(T-t_i)}$ to terminal-state sensitivity. At $T = 3$ with terminal

residual 2, the gradient for a pulse at $t = 1$ is $2e^{-2} = 0.271$; for a pulse at $t = 2$ it is

> $$2e^{-1} = 0.736.$$

Later forcing has greater leverage because it has less time to decay.

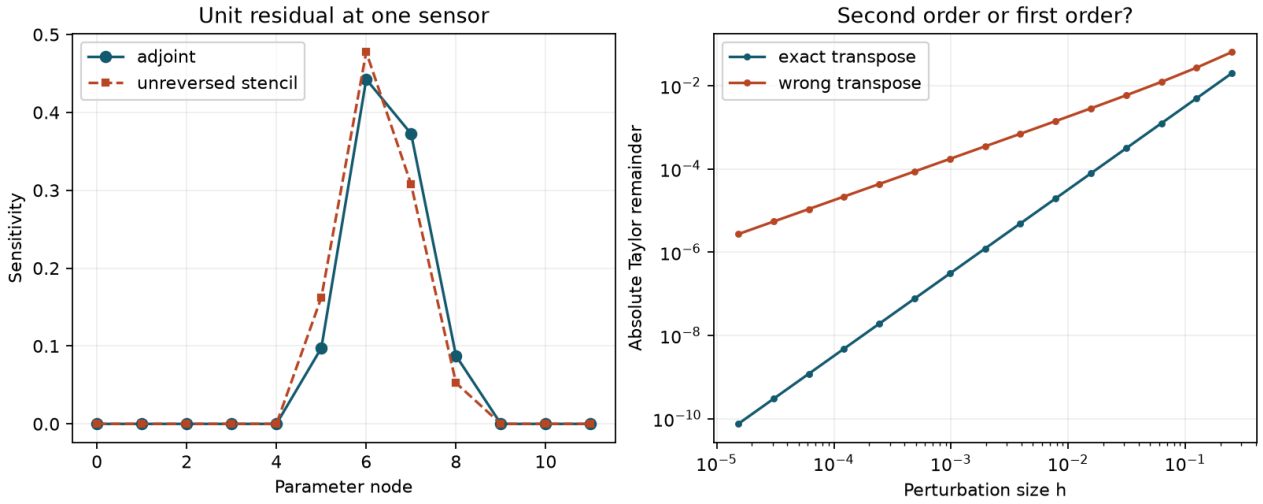
Does the gradient see where the instruments were actually placed?

A periodic ring of parameter nodes is mapped to fewer instrument readings in two operations. First a narrow, slightly asymmetric sensor spreads each nodal signal to its neighbours. Write this as a matrix S. Next the instruments sample between nodes. For an instrument a fraction f of the way from node j to $j + 1$, the corresponding row of B has entries $1 - f$ and f at those nodes. The complete forward model is $A = BS$. It is rarely necessary to assemble this product on a large mesh, but a small explicit example makes it possible to audit every entry.

Take Euclidean inner products so the conventions are clear. An adjoint must satisfy $(Av)^T w = v^T A^T w$ for every parameter direction v and sensor residual w. Moving each matrix across the scalar product yields

$$A^T w = S^T (B^T w).$$

The interpolation transpose scatters a sensor residual back to both neighbouring nodes with exactly the original weights. The smoothing transpose then exchanges the left and right coefficients of the stencil. Applying the forward smoothing again can look plausible and still shift the inferred sensitivity footprint.



Reusing a slightly asymmetric forward stencil produces a plausible-looking sensitivity. Its first-order remainder betrays the error when the objective is perturbed in an independent parameter direction.

To test the gradient choose the merit function $J(m) = \frac{1}{2}\|Am - y\|^2$. Its directional derivative is $v^T A^T (Am - y)$. Expand the quadratic objective exactly to find

$$J(m + hv) - J(m) - hv^T A^T (Am - y) = h^2 \|Av\|^2 / 2.$$

Halving h divides the remainder by four until rounding intervenes. Substituting a wrong derivative leaves an extra term linear in h. The notebook checks both the observed scaling and the coefficient of the quadratic term at one finite step. Several independent directions help avoid a chance cancellation in a single test.

With a fraction $f = 0.35$, one unit of observation residual is split into 0.65 and 0.35 before reverse smoothing. Its total remains one because the row weights sum to one, but conservation of that sum is a weak test: using the wrong asymmetric stencil also preserves it. If the underlying inversion uses volume-weighted mesh or measurement inner products, those weights must be included consistently; the plain transpose in this experiment applies to its explicitly chosen Euclidean metric.

The pictured footprint is obtained by sending a single residual backwards. It can also be reconstructed element by element by perturbing one parameter node forwards and recording just that sensor. The two pictures should agree for every node. This costs more forward evaluations, but on a small test ring it checks the location of a discrepancy directly.

Suppose the continuous forward code records a sensor at fraction 0.35 while its adjoint implementation rounds that sensor to the nearest node. Can an exact solution of the adjoint linear system repair the discrepancy?

▼ **Work through the solution**

The error occurs in the observation operator itself. For a unit residual the true transpose must assign 0.65 and 0.35, whereas the rounded version assigns one and zero. Accurate solves propagate these different inputs accurately. An isolated dot-product test of the observation operator followed by a test of the full composition locates the discrepancy.

Change parameters · Explore this notebook example

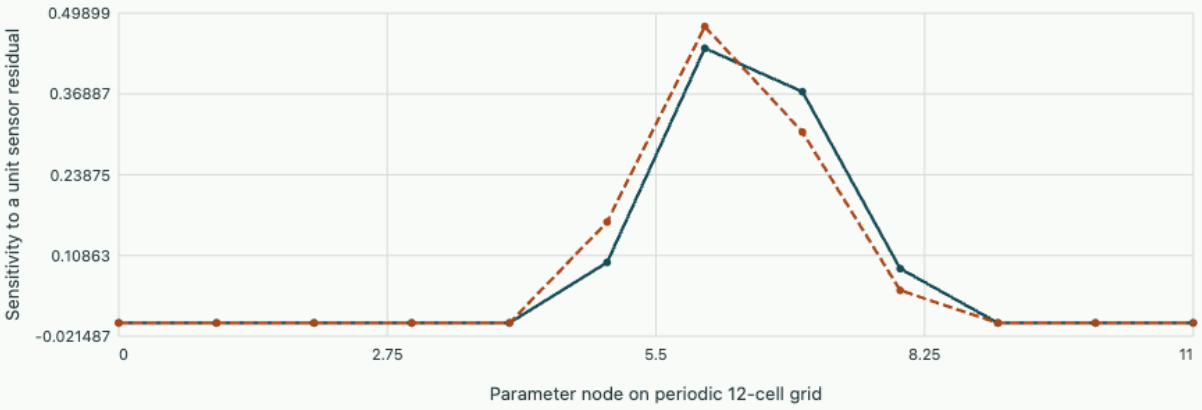
Change one input at a time and compare the recalculated curves with the argument above. The browser evaluates this small model locally; the accompanying numerical values state its assumptions and reference checks. Reset restores the reference inputs, and the data can be downloaded as CSV.

Open this calculation with adjustable inputs

Try the calculation

Sensor interpolation fraction **0.35**

Left smoothing weight **0.25**



Exact transposed smoothing and sensor map | Unreversed smoothing stencil

The sensitivity sums to 1. Asymmetric smoothing must reverse its direction in the adjoint; the mismatch vanishes only at symmetric weights.

Open editable notebook Download .ipynb

The state is also a function of the parameters

Suppose a discretised forward model solves $K(p)u = f(p)$. Data enter through Hu, and observations have been whitened. The objective is $J = \frac{1}{2}\|Hu - y\|^2 + R(p)$. A direct perturbation of a material parameter changes the predicted state even when that parameter never appears in the observation operator. For fixed H, the state derivative contribution is $g_u^{\mathsf{T}} du$, with $g_u = H^{\mathsf{T}}(Hu - y)$. Differentiating the state equation gives

$$K\,du = df - (dK)u.$$

Solving this equation once per parameter would be prohibitively expensive for a fine material field.

Transpose the solve

Define an adjoint state by $K^{\mathsf{T}}\lambda = g_u$. Then

$$g_u^{\mathsf{T}} du = \lambda^{\mathsf{T}} K\, du = \lambda^{\mathsf{T}}[df - (dK)u].$$

For each parameter component the gradient is

$$\frac{\partial J}{\partial p_j} = \frac{\partial R}{\partial p_j} + \lambda^{\mathsf{T}}\left(\frac{\partial f}{\partial p_j} - \frac{\partial K}{\partial p_j}u\right).$$

This uses one forward and one transposed solve for one scalar objective, independent of the number of gradient components. Forming the contractions for all components still costs work, often local to mesh elements. If sensors or boundary conditions also depend explicitly on parameters, their derivative terms must be included. A symmetric diffusion operator may reuse a factorisation; an advection operator need not equal its transpose.

Work a two-cell example

Consider $K = \begin{pmatrix} p+1 & -1 \\ -1 & 2 \end{pmatrix}$, $f = (1, 0)$, and observe the second component with value $y = 0.4$. At $p = 1$, $u = (2/3, 1/3)$, so the residual is $-1/15$. The adjoint load is $(0, -1/15)$; solving gives $\lambda = (-1/45, -2/45)$. Only the first diagonal entry of K depends on p. With no regulariser the derivative is $-\lambda_1 u_1 = 2/135$. Check independently: $u_2 = 1/(2p+1)$, whose derivative at one is $-2/9$. Multiplying by the residual gives the same positive gradient. Increasing p lowers an already underpredicted second reading.

Test the computed gradient in a direction

Pick a physically scaled perturbation v. A correct gradient gives a linear residual $J(p + hv) - J(p) - h\, g^{\mathsf{T}} v$ of order h^2 for a smooth problem. Halving h should roughly quarter this remainder across a useful range. At very small h, forward-solver tolerances and floating-point cancellation can dominate. Repeat with directions touching different components and with the same boundary-condition handling used in production. A close fit to the observed data is a poor gradient test when residuals nearly vanish: a wrong data gradient can then look small by accident. Evaluate at a parameter away from the best fit, solve the state again for each perturbation and compare the resulting objective values. To detect a missing transpose, include a nonsymmetric forward operator in the test. The symmetric two-cell example above can expose a sign error but cannot distinguish the forward matrix from its transpose.

Reverse a sequence of state updates

Suppose the forward solver evolves $u_{n+1} = G_n(u_n, p)$ for a fixed number of steps, and an objective sums observation losses $\ell_n(u_n, p)$ from time zero to the final time. The direct chain rule carries a state

derivative with respect to every parameter through the entire trajectory. To reverse the calculation, start with $\lambda_N = \partial_{u_N} \ell_N$ as a column vector. A perturbation du_n affects both the loss at time n and the following state by $G_{n,u} du_n$. Moving the latter action across a dot product gives the recursion $\lambda_n = \partial_{u_n} \ell_n + G_{n,u}^T \lambda_{n+1}$.

The parameter gradient is the sum of explicit derivatives of all losses, the initial term $(\partial_p u_0)^T \lambda_0$, and the contributions $G_{n,p}^T \lambda_{n+1}$ from every step. Each reversed step needs the state information used by its forward counterpart. Storing the whole trajectory makes that information available immediately. Recomputing segments between stored checkpoints trades extra forward evaluations for memory. These formulas differentiate the actual discrete transitions, so an adaptive solver with parameter-dependent step rejection or event switches needs a careful interpretation of the differentiated computation.

Make the transpose observable in a two-step calculation

Take a two-coordinate forward transition $G(u, p) = Mu + p e_2$, with $M = \left(\begin{smallmatrix} 1 & 2 \\ 0 & 1 \end{smallmatrix}\right)$, $u_0 = 0$, and $p = 1$. Two steps give

$$u_1 = (0, 1)^T,$$

$$u_2 = (2, 2)^T.$$

Observe the first coordinate only at the end, with target value one. The quadratic loss is one half and its final state derivative is

$$\lambda_2 = (1, 0)^T.$$

One reverse step gives

$$\lambda_1 = M^T \lambda_2 = (1, 2)^T.$$

The two parameter contributions are $e_2^T \lambda_2 = 0$ and $e_2^T \lambda_1 = 2$, so the gradient is two.

The zero contribution from the last injection has a physical meaning: the injected second coordinate can influence the first only during a subsequent transition. The first injection has time to do that through the off-diagonal element two. Using M in place of M^T in the backwards calculation would instead keep $\lambda_1 = (1, 0)^T$, predicting zero total sensitivity. This example exercises a distinction that disappears when every state matrix is symmetric.

For a scalar parameter perturbation h, solving both steps again gives terminal first coordinate $2(1 + h)$ and loss $J(1 + h) = (1 + 2h)^2 / 2$. The remainder after subtracting the base loss and the predicted linear term is exactly $2h^2$. For $h = 0.1$ and 0.05 the remainders are 0.02 and 0.005. The erroneous zero gradient instead gives 0.22 and 0.105. The two sequences have very different observed convergence factors even though they use identical forward trajectories and observation data.

More generally an individual parameter may enter the map at several times with different signs. The corresponding gradient terms must be accumulated rather than overwritten during the backward pass. In this example overwriting after the earliest step happens to return the correct value because the last

contribution is zero. Adding a simultaneous observation of the final second coordinate makes both contributions nonzero and provides a complementary regression case for the time-index accumulation.

For a terminal loss summing both coordinate discrepancies, the adjoint load contains both sensor contributions before the first reverse transition.

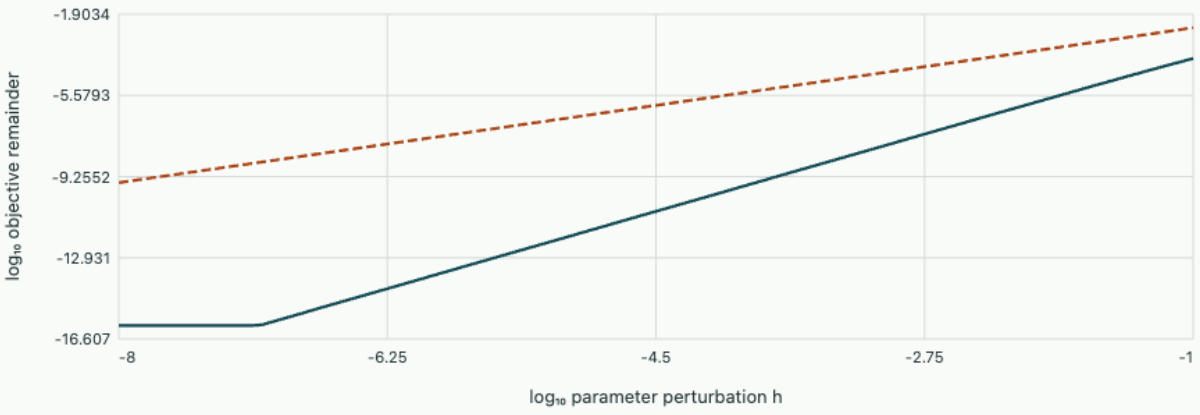
Change parameters · Adjoint gradients through a forward solver

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

Material parameter p	1
Observed second state	0.4

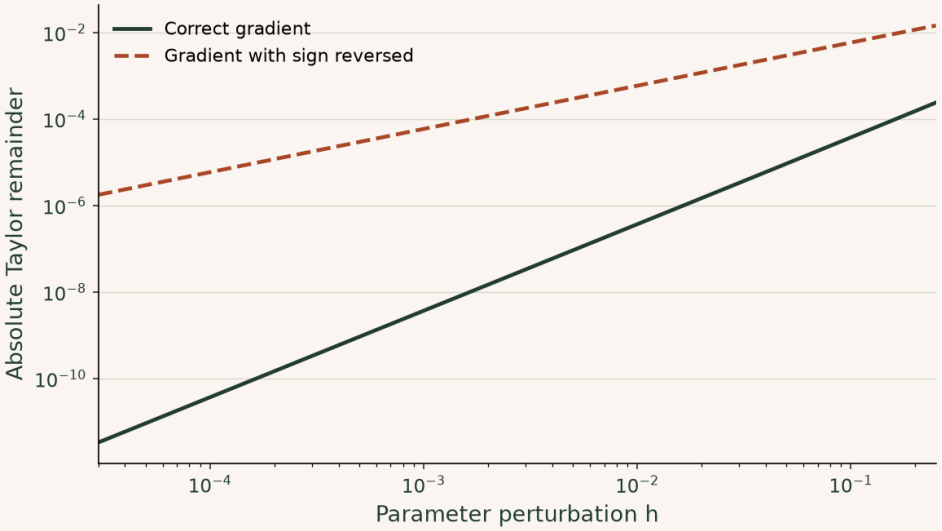


Subtract correct adjoint gradient Subtract wrong-sign gradient

Analytic and adjoint derivative = 0.014815. A correct gradient leaves a quadratic remainder until cancellation dominates; at an exact fit this sign test loses sensitivity.

Recognise an incorrect gradient by the rate of its remainder

Verify the reduced gradient for $J(p) = (1/(p+2) - 0.6)^2/2$ at $p = 1$. For each perturbation solve the scalar equilibrium again and subtract the original objective and the candidate linear term. The sign of the state sensitivity is essential.



In the range shown the Taylor remainder falls quadratically with the correct gradient and approximately linearly with the reversed one. The use of an analytic scalar model isolates this diagnostic; tolerances in a larger forward solve can mask the small-h regime.

▶ **Read values from the plot**

Chapter 10 · Iteration as a regulariser

Derive the spectral action of a matrix-free iteration and explain why more iterations can eventually make an inverse estimate worse.

FROM FIRST PRINCIPLES

Derive a matrix-free correction and its singular-mode filter

For measured data b, minimize $\frac{1}{2}\|Ax-b\|_2^2$. Its negative gradient is $A^T(b-Ax)$, so a step of fixed length ω requires just one forward and one transpose application. Starting from zero gives the Landweber iteration,

$$x_{k+1} = x_k + \omega A^T(b - Ax_k),$$
$$x_0 = 0.$$

For a singular pair $Av_j = \sigma_j u_j$ with $\sigma_j > 0$, project onto v_j. Each coefficient then follows a scalar geometric recurrence. Summing it gives

$$v_j^T x_k = \left[1 - (1-\omega\sigma_j^2)^k\right] \frac{u_j^T b}{\sigma_j}.$$

Choosing $0 < \omega \le 1/\|A\|_2^2$ makes every bracket grow monotonically from zero toward one. Strongly sensed components appear early. Weakly sensed components require many steps, and their final coefficient can greatly amplify noise. This separation of timescales explains why a finite stopping iteration can regularize a noisy reconstruction.

DERIVATION AND APPLICATIONS

Bound the noise admitted by a finite iteration count

Suppose the data are corrupted by an unknown vector e with $\|e\|_2 \le \delta$. Compare two Landweber runs starting at zero, one with exact data and one with noisy data, with fixed $0 < \omega \le 1/\|A\|_2^2$. Orthogonality of the singular vectors makes the worst possible data amplification at iteration $k \ge 1$ the largest singular-mode gain

$$g_k(\sigma) = \frac{1-(1-\omega\sigma^2)^k}{\sigma}, \qquad \sigma > 0.$$

Put $t = \omega\sigma^2 \in [0,1]$. The numerator lies below one and also below kt, since the sum $1 + (1-t) + \cdots + (1-t)^{k-1}$ contains k terms each at most one. These two estimates control opposite ends of the spectrum,

$$0 \le g_k(\sigma) \le \min\{1/\sigma,\ k\omega\sigma\}.$$

The largest value of this upper envelope occurs where its increasing and decreasing branches meet. At $\sigma = 1/\sqrt{k\omega}$, both equal $\sqrt{k\omega}$. The noise passed to the reconstruction is therefore bounded by

$$\|x_k(b+e) - x_k(b)\|_2 \le \delta\sqrt{k\omega}.$$

If that crossover is outside the actual singular spectrum, the bound can be conservative. This estimate describes noise only: the bias remaining after an early stop depends on the true solution's distribution among singular directions. A useful stopping rule considers both effects and the quantity being reported.

WORKED EXAMPLE

Budget the iterations allowed by a noise target

In scaled coordinates let $\delta = 0.010,\ \omega = 0.25$, and require the noise-induced solution change to be at most 0.050. The envelope guarantees this up to $k = (0.050/0.010)^2/0.25 = 100$. At 25 steps the guarantee is 0.025; quadrupling the count doubles the guaranteed noise allowance.

SECOND APPLICATION

Compare the envelope to an actual weak mode

For another run take $\omega = 1,\ k = 10,\ \sigma = 0.10$. The exact gain is $[1 - 0.99^{10}]/0.10 = 0.956$, smaller than the spectrum-independent bound $\sqrt{10} = 3.16$. A perturbation of 0.020 entirely along this data mode produces error 0.0191 after ten steps. A fully inverted fit would amplify the same perturbation to 0.200.

Multiply by the operator without storing its matrix

For large imaging systems a forward projection Bx and its adjoint $B^{\mathsf{T}}r$ may be available as procedures while a matrix with one column per pixel is too large to store. Landweber iteration starts from zero and uses the least-squares gradient:

$$x_{k+1} = x_k + \omega B^{\mathsf{T}}(b - Bx_k).$$

An implementation requires both procedures to be adjoints in the same inner products. Check $\langle Bv, w\rangle = \langle v, B^{\mathsf{T}}w\rangle$ on a variety of inputs; quadrature weights in a discretised integral can change the correct transpose. For the Euclidean norm choose $0 < \omega < 2/\sigma_{\max}^2$ to ensure contraction in every nonzero singular direction. Choosing $\omega \le 1/\sigma_{\max}^2$ additionally avoids alternating scalar transients in these directions.

Derive the filter after a finite number of steps

Let $z_{j,k} = v_j^{\mathsf{T}}x_k$, $\beta_j = u_j^{\mathsf{T}}b$. The recurrence becomes

$$z_{j,k+1} = (1 - \omega\sigma_j^2)z_{j,k} + \omega\sigma_j\beta_j.$$

Summing its geometric series at $x_0 = 0$ gives

$$z_{j,k} = \left[1 - (1 - \omega\sigma_j^2)^k\right] \frac{\beta_j}{\sigma_j}.$$

Components with large singular values reach their least-squares values quickly. Small-singular-value components remain suppressed for many iterations because $1 - (1 - \omega\sigma_j^2)^k \simeq k\omega\sigma_j^2$ while $k\omega\sigma_j^2$ is small. Stopping early is therefore a choice of spectral filter. Changing the initial guess affects the pattern retained in weak or null directions; with initial zero the iteration never invents a null-space component.

Regularise with an artificial evolution time

There are two clocks in a time-dependent inverse experiment. One describes the physical process being measured. A second, entirely artificial clock can limit how far we move towards a data fit. For a fixed real matrix A and fixed observed data b, evolve the negative least-squares gradient from zero,

$$\dot{x}(t) = A^T(b - Ax(t)),$$
$$x(0) = 0.$$

The data misfit $F = \|Ax - b\|_2^2/2$ satisfies

$$dF/dt = -\|A^T(Ax - b)\|_2^2 \leq 0.$$

Monotone improvement of the fit alone does not establish improvement of the reconstruction. In singular coordinates $b_j = u_j^T b$, each positive singular value gives the scalar equation $\dot{x}_j + \sigma_j^2 x_j = \sigma_j b_j$. Multiplying by $e^{\sigma_j^2 t}$ and integrating from zero gives

$$x_j(t) = \frac{1 - e^{-\sigma_j^2 t}}{\sigma_j}\, b_j.$$

A null component remains zero. A mode reaches half of its unregularised amplitude at $t = (\log 2)/\sigma_j^2$. At a fixed finite t, weak modes have small noise gain approximately $t\sigma_j$. At very long times the gain tends to $1/\sigma_j$. This is asymptotic regularisation, with stopping time acting as a spectral filter. As the noise bound decreases one can allow longer runs, subject to continued noise control. A single fixed noisy measurement offers no justification for sending t to infinity.

Two modes fitted on different clocks

Let $A = \mathrm{diag}(1, 1/4)$ and $b = (1, 0.04)^T$. At $t = 16 \log 2$, the weak exponential is exactly $1/2$ while the strong exponential is 2^{-16}. The reconstruction is $(1 - 2^{-16},\ 0.08)^T$. The fully inverted weak datum would give 0.16. The strong measurement is fitted to within $2^{-16} \approx 0.0000153$ in amplitude while half the weak amplitude is still suppressed. If some of the 0.04 is noise, the same factor suppresses that contribution.

Forward Euler with step h in this artificial time recovers the Landweber iteration. It replaces each remaining exponential by $(1 - h\sigma_j^2)^k$. To make every positive mode decay we need $0 < h < 2/\sigma_{\max}^2$. An

explicit step beyond this limit can diverge even though the original evolution dissipates misfit. Comparing stopping iterations therefore requires reporting h as well as k.

Practise on sheet 5 · When explicit reconstruction has its own stability limit. Further reading · University of Cambridge · Inverse Problems, Mathematical Tripos Part III · §3.2.5 Asymptotic regularisation and §3.2.6 Landweber iteration, pp. 39–40.

See semiconvergence in a two-mode experiment

Let $B = \mathrm{diag}(1, 0.1)$, true $x = (1, 0)$ and observed $b = (1, 0.02)$. Set $\omega = 1$. After one step the estimate is $(1, 0.002)$, very close to truth, and the first mode is already exact. After 100 steps the second coordinate is

$$0.2(1 - 0.99^{100}) = 0.1268.$$

At convergence it is 0.2. The residual shrinks steadily toward zero while the actual reconstruction has deteriorated after the first step. With a genuinely nonzero second component the error curve may initially improve for longer. Which iteration is best depends on the unknown signal and noise, so a solver's normal residual criterion is not a suitable scientific stopping rule by itself.

Select an observable stopping test

When a data error bound is known, stop the first time the residual falls below a noise-compatible level. With independent validation measurements monitor their predictive error across iterations. A fast Krylov method can bring useful modes in much sooner but may also reach noise-dominated directions sooner. Compare algorithms by scientific accuracy and forward/adjoint applications at a defensible stopping point. If a prior-weighted variable or preconditioner changes how quickly individual modes enter, the early-stopped result changes as well. Record that choice alongside the step size, starting state and stopping rule so the reconstruction can be interpreted.

Read the polynomials selected by a Krylov inverse

Landweber iteration applies a predetermined residual polynomial once the step size and count are chosen. Conjugate gradients on the normal equations instead choose each update from a growing search space using the actual data. With $x_0 = 0$, write $g = B^T b$ and $C = B^T B$. The first Krylov approximation is $x_1 = \alpha g$. Minimising $\|b - \alpha B g\|^2$ gives $\alpha = (Bg)^T b / \|Bg\|^2 = \|g\|^2 / \|Bg\|^2$ when $g \neq 0$. The first step is thus a data-dependent scaling of the backprojected measurements. It uses only products by the forward operator and its transpose.

At iteration k, the corrections lie in the span of $g, Cg, \ldots, C^{k-1}g$. In singular coordinates the solution coefficient is a polynomial in σ_j^2 times $\sigma_j b_j$. In exact arithmetic, once the span contains every excited singular direction the least-squares fit is exact in the measured range. Faster convergence of the data

residual is valuable for well-determined components. It can also admit a noise-dominated direction much more abruptly than an investigator expects from watching just the residual.

See the second iteration introduce a spurious peak

Let $B = \mathrm{diag}(1, 0.05)$, true state $(2, 0)^T$, and observed $b = (2, 0.10)^T$. The second reading is entirely noise in this constructed test. Then $g = (2, 0.005)^T$ and $Bg = (2, 0.00025)^T$. The optimal scalar is $(4 + 0.000025)/(4 + 0.0000000625) \approx 1.00000623$. The first state is approximately $(2.0000125, 0.00500003)^T$, closely matching the known truth. Its predicted second reading is only about 0.000250, and the total residual is consequently just below 0.10. A discrepancy threshold of 0.11 would accept this result.

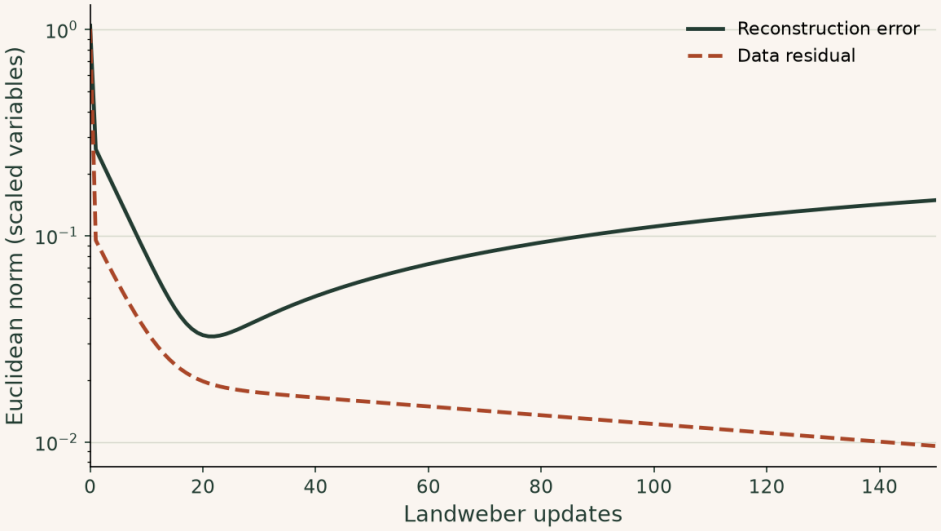
Both nonzero singular directions are excited by the data. The second Krylov space is the whole plane because g and $Cg = (2, 0.0000125)^T$ are independent. Its minimiser is therefore the exact inverse state $(2, 2)^T$, with zero data residual. The second true coordinate is zero, so this last step increases the state error from about 0.005 to two. The improvement in the objective is an improvement in fitting the particular noise draw. It has supplied no additional physical information about a weakly measured component.

For a real experiment the true state is unavailable, so a plot of state error cannot choose the iteration. Use measured uncertainty, predicted quantities on held-out readings or credible prior constraints to decide about stopping. If independent repeats show the same signed second reading with uncertainty far below 0.10, the case for treating it as pure noise would weaken. A stopping threshold calibrated for the present example should then change with the measurement evidence.

The data-based early step also changes under a rescaling of the parameter coordinates. A different scaling changes the shape of the first search space after backprojection, although the final unregularised solution expressed in the original coordinates remains the same. For that reason the iteration number by itself does not fully specify an iterative regulariser. Record the scaling, preconditioner and starting estimate as part of the reconstruction.

Observe semiconvergence before the residual is exhausted

Run unit-step Landweber iteration from zero on a diagonal operator with singular values 1, 0.35 and 0.07. The true modes are (1, 0.3, 0); add measurement errors (0, 0.001, 0.02). Track both norms using the known synthetic truth.



The middle mode is initially under-recovered. Later the weakest mode starts fitting a perturbation that contains no true signal. The data residual keeps falling while the reconstruction passes its best accuracy and deteriorates.

▶ **Read values from the plot**

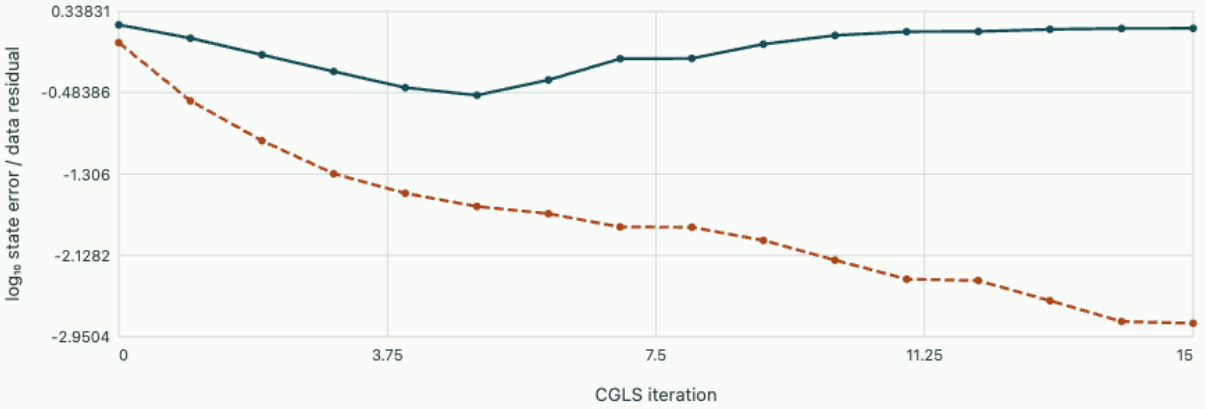
Change parameters · Observe semiconvergence before the residual is exhausted

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

Noise amplitude	0.01
Singular-value decay power	1.5
Displayed iterations	15



Smallest displayed state error occurs at iteration 5. The truth is available only because this is a manufactured experiment; residual decrease alone cannot identify that stopping point.

Study the full derivation

Recovering sparse coefficients without hiding amplitude bias · Inferring parameters and fields from observations

A sensor records a linear mixture of a few active components. Derive the soft-threshold map, use it in an iterative reconstruction and decide when correlated response columns make an apparently sharp sparse answer ambiguous.

Problem sheet 5 · Derivatives and early stopping

Check a forward solver gradient independently, then decide when its iterative inverse has started fitting noise.

1. Differentiate through a scalar equilibrium

A steady sensor reads u satisfying $pu = 2$. The normalised measured value is 0.8 and $J = (u - 0.8)^2/2$. (a) At $p = 2$, compute the adjoint, gradient and direct derivative. (b) What sign would a one-step gradient descent change have? (c) Does that sign make physical sense?

▼ **Check the sign against the forward response**

The state is 1, residual 0.2. The adjoint equation $p\lambda = 0.2$ gives 0.1. With no forcing derivative the gradient is $-\lambda u = -0.1$. Directly $du/dp = -2/p^2 = -0.5$, yielding the same gradient. A descent step increases p; increasing p lowers the model output toward the observation.

Explore the related calculation with adjustable inputs. Inspect the Taylor remainder with the correct and wrong-sign adjoint. The explorer uses a two-state equilibrium; apply the stated scalar or nonsymmetric operator to obtain this exercise's exact gradient.

---

2. A directional derivative check

At $p = 1$ for $J(p) = p^3$, choose direction $v = 2$. (a) Derive the Taylor remainder after subtracting the predicted linear term. (b) Evaluate it for $h = 0.10$ and 0.05. (c) Predict what halving h would do if the component gradient 3 were accidentally coded as 2.

▼ **Examine the order of the diagnostic**

The exact directional slope is $3v = 6$. Expanding $(1 + 2h)^3$ leaves $12h^2 + 8h^3$ after subtracting 1 and $6h$. The remainders are 0.128 and 0.031, close to a factor of four apart. With incorrect slope 4 the error becomes $2h + 12h^2 + 8h^3$. At small enough h it will fall by about a factor of two instead, until numerical noise intervenes.

Explore the related calculation with adjustable inputs. Change the parameters in the corresponding calculation and compare its independently recomputed values with the equation derived in this solution.

---

3. Stop before the weakest direction is inverted

Use $B = \mathrm{diag}(1, 0.2),\ b = (1, 0.1),\ x_0 = 0,\ \omega = 1$. (a) Derive the second coordinate and residual after $k \geq 1$ iterations. (b) Find the first iteration whose residual norm is at most 0.09. (c) Compare the result there with the converged second coordinate and explain when early stopping helps.

▼ **Solve the scalar recurrence**

The first coordinate reaches 1 in one step. The second tends to 0.5 as $x_{2,k} = 0.5(1 - 0.96^k)$. Thus $\|r_k\| = 0.1(0.96)^k$, and crossing 0.09 requires $k \geq \ln(0.9)/\ln(0.96) = 2.58$, or three updates. The early estimate is 0.0576 while the converged one is 0.5. If most of the observed weak-mode coefficient is noise the early estimate is protective; if it mostly represents signal the same stopping rule underestimates it.

Explore the related calculation with adjustable inputs. Vary noise and iteration count to distinguish a decreasing training residual from a worsening reconstruction. CGLS in the explorer and the exercise's Landweber recurrence have different polynomial filters.

When explicit reconstruction has its own stability limit

Let $A = \mathrm{diag}(2, 1/2)$. Explain why the continuous residual decays in both modes but a Landweber step $h = 0.6$ can fail. Repeat with $h = 1/4$. Find when the weak component has first acquired at least half its unregularised amplitude, and compare with the continuous half-time.

▼ **Worked solution**

The continuous residual factors are e^{-4t} and $e^{-t/4}$. At $h = 0.6$ the strong discrete factor is

$$1 - 0.6(4) = -1.4.$$

Any excited strong residual changes sign and grows. The step exceeds the upper bound $2/4 = 0.5$.

At $h = 1/4$ the strong factor is zero, and the weak factor is

$$1 - (1/4)(1/4) = 15/16.$$

Solve $(15/16)^k \leq 1/2$: the real threshold is 10.7401, hence the first integer is 11. Its artificial time is $11/4 = 2.75$. The continuous half-time is $4 \log 2 = 2.7726$. The small difference reflects the exponential approximation and inspection at discrete steps.

Explore the related calculation with adjustable inputs. Vary noise and iteration count to distinguish a decreasing training residual from a worsening reconstruction. CGLS in the explorer and the exercise's Landweber recurrence have different polynomial filters.

Revisit chapter 10 · Regularise with an artificial evolution time.

Notebook practice · Does the gradient see where the instruments were actually placed?

Chapter 11 · Constraints, sparse signals and edge-preserving penalties

Derive a solution at a bound, a soft-threshold rule and the behaviour of a total-variation penalty at a material interface.

FROM FIRST PRINCIPLES

Derive thresholding from an absolute-value penalty

A coefficient that can physically vanish can be estimated by balancing squared measurement mismatch against a penalty on its absolute magnitude. For a scalar reading y, consider $F(x) = \frac{1}{2}(x-y)^2 + \lambda|x|$ with $\lambda \geq 0$. On either open half-line the derivative is ordinary, giving candidates

$$x = y - \lambda \quad (x > 0),$$
$$x = y + \lambda \quad (x < 0).$$

When neither candidate has the assumed sign, the kink at zero must be checked. The left derivative at zero is $-y - \lambda$ and the right derivative is $-y + \lambda$. A convex minimum occurs there if these enclose zero, namely if $|y| \leq \lambda$. Combining the cases gives

$$x_* = \operatorname{sign}(y)\max(|y| - \lambda, 0).$$

The shrinkage creates exact zeros and biases surviving amplitudes toward zero. Applying a similar penalty to differences between adjacent field values instead of to the field itself favours neighbouring sites that share one value while permitting isolated jumps.

DERIVATION AND APPLICATIONS

Shrink a two-dimensional gradient while preserving its direction

At an image pixel or a mesh cell the inferred gradient can have two measured components. If the physical surface has no preferred coordinate axes, a rotationally invariant local penalty is proportional to the Euclidean length of the gradient vector. Consider the local optimization with equally scaled and equally precise components,

$$\min_{g \in \mathbb{R}^2} \frac{1}{2}\|g - y\|_2^2 + \lambda\|g\|_2, \qquad \lambda \geq 0.$$

For a fixed trial length $r = \|g\|$, the data mismatch is minimized by making g parallel to y, since this maximizes their dot product. For $y \neq 0$ write $g = ry/\|y\|$. The remaining one-dimensional constrained problem is

$$\min_{r\ge 0}\ \frac{1}{2}(r-\|y\|)^2+\lambda r, \qquad r_*=\max(\|y\|-\lambda,0).$$

Thus both gradient components are attenuated by a common factor; if $y=0$, the solution is zero directly. The componentwise absolute-value penalty gives a different rule,

$$g_*=\max(1-\lambda/\|y\|,0)y, \qquad (g_{\text{component}})_i=\operatorname{sign}(y_i)\max(|y_i|-\lambda,0).$$

The first expression is for $y\neq 0$. A local gradient shrinkage step is used inside some total-variation algorithms, but arbitrary independently shrunk gradient vectors need not be differences of one global image. A full reconstruction must enforce consistency with the field and the measurement equations. Different component units or uncertainties likewise require a corresponding metric.

WORKED EXAMPLE

Compare an inclined edge under two penalties

For measured gradient $y=(3,4)$ and $\lambda=2$, the isotropic solution has length $5-2=3$ and vector $(1.8,2.4)$; its angle is unchanged. Applying componentwise shrinkage gives $(1,2)$ with angle 63.4°, compared with the measured 53.1°. The choice of penalty can therefore affect an estimated orientation.

SECOND APPLICATION

Test whether a detection threshold depends on the grid axes

A second gradient has length 1.2 with $\lambda=1$. Along an axis, both rules return length 0.2. Rotating it by 45° gives coordinates $(0.849,0.849)$. The isotropic rule still returns length 0.2, whereas componentwise thresholding returns zero because neither coordinate reaches one. An orientation-dependent disappearance can therefore be introduced by the estimator itself.

Physical constraints change the optimality test

An attenuation coefficient cannot be negative. In a convex inversion with lower bound zero, the unconstrained least-squares normal equations need not hold at a valid constrained solution. Consider minimising a differentiable $F(x)$ with $x_j\ge 0$. At a free component the derivative must be zero. At a zero component only positive feasible perturbations exist, so the derivative must be nonnegative. Negative reconstructed cells cannot generally be repaired by clipping after inversion because the forward operator couples neighbouring parameters.

For example minimise $\frac{1}{2}[(x_1+x_2-1)^2+(x_1+1)^2]$ under $x_i\ge 0$. The unconstrained solution is $(-1,2)$. Clipping gives $(0,2)$, which overpredicts the first observation. On the active face $x_1=0$, the minimising x_2 is 1. At $(0,1)$ the gradient is $(1,0)$ and the bound conditions hold. The price of feasibility appears in the data residual; it is important to show which observations the constrained model cannot fit.

Derive the soft-threshold rule

For a scalar coefficient with observation b, minimise $\frac{1}{2}(x-b)^2+\lambda|x|$. On $x>0$ the stationary point is $b-\lambda$, feasible there only when $b>\lambda$; on $x<0$ it is $b+\lambda$. At zero, the admissible slopes of the absolute value range from -1 to 1, so zero solves the condition when $|b|\le\lambda$. Together these give

$$x=\operatorname{sign}(b)\max(|b|-\lambda,0).$$

Even retained coefficients shrink by λ, a real amplitude bias. With $b=(-0.2,0.7,2.4)$ and $\lambda=0.5$, the estimate is $(0,0.2,1.9)$ for an identity forward map.

Penalise changes in a field

In one dimension discrete total variation is $\sum_j|x_{j+1}-x_j|$. A single large jump and several monotone small steps spanning the same overall change have equal total variation. In contrast a squared difference penalty favours spreading the jump over several sites. That explains why total variation can preserve an abrupt edge when the data localise it. It can also create nearly constant plateaus in a field that really changes smoothly.

For two observed values b_1,b_2, use the variables $m=(x_1+x_2)/2,\ d=(x_2-x_1)/2$. With objective $\frac{1}{2}\sum(x_i-b_i)^2+\lambda|x_2-x_1|$, the optimal mean is the observed mean and half-difference is the soft-thresholded observed half-difference with threshold λ. When $|b_2-b_1|\le 2\lambda$ the two estimated levels fuse. This concrete loss of resolution should be reported when interpreting a uniform patch in a reconstructed image.

Make an edge claim testable

Generate test fields both with a sharp jump and with a smooth ramp. Pass them through the same sensor model and noise level, then fit with the same strength-selection procedure used for real data. Measure contrast bias and edge-location error separately; an image can put an edge in the correct place but suppress its amplitude. For a nonnegative material field inspect the gradient at active zeros alongside the residual. The regulariser and the bounds each supply information that the instrument may not have measured.

Follow total variation through three coupled cells

Consider three ordered cell values observed directly with equal independent precision. Minimise $\frac{1}{2}\sum_i(x_i-b_i)^2+\lambda(|x_2-x_1|+|x_3-x_2|)$. Introduce edge variables s_1,s_2 equal to the signs of nonzero differences. At a zero difference either variable may take any value between minus one and one. Differentiation with these admissible subgradients gives the three balances $x_1-b_1-\lambda s_1=0,\ x_2-b_2+\lambda(s_1-s_2)=0,\ x_3-b_3+\lambda s_2=0$. A candidate solution must satisfy all three balances and must agree with the sign choices that produced it.

Adding the balances eliminates both internal edge forces. Consequently the reconstructed values have exactly the same sum as the observations, regardless of the penalty strength. The quadratic part is strictly convex, so any feasible sign assignment satisfying the balances proves that the state is the unique minimiser. Adding lower bounds can introduce reactions and break the sum identity. Unequal precision also changes which weighted average is conserved. The calculation depends on the observation operator being the identity; after blurring, the first derivative of the data term already couples all cells.

Derive the complete fusion path for a short profile

Take $b = (0, 1, 5)^T$. For small positive λ, both differences remain positive, so

$$s_1 = s_2 = 1.$$

The equations give

$$x = (\lambda, 1, 5 - \lambda)^T.$$

The first two coordinates meet at $\lambda = 1$, while the second and third are still unequal. Assume those first two now form a plateau of height a, with $s_2 = 1$. Adding their balances gives $2a - 1 - \lambda = 0$, hence $a = (1 + \lambda)/2$. The first equation then requires

$$s_1 = a/\lambda = (1 + \lambda)/(2\lambda),$$

which lies in the permitted interval exactly when $\lambda \geq 1$.

The third coordinate stays $5 - \lambda$. Equating it to the plateau gives $(1 + \lambda)/2 = 5 - \lambda$, or $\lambda = 3$. Beyond this value all three cells coincide, and sum conservation fixes them at two. The required subgradients are $s_1 = 2/\lambda$ and $s_2 = 3/\lambda$, both admissible for $\lambda \geq 3$. This completes the path without numerical optimisation. At strengths 0.5, 2 and 4 the respective profiles are $(0.5, 1, 4.5)$, $(1.5, 1.5, 3)$ and $(2, 2, 2)$.

The thresholds record when the data imbalance inside a proposed plateau can be supported by the admissible edge subgradients. A flat patch at the second strength is therefore compatible with unequal raw values of zero and one. Conversely, a visible third-cell edge at that strength has already lost much of its amplitude. A reconstruction report that uses one threshold for declaring the existence of a jump and another for estimating its amplitude should explain both decisions.

Checking the final all-fused solution using only its observed mean would miss the second edge constraint. For example at strength two that constant candidate would require $s_2 = 1.5$, outside the admissible unit interval, despite reproducing total mass perfectly. The failed local balance tells which boundary of the candidate plateau has to reopen. This provides a simple hand diagnostic for more complicated one-dimensional reconstructions when an apparently reasonable flat interval has been proposed.

Choose how weak coefficients enter a reconstruction

For unit forward gain, minimise half squared data error plus either $0.5|x|$ or $0.5x^2/2$. Solve the derivative equation away from zero and use the subgradient at zero for the absolute-value penalty. Here the plotted input is the measured datum.



A reading of 0.4 is suppressed entirely by the absolute-value penalty but becomes about 0.267 with quadratic shrinkage. A reading of 2 becomes 1.5 or 1.333 respectively. The weighting conventions matter when comparing penalties.

▶ **Read values from the plot**

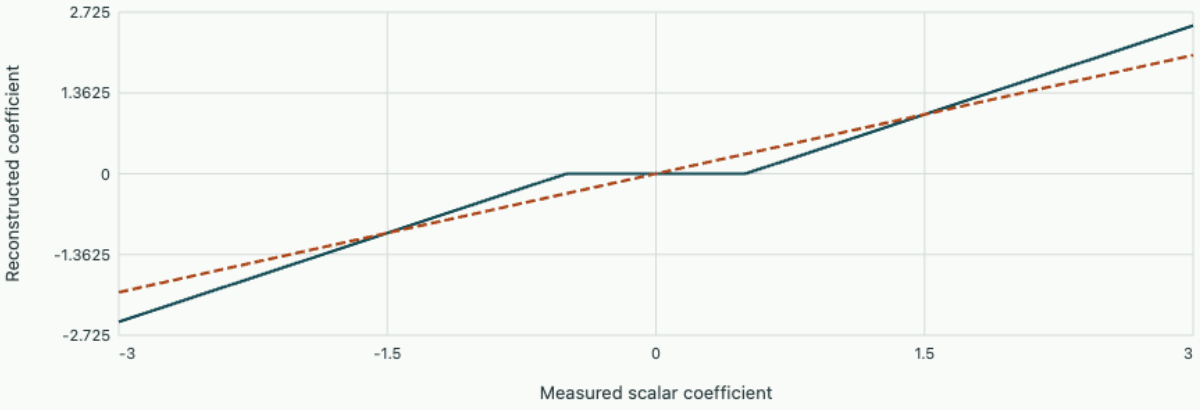
Change parameters · Choose how weak coefficients enter a reconstruction

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

Penalty weight λ **0.5**



The objectives are $(x - y)^2/2 + \lambda|x|$ and $(x - y)^2/2 + \lambda x^2/2$. The first sets all $|y| \le \lambda$ exactly to zero.

Study the full derivation

Recovering sparse coefficients without hiding amplitude bias · Inferring parameters and fields from observations

A sensor records a linear mixture of a few active components. Derive the soft-threshold map, use it in an iterative reconstruction and decide when correlated response columns make an apparently sharp sparse answer ambiguous.

Chapter 12 · A reconstruction audit from observations to a defensible claim

Combine residuals, resolution, uncertainty and independent predictions to decide what one spatial experiment establishes.

FROM FIRST PRINCIPLES

Audit a recovered contrast in measurement units

Suppose two equally calibrated channels sense neighbouring state values with symmetric cross-talk $0 \le a < 1/2$. Their noiseless observations are $y_1 = (1-a)x_1 + a x_2$ and $y_2 = a x_1 + (1-a)x_2$. Subtracting removes the common level and reveals how strongly the desired contrast $d = x_1 - x_2$ appears in the readings,

$$y_1 - y_2 = (1-2a)d.$$

Independent zero-mean channel noise of common standard deviation s gives difference variance $2s^2$. Inverting only this observable therefore yields the data-based estimate and uncertainty

$$\widehat{d} = \frac{y_1 - y_2}{1-2a},$$

$$\mathrm{sd}(\widehat{d}) = \frac{\sqrt{2}\,s}{|1-2a|}.$$

The noise is amplified as the channels approach identical response. A useful resolution audit reports the measured difference, the calibrated contrast gain and the relevant uncertainty sources. If the gain itself is uncertain or channel noise is correlated, those effects must enter the audit before a physical contrast is claimed.

Background · Resolution matrix

$$R = G^{\dagger}_{\mathrm{reg}} G$$

For a linear regularised inverse, a row describes how the estimated parameter mixes the actual components. A sharp-looking result may still have broad resolution kernels.

DERIVATION AND APPLICATIONS

Quantify signal leakage into an estimate of a local field

A reconstructed pixel is often a weighted average of several true cells. A resolution audit can predict that mixing and bound its effect using independent knowledge of admissible field gradients. Consider a linear

model $y = Ax + \epsilon$ with zero-mean measurement error of covariance Σ, and a chosen linear estimator for the target cell j, $\widehat{x}_j = w^T y$. Define its resolution weights by the response to unit signals in individual true cells,

$$k^T = w^T A,$$
$$\mathrm{E}\,\widehat{x}_j = \sum_i k_i x_i,$$
$$\operatorname{var} \widehat{x}_j = w^T \Sigma w.$$

To reproduce a uniform field exactly, require $\sum_i k_i = 1$. The bias relative to the target is then a sum of differences from that cell. Suppose independently supported spatial smoothness gives

$$|x_i - x_j| \le L d_{ij},$$

where d_{ij} is cell distance. Taking absolute values yields

$$|\mathrm{E}\,\widehat{x}_j - x_j| = \left| \sum_i k_i (x_i - x_j) \right| \le L \sum_i |k_i| d_{ij}.$$

This weighted distance has units of length even when resolution weights have negative lobes. If Gaussian measurement noise is justified, a conservative two-sided interval with at least the usual 95 percent coverage for the true cell is

$$\widehat{x}_j \pm \left[L \sum_i |k_i| d_{ij} + 1.96 \sqrt{w^T \Sigma w} \right].$$

The bias bound is added to the statistical margin because it represents an unknown fixed error, not an independent random draw. The guarantee depends on the supplied smoothness information and calibrated forward model. At a genuine sharp interface the assumed gradient bound may be inappropriate.

WORKED EXAMPLE

Audit a local mean on a uniform grid

Three true cell centres lie $0.5\,\mathrm{m}$ apart. An estimator of the middle temperature has weights $(0.2, 0.6, 0.2)$, output noise deviation $0.10\,\mathrm{K}$ and independently verified gradient bound $2.0\,\mathrm{K/m}$. Its effective weighted distance is $0.20\,\mathrm{m}$ and maximal mixing bias is $0.40\,\mathrm{K}$. An estimate of $18.0\,^{\circ}\mathrm{C}$ has interval half-width $0.40 + 1.96(0.10) = 0.596\,\mathrm{K}$.

SECOND APPLICATION

Compare a sharper reconstruction at a noisy boundary

An alternative three-cell estimator has weights $(-0.1, 1.2, -0.1)$ at distances $1, 0, 1\,\mathrm{m}$. For $L = 0.50\,\mathrm{K/m}$ its bias bound is $0.10\,\mathrm{K}$. With noise deviation $0.30\,\mathrm{K}$, the total half-width is $0.688\,\mathrm{K}$. A smoother estimator of the same target with weighted distance $0.50\,\mathrm{m}$ and noise deviation $0.12\,\mathrm{K}$ gives

$$0.25 + 0.235 = 0.485\,\mathrm{K}.$$

The sharper-looking estimator yields the wider interval for this stated field class.

Set a question that can fail

Two adjacent thermal zones are blurred by the sensor matrix $A = \left(\begin{smallmatrix} 0.55 & 0.45 \\ 0.45 & 0.55 \end{smallmatrix}\right)$. Let the observations be $y = (21.08, 20.92)$ °C with independent standard deviations 0.10 K. The engineering question is whether the first zone is at least one kelvin hotter than the second. Work in mean and difference coordinates. The mean passes unchanged, while the difference is attenuated by 0.10. Before fitting, predict that uncertainty in the contrast will dominate this decision.

Derive the data-only estimate and interval

The observed mean is 21 °C, with standard deviation $0.10/\sqrt{2} = 0.0707$ K. The estimated true contrast is

$$\hat{d} = (21.08 - 20.92)/0.10 = 1.60\,\mathrm{K}.$$

Its standard deviation is

$$\sqrt{2}(0.10)/0.10 = 1.414\,\mathrm{K}.$$

For known Gaussian noise, a two-sided 95% data-only interval is about $[-1.17, 4.37]$ K. The inverse values $(21.8, 20.2)$ exactly reproduce both readings. Nevertheless, even the sign of the contrast is uncertain at this level. The exact residual consumed both degrees of freedom and cannot test the instrument model.

Compare a prior-based inference on the same scale

Assume an independent prior for the contrast $d \sim N(0, 1\ \mathrm{K}^2)$, leaving the mean very weakly constrained. In contrast coordinates the data variance is 2 and the observed estimate is 1.60. Adding precisions gives posterior variance $1/(1 + 1/2) = 2/3$ K² and mean $(2/3)(1.60/2) = 0.533$ K. The probability $P(d > 1 \mid y)$ is $1 - \Phi[(1 - 0.533)/\sqrt{2/3}] \simeq 0.28$. The prior changed both the centre and

the width; this probability is explicitly conditional on the zero-centred prior. Repeating with a wider prior shows how much the conclusion depends on that assumption.

Buy information in the direction the decision needs

Suppose one additional test is available. Another independent repeat with the same blurred sensor reduces the data-only standard deviation of the contrast by $\sqrt{2}$, to 1.0 K. A new differential sensor reading $d + \epsilon$ with standard deviation 0.50 K and errors independent of the original instrument directly adds precision 4 K⁻² to the existing 0.5 K⁻², giving combined standard deviation $1/\sqrt{4.5} = 0.471$ K. This pre-data comparison depends on geometry and precision alone, not on an unknown future reading. Use held-back calibration objects to verify that the new sensor's stated precision and averaging model hold near an actual interface.

Communicate the final claim

Retain the observable mean and the uncertain contrast as separate reported quantities. State the noise law and model assumptions behind any interval. Where a numerical solver contributed, distinguish verified numerical precision from the statistical precision of the data. The difference is also visible in the temperature-design case: making the optimisation converge more tightly cannot restore information lost when all temperatures are close to the reference. A successful analysis may conclude that a particular observation needs redesign while still supplying an accurate prediction of another quantity.

Propagate a spatially varying calibration error

The preceding contrast calculation assumes a known blur and unbiased independent channel errors. Add two possible calibration terms: an offset b shared by both readings and a differential offset c that enters the first reading as $+c/2$ and the second as $-c/2$. The measurement equations then give mean reading equal to the true mean plus b and mean random error. Their difference is $0.10d + c + e_1 - e_2$. A shared offset cancels exactly from the contrast, whereas a differential offset is amplified tenfold when the contrast is reconstructed.

If the differential offset has an independently calibrated mean zero and known standard deviation s_c, variance propagation gives $\mathrm{Var}(\hat{d}) = (2s_e^2 + s_c^2)/0.10^2$ for equal individual standard deviations s_e. Treating the offset as random is appropriate for an ensemble of independently calibrated experiments with the stated calibration model. Within one installation it is a common unknown across all repeated exposures. If only a deterministic bound on c is justified, its effect is better reported as a bounded systematic displacement in addition to random sampling uncertainty.

Test the decision after averaging a hundred exposures

Suppose the original pair is observed one hundred independent times with the same differential calibration. The observed average difference remains 0.16 K, individual channel noise is 0.10 K per

exposure, and calibration tests support $|c| \le 0.04$ K. The standard deviation of the averaged difference from independent channel noise alone is

$$\sqrt{2}\,0.10/10 = 0.01414\,\mathrm{K}.$$

In physical contrast units the random standard deviation is 0.1414 K. The possible systematic error is $0.04/0.10 = 0.40$ K in either direction, undiminished by the additional exposures.

A conservative one-sided 95 percent lower limit under Gaussian independent channel noise subtracts both the worst allowed positive offset and the sampling margin. In contrast units it gives $1.60 - 0.40 - 1.645(0.1414) \approx 0.967$ K. Even with one hundred exposures this calculation falls slightly below the one-kelvin engineering threshold. Omitting the differential offset would give a lower limit about 1.367 K and reverse the decision. The discrepancy originates in the calibration model, although all raw measurements and all random-noise calculations are unchanged.

What new evidence would change the result? A reference specimen known to have zero contrast estimates c directly through the observed channel difference. If calibration tightens the guaranteed bound to 0.01 K, the same calculation gives $1.60 - 0.10 - 0.233 \approx 1.267$ K. If the reference itself has uncertain contrast, propagate that uncertainty through its own blur factor before claiming the tighter bound. This comparison assigns the new measurement a clear purpose: it reduces the component of uncertainty aligned with the decision, rather than collecting more observations of a systematic offset that cancels in neither channel difference nor repeated averages.

An orientation-reversal test can also distinguish true contrast from a fixed differential channel offset. Reversing a specimen of equal geometry reverses the sign of its true contrast but preserves the channel offset. Half the difference of the two measured contrasts then isolates the blurred physical contrast, provided the specimen is stable and both orientations sample comparable regions.

Design the reversal and propagate its uncertainty

Write the normal and reversed orientation mean differences as $D_+ = 0.10d + c + \epsilon_+$ and $D_- = -0.10d + c + \epsilon_-$. Solving for the two unknowns gives

$$\hat{d} = (D_+ - D_-)/0.20,$$

$$\hat{c} = (D_+ + D_-)/2.$$

The independent information comes from changing the sign of the physical response while keeping the offset fixed. Repeating the normal orientation alone produces another row proportional to the first and cannot separate the two quantities, however many times it is repeated.

Suppose each orientation is averaged over 100 exposures with the same independent channel noise as before. The standard deviation of either orientation difference is 0.01414 K. If the two averages have independent random errors, then $\mathrm{sd}(\hat{d}) = \sqrt{2}(0.01414)/0.20 = 0.10$ K and $\mathrm{sd}(\hat{c}) = 0.010$ K. With readings 0.18 K and −0.14 K, one obtains $d = 1.60$ K and $c = 0.02$ K as estimates. The one-sided lower contrast limit is $1.60 - 1.645(0.10) \approx 1.435$ K. Unlike the earlier bounded-offset calculation, this inference estimates the shared differential offset from the design itself.

If the specimen cools differently between orientations, or rotating it changes the sensor footprint, the two signs might no longer describe equivalent regions. A reversal carried out in alternating order across independent blocks can expose drift. Check the recovered offsets for consistency across these blocks before claiming that orientation has separated the physical contrast from an instrument artifact.

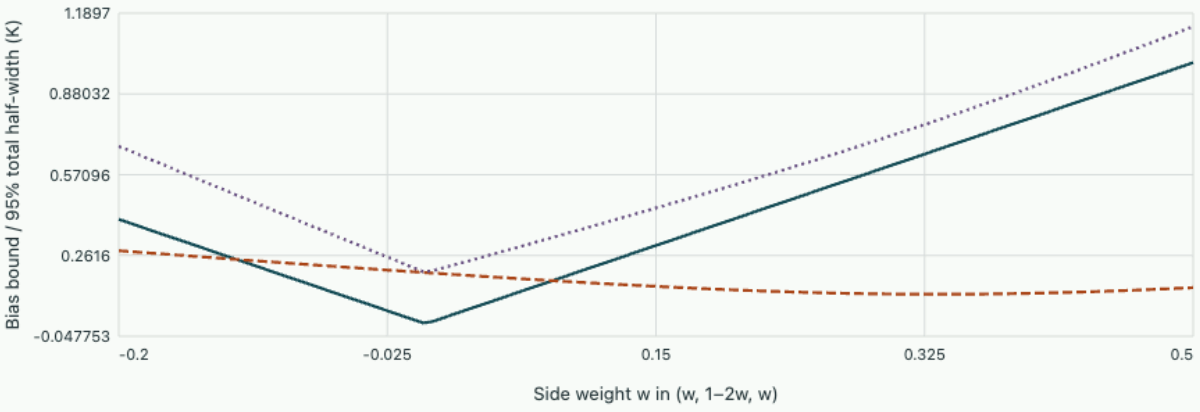
Change parameters · Explore the worked calculation

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

Verified gradient bound (K/m)	2
Independent noise SD (K)	0.1
Cell spacing (m)	0.5



The weights sum to one. Negative side weights sharpen some features but increase the absolute-distance bias bound and noise gain; a smaller variance alone does not establish a sharper reliable reconstruction.

Problem sheet 6 · Edges, decision thresholds and validation

Determine whether an apparent feature survives a change of penalty, a noise calculation and an independent test.

1. Fusing two adjacent estimates

Take identity observations $b = (1, 2.2)$. Minimise $\frac{1}{2}\sum(x_i - b_i)^2 + \lambda|x_2 - x_1|$. (a) Compute both fitted values for $\lambda = 0.2$ and 0.7. (b) Identify the fusion threshold. (c) Does observing a flat fit establish equal true values?

▼ **Retain the mean and shrink the half-difference**

The observed mean is 1.6 and half-difference 0.6. At $\lambda = 0.2$, the fitted half-difference is 0.4 and the fit is $(1.2, 2.0)$. At 0.7 it is zero, giving $(1.6, 1.6)$. Fusion starts at $\lambda = 0.6$. A flat solution can be induced by the penalty; report the relevant contrast, noise and sensitivity to penalty strength before making a physical equality claim.

Explore the related calculation with adjustable inputs. Inspect soft-thresholding and quadratic shrinkage. For two-cell total variation, apply the threshold to the half-difference and retain the mean as derived here.

---

2. A blur and a threshold

A sensor attenuates a physical contrast by 0.20 and gives a measured difference 0.60 K with standard deviation 0.20 K. (a) Find the data-only estimated contrast and standard deviation. (b) Can a one-sided 95% lower limit establish that the contrast exceeds 2 K? Use the Gaussian multiplier 1.645. (c) Assuming independent repeats with identical mean, how many would make that lower limit exceed 2 K?

▼ **Propagate the contrast scale only once**

The estimate is 3 K with standard deviation 1 K. The lower limit $3 - 1.645 = 1.355$ K is below 2 K. If the sample mean remains 3 K when expressed in physical units, require $1.645/\sqrt{N} < 1$, so three independent repeats suffice for that projected precision. Future means are random, hence this is a conditional planning calculation rather than a guaranteed outcome.

Explore the related calculation with adjustable inputs. Change the parameters in the corresponding calculation and compare its independently recomputed values with the equation derived in this solution.

---

3. Challenge a smooth reconstruction

A smoothed image fits all training readings within their stated noise and shows one broad transition. (a) Describe two distinct synthetic truths to use in a resolution audit. (b) Specify two numerical metrics and one measurement to hold out. (c) Explain how a code error, statistical uncertainty and regularisation bias would appear differently across the tests.

▼ **Use tests that can disagree with the interpretation**

Choose a sharp step and a gradual ramp with similar large-scale contrast. Generate observations using a finer or independent forward discretisation, realistic noise and the actual sensor footprint. Record recovered edge location and recovered contrast or width. Reserve an independent measurement that is sensitive to the transition, ideally at a finer spatial scale. Failure on a noiseless manufactured case can expose an implementation mistake. Scatter across noise realisations quantifies sampling variation. A consistently broadened step across seeds at an otherwise verified discretisation indicates regularisation or finite sensor resolution; vary both separately to distinguish them.

Choose a sensor direction that resolves an unobservable combination

A prior for two parameters is Gaussian with zero mean and identity covariance. A first reading measures $x_1+x_2 = 2.0$ with noise standard deviation 0.1. Find the posterior mean and the variance along each normalised sum and difference direction. A second sensor of the same noise level can measure the sum again or the difference. Which design constrains both parameters, and why?

▼ **Worked solution**

In the orthonormal directions $s = (1,1)/\sqrt{2}$ and $d = (1,-1)/\sqrt{2}$, the first measurement row is $(\sqrt{2},0)$. Its precision contribution has diagonal $(200,0)$. Adding the unit prior precision gives posterior variances $1/201$ and 1. Multiplying the precision-inverse by $A^{\mathsf{T}}b/0.01$ gives a posterior mean with both components $200/201 \approx 0.995$.

Repeating the sum doubles the sum-direction information, giving variances $1/401$ and 1. Measuring the difference instead adds information along d, giving $1/201$ for both directions. The trace of the second covariance is about 0.00995, versus about 1.0025 for the repeated-sum design. The second measured value changes the posterior mean but, for this linear Gaussian model, not the covariance.

An objective concerned only with the sum would favour the repeated sensor. For reconstructing both parameters, sensitivity in independent directions matters. A very small residual for the original measurement is compatible with a whole line of plausible parameter pairs, which is why reporting only that residual would conceal the design difference.

Explore the related calculation with adjustable inputs. Change the parameters in the corresponding calculation and compare its independently recomputed values with the equation derived in this solution.

Chapter 13 · Choosing regularisation when the true answer is hidden

Derive two parameter choices from different information, then see how an excellent fit to sensor data can restore the noise in a weakly visible mode.

Each observation competes with a penalty

In orthogonal singular coordinates, let $d_i = \sigma_i x_i + \eta_i$. For a candidate $\lambda > 0$, minimise $\sum_i (\sigma_i x_i - d_i)^2 + \lambda^2 \sum_i x_i^2$. Setting each derivative to zero gives

$$\widehat{x}_i = \sigma_i d_i / (\sigma_i^2 + \lambda^2).$$

Predicted data are $H_i d_i$, where $H_i = \sigma_i^2/(\sigma_i^2 + \lambda^2)$. These factors range from zero to one. Their sum is the effective number of fitted data directions.

If an independently obtained bound for the whitened noise norm is δ, compare $\sum_i [(1 - H_i) d_i]^2$ with δ^2. The residual rises as λ increases. A discrepancy choice seeks the largest smoothing consistent with the stated noise level. If there is no crossing, either the search range, model or noise estimate needs examination.

A prediction criterion without a noise bound

For a linear smoother $\widehat{d} = Hd$, deleting observation i gives the leave-one-out residual $r_i/(1 - H_{ii})$ under the usual fixed-design ridge fit. Generalised cross-validation replaces the individual leverages by their mean. This produces the computable quantity

$$\mathrm{GCV}(\lambda) = \frac{\|d - \widehat{d}\|^2/n}{[1 - \operatorname{trace}(H)/n]^2}.$$

The denominator penalises fits that spend nearly one degree of freedom per datum. The criterion estimates predictive performance under assumptions on noise and design. A selected value need not minimise the error in the unknown parameter, and highly unequal leverages can defeat the approximation.

A sensor that barely sees a mode

Take sensitivities $1, 0.01$, true parameters $(1, 1)$ and data $(1, 0.02)$. Unregularised inversion returns $(1, 2)$ with zero residual. A perturbation of one hundredth in the second observation doubled the second parameter. With $\lambda = 0.01$ the second gain is $\sigma/(\sigma^2 + \lambda^2) = 50$, so the same datum yields 1. This exact agreement with the hidden truth is accidental.

For $\lambda = 0.1$, the second estimate drops to about 0.0198. A weak direction has been almost removed. Meanwhile the first estimate remains about 0.9901. A single residual norm does not tell you which parameter direction was suppressed.

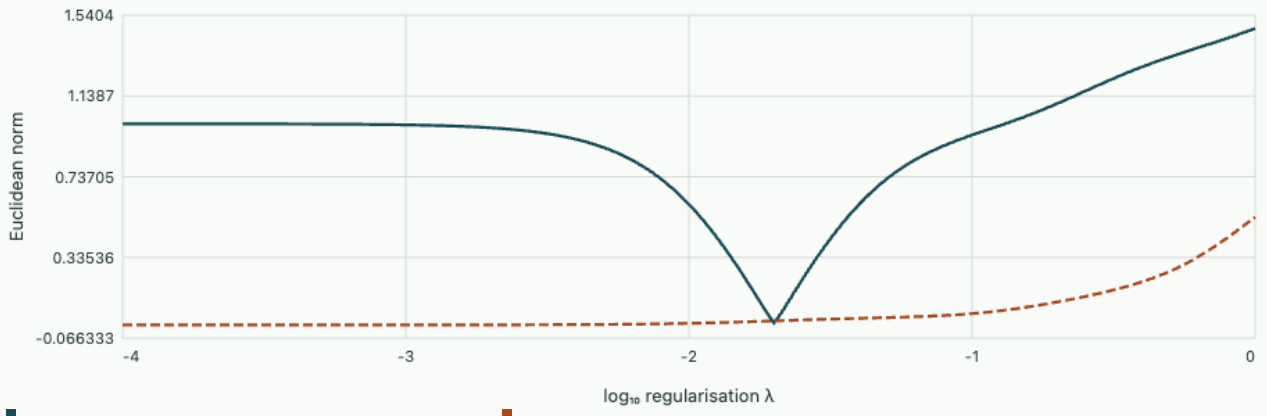
Inspect an entire regularisation path

Three synthetic sensors have sensitivities $1, 0.2, 0.02$, true parameters all equal to one, and an adjustable perturbation only in the last sensor. This deliberately uneven noise makes it possible to test when the GCV assumptions are plausible. Inspect parameter error and data residual over a logarithmic range of λ. The result reports the best GCV value on the sampled grid and the residual at that choice.

Open this experiment online

Try the calculation

Error in weakest observation **0.02**



Grid minimum GCV 0.0046083 at λ = 0.035481, giving residual 0.030988. Actual data noise norm 0.02.

Tutorial on choosing a criterion

If all H_i equal a common $h < 1$, can GCV distinguish different values of h for fixed data?

▼ **Worked solution**

The numerator contains $(1-h)^2\|d\|^2/n$ and the denominator $(1-h)^2$. They cancel exactly. There is no identifiable minimum from this score.

If residuals remain much larger than a well-calibrated noise norm even at negligible regularisation, should increasing the penalty help?

▼ **Worked solution**

Increasing the penalty shrinks the fit and usually increases residual. Check the forward model, its range and data preparation, including the uncertainty units used in whitening.

Read the university treatment

Compare the information each stopping rule needs about measurement noise with the information available in the three-sensor experiment. Per Christian Hansen · Stopping rules for noisy data

Study the full derivation

Choose a regularisation scale by predicting omitted data · Inferring parameters and fields from observations

Regularised least squares can fit each observation partly with itself. Derive leave-one-out residuals from a single factorisation, approximate them by a trace and identify when a generalised cross-validation score misses a highly influential sensor.

Chapter 14 · Learning a drifting temperature from a sequence of readings

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

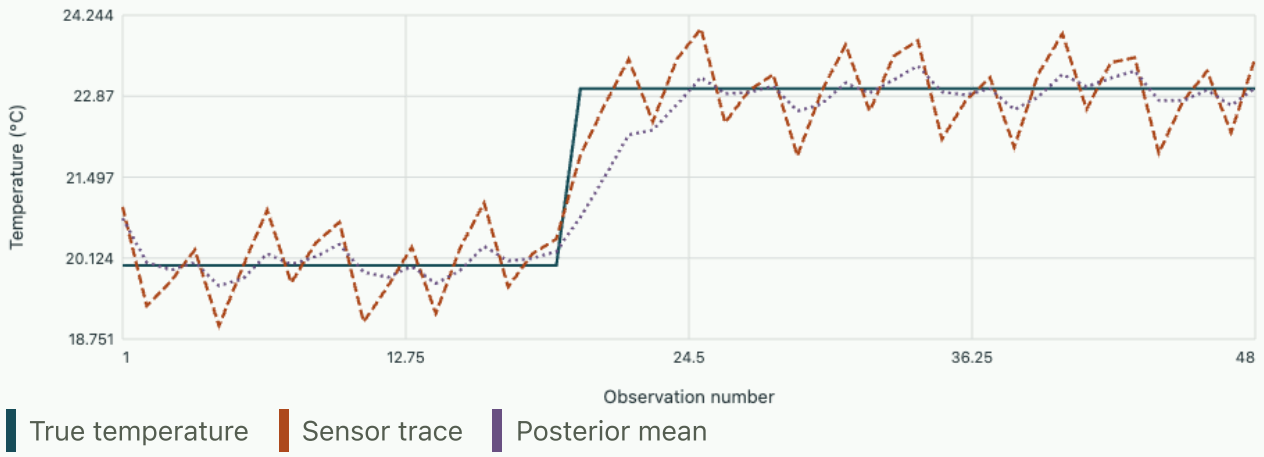
For a reproducible illustration, a fixed synthetic thermometer trace contains a smooth periodic disturbance and a step in the hidden temperature. It is deliberately held fixed while the filter's assumed Q and R change. Compare the lag just after the step with the wiggle once the temperature settles.

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

Study the full derivation

What each new sensor reading changes in a Gaussian inverse problem · Inferring parameters and fields from observations

Start with an uncertain vector, condition on one noisy linear observation, and reconstruct the posterior mean and covariance without repeating a batch inversion. Check shared calibration errors before treating repeated measurements as independent evidence.

Chapter 15 · When a single uncertainty interval hides two explanations

Start from a squared-amplitude measurement, derive the full posterior and design a second observation capable of identifying the sign.

Symmetry is an identifiability statement

A sensor measures $d = x^2 + \eta$ with Gaussian noise of standard deviation s. A symmetric Gaussian prior of variance τ^2 puts equal weight on opposite signs. Bayes' rule gives the density

$$p(x \mid d) \propto \exp\left[-\frac{(d - x^2)^2}{2s^2} - \frac{x^2}{2\tau^2}\right].$$

The exponent is even, so the posterior mean is exactly zero whenever the integral exists. That mean may make an extremely poor prediction of the observed quantity. To locate modes, differentiate the negative log density to get $x[2(x^2 - d)/s^2 + 1/\tau^2]$. Two off-centre modes appear precisely when $d > s^2/(2\tau^2)$. At smaller d, the prior and noise merge the explanations into one central mode.

Numbers and competing summaries

Set $d = 1$, $s = 0.2$, and $\tau = 2$. The modes are at $\pm\sqrt{0.995}$, close to ± 0.9975. The density ratio of $x = 0$ to $x = 1$ is $\exp[-1/(2s^2) + 1/(2\tau^2)]$, about 4.2×10^{-6}. The reported mean of zero sits in a very low-density region.

For $d = -0.05$, there is no real parameter that exactly fits the measured amplitude squared. A noisy sensor can still produce this value. The distribution has a single mode at zero, and there is no justification for taking the square root of the negative datum or discarding the observation.

Buy a measurement that identifies the sign

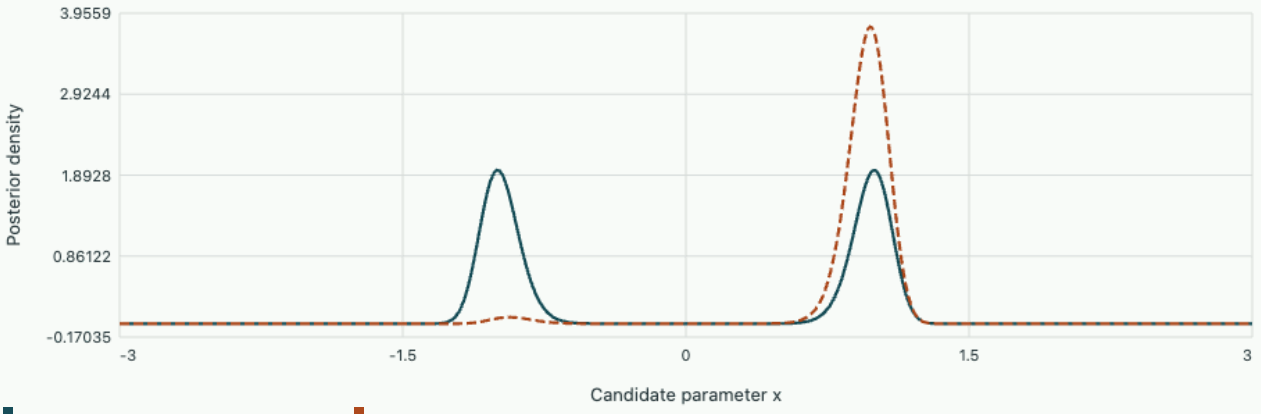
Assume an independent second sensor reports $b = x + \epsilon$ with known standard deviation 0.5. Add $-(b - x)^2/(2 \cdot 0.5^2)$ to the log density. The original two signs now receive different weights. Both curves in the plot are independently normalised by quadrature on the displayed interval; the sign probability is conditional on that interval. Compare the observation $b = 0$ with $b = 1$.

Open this experiment online

Try the calculation

Squared-amplitude datum d **1**

Signed sensor reading b **0.5**



Probability of positive x on [−3,3]: 50% before the signed measurement, 97.698% afterwards.

Tutorial on designing another experiment

Repeat the squared-amplitude measurement a hundred times with independent noise. What can become precise?

▼ **Worked solution**

The magnitude x^2 can become very precise. Every likelihood factor remains even in x, so the sign probabilities stay equal with this prior. More data of the same kind cannot break that symmetry.

If the two candidate states were exactly $x = \pm 1$, what likelihood ratio does $b = 0.5$ give in favour of the positive state?

▼ **Worked solution**

The exponent difference is

$$((0.5 + 1)^2 - (0.5 - 1)^2)/(2 \cdot 0.5^2) = 4.$$

The likelihood ratio is e^4, about 54.6. A continuous posterior also averages over the spread around each candidate.

Chapter 16 · How model error masquerades as a fitted physical parameter

Project an omitted spatial effect into the fitted parameter space. Work out when residual checks reveal the omission and when they are silent.

A misspecified instrument

Suppose $d = Ax_* + b + \eta$, with zero-mean measurement noise and a deterministic but omitted contribution b. Assume A has independent columns. Least squares gives

$$\widehat{x} = (A^T A)^{-1} A^T d.$$

Taking expectations leaves the parameter bias $A^+ b$. The fitted residual is $(I - AA^+)(b + \eta)$. Thus the same model error separates into a component that biases the parameters and an orthogonal component that remains visible to residual checks.

For a single sensitivity column a, decompose $b = b_{\|} + b_{\perp}$, where $b_{\|} = a(a^T b)/(a^T a)$. Repeating the measurement reduces independent random error; the deterministic projection remains. Error bars computed solely from measurement variance can become more misleading as more readings are acquired.

Calibration with an omitted offset

Fit $d_i = mt_i$ at $t = (1, 2, 3)$ to observations actually generated by $d_i = 2t_i + 0.3$. The unmodelled constant has nonzero projection onto t. The bias is $0.3(1 + 2 + 3)/(1 + 4 + 9) = 0.12857$, giving fitted slope 2.12857. The residuals at the three points are approximately 0.17143, 0.04286 and −0.08571. Their signed trend suggests a missing intercept. At $t = 6$ the model predicts 12.7714 while the actual law gives 12.3; an apparently modest training residual has grown into a substantial extrapolation error.

Now use a discrepancy exactly proportional to time, $b_i = 0.1t_i$. The fitted slope becomes 2.1 with zero residual at every training point. Additional times under the same model cannot separate the true slope from this contribution. A separate calibration that suppresses the extra physical effect can.

Rotate the missing physics

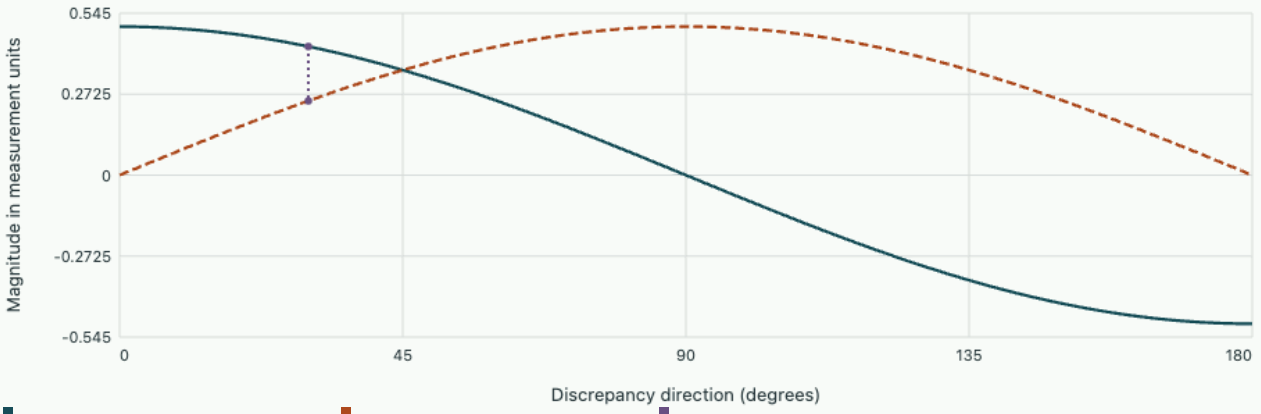
In a two-reading instrument, the model sensitivity is the unit horizontal vector. Vary the magnitude and direction of the omitted contribution. Observe parameter bias and visible residual separately, and locate cases with the same discrepancy norm but completely different fit diagnoses.

Open this experiment online

Try the calculation

Omitted-effect norm **0.5**

Direction (degrees) **30**



Signed parameter bias | Visible residual norm | Selected direction

At 30° the parameter bias is 0.43301 and residual norm 0.25.

Tutorial on experimental design

Repeat the offset example at symmetric times $(-1, 0, 1)$. Which part of the fitted result changes?

▼ **Worked solution**

The constant is now orthogonal to t, giving zero slope bias. The residual is a constant 0.3 in every reading. The experiment has made an offset visible without corrupting the slope.

Can adding an unrestricted discrepancy parameter for every datum solve the identifiability problem?

▼ **Worked solution**

It permits exact fits for any physical parameter by assigning the remaining signal to the discrepancy. Constrain the discrepancy from theory, independent data or a justified prior to extract the physical parameter.

Read the university treatment

Compare errors that project into a fitted singular direction with errors seen in the orthogonal data residual. Per Christian Hansen · Discrete Inverse Problems

Chapter 17 · Predicting a new measurement when the fitted parameters are correlated

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

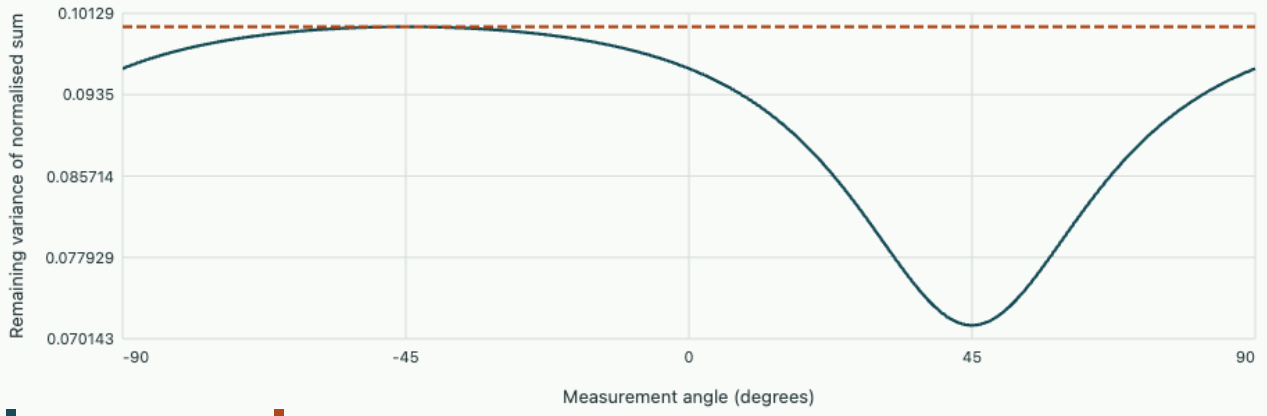
For a fairer geometric comparison, restrict both measurement and prediction directions to unit length. Fix the target along the normalised sum. Sweep candidate observation angles and read the remaining prediction variance. Change the parameter correlation: when does a coordinate sensor improve the target 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

Chapter 18 · Calibrating two instruments when both axes fluctuate

The horizontal axis of a calibration plot is often described as an input. In this experiment it is itself a reading from a second noisy instrument. Infer a straight calibration relation from paired data, derive the metric in which the fit should be performed, and use a change of units to reveal an easily overlooked modelling error.

What was actually held fixed during the experiment?

At setting i, let the unknown true first-channel signal be ξ_i, measured in millivolts. Suppose the true second-channel value, in kelvin, is $a + b\xi_i$ over the limited operating range. What the laboratory records is the pair $x_i = \xi_i + \epsilon_{xi}$,

$$y_i = a + b\xi_i + \epsilon_{yi}.$$

Repeatability tests supply standard deviations $s_x, s_y > 0$. For now fix them across observations, with independent zero-mean Gaussian errors in both channels. The physical settings ξ_i are treated as unknown fixed quantities, one per record.

Ordinary vertical least squares implicitly treats the observed x_i as the exact horizontal values. That choice would make sense if the settings had been specified by a sufficiently accurate standard and x_i denoted those settings. It changes the inference when x_i comes from a noisy meter. Every displayed point can then move both horizontally and vertically as we estimate its underlying position on the calibration line.

A geometric phrase like “shortest distance to the line” has no unique experimental meaning yet. A displacement of one millivolt and a displacement of one kelvin cannot be compared by drawing the points on square graph paper. Measurement uncertainty supplies a metric. Divide each horizontal adjustment by s_x and each vertical one by s_y. Under the stated Gaussian assumptions their squared sum is proportional to twice the negative log likelihood, up to additive constants independent of the parameters. This also clarifies why the relative scale requires a separate calibration: one cloud does not automatically reveal how much of its spread belongs to each axis.

Eliminate the unobserved setting algebraically

For a proposed line, estimate each unknown ξ_i by minimising its pair of squared corrections. Define $r_i = y_i - a - bx_i$ and write $\xi_i = x_i + \delta_i$. Then the horizontal term is δ_i^2 / s_x^2, and the vertical term is $(r_i - b\delta_i)^2 / s_y^2$. Differentiating their sum with respect to δ_i leaves a scalar linear equation. Its solution apportions the original vertical residual between the two measurement errors.

$$\delta_i = \frac{b s_x^2 r_i}{s_y^2 + b^2 s_x^2},$$

$$y_i - a - b\xi_i = \frac{s_y^2 r_i}{s_y^2 + b^2 s_x^2}.$$

For example, let $b = 2\,\mathrm{K/mV}$, $s_x = 0.5\,\mathrm{mV}$, $s_y = 1\,\mathrm{K}$, and $r_i = 1\,\mathrm{K}$. The inferred horizontal correction is 0.25 mV and the remaining vertical error is 0.5 K. In uncertainty units each has magnitude one half and the squared cost is one half. Keeping the first reading fixed would assign the whole kelvin to the second channel and produce cost one. The formula decides the split from the line sensitivity and the two instrument uncertainties.

Substitute the optimal correction into the objective. The profiled contribution simplifies to $r_i^2/(s_y^2 + b^2 s_x^2)$. For fixed slope the denominator is common to all records. Differentiating with respect to a therefore gives

$$a = \bar{y} - b\bar{x}.$$

Every candidate slope now passes through the sample centroid. The remaining scalar optimisation is an independent way to check a geometric solution.

$$J(b) = \frac{\sum_i [(y_i - \bar{y}) - b(x_i - \bar{x})]^2}{s_y^2 + b^2 s_x^2}.$$

Do not attach a slope-dependent logarithm of this denominator to the present objective. We obtained it by maximising over fixed nuisance coordinates, each with its own original pair of measurement densities and fixed noise normalisation. A different experiment might specify a distribution for the latent settings and integrate them out instead. That would be a different likelihood construction. The fitting rule follows from how the unknowns and observations have been modelled.

Fit in coordinates in which the noise is isotropic

Make a data matrix with one row per observation. Subtract the centroid and divide the first column by s_x, the second by s_y. Call the result Z. A line through the origin in these whitened coordinates is completely specified by a unit normal n, including vertical lines that would cause a slope parameter to diverge. The squared perpendicular distances sum to $\|Zn\|^2 = n^T Z^T Z n$. Minimising over unit normals selects the eigenvector of the smallest eigenvalue of $Z^T Z$. With a singular value decomposition $Z = U\Sigma V^T$, it is the last right singular vector.

The other right singular vector points along the fitted line. Transform that direction back to the physical instrument units by multiplying its components by s_x and s_y. The slope is the ratio of those transformed components. The sign of the singular vector is mathematically arbitrary and has no effect on the line. The feet of the perpendiculars are simple in whitened coordinates: subtract $(z_i \cdot n)n$ from each whitened point. Transform them back and add the centroid. In a physical plot with unlike axis scales those connecting segments need not look perpendicular by eye.

A three-point calibration you can check without a computer

Use dimensionless readings $(-1, -2), (0, 1), (1, 1)$ and equal measurement standard deviations. Both column means are zero. The centred sums are

$$S_{xx} = 2, S_{yy} = 6, S_{xy} = 3.$$

Ordinary vertical least squares gives slope $3/2$. For a line direction $(1, b)$, requiring that this direction be an eigenvector of the scatter matrix yields

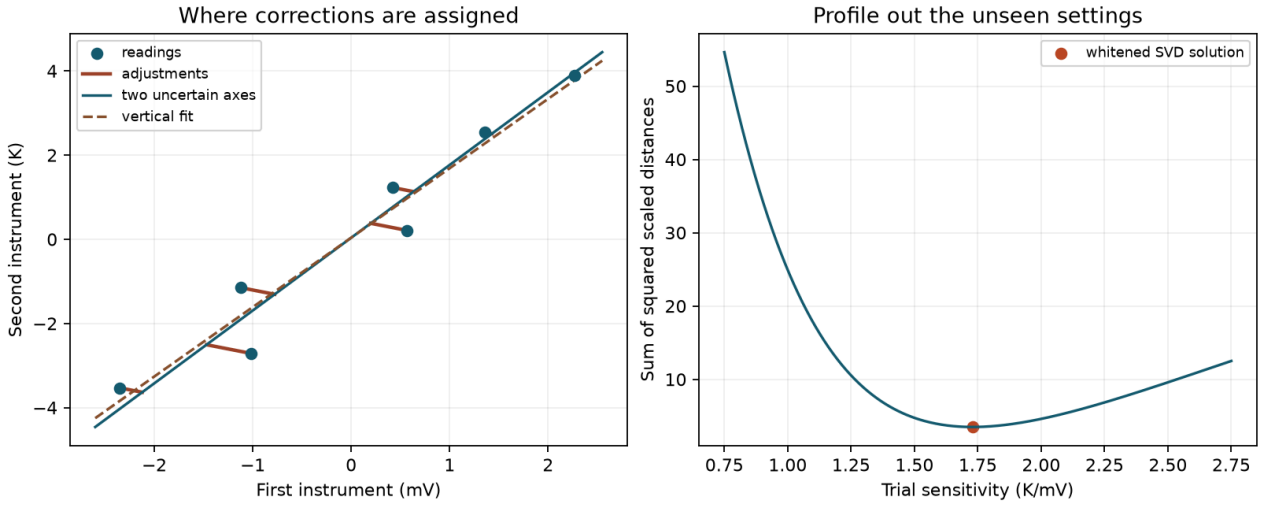
$$3 + 6b = b(2 + 3b).$$

There are two roots. The direction along the major axis corresponds to the positive one,

$$b = (4 + \sqrt{52})/6 \simeq 1.8685.$$

The negative root is the perpendicular minor-axis direction.

The residual cost of the vertical fit, evaluated in its own vertical metric, should not be directly compared with the perpendicular cost of the other fit. Evaluate both using the same profiled cost instead. With $s_x = s_y = 1$, the minimum equals the smaller scatter eigenvalue $(8 - \sqrt{52})/2 \simeq 0.39445$. Substituting 1.5 into the profiled formula yields $(6 - 9 + 4.5)/(1 + 2.25) \simeq 0.46154$. The orthogonal fit has achieved the smaller value in the metric actually specified by the instrument model.



The marked minimum comes from the small singular vector of the whitened observations. The curve calculates the same score by independently eliminating the unknown horizontal reading associated with each data point.

When is a best direction too weak to interpret?

The separating quantity is the gap between the two eigenvalues of the whitened scatter matrix. If they are nearly equal, the data cloud is nearly round in the metric that matters. Rotating the line then changes the score very little. If they are exactly equal, every line through the centre has the same score. A numerical routine can return a perfectly normalised eigenvector in this situation. Its normalisation and tiny algebraic residual say nothing about whether the data determine a preferred direction.

Try the notebook's seven paired readings first with its default uncertainty scales. The weighted slope is approximately 1.7290 K/mV and the vertical slope about 1.6490 K/mV. Decrease the declared uncertainty of the first channel while retaining the numerical observations. The fit approaches the vertical least-squares answer because horizontal adjustments become expensive. Increasing this uncertainty instead assigns more of a discrepancy to the first instrument. When the recorded inputs span too narrow a range, even very accurate readouts cannot robustly establish a sensitivity over a wide operating range.

A reliable implementation should survive changes of units. Convert millivolts to volts, and divide both the first coordinates and their standard deviation by one thousand. The whitened first column and the minimum score remain unchanged. The newly reported slope has units of kelvin per volt and is one thousand times the original physical number. If the slope fails this conversion test, inspect noise scaling, plotting aspect ratios and any normalisation by sample variance that was used in place of a known measurement uncertainty.

One may also ask for a confidence interval on the inferred line. That requires a sampling model for repeat experiments. In particular, treating each corrected pair as a known physical truth ignores uncertainty in the inferred settings. A simple exploratory test simulates new independent measurement errors at plausible fitted settings, refits the complete procedure and examines the spread of slopes over many trials. The result is conditional on those settings and the stated noise levels. A common calibration error shared by every specimen or correlated errors within a reading would require a different covariance model before interpreting such an interval.

A useful practical control is to interchange the two instruments, fit again with interchanged uncertainties, and convert the line to the original orientation. Whitened orthogonal regression should recover the same geometry. In contrast, two separate vertical regressions generally give different lines after their slopes are inverted. Directional regression can be an appropriate predictor for new observations under another sampling model; this reversal test specifically examines the symmetry of the latent-line calibration adopted here.

Open editable notebook Download .ipynb

Tutorial on calibration uncertainty and identifiability

1. With slope 2 and a positive vertical residual of 1 in the units of the hand example, change s_x to 0.25 and retain $s_y = 1$. What correction is assigned to each instrument?

▼ **Work through the solution**

The denominator becomes

$$1 + 4(0.25)^2 = 1.25.$$

The latent first coordinate increases by $2(0.25)^2/1.25 = 0.1$. The unexplained vertical error is

$$1/1.25 = 0.8.$$

Scaled corrections of 0.4 and 0.8 give a total cost of 0.80. Improving the horizontal meter shifts more of the inferred error to the vertical channel.

---

2. Multiply both standard deviations by three while retaining the same actual data. Which reported quantities change?

▼ Work through the solution

The centred whitened matrix is divided by three. Both singular values scale by one third, but the right singular directions and fitted physical line stay fixed. The squared score decreases by a factor of nine. A fixed fit can therefore be more or less consistent with an instrument specification even when its slope is unchanged.

---

3. Four measurements in whitened coordinates lie at $(1, 0), (-1, 0), (0, 1), (0, -1)$. What slope will a physically meaningful calibration report?

▼ Work through the solution

The scatter matrix is twice the identity, so every line through the origin has squared score two. A unique calibration slope cannot be recovered from these measurements under the stated equal-uncertainty model. To break the symmetry, obtain additional measurements spanning a wider range of true settings or supply new information about instrument uncertainties.

Further reading · NIST, orthogonal distance regression and errors in the explanatory variables

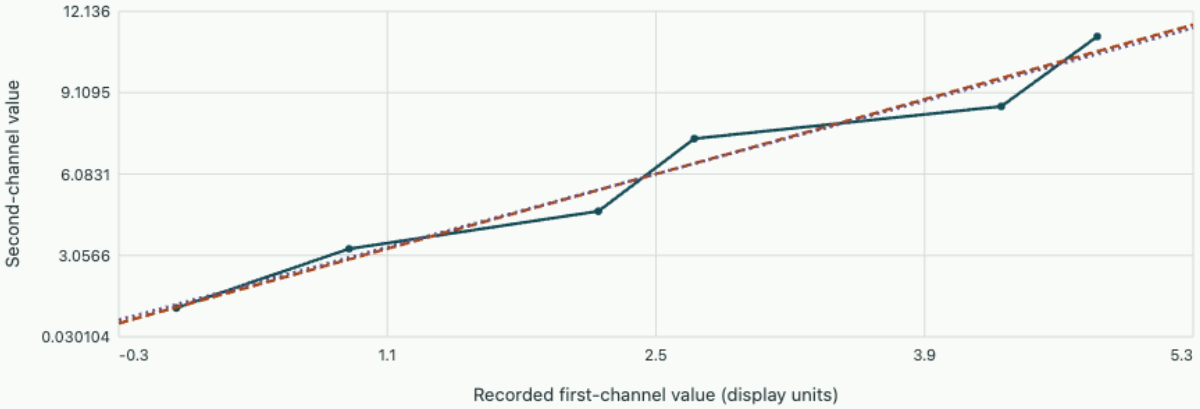
Change parameters · Calibrating two instruments when both axes fluctuate

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

First-instrument noise SD	**0.5**
Second-instrument noise SD	**1**
Horizontal units multiplier	**1**



Physical Deming slope 1.9833; displayed slope 1.9833. The noise metric is transformed with the horizontal units, preserving the physical fit.

Problem sheet 7 · Project · Which thermometer should the laboratory buy?

Separate an unknown common temperature, a calibration contrast and temporal drift. Design a measurement that improves the weakly determined quantity and check whether a small residual signals a trustworthy reconstruction.

Resolve two independent directions

Two calibrated outputs in orthogonal data coordinates are $d_1 = x_1 + \eta_1$, $d_2 = 0.02x_2 + \eta_2$. Both noise standard deviations are 0.01 °C. The observed outputs are 1.00 and 0.03. Compute the unregularised estimates and their standard deviations. Which parameter does the experiment measure reliably?

▼ **Work through the solution**

The estimates are 1 and 1.5. Independent noise standard deviations propagate by division by the sensitivities, giving 0.01 and 0.5. Both fitted data residuals are zero. Sensor fit alone has hidden the factor of fifty between parameter uncertainties.

Explore the related calculation with adjustable inputs. Change the parameters in the corresponding calculation and compare its independently recomputed values with the equation derived in this solution.

Separate an assumption from a measurement

With $\lambda = 0.02$ apply the zero-centred Tikhonov penalty. Calculate both fitted parameters, their predicted data and the residual. Compare with the regularisation path. Its three-channel test adds a middle sensitivity of 0.2. Set the weak-channel error to 0.010 for the same weak observation and compare the weak-channel prediction. The third sensor changes the total residual norm. Would another identical weak measurement halve the unregularised standard deviation?

▼ **Work through the solution**

The first estimate is $1/1.0004 \approx 0.9996$. The second gain is $0.02/(0.0004 + 0.0004) = 25$, hence $x_2 = 0.75$. The two predicted data are about 0.9996 and 0.015. Subtracting these from the observed data gives residuals 0.0004 and 0.015, with Euclidean norm about 0.0150. Two independent readings at the same design divide standard deviation by $\sqrt{2}$. A change of sensitivity from 0.02 to 0.04 would halve it directly.

Reuse the sensors over time

For a common temperature whose prior variance is 0.04 °C², a time step adds process variance 0.01 °C². A new independent reading has variance 0.04 °C² and lies 0.3 °C above prediction. Calculate the posterior mean shift and standard deviation. Use the temperature filter to predict what increasing process variance does to lag.

▼ **Work through the solution**

The predicted variance is 0.05. Gain is $0.05/(0.05+0.04) = 5/9$, hence the mean shifts by 0.1667 °C. The posterior variance is $0.05(4/9) = 0.02222$, giving standard deviation 0.1491 °C. Larger process variance gives a higher gain, tending to reduce lag while allowing more measurement variability into the estimate.

Audit the purchase decision

A cheaper replacement responds to x_2^2 with very low noise. Is it a good replacement when the sign of the calibration contrast decides whether to add or subtract a correction? Explore the signed measurement experiment.

▼ **Work through the solution**

A precise squared measurement alone still cannot distinguish equal positive and negative contrasts under a symmetric prior. A weak signed reading can split the two explanations. Evaluate loss in the corrected temperature and the cost of choosing the wrong sign before choosing hardware.

Problem sheet 8 · Estimate optical thickness when both measurements count photons

A dim red calibration source is viewed through an absorbing filter and then with the filter removed. Both observations fluctuate. Build the inverse problem directly from the photon counts, profile out the unknown source brightness, and decide how to distribute scarce observation time.

Choose the generative model before transforming the data

Take a linear counter with negligible dark counts and a stable monochromatic source. In a blank exposure of duration t_b the photon rate is an unknown λ; with the sample inserted the rate is λe^{-a}, where optical thickness $a = \mu L \geq 0$ is the desired parameter. Counts b and s in disjoint exposure windows are independent Poisson random variables. At only a few photons it is quite possible to record zero. Consider a glass sample of thickness 2.00 mm and ask which part of the inference requires this geometric information.

Profile out the source rate

Write the joint log likelihood at finite counts, then find the best source brightness for each proposed optical thickness.

▼ **Work through the solution**

Independence adds the log probabilities. Ignoring only terms independent of both unknown parameters gives

$$\ell(\lambda, a) = (b + s)\ln\lambda - sa - \lambda(t_b + t_s e^{-a}) + \text{constant}.$$

Holding a fixed, set $\partial_\lambda \ell = 0$ to obtain $\hat{\lambda}(a) = (b + s)/(t_b + t_s e^{-a})$ whenever the total count is positive. Substitution gives

$$\ell_p(a) = -sa - (b + s)\ln(t_b + t_s e^{-a}) + \text{constant}.$$

The specimen thickness L has yet to enter the calculation. One measures optical thickness directly; reporting an absorption coefficient $\mu = a/L$ additionally assumes a known geometrical length and an appropriate single-pass transmission model. Scattering out of the acceptance or reflections can contribute to apparent extinction if the apparatus admits them.

Explore the related calculation with adjustable inputs. Change the parameters in the experiment on this same project page and compare its output with this part of the derivation.

Check the result with an independent probability model

Condition on the observed total count

Show that the sample count given total $n = b + s$ is binomial. Derive the interior maximum and evaluate it for equal exposure times with $b = 40,\ s = 10$.

▼ **Work through the solution**

Conditional on a recorded photon belonging to either run, its probability of being in the sample run is $p(a) = t_s e^{-a}/(t_b + t_s e^{-a})$. The binomial probability of observing s such assignments among n is proportional to $p^s(1-p)^b$. Substitution reproduces the same profile likelihood as eliminating brightness analytically. The two routes make the disappearance of the source rate transparent. This conditioning removes its information about brightness while keeping the information about relative attenuation when brightness itself is free.

For positive interior counts the best fraction is s/n, giving $\hat{a} = \ln[(b/t_b)/(s/t_s)]$. With equal times it is $\ln 4 = 1.3863$, and the reported coefficient would be $693.1\ \mathrm{m}^{-1}$. If the sample rate happens to exceed the blank rate, the constrained maximum is $a = 0$. A negative unconstrained estimate is still informative as a check on fluctuations or alignment. Hiding it with a silent positivity correction makes repeated calibrations difficult to interpret.

Explore the related calculation with adjustable inputs. Change the parameters in the experiment on this same project page and compare its output with this part of the derivation.

Ask what the experiment establishes when no photon arrives

Contrast a curvature estimate with an exact one-sided statement

Evaluate the local standard error for the positive-count example. For a second equal-time experiment with $b = 8,\ s = 0$, obtain a one-sided conditional 95% lower bound on optical thickness.

▼ **Work through the solution**

The profile curvature at the interior maximum has magnitude $np(1-p) = bs/(b+s)$. Its reciprocal gives the local variance $1/b + 1/s$, so the standard error is $\sqrt{0.125} = 0.354$ for the first run. This local quadratic approximation is a high-count diagnostic. It is weak near a boundary and can miss asymmetry. A direct profile plot of twice the log likelihood loss reveals that asymmetry without taking the logarithm of an observed zero.

With zero sample counts and eight total, the MLE tends to infinity. Yet the data carry finite information. For a candidate p, the conditional probability of observing zero is $(1-p)^8$. Solve $(1-p_U)^8 = 0.05$ to obtain

$$p_U = 1 - 0.05^{1/8} = 0.3123.$$

At equal exposure $a = \ln[(1-p)/p]$, giving the exact one-sided conditional lower bound $a \geq 0.789$ after rounding. There is no finite upper bound from these counts alone. Adding one artificial count to numerator and denominator would manufacture a finite point estimate tied to that convention. A specified prior or a measured background can change the inference, but must be included in the generative model.

Explore the related calculation with adjustable inputs. Change the parameters in the experiment on this same project page and compare its output with this part of the derivation.

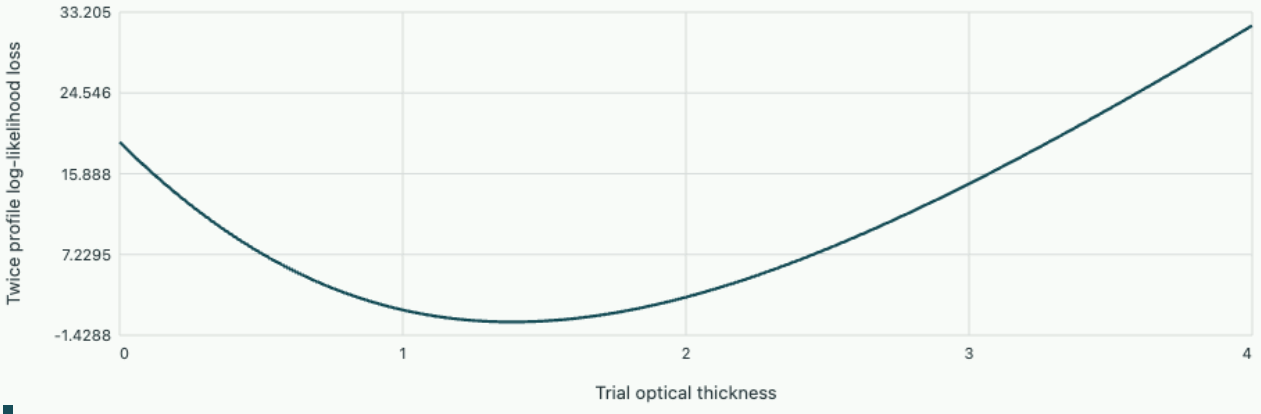
Allocate the two exposures and interpret the profile graph

For a total time $T = t_b + t_s$ and a provisional transmission $q = e^{-a}$, use expected counts to approximate $\operatorname{var}(\hat{a}) \simeq 1/(\lambda t_b) + 1/(\lambda q t_s)$. Optimisation requires a provisional attenuation. After a short pilot run its uncertainty can be propagated through a range of candidate designs. The graph below instead holds observed counts fixed and recomputes the profile loss over trial optical thickness. Use its time-ratio slider to correct the transmission for unequal known exposures. It is a reanalysis of those counts under the selected acquisition timing, rather than a simulated independent observation.

Open this experiment online

Try the calculation

Blank count	40
Sample count	10
Sample time / blank time	1



Conditional Poisson profile

Best optical thickness 1.386. Interior curvature standard error 0.354.

Find the best time ratio and distinguish design from estimation

If $q = 1/4$, find the variance-minimising fraction of time spent on the specimen and compare with the strategy of aiming for equal expected counts. How would a fluctuating source alter the protocol?

▼ **Work through the solution**

Substituting $t_b = T - t_s$ and differentiating gives $(T - t_s)^{-2} = 1/(qt_s^2)$, hence

$$t_s/t_b = 1/\sqrt{q} = 2.$$

Spend two thirds of the time on the specimen. The equal-count rule would require a ratio of four, overusing sample time while leaving the blank reference noisy. For $\lambda T = 120$, the optimum expected counts are 40 blank and 20 sample; the predicted variance is 0.075. At equal expected counts it becomes

$$2/24 = 0.0833.$$

The pilot estimate of transmission affects the best allocation but has no role in the profile formula once exposure times are fixed.

If source brightness varies systematically between the two runs, a lower sample rate can represent source drift rather than attenuation. Short interleaved blank and sample segments, or a simultaneously monitored reference, provide observations that identify this nuisance effect. More photons within a pair of separately drifting runs reduce counting error while leaving the confounding intact. At appreciable independently measured background, replace each Poisson mean by signal plus background and fit the counts jointly; the simple conditional probability above then also depends on the background rates.

Explore the related calculation with adjustable inputs. Change the parameters in the experiment on this same project page and compare its output with this part of the derivation.

Further reading

- University of Michigan · Jeffrey Fessler on exponential transmission models with Poisson counts

Notebook practice · Calibrating two instruments when both axes fluctuate

Inferring parameters and fields from observations

Follow information from a forward model back to its unknowns. Evaluate identifiability, numerical sensitivity and how an additional observation changes the inference.

1. ## Eliminate linear nuisance parameters before fitting a decay

   A measured relaxation has an unknown rate, amplitude and background. Derive a one-dimensional profile fit without fixing the nuisance parameters, explain why short observation windows can hide the rate and test an actual numerical dataset.

2. ## Choose a regularisation scale by predicting omitted data

   Regularised least squares can fit each observation partly with itself. Derive leave-one-out residuals from a single factorisation, approximate them by a trace and identify when a generalised cross-validation score misses a highly influential sensor.

3. ## Recovering sparse coefficients without hiding amplitude bias

   A sensor records a linear mixture of a few active components. Derive the soft-threshold map, use it in an iterative reconstruction and decide when correlated response columns make an apparently sharp sparse answer ambiguous.

4. ## What each new sensor reading changes in a Gaussian inverse problem

   Start with an uncertain vector, condition on one noisy linear observation, and reconstruct the posterior mean and covariance without repeating a batch inversion. Check shared calibration errors before treating repeated measurements as independent evidence.

5. ## What can unfinished endurance tests reveal about lifetime?

   When a test stops, a surviving bearing has supplied information. Derive the likelihood from first principles and examine how changing the observation deadline moves the rate inferred from a small batch.

6. ## How an integral instrument becomes a matrix

   Construct two discrete models of the same measurement, diagnose noise with a Picard plot and test a regularisation parameter against a deliberately imperfect instrument model.

7. ## Early stopping and penalties that leave part of a field free

   Derive two matrix-free least-squares iterations, find where noise first enters their solution and separate the measured and penalised directions of a nonstandard Tikhonov problem.

Inferring parameters and fields from observations · 1. Eliminate linear nuisance parameters before fitting a decay

A measured relaxation has an unknown rate, amplitude and background. Derive a one-dimensional profile fit without fixing the nuisance parameters, explain why short observation windows can hide the rate and test an actual numerical dataset.

Separate the unknowns by how they enter the observations

Suppose observations at known times t_i have mean $y_i = b + ae^{-kt_i}$ with $k > 0$. Both a and b enter linearly once k is specified. For independent equal-variance measurement noise, minimising squared residuals is a natural fit. More general known covariance can be incorporated by whitening the data and model columns together. Fitting all three unknowns by nonlinear iteration is possible, but every candidate rate should be judged against the best available amplitude and background at that rate. Otherwise an apparent improvement or failure of the rate may merely reflect a poor nuisance estimate.

Let $s_i = e^{-kt_i}$, bars denote sample means, and regress y_i on $1, s_i$. Differentiating the sum of squared residuals with respect to b gives

$$b = \bar{y} - a\bar{s}.$$

Substitution and differentiation with respect to a yield

$$a(k) = \frac{\sum_i (s_i - \bar{s})(y_i - \bar{y})}{\sum_i (s_i - \bar{s})^2},$$

$$b(k) = \bar{y} - a(k)\bar{s}.$$

The denominator vanishes when all times coincide or $k = 0$, at which point the two columns are identical. A QR or SVD of the whitened design reveals this rank loss. The simple centred formulas are transparent for analysis; for large poorly scaled datasets a stable factorisation remains preferable to unmonitored normal equations. Define profile objective $\Phi(k) = \frac{1}{2}\sum_i r_i^2$,

$$r_i = y_i - b(k) - a(k)s_i.$$

A one-dimensional search can now explore its shape, bounds or competing minima.

Differentiate the profiled objective without differentiating the whole fit

Write $X = [\mathbf{1}, s]$, $c = (b, a)^T$ and $r = y - Xc$. At the least-squares solution $X^T r = 0$. Taking the total derivative of $\Phi = r^T r/2$ gives $-r^T(X'c + Xc')$. The term with c' vanishes by orthogonality, leaving $\Phi' = -r^T X' c = a\sum_i r_i t_i e^{-kt_i}$. This simplification applies to the scalar objective gradient, not necessarily to the full residual Jacobian or second derivative. The projected component $(I - P_X)X'c$ is the local model change with rate that cannot be absorbed by a linear-nuisance adjustment. Its whitened squared norm characterises rate information near a well-fitting model.

At short times,

$$b + ae^{-kt} \simeq (a + b) - akt + ak^2t^2/2 + \cdots.$$

Data resolving only a constant and a slope constrain $a + b$ and ak, not a, b, k individually. Making the inner least-squares solve arbitrarily accurate cannot create curvature information absent from the acquisition. Include times approaching the decay time and a late observation near the baseline if feasible. Alternatively supply an independently characterised background with uncertainty rather than fixing it silently. A flat profile exposes this identifiability issue more clearly than a single successful three-parameter optimiser status.

Recover a rate exactly before introducing noise

At $t = 0, 1, 2, 3$ s consider the noiseless data $5, 3, 2, 1.5$ in sensor units. Successive differences are $-2, -1, -0.5$, a geometric sequence with ratio $q = e^{-k} = 1/2$. The rate is $k = \ln 2\,\mathrm{s}^{-1}$, and substituting the first two points gives

$$a = 4, b = 1.$$

At the correct rate the profile residual vanishes. An independent numerical implementation should reproduce both the rate and these coefficients and show $X^T r$ near zero for every rate, including incorrect candidate rates. These checks distinguish errors in profiling from failures of the outer search.

The difference ratio gives a poor general estimator when the later differences approach the measurement noise. Its numerator and denominator share observed samples, so their errors are correlated. Profiling the original observations uses all data without creating these ratio singularities and permits weighting by known variances. A late-time outlier can nonetheless bias a nonlinear fit with only a handful of samples. Examine the four individual residuals and compare predictions at a held-out time before interpreting a narrow numerical minimum as an identified physical mechanism.

Inspect the rate profile after refitting amplitude and background

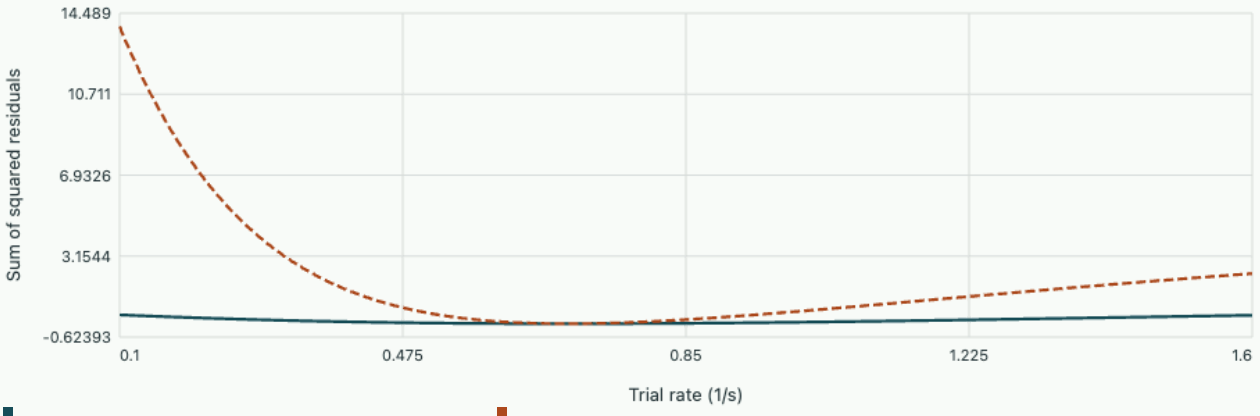
The calculation uses the four times above and adds a controlled perturbation to one late observation. It solves the two linear nuisance parameters afresh at each trial rate; a second curve evaluates the same rate with amplitude and background fixed to their original noiseless values. Comparing them makes visible how allowing nuisance parameters to adapt changes both the best rate and the apparent uncertainty. At zero perturbation both curves share $k = \ln 2$, but only the profiled curve describes the stipulated joint fit.

Open this experiment online

Try the calculation

Change in last observation **0**

Rate under inspection (1/s) **0.69**



At the inspected rate: amplitude 4.006, baseline 0.993, profiled residual sum 0.00000915. Profile minimum on the plot near 0.693 1/s.

What precision would the four measurements support?

Let the four observations have independent, equal standard deviation σ. A small change in rate has derivative vector $v_i = -a t_i e^{-k t_i}$. Refitting amplitude and background can mimic any change in the span of $s_i = e^{-k t_i}$ and the constant vector. Only the orthogonal component $v_\perp = (I - P_X)v$ distinguishes the rate. At the baseline $a = 4,\ k = \log 2$, the unprojected derivative is $(0, -2, -2, -1.5)$. Regressing it on s and a constant gives slope 1.98261 and intercept -2.30435. Subtract that fit and add the squared remainders to obtain

$$\|v_\perp\|^2 = 0.92174.$$

The local least-squares change in rate due to a noise vector is $v_\perp^T \epsilon / \|v_\perp\|^2$. Its standard deviation is therefore $\sigma / \|v_\perp\|$. For $\sigma = 0.10$ this is approximately $0.104\ \mathrm{s}^{-1}$. Tiny tolerances in the scalar optimiser cannot supply additional information. The estimate assumes small perturbations and an adequate decay law. A visibly asymmetric profile requires inspecting the actual objective on both sides before attaching symmetric error bars.

Tutorial · distinguish a narrow numerical minimum from identified data

Check a false rate numerically

For the noiseless observations take $q = e^{-k} = 1/4$. Compute the best linear background and amplitude approximately. Verify that the sum of residuals vanishes although the fit is wrong.

▼ **Work through the solution**

Here $s = (1, 0.25, 0.0625, 0.015625)$. Using the centred expressions gives $a \simeq 3.325,\ b \simeq 1.771$. The mean prediction equals the observed mean 2.875 by construction, forcing the residual sum to zero. The individual residuals are not all zero because their successive differences have ratio $1/2$ instead of $1/4$. Orthogonality to fitted columns certifies the conditional linear solve, not the outer rate.

Measure how the profile changes if the sampling interval doubles

Suppose the same four numerical readings are obtained at $0, 2, 4, 6$ s. Which parameters change, and which ratio argument still works?

▼ **Work through the solution**

The decay per sample remains $q = 0.5$, but now $q = e^{-2k}$, so

$$k = (\ln 2)/2 = 0.3466\,\mathrm{s}^{-1}.$$

Amplitude and baseline stay 4 and 1. Reporting q as a per-second rate without including time spacing would create a factor-of-two physical error even though every fitted residual vanishes.

Identify a structurally invisible rate

Take a preparation with true amplitude $a = 0$, independent equal-variance noise and an unknown baseline. Can any choice of times identify k from the exact noiseless data?

▼ **Work through the solution**

For every $k > 0$ the inner fit permits $a = 0, b = \bar{y}$, giving zero residual for constant data. The local rate derivative $X'c$ vanishes and the profile carries no rate information. Random noise may select a numerical minimum, but it does not turn the true no-decay signal into an identified relaxation rate. The experiment needs a preparation with measurable nonzero excitation or a different observable sensitive to the rate.

Further reading

- O’Leary and Rust · projection derivatives, implementation and conditioning
- Golub and Pereyra · separable nonlinear least squares and variable projection
- Course investigation · profiles and predictive uncertainty
- Test which parameters a temperature series identifies

Inferring parameters and fields from observations · 2. Choose a regularisation scale by predicting omitted data

Regularised least squares can fit each observation partly with itself. Derive leave-one-out residuals from a single factorisation, approximate them by a trace and identify when a generalised cross-validation score misses a highly influential sensor.

Express a regularised fit as a map on measured data

Let whitened observations $y \in \mathbb{R}^n$ be modelled by Ax, and for $\lambda > 0$ minimise $\|Ax - y\|^2 + \lambda\|x\|^2$. The minimiser satisfies

$$(A^T A + \lambda I)x = A^T y.$$

Predicted data are $\hat{y} = H_\lambda y$, where $H_\lambda = A(A^T A + \lambda I)^{-1} A^T$. This "hat" matrix maps the observations to their own fitted values. If $A = U\Sigma V^T$, its nonzero eigenvalues are

$$h_j = \sigma_j^2/(\sigma_j^2 + \lambda).$$

Their sum $d(\lambda) = \operatorname{tr} H_\lambda$ is an effective degrees-of-freedom count. A weakly determined direction with small singular value can contribute a fraction of an independently fitted parameter.

The norm of the training residual decreases or stays constant as the penalty weakens. For a square full-rank model it tends to zero as the penalty tends to zero, even with noisy data. Individual residuals can change in either direction. It does not measure out-of-sample error. Instead ask how well the fitted model predicts each observation when it was left out. Repeating the full solve n times conceptually works but wastes structure. An exact deletion identity gives each left-out residual $r_i^{(-i)} = (y_i - \hat{y}_i)/(1 - H_{ii})$, where H_{ii} is the observation's leverage. To derive the formula, remove row a_i^T, subtracting $a_i a_i^T$ from the penalised normal matrix. Multiplying its inverse update by the original residual along a_i gives the factor $1/(1 - a_i^T G^{-1} a_i)$, with $G = A^T A + \lambda I$. This derivation assumes the same penalty strength in the objective for the held-out and full fits.

Replace individual influences by their average only once

The leave-one-out score is the mean of $[r_i/(1 - H_{ii})]^2$. Generalised cross-validation replaces the individual diagonal elements by the common mean d/n, giving

$$\mathrm{GCV}(\lambda) = \frac{\|y - \hat{y}\|^2/n}{(1 - d(\lambda)/n)^2}.$$

When every leverage is equal, this approximation reproduces leave-one-out exactly. Its expression is invariant to orthogonal transformations of the whitened data space because both trace and residual norm are invariant. Exact leave-one-observation-out instead depends on which physical observations are removed, which matters for a particular instrument arrangement. GCV is useful when a rotationally averaged prediction criterion is appropriate; it is not a universal interpretation of removing an actual sensor or predicting an unmeasured parameter direction.

Consider $A = (1, 1, 1)^T$ fitting a scalar common signal, $y = (1.1, 0.9, 1.0)$, and $\lambda = 1$. The fit is $\hat{x} = 3/4 = 0.75$, so residuals are $0.35, 0.15, 0.25$. All leverages are $1/4$. Exact prediction errors on deleting each datum are residuals divided by 0.75. For example removing 1.1 leaves $(0.9 + 1.0)/(2 + 1) = 0.6333$, giving 0.4667 error, equal to $0.35/0.75$. GCV equals exact leave-one-out here. Clear agreement in this symmetric scalar case should be a first numerical verification before interpreting a higher-dimensional score whose influences vary.

Background · What counts as an independent observation when validating?

Suppose each specimen is read twice on the same instrument, and its calibration offset changes only between specimens. Omitting one reading leaves its twin available to the fit. A low single-reading validation error then describes another reading of an already observed specimen. It can be optimistic for the next specimen, whose offset is unknown.

For an omitted group G, the scalar denominator $1 - h_{ii}$ from leave-one-out becomes a small matrix. With fixed regularisation, unchanged data weights and squared loss, the group prediction residual satisfies

$$(I - H_{GG})\, r_G^{(-G)} = r_G.$$

The off-diagonal entries account for the influence that the two readings exert on each other. If removing an entire group leaves a poorly constrained direction, the solve amplifies that direction. Match the omitted group to the intended future experiment before choosing a score.

See how one sensitive observation can dominate a fit

For another scalar design $A = (1, 1, 4)^T$ with $\lambda = 2$, $G = 20$ and $H_{ii} = (0.05, 0.05, 0.80)$. A residual on the third sensor is multiplied by five in a true deletion while the first two are amplified only slightly. GCV uses the same factor $1/(1 - 0.9/3) = 1.429$ for all. If the model discrepancy is concentrated in that high-leverage reading, trace averaging can underestimate the consequence of losing it. Conversely if errors fall only on low-leverage sensors it can overstate their delete-one loss. Inspecting the leverage distribution is inexpensive compared with attributing precision to a single scalar tuning criterion.

Whitening rotates and rescales observations when errors are correlated. Leaving out one row after a dense whitening transformation is generally not the same experiment as removing one physical sensor before whitening. A time-correlated series may demand holding out an entire time block so training data cannot predict a held-out noise realisation through short-range correlation. Predictive selection should specify whether the intended prediction concerns a new repeat at known conditions, a missing detector or a new range of the physical parameters. A penalty that works well for a data prediction need not minimise reconstruction error in a weakly observable component of x.

When does using the average leverage change the answer?

Four sensors have sensitivity one. Their readings are 1.10, 0.90, 1.05 and 0.95. The fifth reading equals its sensitivity plus an adjustable offset. First give all five equal sensitivity. The two scores coincide even with a nonzero offset. Now set the fifth sensitivity to 5 and its offset to 0.8. At λ = 1, compare the fifth leverage with the mean leverage. Predict the fifth reading from the other four, then use this single omitted-sensor error to check which score can reach 0.66.

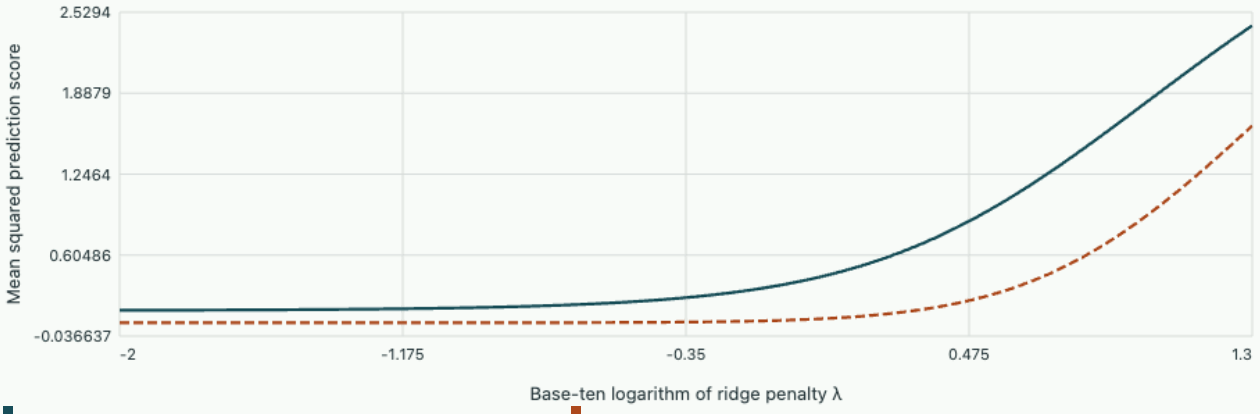
The OneSimulation model refits five separate reduced data sets for every penalty. The first CSV score therefore measures actual omitted-sensor predictions. Compare it point by point with the faster identity used in the browser.

Open this experiment online

Try the calculation

Sensitivity of the fifth sensor **3**

Fifth reading offset **0.8**



At λ = 1, the fifth sensor has leverage 0.64286. The leave-one-out and generalised scores are 0.40708 and 0.095014.

Download the OneSimulation model

In your OneSimulation environment, run the model with these inputs.

```
python networks-and-calibration.py validation --sensitivity 3 --offset 0.8
```

Tutorial · recover two kinds of influence from one factorisation

Compute the symmetric-case score

Using the common-signal example, evaluate GCV and verify it by explicit deletion of all three observations.

▼ **Work through the solution**

The residual squared sum is

$$0.35^2 + 0.15^2 + 0.25^2 = 0.2075.$$

Its mean divided by 0.75^2 gives 0.12296. The three exact predictive residuals are $0.4667, 0.2000, 0.3333$; squaring and averaging reproduces 0.12296. The unit of this score is squared whitened observation units. Taking its square root gives an rms predicted datum error, not a standard deviation of the fitted underlying parameter.

Recover effective dimension from singular values

A whitened model has singular values $4, 1, 0.1$ and six readings. For $\lambda = 1$, calculate d and the GCV multiplier applied to mean square residual.

▼ **Work through the solution**

The gains are

$$16/17 = 0.9412,\ 1/2,\ 0.01/1.01 = 0.00990.$$

Their sum is 1.4511 and $1/[1 - d/6]^2 \simeq 1.740$. The weak singular direction barely enters the prediction even though the algebraic rank is three. With vanishing penalty all three nonzero gains tend to one; with very large penalty they approach zero. This continuous change is why a count of nonzero columns cannot replace the trace in the score.

Diagnose misleading use of the diagonal basis

For $A = I$ of size n and the same ridge penalty, compute GCV for every positive λ. Can these data identify a preferred penalty from that score?

▼ **Work through the solution**

Here $H = I/(1 + \lambda)$,

$$r = \lambda y/(1 + \lambda),$$
$$1 - d/n = \lambda/(1 + \lambda).$$

Cancellation gives $\mathrm{GCV} = \|y\|^2/n$, independent of λ. Leaving one coordinate out leaves no data at all about that independent parameter, so the prediction by the penalised system for that coordinate is always zero. A numerical minimum created by roundoff does not select a meaningful smoothing strength. Structural correlations, replication, a noise model or external validation are needed to decide how much shrinkage helps the intended task.

Further reading

- Golub, Heath and Wahba · the original generalised cross-validation derivation
- Course investigation · regularisation with singular data
- Test which parameters a temperature series identifies

Inferring parameters and fields from observations · 3. Recovering sparse coefficients without hiding amplitude bias

A sensor records a linear mixture of a few active components. Derive the soft-threshold map, use it in an iterative reconstruction and decide when correlated response columns make an apparently sharp sparse answer ambiguous.

Derive a nonsmooth optimum before choosing a solver

Let a whitened linear model be $y = Ax + \eta$ and consider minimising $F(x) = \|Ax - y\|^2/2 + \lambda \sum_j |x_j|$. The parameter $\lambda \geq 0$ balances fit with sparsity in this chosen coordinate system. Rescaling one column of A while leaving its penalty unchanged changes how expensive the corresponding physical component is; normalise columns or use documented coefficient-specific weights before comparing support selection across different quantities. Sparsity of a signal in one basis is a modelling assumption, not evidence generated by the penalty itself.

For an orthonormal design $A^T A = I$, expanding the square shows that minimisation separates into scalar terms $(x_j - z_j)^2/2 + \lambda |x_j|$, with $z = A^T y$. For a positive solution the derivative is $x - z + \lambda = 0$; for a negative one it is $x - z - \lambda = 0$. At zero the absolute-value subgradient is the interval $[-1, 1]$, so zero is optimal when $|z| \leq \lambda$. These cases give

$$S_\lambda(z) = \operatorname{sign}(z) \max(|z| - \lambda, 0).$$

Unlike simply removing small coefficients, soft thresholding also shrinks every retained amplitude by λ. This predictable bias can matter even if the selected support is physically correct.

For $z = (2.4, -0.7, 0.18)$ and $\lambda = 0.5$, the solution is $(1.9, -0.2, 0)$. The residual correlations $A^T(y - Ax)$ must equal λ times the sign on active coordinates and stay in $[-\lambda, \lambda]$ on inactive coordinates. Here they are $(0.5, -0.5, 0.18)$. These conditions independently certify optimality for the convex objective, without proving the true sample actually contains only two active mechanisms.

Majorise the smooth loss and solve its local model exactly

For a general A, write $f(x) = \|Ax - y\|^2/2$ and $\nabla f = A^T(Ax - y)$. Taylor expansion is exact with second-order term $\|A(z - x)\|^2/2 \leq L\|z - x\|^2/2$, $L = \|A\|_2^2$. Replace the true quadratic term by this upper bound, keep the penalty exact and minimise over z. Completing the square produces

$$x^{k+1} = S_{\lambda/L}\left(x^k - \frac{1}{L} A^T(Ax^k - y)\right)$$

componentwise. This proximal-gradient or ISTA step decreases the objective for this step bound. An under-estimated L can break the majorisation. A backtracking check evaluates the true smooth loss at the candidate instead of requiring an exact power-method estimate of the largest singular value. Acceleration can reduce the number of iterations, while a numerical termination test based on the subgradient or objective gap remains necessary.

Take $A = \left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right), y = (1, 0)^T, \lambda = 0.2$. Its spectral $L = (3 + \sqrt{5})/2 = 2.618$. Starting at zero, $A^T y = (1, 1)$, so the first iterate is

$$(0.8/L, 0.8/L) \simeq (0.3056, 0.3056).$$

Both components initially enter equally even though the first column alone fits the pattern exactly. The second measurement penalises a nonzero second coefficient. At the optimum $(0.8, 0)$, residual correlations are $(0.2, 0.2)$, so the inactive second coefficient sits exactly on its subgradient boundary. A small data perturbation can change its selected status, despite a well-defined convex optimum for this full-rank matrix. Stability of objective value and stability of support are different questions.

Interpret support and debias only after validating it

If two columns of A are identical, two nonnegative amplitudes with the same sum give the same predictions and the same ℓ^1 penalty. A solver may split an identifiable total between the columns or report one arbitrary active component. Lowering the residual tolerance will not resolve this ambiguity. Acquire a measurement where the columns respond differently, merge them if only the total is scientifically meaningful or report a range of admissible allocations. Near-collinear columns give an approximate version of this effect that can be revealed by perturbing the data within their measured noise or holding out a sensor condition.

A common second step is to refit by ordinary or weighted least squares on the selected support to remove shrinkage bias. For the orthonormal example it would restore the retained amplitudes to $2.4, -0.7$. This is a conditional estimate after choosing that support; it does not undo selection effects and can amplify noise strongly for a nearly singular active submatrix. Test the selected support and refitted prediction on held-out observations. An uncertainty interval that assumes the support known in advance can be too optimistic when small perturbations switch the active set.

Can support and amplitudes both be correct?

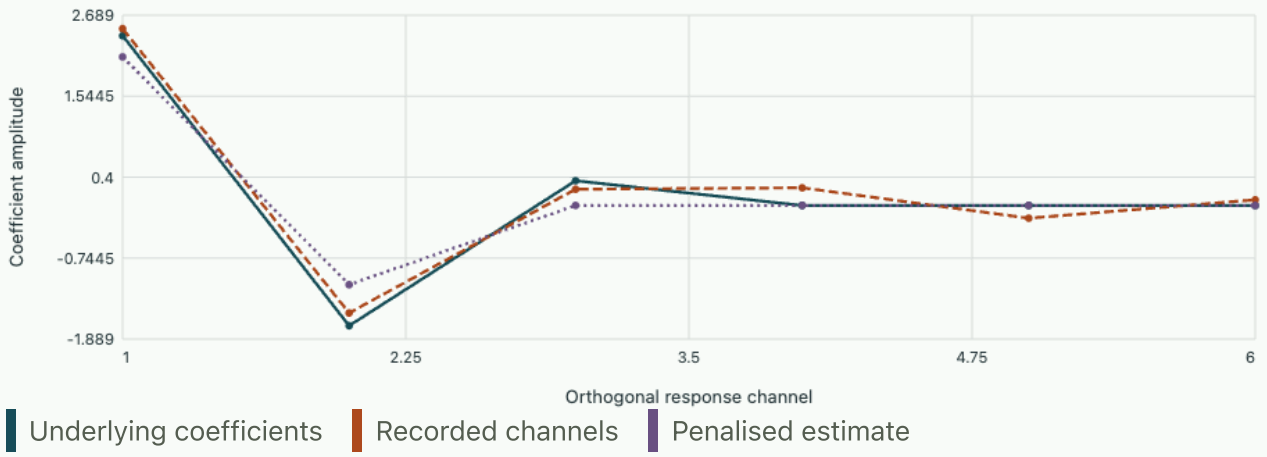
The six orthogonal channels have three true nonzero coefficients. Set the perturbation multiplier to one and start with zero threshold. Increase the threshold just far enough to remove the three noise-only channels. Does the weak third signal survive? Now set the perturbation to zero while keeping that threshold fixed. Compare the third estimate with its true value to measure the shrinkage introduced by the penalty itself.

Open this experiment online

Try the calculation

Penalty λ **0.4**

Multiplier for measurement perturbations **1**



Recovered 2 nonzero coefficients of six. Euclidean reconstruction error 0.74088.

Tutorial · verify an inverse solution using subgradients

Find the penalty where everything switches off

For a general A, y, derive the smallest λ for which $x = 0$ is optimal. Evaluate for the small matrix above.

▼ Work through the solution

At zero, stationarity requires $A^T y \in \lambda[-1, 1]^p$. Thus $\lambda_{\max} = \|A^T y\|_\infty$, which equals 1 in the example. At any larger penalty no component is supported by its residual correlation relative to the threshold. At exact equality an inactive coordinate can lie on the boundary. The criterion concerns the scale of whitened data and normalised design and changes when units or weights change.

Compare soft thresholding with hard component deletion

For a scalar reading 1.2 and $\lambda = 0.5$, compare objectives at the soft estimate, raw reading and zero. Which is lowest and what does it show?

▼ Work through the solution

The soft estimate is 0.7, with objective $0.5(0.5)^2 + 0.5(0.7) = 0.475$. At 1.2 the residual vanishes but penalty yields 0.600; at zero the loss is 0.720. The shrinkage is a deliberate consequence of the stated objective rather than an error of an iterative algorithm. If amplitude preservation is important, a separate support-conditioned refit answers a different optimisation question.

Prove nonidentifiability with two equal columns

Take $A = (1, 1)$, $y = 3$ and $\lambda = 1$. Characterise an entire family of optimal nonnegative solutions and the unique total that data and penalty identify.

▼ **Work through the solution**

On nonnegative inputs the objective depends only on $s = x_1 + x_2$ as $(s - 3)^2/2 + s$. Its minimum is at $s = 2$. The triangle inequality shows no signed decomposition of that same sum can lower its penalty. Every pair $(q, 2 - q)$ for $0 \le q \le 2$ is therefore globally optimal. An algorithm returning $(2, 0)$ has not established that the first physical species is present and the second absent. Only their aggregate is determined without a distinguishing observation or another justified prior.

Further reading

- Beck and Teboulle · iterative shrinkage-thresholding for linear inverse problems
- Course investigation · inverse problems with nonsmooth constraints
- Test which parameters a temperature series identifies

Inferring parameters and fields from observations · 4. What each new sensor reading changes in a Gaussian inverse problem

Start with an uncertain vector, condition on one noisy linear observation, and reconstruct the posterior mean and covariance without repeating a batch inversion. Check shared calibration errors before treating repeated measurements as independent evidence.

Condition a joint Gaussian by eliminating a correlated component

Let x have a Gaussian prior with mean m and positive definite covariance P. A scalar instrument returns $y = h^T x + \epsilon$, with zero-mean Gaussian noise independent of x of variance $r > 0$. The predicted observation mean is $h^T m$; the innovation $d = y - h^T m$ has variance $S = h^T P h + r$, and cross covariance of x with this innovation is Ph. Seek a corrected error $(x - m) - Kd$ uncorrelated with d. Its covariance with d vanishes when $K = Ph/S$. Because the joint distribution is Gaussian, uncorrelated then implies independent, so conditioning on the observed d gives

$$m^+ = m + Kd,$$

$$P^+ = P - \frac{Phh^T P}{S}.$$

The covariance correction is rank one and positive semidefinite when subtracted. It depends on geometry and declared noise, not the particular measurement value. The new mean can move strongly when a large innovation occurs even if the predicted covariance reduction was modest. A very unlikely innovation is reason to check the physical model and reported uncertainty before proceeding as though the calibration were known. Recomputing an ad hoc smaller measurement variance after seeing the data would change the probabilistic assumptions.

An equivalent route minimises $(x - m)^T P^{-1} (x - m) + (y - h^T x)^2 / r$, yielding posterior precision $P^{-1} + hh^T / r$. Multiplying by the proposed covariance above verifies its inverse directly, a one-row Woodbury identity. For several independent readings, applying rank-one updates sequentially produces the same posterior as one batch solve in exact arithmetic. If the instrument errors are correlated, either whiten the full observation model or explicitly include shared error components in the state before applying independent updates. The matrix identity is exact only for the covariance actually specified in the statistical model.

A correlated prior transfers information to an unmeasured coordinate

Set $m = (0, 0)^T$, $P = \left(\begin{smallmatrix} 4 & 1 \\ 1 & 1 \end{smallmatrix}\right)$. Measure only the first coordinate with $h = (1, 0)^T,\ r = 1,\ y = 3$. The innovation variance is 5 and the gain is $(0.8, 0.2)^T$. Hence

$$m^+ = (2.4, 0.6),$$

$$P^+ = \left(\begin{smallmatrix} 0.8 & 0.2 \\ 0.2 & 0.8 \end{smallmatrix}\right).$$

The second coordinate moved even though the instrument does not directly sense it. That inference uses prior correlation. If prior off-diagonal covariance were zero, its posterior mean and variance would remain unchanged. A conclusion about an unobserved material parameter should therefore report whether sensitivity comes from data or from that imposed correlation structure.

The innovation itself is $3/\sqrt{5} = 1.34$ predicted standard deviations, plausible under the original model. Conditioned residual at the same reading becomes $3 - 2.4 = 0.6$, much smaller because that observation helped fit the estimate. A model check should use the pre-update innovation or an independent held-out observation. In a sequence of independent conditionally generated data, correctly normalised innovations provide a useful diagnostic for error scale and neglected correlation. An extremely small training residual after every assimilation can simply reflect a model that overweights its own observations.

Maintain a valid covariance over many updates

In exact arithmetic $P^{+} = P - Kh^{T}P$. In finite precision subtracting two nearly equal matrices can erode symmetry or produce a small negative eigenvalue in a direction known very precisely. The equivalent Joseph form $P^{+} = (I - Kh^{T})P(I - Kh^{T})^{T} + KrK^{T}$ expresses the result as a sum of positive semidefinite pieces. It also derives directly from the prior state error and independent measurement noise entering the update. A square-root covariance or information factor can improve numerical robustness over repeated strongly informative observations. Simply setting negative diagonal entries to zero after a bad update may still leave an indefinite correlated matrix.

If the state evolves between observations with $x_{\text{next}} = Fx + w$, independent process noise covariance Q gives predicted mean Fm^{+} and covariance $FP^{+}F^{T} + Q$. Choosing $Q = 0$ asserts exact dynamics; it is not a harmless way to remove an implementation parameter. Repeated observations can then drive covariance toward zero in a falsely rigid physical model. Conversely arbitrarily increasing process noise to make any data appear plausible sacrifices the predictive test. Estimate or validate process variability independently when deciding whether a sequential inverse problem represents one stationary parameter or an evolving hidden state.

One hundred readings with one shared calibration

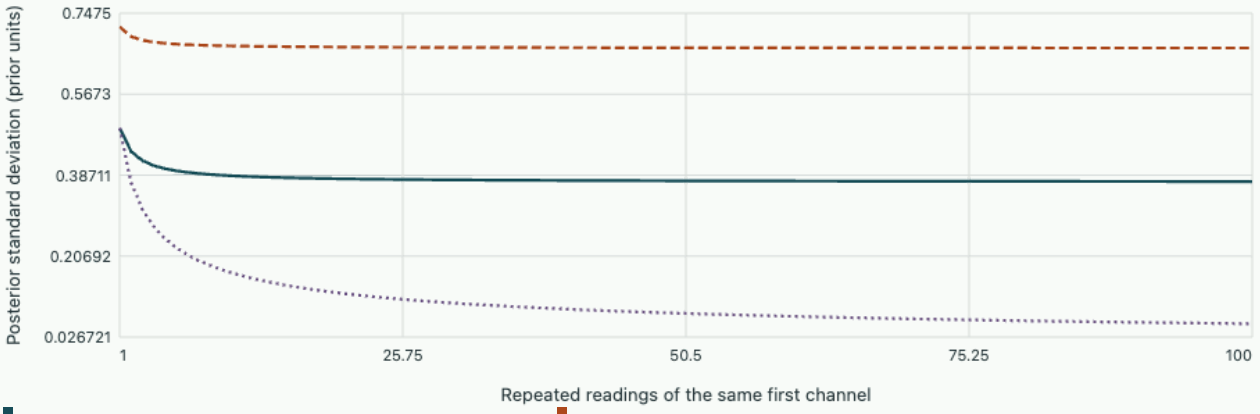
Two parameters have unit prior variance and adjustable correlation. Observe the first parameter repeatedly. Each reading has independent Gaussian noise of standard deviation 0.4 plus the same Gaussian calibration offset, drawn independently of the parameters and the individual noise. First set the calibration uncertainty to zero. Restore it and compare the limiting widths. Finally set the prior correlation to zero and decide which reading could possibly update the second parameter.

Open this experiment online

Try the calculation

Prior parameter correlation **0.8**

Shared calibration standard deviation **0.4**



Observed parameter, shared calibration

Unobserved parameter, shared calibration

Observed parameter, independent-noise assumption

After 100 observations the true posterior widths are 0.37299 and 0.6701. Single-reading noise standard deviation 0.56569.

Tutorial · separate new information from repeated calibration

Assimilate an independent second measurement

Continue from the two-coordinate example and measure the second coordinate with $h = (0, 1)^T$, $r = 0.2$, $y = 1.1$. Compute the updated mean and covariance.

▼ Work through the solution

The predicted reading is 0.6, innovation 0.5 and $S = 0.8 + 0.2 = 1$. The gain is $(0.2, 0.8)^T$, giving mean $(2.5, 1.0)$. Subtract $\left(\begin{smallmatrix} 0.04 & 0.16 \\ 0.16 & 0.64 \end{smallmatrix}\right)$ from the preceding covariance, giving $\left(\begin{smallmatrix} 0.76 & 0.04 \\ 0.04 & 0.16 \end{smallmatrix}\right)$. Its determinant is 0.12 and both diagonal entries are positive, consistent with a positive definite remaining uncertainty.

Quantify a shared calibration floor

A scalar prior has variance 4. Two or more readings each have independent noise variance 1 and share the same unknown additive calibration bias of variance 0.25. Find posterior variance of the scalar from n readings and its limit.

▼ Work through the solution

The sample mean measures the scalar plus shared bias with variance $0.25 + 1/n$; independent deviations of readings from their mean contain no information about the scalar in this Gaussian model. Thus posterior variance is $[1/4 + 1/(0.25 + 1/n)]^{-1}$, tending to $1/4.25 = 0.2353$. Incorrectly assigning all errors independent would make the posterior variance vanish. An independent reference sensitive to the calibration bias can lower this floor.

Choose a sensor for a requested derived quantity

Two candidate sensors of the same noise cost are available. Derive the covariance reduction in $q^T x$ and evaluate for $q = (0, 1)$, the original P, $r = 1$ and direct measurements of either coordinate.

▼ **Work through the solution**

The variance reduction is $(q^T P h)^2/(h^T P h + r)$. Measuring coordinate 1 lowers the desired variance by $1/5 = 0.2$; measuring coordinate 2 lowers it by $1/2 = 0.5$. The larger variance of coordinate 1 could make it an attractive sensor target under a different design criterion, yet the specified prediction favours direct measurement of coordinate 2. State the scientific quantity whose uncertainty the design should reduce.

Further reading

- Kalman · A New Approach to Linear Filtering and Prediction Problems
- Course investigation · Gaussian priors and posterior uncertainty
- Test which parameters a temperature series identifies

Inferring parameters and fields from observations · 5. What can unfinished endurance tests reveal about lifetime?

When a test stops, a surviving bearing has supplied information. Derive the likelihood from first principles and examine how changing the observation deadline moves the rate inferred from a small batch.

Count failures and add the time spent on test

Assume identical bearings fail independently at a constant rate λ per hour. Over a short interval dt, the conditional probability of failure for an operating bearing is $\lambda\,dt + o(dt)$. This gives the differential equation $S'(t) = -\lambda S(t)$ with $S(0) = 1$ for the fraction surviving beyond t. Its solution is $S(t) = e^{-\lambda t}$ and the failure-time density is

$$f(t) = \lambda e^{-\lambda t}.$$

Choose a fixed administrative deadline c in advance. Each observed duration is the earlier of the failure time and c. Write $\delta_i = 1$ for a recorded failure, zero for an unfinished run. A failure contributes its density. An unfinished run contributes the probability of lasting at least its recorded duration. Multiplying these quantities over the bearings gives

$$L(\lambda) = \prod_i f(t_i)^{\delta_i} S(t_i)^{1-\delta_i} = \lambda^d e^{-\lambda T},$$

$$d = \sum_i \delta_i,$$

$$T = \sum_i t_i.$$

For positive d, differentiating $d \log \lambda - \lambda T$ gives a score $d/\lambda - T$. It vanishes at $\hat{\lambda} = d/T$. The curvature $-d/\lambda^2$ is negative. With zero failures the likelihood falls immediately to the right of zero, making zero the boundary maximum. A survivor contributes no failure to d and still adds positive exposure to T. Removal triggered by a bearing's imminent failure would require a model for that mechanism. A predeclared deadline avoids this selection issue.

Audit a batch stopped after seven hours

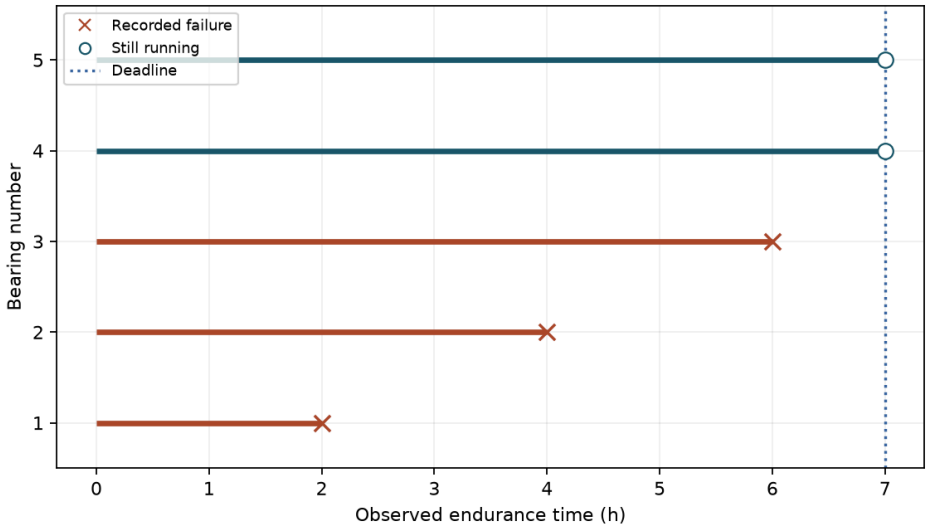
Suppose the simulated complete lifetimes are 2, 4, 6, 9 and 14 hours. At seven hours the actual records are three failures and two unfinished tests. Their total exposure is $2 + 4 + 6 + 7 + 7 = 26$ bearing-hours. The rate estimate is $3/26 = 0.11538$ per hour and the fitted mean lifetime is $26/3 = 8.667$ hours. Discarding the two survivors leaves 12 bearing-hours and inflates the apparent rate to $3/12 = 0.25$ per hour.

For a candidate rate of 0.20 per hour the log likelihood relative to the full-data maximum is

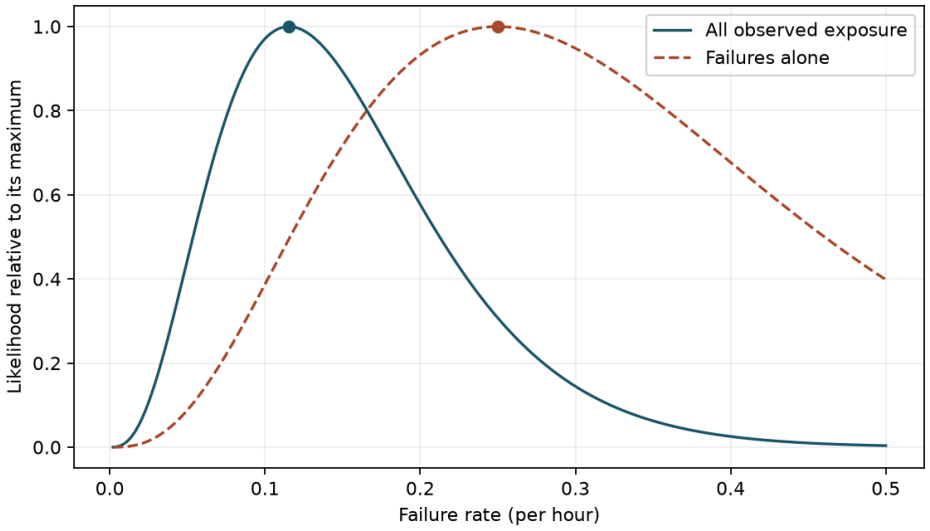
$$3 \log(0.20/(3/26)) - 26(0.20 - 3/26) \approx -0.550.$$

This is a relative likelihood near 0.58. The absolute density changes when time units change, while this likelihood ratio is invariant. In the notebook change the deadline to five hours. Each failure that now

occurs after the deadline becomes an unfinished observation. Inspect the time lines before interpreting the new estimate.



Every horizontal length enters the denominator of the estimate. Only the crosses add to the failure count in its numerator.



Each curve is divided by its own maximum. Moving the deadline to before the first failure makes the complete-observation likelihood attain its maximum at a rate of zero.

Exercise · change the stopping time

What does a five-hour deadline yield for the same five bearings? Can extending the deadline without any additional failure increase the estimated rate?

▼ Work through the solution

The durations are 2, 4, 5, 5 and 5 hours with two observed failures. The new estimate is $2/21 = 0.09524$ per hour. While additional operating time accumulates with the same failure count, the ratio d/T falls for positive d. The estimate can jump upward at an actual failure event. A generally increasing or decreasing trend with a moving deadline is therefore not guaranteed.

Exercise · infer an upper rate after zero failures

Five fresh bearings all survive ten hours. Find the rate for which observing no failures would have probability 0.05 under the model.

▼ **Work through the solution**

All five must survive ten hours. Independence gives probability $e^{-50\lambda}$. Setting this equal to 0.05 gives $\lambda = -\log(0.05)/50 \approx 0.05991$ per hour. Inverting this zero-failure event gives a one-sided 95% upper confidence bound for the rate in this experiment. The likelihood maximum itself remains at zero.

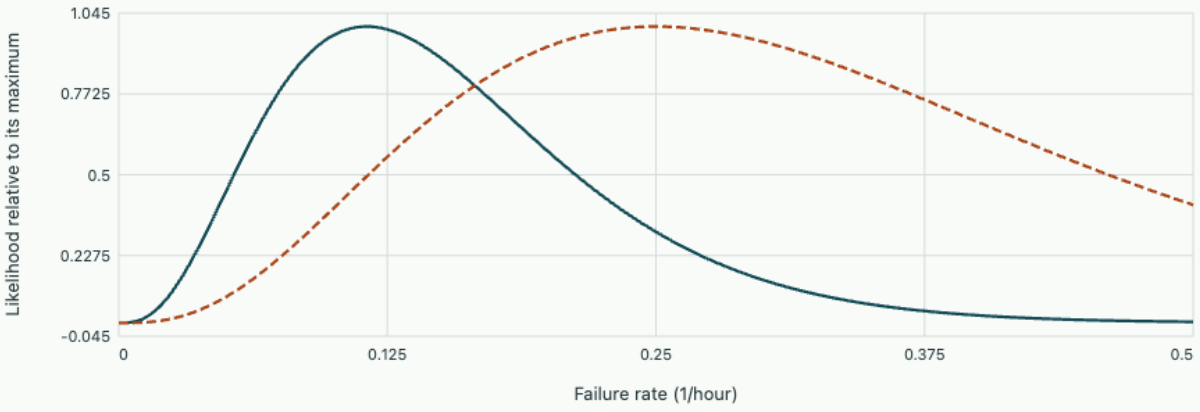
Change parameters · Explore this notebook model

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

Test deadline (hours) 7



All observed exposure Discard unfinished tests

3 failures in 26 bearing-hours; rate estimate 0.11538 per hour, mean lifetime 8.6667 hours.

Open editable notebook Download .ipynb

Inferring parameters and fields from observations · 6. How an integral instrument becomes a matrix

Construct two discrete models of the same measurement, diagnose noise with a Picard plot and test a regularisation parameter against a deliberately imperfect instrument model.

Derive the matrix twice from the same kernel

Suppose an instrument measures $b(s) = \int_0^1 K(s,t)x(t)\,dt + e(s)$. With square-integrable kernel this forward operator is compact on L^2. Small singular values suppress some source directions. For a midpoint rule with source nodes t_j and weights w_j, collocation at receiver nodes s_i gives $b_i \simeq \sum_j K(s_i,t_j)w_j x(t_j)$. To compare spectra over nonuniform meshes, use the variables $\sqrt{w_j}x(t_j)$ and data $\sqrt{v_i}b(s_i)$. The Euclidean matrix is $A_{ij} = \sqrt{v_i}K(s_i,t_j)\sqrt{w_j}$. Noise whitening is an additional operation using the measurement covariance.

Projection instead represents $x = \sum_j c_j\phi_j$ and tests against ψ_i. For orthonormal bases the matrix and data are

$$G_{ij} = \int_0^1 \int_0^1 \psi_i(s)K(s,t)\phi_j(t)\,dt\,ds,$$

$$d_i = \int_0^1 \psi_i(s)b(s)\,ds.$$

For cell indicators normalised by the square root of cell length this double integral averages the response over both the detector and the source footprint. For nonorthonormal bases use mass matrices to measure physical norms. Omitting these metrics changes the meaning of singular values as the mesh changes.

Take $K = 1 + st$, two equal cells and coincident midpoint receivers. The weighted midpoint and cell-average projection agree because the kernel is affine in each argument. The two matrices equal one half times the matrix with rows $(17/16, 19/16), (19/16, 25/16)$. Their determinant is $1/16$. Changing to $K = 1 + s^2t^2$ gives midpoint squares $1/16, 9/16$ while their cell averages are $1/12, 7/12$. The resulting spectra differ even with exact arithmetic. Refining both grids and inspecting the spectrum that stays above the data-noise floor distinguishes discretisation from information.

Read data and singular values together

For $A = U\Sigma V^T$, the noise-free continuum solution has finite energy if $\sum |u_i^T b_{\text{exact}}/\sigma_i|^2 < \infty$, with no inaccessible data component. With independent whitened noise of similar variance in each left singular direction, $|u_i^T b|$ eventually follows a roughly flat noise floor. A Picard diagram plots σ_i, $|u_i^T b|$ and their quotient against index. For an exact quotient proportional to i^{-1} the squared sum converges. If the noise coefficients have typical size δ and $\sigma_i \sim i^{-2}$, their quotients typically grow like δi^2. The index where the data cease to decay depends on the observation and its noise.

Derive what an L-curve actually plots

For identity Tikhonov and strength μ, write $b_i = u_i^T b$ and $f_i = \sigma_i^2/(\sigma_i^2 + \mu^2)$. The residual and solution coordinates of the L-curve are

$$R^2 = \sum_i (1 - f_i)^2 b_i^2 + \|b_\perp\|^2,$$

$$S^2 = \sum_i (f_i b_i/\sigma_i)^2.$$

Mark $(\log R, \log S)$ against $\log \mu$. A bend can reveal the transition from fitting signal to amplifying noise. With $\eta = \log \mu$ its curvature magnitude is $|X'Y'' - Y'X''|/(X'^2 + Y'^2)^{3/2}$ for $X = \log R, Y = \log S$. Sparse sampling or multiple spectral clusters can give competing curvature peaks. If b has energy in only one singular direction, the curve reflects the scale of that mode even with zero noise. A corner by itself then provides no data-driven noise estimate. Compare with a discrepancy level justified by instrument noise, or held-out predictive error and GCV.

For a reproducible ambiguous corner choose singular values $(1, s)$ and data coefficients $(1, \sqrt{s})$, with $s = 10^{-3}$. Replacing μ by s/μ swaps each filter with the complement of the other. Substitution gives $R(s/\mu) = \sqrt{s}\, S(\mu)$ and $S(s/\mu) = R(\mu)/\sqrt{s}$. In log coordinates this transformation is a reflection and preserves curvature magnitude. The two outward bends near $\mu = s$ and $\mu = 1$ therefore have equal curvature. These choices differ by roughly a thousand in regularisation strength. If the second measured coefficient is reliable, damping it away costs information. If that coefficient is mostly noise, keeping it produces a state component amplified by $1/s$. The shape of this curve cannot decide which measurement assumption applies.

OneSimulation experiment · Reconstruct data from an independently discretised instrument

Vary the true blur width or noise seed. Decide which feature persists across plausible instruments.

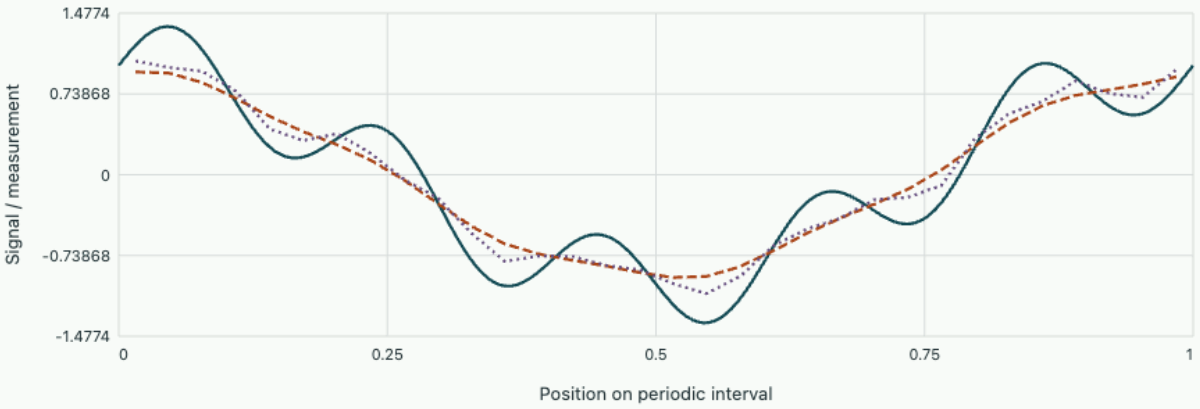
Download the executable notebook · Python source

Adjust the experiment in the browser. Start with matching blur widths and zero noise. Increase the noise, then vary μ to trade oscillatory reconstruction against loss of the small-scale signal. Finally change only the true width. A small data residual can now coexist with a biased reconstruction. Every adjustment regenerates the data on 1024 source cells and solves the 32-variable regularised normal system by Cholesky factorisation. The browser keeps one reproducible noise pattern so a slider movement isolates a parameter change. The downloadable OneSimulation model below makes the same distinction between data and inversion grids and adds spectral diagnostics; its random generator produces a different noise sample.

Open the live experiment

Try the calculation

True blur width	0.065
Assumed blur width	0.045
Noise standard deviation	0.002
$\log_{10}$ regularisation μ	-2



True signal | Noisy instrument data | Tikhonov reconstruction

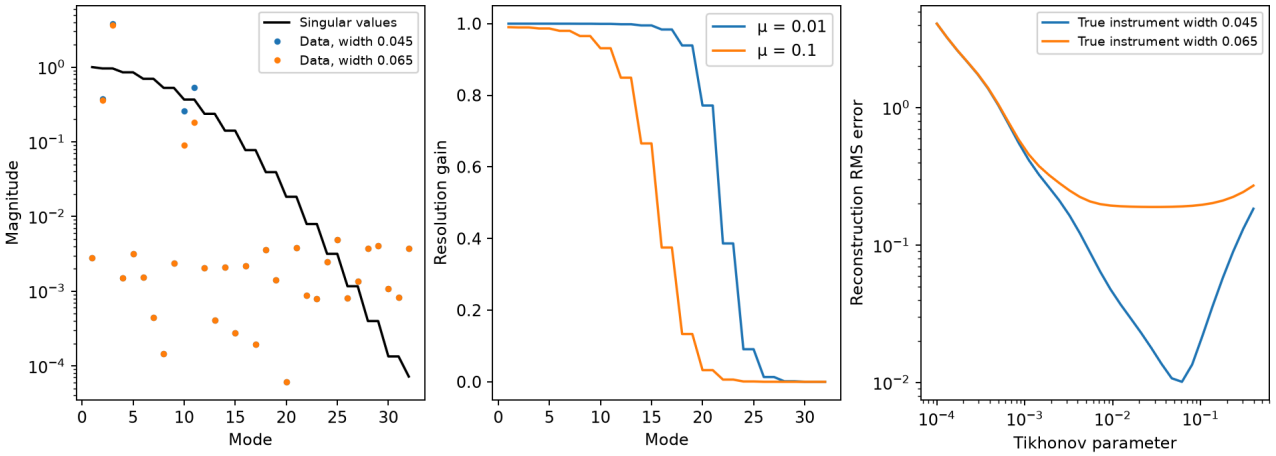
Reconstruction RMS error = 0.19639; data-fit RMS residual = 0.00070773; μ = 0.01. Data use 1024 source cells; the inverse model uses 32. The fixed browser noise realisation is reproducible but differs from the notebook seed implementation.

```
"""Observe model error in a periodic deblurring experiment on different grids."""
import matplotlib.pyplot as plt
import numpy as np
from onesim.numerics.linear.matrices.dense_linear_algebra import solve_dense

def instrument(receivers, sources, width):
    delta = (receivers[:, None] - sources[None, :] + 0.5) % 1 - 0.5
    kernel = np.exp(-(delta/width)**2/2)
    return kernel / kernel.sum(axis=1)[:, None]

def signal(x):
    return np.cos(2*np.pi*x) + 0.4*np.sin(10*np.pi*x)

fine = (np.arange(1024)+0.5)/1024
coarse = (np.arange(32)+0.5)/32
design = instrument(coarse, coarse, 0.045)
rng = np.random.default_rng(17)
noise = rng.normal(0, 0.002, coarse.size)
strengths = np.logspace(-4, -0.4, 32)
fig, (picard, resolution, ax) = plt.subplots(1, 3, figsize=(11.2, 4), layout="constrained")
left, singular, _ = np.linalg.svd(design, full_matrices=False)
indices = np.arange(1, len(singular)+1)
picard.semilogy(indices, singular, color="black", label="Singular values")
for trial in [0.01, 0.1]:
    gain = singular**2 / (singular**2 + trial**2)
    resolution.plot(indices, gain, label=f"μ = {trial:g}")
resolution.set(xlabel="Mode", ylabel="Resolution gain")
resolution.legend()
for width in [0.045, 0.065]:
```

```
    data = instrument(coarse, fine, width) @ signal(fine) + noise
    picard.semilogy(indices, abs(left.T @ data), ".", label=f"Data, width {width}")
    errors = []
    for mu in strengths:
        state = solve_dense(design.T @ design + mu**2*np.eye(coarse.size), design.T @ data)
        errors.append(np.linalg.norm(state-signal(coarse))/np.sqrt(coarse.size))
    print(f"True width {width:.3f}: minimum grid RMS error {min(errors):.5f}")
    ax.loglog(strengths, errors, label=f"True instrument width {width}")
ax.set(xlabel="Tikhonov parameter", ylabel="Reconstruction RMS error")
ax.legend(fontsize=8)
picard.set(xlabel="Mode", ylabel="Magnitude")
picard.legend(fontsize=8)
fig.savefig("numerics-inverse-blur.png", dpi=170)
```

Output from the reference run

```
True width 0.045: minimum grid RMS error 0.01013
True width 0.065: minimum grid RMS error 0.18949
```



Noise, spectral resolution and reconstruction error expose different limitations of two blur models.

Why separate the generation and inversion grids?

Matrices on the same grid share quadrature and representation errors. Synthesising $b = Ax$ and inverting with exactly that A omits those errors from the test. A finer source grid and slightly perturbed blur width create two separately controllable discrepancies. The reconstruction notebook shows how the best achievable error changes when only the assumed width is wrong.

Tutorial on weighted spectra and resolution

Calculate the two singular values of the midpoint example

▼ **Work through the solution**

The matrix is symmetric positive definite with trace $21/16$ and determinant $1/16$. Its eigenvalues are $(21 \pm \sqrt{377})/32$, also its singular values. The smaller mode is much more sensitive to perturbations in the measurements.

Distinguish variance reduction from regained resolution

▼ **Work through the solution**

In singular direction i the mean reconstruction of a true coefficient x_i is $f_i x_i$ and independent data noise of variance δ^2 contributes variance $\delta^2 f_i^2/\sigma_i^2$. Larger μ lowers both gain and variance. A small repeat-to-repeat spread can coexist with a strongly suppressed feature.

Hansen · discrete inverse problems and teaching materials

Inferring parameters and fields from observations · 7. Early stopping and penalties that leave part of a field free

Derive two matrix-free least-squares iterations, find where noise first enters their solution and separate the measured and penalised directions of a nonstandard Tikhonov problem.

Obtain a full iteration without forming the normal matrix

Let $r = b - Ax,\ s = A^T r$. Applying conjugate gradients to $A^T A x = A^T b$ starting at $x = 0$ works on the range of A^T, where the operator is positive definite even if A has a nullspace. Initialise $p = s$ and $\gamma = s^T s$. Along p the residual minimisation gives $\alpha = (Ap)^T r / \|Ap\|^2$. Conjugacy of earlier directions implies $p^T s = s^T s$. Thus one CGLS iteration is

$$q = Ap,$$

$$\alpha = \gamma/(q^T q), \quad x \leftarrow x + \alpha p, \quad r \leftarrow r - \alpha q, \quad s \leftarrow A^T r,$$

$$\gamma_{\text{new}} = s^T s, \quad p \leftarrow s + (\gamma_{\text{new}}/\gamma)p.$$

Replace γ by γ_{new} and continue. A vanishing gradient means stationarity. Handle a zero denominator before division and recompute true residuals when finite precision matters. The recurrence needs only forward and transpose products. Exact-arithmetic search spaces are nested, so the training residual falls monotonically. The distance to the unknown state can increase.

Solve a sequence of projected rectangular problems

Normalise $u_1 = b/\beta_1$, $\beta_1 = \|b\|$, and $v_1 = A^T u_1/\alpha_1$, $\alpha_1 = \|A^T u_1\|$. Alternate orthogonal projections:

$$\beta_{j+1} u_{j+1} = A v_j - \alpha_j u_j,$$

$$\alpha_{j+1} v_{j+1} = A^T u_{j+1} - \beta_{j+1} v_j.$$

These are the Golub–Kahan recurrences, with signs fixed by positive norms. In exact arithmetic induction using transposition shows both sets are orthonormal until breakdown. Collect the coefficients in a lower bidiagonal $(k+1)$-by-k matrix B_k. Since $AV_k = U_{k+1} B_k$, minimise $\|\beta_1 e_1 - B_k y\|$ and put $x_k = V_k y$. LSQR applies plane rotations as each bidiagonal row is appended. Its exact-arithmetic iterates match CGLS started at zero. Finite precision, explicit residual checks and stopping rules affect their returned solutions.

Take $A = \operatorname{diag}(1, 0.01)$, true state $(1, 0)$, measured $b = (1, 0.01)$. The first CGLS step uses $p = (1, 10^{-4})$, with α essentially one, leaving the error about 10^{-4}. The second step can drive both residuals to zero by returning $(1, 1)$, whose state error is one. A discrepancy threshold near the known noise norm 0.01 can stop before fitting the weak direction.

Diagonalise a matrix pair in the right metric

Consider $\|Ax - b\|^2 + \mu^2 \|Lx\|^2$. Uniqueness requires $\ker A \cap \ker L = \{0\}$. Then $M = A^T A + L^T L$ is positive definite. Factor $M = R^T R$. The matrices $\widetilde{A} = AR^{-1}$, $\widetilde{L} = LR^{-1}$ satisfy $\widetilde{A}^T \widetilde{A} + \widetilde{L}^T \widetilde{L} = I$.

Orthogonally diagonalise the first Gram matrix. In that same basis the second is diagonal with complementary nonnegative entries c_i^2, s_i^2. Thus the associated right directions $x_i = R^{-1}v_i$ satisfy $\|Ax_i\| = c_i,\ \|Lx_i\| = s_i$ and $c_i^2 + s_i^2 = 1$. Their images separately are pairwise orthogonal. This construction derives the GSVD geometry. A production algorithm avoids explicitly forming ill-conditioned Gram matrices.

For $c_i > 0$ put $u_i = Ax_i/c_i$. With $x = \sum \xi_i x_i$, the component objective is $(c_i\xi_i - u_i^T b)^2 + \mu^2 s_i^2 \xi_i^2$. Minimisation gives

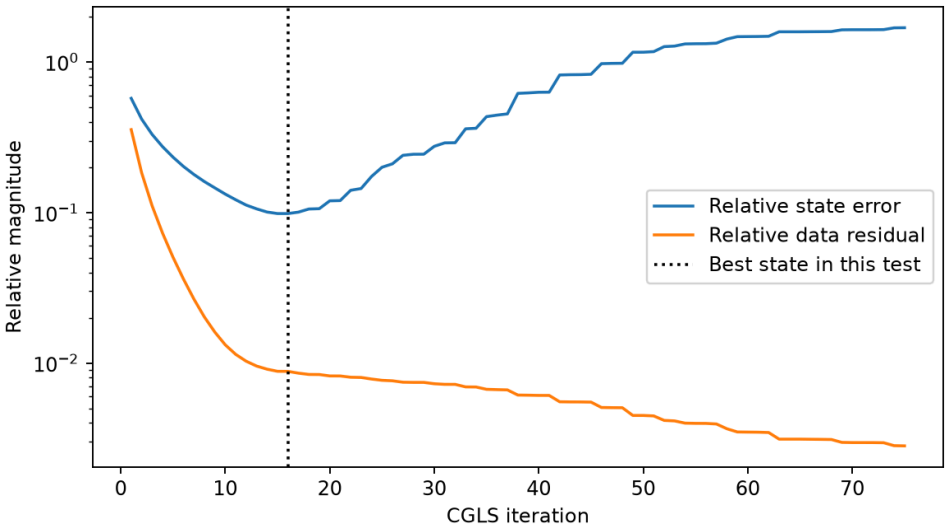
$$\xi_i = \frac{c_i}{c_i^2 + \mu^2 s_i^2}\, u_i^T b,$$

f_i=\frac{c_i^2}{c_i^2+\mu^2s_i^2}.

An unpenalised but measured mode has $s_i = 0$ and filter one. A measured-data null direction with $c_i = 0$ has coefficient zero at the minimum. With a nonzero reference x_{ref}, apply the same derivation to its correction and the data $b - Ax_{\mathrm{ref}}$.

What if both operators miss the field average?

A gradient penalty and a differential measurement can have the same constant nullspace. Adding any constant changes neither term of the objective. An extra measurement or physically specified mean is needed to fix the offset. An arbitrary minimum-norm convention changes the interpretation of the result.



The residual decreases after the best state has been passed. Weakly transmitted components eventually fit the fixed measurement noise. The state-error minimum is observable here because the source used to generate the example is known.

Tutorial on filtered and unfiltered directions

Interpret a first-difference penalty in a two-cell experiment

▼ **Work through the solution**

With $A = I,\ L = (1, -1)$, normalised sum and difference are Euclidean eigenvectors. Their fitted gains are 1 and $1/(1 + 2\mu^2)$. Increasing μ enforces nearly equal cells while retaining their experimentally measured mean.

Choose a fair comparison between LSQR and CGLS

▼ **Work through the solution**

Start both at zero, use identical operators and noisy data, count one forward and one transpose product per iteration, and plot reconstruction error and independent prediction error alongside the training residual. Equal iteration counts isolate finite-precision effects. Equal stopping rules test what an investigator actually reports.

Hansen · iterative and general-form regularisation

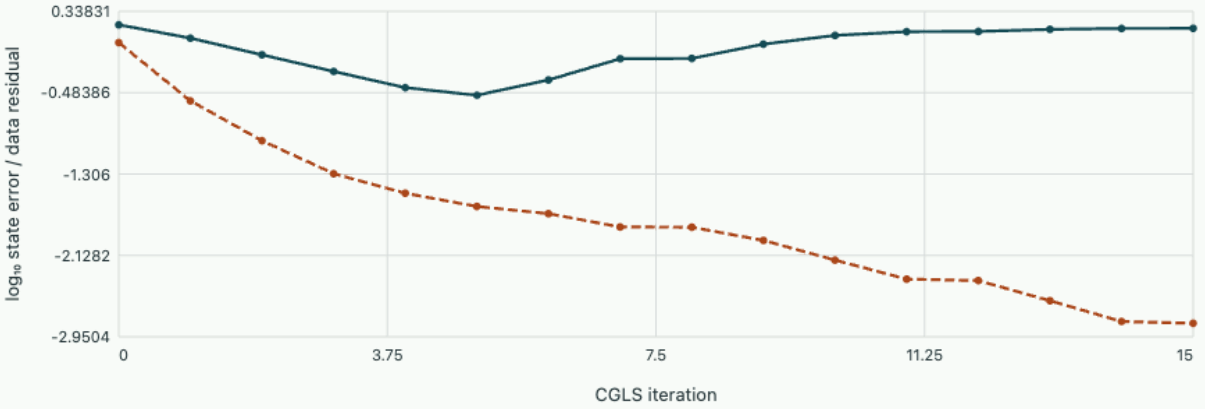
Change parameters · Early stopping and penalties that leave part of a field free

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

Noise amplitude	0.01
Singular-value decay power	1.5
Displayed iterations	15



Smallest displayed state error occurs at iteration 5. The truth is available only because this is a manufactured experiment; residual decrease alone cannot identify that stopping point.

Which field can be recovered from boundary data?

Relate an observation to its spatial sensitivity. Use reciprocity, smoothing with distance and changes of electrode configuration to identify recoverable features.

1. Green functions, the Poisson kernel and what a small electrode can influence
2. Normalised spherical harmonics and the field of an arbitrary three-dimensional source
3. Capacitance matrices, floating nodes and controlled approximation bounds
4. Reconstructing a source from noisy voltage measurements
5. Which part of a sample changes a four-electrode impedance measurement?
6. Designing a displacement bridge at its noise limit
7. Least squares and orthogonal factorisation
8. Singular values and minimum-norm solutions
9. Removing unequal scales before an iteration

Green functions, the Poisson kernel and what a small electrode can influence

Electrostatic field theory and measurement

A Green function is the response of an entire boundary-value problem to a unit source. Once that response is known, a continuous charge distribution and a patterned boundary voltage can both be treated by integration. The two integrations require different kernels.

Define a point response with its boundary conditions

Let Ω be a vacuum domain with boundary on which the voltage is given. Use the convention $-\nabla_{\mathbf{y}}^2 G(\mathbf{x},\mathbf{y}) = \delta^3(\mathbf{y}-\mathbf{x})$ and $G = 0$ on the boundary as a function of $\mathbf{y}$. In free space $G = 1/(4\pi|\mathbf{x}-\mathbf{y}|)$ if decay is required at infinity. This definition places no permittivity in G; physical charge enters as ρ/ϵ_0. A different frequent convention puts 4π in the defining equation, changing all subsequent prefactors. Comparing the near-source flux to unity is a straightforward way to fix which convention is being used.

Begin with two differentiable functions u, v. The difference $\nabla\cdot(u\nabla v - v\nabla u) = u\nabla^2 v - v\nabla^2 u$ has no mixed-gradient term. Integrate this identity with $u = V(\mathbf{y})$,

$$v = G(\mathbf{x},\mathbf{y}).$$

The source equation for G samples V at $\mathbf{x}$, while Poisson's equation for V supplies a volume density. With the outward normal taken from Ω,

$$V(\mathbf{x}) = \int_\Omega G(\mathbf{x},\mathbf{y})\,\frac{\rho(\mathbf{y})}{\epsilon_0}\,d^3y - \oint_{\partial\Omega} V(\mathbf{y})\,\partial_{n_{\mathbf{y}}} G(\mathbf{x},\mathbf{y})\,dA_{\mathbf{y}}.$$

The $G\partial_n V$ boundary term vanishes because $G = 0$ there. If the singularity causes concern, exclude a small sphere about $\mathbf{x}$, apply the ordinary identity, and shrink the sphere. The inward normal on the excised sphere produces the same $V(\mathbf{x})$ term. The representation separates charges in the volume from voltages maintained by electrode sources; it does not combine them into one fictitious material distribution.

Solve the half-space and differentiate with respect to source position

Consider $\mathbf{x} = (\mathbf{X}, h)$ with $h > 0$ above the plane $z = 0$. For an interior source $\mathbf{y} = (\mathbf{Y}, z)$, denote its reflection by $\mathbf{y}^* = (\mathbf{Y}, -z)$. The grounded half-space Green function is

$$G = \frac{1}{4\pi}\left[\frac{1}{\sqrt{|\mathbf{X}-\mathbf{Y}|^2 + (h-z)^2}} - \frac{1}{\sqrt{|\mathbf{X}-\mathbf{Y}|^2 + (h+z)^2}}\right].$$

It vanishes as the source coordinate approaches the plane. The outward domain normal there is $-\hat{\mathbf{z}}$; therefore the coefficient of a prescribed boundary voltage is

$$P_h = -\partial_n G = \partial_z G|_{z=0}.$$

Differentiating the two terms gives equal positive contributions:

$$P_h(\mathbf{X}-\mathbf{Y}) = \frac{h}{2\pi[|\mathbf{X}-\mathbf{Y}|^2+h^2]^{3/2}}.$$

Its units are inverse area, while G has inverse length in this three-dimensional convention. For no interior charge $V(\mathbf{X},h) = \int P_h(\mathbf{X}-\mathbf{Y}) f(\mathbf{Y})\, d^2Y$. The kernel is positive. Integrate over the whole plane using $2\pi s\, ds$; the result is one. Hence a constant boundary voltage is reproduced, and an arbitrary voltage bounded between two numbers produces values between the same bounds inside. The integral solves the bounded half-space problem with the far-field condition appropriate to the data. Demanding decay for a nonzero constant boundary value would be inconsistent.

A detector at height h takes a weighted average over a neighbourhood of width comparable with h. The kernel has algebraic tails, so its support is not a sharp circular footprint. The contribution from radii smaller than b is $1 - h/\sqrt{b^2+h^2}$. Half the total weight lies within $b = \sqrt{3}h$. Ninety per cent requires

$$b = \sqrt{99}h \simeq 9.95h.$$

These quite different radii show why a full-width-at-half-maximum measurement of the kernel cannot by itself define the region that contains most of its influence.

A finite driven electrode and a reciprocity experiment

Fix a circular disk of radius b in the plane at voltage V_d while grounding the rest of the plane, with ideal thin insulating seams. Directly over its centre,

$$V(0,h) = V_d\left(1 - \frac{h}{\sqrt{h^2+b^2}}\right),$$

$$E_z(0,h) = V_d \frac{b^2}{(h^2+b^2)^{3/2}}.$$

At $h \to 0^+$ the electrode voltage is recovered at its centre. For $h \gg b$ the voltage scales as $V_d b^2/(2h^2)$. A small electrode surrounded by ground does not behave at long distance as an isolated charged electrode with a $1/h$ potential. Its surrounding return charge makes a combined disturbance whose far field has a different order. At the sharp ideal edge the boundary voltage jumps, and local fields there require an actual seam geometry if one wants a bounded peak or a breakdown prediction.

For $b = 5.0$ mm, $V_d = 4.0$ V and $h = 5.0$ mm the axial voltage is 1.17 V and field is 283 V/m. A separate experiment grounds every electrode and places a point charge q at that observation point. Green reciprocity relates the net charge on the disk to the previous weighting potential $w = V/V_d$: $Q_d = -q\,w$. With $q = 20$ pC the disk collects -5.86 pC. Direct integration of the image-plane surface charge $\sigma(s) = -qh/[2\pi(s^2+h^2)^{3/2}]$ gives the same result. It is precisely $-qP_h$. Thus the boundary kernel has both a voltage-producing interpretation and a charge-collection interpretation measurable with different instrumentation.

Recover the kernel of an electrode strip

A boundary pattern invariant in y should not require a two-dimensional source integral. Integrate P_h over the unused coordinate. With lateral separation $s = X - X'$, the result is

$$p_h(s) = h/[\pi(s^2 + h^2)].$$

It integrates to one over the real line. A strip between $-b, b$ at V_d gives on its central normal $V(0, h) = 2V_d \arctan(b/h)/\pi$. At $h = b$ the value is half the applied voltage, distinct from the circular electrode's $1 - 1/\sqrt{2}$. The strip extends in one more direction and occupies a larger angular range of boundary influence.

For sinusoidal boundary data $f(X) = \cos kX$ there is an independent way to determine the result without evaluating a difficult oscillatory integral. Translational symmetry and linearity preserve the sinusoid. Solving Laplace's equation with vertical factor $Z(h)$ gives

$$Z'' = k^2 Z.$$

The bounded decaying solution with unit boundary amplitude is $e^{-|k|h}$. Thus the Fourier transform of the strip kernel is $e^{-|k|h}$. Convolving twice with heights h_1, h_2 is equivalent to convolution once with height $h_1 + h_2$. This composition property checks a numerical propagation code without requiring separate analytic solutions for many different boundary patterns.

Which symmetries does an influence table possess?

Exchange source and detector in the defining equations for $G(\mathbf{x}, \cdot)$ and $G(\mathbf{z}, \cdot)$. Their difference identity gives $G(\mathbf{x}, \mathbf{z}) - G(\mathbf{z}, \mathbf{x}) = 0$, since both vanish on the same boundary. Symmetry extends to scalar or symmetric-tensor positive permittivity when the operator is defined as $-\nabla \cdot \epsilon \nabla$. It refers to exactly the same material, boundary preparation and domain in both experiments. Replacing one experiment's grounded conductor by a floating conductor changes the Green function.

On a nonuniform mesh it matters whether influence entries correspond to equal point charges or equal charge densities in unequal cells. If a finite-volume solver stores sources as densities, a column's total injected charge scales with its cell volume. An unweighted table of sampled voltages can look nonsymmetric even when the discrete physics is reciprocal. Form the matrix with unit total source and consistent detector averaging, or retain the correct volume-weighted inner product. Test positivity for positive sources as a second check. Symmetric influence entries by themselves cannot expose a common incorrect normalisation or a misplaced outer boundary.

A kernel that reproduces every harmonic around a circle

In a two-dimensional circular domain of radius a, prescribe a voltage $f(\phi')$ around the wall. For a detector at (r, ϕ) let $t = r/a < 1$. The angular Fourier component with order m is multiplied by $t^{|m|}$ on continuation into the disk. This follows directly from the regular radial power solution of Laplace's

equation. Substitute the Fourier coefficient integral into the series and interchange the integral and the absolutely convergent series at fixed $t < 1$. The boundary response per unit angle is then

$$P_t(\delta) = \frac{1}{2\pi} \sum_{m=-\infty}^{\infty} t^{|m|} e^{im\delta} = \frac{1 - t^2}{2\pi(1 - 2t\cos\delta + t^2)},$$

$$\delta = \phi - \phi'.$$

To obtain the rational form, separate the $m = 0$ term and sum the two conjugate geometric series $te^{\pm i\delta}/(1 - te^{\pm i\delta})$. The result is positive for every interior point. Integrating it over angle leaves only the constant Fourier mode, so it is normalised to unity. At the centre the kernel is simply $1/(2\pi)$, recovering the exact mean-value theorem for harmonic functions. Near the wall it becomes very sharply concentrated around the detector's azimuth while retaining a small nonlocal tail. Using prescribed boundary data as a flat average at every interior point would reproduce only the centre reading.

For a concrete experiment put a short arc of half-angle α centred on $\phi = 0$ at 1 V and ground the remaining circular boundary. Along $\phi = 0$ the potential equals the integral of P_t between $-\alpha, \alpha$. With $u = \tan(\phi'/2)$ one has $\cos\phi' = (1 - u^2)/(1 + u^2)$, $d\phi' = 2du/(1 + u^2)$. Substitution gives the weight

$$w(t) = \frac{2}{\pi} \arctan\left[\frac{1 + t}{1 - t} \tan(\alpha/2)\right].$$

At $t = 0$ it gives α/π, exactly the fraction of the circumference occupied by the arc. For a 60°-wide arc $\alpha = \pi/6$, hence the centre reads $1/6$ V. At $t = 0.6$ on the centreline it reads about 0.522 V. Placing a small positive line charge there while grounding the whole cylindrical wall induces -0.522 times the line charge on that arc by the same reciprocity argument used on the flat disk. The angular share depends on position inside the circle even though every point on the conductor has the same actual grounded voltage.

An independent check takes the limit of a very large circle with the probe only a distance $h = a - r$ from the nearest boundary point. For small angular displacement $\delta = s/a$, $1 - 2t\cos\delta + t^2 \simeq (h/a)^2 + (s/a)^2$ and $1 - t^2 \simeq 2h/a$. The angular probability element $P_t(\delta)d\delta$ therefore tends to $h\,ds/[\pi(h^2 + s^2)]$, precisely the translation-invariant strip kernel obtained by integrating the half-space formula over one coordinate. Curvature corrections can be estimated by increasing a/h at fixed physical arc width and comparing both predictions.

Why an insulated Green function needs a zero-mode subtraction

Consider the simpler interval $0 < x < L$ with prescribed zero flux at both ends. We would like a point response satisfying $-G'' = \delta(x - s)$, $G'(0) = G'(L) = 0$. Integrate this proposed equation over the interval: the left-hand side is $-G'(L) + G'(0) = 0$, whereas the right-hand side is 1. The attempted kernel violates compatibility. Physically a sealed insulating region cannot take a net free test charge without changing its prescribed outer flux. Adding an arbitrary constant to G cannot fix this contradiction.

A useful alternative removes the constant source mode and fixes the remaining voltage reference. Define $-G_N'' = \delta(x - s) - 1/L$, $G_N'(0) = G_N'(L) = 0$, and $\int_0^L G_N(x, s)\,dx = 0$. Away from the point the

curvature is $+1/L$. The left derivative must be x/L to vanish at zero, the right derivative $x/L-1$ to vanish at L. They jump by -1 at $x=s$. Integrating, imposing continuity and subtracting the mean gives the compact symmetric expression

$$G_N(x,s)=\frac{x^2+s^2}{2L}-\max(x,s)+\frac{L}{3}.$$

Differentiating verifies both end conditions and the jump. Integrating the quadratic part and the piecewise maximum over x verifies the mean condition. Its sign can be negative because the response includes an opposite uniform return source and uses a zero-mean reference. Unlike the positive grounded Dirichlet kernel, it is not a positive voltage weight to use for a maximum-principle test.

Apply a neutral pair of sheets $+\sigma$ at $s_1=0.2L$ and $-\sigma$ at $s_2=0.7L$. Their voltage is $\sigma[G_N(x,s_1)-G_N(x,s_2)]/\epsilon$. The two artificial uniform return sources cancel exactly. To the left of both sheets the field vanishes, between them it equals $+\sigma/\epsilon$, and to the right it vanishes again. The potential drops by $0.5L\sigma/\epsilon$ across the two sheets. This matches direct Gauss-law jumps, and the zero-mean potential fixes the constant that Gauss's law could not determine. For more complex closed domains numerical solvers implement the same idea by projecting out the constant eigenvector and checking the compatible total charge before inverting. An arbitrary tiny diagonal regularisation of a singular matrix can hide a physically impossible net-source input.

Construct a point response when the medium itself screens charge

A linear ionic or phenomenological screening model between two grounded parallel electrodes leads in one dimension to $[-\partial_x^2+\kappa^2]V=\rho_{\mathrm{imposed}}/\epsilon$, $0<x<L$. Here ρ_{imposed} excludes the mobile linear response already represented by κ^2V. We seek $G_\kappa(x,s)$ with zero voltage at both ends. Away from its unit source the equation has hyperbolic solutions. Use $A\sinh(\kappa x)$ on the left to vanish at zero and $B\sinh[\kappa(L-x)]$ on the right to vanish at L. Continuity at $x=s$ removes one constant. Integrating across the source gives $G'_\kappa(s^+)-G'_\kappa(s^-)=-1$; the finite mass term integrates to zero in a shrinking source interval.

To solve transparently use the functions $u(x)=\sinh(\kappa x)$,

$$v(x)=\sinh[\kappa(L-x)].$$

Their Wronskian $u'v-uv'$ is $\kappa\sinh(\kappa L)$, independent of x. Then continuity, the derivative jump, and symmetry under $s\leftrightarrow x$ are manifest in

$$G_\kappa(x,s)=\frac{\sinh(\kappa x_<)\sinh[\kappa(L-x_>)]}{\kappa\sinh(\kappa L)},$$

$$x_<=\min(x,s),$$

$$x_>=\max(x,s).$$

For $\kappa\to0$, expand each hyperbolic sine to first order. The result is $x_<(L-x_>)/L$, the grounded unscreened response. When both observation and source lie many screening lengths from every

boundary, the kernel tends to $e^{-\kappa|x-s|}/(2\kappa)$, the infinite-line screened solution. Thus one formula contains both geometric screening by grounded conductors and material screening by mobile charges.

At the midpoint source and detector coincide. The finite kernel reduces to $\tanh(\kappa L/2)/(2\kappa)$. Compared with the unscreened midpoint value $L/4$, its ratio is $2\tanh(\kappa L/2)/(\kappa L)$. At $\kappa L = 2$ a fixed inserted sheet therefore creates only 76.2% of its unscreened midpoint voltage. A naïve replacement of plate separation by the screening length would give a constant ratio and miss the smooth small-κ crossover. For long gaps the measured voltage saturates at fixed sheet charge, because the countercharge response near the sheet now sets the local scale.

Check the convolution using a uniform positive imposed charge density ρ_0. Without integrating the kernel directly, seek a constant particular solution plus a symmetric homogeneous correction. The result is

$$V(x) = [\rho_0/(\epsilon\kappa^2)]\{1 - \cosh[\kappa(x - L/2)]/\cosh(\kappa L/2)\}.$$

Twice differentiation reconstructs the source and zero endpoints. In the small-κL limit it tends to $\rho_0 x(L - x)/(2\epsilon)$. A numerical integration of the point response against a uniform source should match this independent closed solution throughout the interval, including the near-electrode regions. It is a stronger test than checking the point kernel at the centre alone.

Integrate the original equation across the gap. The sum of outward displacement fluxes equals imposed charge minus $\epsilon\kappa^2 \int V\,dx$, the magnitude of the compensating mobile charge. For the uniform source the fraction of imposed charge balanced by electrode rather than mobile charge is $2\tanh(\kappa L/2)/(\kappa L)$. For a long gap this fraction falls like twice the screening length divided by gap width, because only the two boundary regions communicate appreciably with the electrodes. A current meter attached to the electrodes and a chemical measurement of accumulated mobile ions can therefore close the same source budget independently.

Inspect the response to a localised source

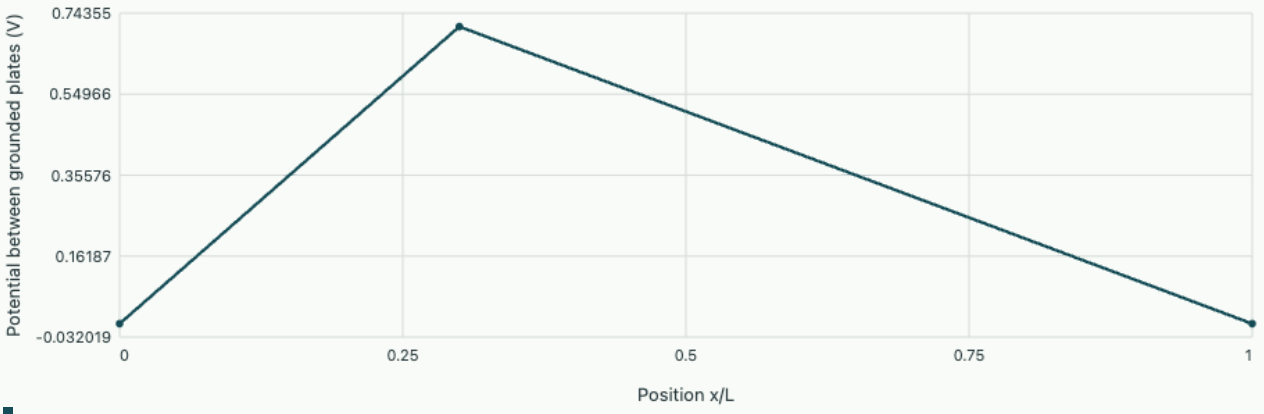
Move the source while holding the boundaries fixed. Test the symmetry between source and observation point predicted by reciprocity.

Open this experiment online

Try the calculation

Sheet position s/L **0.3**

Sheet charge density (nC/m²) **3**



Grounded 10 mm gap

Induced densities are -2.1 nC/m² on the left and -0.9 nC/m² on the right. Potential at the sheet is 0.712 V.

Design a differential collecting electrode from its weighting functions

Two concentric electrodes supply one signed position signal

A grounded plane is divided into a central disk of radius a, a surrounding annulus from a to $2a$, and the remainder. A small charge moves along the central normal. Two charge amplifiers measure the disk and annulus. Construct an output that vanishes at $h = a$, then find its local height sensitivity. Why can that output alone not distinguish uncertain charge from height away from the null?

▼ **Derive and check the result**

Let $w_a = 1 - h/\sqrt{h^2 + a^2}$ and $w_{2a} = 1 - h/\sqrt{h^2 + 4a^2}$. The disk reads $Q_d = -q w_a$; the annulus reads $Q_r = -q(w_{2a} - w_a)$. An output $S = Q_d - \gamma Q_r$ vanishes at $h = a$ if $\gamma = (1 - 1/\sqrt{2})/(1/\sqrt{2} - 1/\sqrt{5}) = 1.127$ approximately. Both readings are negative for positive q, so it is their difference, not their sum, that rejects the reference position. The ideal zero does not depend on the magnitude of any nonzero q.

Differentiate with respect to height using $w_b' = -b^2/(h^2 + b^2)^{3/2}$. At the null $a\,w_a' = -1/(2\sqrt{2})$ and $a\,w_{2a}' = -4/(5\sqrt{5})$. Hence

$$a\,d(S/q)/dh = -(1 + \gamma) a w_a' + \gamma a w_{2a}' \simeq 0.349.$$

For $q = 10$ pC, $a = 2.0$ mm, a small upward displacement of 0.10 mm produces approximately $+0.175$ pC. Finite differences of the full two-disk formula at heights on both sides of a check the linear slope. For displacements that are no longer small compared with a, use the full expression rather than extrapolating that single sensitivity.

An additional sum channel $Q_d + Q_r = -qw_{2a}$ permits a dimensionless output $S/(Q_d + Q_r)$ that cancels the magnitude and sign of q while retaining height dependence. This ratio becomes noise sensitive far from the electrodes, where collected charge is small. A null by itself can also mean that there is no charge at all, so a usable position detector checks the sum channel at the same time. The exercise combines electrostatic reciprocity, electronic weighting and a finite signal-to-noise constraint while requiring only two precomputed boundary responses.

Tutorial · predict before numerically integrating the kernel

How much charge belongs to an annular collector?

An annular electrode with inner radius h and outer radius $3h$ lies in a grounded plane. A charge q sits at height h over its centre. Derive its induced charge. If a numerical integration omits all radii beyond $3h$, where does the missing charge go?

▼ Work through the solution

Subtract two circular weights. The annulus receives $Q/q = -[1/\sqrt{2} - 1/\sqrt{10}] = -0.3909$. The central disk receives $-(1 - 1/\sqrt{2})q = -0.2929q$. The whole plane should sum to $-q$, so the part beyond the outer edge supplies

$$-q/\sqrt{10} = -0.3162q.$$

Nothing has been lost by the physical system; a finite integration patch excludes a considerable piece of the long-tailed kernel. Radial extrapolation with the analytic far tail or an expanding integration domain distinguishes missing area from poor quadrature within the patch.

Use resistance to a modelling mistake

A student proposes G itself as the weight for setting electrode voltages on the plane. Another replaces it by $2P_h$, arguing that an image doubles the response. Predict two different tests that diagnose these errors.

▼ Work through the solution

The first proposal gives zero contribution since the Dirichlet Green function is zero at the boundary. It confuses volume source response with a boundary normal derivative. The second produces twice the prescribed constant voltage because the kernel would integrate to two. Independent electrostatic

reciprocity would assign $-2q$ to the whole grounded plane around a charge q, violating the charge sum derived from the real and image fields. For a numerical test, set successively a constant, a sinusoid, and a bounded disk on the same plane. Those cases examine normalisation, spectral attenuation, and spatial tails separately.

Further reading

- Move a source between two grounded electrodes
- Compare vertical attenuation for two boundary harmonics
- Fließbach, chapter 7 · electrostatic boundary problems

Normalised spherical harmonics and the field of an arbitrary three-dimensional source

Electrostatic field theory and measurement

An axial voltage scan can determine Legendre coefficients only if the source is rotationally symmetric. A general spherical electrode or molecule needs angular modes with two indices. Their normalisation fixes the conversion from a measured multipole coefficient to a physical voltage.

Separate three coordinates without imposing axial symmetry

In spherical coordinates the source-free Laplace equation is $r^{-2}\partial_r(r^2\partial_r V) + r^{-2}\Delta_\Omega V = 0$, where $\Delta_\Omega = (\sin\theta)^{-1}\partial_\theta(\sin\theta\,\partial_\theta) + (\sin^2\theta)^{-1}\partial_\phi^2$. Assume $V = R(r)Y(\theta,\phi)$. Separation requires

$$-\Delta_\Omega Y = \lambda Y.$$

Multiply by Y^* and integrate over the unit sphere. Integration by parts has no actual surface boundary for globally smooth fields and gives

$$\lambda \int |Y|^2 d\Omega = \int |\nabla_\Omega Y|^2 d\Omega \geq 0.$$

The constant mode has eigenvalue zero.

Separate azimuth as $e^{im\phi}$. Continuity after a full rotation forces integer m. With $u = \cos\theta$, the remaining equation becomes

$$[(1-u^2)P']' + [\lambda - m^2/(1-u^2)]P = 0.$$

Requiring regular behaviour at both poles yields $\lambda = \ell(\ell+1)$ with integers $\ell \geq |m|$. The corresponding angular functions are the associated Legendre functions. For each ℓ there are $2\ell+1$ independent m's, all with identical radial behaviour. A rotation mixes coefficients within this subspace but cannot convert a dipole into a quadrupole.

Fix one convention and verify its first modes

Choose $P_\ell(1) = 1$ and for $m \geq 0$ include the Condon–Shortley phase in $P_\ell^m(u) = (-1)^m(1-u^2)^{m/2}d^m P_\ell/du^m$. Define

$$Y_{\ell m} = \sqrt{\frac{2\ell+1}{4\pi}\frac{(\ell-m)!}{(\ell+m)!}}\,P_\ell^m(\cos\theta)e^{im\phi},$$

$$Y_{\ell,-m} = (-1)^m Y_{\ell m}^*.$$

With this convention $\int Y_{\ell m}^* Y_{\ell' m'}\,d\Omega = \delta_{\ell\ell'}\delta_{mm'}$, using $d\Omega = \sin\theta\,d\theta\,d\phi$. In particular $Y_{00} = 1/\sqrt{4\pi}$,

$$Y_{10} = \sqrt{3/(4\pi)}\cos\theta,$$

$$Y_{11} = -\sqrt{3/(8\pi)}\sin\theta e^{i\phi}.$$

To check the dipole normalisation independently,

$$\int \cos^2\theta \, d\Omega = 4\pi/3.$$

For Y_{11},

$$\int \sin^2\theta \, d\Omega = 8\pi/3.$$

The complex basis is an accounting device. A real potential obeys $v_{\ell,-m} = (-1)^m v^*_{\ell m}$, leaving $2\ell+1$ real angular degrees of freedom at each order.

On a sphere of radius a suppose segmented electrodes prescribe $f(\theta,\phi)$. Its coefficients are $v_{\ell m} = \int Y^*_{\ell m} f \, d\Omega$. These have units of volts because the angular functions and solid angle are dimensionless. The radial equation $(r^2R')' - \ell(\ell+1)R = 0$ has solutions $r^\ell, r^{-\ell-1}$. For an empty interior, regularity selects $V_{\text{in}} = \sum v_{\ell m}(r/a)^\ell Y_{\ell m}$. For empty exterior decaying at infinity,

$$V_{\text{out}} = \sum v_{\ell m}(a/r)^{\ell+1} Y_{\ell m}.$$

Nonuniform prescribed surface voltage requires electrically distinct electrode patches or a specified ideal boundary; one continuous metal conductor could not sustain that static variation.

Represent a tilted dipole in the laboratory frame

Impose $f = V_0 \hat{\mathbf{p}}\cdot\hat{\mathbf{r}}$ with $\hat{\mathbf{p}} = (\sin\alpha, 0, \cos\alpha)$. In the rotated frame with polar axis along $\hat{\mathbf{p}}$ this is a pure $m=0, \ell=1$ pattern. In the laboratory it is $V_0[\cos\alpha\cos\theta + \sin\alpha\sin\theta\cos\phi]$. Using the explicit harmonics,

$$v_{10} = V_0\sqrt{4\pi/3}\cos\alpha,$$

$$v_{11} = -V_0\sqrt{2\pi/3}\sin\alpha,$$

$$v_{1,-1} = +V_0\sqrt{2\pi/3}\sin\alpha.$$

The sum of their squared magnitudes is $4\pi V_0^2/3$, independent of tilt. The interior solution simplifies to $V_0\hat{\mathbf{p}}\cdot\mathbf{r}/a$, giving uniform $\mathbf{E} = -V_0\hat{\mathbf{p}}/a$. At $a = 40$ mm, $V_0 = 2.0$ V and $\alpha = 30^\circ$, $E_x = -25.0$ V/m, $E_z = -43.3$ V/m. An axis-only fit would infer the correct axial component while losing the transverse one. A full vector field measurement at one interior point distinguishes the rotated dipole from a weaker aligned one.

Tie angular modes to an actual charge distribution

The addition theorem expresses $P_\ell(\hat{\mathbf{r}}\cdot\hat{\mathbf{s}}) = [4\pi/(2\ell+1)]\sum_m Y_{\ell m}(\hat{\mathbf{r}})Y^*_{\ell m}(\hat{\mathbf{s}})$. Rotational invariance explains its structure: the sum is unchanged by simultaneous rotation of both directions, so it depends only on their angle. Put $\hat{\mathbf{s}}$ at the north pole, where only $m=0$ survives, and the proportionality constant follows from $Y_{\ell 0}$ and $P_\ell(1)=1$. Checking $\ell=1$ recovers the dot product of the two unit vectors.

Use the generating series $1/|\mathbf{r}-\mathbf{s}| = \sum_{\ell\ge 0}(r_<^\ell/r_>^{\ell+1})P_\ell(\hat{\mathbf{r}}\cdot\hat{\mathbf{s}})$. For an observation radius outside every source point, define $q_{\ell m} = \int \rho(\mathbf{s}) s^\ell Y^*_{\ell m}(\hat{\mathbf{s}})\, d^3s$. Inserting the addition theorem into the Coulomb integral gives

$$V(\mathbf{r}) = \frac{1}{\epsilon_0}\sum_{\ell,m}\frac{q_{\ell m}}{(2\ell+1)r^{\ell+1}}Y_{\ell m}(\hat{\mathbf{r}}).$$

Unlike boundary voltage coefficients, $q_{\ell m}$ carries units C m$^\ell$. At $\ell=0$, $q_{00} = Q/\sqrt{4\pi}$ and multiplication by Y_{00} returns $Q/(4\pi\epsilon_0 r)$, fixing the prefactor. At $\ell=1, m=0$,

$$q_{10} = \sqrt{3/(4\pi)}\,p_z.$$

The two transverse coefficients reconstruct p_x, p_y. Measuring the voltage coefficient at two radii can distinguish a decaying exterior multipole from a regular potential imposed by distant equipment, whose local expansion instead grows as r^ℓ.

Design an angular quadrature that measures the claimed coefficient

Uniformly spaced latitude rings do not have equal area. In angular integrals substitute $u=\cos\theta$, giving $d\Omega = du\,d\phi$. Equal spacing in u and ϕ already avoids overweighting the polar caps. More accurate quadrature uses Gauss–Legendre nodes in u with their weights, combined with equispaced azimuthal samples for the Fourier factors. If a known angular pattern contains $|m|\le M$, at least $2M+1$ azimuthal samples are needed to distinguish these signed orders on a fixed ring. A higher order present in an unknown pattern can alias into the fitted low orders, so adding more sample angles and reserving independent test directions is informative.

A robust reconstruction checks the constant mode on $f=1$, the total power $\sum|v_{\ell m}|^2 = \int |f|^2 d\Omega$ for an exactly represented pattern, and the predicted angular field at withheld directions. With noisy overdetermined data use a weighted linear least-squares fit to the modes, including weights for sampling geometry or measurement uncertainty as appropriate. The two choices are physically different: one approximates an angular integral, the other quantifies independent noisy observations. State which objective your fit minimises. An angular reconstruction can predict a voltage while still having very poor sensitivity to high-order exterior source moments, whose signals decay rapidly with radius.

A displaced charge supplies an exact all-orders convergence experiment

Place a point charge q at $s\hat{\mathbf{z}}$ within an otherwise empty sphere used as a measurement surface. At radius $R > s$ the potential is $q/[4\pi\epsilon_0\sqrt{R^2 + s^2 - 2Rs\cos\theta}]$. The axial symmetry makes only $m = 0$ coefficients nonzero, giving the exact Legendre series $[q/(4\pi\epsilon_0 R)]\sum_{\ell\geq 0}(s/R)^\ell P_\ell(\cos\theta)$. This example exhibits the convergence parameter directly. It is the radius of the enclosing source relative to the measurement radius, not the magnitude of the charge or the number of sampled angles.

On the ray nearest the point charge, $\theta = 0$, every Legendre polynomial is unity. Truncation at order L there gives a geometric partial sum and a relative error exactly $(s/R)^{L+1}$ when compared with the exact voltage at that point. At $s/R = 0.40$, six retained orders $\ell = 0, \ldots, 5$ leave 0.410% relative error. At $s/R = 0.85$ the same approximation leaves 37.7%. Achieving 1% in the latter case requires $L + 1 > \ln(0.01)/\ln(0.85) = 28.3$, so at least 29 orders. A claim that retaining through quadrupole always suffices has no physical meaning without an observation radius and a desired error.

On the opposite ray $P_\ell(-1) = (-1)^\ell$, making the terms alternate. The absolute voltage error there has a smaller prefactor $1/(1 + s/R)$, in contrast to $1/(1 - s/R)$ on the near side. For this particular geometric series its relative error magnitude is still $(s/R)^{L+1}$, despite the alternating signs. At other angles the angular polynomial oscillations can produce much smaller error but not a uniformly improved bound. The simple inequality $|P_\ell(u)| \leq 1$ gives an absolute bound $|q|(s/R)^{L+1}/[4\pi\epsilon_0 R(1 - s/R)]$ for the entire angular surface. For an arbitrary signed source replace $|q|$ by integrated absolute charge and s by a radius enclosing all source points. A nearly neutral distribution can have a small exact voltage at some angles, making any comparable relative bound much less informative.

Now rotate the point charge through polar and azimuthal angles θ_s, ϕ_s without changing its radius. The coefficient is exactly $q_{\ell m} = q s^\ell Y_{\ell m}^*(\theta_s, \phi_s)$. Summing $|q_{\ell m}|^2$ over m gives $q^2 s^{2\ell}(2\ell + 1)/(4\pi)$, by setting the two angular arguments equal in the addition theorem. This rotational invariant is a strong test of a numerical angular projection: a rotation changes individual complex phases, but the total power at each angular order remains fixed. Rotating a sensor array rather than the source should produce the equivalent opposite rotation of the fitted description.

A spherical split electrode drives an interior field without net charge

Construct a spherical boundary of radius a whose northern half is held at $+V_0$ and southern half at $-V_0$, separated by a thin insulating equatorial seam. In the mathematical limit take $f(u) = V_0 \operatorname{sgn} u$, $u = \cos\theta$. The problem is axisymmetric and changes sign under reflection through the equator, so only odd ℓ occur. Using the Legendre (rather than unit-normalised harmonic) coefficients, the series inside is $V = \sum b_\ell (r/a)^\ell P_\ell(u)$, where $b_\ell = (2\ell + 1)V_0 \int_0^1 P_\ell(u)\, du$ for odd ℓ. There is no monopole, hence no change to the net charge when the two voltages are applied symmetrically in a symmetric environment.

The first two nonzero coefficients can be found without a general table. $P_1 = u$ gives

$$b_1 = 3V_0/2.$$

$P_3 = (5u^3 - 3u)/2$ integrates from zero to one to $-1/8$, so $b_3 = -7V_0/8$. At the centre all terms with $\ell \geq 2$ have vanishing gradient, leaving an exact central field $E_z(0) = -3V_0/(2a)$. Even though the electrode voltages jump abruptly at the equator, a single low-order coefficient determines this point measurement exactly. A finely resolved electrode charge density near the seam requires a very different number of modes and a model of the seam width.

Set $a = 30$ mm and $V_0 = 2.0$ V. The central field is -100 V/m. At $z = 0.20a$ on the northern axis, the $\ell = 1$ voltage contributes 0.600 V and $\ell = 3$ adds -0.0140 V. The first omitted odd term scales as $(0.2)^5$ times its angular coefficient. At $z = 0.9a$ the same rapidly convergent two-term approximation is no longer trustworthy. A student can compare the actual surface voltages estimated from a truncated expansion to predictions at selected interior radii. A converged central field and a conspicuously rounded numerical step can coexist in the same exact analytic model.

For a finite insulating angular seam suppose instead the boundary interpolates continuously and antisymmetrically near $u = 0$. Only a thin band changes relative to the sign function. The dipole coefficient depends on $\int f(u) u \, du$; the extra weight u suppresses the influence of perturbations concentrated around the equator. If the band spans $|u| < \delta$ and the change in potential is bounded by V_0, the dipole coefficient error is at most $3V_0\delta^2/2$. High-order edge charge can nevertheless change markedly. This estimate shows how a real imperfection can have a small impact on one observable and a large impact on another. It also provides a quantitative tolerance when using the ideal split boundary to predict a centre-field calibration.

Separate material inside a measurement sphere from fields imposed outside it

A laboratory sometimes measures voltages on two concentric source-free spherical surfaces, without knowing whether the disturbance comes from an object inside both surfaces or from distant electrodes outside both. A single angular map determines the coefficient of each measured harmonic at that radius but cannot in general distinguish these two origins. In an annulus lacking sources, each angular component takes the radial form $v_{\ell m}(r) = A_{\ell m} r^{\ell} + B_{\ell m} r^{-\ell-1}$. The regular term is compatible with external forcing, and the decaying term with sources inside the inner radius. Both are harmonic in the measurement shell.

For a single $\ell = 1$ component, denote its coefficient of an ordinary $\cos\theta$ pattern by $v(r) = a_1 (r/R_1) + b_1 (R_1/r)^2$, so a_1, b_1 are voltages at the first radius. If two scans give $v(R_1) = 2.00$ V and $v(2R_1) = 1.30$ V, the equations are

$$a_1 + b_1 = 2.00, \; 2a_1 + b_1/4 = 1.30.$$

Multiply the first by one quarter and subtract, giving $a_1 = 0.4571$ V and $b_1 = 1.5429$ V. At a third radius $1.5R_1$ the predicted value is 1.371 V. A pure internal dipole would have returned 0.50 V on the second sphere, while a pure uniform external field would have returned 4.00 V. Fitting a single dipole to the first reading alone would grossly misdescribe the second.

For general order the determinant of the two normalised equations is $t^{-\ell-1} - t^{\ell}$ with $t = R_2/R_1$. As the two radii approach each other, this determinant tends to zero and the coefficients cannot be separated

reliably. Increasing separation improves discrimination of the two radial powers up to the point where one component falls below the measurement noise or additional sources enter the annulus. The monopole case deserves separate attention: its growing branch is just a constant voltage offset, while its decaying branch reveals total charge. A voltage reference common to both scans is therefore indispensable when using their difference to infer monopole charge.

Suppose the field was measured rather than voltage. Radial field coefficients are $-\ell A_{\ell m} r^{\ell-1} + (\ell+1) B_{\ell m} r^{-\ell-2}$. Measuring both radial field and voltage on one sphere can, in principle, supply the two independent equations needed to separate regular and decaying parts. For $\ell = 0$ the constant reference disappears from the radial field. An independently calibrated force probe then constrains total internal charge even when the voltmeter has a common offset. Numerically estimated radial derivatives from two very close voltage scans, however, share measurement errors with those voltages; treating them as two independent precise observations overstates the information gained.

A final issue is the coordinate origin. Moving it can create higher moments from lower ones for a nonneutral distribution. If the actual source is a displaced point charge and the coordinate origin shifts toward it, the series converges faster at a given observation sphere. If the charge distribution is unknown, choosing the origin solely to minimise apparent high-order coefficients can hide genuine structure or amplify noise. Record the physical origin of the measuring geometry, report withheld predictions, and compare the integrated monopole flux, which is independent of the choice of origin as long as the source remains enclosed. This separates a useful coordinate convention from an inferred physical symmetry.

Propagate measured boundary harmonics

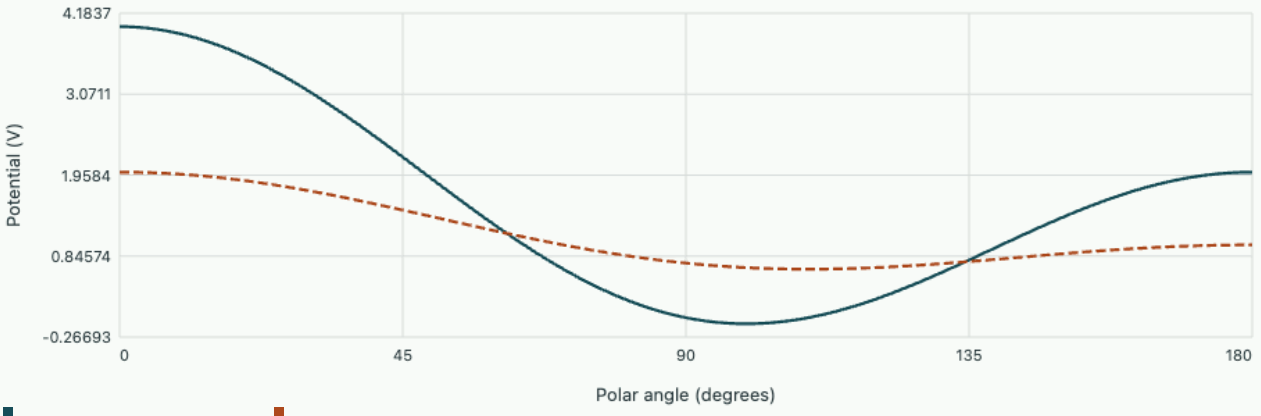
Choose angular boundary data and examine its attenuation away from the spherical surface. Relate the relative loss of detail to harmonic degree.

Open this experiment online

Try the calculation

Observation radius r/a **0.5**

Quadrupole on the sphere (V) **2**



The north pole and equator read 2 V and 0.75 V at this radius. Both branches coincide on the boundary.

Predict an angular sampling requirement before buying more probes

How many angular measurements are needed for a declared band limit?

A real boundary pattern is known, independently of the current measurement, to contain only harmonics with $\ell \le 3$. Count its real degrees of freedom. Design a separable quadrature that exactly projects such a pattern onto the same harmonics in exact arithmetic. Explain what can still go wrong when the true band limit was assumed incorrectly.

▼ **Derive and check the result**

There are $\sum_{\ell=0}^{3}(2\ell+1) = 16$ independent real degrees of freedom. Sixteen arbitrary readings need not determine them: their locations must give a full-rank observation matrix. Equatorial-only data, for example, miss every axisymmetric odd pattern there. A straightforward overdetermined grid uses uniform azimuths and Gauss–Legendre nodes in $u = \cos\theta$. For $L = 3$ use at least $2L + 1 = 7$ equally spaced azimuths so that differences $m - m'$ up to 6 cannot alias a nonzero azimuthal difference into the constant component of a projection.

After that exact azimuthal sum, only equal m terms survive. Their products in u are proportional to $P_\ell^m(u)P_{\ell'}^m(u)$. Each contains $(1-u^2)^{|m|}$ times a polynomial and is itself a polynomial of degree at most $\ell + \ell' \le 6$. A four-point Gauss–Legendre rule integrates polynomials through degree 7 exactly. Thus four latitudes times seven azimuths, 28 weighted samples, suffice to reconstruct all 16 coefficients for the

declared band limit. Both polar caps are represented by quadrature weights; no point must lie precisely on either pole.

If the true field contains order $m = 7$, then $e^{i7\phi} = 1$ at the seven uniformly spaced azimuths starting at zero. Its azimuthal variation can masquerade as a constant on each sampled circle. A second scan shifted by half an azimuthal spacing flips the sign of this harmonic but preserves a true constant, exposing the alias. Unmodelled high latitudinal orders can similarly corrupt low coefficients through the finite radial quadrature. With noisy measurements evaluate withheld off-grid directions and report both the sensitivity to additional modes and the precision assumed in the protected low orders. An exactly integrating grid under a band-limit hypothesis is not a proof that the physical source obeyed that hypothesis.

Tutorial · derive a surface charge and test identifiability

Two spheres agree on their voltage patterns

The same prescribed $v_{\ell m}$ is realised on an infinitesimally thin spherical electrode array between identical dielectric media of permittivity ϵ. Find the coefficient of its free charge density. Which integrated measurement isolates the monopole?

▼ **Work through the solution**

The interior and exterior radial derivatives on $r = a$ are $\ell v_{\ell m} Y_{\ell m}/a$ and $-(\ell+1) v_{\ell m} Y_{\ell m}/a$. With outward normal pointing from interior to exterior, $\sigma_f = \epsilon(E_{\mathrm{out},r} - E_{\mathrm{in},r})$, giving coefficient $\epsilon(2\ell+1)v_{\ell m}/a$. A whole-surface integral of $\sigma_f a^2 d\Omega$ retains only the monopole since all nonconstant harmonics are orthogonal to Y_{00}. For uniform boundary voltage V_0, $v_{00} = \sqrt{4\pi} V_0$ and total charge $4\pi\epsilon a V_0$, the capacitance of an isolated sphere times voltage. High-order voltage components can instead demand strong local charge densities despite zero net added charge.

A ring of detectors misses an entire pattern

Twelve probes placed uniformly on the equator read zero voltage. Can this exclude a substantial $\ell = 1$ component? Give an explicit counterexample and a second configuration still invisible after adding a north-pole probe.

▼ **Work through the solution**

A pure Y_{10} vanishes all along the equator, whatever its amplitude. A north-pole reading would detect it. But the real quadrupole proportional to $\sin\theta\cos\theta\cos\phi$, corresponding to xz/r^2, vanishes on both the equator and at either pole. An intermediate ring or a set of oblique directions is needed. Increasing the number of equatorial samples alone cannot resolve a lack of checks at other latitudes. Under a declared band limit compute the rank and singular values of the probe-by-harmonic matrix before interpreting a null reading as an absent multipole.

Further reading

- Compare interior and exterior harmonics with a polar scan
- Infer a tilted quadrupole from six independent probe readings
- Fließbach, chapter 11 · spherical harmonics
- Griffiths, 5th ed., §§3.3–3.4 · spherical separation and multipoles

Capacitance matrices, floating nodes and controlled approximation bounds

Electrostatic field theory and measurement

A multi-electrode device generally has no single capacitance until the other terminals have been connected or isolated. An energy identity determines the matrix that relates voltages to charges, and a complementary flux identity can certify estimates even when the exact field is unknown.

Build the matrix from unit-voltage experiments

Consider n isolated conductor surfaces embedded in a connected linear dielectric, plus a grounded reference conductor or decay relative to a reference at infinity. There is no free charge in the dielectric. Define the weighting function $w_j(\mathbf{x})$ by putting conductor j at 1 V, all other conductors and the reference at zero, then dividing by 1 V. For arbitrary voltages V_j, linearity and uniqueness give

$$\phi(\mathbf{x}) = \sum_j V_j w_j(\mathbf{x}).$$

The charge on conductor i is an integral of displacement leaving its metal into the dielectric. Using a normal $\mathbf{n}$ pointing out of the dielectric into the metal gives

$$Q_i = \int_{S_i} \epsilon \partial_n \phi \, dA.$$

Therefore

$$Q_i = \sum_j C_{ij} V_j,$$

$$C_{ij} = \int_{S_i} \epsilon \, \partial_n w_j \, dA = \int_\Omega \nabla w_i \cdot \epsilon \nabla w_j \, dV.$$

The last equality follows by multiplying the charge-free field equation for w_j by w_i and integrating by parts. The boundary integral selects just conductor i because its weighting function is unity there and zero on every other conductor. The volume expression proves reciprocity $C_{ij} = C_{ji}$ for a symmetric dielectric tensor. The diagonal terms are positive, since they integrate the squared fields of each independent one-volt experiment. In an ordinary connected isotropic configuration a positive voltage on j with $i \neq j$ grounded puts negative induced charge on i; the corresponding mutual coefficient C_{ij} is nonpositive.

The stored energy is

$$U = \tfrac{1}{2} \sum_{ij} V_i C_{ij} V_j.$$

Positive material response makes the reduced matrix with an effective reference positive definite under usual connected boundary conditions. If all conductors including the external closing conductor are retained without fixing a reference, changing all voltages by a common constant produces no field. The

full matrix then has a zero common-voltage mode and each row sums to zero. Removing the grounded reference from the matrix removes this gauge null direction; its remaining rows need not sum to zero. Treating both statements as universal would create a spurious constraint on measured terminal data.

Convert a field matrix to a circuit measurement

A two-port device might have reduced matrix in picofarads $C = \begin{pmatrix} 12 & -3 \\ -3 & 8 \end{pmatrix}$ relative to its grounded enclosure. With conductor 1 at 2 V and conductor 2 at 0 V, charges are 24 pC and -6 pC. The grounded enclosure takes the compensating -18 pC. Reciprocity predicts that holding conductor 1 at 0 V and conductor 2 at 2 V puts -6 pC on conductor 1. The diagonal 12 pF includes both the coupling to conductor 2 and the path to the enclosure. In an equivalent lumped model the positive pair capacitance between 1 and 2 is 3 pF, while the two separate couplings to ground are 9 and 5 pF. The sign of a Maxwell matrix off-diagonal is opposite to a positive capacitor drawn between two nodes.

Now disconnect conductor 2 after neutralising it. Its total charge stays zero while its voltage becomes an unknown. The second matrix row demands $-3V_1 + 8V_2 = 0$. Thus $V_2 = 3V_1/8$, giving apparent capacitance seen at terminal 1 equal to $12 - 9/8 = 10.875$ pF. Grounding conductor 2 would instead yield 12 pF. The change is a measurable effect of the prepared state, not a change in the dielectric constant of the material. If conductor 2 already carries charge Q_2^0, its voltage acquires a constant offset Q_2^0/C_{22} in addition to the response to V_1.

For several floating nodes write their block equation as $Q_f = C_{fa}V_a + C_{ff}V_f$. If the reference fixes the potential freedom, solve $V_f = C_{ff}^{-1}(Q_f - C_{fa}V_a)$ and substitute into the active-node block. The differential active capacitance is the Schur complement $C_{aa} - C_{af}C_{ff}^{-1}C_{fa}$. It is symmetric and positive for a positive full reduced matrix. With fixed floating charges the constant term shifts active charges but leaves this differential matrix unchanged in a linear medium. An actual cable measurement that results from leaving a neighbouring shield unconnected should be compared to this elimination, not to the original diagonal.

Obtain an upper capacitance bound by trying a voltage

For a two-terminal charge-free dielectric at prescribed voltage difference V, every admissible continuous trial potential satisfying the electrode voltages gives an energy at least as large as the exact electrostatic energy. This follows by writing the trial as exact plus a perturbation vanishing on fixed-voltage electrodes and expanding the energy. The mixed term integrates to zero because the exact displacement is divergence free. If exterior walls are insulated, the exact normal displacement vanishes there and the trial need not be assigned a fixed potential on them. Thus $C \le 2U_{\mathrm{trial}}/V^2$. The fields must genuinely arise as a gradient of one trial voltage. Guessing unrelated components of an electric field does not provide this upper bound.

For a capacitor of plate area A, separation d, and scalar permittivity varying in the gap, choose a linear voltage along the plate normal. The trial energy gives

$$C/A \le \langle \epsilon \rangle / d,$$

where the average is over the gap volume divided by Ad. For dielectric columns that run parallel to the applied field with no change of material along the field direction, the linear potential can be exact. For layers stacked normal to the field it usually overestimates capacitance. The same arithmetic average thus serves as the actual result or a strict upper bound depending on geometry, even when the material fractions are identical.

A complementary trial with a divergence-free displacement

Fix total plate charges $\pm Q$ instead, taking no volume free charge. An admissible displacement trial must have the prescribed flux into each electrode, no flux through insulating side walls, continuous normal component across free-charge-free interfaces, and zero divergence inside. It need not correspond to an exact potential-gradient electric field. Consider $U_D[\mathbf{D}] = \frac{1}{2}\int \mathbf{D}\cdot\epsilon^{-1}\mathbf{D}\,dV$. Write a trial as $\mathbf{D}_* + \delta\mathbf{D}$. The cross term is $\int(-\nabla\phi_*)\cdot\delta\mathbf{D}\,dV = 0$: integration by parts gives zero bulk divergence and each conductor has constant voltage multiplying zero net flux change. Thus $U_D \geq Q^2/(2C)$ and

$$\frac{Q^2}{2U_D^{\mathrm{trial}}} \leq C \leq \frac{2U_V^{\mathrm{trial}}}{V^2}.$$

Unlike two competing estimates, these inequalities identify which side contains the exact answer. For the same parallel-plate geometry choose constant displacement $D_z = Q/A$. It yields $C/A \geq 1/[d\langle 1/\epsilon\rangle]$, the harmonic-mean lower bound. For layers stacked normal to the plates that trial is exact. For columns parallel to the field the upper bound is exact. An arbitrarily shaped two-phase composite satisfies both under the stated plate and insulating-side conditions. This gives a strong physical check before solving detailed microstructure on a mesh.

Use measured numbers to rule out a material model

Take equal volume fractions of relative permittivities 2 and 8 in a cell of area 100 mm² and thickness 1.0 mm. The harmonic mean is 3.2, arithmetic mean 5.0. Since $\epsilon_0 A/d = 0.8854$ pF, the admissible capacitance interval is 2.83–4.43 pF. A measured 3.3 pF is consistent with many internal geometries and does not determine how the two materials are arranged. A calibrated value of 5.2 pF would require examining fringing, area and thickness, parallel cable capacitance, dispersive or conducting response, and the assumed constituent data. Merely refining an interior mesh cannot produce a static value outside the proven interval if it faithfully solves the same ideal model.

For a field computation use the matrix reciprocity relation as one diagnostic and the two variational bounds as another. An electrostatic finite-element potential may produce a reliable energy estimate while numerical differentiation at a sharp electrode edge produces poor charge density. Integrating boundary fluxes gives each terminal charge, but reconstructing C_{ij} by the mixed energy integral often has a different error sensitivity. Comparing both over a grid sequence can indicate whether the error is concentrated at edges or in the field volume. Symmetrising a noisy measured matrix by averaging its transpose removes an antisymmetric discrepancy but cannot establish positivity or explain an inaccurate common scale.

Construct a complete measurable matrix without numerical differentiation

Three thin spherical conductors have radii $a < b < c$. The outer one is tied to the reference potential. Homogeneous material of permittivity ϵ_1 fills the inner annulus $a < r < b$ and another material ϵ_2 fills $b < r < c$. Spherical symmetry gives each annulus its own independent radial displacement field fixed by its electrode voltage difference. Define the ordinary two-sphere gap capacitances $K_1 = 4\pi\epsilon_1 ab/(b-a)$,

$$K_2 = 4\pi\epsilon_2 bc/(c-b).$$

The inner conductor takes $K_1(V_1 - V_2)$. The middle one has two faces, carrying the opposite inner gap charge and an outer gap charge $K_2 V_2$. Thus $Q_1 = K_1 V_1 - K_1 V_2$ and $Q_2 = -K_1 V_1 + (K_1 + K_2)V_2$. Every sign and every term can be read from a physical face.

In vacuum choose $a = 10$ mm, $b = 20$ mm, $c = 40$ mm. Then $K_1 = 2.225$ pF and $K_2 = 4.451$ pF. Driven at 3.0 V with the middle conductor grounded, the inner sphere carries 6.68 pC and the outer gap has zero field. With the middle conductor floating and neutral, $V_2 = K_1 V_1/(K_1 + K_2) = 1.0$ V. The inner sphere now carries 4.45 pC. The inner and outer faces of the floating sphere carry exactly opposite charges, although the conductor has a nonzero voltage. Integrating a current meter attached to the conductor during charging need not tell you each face's charge separately.

Now electrically connect the inner two conductors. Their common voltage sets the inner field to zero. The combined conductor has capacitance K_2 to the exterior, independent of a. This third preparation is distinct from grounding or floating the middle shell. Before calculating an experimental 'device capacitance', draw the electrical connections as well as the geometric cross section. In a larger matrix a driven conductor and a driven guard at the same voltage should have their charges combined only if the experimental ammeter actually measures their total current.

One can also recover the two constituent permittivities from just two carefully selected scalar measurements. Ground the middle shell to determine K_1 directly. Then float it neutral and measure $K_{\mathrm{ser}} = K_1 K_2/(K_1 + K_2)$, giving $K_2 = K_{\mathrm{ser}} K_1/(K_1 - K_{\mathrm{ser}})$. If K_2 is very large, the floating and grounded readings nearly coincide and their difference becomes noise sensitive. An alternative is to drive inner and middle together and measure the outer-gap current directly. The algebra selects a better experimental protocol instead of hiding ill-conditioning in a formula.

Tighten a variational bound without solving the entire field

Take the stacked two-layer capacitor with no space charge, equal layer thickness and isotropic permittivities $\epsilon_1 = 2\epsilon_0, \epsilon_2 = 8\epsilon_0$. A single linear trial produces the arithmetic-mean upper bound. Insert one interior voltage degree of freedom at a point that coincides with the material interface. The resulting piecewise-linear trial space contains the exact solution. Minimising its energy makes the normal displacements equal, reproducing the harmonic-mean capacitance exactly. A global quadratic trial with the same number of extra parameters will usually perform worse because its derivative is continuous while the exact normal field jumps.

To see this quantitatively write $x = z/d$ on the unit interval and $V(x) = V_0[x + \eta x(1-x)]$, with the lower half at relative permittivity 2 and upper half at 8. This trial fixes both endpoints. Its dimensionless capacitance upper estimate is

$$c(\eta) = \int_0^1 \epsilon_r(x)[1 + \eta(1-2x)]^2 dx.$$

Integrating gives

$$c(\eta) = 5 - 3\eta + (5/3)\eta^2.$$

The best quadratic trial has $\eta = 0.9$ and $c = 3.65$, compared with 5.00 for the straight line and 3.20 for the exact layered solution. The unrefined constant-displacement lower trial already supplies 3.20, so the two-field comparison certifies an interval before any field-map visualisation.

Why did the smooth trial still overestimate by 14%? At its endpoints it predicts slopes 1.9 and 0.1 times the mean field magnitude, whereas the exact field has slopes 1.6 and 0.4 in the two halves. Optimising a volume energy distributes a global smooth approximation across both layers. It has no mechanism to jump at the known interface, so an error persists even with its one coefficient optimised perfectly. The example explains why fitting material interfaces in a mesh, or using basis functions that admit correct normal derivative jumps, can matter more than merely increasing polynomial degree.

In an irregular two-dimensional composite an exactly divergence-free complementary trial can be constructed from a stream function Ψ: take $\mathbf{D} = (\partial_y \Psi, -\partial_x \Psi)$ in Cartesian coordinates. Its divergence vanishes identically. Total flux through an oriented path is a difference of Ψ at its endpoints, so carefully chosen boundary values enforce required electrode charges and zero leakage on insulating sides. Optimising $\int \mathbf{D} \cdot \epsilon^{-1}\mathbf{D}/2$ over such trial functions tightens the lower capacitance bound. Optimising the independent voltage trial tightens the upper one. The difference between their resulting bounds quantifies what the chosen approximation spaces still fail to resolve.

What an AC bridge measures when a nominally floating terminal leaks

A real high-impedance lead is neither perfectly floating nor an ideal ground at every frequency. Add a leakage resistance R from conductor 2 of the two-port example to its grounded enclosure. Keep the reduced static capacitance matrix C fixed, linear and quasistatic at the frequencies of interest. Let terminal 1 be driven sinusoidally and terminal 2 have no other external drive. Using $e^{i\omega t}$ here for circuit phasors, charge continuity at terminal 2 gives

$$i\omega(C_{21}V_1 + C_{22}V_2) + V_2/R = 0.$$

The outward conductive current plus the rate of charge stored in the electrode must vanish. Solving for the active terminal admittance gives

$$Y_{\mathrm{in}} = i\omega C_{11} + \frac{\omega^2 C_{12}C_{21}}{1/R + i\omega C_{22}}.$$

At low frequency the leakage holds terminal 2 close to ground, so the leading measured capacitance is C_{11}. At high frequency the floating condition is recovered and $Y_{\mathrm{in}}/(i\omega) \to C_{11} - C_{12}C_{21}/C_{22}$. The effective admittance also has a positive real part, since the driven voltage on terminal 2 supplies Joule heat in the leakage resistance. This term persists even if the dielectric itself is lossless. An AC capacitance meter reporting a frequency-dependent loss tangent might therefore be detecting a boundary preparation rather than intrinsic material loss.

Take $C_{22} = 8$ pF and $R = 100$ MΩ. The crossover $1/(2\pi R C_{22})$ is 199 Hz. At this frequency the quadrature capacitance is halfway between 12.0 and 10.875 pF, namely 11.4375 pF. The corresponding loss conductance divided by ω is 0.5625 pF. An independent verification uses the port voltage ratio at crossover,

$$|V_2/V_1| = |C_{21}|/(\sqrt{2}C_{22}) = 0.265.$$

Compute dissipated power as $|V_2|^2/(2R)$ for peak phasor amplitudes and compare it with $\mathrm{Re}\, Y_{\mathrm{in}} |V_1|^2/2$; the two results must coincide.

The crossover shifts inversely with the leakage resistance while its two limiting capacitances stay fixed in this simple model. Measuring several shunt resistances or adding a known external one provides a direct way to test this interpretation. If a neighbouring shield is connected to a voltage follower instead, its response includes the follower's own gain and phase; static floating-node elimination is insufficient. Extending the field-derived capacitance matrix by experimentally known terminal circuits makes it possible to distinguish changes in electrostatic geometry from changes in what the surrounding measurement wiring enforces.

Close the charge balance at a floating conductor

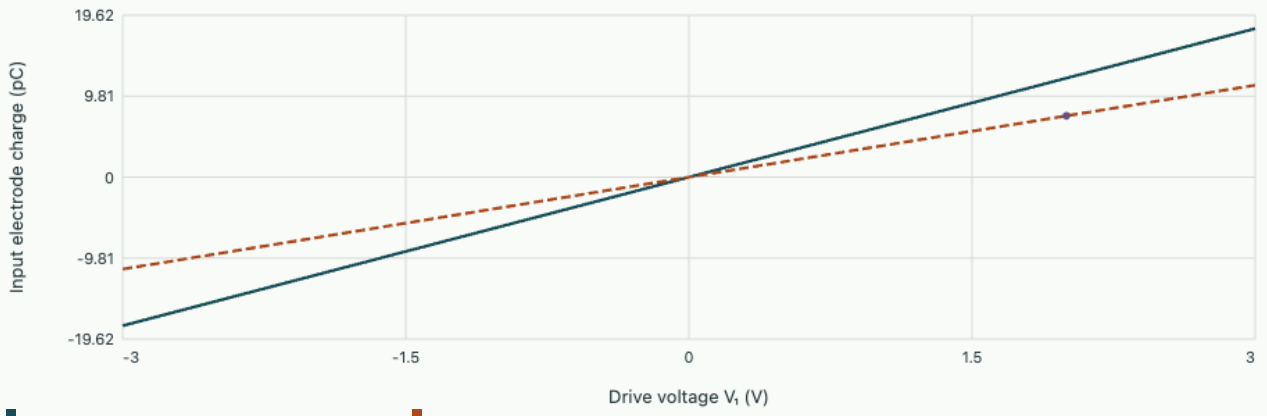
Compare voltages obtained by fixing a conductor potential and by prescribing its net charge. Check the sign of the off-diagonal capacitance terms.

Open this experiment online

Try the calculation

Mutual capacitance (pF) 4

Drive voltage V_1 (V) 2



Second electrode grounded
Second electrode neutral and floating
Selected floating charge on electrode 1

Neutral electrode 2 floats to 1.143 V. Grounded input capacitance 6 pF, floating input capacitance 3.714 pF. Floating-case field energy 7.429 pJ.

Tutorial · distinguish a floating port and an inadmissible trial

What changes after disconnecting a biased guard?

Using the 12, 8 and –3 pF coefficients, start with terminal 1 at 0 V and terminal 2 at 1 V. Disconnect terminal 2, then raise terminal 1 to 2 V. Find the new floating voltage, both charges and the change of active charge.

▼ **Work through the solution**

Initially $Q_2 = 8$ pC, which stays fixed. The new condition is $8 = -3(2) + 8V_2$, giving $V_2 = 1.75$ V. Terminal 1 initially held -3 pC; finally it holds $12(2) - 3(1.75) = 18.75$ pC. Its change is 21.75 pC, or 10.875 pF times the 2 V increment. This reproduces the neutral-floating differential result but a different absolute charge. A charge amplifier reset at the initial state observes the change, whereas a device measuring absolute node potential distinguishes the stored preparation.

Can a lower voltage-trial estimate be trusted?

A potential approximation for the two-phase cell returns 2.70 pF while a constant-displacement trial gives 2.83 pF. The implementer reports the smaller number because its potential looked smoother. Diagnose the inconsistent pair.

▼ **Work through the solution**

For admissible fields on the same problem the flux trial gives a lower capacitance bound and the potential trial an upper one. The reversed interval cannot result from arbitrary coarseness alone. Check that the voltage trial satisfies the exact terminal difference and that all physical material volumes and energy factors are included. Check that the displacement trial has the correct area-normalised total charge and satisfies flux continuity. Finally examine numerical quadrature and unit conversion. Comparing the plots of the two fields does not fix an inadmissible constraint. If both are corrected, their gap gives an actual a posteriori uncertainty interval even without an analytic solution.

Further reading

- Measure a variational energy gap
- Change the geometry and permittivity of a layered capacitor
- Griffiths, 5th ed., §§2.4–2.5 and 4.4 · capacitance and field energy

Reconstructing a source from noisy voltage measurements

Electrostatic field theory and measurement

A forward electrostatic calculation can be unique and stable while the inverse problem constructed from it is unstable or nonunique. A scanned electrode, a finite collecting area and an explicitly chosen noise level make these statements experimentally testable.

Write down what the instrument actually measures

Pattern a long electrode at $z = 0$ with an unknown voltage $f(x)$, periodic over length L for this model. The region above is homogeneous, charge free, and contains a sufficiently weak probe at height h. For a Fourier component $f_m e^{ik_m x}$, $k_m = 2\pi m/L$, the bounded harmonic solution at height h is $f_m e^{-|k_m|h} e^{ik_m x}$. Treat the constant component according to the actual reference. Suppose the probe reports an average over a rectangular window of width w centred at x_j, plus noise e_j. For uniform probe sensitivity the forward relation is

$$y_j = \sum_m H_m f_m e^{ik_m x_j} + e_j,$$

$$H_m = e^{-|k_m|h}\,\mathrm{sinc}(k_m w/2),$$

$$\mathrm{sinc}\, t = \frac{\sin t}{t}.$$

The exponential is physical propagation from boundary to probe; the sinc factor is averaging by the detector. The latter has exact zeros in this idealised model. For modes at those zeros no noise-free reading at the same width and height can recover the lost amplitude. More closely spaced scan positions increase the sampling bandwidth but cannot undo either such instrumental nulls or the exponential decrease of the true high-frequency signal.

A nonuniform probe shape replaces the sinc by its Fourier transform. Unknown height adds an additional calibration problem: increasing h and increasing high-frequency boundary amplitudes can yield the same readings if the amplitudes are unrestricted. Before inferring boundary detail, calibrate the probe width and decide whether height is known, measured with uncertainty or jointly estimated using an independent constraint.

A direct inverse amplifies uncertainty in predictable bands

In an ideal complete periodic scan the discrete Fourier transform separates resolvable modes. Let $\hat{y}_m = H_m f_m + \eta_m$. Inverting one coefficient gives $f_m^{\mathrm{dir}} = \hat{y}_m / H_m$ whenever $H_m \neq 0$. A perturbation $\delta\hat{y}_m$ becomes

$$\delta f_m = \delta\hat{y}_m / H_m.$$

With $H_m = e^{-|k|h}$, the amplification grows exponentially with spatial frequency. The mode values can be uniquely determined in ideal noiseless data of infinite precision, yet the inverse is unbounded as increasingly fine modes are admitted. Finite data can additionally produce exact nonuniqueness.

For $h = 2.0$ mm and negligible aperture, boundary wavelengths 20, 10 and 4 mm have transfers $0.533, 0.285, 0.0432$, respectively. With 1 mV noise in the measured amplitude of each Fourier component, inverse noise levels are $1.87, 3.51, 23.1$ mV. A genuine 10 mV boundary ripple at wavelength 4 mm contributes only 0.432 mV at the probe, smaller than that noise. Plotting its raw inverted coefficient with a smooth line does not improve its information content. Reducing height to 1 mm raises that ripple's signal to 2.08 mV and changes the experiment's resolving power before any reconstruction choice is made.

Derive a regularised inverse and expose its bias

Assume coefficients are inferred separately and use a quadratic penalty. For each m, minimise $J_m = |H_m f - \hat{y}_m|^2 + \alpha^2 w_m^2 |f|^2$ with positive weights w_m and positive regularisation parameter α. Splitting real and imaginary parts, or differentiating with respect to f^*, gives

$$\widehat{f}_m = \frac{H_m^* \hat{y}_m}{|H_m|^2 + \alpha^2 w_m^2},$$

$$W_m = \frac{|H_m|^2}{|H_m|^2 + \alpha^2 w_m^2},$$

$$\mathbb{E}[\widehat{f}_m] = W_m f_m$$

for zero-mean measurement noise. The reconstruction suppresses noise near a small transfer by introducing deterministic bias $(W_m - 1) f_m$. Its variance for known mode-noise variance s_m^2 is $|H_m|^2 s_m^2 / (|H_m|^2 + \alpha^2 w_m^2)^2$. Neither low variance alone nor a small data residual is evidence that a spatial ripple was recovered. Setting α very large produces a smooth, nearly zero estimate regardless of the data. Setting it very small can reproduce noise almost exactly.

For $w_m = 1, \alpha = 0.10$ and wavelength 4 mm in the example above, $W_m \simeq 0.157$. A true 10 mV ripple is estimated at only 1.57 mV on average. The inversion gain is about 3.64 rather than 23.1. At 20 mm, the same penalty retains 96.6% of the true amplitude. If the purpose is to quantify a short ripple, improving the experiment can be preferable to changing the penalty. A derivative penalty $w_m \propto |k_m|$ encodes a stronger preference against fine structure; a constant mode must be handled separately if the voltage reference or the derivative penalty leaves it unconstrained.

What survives away from a uniform grid or simple geometry?

For arbitrary electrode shapes and detector positions, represent source or boundary data in a chosen basis $f = \sum_j a_j b_j$. Solve the forward field problem for each unit coefficient and sample it exactly as the real probe would, giving matrix entries A_{ij}. A noise covariance Σ weights the misfit through $\|\Sigma^{-1/2}(A\mathbf{a} - \mathbf{y})\|^2$. For an identity penalty apply the singular-value decomposition to the whitened matrix,

$$\Sigma^{-1/2} A = U S V^*.$$

Each right singular vector is a source pattern; its singular value tells how strongly it appears in the whitened measurement space. Multiplication by U^* diagonalises the data misfit and reduces the problem to the scalar filters above.

A zero singular value means the chosen sensors are insensitive to an entire source combination. No choice of numerical precision or regularisation changes that observational fact. A small nonzero value means one needs sufficiently small noise and modelling errors to identify the corresponding component. A practical resolution matrix $R = (A^*\Sigma^{-1}A + \alpha^2 I)^{-1}A^*\Sigma^{-1}A$ maps a noiseless true coefficient vector to the regularised estimate. A column of R shows which neighbouring coefficients respond when one ideal source coefficient is present. Testing such impulse responses describes the reconstruction's spatial smoothing more honestly than quoting the spacing of its output grid.

Use withheld measurements and identifiable physical quantities

Reserve some independently measured positions, heights or probe widths. Fit the rest and use the recovered candidates to predict the withheld measurements by applying the forward operator. The test is more informative when the withheld observations have different sensitivity patterns. A second distant scan can verify low modes while remaining insensitive to every high-frequency ambiguity. A second close scan or a different probe width can fill modes attenuated or nulled by the first. Correlated calibration errors, such as one common voltage offset, should be modelled as such, rather than being counted as many independent observations.

A distinct nonuniqueness occurs when trying to infer an arbitrary three-dimensional charge density solely from exterior voltages. Take any smooth compactly supported potential $g(\mathbf{x})$ inside the source region whose value and derivatives vanish outside. Add $\delta\rho = -\epsilon_0\nabla^2 g$ to a candidate charge distribution. Its induced potential is g and therefore vanishes at every external detector. Unless additional sign, support or constitutive assumptions constrain the source, infinitely many internal distributions remain consistent with exact exterior data. This is stronger than amplification of measurement noise. Total charge from a closed-surface flux or low multipole moments can still be identifiable when the internal density is not.

See an ill-conditioned inverse in a two-source instrument

Two possible source patches sit symmetrically beneath two voltage probes. A calibrated forward response, in volts of probe signal per chosen source unit, is $A = \begin{pmatrix} 1 & 0.98 \\ 0.98 & 1 \end{pmatrix}$. Each patch affects both probes almost equally. The sum pattern $(1, 1)/\sqrt{2}$ and difference pattern $(1, -1)/\sqrt{2}$ are orthonormal eigenvectors as well as singular vectors, with gains 1.98 and 0.02. The condition number is 99. The instrument localises total source strength much more reliably than it locates its division between the two patches.

Suppose the actual source vector is $(1, 0)$. The predictions are $(1.00, 0.98)$ V. A measurement perturbation $(+0.01, -0.01)$ V, though small compared with the signal, lies in the weak difference mode. The exact inverse interprets it as source error $(+0.50, -0.50)$, giving a reconstructed distribution $(1.50, -0.50)$. A common perturbation $(+0.01, +0.01)$ instead produces almost negligible source

changes $+0.00505$ each. A single rms error statistic for the voltage readings misses how their spatial correlation changes uncertainty in the reconstruction.

Use an identity Tikhonov penalty $\alpha = 0.10$. The response factors for the two true source directions become

$$W_{+} = 1.98^2/(1.98^2 + 0.10^2) = 0.9975,$$
$$W_{-} = 0.02^2/(0.02^2 + 0.10^2) = 0.0385.$$

In the absence of noise the estimated source from $(1, 0)$ is approximately $(0.518, 0.480)$. Both patches light up almost equally. This is the resolution matrix in action: the model predicts the observed voltages well yet cannot confidently identify which of the two patches generated them. Under the same opposite noise perturbation the additional reconstructed difference is only 0.0192 on each patch in opposite directions. A visually smooth result is reproducible but biased towards nearly equal strengths.

Add a third probe designed to respond as $(0.6, -0.6)$. This linear combination may be implemented physically by a differential electrode, so negative coefficients are meaningful. The squared singular gain of the old difference direction increases from 0.0004 to $0.0004 + 2(0.6)^2 = 0.7204$, while the common gain is unchanged. The new difference singular value is 0.849, reducing the ratio of the largest to smallest singular values to about 2.33 if the three channels have comparable independent noise in the specified units. Adding another copy of the first common-sensitive probe would have been much less effective. Designing measurements against the weak singular vector can improve source resolution far more than collecting many redundant readings.

Separate numerical error from height uncertainty and unknown backgrounds

Return to a sinusoidal source with negligible aperture. If its true height is h but one divides by an assumed transfer computed at $h + \delta h$, then even exact noiseless data return $\widehat{f}/f = e^{|k|\delta h}$. For a 4 mm wavelength and positive height error 0.10 mm, the fractional amplitude bias is

$$e^{0.157} - 1 = 17.0\%.$$

Unlike random noise this bias does not shrink by averaging repeated scans at the same miscalibrated height. A 20 mm calibration pattern at that same error suffers only about 3.2% bias, so testing one long-wavelength standard need not guarantee short-wavelength accuracy.

For an unknown but stable boundary pattern recorded at two known relative probe heights, $\hat{y}_m(h_2)/\hat{y}_m(h_1) = e^{-|k_m|(h_2 - h_1)}$ for modes above noise when the aperture is unchanged. The source amplitude and common aperture factor cancel. This measures a relative vertical displacement but not the absolute height if the true boundary amplitudes are unknown. Add a calibration electrode of known voltage and wavelength, or an independently measured mechanical height, to determine the absolute scale. Near a sensor's aperture zero the ratio is exceptionally noise sensitive and should not be used for the height fit.

Consider also a slowly varying stray background produced by distant conductors. In a source-free layer of finite thickness a Fourier mode can contain both $Ae^{-|k|z}$ and $Be^{+|k|z}$. The growing term is excluded in a bounded half-space continued to infinity, but can represent a field launched from the upper boundary of a finite domain. For a single spatial mode measured at two heights one can solve a 2×2 system for A, B. Its determinant is $2\sinh[|k|(h_2 - h_1)]$, up to a sign fixed by the column order. As the two heights coalesce their rows become nearly dependent. A measurement showing unexpected height scaling may thus indicate the wrong outer boundary in the forward model rather than a faulty inverse algorithm.

A reproducible inverse analysis can be audited along three independent axes. First manufacture noise-free readings from a source whose exact forward voltage is known, then refine only the forward discretisation. Second add noise with a specified amplitude and covariance and repeat enough draws to estimate recovered variation in identified modes. Third perturb physical nuisance parameters such as probe width, height and background independently of the noise draws. For each stage state what quantity is being recovered: total charge, a few spatial coefficients, or a real-space image biased by a chosen penalty. These stages expose different error mechanisms even when they yield similar changes in a plotted voltage trace.

Build an exactly invisible internal charge perturbation

The general construction of a silent source can be made explicit for a spherical support of radius a. Pick a smooth interior potential perturbation $g(r) = V_g[1 - (r/a)^2]^3$ for $0 \le r < a$, and set $g = 0$ outside. At $r = a$ both the potential and its first radial derivative vanish; its second radial derivative also approaches zero, avoiding unmodelled surface sheets. It is regular at the origin. Therefore g itself is the potential relative to infinity of the compact charge distribution $\delta\rho = -\epsilon_0 r^{-2} d(r^2 g')/dr$. With $x = r/a$, expansion gives

$$\delta\rho = \frac{\epsilon_0 V_g}{a^2}(18 - 60x^2 + 42x^4), \qquad 0 \le x < 1.$$

The perturbation is positive near the centre and negative in an outer band if $V_g > 0$. Its enclosed charge is

$$-4\pi\epsilon_0 r^2 g' = 24\pi\epsilon_0 V_g a\, x^3(1 - x^2)^2.$$

It vanishes at both ends of the radius interval and is positive at intermediate radii. Gauss's law outside then gives exactly zero electric field and the chosen reference gives zero exterior voltage. A numerical inverse recovering only zero perturbation from exterior data is observationally consistent with both the absence of charge and this particular structured distribution. Integrating the polynomial with its correct $4\pi r^2 dr$ weight confirms the exact zero total charge independently.

For $a = 20$ mm, $V_g = 0.10$ V, the internal centre voltage differs by 100 mV while every ideal exterior detector, no matter how accurately calibrated, records no difference whatsoever. At $r = a/2$, $g = 0.0422$ V and the enclosed perturbation charge is 9.39×10^{-14} C. An internal measurement can therefore break this particular ambiguity. Moving external probes closer while keeping them outside the sphere cannot. Positivity restrictions on the total density can reduce ambiguity, but adding a sufficiently small signed

perturbation to a sufficiently positive background can preserve nonnegativity of both candidate densities and still leave their exterior data identical.

This example suggests a practical way to report results when full source reconstruction is impossible. Ask whether the quantity of interest changes along a silent perturbation. Total charge and all external moments stay unchanged here, whereas the central voltage and internal charge profile change. A claim about total charge can be appropriate even when detailed shell-by-shell densities are not identifiable. Conversely a fitted compact central peak may depend on the chosen smoothness penalty or positivity constraints even when every reported exterior measurement is reproduced.

A multiscale numerical exercise can demonstrate the effect reproducibly. Discretise both g and its associated $\delta\rho$, solve the independent forward Poisson problem, and evaluate internal and external probes. Coarse grids may show a small exterior signal if source deposition fails to preserve its neutral moments. Refine the source representation, plot the exterior residual separately from the interior voltage difference and compare with the analytic polynomial. A residual that tends to zero while the interior difference remains finite is the expected physical outcome. Using exactly the same discrete operator to manufacture data and to recover them can mask representation errors, so a second mesh, an analytic source integral or a differently implemented forward solver is especially valuable.

See which part of a specimen electrodes can resolve

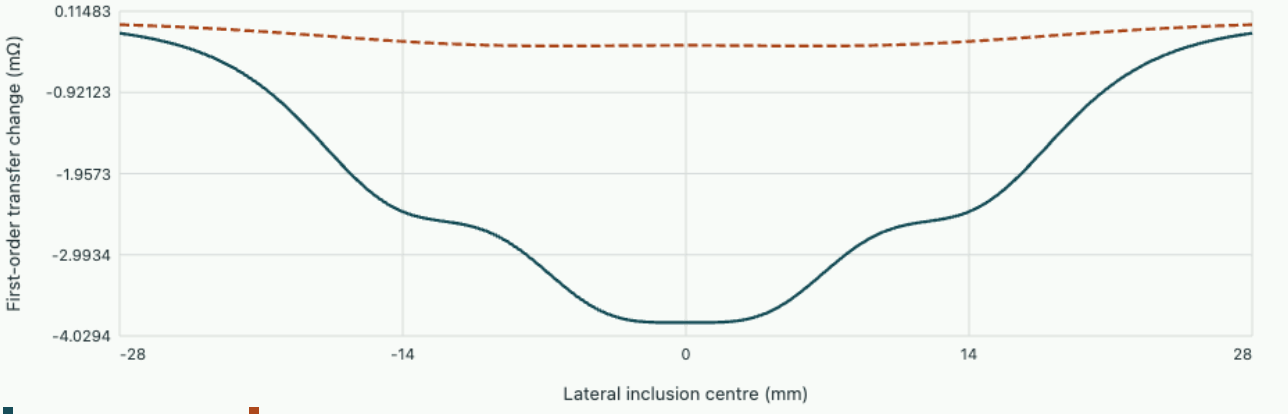
Compare two contact configurations and their sensitivity to a local material change. An accurate potential at the electrodes does not necessarily resolve the interior.

Open this experiment online

Try the calculation

Inclusion depth (mm) 8

Sensing half-spacing (mm) 5



Centre perturbation -3.858 mΩ; unperturbed transfer 79.58 Ω. Sphere radius 1 mm, conductivity increase 0.020 S/m.

Can two independent estimates agree for the wrong reason?

Separate common uncertainty from independent noise in two height scans

A probe measures a 10 mm wavelength mode at heights 1.0 and 2.0 mm with negligible aperture. The two Fourier amplitudes are 5.34 and 2.85 mV, each with independent 0.10 mV random uncertainty, but they share an unknown common voltage gain. Estimate the boundary amplitude and its random error under unit gain. Decide what the agreement can establish about calibration.

▼ **Derive and check the result**

The two transfers are

$$H_1 = e^{-0.2\pi} = 0.5335,$$
$$H_2 = e^{-0.4\pi} = 0.2846.$$

Each individually predicts about 10.0 mV at the boundary. For independent equal-variance amplitude noise, minimise $\sum (y_i - H_i f)^2$. Its solution is $f = (H_1 y_1 + H_2 y_2)/(H_1^2 + H_2^2)$, and random standard deviation $0.10/\sqrt{H_1^2 + H_2^2} = 0.165$ mV. Simply averaging the two directly inverted amplitudes gives the noisier second inversion equal weight and hence a larger variance.

The measured amplitude ratio is about 0.534, as predicted by the additional 1 mm of propagation. A common multiplicative calibration factor cancels from this ratio. If the true voltage gain were 1.10, both unit-gain inversions would agree at about 10 mV even though the physical amplitude would be 9.1 mV. Repeated scans reduce the random uncertainty but cannot remove this systematic scale. A known reference electrode or independent calibrated voltage source is needed.

If instead both probes share a small additive voltage offset in every spatial sample, a full ideal Fourier analysis would place that constant entirely in the zero mode. Incomplete sampling, finite windows or background leakage can mix the offset into the fitted sinusoid. Including a constant nuisance term in a weighted least-squares fit is one way to test that effect directly. Identifying the form of a shared error is as important as quoting its magnitude: common gain, common offset and relative height uncertainty enter the inverse problem in different directions and respond to different calibration experiments.

Tutorial · compare an experimental redesign with a stronger prior

Resolve a window null without changing sample spacing

A voltage camera averages over $w = 4$ mm. A suspected boundary pattern has wavelength 4 mm. Find its measured transfer and propose a second window width that recovers sensitivity. Include the effect of a 1 mm measurement height.

▼ **Work through the solution**

At equal aperture and wavelength the sinc argument is π, giving zero. With a 2 mm window the aperture transfer becomes

$$\sin(\pi/2)/(\pi/2) = 2/\pi.$$

Height attenuation is $e^{-2\pi(1/4)} = 0.2079$, so the combined transfer is 0.132. A 20 mV boundary amplitude therefore gives 2.65 mV in the second scan and exactly zero in the first idealised scan. Collecting more centres with the original width would still give zero. Calibrating the second aperture and its orientation is part of the forward model, especially near a sinc null where small geometry errors strongly change inference.

Choose a regularisation level by its permitted bias

For a mode with transfer 0.20 and unit penalty weight, require the mean reconstructed amplitude to retain at least 80% of the true value. Find the allowed maximum α. For 2 mV noise in the measured coefficient, compare inverse standard deviations with and without this penalty.

▼ **Work through the solution**

Set $W = 0.04/(0.04 + \alpha^2) \geq 0.80$. This gives $\alpha \leq 0.10$. At the limiting value the noise gain is $0.20/(0.04 + 0.01) = 4$, so the reconstructed noise standard deviation is 8 mV. Without regularisation the inverse noise is 10 mV. This modest improvement costs a 20% amplitude bias. The calculated tradeoff may or may not be acceptable for the physical question. Stating both numbers allows a reader to decide whether this experiment can support a claimed amplitude difference.

Further reading

- Measure the transfer of two spatial frequencies
- Reconstruct a tensor from finite electrostatic observations
- MIT, Smith · electric-field inversion and models of interacting electrodes
- Christian Clason · Regularization of Inverse Problems, spectral filters and Tikhonov regularisation

Which part of a sample changes a four-electrode impedance measurement?

Electromechanics, optics and measurement

Driving one electrode pair and measuring another yields a transfer impedance whose spatial sensitivity can be positive, negative or nearly zero. A second field solution supplies the sensitivity without perturbing every element of a numerical mesh.

Distinguish the current terminals from the voltmeter terminals

Consider a bounded two-phase electrical sample in a benign laboratory characterisation cell. Drive harmonic current $+I$ through electrode a and collect $-I$ through b. Two independently selected electrodes c, d read the difference $U_c - U_d$ with negligible voltmeter current. The measured transfer impedance is

$$Z_{cd,ab} = (U_c - U_d)/I.$$

A four-terminal measurement reduces some lead-voltage errors, but finite-area electrodes still alter the field and contact processes can affect it. Reversing either pair reverses the signed transfer reading. Interchanging both pairs leaves it unchanged in a reciprocal linear sample.

Adopt phasors $e^{i\omega t}$ and a local complex admittivity $\gamma = \sigma + i\omega\epsilon$, with real positive scalar σ, ϵ for this first model. The total current density is $\mathbf{J} = -\gamma\nabla u$. In the electroquasistatic regime $\nabla \cdot (\gamma\nabla u) = 0$ away from injected current at the boundary. Magnetic induction and propagation are negligible at the chosen cell dimensions and frequencies; checking electrode and cable dynamics remains necessary even when the sample itself is quasistatic. Perfectly conducting measurement electrodes are equipotentials. Inactive electrodes have zero *net* current, which allows charge redistribution over their finite surfaces; a strictly insulating uncovered wall instead requires zero normal current locally.

Use an outward normal $\mathbf{n}$ from the sample. Define positive terminal current into the sample by $I_i = \int_{S_i} \gamma \partial_n u \, dA$; it equals minus the outward conduction-plus-displacement flux. The pair currents sum to zero, and only voltage differences matter, so fix one reference or a mean voltage. Integrate by parts against any test voltage v constant on each ideal electrode, giving

$$\int_\Omega \gamma \, \nabla u \cdot \nabla v \, dV = \sum_i I_i \, v_i.$$

This is a complex bilinear identity for reciprocity and linear sensitivity. Real average heating is obtained separately by conjugating the physical fields when computing power. Confusing those two products can give an incorrect phase or an incorrect sign for an imaginary admittivity perturbation.

Perturb the sample while keeping injected currents fixed

Let ψ_{ab} be the potential per unit current driven from a to b, and ψ_{cd} the solution when current is instead driven between the measurement electrodes. Using each as test function for the other shows directly that

$Z_{cd,ab} = Z_{ab,cd}$. Now change material admittivity by a small distribution $\delta\gamma$ and let the drive field change by $\delta\psi_{ab}$. Holding all currents and electrode geometry fixed, differentiate the weak equation:

$$\int_\Omega \gamma \nabla \delta\psi_{ab} \cdot \nabla v \, dV = -\int_\Omega \delta\gamma \, \nabla\psi_{ab} \cdot \nabla v \, dV.$$

Choose $v = \psi_{cd}$ and apply the unperturbed measurement-pair equation to $\delta\psi_{ab}$. Its left side becomes the change in voltage difference per drive current, hence

$$\delta Z_{cd,ab} = -\int_\Omega \delta\gamma(\mathbf{x}) \, \nabla\psi_{ab}(\mathbf{x}) \cdot \nabla\psi_{cd}(\mathbf{x}) \, dV.$$

The minus sign has an immediate check in a conducting bar: increasing conductivity lowers its resistance at fixed current. The sensitivity density is the overlap of the drive and reciprocal fields, not simply the magnitude of the drive field. If the fields are locally perpendicular a small isotropic admittivity perturbation is invisible to first order in this channel. If the fields are opposed the local real sensitivity can have the opposite sign to nearby regions. For tensor perturbations replace the product by $-\nabla\psi_{cd} \cdot \delta\boldsymbol{\gamma} \nabla\psi_{ab}$. Reciprocity in the above form requires symmetric material tensors.

A one-dimensional experiment tests the sign, normalisation and scale

Take a uniform real-conductivity bar of length 20 mm, cross-section $A = 100$ mm² and conductivity 0.20 S/m. Assume end electrodes cover its faces and ideal voltage taps select the full length in this limiting four-wire experiment. The impedance is

$$L/(\sigma A) = 1000\ \Omega.$$

The drive and reciprocal gradients per ampere have magnitude $1/(\sigma A)$. Increase conductivity in a full cross-section segment 2.0 mm long by 0.020 S/m. The first-order formula predicts

$$\delta R = -\ell \delta\sigma / (\sigma^2 A) = -10.0\ \Omega.$$

Direct series integration of $1/(\sigma A)$ gives a change from 100.0 to 90.91 Ω in the segment, or $-9.09\ \Omega$.

The 0.91 Ω difference is consistent with truncating at first order for a 10% conductivity change. Recompute at half the material perturbation: the exact change becomes $-4.762\ \Omega$, only 0.238 Ω away from the first-order $-5.00\ \Omega$. The residual decreases by about four when the step halves. Repeating at several positive and negative perturbations and taking a centred difference provides a sharper test of the adjoint code. Varying only the mesh resolution with a fixed large material contrast cannot remove a genuine nonlinear remainder.

The same one-dimensional geometry also demonstrates an absolute identifiability limit. Interchanging two equal-length full-cross-section inserts with different conductivities leaves the end-to-end resistance unchanged, since the integral of reciprocal conductivity is independent of ordering. Perfect measurement

precision cannot recover the order. A voltage tap between the two inserts or an alternative contact geometry is needed to break that symmetry.

Interpret real and imaginary voltage readings together

Even if each local σ, ϵ is real, the fields ψ in an inhomogeneous sample usually carry spatially varying phase. A change $\delta\sigma$ can therefore change both the real and imaginary transfer impedance, as can $i\omega\delta\epsilon$. Treat a complex reading as two real measurements with their calibrated covariance. To first order one finite-element cell with piecewise-constant parameters contributes coefficients $-\int_{\text{cell}} \nabla\psi_{ab} \cdot \nabla\psi_{cd}\, dV$ for $\delta\sigma$ and $i\omega$ times that coefficient for $\delta\epsilon$. The units of the two unknowns differ, so rescale with plausible physical parameter changes before comparing singular values.

As a simple frequency check choose a uniform engineered dielectric-conductor composite with $\sigma = 0.010$ S/m and $\epsilon_r = 1000$. At $f = \sigma/(2\pi\epsilon) \simeq 180$ kHz, $\gamma = \sigma(1 + i)$. The same bar geometry then gives $Z = (L/A)/\gamma = 10.0(1 - i)$ kΩ, a phase of -45°. At low frequency the phase tends to zero and the real impedance to 20.0 kΩ. A constant cable delay produces phase proportional to frequency independent of the sample parameters, so record a calibrated reference at each frequency before attributing an observed phase change solely to the specimen.

Derive a spatial sensitivity that can be checked without a mesh

Consider a homogeneous, thick conducting calibration block with a flat, electrically insulating top except at four small contacts on the same line. Coordinates are x, y along the surface and $z > 0$ into the block. If a unit current enters through a point at the surface, it spreads over hemispheres of radius r. Uniform normal current through a hemisphere requires

$$J_r = 1/(2\pi r^2).$$

The radial field is J_r/σ; integrating from r to infinity gives potential $1/(2\pi\sigma r)$. The coefficient differs from a point source in the interior where an entire sphere carries the current. Subtract a corresponding solution for the terminal extracting the current to satisfy conservation.

Let the drive pair be at $x = \mp a$ and the sense pair at $x = \mp c$, with positive terminal at negative x in each case. At surface points separated from the point contacts the baseline transfer impedance becomes

$$Z_0 = \frac{1}{\pi\sigma}\left(\frac{1}{a - c} - \frac{1}{a + c}\right) \qquad (0 < c < a).$$

For $a = 15$ mm, $c = 5$ mm and $\sigma = 0.20$ S/m this is 79.58 Ω. A 1 mA real current gives 79.58 mV. The same number follows on interchanging the drive and measurement pairs. The large individual contact potentials of an ideal point source are singular at the driving contact itself, yet the transfer potential evaluated at a distinct contact is finite. A two-terminal impedance obtained by setting $c = a$ is divergent in this point-contact approximation. To make such a measurement calculable one must assign physical electrode radii, contact laws and current distributions.

Place a small spherical inclusion of radius 1.0 mm at $x = y = 0$ and depth z, with conductivity just 0.020 S/m above background. At the centreline the vertical field components cancel in each paired solution. Differentiating the two point-source potentials horizontally gives a gradient magnitude $a/[\pi\sigma(a^2 + z^2)^{3/2}]$ for a drive half-spacing a; replace a by c for the reciprocal measurement field. The first-order change in impedance is the product of the two signed gradients times minus the conductivity perturbation and the sphere volume. The centre sample approximates the volume integral when both background gradients vary slowly over the sphere.

At $z = 8$ mm the result is -3.86 mΩ, a $-3.86\,\mu\mathrm{V}$ perturbation at 1 mA, about 48 parts per million of baseline. Increasing depth to 16 mm leaves only -0.320 mΩ. A point midway between one source and one sensing contact can experience a very different gradient alignment from a point on the common centreline. This is why a general inclusion shift should be calculated by differentiating the vector field before taking the dot product. Replacing each gradient by its magnitude would forbid the experimentally possible positive changes. A point inclusion moved twice as deep does not, in general, change the answer by one universal inverse power. Each field samples a different combination of depth and contact separation.

Retain unknown contact impedance as a separate physical parameter

A metal current terminal is nearly equipotential internally but can have an interface voltage drop at the sample. Let U_i be its metal potential, u the potential on the adjacent sample, $j_i = \gamma \partial_n u$ positive into the body and z_i an area-specific contact impedance in Ω m². A simple local interface law is $U_i - u = z_i j_i$, together with $\int j_i \, dA = I_i$. At an inactive metal terminal only this integral is zero. Current can enter on one side and return on the other if the sample potential varies appreciably across its face. A high-impedance voltmeter does not by itself make that metal patch equivalent to an absent electrode.

For sample and electrode test fields (v, W), integration by parts and the contact law give

$$B((u,U),(v,W)) = \int_\Omega \gamma \nabla u \cdot \nabla v \, dV + \sum_i \int_{S_i} \frac{(U_i - u)(W_i - v)}{z_i} \, dA = \sum_i I_i W_i.$$

Here the current and potential gauges are chosen consistently on both drive and reciprocity solves. The contact product is as symmetric as the bulk product, so finite reciprocal contact impedances retain equality upon drive/sense exchange. If the contact changes, differentiating $1/z_i$ introduces a minus, and the minus from differentiating the full field equations cancels it. The additional first-order sensitivity is $+\sum_i \int \delta z_i j_{i,d} j_{i,m} \, dA$, with both current densities normalised to unit terminal current. In a uniform two-terminal contact of area A, the two normalised currents are $1/A$ and one contact contributes $+\delta z/A$, as an ordinary series contact should. This independent limit fixes both the positive sign and the square-metre convention.

For a controlled reconstruction, form one column per candidate sample perturbation and another per uncertain contact coefficient. Weight measured channels by their noise covariance. Two nearly proportional columns then warn that bulk variation and degrading contact cannot be separated by those channels over that bandwidth. Reciprocity catches lead swaps and some drift, but a reciprocal change in

a contact can pass this check. Repeat with a different electrode pattern, or calibrate against a homogeneous reference block at the same frequency. To verify a numerical sensitivity, perturb the contact alone by equal positive and negative amounts and compare the centred impedance difference with the surface integral. Then perturb the sample alone and compare with the volume integral. A successful combined test could otherwise conceal equal and opposite implementation errors.

Move a weak inclusion beneath four surface contacts

The block is treated as a homogeneous half-space except for a 1 mm radius sphere with a positive conductivity perturbation. Each plotted point uses the two unperturbed analytical vector gradients at the sphere centre. The second curve repeats the calculation at twice the selected depth. At 8 mm and 5 mm sensing half-spacing the predicted centre perturbation is −3.86 mΩ against a baseline of 79.58 Ω. Look for lateral positions where the sign changes before interpreting increased conductivity as a uniformly smaller transfer impedance.

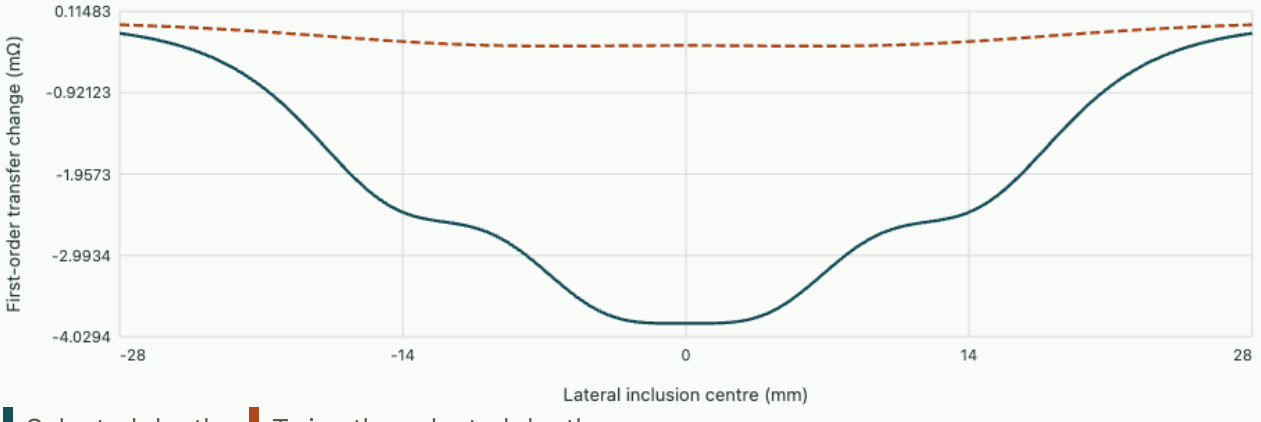
Widen the sensing pair and predict how both the transfer baseline and the local sensitivity will respond. A larger transfer voltage does not guarantee a proportionately larger fractional change for a chosen deep inclusion. The ideal point-contact fields vary steeply close to electrodes. When the inclusion radius approaches the distance over which the gradients change, volume quadrature supplies a better approximation than evaluating a single point. A finite laboratory specimen also needs boundaries sufficiently distant compared with the patterned contact spacing.

Open this experiment online

Try the calculation

Inclusion depth (mm) **8**

Sensing half-spacing (mm) **5**



Centre perturbation -3.858 mΩ; unperturbed transfer 79.58 Ω. Sphere radius 1 mm, conductivity increase 0.020 S/m.

Tutorial · bound a perturbation and design another observation

Predict a transfer change without solving again

A field computation for unit drive and unit reciprocal current in a small inclusion gives approximately parallel real field gradients of magnitudes 1500 and 400 V/(m A). The inclusion volume is 2 mm³. Its conductivity rises by 0.010 S/m. Predict the resistance change and the observed voltage change at a 5 mA drive. What changes if only the measurement electrodes are reversed?

▼ **Work through the solution**

The overlap is $6.0 \times 10^{5}\ \mathrm{V^2/(m^2 A^2)}$. Multiplying by $2 \times 10^{-9}\ \mathrm{m^3}$ and the material perturbation gives $\delta Z = -1.2 \times 10^{-5}\ \Omega$, or -12 µΩ. At 5 mA the transfer voltage changes by -60 nV. Reversing only c, d reverses the reciprocal field and the transfer change. The result assumes nearly uniform overlap inside the inclusion and small perturbation. Subdividing the inclusion or independently perturbing the forward model tests these assumptions differently. A voltmeter's microvolt accuracy would be inadequate for the proposed single measurement.

Do more electrodes resolve two unknown defects?

Two calibrated real channels have sensitivity rows proportional to $(1, 0.8)$ and $(0.8, 1)$ for two dimensionless defect amplitudes. Which combination is measured poorly? Would two additional measurements that simply reverse the existing electrode signs supply a new material constraint?

▼ **Work through the solution**

The common pattern has gain 1.8 and the difference pattern gain 0.2; their singular-value ratio is nine in these units. Distinguishing which defect changed is harder than detecting a common material change. Sign reversals duplicate the same rows up to a minus sign. They can cancel offsets or improve precision by averaging independent noise, but do not increase the rank of the noise-free sensitivity matrix. A drive/measurement pair with opposing sensitivities to the two spatial regions would directly strengthen the weak difference direction.

Further reading

- Derive electrostatic reciprocity from a boundary-value problem
- Check complex field conventions with an independent power balance
- Somersalo, Cheney and Isaacson · existence and uniqueness for the complete electrode model

Designing a displacement bridge at its noise limit

Electromechanics, optics and measurement

A moving electrode changes two capacitances in opposite directions. Derive the displacement signal, reconstruct its noise in metres and select a carrier voltage consistent with measured mechanical stiffness and electrical losses.

Find the voltage of the moving electrode before calculating sensitivity

Consider three plane electrodes with common active area. At the centred position the middle conductor is separated by g from each outer electrode and each gap has capacitance C_0. Displacement x toward the positive drive electrode reduces one gap and increases the other. Use $s = x/g$, where $|s| < 1$. Drive the outer electrodes with coherent sinusoidal voltages $+V_d$ and $-V_d$, with V_d the rms amplitude of either electrode relative to ground. The middle conductor leads to a high-impedance amplifier. Its cable and input provide a fixed total capacitance C_p to ground. Leakage is initially negligible over one carrier period; the input draws negligible signal current. All formulas for carrier voltages in this project use rms amplitudes.

$$C_+ = \frac{C_0}{1 - s},$$

$$C_- = \frac{C_0}{1 + s}.$$

The middle node has zero net charge at the carrier frequency. Its instantaneous ac charge is $C_+(V_n - V_d) + C_-(V_n + V_d) + C_p V_n$. Set the sum to zero and solve the resulting linear equation for V_n. After combining the gap capacitances the result simplifies considerably:

$$\frac{V_n}{V_d} = \frac{C_+ - C_-}{C_+ + C_- + C_p} = \frac{2C_0 s}{2C_0 + C_p(1 - s^2)}.$$

For zero parasitic capacitance the response is exactly $V_d x/g$ throughout the ideal geometry. Each individual capacitance is strongly curved near an electrode. Charge conservation divides one nonlinear combination by another and removes that curvature from their voltage ratio. Additional capacitance reduces the small-signal gain and introduces an odd cubic term. Let $\eta = 2C_0/(2C_0 + C_p)$. Expanding about the centre gives

$$V_n/V_d = \eta s[1 + (1 - \eta)s^2 + \cdots].$$

The derivative at zero is $K = \eta V_d/g$ in volts per metre. A small-signal calibration taken only at the origin implicitly accepts this cubic distortion at larger displacements.

For $C_0 = 2.00$ pF, $C_p = 4.00$ pF, $g = 100\,\mu\mathrm{m}$ and $V_d = 1.00$ V, $K = 5000$ V/m. A displacement of 100 nm produces 0.5000 mV. At 20 µm, the exact signal is 0.10204 V against the linear prediction 0.10000 V. The 2.04% curvature would be apparent in a calibration over the wider range. A single proportional calibration cannot remove it. A fit to the ratio itself has a physical parameter, $C_p/(2C_0)$, which can also be checked by measuring total input capacitance with the mechanical element held at centre.

Translate each readout fluctuation into equivalent displacement

Represent the amplifier near the carrier by an input series voltage noise density e_n and a shunt current noise density i_n. With the drives treated as small-signal grounds for a noise perturbation, the central node presents $C_T = 2C_0 + C_p$ at the centre. The impedance magnitude for current fluctuations is $1/(\omega C_T)$. For mutually independent noise sources and a narrow demodulation band the corresponding input-referred displacement amplitude spectral density is

$$S_{x,e}^{1/2} = \frac{1}{K}\sqrt{e_n^2 + \left(\frac{i_n}{\omega C_T}\right)^2}.$$

Use noise densities and gain referred to the same detection channel. An rms sine amplitude, a peak-amplitude display and a lock-in reading with an unreported normalisation can differ by factors involving $\sqrt{2}$. Here noise figures denote the effective one-sided densities referred to the recovered rms carrier channel. For a white spectrum and an equivalent noise bandwidth B_e the displacement standard deviation equals the density times $\sqrt{B_e}$. The displayed cut-off frequency of a multi-pole lock-in filter need not equal its equivalent noise bandwidth. A calibration using a known injected sideband measures both the gain convention and the actual band response.

At 100 kHz, $e_n = 10\,\mathrm{nV}/\sqrt{\mathrm{Hz}}$, $i_n = 10\,\mathrm{fA}/\sqrt{\mathrm{Hz}}$ and $C_T = 8.00$ pF, current noise produces $1.99\,\mathrm{nV}/\sqrt{\mathrm{Hz}}$. Adding variances gives $10.20\,\mathrm{nV}/\sqrt{\mathrm{Hz}}$, or $2.04\,\mathrm{pm}/\sqrt{\mathrm{Hz}}$ at the stated gain. A 100 Hz equivalent noise bandwidth gives 20.4 pm rms of electrical readout noise. Cable capacitance reduces the voltage made by current noise but also reduces signal gain. In fact $C_T K = 2C_0 V_d/g$, so their ratio for current noise is independent of C_p in this model. The voltage-noise contribution grows with cable capacitance. Higher carrier frequency initially reduces the current contribution but reaches a plateau set by voltage noise.

Derive backaction for a floating sensing plate

Electrical drive can change the mechanical stiffness of the sensor. For each gap take half the squared rms gap voltage times its capacitance derivative. This follows either by averaging the instantaneous plate pressure over a carrier cycle or by differentiating electrical coenergy with the middle node at its charge-balanced voltage. Mechanical motion must be slow compared with the carrier. Since C'_- is negative, the two signed forces cancel at the centre. For a small positive s, write $V_n \simeq \eta V_d s$,

$$C'_+ \simeq (C_0/g)(1+2s),$$
$$C'_- \simeq -(C_0/g)(1-2s).$$

The leading imbalance is

$$F = \tfrac{1}{2}C'_+(V_d - V_n)^2 + \tfrac{1}{2}C'_-(-V_d - V_n)^2 \simeq \frac{2C_0V_d^2}{g^2}(1-\eta)x.$$

The electromagnetic stiffness $K_e = 2C_0V_d^2(1-\eta)/g^2$ subtracts from the mechanical restoring stiffness k. If $C_p = 0$, the equal field magnitudes on both sides cancel exactly for every displacement in ideal plane

geometry. To check directly, the positive gap field is

$$[V_d(1-s)]/[g(1-s)] = V_d/g.$$

The negative gap field has the same magnitude. A charge amplifier holding the central electrode at virtual ground would change this condition and have a different backaction; its calibration must use its own circuit. For a finite C_p, the measured change of resonance with carrier voltage can identify electrical coupling independently of the signal voltage.

With $k = 0.020$ N/m the nominal configuration has $K_e = 0.00020\,\mathrm{N/m}$ at 1 V and $0.00080\,\mathrm{N/m}$ at 2 V. At 2 V the resonance ratio for the same effective mass is

$$\sqrt{1 - 0.0008/0.020} = 0.980.$$

A 2% frequency decrease measures roughly a 4% reduction of stiffness. If the displacement is imposed by a very stiff external positioner this effect changes the force needed to hold position. If the sample is itself a force sensor, the same effect changes the mechanical conversion between applied force and displacement. Quoting an electronic noise floor without specifying which mechanical situation is intended can hide the larger uncertainty.

Choose a feasible drive and include Brownian motion

Raising the carrier at fixed frequency increases displacement signal. It may also heat the dielectrics, exceed amplifier headroom or significantly soften the spring. Suppose each gap at centre has an effective loss tangent $\tan\delta$, with a real conductance $\omega C_0 \tan\delta$ at the carrier. Its dissipated power is conductance times the squared rms voltage. Add both gaps to find

$$P = 2\omega C_0 \tan\delta V_d^2.$$

Fixed parasitic capacitance at the stationary middle node has negligible carrier voltage at exact balance in this model. An actual shield or cable dielectric connected directly to one of the drive nodes will have its own loss. A calorimetric or current phase measurement should include that extra load.

With effective $\tan\delta = 0.010$, the two active gaps dissipate 25.13 nW per squared volt at 100 kHz. A 100 nW heat budget gives $V_d \le 1.995$ V. A driver that remains linear to 2.50 V and a softening allowance $K_e \le 0.1k$ permit 2.50 V and 3.16 V respectively. The thermal bound sets the largest practical drive under these assumptions. For the electrical readout alone, the best feasible setting is thus 1.995 V and the electrical floor is about $1.02\,\mathrm{pm}/\sqrt{\mathrm{Hz}}$. A separate limit on the electric field or measurement amplitude could be introduced as another experimentally justified constraint.

Now let the mechanical element have viscous damping $b = 1.00 \times 10^{-6}\,\mathrm{N\,s/m}$ at $T = 300$ K. In thermal equilibrium the one-sided Brownian force density is

$$\sqrt{4k_B T b} = 1.29 \times 10^{-13}\,\mathrm{N}/\sqrt{\mathrm{Hz}}.$$

Far below mechanical resonance its displacement density is this force divided by $k - K_e$. At 1 V it is $6.50\,\mathrm{pm}/\sqrt{\mathrm{Hz}}$, already greater than the electrical term. At 1.995 V it is about $6.70\,\mathrm{pm}/\sqrt{\mathrm{Hz}}$. The combined density adds in quadrature, giving about 6.81 at 1 V and 6.78 at the heat limit. A minimum of the total can occur between them. For an imposed position of a rigid stage the Brownian motion of this compliant probe may be outside the measurement model; for a freely suspended sensing mass it is part of the signal uncertainty. The optimisation criterion follows the physical quantity the instrument measures.

Use the phase and polarity of a calibration to track false signals

Middle-node leakage to ground with conductance G changes the circuit equation to $[G + i\omega C_T]V_n = i\omega(C_+ - C_-)V_d$ at small displacement. The ideal gain acquires a factor $i\omega C_T/(G + i\omega C_T)$. For a 10 MΩ leakage path and 8 pF, the corner is 1.99 kHz. At 100 kHz the magnitude reduction is about 0.02% and the phase lead about 1.14°. At 2 kHz both quadratures matter. A drift in humidity that lowers the resistance can look like a gain change if only the in-phase channel is recorded. Measure a known displacement at two carrier frequencies and fit both quadratures before interpreting it as a change in geometry.

Suppose the two actual drives are $V_d(1 + a)$ and $-V_d(1 - a)$. At centre the output becomes $\eta V_d a$, equivalent in the small-signal calibration to displacement ag. An amplitude imbalance of one part in ten thousand gives a 10 nm offset at $g = 100\,\mu\mathrm{m}$, many orders above a short-band noise floor expressed in picometres. Reverse the mechanical calibration displacement to isolate odd signal, measure the voltages at the electrode terminals and compare the nulled bridge under exchange of the drive channels. Reversing both carrier phases simply reverses both desired signal and this coherently driven offset; by itself it cannot distinguish them. An offset stable over a short scan can be estimated from two symmetric calibration displacements, whereas a drifting imbalance requires a reference or simultaneous drive measurement.

Find the minimum for a compliant probe instead of guessing at it

It is possible to optimise the free probe analytically up to a scalar root. Write $y = V_d^2$, $K_e = hy$, and a for the readout displacement density at unit drive multiplied by one volt. Let $d_F = \sqrt{4k_B T b}$ be the Brownian force density. At displacement frequencies well below the resonance, the total variance density has an electrical term falling as $1/y$ and a mechanical term that grows as the spring softens:

$$S_x(y) = \frac{a^2}{y} + \frac{d_F^2}{(k - hy)^2},$$

$$\frac{dS_x}{dy} = -\frac{a^2}{y^2} + \frac{2h d_F^2}{(k - hy)^3}.$$

On the stable interval the first derivative strictly increases when $h > 0$. The first term starts at negative infinity and the second diverges at the stiffness boundary. Hence there is one unconstrained minimum. Solve a bracketed scalar equation for y, then compare with each experimental upper bound in volts squared. At zero parasitic capacitance the ideal floating plate has $h = 0$; total noise then falls monotonically toward an asymptotic Brownian floor. A numerical routine that divides by h would fail

precisely at this informative physical limit. In either case, a force-noise specification referred to the external sample can produce a different optimum because force conversion multiplies displacement noise by stiffness.

With the baseline values $a = 2.039 \times 10^{-12}\ \mathrm{m\,V/\sqrt{Hz}}$, $h = 2.00 \times 10^{-4}\ \mathrm{N/(m\,V^2)}$, $d_F \simeq 1.287 \times 10^{-13}\ \mathrm{N/\sqrt{Hz}}$. Substitution brackets the stationary point between 1.45 and 1.50 V. At roughly 1.47 V the density is about $6.72\,\mathrm{pm/\sqrt{Hz}}$, a shallow minimum compared with the heat-limited result. A tiny uncertainty in amplifier noise or in the support damping could shift the exact optimum appreciably while having very little effect on achievable resolution. Reporting a numerically optimised voltage to six decimals would imply a knowledge of the physical system absent from these inputs.

A direct bench experiment can test whether the excess really comes from compliance. Compare the demodulated spectrum with the centre plate tightly held and with it free. The electrical floor should be similar if the clamping keeps cable capacitance and carrier coupling fixed. The free probe should acquire a mechanical resonance which moves with V_d^2, and the low-frequency component should track the inferred static stiffness. A large peak locked to carrier-drive amplitude modulation in both experiments suggests coherent feedthrough. Also check that the noise estimate comes from a frequency band above drift yet below resonance, and that the inferred amplitude density times the square root of the bandwidth predicts repeated stationary readings. These practical checks distinguish a readout-limited displacement estimate from a calibrated measurement of actual thermal motion.

Explore the usable voltage range of the bridge

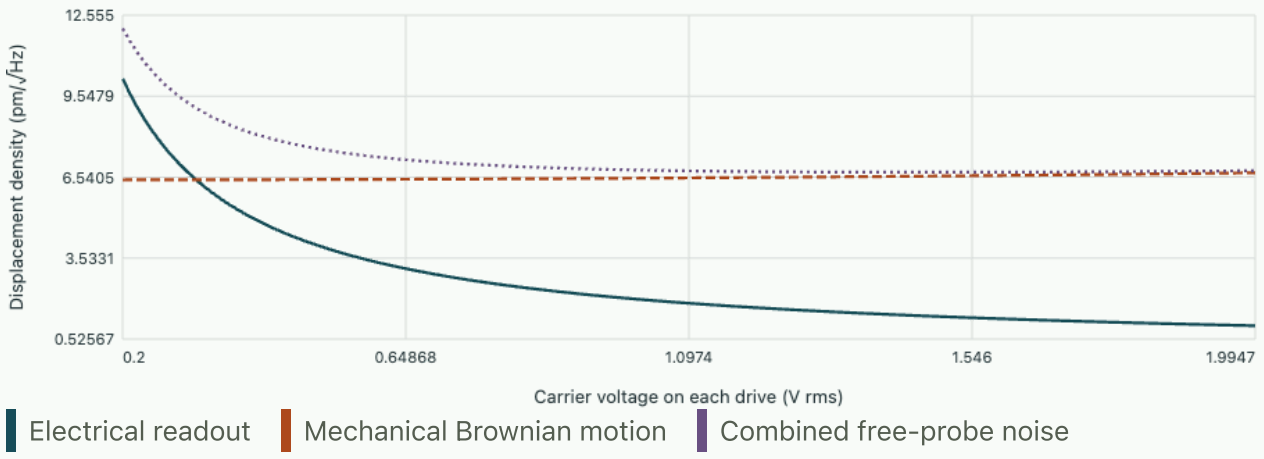
The horizontal axis ends at the first of the heat, driver and mechanical limits. With 4 pF fixed parasitic capacitance at 100 kHz the heat ceiling is 1.995 V rms, while a freely suspended probe reaches its minimum combined noise near 1.47 V rms. Reduce the added capacitance to zero and explain why a finite optimum arising from mechanical softening disappears. Then lower carrier frequency. The current-noise term grows and dielectric heating at a given voltage falls. Observe which effect changes the allowed voltage range and which changes the height of the electrical noise curve.

Open this experiment online

Try the calculation

Input and stray capacitance (pF) 4

Carrier frequency (kHz) 100



Maximum allowed drive 1.995 V rms. Minimum combined noise 6.72 pm/√Hz at 1.472 V rms.

Tutorial · separate electronic resolution from mechanical response

Find the frequency at which current and voltage noise contribute equally

Retain the baseline input capacitances and noise sources. Calculate the crossover frequency and the effect of doubling cable capacitance from 4 pF to 8 pF. What happens to input-referred current-noise displacement at fixed frequency?

▼ **Work through the solution**

Set $i_n/(2\pi f C_T) = e_n$. With $C_T = 8$ pF the crossover is 19.9 kHz. The longer cable gives $C_T = 12$ pF and a crossover at 13.3 kHz. At 100 kHz, the current noise at the input falls to $1.33\,\mathrm{nV}/\sqrt{\mathrm{Hz}}$ but gain falls from 5000 to 3333 V/m. The current contribution remains $0.398\,\mathrm{pm}/\sqrt{\mathrm{Hz}}$. The voltage contribution increases from 2.00 to 3.00 pm/√Hz. Both results follow independently by substituting $K = 2C_0V_d/(gC_T)$ into the noise formula and cancelling C_T in only one of the two terms.

Diagnose a frequency shift and a noise claim

At 2.00 V the resonance changes from 500.0 Hz to 490.0 Hz. Determine the implied stiffness reduction. A second experiment quotes 10 pm rms in a 100 Hz noise bandwidth for the freely suspended probe at 300 K. Can thermal equilibrium and the measured spring and damping account for this number?

▼ **Work through the solution**

At unchanged moving mass $k_{\mathrm{eff}}/k = (490/500)^2 = 0.9604$. The measured electrical stiffness is 0.000792 N/m, consistent with the 0.00080 N/m prediction. At frequencies well below resonance the Brownian density is approximately 6.70 pm/√Hz, and a flat 100 Hz equivalent band would contain approximately 67 pm rms of thermal motion alone. Ten picometres rms would correspond to about 2.2 Hz equivalent noise bandwidth at this density, or require a changed mechanical condition, frequency band or calibration. To discriminate experimentally, measure the mechanical resonance spectrum with a known input-force calibration and integrate the noise with the actual filter transfer. Include mechanical susceptibility over the band if the filter extends close to resonance.

Further reading

- Maxwell stress and electrical coenergy in mechanical force balances
- MIT · Woodson and Melcher, lumped-parameter electromechanics, chapter 3

Least squares and orthogonal factorisation

Numerical linear algebra

Fit a straight line to inconsistent observations, derive its residual conditions and explain why an orthogonal factorisation can preserve accuracy better than the normal equations.

FROM FIRST PRINCIPLES

Derive the orthogonality condition for a fit

With more observations than unknown coefficients, the linear prediction Ax can miss the observed vector b. Suppose squared residuals carry equal weight and choose coefficients to minimize $\|b - Ax\|_2^2$. Try any change $x \mapsto x + td$, where the direction d is fixed but the scalar t can have either sign. Expand with $r = b - Ax$,

$$\|r - tAd\|_2^2 = \|r\|_2^2 - 2t\, d^T A^T r + t^2 \|Ad\|_2^2.$$

At a minimum the linear term must vanish in every direction, so $A^T r = 0$. This says the residual is perpendicular to every vector the model can generate. Rearranging gives the normal equations,

$$A^T A x = A^T b.$$

If A has independent columns then the quadratic term is positive for every nonzero direction and the minimizer is unique. A thin QR factorization with orthonormal columns Q and triangular R yields the equivalent equation $Rx = Q^T b$ and avoids explicitly forming products that can worsen numerical conditioning.

Background · Orthogonal projector

$$P = QQ^{\mathsf{T}}$$

Columns of Q must be orthonormal in the stated real inner product. Least squares removes the component of the data orthogonal to the fitted range.

DERIVATION AND APPLICATIONS

Impose one exact calibration condition on a noisy fit

Some data describe noisy readings while another relationship is an exact conservation or calibration requirement. Model noisy data by $Ax \simeq b$ with independent equal-variance errors, but require $c^T x = d$ exactly. Assume A has full column rank and $c \neq 0$. Write $H = A^T A$, and let x_0 denote the ordinary least-

squares solution. Expanding around x_0 removes the cross term because the residual there is orthogonal to every column of A,

$$\|Ax - b\|^2 = \|Ax_0 - b\|^2 + (x - x_0)^T H (x - x_0).$$

Thus the constrained fit is the closest feasible point in the metric determined by H. Introduce a multiplier λ for the exact condition in half the squared objective. Setting its derivatives to zero and solving the first equation gives

$$H(x - x_0) + c\lambda = 0,$$
$$x = x_0 - H^{-1} c \lambda.$$

Applying the constraint determines the multiplier and the increase in squared residual,

$$\lambda = \frac{c^T x_0 - d}{c^T H^{-1} c},$$
$$\Delta \|r\|^2 = \frac{(c^T x_0 - d)^2}{c^T H^{-1} c}.$$

The denominator is positive. In computation the expression $H^{-1}c$ denotes a solve, preferably using existing QR factors. An already satisfied constraint leaves the fit unchanged. Directions weakly constrained by the measured data can accommodate an exact calibration with little increase in residual.

WORKED EXAMPLE

Reconcile two fitted quantities with an exact total

An ordinary fit gives $x_0 = (2, 5)^T$ with $H = \operatorname{diag}(4, 1)$. The correct total is $x_1 + x_2 = 6$. Now $H^{-1}c = (0.25, 1)^T$, the denominator is 1.25, and $\lambda = 0.8$. The adjusted pair is $(1.8, 4.2)$. Its extra squared residual is $4(0.2)^2 + (0.8)^2 = 0.80$, with the less precisely measured second coordinate absorbing most of the correction.

SECOND APPLICATION

Enforce a zero offset in a calibration line

Three equally weighted readings at $t = 0, 1, 2$ are $1, 3, 5$. The free fit is $1 + 2t$. A separate exact zero calibration demands an intercept of zero. Along that feasible line the slope obeys $(0^2 + 1^2 + 2^2)m = 0 + 3 + 10$, giving $m = 2.6$. The three residuals $1, 0.4, -0.2$ have squared norm 1.20, a directly observable cost of the added calibration condition.

Replace exact consistency by an optimisation problem

Three thermometers at different positions need not lie on a common straight temperature profile. Let the rows of A contain the values of candidate basis functions at those positions and let b contain the

observations. With more independent measurements than unknown coefficients, $Ax = b$ may have no solution. A least-squares fit chooses x to minimise $\|b - Ax\|_2^2$. Equal unweighted terms assume the deviations are comparable in units and importance. If one instrument is much noisier than another, that assumption deserves review before interpreting a fitted slope.

Take any coefficient perturbation sd. Expanding the squared mismatch gives $\|b - Ax\|^2 - 2s\, d^T A^T(b - Ax) + s^2\|Ad\|^2$. At a minimum the coefficient of s must vanish for every d. Hence $A^T Ax = A^T b$. This is also sufficient: when the linear term vanishes, every perturbation adds a nonnegative square. Geometrically the residual is perpendicular to the column space of A. Full column rank makes the minimiser unique, since no nonzero coefficient perturbation then leaves the fitted values unchanged.

Background · When is a fitted residual independent of the fitted coefficients?

Assume the design matrix has full column rank and the observation errors have covariance $\sigma^2 I$. Write the parameter error as $A^+\epsilon$ and the residual as $(I - P)\epsilon$, with $P = AA^+$. Their cross-covariance is

$$\operatorname{Cov}(\hat{x} - x, r) = \sigma^2 A^+(I - P)^T = 0.$$

The last equality follows because $I - P$ maps every vector into the orthogonal complement of the columns of A, where A^+ vanishes. For Gaussian errors the two quantities are jointly Gaussian, so zero cross-covariance also gives independence. With non-Gaussian noise their projections may still be statistically dependent. Orthogonality is always an algebraic property of the least-squares fit; independence needs an assumption about how errors were generated.

Fit three imperfect readings

Take positions $t = -1, 0, 1$ and deviations from a baseline temperature of $1, 2, 5$ K. Fit $b(t) \simeq c + mt$, using a constant and a linear column. Their dot product is zero because the positions sum to zero. Their squared lengths are three and two. Multiplying the observation vector by each column gives 8 and 4, so the normal equations are diagonal and yield $c = 8/3$ K and $m = 2$ K per position unit.

The predicted values are $2/3, 8/3, 14/3$ K. The residual $b - Ax = (1/3, -2/3, 1/3)^T$ has zero sum and zero dot product with $(-1, 0, 1)^T$, as required. Changing the intercept cannot lower the sum of squares because the positive and negative residuals already balance; changing the slope has the analogous weighted balance. The squared residual norm is $2/3$ K². This measures fit to these readings. It does not tell whether a line will predict temperatures well outside the sampled interval.

Remove redundant precision loss

If A has full column rank and singular values σ_i, then $A^T A$ has eigenvalues σ_i^2. Consequently

$$\kappa_2(A^T A) = \kappa_2(A)^2.$$

When two model columns are nearly parallel, forming the products explicitly can lose significant information about their small difference. For example, a straight-line fit on positions clustered around a very large offset has nearly proportional constant and position columns. Centering and scaling positions improves the parameterisation and gives an intercept with a local physical interpretation.

Alternatively, factor $A = QR$ with Q having orthonormal columns and R upper triangular. Decompose the residual into its orthogonal parts inside and outside the column space:

$$\|b - Ax\|^2 = \|Q^T b - Rx\|^2 + \|(I - QQ^T)b\|^2.$$

The second term is independent of the fit. Solve $Rx = Q^T b$ by back substitution. Householder reflectors build this factorisation using length-preserving transformations. In the three-temperature example, dividing the constant column by $\sqrt{3}$ and the position column by $\sqrt{2}$ already produces Q. The transformed data are $8/\sqrt{3}$ and $4/\sqrt{2}$. Dividing by the diagonal of R recovers exactly the same fit. When the columns are nearly dependent, inspect the rank explicitly, using column pivoting or singular values, before giving a physical interpretation to individual coefficients.

Include measurement weights explicitly

If an instrument gives independent errors with standard deviations s_i, divide row i of both A and b by s_i. The sum of squares now measures squared differences in units of each sensor’s own standard deviation. For two readings 0 and 10 with standard deviations 1 and 3, a constant fit minimises $c^2 + (10 - c)^2/9$. Differentiation gives $c = 1$, versus the unweighted mean of 5. A reliable low reading exerts more influence on the weighted result.

This weighting still assumes independence and an appropriate error description. If both readings share an uncertain offset, repeated observations can reduce independent fluctuations without removing the offset. Near a rank deficiency, a small weighted residual likewise cannot establish separate physical identifiability of the fitted coefficients. Examine uncertainty or singular directions in the parameter map as well as the final mismatch.

Derive a prediction after withholding one observation

Let A have full column rank and use equal independent data variances. The fitted vector is $\hat{y} = Hy$ with $H = A(A^T A)^{-1} A^T$; implement the projection with a stable factorisation. Suppose row i is withheld and the reduced design stays full rank. Its new prediction at that location, $\hat{y}_i^{(-i)}$, can be obtained without refitting the entire data set. Replace y_i in the full data by this predicted value. The model fitted to the reduced observations now also satisfies the full normal equations because the added row has precisely zero residual at that fit.

Apply the original hat matrix to the modified vector. Its i-th fitted value is $\hat{y}_i + h_{ii}(\hat{y}_i^{(-i)} - y_i)$, which must equal $\hat{y}_i^{(-i)}$. Rearranging gives

$$y_i - \hat{y}_i^{(-i)} = (y_i - \hat{y}_i)/(1 - h_{ii}).$$

This relation shows how leverage can mask an error inside the training data. If $h_{ii} = 1$, the denominator vanishes because that observation alone identifies some parameter combination. Withholding it then requires a different rank convention or additional modelling information.

Predict the end of a calibration line from its other readings

Fit $y = a + bt$ to $(t, y) = (0, 0), (1, 1), (2, 4)$. The mean time is one and the mean response is $5/3$. The slope from centred data is $[(-1)(-5/3) + (1)(7/3)]/[1 + 1] = 2$, giving intercept $-1/3$. The fitted values are $(-1/3, 5/3, 11/3)$ and residuals are $(1/3, -2/3, 1/3)$. Taken individually, either endpoint looks better fitted than the midpoint. But each endpoint also has disproportionate influence on the slope used to fit itself.

For a straight line the intercept and centred-time columns are orthogonal, so $h_{ii} = 1/3 + (t_i - 1)^2/2$. The endpoint leverages are $5/6$, the central leverage is $1/3$, and their sum is two, the number of fitted parameters. The leave-one-out residuals are consequently $2, -1, 2$. Verify the last directly: the first two readings define the line $y = t$; it predicts 2 at the final time, while the held-out observation is 4. For the central reading the two endpoints define $y = 2t$, predicting 2 instead of 1.

The scientific question determines whether this test is relevant. Dropping an endpoint forces extrapolation, whereas dropping the midpoint tests interpolation between two retained settings. Averaging these three squared errors mixes two distinct prediction tasks. If future measurements will be taken only inside the calibrated interval, a separate evaluation there is more informative. Serially correlated measurements also invalidate the interpretation of neighbouring held-out readings as independent validation, although the algebraic leave-one-out identity for a fixed design remains valid.

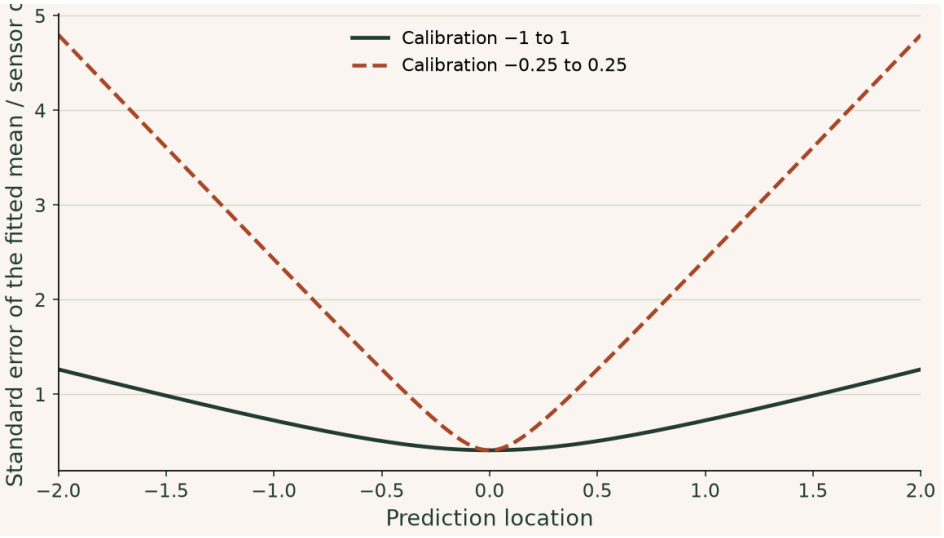
For these three observations the sum of squared training residuals is

$$1/9 + 4/9 + 1/9 = 2/3.$$

The sum after leaving each observation out in turn is $4 + 1 + 4 = 9$. The discrepancy arises even with an exactly implemented fitting routine. Each training prediction has already used the observation against which it is compared. The difference becomes acute near the boundary because that observation has supplied much of the information about its own fitted value.

Estimate a fitted line beyond its calibration interval

Six independent sensors of equal variance calibrate an intercept and slope. For each proposed location use the row $c = (1, x)$. The standard error of the fitted mean is $\sigma \| R^{-\mathsf{T}} c \|$ from the QR factorisation of the design.



Both designs give $1/\sqrt{6}$ times sensor σ at the centre. Narrowing the calibration range by four multiplies the slope standard error by four. To predict an individual future reading add its variance before taking the square root.

▶ **Read values from the plot**

Change parameters · Estimate a fitted line beyond its calibration interval

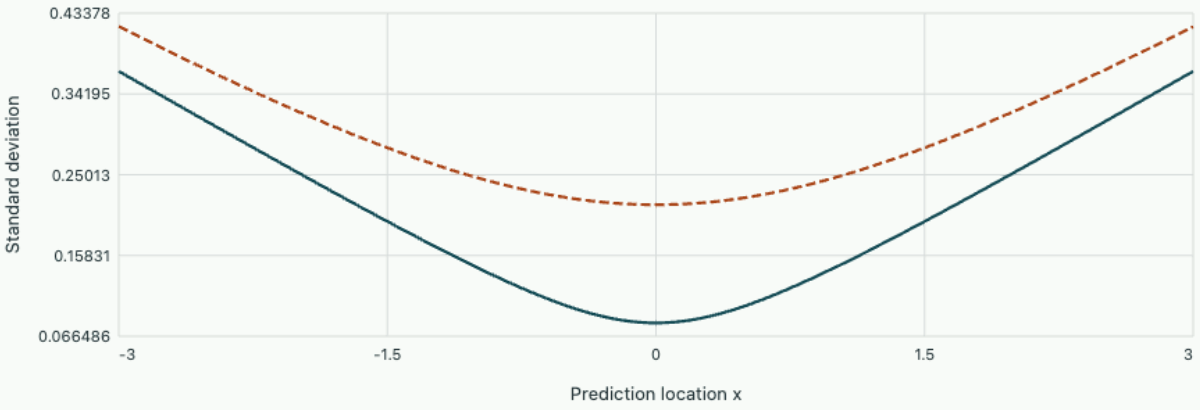
Change one input at a time and compare the recalculated curves with the argument above. The browser evaluates this small model locally; the accompanying numerical values state its assumptions and reference checks. Reset restores the reference inputs, and the data can be downloaded as CSV.

Open this calculation with adjustable inputs

Try the calculation

Calibration half-width **1**

Independent sensor SD **0.2**



Uncertainty of fitted mean Uncertainty of a new reading

Slope SD 0.11952; intercept SD 0.08165. Six equally spaced calibration readings have equal, independent variance.

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 $u_j = Av_j/\sigma_j$. 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

$$Av_i = \sigma_i u_i$$

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

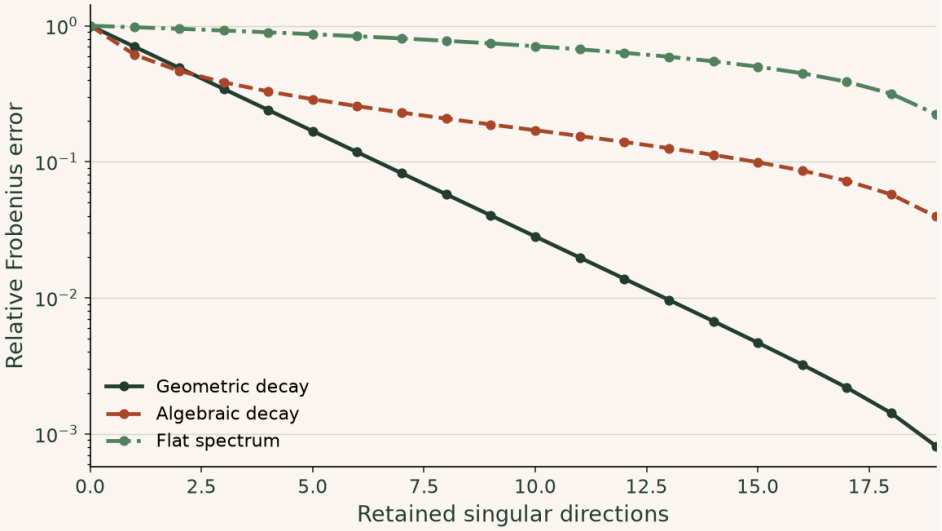
$$3^2 + 1^2 = 10.$$

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

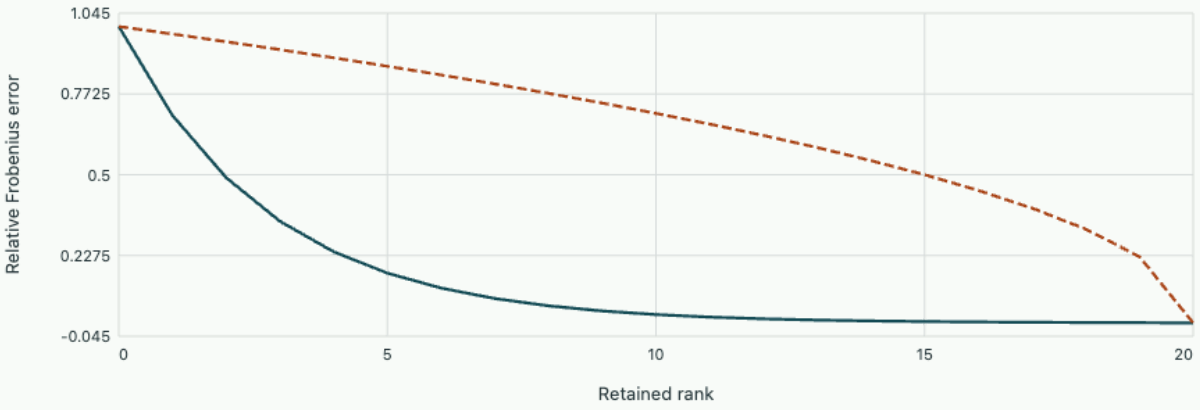
Change one input at a time and compare the recalculated curves with the argument above. The browser evaluates this small model locally; the accompanying numerical values state its assumptions and reference checks. Reset restores the reference inputs, and the data can be downloaded as CSV.

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.

Removing unequal scales before an iteration

Numerical linear algebra

Derive the change of variables behind diagonal preconditioning, then measure what it can and cannot repair in a two-spring system.

Coordinates in the energy

Suppose one displacement is measured in millimetres and the other in metres. A steepest-descent direction depends on the Euclidean lengths of the coordinates. The physical minimum does not. Let A be symmetric positive definite, $D = \operatorname{diag}(A)$, and introduce $y = D^{1/2}x$. Substituting $x = D^{-1/2}y$ in the energy produces

$$\widetilde{A} = D^{-1/2} A D^{-1/2},$$

$$\tilde{b} = D^{-1/2} b.$$

The diagonal of the new matrix is one. Its off-diagonal entry is the original coupling divided by the geometric mean of the two direct stiffnesses. This is a dimensionless measure of interaction. The transformed gradient step maps back to $x_{j+1} = x_j + \alpha D^{-1}(b - A x_j)$.

Choosing a common step

For any eigenvector of the transformed matrix the iteration multiplies the error by $1 - \alpha\lambda$. With eigenvalues in $[m, M]$, equalise the two extreme absolute factors by taking $1 - \alpha m = -(1 - \alpha M)$. This gives $\alpha = 2/(M + m)$ and worst-case contraction $(M - m)/(M + m)$. If m approaches zero the direction barely moves, however accurately the linear algebra is performed.

Two quite different causes of a slow solve

For $A = \operatorname{diag}(100, 1)$, an optimal unscaled step shrinks the error by only $99/101$. Diagonal scaling gives the identity. One step of length one then solves both components exactly. This is the effect of unlike units or stiffness scales.

For a unit-diagonal matrix with coupling 0.95, scaling leaves the matrix unchanged. Its eigenvalues are 1.95 and 0.05. Each optimal step can retain 95% of the error. The nearly free opposing motion is part of the problem itself. A preconditioner that represents the coupling, or a different physical constraint, is needed for faster convergence in that direction.

Compare convergence factors

Vary the stiffness ratio and coupling independently. The curves are worst-case error factors raised to the iteration count. They give bounds attained by an extreme eigenmode with this particular fixed optimal step.

Open this experiment online

Try the calculation

Direct stiffness ratio **25**

Dimensionless coupling **0.3**



Optimal fixed-step contraction 0.93026 before scaling and 0.3 afterwards.

Tutorial on interpreting convergence

If a bound decreases from 1 to 10^{-3} in k steps, how do you determine k from a known factor 0.8?

▼ **Worked solution**

Take logarithms in $0.8^k \le 10^{-3}$. Because $\log(0.8)$ is negative, division reverses the inequality. The smallest integer is 31.

Why should a conjugate-gradient implementation use a positive definite preconditioner?

▼ **Worked solution**

In transformed variables CG minimises an energy for a symmetric positive definite operator. A positive definite preconditioner admits an invertible square root and transfers this property by congruence. An indefinite matrix does not define the corresponding norm.

Read the university treatment

Track which linear system must be solved at each iteration. Compare the cost of a preconditioner with the reduction in the iteration count. Driscoll and Braun · Preconditioning

A measured curve and competing mechanisms

Fit material response, decay and resonant splitting. Design at least one new observation that can distinguish explanations left unresolved by the first experiment.

1. Measure how quickly a dielectric follows an electric field
2. Identify mobile charge with a transverse field
3. Measure equilibrium field fluctuations with a capacitor
4. Move the reference plane of a wave measurement
5. Estimate a cavity linewidth from a finite, biased record
6. Choose a frequency that identifies a coil's stored and dissipated energy
7. Measuring magnetic core loss by a hysteresis loop and by calorimetry
8. Optical trapping and Brownian calibration of a dielectric bead
9. Coupled optical and mechanical resonances
10. Scattering from a central potential and what angular data can recover
11. Which part of a Bell correlation survives imperfect state preparation?
12. Recover a transition frequency from two separated pulses
13. Transport coefficients from equilibrium fluctuations

Measure how quickly a dielectric follows an electric field

Fields and the response of matter

Derive the time and frequency responses of a relaxing polarisation, locate the loss maximum and decide whether a measured sample contains more than one relaxation process.

From a delayed polarisation to a constitutive equation

A slab of insulating material is placed between parallel electrodes. At first the measured charge increases very quickly when a small voltage is switched on. It then continues to rise slowly even though the applied voltage is fixed. A constant real relative permittivity can describe the final charge but cannot predict this record. The experiment reveals at least two relevant time scales. For a measurement window that resolves only the slower process, treat the fast electronic response as instantaneous and describe the slower polarisation P by its approach to an equilibrium value.

At fixed weak field the equilibrium slow polarisation is $\epsilon_0 \Delta\epsilon E$, with $\Delta\epsilon > 0$. Assume the rate of change is proportional to the distance from that equilibrium, with one positive relaxation time τ. Then

$$\tau \dot{P} + P = \epsilon_0 \Delta\epsilon E(t).$$

This is a phenomenological material law. Its assumptions include linear response about the prepared state, a time-independent τ, and a sample whose constituents can be described by one dominant relaxation. Testing those assumptions is part of the experiment.

For $E(t) = E_s$ switched on at zero from an unpolarised state, solve using the integrating factor $e^{t/\tau}$. The result is

$$P = \epsilon_0 \Delta\epsilon E_s (1 - e^{-t/\tau}).$$

After one τ, 63.21% of the slow response has developed; after 3τ, 95.02%. Measuring an absolute fraction requires subtracting the fast initial jump and knowing the eventual plateau. A long baseline drift can look like a second relaxation if those references are not controlled.

Derive the complex dielectric response

Now use peak phasors with time dependence $e^{-i\omega t}$. Differentiation contributes $-i\omega$, so

$$P/E = \epsilon_0 \Delta\epsilon / (1 - i\omega\tau).$$

Include the high-frequency relative permittivity ϵ_∞, where “high” means above the slow process but below additional electronic resonances. The real and imaginary parts of relative permittivity are

$$\epsilon' = \epsilon_\infty + \frac{\Delta\epsilon}{1 + (\omega\tau)^2},$$

$$\epsilon'' = \frac{\Delta\epsilon\, \omega\tau}{1 + (\omega\tau)^2}.$$

At fixed field amplitude the polarisation current is

$$\dot{P} = -i\omega P.$$

The mean electrical work per unit time and volume is

$$\tfrac{1}{2}\mathrm{Re}(E^{*}\dot{P}) = \tfrac{1}{2}\omega\epsilon_0\epsilon''|E|^2.$$

Loss therefore has positive ϵ'' in the chosen time convention. Using $e^{+i\omega t}$ reverses the sign of the imaginary part of permittivity but leaves mean absorbed energy positive when the current convention is changed consistently.

Set $u = \omega\tau$. Differentiating $u/(1+u^2)$ gives $(1-u^2)/(1+u^2)^2$. Hence ϵ'' peaks at $u = 1$, with height $\Delta\epsilon/2$, and the real part lies halfway between its low- and high-frequency plateaux. The peak identified here is in the imaginary *permittivity*. A plot of dissipated power at fixed field includes an additional factor of frequency and need not have a peak at the same location.

Tie a frequency sweep to a step test

Suppose a material has $\epsilon_\infty = 2$, $\Delta\epsilon = 6$, and a measured loss peak at 1 kHz. Then $\tau = 1/(2\pi\,1000) \simeq 159\ \mu$s. At the peak the relative permittivity is $5 + 3i$ in our convention. A time-domain voltage step should show a slow charge transient reaching 63.2% of its added charge in roughly 159 µs. This comparison connects two experiments that weight electrode offsets and parasitic impedances differently.

An independent shape check follows by eliminating u. Let $X = \epsilon' - \epsilon_\infty$ and $Y = \epsilon''$. From the formulas $X^2 + Y^2 = \Delta\epsilon X$. Thus $(X - \Delta\epsilon/2)^2 + Y^2 = (\Delta\epsilon/2)^2$: a single Debye relaxation traces a semicircle in the complex plane. Departures from a semicircle can reflect several relaxation times, spatially inhomogeneous samples, leakage or imperfect fixture correction. They do not uniquely identify the microscopic mechanism.

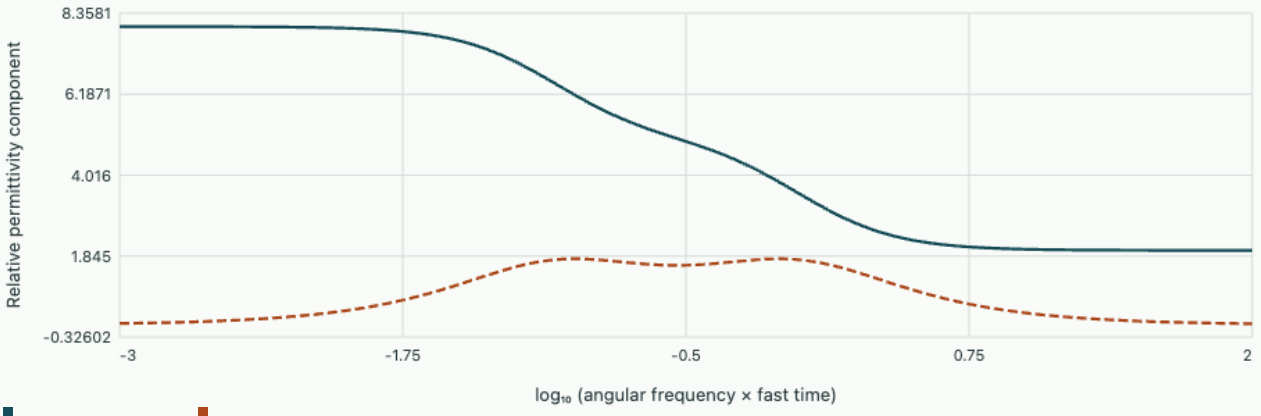
Resolve a second process only when the data require it

Open the live experiment

Try the calculation

Slow fraction of total dielectric strength **0.5**

Slow / fast relaxation time **12**



Largest sampled loss 1.7724 near ωτ_fast = 0.1. Relative permittivity tends to 8 at low frequency and 2 at high frequency.

The synthetic sample has two groups contributing a fixed total $\Delta\epsilon = 6$. Change the fraction in the slower group and its relaxation time relative to the faster group. At equal relaxation times the groups give exactly the single-process response whatever their fractions. When the times differ, the imaginary curves broaden and may separate. Observe the real dispersion steps as well. A broad lossy peak alone can be fitted by more than one pair of times and strengths.

In a practical measurement the low-frequency tail can contain DC conductivity. In the $e^{-i\omega t}$ convention,

$$\sigma E - i\omega\epsilon_0\epsilon E = -i\omega\epsilon_0[\epsilon + i\sigma/(\epsilon_0\omega)]E.$$

This added imaginary part is proportional to $1/\omega$, whereas the low-frequency side of an isolated Debye relaxation rises as ω. Change electrode separation and use a DC conduction measurement to challenge the interpretation. An electrode interface can add a further time scale, so an observed fit parameter deserves an uncertainty and a physical validation step.

Tutorial

The loss maximum is at 2 kHz and its height is 4 in relative-permittivity units. The high-frequency plateau is 3. Predict the low-frequency plateau and the slow response at $t = 100\ \mu$s after a step.

▼ **Work through the solution**

The single-time model gives $\Delta\epsilon = 8$ and the low-frequency plateau 11. The relaxation time is $1/[2\pi(2000)] \simeq 79.6\,\mu$s. The slow fraction after 100 µs is $1 - e^{-100/79.6} \simeq 0.715$. Repeat at another step amplitude to check linearity.

At the loss peak, is the phase difference between the *total* displacement and the electric field always 45°?

▼ Work through the solution

The slow polarisation alone has a 45° phasor angle at $\omega\tau = 1$. The total displacement includes $\epsilon_\infty E$. Its angle is $\arctan[(\Delta\epsilon/2)/(\epsilon_\infty + \Delta\epsilon/2)]$, generally less than 45°.

Why does observing only a high-frequency tail constrain a ratio of parameters more tightly than the individual parameters?

▼ Work through the solution

For large $\omega\tau$,

$$\epsilon'' \simeq \Delta\epsilon/(\omega\tau).$$

Many values with the same ratio $\Delta\epsilon/\tau$ have very similar tails. Measure the peak or both plateaux to distinguish the two parameters.

Compare with material measurements

Inspect the orientational response and dielectric-loss treatment in MIT's material-properties course. Use the phasor convention of each source to predict the sign assigned to loss.

Identify mobile charge with a transverse field

Fields and the response of matter

Derive a Hall-voltage measurement from transverse force balance, reverse current and field to reject offsets and explore what fails when two carrier types conduct.

Choose signs from a drawing

A rectangular conducting sample has its long axis along x, width w along y, and thickness t along z. A uniform magnetic field $B\hat{\mathbf{z}}$ is applied. An external circuit maintains conventional current $I > 0$ in $+x$. High-impedance transverse probes draw negligible current. Define the measured signal as $V_H = V(y = 0) - V(y = w)$. For uniform transverse field this convention gives

$$V_H = E_y w.$$

Reversing the voltmeter leads reverses the reported sign; writing a carrier sign without a lead convention is incomplete.

With one charge carrier species of number density n and signed charge q, the drift velocity satisfies

$$j_x = nqv_x = I/(wt).$$

If the transverse velocity vanishes, its magnetic force component is $-qv_xB$, opposed by qE_y. Thus $E_y = v_xB = j_xB/(nq)$. The resulting port measurement is

$$V_H = \frac{IB}{nqt}.$$

The width cancels for a uniform bar; changing the contact geometry or using a current source with a nonuniform injection profile can spoil that simplification near the terminals. The cancellation also exposes the sign. For electrons $q < 0$, positive conventional current means that the mobile charge drifts in $-x$. The inferred transverse field is negative for positive B in the chosen convention.

Reconstruct density and mobility from separate observations

Take $t = 0.50$ mm, $I = 5.0$ mA, $B = 0.20$ T and a transverse signal of -1.25 mV. The inferred signed Hall coefficient is

$$R_H = tV_H/(IB) = -6.25 \times 10^{-4}\ \mathrm{m^3/C}.$$

Using singly charged electrons gives

$$n = 1/(e|R_H|) \simeq 1.00 \times 10^{22}\ \mathrm{m^{-3}}.$$

A longitudinal voltage measurement, after correcting for probe separation, can yield the material conductivity. If $\sigma = 80$ S/m, the single-carrier mobility is

$$\mu = \sigma |R_H| = 0.050\ \mathrm{m^2/(V\,s)}.$$

These operations use different pieces of information. The longitudinal measurement alone gives only the product $n|q|\mu$. The transverse measurement supplies a sign and, in the one-carrier model, the density independently of mobility. This is an inverse experiment whose two observables have different sensitivities, not a more precise measurement of the same scalar.

One can derive the same result from a stationary Drude equation $0 = q(\mathbf{E} + \mathbf{v} \times \mathbf{B}) - m\mathbf{v}/\tau$. The longitudinal drag gives

$$E_x = m v_x / (q\tau);$$

force balance in y gives the transverse field just obtained. In a homogeneous single-carrier sample, $E_y/E_x = q\tau B/m$, the signed Hall angle tangent. At fixed sample and field, doubling I doubles both voltages in this linear regime and leaves their ratio unchanged. Significant current-dependent apparent density can indicate heating or nonlinear contacts.

Isolate the transverse voltage from a misaligned contact

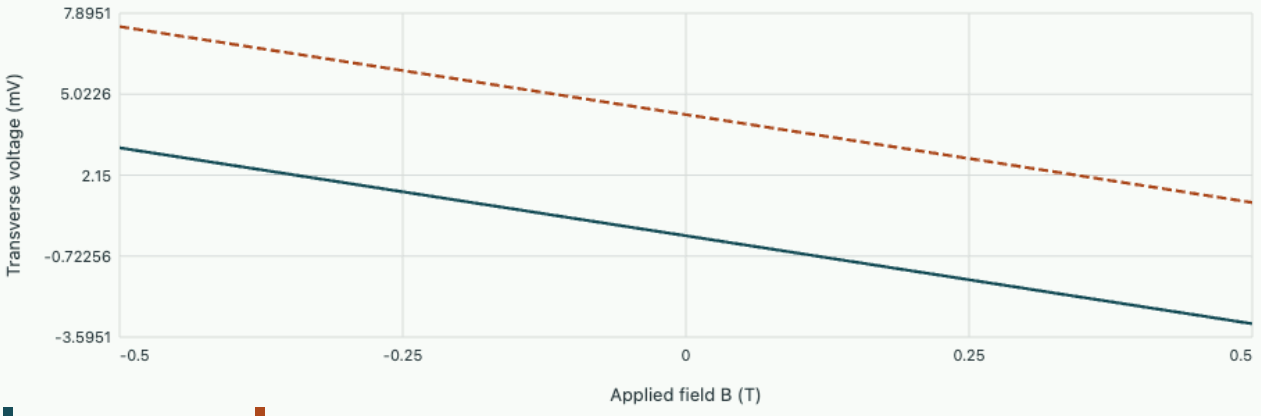
Practical side contacts seldom sit at exactly the same x. A fraction of the longitudinal voltage can leak into the transverse pair. Write the readout at a fixed current as $V_\mathrm{m}(B) = V_0 + \alpha V_{xx} + V_H(B)$. A background that is even in field cancels in $[V_\mathrm{m}(B) - V_\mathrm{m}(-B)]/2$. For example, readings of 3.05 mV and 5.55 mV at $+B, -B$ return -1.25 mV Hall signal and 4.30 mV of even background. Taking a single measured sign would have diagnosed the wrong carriers.

Open the live experiment

Try the calculation

Electron density (10^{22} m^{-3}) 1

Current (mA) 5



At +0.2 T the instrument reads 3.0517 mV; at −0.2 T it reads 5.5483 mV. Half-difference -1.2483 mV.

The plot deliberately superposes a constant 4.3 mV pickup on a signed Hall signal. Vary density and current independently before reading the pair at $B = \pm 0.2$ T. Large density weakens the signal; current increases it. Verify that field reversal recovers the correct transverse voltage even when both instrument readings have the same sign. If the unwanted background changes between readings, reverse the field several times and check for repeatable drift.

With both current and field reversal, the double-odd combination $[V(+I,+B) - V(+I,-B) - V(-I,+B) + V(-I,-B)]/4$ isolates terms odd in both. A thermal emf that is independent of current cancels along with a stable contact offset. This symmetry decomposition is a physical assumption about the parasitics, so preserve all four readings and inspect the rejected channels too.

When a Hall density stops being a particle count

With electrons and holes, longitudinal conductivities add. In a weak magnetic field their transverse contributions also depend on squared mobilities. For positive mobilities μ_e, μ_h and electron and hole densities n, p, the two-carrier weak-field coefficient in the sign convention above is $R_H = (p\mu_h^2 - n\mu_e^2)/[e(p\mu_h + n\mu_e)^2]$. To see where the square appears, note that one factor of mobility converts the longitudinal field into drift; a second converts the resulting transverse magnetic force into a drift contribution. Summing species currents and imposing zero total transverse current gives the denominator management through the summed longitudinal conductivity.

Equal electron and hole densities with equal mobilities give zero Hall signal while current still flows readily. A single-species inversion would suggest unbounded density. At equal densities but unequal mobilities, the faster species can set the sign even though neither species is more numerous. Additional field

dependence, temperature variation and independent structural information are needed before treating a Hall number as a carrier census.

Tutorial

A measurement doubles the width w at fixed thickness, total current and field in a uniform sample. Predict both E_y and V_H.

▼ **Work through the solution**

The current density halves, so v_x and E_y halve. The voltage integrates that halved field over twice the width and stays the same. If fixed current density were held instead, the transverse voltage would double.

A nominal thickness is overestimated by 5%. What bias does this cause in the single-carrier density inferred from a fixed transverse signal?

▼ **Work through the solution**

$n = |IB|/(et|V_H|)$, so using $1.05t$ gives $n_{\mathrm{fit}} = n/1.05$, approximately 4.8% low. This systematic error is common to all repeated field sweeps.

Why might a metal yield a smaller Hall voltage than a semiconductor with the same dimensions at the same current?

▼ **Work through the solution**

In the one-carrier regime the signal is inversely proportional to carrier density. A large density allows the imposed current at a very small drift speed, requiring a correspondingly small transverse balancing field. Conductivity alone cannot determine this comparison without mobility information.

Follow the transport model

Compare the one-carrier assumptions with the treatment of mobile carriers in David Tong's solid-state physics course. Connect an independently measured conductivity with the signed transverse response before interpreting an effective mass or scattering time.

Measure equilibrium field fluctuations with a capacitor

Fields and the response of matter

Derive an electrical noise measurement in two ways, determine what fraction reaches a finite-bandwidth detector and use agreement to calibrate an experiment.

A zero-average voltage can store measurable energy

Connect a capacitor C across a resistor R at a uniform absolute temperature T, with no external drive. Ideal circuit equations predict that any deliberately deposited charge decays on time scale RC. At thermal equilibrium, microscopic fluctuations continually exchange charge and energy with the capacitor. Symmetry gives zero mean voltage, but the stored energy involves V^2. In the classical limit a quadratic degree of freedom has mean energy $k_B T/2$, giving $\langle V^2 \rangle = k_B T/C$. The resistor sets the time scale of equilibration; equilibrium mean-square voltage does not depend on resistance in this ideal model.

Check the dimensions: $k_B T$ is an energy in joules and capacitance has units coulombs per volt. Their ratio is volts squared. At $T = 300$ K and $C = 100$ pF the RMS voltage is $\sqrt{k_B T/C} \simeq 6.44\ \mu\mathrm{V}$. A voltmeter that averages the signed voltage for a long time reports a value near zero. Measuring the noise requires an RMS, variance or appropriately calibrated spectral measurement.

Derive the same variance from a spectral measurement

Use a one-sided voltage power spectral density per hertz at positive frequencies. The classical Johnson–Nyquist open-circuit result for a resistor is $S_R(f) = 4k_B T R$ V²/Hz. Its square root is an amplitude spectral density in V/$\sqrt{\mathrm{Hz}}$, not a voltage variance. Use an equivalent series voltage noise source and the RC transfer function $H(f) = 1/[1 + 2\pi i f RC]$. A stationary linear system multiplies the voltage power spectrum by $|H|^2$, so

$$S_C(f) = \frac{4k_B T R}{1 + (2\pi f RC)^2},$$

\int_0^\infty S_C(f)\,df=\frac{k_BT}{C}.

The integral follows by $u = 2\pi RCf$ and $\int_0^\infty du/(1+u^2) = \pi/2$. It furnishes a detailed check on the factor four, the use of one-sided density, the conversion from hertz to angular frequency and the amplitude-versus-power distinction. If your result disagrees by a constant factor of two or 2π, audit these conventions before assigning a new noise temperature.

For $R = 10$ kΩ, $T = 300$ K, the low-frequency amplitude spectral density is approximately 12.9 nV/$\sqrt{\mathrm{Hz}}$. The RC corner at $C = 100$ pF is $f_c = 1/(2\pi RC) \simeq 159$ kHz. The open-circuit low-frequency density rises with R while the bandwidth shrinks in inverse proportion. Their integrated product gives the earlier resistance-independent variance.

Include a real observation bandwidth

Suppose a first measurement adds an ideal low-pass cutoff B. The recorded fraction of equilibrium variance is $F = (2/\pi)\arctan(B/f_c)$. Half of the noise variance lies below the RC corner. That may be surprising if the corner was introduced as the frequency at which the *spectral density* reaches half its zero-frequency value. These two properties coincide for this single-pole spectrum for different mathematical reasons.

At $B = 0.1 f_c$, $F \simeq 0.06345$, so the recorded RMS is about 25.2% of the infinite-band RMS. At $B = 10 f_c$, $F \simeq 0.93655$, giving about 96.8% of the RMS. In an experiment, taking a square root one step too early changes this comparison. A real detector transfer function replaces the sharp cutoff by an integral of $S_C(f)|H_{\mathrm{det}}(f)|^2$.

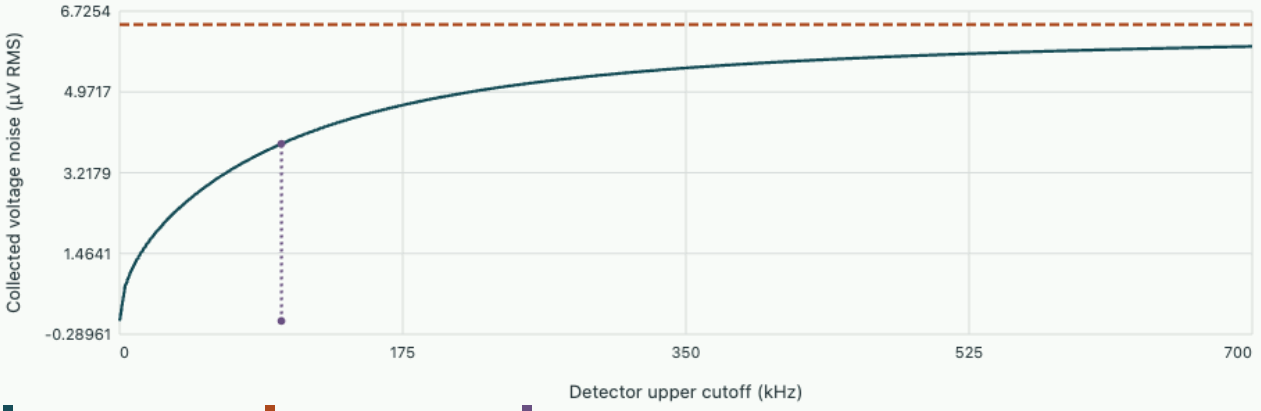
Open the live experiment

Try the calculation

Resistance (kΩ) **10**

Temperature (K) **300**

Detector bandwidth (kHz) **100**



RC corner 159.15 kHz. Detector reads 3.8461 µV RMS, collecting 35.713% of equilibrium variance.

Sweep the resistance by a factor ten. Observe how the corner moves along the frequency axis and how the fraction measured in a fixed bandwidth changes. Then vary temperature. In the classical regime spectral density and integrated variance are proportional to T, while RMS voltage grows only as $\sqrt{T}$. A plot labelled “noise” without units cannot distinguish these two dependencies.

Background · How does a longer voltage exposure alter the measured noise?

An equilibrium resistor–capacitor circuit has voltage covariance $C_v(s) = (k_B T/C)\exp(-|s|/\tau)$, where $\tau = RC$. Suppose a digitiser returns the mean voltage during an exposure T_m. Squaring a sample of this mean measures a different variance from averaging the instantaneous voltage squared. In the classical stationary model the two-time integral becomes

$$\operatorname{var}(\bar{v}) = \frac{2}{T_m^2}\int_0^{T_m}(T_m - s)C_v(s)\,ds = \frac{2k_BT}{C}\,\frac{u - 1 + e^{-u}}{u^2},$$

$$u = \frac{T_m}{\tau}.$$

Short exposures recover k_BT/C. At ten correlation times only about 18% of that variance remains. For very long exposures the variance decreases as $2\tau/T_m$. Taking more closely spaced samples of one exposure cannot undo that averaging.

A physical check and a calibration check

An RC time-domain experiment provides a second observable beyond integrated power. Under the same single-pole equilibrium model the voltage autocovariance is

$$\langle V(t)V(t+\tau)\rangle = (k_BT/C)e^{-|\tau|/(RC)}.$$

Its variance at zero lag agrees with equipartition, and its decay time is obtained from the homogeneous charge-relaxation equation. Repeating the measurement with a known extra parallel capacitance lowers the RMS and lengthens the correlation time. Seeing only one of those changes suggests that the observed noise may include substantial amplifier input voltage or an independent background.

The classical approximation assumes $hf \ll k_BT$ over the relevant band and a resistor that can be treated as frequency independent. A MHz room-temperature experiment is far inside this regime. At sufficiently high frequencies or cryogenic temperatures, a quantum noise model, an explicit definition of the measured ordering of the spectrum and the detector's energy exchange are needed. Extending an ideal classical white source to infinite frequency is only permissible as an approximation inside the convergent low-frequency RC calculation; it is not a microscopic model at arbitrarily high frequency.

Tutorial

At fixed C, T, doubling R in a detector with $B \ll f_c$ changes the measured RMS by what factor? Will the same factor apply when $B \gg f_c$?

▼ **Work through the solution**

At very small bandwidth, $S_C \simeq 4k_BTR$ is nearly constant, so variance doubles and RMS increases by $\sqrt{2}$. At large bandwidth the measurement approaches k_BT/C and becomes independent of R.

A two-sided spectrum in hertz is used instead. What density and integration limits preserve the same variance?

▼ **Work through the solution**

For this real voltage, the density for each sign of frequency is half the one-sided positive-frequency value. Integrate it over $(-\infty, \infty)$. Integrating the one-sided density over both signs would count the power twice.

A measured RMS is 10 µV, including independent amplifier noise of 6 µV RMS. Can 6 µV simply be subtracted to infer the sample contribution?

▼ **Work through the solution**

Independent zero-mean voltages add in variance. The inferred sample RMS is

$$\sqrt{10^2 - 6^2} = 8\ \mu\text{V}.$$

A deterministic offset or correlated noise would require a different treatment.

Compare with an actual laboratory

The MIT Junior Lab noise experiment treats the detector as part of the measurement. Identify its calibration observable before comparing the theoretical density with an instrument reading.

Move the reference plane of a wave measurement

Numerical field laboratory

Recover a load reflection coefficient from a noncontact voltage measurement and explain why identical amplitude data can represent physically different terminations.

Specify which plane defines the load

A uniform lossless transmission line has real characteristic impedance $Z_0 = 50\ \Omega$. Drive it sinusoidally at one frequency. Choose peak phasors with time factor $e^{i\omega t}$ and measure distance d from the load towards the source. Incident and reflected waves move in opposite directions. Taking the load as phase origin, the ratio of the reflected to incident voltage at the load is

$$\Gamma_L = (Z_L - Z_0)/(Z_L + Z_0).$$

The ratio at a sensor plane a distance d away is $\Gamma_{\mathrm{in}} = \Gamma_L e^{-2i\beta d}$. The factor two is the phase difference of two opposite travellers. Moving the sensor does not change a physical resistor.

For a real resistive load, Γ_L is real and ranges from −1 for a short to +1 for an open. A 75 Ω load gives +0.2 and a $100/3$ Ω load gives −0.2. These loads have the same reflected power fraction of four per cent. Distinguishing them requires phase information or positioning the sensor at more than one known location.

Derive the envelope and the ambiguity

The local voltage is the sum of incident and reflected phasors. Factor out the incident phase and take the magnitude:

$$\frac{|V(d)|^2}{|V^+|^2} = 1 + |\Gamma_L|^2 + 2\,\mathrm{Re}\{\Gamma_L e^{-2i\beta d}\}.$$

Its maxima and minima have amplitude ratios $1 + |\Gamma|$ and $1 - |\Gamma|$ for a passive load, giving $\mathrm{VSWR} = (1 + |\Gamma|)/(1 - |\Gamma|)$. The distance between successive maxima is $\lambda/2$, because the measured intensity involves a relative wave number 2β. A measured ratio 1.5 yields $|\Gamma| = 0.2$ but leaves its sign and, for general impedances, its phase undetermined. Extremum positions contain information that a ratio discards.

At one eighth of a wavelength from a 75 Ω load the phase rotation is $-\pi/2$, giving $\Gamma_{\mathrm{in}} = -0.2i$. If a researcher plugs that complex coefficient directly into $Z = Z_0(1 + \Gamma)/(1 - \Gamma)$, the inferred impedance is about $46.15 - 19.23i\ \Omega$. That is the correct input impedance at the sensor plane. To infer the actual element multiply back by $e^{2i\beta d}$ before the inversion. The recovered load is then 75 Ω with zero imaginary part.

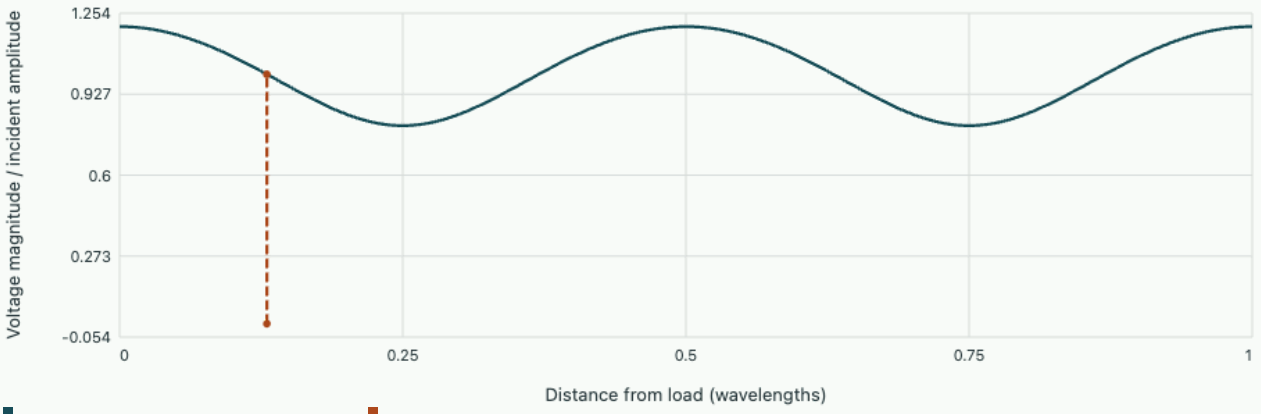
Put a probe between voltage extrema

Open the live experiment

Try the calculation

Load resistance (Ω) **75**

Probe plane from load (wavelengths) **0.13**



Standing-wave envelope Selected reference plane

Load Γ = 0.2. Probe-plane phase -93.6°. Voltage standing-wave ratio 1.5.

Set the load to 75 Ω and move the sensor from the load through a quarter wavelength. The envelope changes from 1.2 to 0.8 times the incident amplitude. At $\lambda/8$ it is $\sqrt{1+0.2^2}$. Centre your first trial on a voltage extremum and estimate how much the reading changes for a small positioning error. Repeat around the steepest part of the envelope. A high-slope position is sensitive to phase; a stationary amplitude extremum is useful when estimating the reflectance magnitude.

Use the matched case next. The field envelope becomes flat. The program reports that reflection phase is undefined. A zero complex amplitude has no observable phase angle. An instrument that prints a quickly varying phase for a nearly matched load can simply be reporting the angle of noise around the origin. Propagate uncertainty in real and imaginary parts before interpreting such a number.

Recover a complex reflection from several amplitude measurements

If $|V^+|$ is independently calibrated, sample squared voltage at $d = 0, \lambda/8, \lambda/4, 3\lambda/8$. Denote normalised intensity by I_d. Writing $\Gamma = x + iy$ gives $I_0 - I_{\lambda/4} = 4x$ and $I_{\lambda/8} - I_{3\lambda/8} = 4y$ with the convention above. Constant detector offsets cancel in these differences. A multiplicative gain error remains and must be calibrated. A mismatch in the assumed wavelength rotates or distorts the reconstructed coefficient rather than merely adding random noise.

For $\Gamma = 0.2$, the four intensities are 1.44, 1.04, 0.64 and 1.04. The reconstruction returns $x = (1.44 - 0.64)/4 = 0.2$, $y = 0$. The observation probes have provided both quadratures without a phase-sensitive receiver. For a lossy line the round-trip propagation factor also attenuates magnitude; the lossless envelope formula and phase-only de-embedding then require modification.

Tutorial

A reference plane is misplaced by $\lambda/100$. Estimate the resulting phase error of the reconstructed load.

▼ **Work through the solution**

The round-trip phase is $4\pi d/\lambda$. A placement error of $\lambda/100$ therefore rotates the result by 0.04π rad, or 7.2°. Its sign depends on whether distance was over- or underestimated.

Why is recovering a very large resistance from reflection poorly conditioned?

▼ **Work through the solution**

$$dZ/d\Gamma = 2Z_0/(1-\Gamma)^2.$$

Near the open-circuit value +1 this derivative becomes large. A small error in the complex coefficient can imply a large change in estimated resistance while reflected power barely changes.

A standing-wave ratio is measured with a detector displaying power rather than amplitude. Its largest-to-smallest reading is 2.25. What is the correct reflection magnitude?

▼ **Work through the solution**

Take the square root to get a voltage ratio of 1.5, then

$$|\Gamma| = (1.5-1)/(1.5+1) = 0.2.$$

Substituting 2.25 directly into the voltage formula would yield a biased result.

Estimate a cavity linewidth from a finite, biased record

Numerical field laboratory

Connect the time and frequency descriptions of one damped mode, then diagnose why a weak detector background systematically biases the inferred quality factor.

Know whether the detector reads field or energy

After a drive switches off, let a single cavity mode have complex analytic amplitude $a(t) = a_0 e^{-\kappa t/2} e^{-i\omega_0 t}$ for $t \geq 0$. The energy is proportional to $|a|^2$, giving $U = U_0 e^{-\kappa t}$. Its instantaneous decay rate $-U'/U = \kappa$ implies

$$Q = \omega_0 U / P_{\text{loss}} = \omega_0 / \kappa.$$

A phase-sensitive receiver measuring field amplitude sees a time constant twice as long as the energy lifetime. Before fitting the data, ask how the measurement transducer converts the field into a reported number.

Fourier transforming the causal decaying amplitude gives an integral proportional to $1/[\kappa/2 - i(\omega - \omega_0)]$, up to the time convention. Its squared magnitude has half-power points at $|\omega - \omega_0| = \kappa/2$. Therefore the full power linewidth in hertz is $\Delta f = \kappa/(2\pi)$. A measurement of either the energy decay time $\tau_U = 1/\kappa$ or the spectral width can test the other without having to calibrate the absolute detector gain.

At $f_0 = 2.0$ GHz and $\tau_U = 0.50$ μs, we obtain $\kappa = 2.0 \times 10^6$ s^{-1}, $\Delta f \simeq 318.3$ kHz and $Q \simeq 6283$. A field receiver would report an amplitude time constant of 1.0 µs. Substituting that time directly into $Q = 2\pi f_0 \tau_U$ would double the apparent quality.

Derive the two-point fit with a detector offset

Let a power detector report $S(t) = A e^{-\kappa t} + B$, with a constant background $B > 0$. In the absence of background two samples separated by Δt give

$$\widehat{\kappa} = \log[S(t)/S(t+\Delta t)]/\Delta t.$$

With background, the logarithmic derivative is

$$-\frac{d}{dt}\log S = \kappa \frac{A e^{-\kappa t}}{A e^{-\kappa t} + B}.$$

It starts below the true rate and tends to zero in the tail. A straight line fitted to late log data can therefore look impressively precise while overestimating the lifetime. Increasing the acquisition duration can worsen this bias if the entire record is forced into a straight-line model.

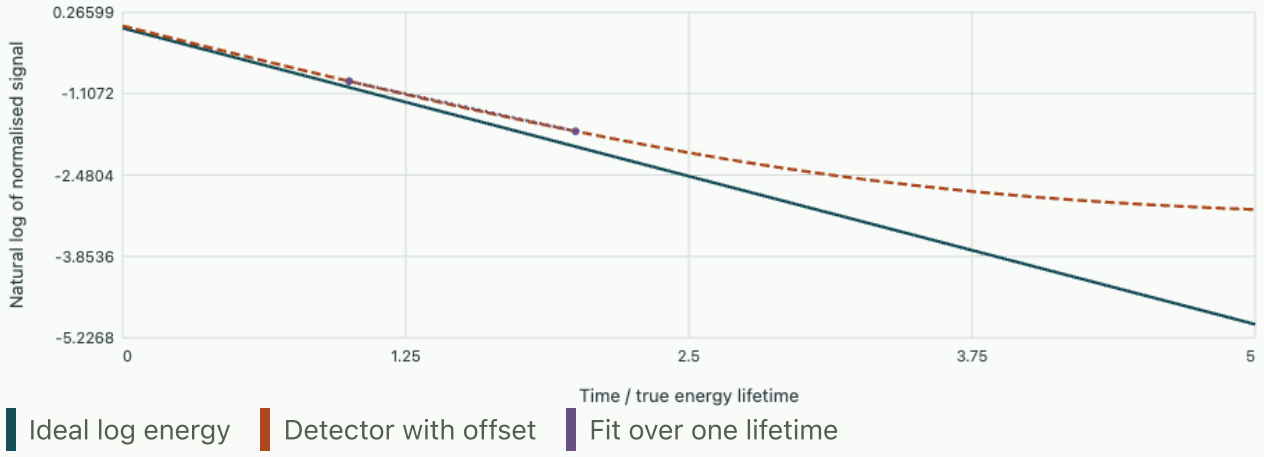
In the experiment set $b = B/A = 0.04$. A two-point fit from one true lifetime to two has dimensionless rate $\log[(e^{-1} + 0.04)/(e^{-2} + 0.04)]$, approximately 0.844. Its apparent lifetime is about 18.5% high. The source, walls and coupling have not changed. Only the interpretation of detector output has changed. Move the start of the fit towards three lifetimes and inspect the increasing gap from the ideal log line.

Open the live experiment

Try the calculation

Additive detector offset / initial energy **0.04**

Start of fitting window (lifetimes) **1**



Two-point inferred lifetime 1.1845 times the true lifetime. The same fractional bias transfers to Q for a known mode frequency.

Design a second independent inference

With three equally spaced, noiseless readings S_0, S_1, S_2 of one exponential plus a constant, subtract consecutive values before taking a logarithm. If $a = e^{-\kappa \Delta t}$, the differences are $A'(1-a)$ and $A'a(1-a)$, giving $a = (S_1 - S_2)/(S_0 - S_1)$. This cancels the constant background. At an interval of one lifetime, a first signal of unit energy and background 0.04 gives approximately $1.04, 0.407879, 0.175335$. Their ratio of differences recovers e^{-1}. In a real noise floor two nearly equal readings make this division unreliable. Use multiple windows or fit the raw signal with a measured noise model.

Finite spectral resolution creates a separate challenge. Multiplying the measured field by a rectangular window of duration T convolves its spectrum with a sinc-like response of width of order $1/T$. For the 318 kHz linewidth in the example, $T = 1$ µs gives frequency bins spaced by 1 MHz in a basic Fourier transform. Zero padding can interpolate its curve, but cannot restore the unmeasured late-time signal. A parametric fit may extract a sub-bin frequency or decay rate by using its model assumptions; its uncertainty then depends on signal-to-noise ratio and on whether the one-mode theory is adequate.

Check the single-mode model before refining its fit

If two nearby modes reach the same detector, their fields add before squaring. The result contains an interference term oscillating at their frequency difference on top of a decaying envelope. A scalar positive background cannot explain that alternating residual. Test by narrowing the excitation band, moving the

pickup to a symmetry node of one mode, or examining a complex receiver's quadratures. A model with two exponentials has more degrees of freedom and can fit noise as well as signal; require consistency across independent records.

An honest linewidth report specifies the measured quantity, acquisition window, background treatment, coupling state, reference frequency and uncertainty. A fit to stored energy measures total loss. Inferring internal absorption alone additionally requires a known port coupling or a measurement that changes it separately, as in the course's cavity project.

Tutorial

A detector background is independently measured but subtracted too aggressively. What feature can appear in the logarithmic tail?

▼ **Work through the solution**

The corrected signal can approach zero and cross below it. The logarithm then becomes undefined. Even before that point its apparent local slope can become arbitrarily large. Inspect the raw signed residual and its uncertainty instead of dropping nonpositive points without accounting for selection bias.

A second experiment doubles the peak stored energy while the absolute detector offset is fixed. What should happen to the shape of the normalised log tail?

▼ **Work through the solution**

$b = B/A$ halves. Background contamination becomes important later by about $\tau_U \ln 2$. A real change in the physical lifetime need not follow that shift when the response remains linear.

Can a passive resonator's intrinsic losses be determined from a single ring-down measured at one port?

▼ **Work through the solution**

The observed decay is the sum of internal and external rates. A single accurately measured lifetime identifies their sum. A calibrated change in coupling or an independent power budget is needed to separate them.

Choose a frequency that identifies a coil's stored and dissipated energy

Numerical field laboratory

Design a small-signal experiment to estimate resistance and inductance, propagate amplitude and phase uncertainty and decide when a distributed field model is needed.

A measured impedance is one complex number

A laboratory sensing coil is driven by a small sinusoidal current at angular frequency ω. In its simplest usable range, a series model predicts $Z = R + i\omega L$. The voltage across a port defines the sign of current entering the positive terminal. With peak phasors, complex power is

$$S = VI^{*}/2 = |I|^{2}(R + i\omega L)/2.$$

Thus the real part measures average irreversible loss. Magnetic energy at peak current is $L|I|^{2}/2$; reactive power equals ω times this peak energy. These equalities supply two routes to the same parameters provided parasitic capacitance and frequency-dependent losses remain negligible.

At 100 Hz, a coil with $R = 20\ \Omega$ and $L = 40$ mH has reactance $2\pi(100)(0.040) = 25.13\ \Omega$, magnitude 32.12 Ω and phase about 51.49 degrees. A 10 mA peak current produces about 0.321 V peak and 1 mW average resistive loss. The peak magnetic energy is only $2\,\mu\mathrm{J}$. Check that the reactive power obtained from the phase and current is about 1.26 mvar. Confusion between RMS and peak phasors easily introduces a factor two while still leaving the impedance ratio correct.

Derive what an amplitude error and a timing error do

Suppose the readout gives magnitude M and phase ϕ. Then $R = M\cos\phi$, $L = M\sin\phi/\omega$. For small independent uncertainties and known ω, differentiate both expressions. The relative first-order perturbations are

$$\frac{\delta R}{R} = \frac{\delta M}{M} - \tan\phi\,\delta\phi,$$

\frac{\delta L}{L}=\frac{\delta M}{M}+\cot\phi\,\delta\phi .

The phase uncertainty must be expressed in radians. For independent random errors their variances add with the squared coefficients; the signed equations above describe individual perturbations. Near zero phase even a very accurate amplitude tells little about inductance, while near 90 degrees inferring resistance requires exceptional phase accuracy. A near-45-degree measurement balances these two sensitivities when the uncertainty budget treats them similarly.

If the phase uncertainty is 0.5° and relative magnitude uncertainty is 1%, the 45-degree standard deviation in either recovered parameter is approximately $\sqrt{0.01^{2} + 0.00873^{2}} = 0.0133$, or 1.33%. At five degrees the relative inductance contribution from phase alone is $\cot 5^{\circ} \times 0.00873 \simeq 10\%$. Taking data at more digits does not improve phase calibration.

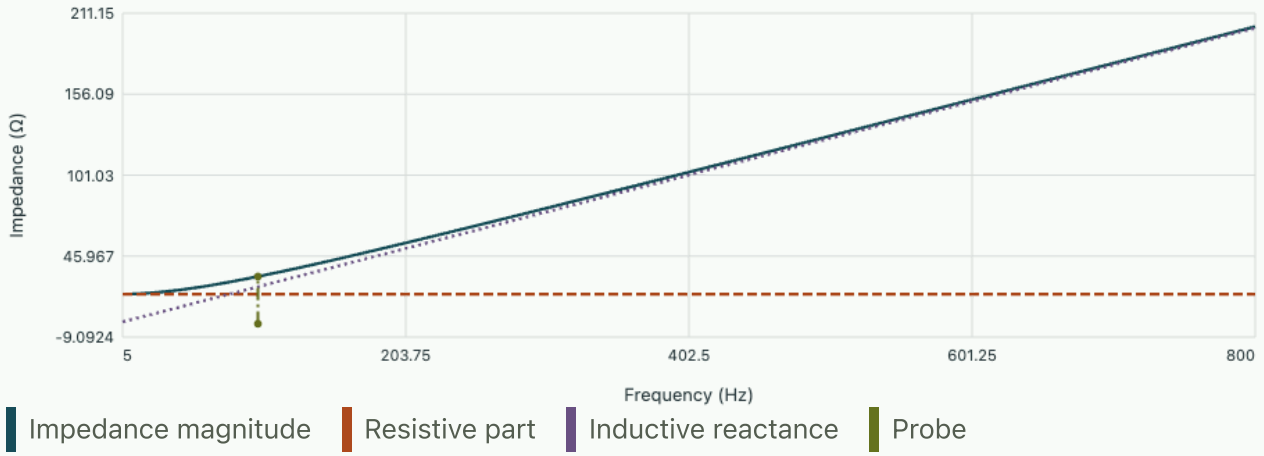
Choose the operating point, then challenge the model

Open the live experiment

Try the calculation

Probe frequency (Hz) 100

Coil inductance (mH) 40



At the probe: |Z| = 32.119 Ω, phase 51.488°. Equal resistive and reactive contributions at 79.577 Hz.

Vary frequency at fixed inductance and read where resistive and reactive contributions are equal. The crossover is $f_c = R/(2\pi L)$, about 79.58 Hz for the default coil. Now double the inductance without changing R; the crossover halves. The graph is a prediction with known parameters. In an actual inverse experiment you know only rough bounds initially. Begin with a small frequency sweep that brackets the crossover, then allocate repeated measurements where the target parameter is most sensitive.

Measure at two or more separated frequencies and reconstruct R and L independently at each. A frequency-dependent apparent R could be caused by skin effect or core loss. A falling apparent inductance near a self-resonance can signal stray capacitance. Simply allowing R, L to take arbitrary values at each frequency fits all data without explaining the mechanism. Add a candidate physical effect and check an independent observable such as loss under a different current amplitude, ring-down frequency or a change in lead geometry.

Close the loop from a fitted parameter to a field calculation

The lumped magnetic energy $LI^2/2$ should match the field integral $\int \mathbf{B}\cdot\mathbf{H}\,dV/2$ for a linear reversible system that includes the relevant surrounding field. This is why the exterior air region and its imposed boundary can affect a numerical estimate even when the measured coil is compact. Increase the domain and refine the mesh separately. A converged mesh in a small box can deliver a very repeatable value for the wrong boundary problem. The earlier chapter on mutual-inductance matrices generalises this check to cross terms between currents.

Another independent measurement uses a slowly executed current ramp with a known winding resistance. Subtract the measured resistive voltage from terminal voltage and integrate over time. Under the linear model,

$$\int (V - RI)\, dt = L\Delta I.$$

This test weights offsets very differently from sinusoidal phase detection: a constant voltmeter offset integrates into a bias proportional to record duration, while a high-frequency phase error drops out. Agreement of the two methods is stronger evidence than repeating a single instrument setup at the same frequency.

Tutorial

Only voltage and current magnitudes are available, but two frequencies are accurately known. Can a constant R and L be recovered, and what makes the inference fragile?

▼ **Work through the solution**

$$M_i^2 = R^2 + \omega_i^2 L^2.$$

Subtract the equations to get

$$L^2 = (M_2^2 - M_1^2)/(\omega_2^2 - \omega_1^2).$$

Insert this in either equation for R^2 and take positive roots. Nearly equal frequencies produce a small denominator; nearly equal magnitudes amplify measurement noise. Under a misspecified model the inferred square can even become negative.

At a fixed frequency the current reverses direction in a repeat measurement but its reference voltage polarity is held fixed. What measurement error could this create?

▼ **Work through the solution**

Reversing the defined port-current orientation multiplies the inferred impedance by –1 unless the voltage reference is reversed consistently. An apparent negative resistance in passive hardware may be a port convention issue. Check the direction of measured power flow with a known reference resistor.

What extra evidence would justify carrying a fitted inductance into a time-domain prediction?

▼ **Work through the solution**

Make the prediction for a current waveform whose spectrum lies within the validated range of the measurements. Test at another amplitude for nonlinearity or hysteresis and record the residual voltage

after subtracting $RI + L\,dI/dt$. A successful fit at one frequency alone does not establish a broadband constitutive law.

Measuring magnetic core loss by a hysteresis loop and by calorimetry

Electromechanics, optics and measurement

A closed loop on a graph of magnetic induction against field encloses an area with units of energy per unit volume. A thermometer measures kelvins. Bring these two observations onto the same scale, account for the winding heat, and determine when the two experiments actually test the same material loss.

Infer the field and the induction from independent windings

Wind a toroidal specimen with a primary winding of N_p turns and a lightly loaded pickup of N_s turns. Let A be the representative iron cross-sectional area, ℓ the mean magnetic path length and $V_c = A\ell$ the magnetic volume. Neglect a gap and leakage initially. Ampère's law around a path through the core gives

$$H \simeq N_p i_p / \ell.$$

This relation is accurate when the toroid is slender enough that the spread of path radii is small and the pickup current is negligible. In a thick toroid $H(r) = N_p i_p/(2\pi r)$, so one point on a nominal loop can actually represent a range of local field values.

Faraday's law supplies the second observable. Choose winding polarity so the reported voltage is

$$v_s = N_s A \, dB/dt.$$

Integrating the voltage gives the change in cross-section-averaged magnetic induction. The pickup cannot determine a constant induction offset, because static flux induces no voltage. A closed loop area is insensitive to adding a constant to B or H, but a dc offset in the voltage integrates into a time-dependent ramp and prevents the reconstructed trajectory from closing. In steady periodic operation one can subtract the voltage mean over a complete period before integration. Applying that correction to a genuinely evolving magnetisation process would remove part of a real flux change.

Multiply the two measured field quantities in the appropriate order. The field work supplied per unit material volume on a small induction change is $H\,dB$. The source term associated with the magnetising flux is $i_p v_{p,\mathrm{induced}}\,dt$. Using $v_{p,\mathrm{induced}} \simeq (N_p/N_s) v_s$, its value is

$$i_p N_p A \, dB = V_c H \, dB.$$

Integrating around a settled cycle gives the magnetic loss

$$W_c = V_c \oint H \, dB,$$

$$P_c = f V_c \oint H \, dB.$$

The field can absorb positive work and return some of it later in the same cycle. The area measures the algebraic remainder. Reversing the polarity of a single signal flips the loop orientation. A passively heated specimen must receive positive net energy when signs are defined for power entering the core. Simply

taking the absolute value of all instantaneous products would count reversible storage twice and can grossly overestimate the heat.

Work an elliptical loop from time-domain signals

Suppose a nearly sinusoidal field is $H = H_p \cos \omega t$ and the induction lags it by a positive angle δ, $B = B_p \cos(\omega t - \delta)$. Differentiate the latter and multiply by H. Over one cycle $\cos \omega t \sin \omega t$ averages to zero, and $\cos^2 \omega t$ averages to one half. The resulting mean power density is $\omega H_p B_p \sin \delta / 2$, giving $\oint H \, dB = \pi H_p B_p \sin \delta$. With zero lag the two signals trace a line and the area vanishes, even if the peak values are large. A saturation-distorted loop need not be elliptical; the original time integral remains valid with synchronised data.

Take $H_p = 120\,\mathrm{A/m}$, $B_p = 0.200\,\mathrm{T}$ and $\delta = 12^\circ$. The energy per cycle and volume is $15.68\,\mathrm{J/m^3}$. A toroid with $\ell = 0.20\,\mathrm{m}$, $A = 2.0 \times 10^{-4}\,\mathrm{m^2}$ and $V_c = 40\,\mathrm{cm^3}$ dissipates $0.627\,\mathrm{mJ}$ per cycle. At $f = 200\,\mathrm{Hz}$ its power is $0.125\,\mathrm{W}$. With $N_p = 100$, the necessary current peak is $0.240\,\mathrm{A}$. With $N_s = 50$, the pickup voltage has peak $N_s A (2\pi f) B_p = 2.51\,\mathrm{V}$. These last two predictions allow the raw oscilloscope traces to be checked before a calculated loss in watts is trusted.

For sampled periodic data use $w = \sum_n (H_n + H_{n+1})(B_{n+1} - B_n)/2$, including the segment from the last point back to the first. This trapezoidal line integral computes an oriented polygon area. Plot the starting and ending points before adding the closing segment. Connecting widely separated endpoints by hand hides drift and can generate a fictitious wedge whose area depends on acquisition timing. Repeat at higher sampling rates to see whether corners or switching features carry appreciable extra area. When a periodic excitation has harmonics the time increment must resolve the highest significant harmonic in both measured signals.

Compare a thermal measurement with the magnetic area

Consider first a calorimeter that receives heat only from the core. Let the heat capacity of specimen, holder and sensor be C_{eff}. Let the leakage conductance to an ambient bath be G, and let $\Delta T = T - T_a$. The thermal balance $C_{\mathrm{eff}} \dot{\Delta T} = P_c - G \Delta T$ predicts a steady offset P_c/G and an initial heating slope P_c / C_{eff}. A small thermometer lag or poor thermal contact can make the earliest measured slope misleading even though environmental losses are then small. A separate electrical heater placed where the core deposits its heat can calibrate the effective thermal response.

For $C_{\mathrm{eff}} = 20\,\mathrm{J/K}$, $G = 0.015\,\mathrm{W/K}$, and a $2.0\,\mathrm{K}$ rise above ambient, the magnetic loss $0.125\,\mathrm{W}$ would give a slope of about $0.00477\,\mathrm{K/s}$. The heat already leaking to the bath is $0.030\,\mathrm{W}$, too large to discard in a ten-percent comparison. If the winding is inside the same calorimeter, its Ohmic heat must enter the balance as well. With primary resistance $1.50\,\Omega$ and the computed sinusoidal current, winding heating is

$$(0.240/\sqrt{2})^2 (1.50) = 0.0432\,\mathrm{W}.$$

The expected slope of the combined package at the same temperature is now $0.00693\,\mathrm{K/s}$. Subtraction of winding heat recovers the core result only when the estimated resistance represents the winding at its operating temperature and frequency.

A thermal estimate drawn from a whole interval is often more robust than a derivative from two adjacent temperature samples. Integrating gives

$$Q_{\mathrm{deposited}} = C_{\mathrm{eff}}[\Delta T(t_2) - \Delta T(t_1)] + G\int_{t_1}^{t_2} \Delta T\, dt.$$

Divide by the duration for mean power, then remove other measured heat sources. A drifting ambient must be recorded explicitly because the calorimeter stores energy relative to its own absolute temperature. If the core and coil equilibrate slowly with each other, a single fitted heat capacity can depend on the fitting time window. A two-temperature model or a longer dwell then supplies an independent diagnostic.

Separate reversible exchange from hysteretic and induced currents

A single measured loop at finite frequency gives the total energy dissipated by the core at that prescribed waveform and temperature. The loop can include irreversible domain processes, eddy currents and other rate-dependent effects. It does not by itself split these contributions. To probe their different scaling hold the peak induction and waveform shape fixed, then change frequency. Under a restricted low-frequency model the quasi-static hysteresis energy per cycle is approximately constant, while classical thin-lamination eddy-current power scales as $f^2 B_p^2$. Its energy per cycle therefore scales roughly as f. Plotting the measured loss per cycle and volume against f tests whether such an intercept-plus-slope description fits the chosen range.

Suppose the area is $12\,\mathrm{J/m^3}$ at $100\,\mathrm{Hz}$ and $20\,\mathrm{J/m^3}$ at $300\,\mathrm{Hz}$, at the same peak induction. A straight line gives intercept $8\,\mathrm{J/m^3}$ and slope $0.040\,\mathrm{J\,s/m^3}$. It predicts $16\,\mathrm{J/m^3}$ at $200\,\mathrm{Hz}$, near the elliptical-loop calculation. Two points always determine a line. A third and fourth frequency, together with uncertainties, are required to test the law. Different losses can share similar frequency scaling over a short interval, and temperature-dependent conductivity can bend a nominally simple curve.

Use the pickup to reject the reversible part of terminal power

Measuring primary terminal voltage and current with a power analyser provides a separate electrical route to loss, but the wanted number can be a small difference between terms. The applied voltage contains winding resistance, induced magnetising voltage, leakage-inductance voltage and possible parasitic contributions. In steady periodic operation the energy stored in a fixed linear leakage inductance returns after each cycle, while winding resistance dissipates energy permanently. The secondary pickup preferentially measures the core-linked flux. Forming $p_c(t) = (N_p/N_s) i_p(t) v_s(t)$ and averaging over a full cycle estimates core power without directly including the primary's copper voltage drop. It relies on both windings linking the same relevant flux and on negligible power drawn by the pickup.

For the elliptical example the primary magnetising voltage has peak $5.03\,\mathrm{V}$ and leads the induction by ninety degrees. Its phase relative to the current is

$$90^\circ - \delta = 78^\circ.$$

The core branch's apparent power is $V_{\mathrm{rms}} I_{\mathrm{rms}} \approx 0.603\,\mathrm{VA}$; its mean dissipative part is about $0.125\,\mathrm{W}$, and the reactive part is about $0.590\,\mathrm{var}$. Only the mean part contributes net heating of the magnetic volume. Including the winding voltage adds about $0.043\,\mathrm{W}$ of real power to the terminal reading. A simple product of peak current and peak secondary voltage, without the turns ratio, factor one half and phase factor, has no reason to agree with either measured loss.

Consider a nearly ideal core with the same peak signals but a lag of only one degree. Its loss is about $10.5\,\mathrm{mW}$, while the much larger reversible power exchange is almost unchanged. A time offset of $10\,\mu\mathrm{s}$ at $200\,\mathrm{Hz}$ adds 0.72°, causing a relative loss error of approximately 72% for that small lag. Calibrate the relative channel delays using a reference signal through the actual current transducer and voltage chain, including their filters. Adjusting traces on screen until a core loop appears narrow assumes the very loss one is trying to measure. A synchronous measurement over many periods suppresses random noise but does not remove a persistent delay bias.

Test reconstruction on a waveform with a third harmonic

Let $H(t) = H_1 \cos \omega t + H_3 \cos 3\omega t$ while $B(t) = B_1 \cos(\omega t - \delta_1) + B_3 \cos(3\omega t - \delta_3)$. With these definitions δ_n is the phase lag of the nth harmonic in its own sinusoid. Multiplying H by $\dot{B}$ and integrating over an integer number of fundamental periods eliminates products of different harmonics. Each surviving harmonic contributes the number of its own cycles in one fundamental period. The area per fundamental period is therefore $\pi[H_1 B_1 \sin \delta_1 + 3 H_3 B_3 \sin \delta_3]$. This identity tests whether a numerical area calculation tracks the full waveform. It does not assume that a real material is linear; here the two signals are supplied as measured periodic functions.

Use the earlier fundamental values and add $H_3 = 20\,\mathrm{A/m}$, $B_3 = 0.030\,\mathrm{T}$ and $\delta_3 = 30^\circ$. The third harmonic contributes $2.83\,\mathrm{J/m^3}$ in each fundamental period. This represents an additional $0.0226\,\mathrm{W}$ for the chosen specimen and frequency. A fundamental-only sine fit would miss nearly fifteen percent of the resulting total area. An apparent jagged loop can also arise from asynchronous sampling or an offset. Distinguish genuine harmonics by checking that they are phase locked to the drive and persist under a change in sampling rate and analogue bandwidth.

Sweep phase and read the oriented loop

The live experiment reconstructs the H, B trajectory and compares it with a zero-lag reference. It uses the same peak field and induction for the two curves. At fixed peaks changing frequency leaves an assigned phase lag and therefore the prescribed energy per cycle unchanged, while watts rise with the number of cycles per second. A real core can change its lag and waveform with frequency; the graph separates this independent variable so that the meaning of the two scaled quantities is clear. Add a timing offset and predict the direction of the area error before checking its reported size. Near zero lag even the direction of an apparent loop can be controlled by a small electronic offset.

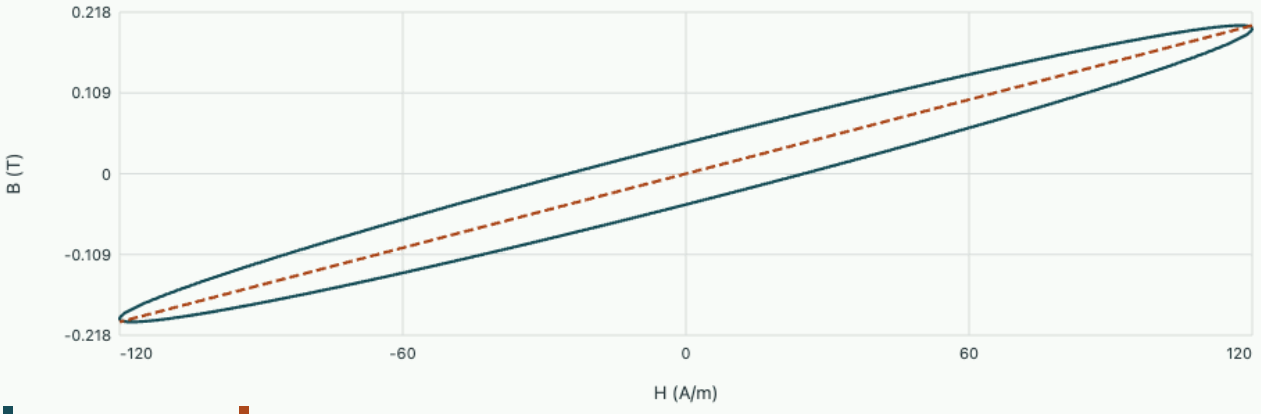
Open the live experiment

Try the calculation

Physical induction lag (°) **12**

Frequency (Hz) **200**

Induction channel delay (µs) **0**



True loop area 15.68 J/m³ per cycle. The recorder indicates 15.68 J/m³. Physical heating of the 40 cm³ specimen is 0.125 W.

Two toroids are made of the same nominal material with equal cross sections but different mean path lengths. How should a result reported only as watts be compared?

▼ **Work through the solution**

Determine actual magnetic volume for each toroid and calculate the energy per unit volume per cycle, $P/(fV_c)$. Also check that both experienced the same induction waveform, frequency and temperature. Equal primary currents need not give equal field when the path lengths differ; equal primary applied voltages need not give equal core induction if turns, winding voltage drops or area differ. Reproducing the actual $B(t)$ with feedback on the pickup is a more direct way to match excitation. Report dimensions and peak induction with the result so a reader can reconstruct what material condition was tested.

A saturated sample shows a small loop during a biased excitation that never crosses zero induction. Does the area formula still apply?

▼ **Work through the solution**

A settled periodic minor loop encloses an area representing the net cyclic energy under exactly that bias and excursion. Adding a static induction offset does not change $\oint H\,dB$, but the bias can alter the trajectory, differential permeability, and width of the minor loop itself. Do not infer the full major-loop energy from its area. The pickup returns only changes of induction unless an independent dc measurement or a known magnetic preparation supplies the offset. A bias winding may additionally heat the calorimeter, so include its measured power when comparing with thermal loss.

Tutorial on comparing two loss measurements

A data logger reports H_p and B_p correctly but adds an uncorrected $10\,\mu\mathrm{s}$ delay to the induction channel. Estimate the fractional area error of the elliptical example.

▼ Work through the solution

At $200\,\mathrm{Hz}$ the extra apparent lag is

$$2\pi f \Delta t = 0.01257\,\mathrm{rad} = 0.72^\circ.$$

The reported area is multiplied by $\sin(12.72^\circ)/\sin(12^\circ)$, about 1.059. An uncorrected delay of only one hundred-thousandth of a second overstates the core loss by about six percent. For a very low-loss specimen with much smaller physical lag, the same delay produces a far larger relative error.

A thermal run on the combined winding and core reports a slope $0.0070\,\mathrm{K/s}$ at $\Delta T = 2.0\,\mathrm{K}$. Recover the core power and the volume-specific loss area at $200\,\mathrm{Hz}$.

▼ Work through the solution

Heat deposition in the package is

$$20(0.0070) + 0.015(2.0) = 0.170\,\mathrm{W}.$$

Subtract the estimated $0.0432\,\mathrm{W}$ of winding heating to obtain $0.1268\,\mathrm{W}$ in the core. Dividing by $200 \times 4.0 \times 10^{-5}$ gives $15.85\,\mathrm{J/m^3}$ per cycle. Its closeness to $15.68\,\mathrm{J/m^3}$ should be evaluated against the uncertainties of heat capacity, thermal conductance, resistance, dimensions and electrical phase alignment.

Vary the magnetic circuit and identify the induction scale used by the specimen.

Optical trapping and Brownian calibration of a dielectric bead

Electromechanics, optics and measurement

An image of a bead remaining near a focus does not establish the strength of its trap. Derive the force pulling a small dielectric sphere into a bright region, compare it with the forward force from scattering and learn what a motion record can tell you about stiffness, drag and temperature. The model here uses an inert dielectric sphere in a homogeneous Newtonian liquid.

From two dielectric boundaries to the energy of an induced dipole

A sphere of radius a and relative permittivity ϵ_b sits in a liquid of relative permittivity ϵ_m. Assuming the sphere is small compared with a wavelength in the liquid, solve its local electric problem in an almost uniform applied field E. Rotational symmetry about the field admits an interior potential proportional to $r\cos\theta$. The exterior consists of the imposed potential $-Er\cos\theta$ and an induced dipolar term proportional to $\cos\theta/r^2$. Equality of potential and of normal displacement at the surface gives the local-field and dipole factors

$$E_{\mathrm{in}} = \frac{3\epsilon_m}{\epsilon_b + 2\epsilon_m} E,$$

$$p = \alpha E,$$

$$\alpha = 4\pi\epsilon_0\epsilon_m a^3 \frac{\epsilon_b - \epsilon_m}{\epsilon_b + 2\epsilon_m}.$$

The field doing the polarising work is the external field in the host at the particle position before inserting it. To turn that work into a potential, build up the dipole from zero rather than assuming a permanent dipole moment. Quasistatic assembly gives

$$-\int_0^E \alpha E'\, dE' = -\alpha E^2/2.$$

For a harmonic peak field average the square over an optical period. In a nearly lossless liquid of refractive index n_m, the peak field is related to travelling-wave intensity by $I = n_m\epsilon_0 c|E|^2/2$. The resulting conservative force is

$$U = -\frac{\mathrm{Re}\,\alpha}{4}|E|^2 = -\frac{\mathrm{Re}\,\alpha}{2n_m\epsilon_0 c} I,$$

$$\mathbf{F}_{\mathrm{grad}} = -\nabla U.$$

The two factors of two have different origins: induced-dipole assembly and time averaging of a sinusoid. Using a permanent-dipole potential with a peak field would miss both. A bead of higher refractive index than the liquid has positive real polarizability and is attracted to increasing intensity. A lower-index inclusion reverses this local tendency. For large particles, strong absorption or rapidly structured fields, a dipole with real static polarizability no longer specifies the force; a field or multipole calculation of the complete momentum balance is required.

Investigate the small-sphere dipole response and its dependence on particle size

Compute a transverse stiffness and test how local it is

Assume the beam near its focal plane has intensity $I(x) = I_0 e^{-2x^2/w^2}$, with power P, waist w defined at $1/e^2$ intensity and $I_0 = 2P/(\pi w^2)$. Put $U_0 = \mathrm{Re}\,\alpha I_0/(2n_m\epsilon_0 c) > 0$. The potential measured above its minimum is $U_0[1 - e^{-2x^2/w^2}]$. Expanding to second order and equating its leading term to $\kappa_x x^2/2$ gives

$$\kappa_x = 4U_0/w^2.$$

Stiffness therefore scales linearly with power, as the third power of sphere radius and as the inverse fourth power of waist at fixed indices. Each scaling applies within the stated approximations.

For a 40 nm radius bead with $n_b = 1.59$, take $n_m = 1.33$, $w = 0.65$ micrometres, vacuum wavelength 1064 nm, $P = 90$ mW and $T = 298$ K. The index contrast factor is 0.1252 and the waist intensity is 1.36×10^{11} watts per square metre. The predicted depth 3.03×10^{-20} J is 7.36 times the thermal energy. The lateral stiffness is 2.87×10^{-7} N/m. A displacement of 50 nm then requires approximately 14.3 fN of transverse balancing force. At a displacement of half a waist, the exact restoring force is the harmonic estimate multiplied by $e^{-1/2} = 0.607$; a Hookean calibration cannot be extrapolated that far.

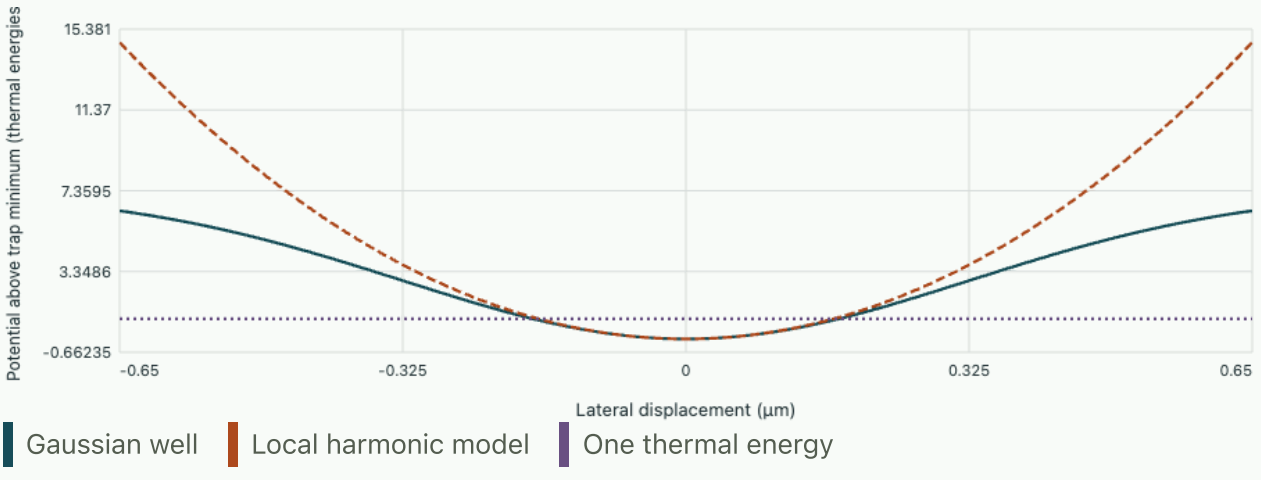
Open the live experiment

Try the calculation

Bead radius (nm) **40**

Optical power (mW) **90**

Beam waist (µm) **0.65**



The well depth is 7.36 kBT. The local model gives an rms displacement of 119.8 nm and a corner frequency of 60.5 Hz at 298 K.

Use the two curves to identify the interval in which the local quadratic energy is within ten percent of the Gaussian prediction. First lower power at fixed waist. Both curves shrink together in energy units while the harmonic thermal displacement grows. Next increase bead radius at fixed power and watch the well deepen, but also evaluate the size parameter $2\pi n_m a/\lambda_0$. It is 0.314 at the default and grows linearly with

radius. Stronger trapping predicted by this simplified model does not establish the accuracy of the dipole approximation at arbitrary sizes.

Compare axial restoring force with radiation pressure

An induced dipole reradiates a portion of the incident field. In a uniform host use the Rayleigh scattering cross section $\sigma_s = (8\pi/3) k_m^4 a^6 |(\epsilon_b - \epsilon_m)/(\epsilon_b + 2\epsilon_m)|^2$. In a common travelling-wave momentum-flux model for a bead in a stationary nondispersive liquid the axial scattering force is approximately $F_s = n_m I \sigma_s / c$; absorption would add a further contribution. This estimate uses the total scattering cross section because the leading dipole pattern has vanishing mean forward cosine. For a larger forward-scattering sphere a transport cross section weights the deflected momentum instead.

In the default case the cross section is roughly 2.05×10^{-18} square metres. The forward force at the waist is about 1.23 fN. A Gaussian axial profile has $I(0,z) = I_0/[1 + (z/z_R)^2]$ with $z_R = \pi n_m w^2/\lambda_0 = 1.66$ micrometres. Its local axial stiffness $2U_0/z_R^2 = 2.20 \times 10^{-8}$ N/m is much smaller than its transverse stiffness. Balancing the constant forward force with that local stiffness predicts an equilibrium about 56 nm downstream of the waist. The small displacement relative to the Rayleigh length supports the linearisation.

Both the axial stiffness and forward scattering force scale with power in this simple model. A power sweep can thus increase confinement while barely shifting the equilibrium offset, a useful discriminating prediction. At fixed focus the offset grows rapidly with bead size because the scattering force scales with the sixth power and the gradient force with the third power. A stable lateral position alone does not establish stable three-dimensional capture. Check the axial restoring slope at a force zero and verify that thermal escape and fluid flow are negligible over the acquisition interval.

Extract stiffness from fluctuations or from a relaxation time

Near a stable point in one Cartesian direction, let x denote displacement from the mean and ζ the viscous drag. At sufficiently low frequencies the inertial term is negligible and the overdamped Langevin equation is

$$\zeta \dot{x} = -\kappa x + \xi.$$

At equilibrium the fluctuating force has zero mean and white covariance $2k_B T \zeta \delta(t - t')$ over the frequencies considered. The deterministic relaxation time is

$$\tau = \zeta/\kappa.$$

Multiplying the equation at positive delay by $x(0)$ and averaging removes the uncorrelated later noise and gives the decaying covariance. Equipartition sets its initial value.

$$C(t) = \frac{k_B T}{\kappa} e^{-|t|/\tau},$$

$$f_c = \frac{\kappa}{2\pi\zeta},$$

$$S_x^{(1)}(f) = \frac{4 k_B T \zeta}{\kappa^2 + (2\pi f \zeta)^2}, \quad f \geq 0.$$

The stated spectrum is one-sided in ordinary frequency. Its integral from zero to infinity equals $C(0)$; doubling another time would double the variance. In an unbounded fluid with viscosity 1.00 mPa s, Stokes drag is $6\pi\eta a = 7.54 \times 10^{-10}$ N s/m. The default predicts $f_c = 60.5$ Hz and rms displacement 120 nm. The latter is roughly 0.18 waist, so the outer part of a long thermal trajectory may sample appreciable anharmonicity. A finite-depth well in an unbounded bath does not have a globally normalisable trapped equilibrium. These local equilibrium predictions apply to a sufficiently long-lived, locally sampled trapped state.

If a voltage detector gives $V = gx$, the spectral corner and normalised autocorrelation time do not depend on the displacement gain g. They determine only κ/ζ. Equipartition using a calibrated spatial variance gives κ independently of drag, provided mechanical fluctuations are in equilibrium and sufficiently resolved. If the viscosity or nearby-wall drag is uncertain, combining a known stage oscillation with the Brownian spectrum can determine further calibration quantities. Keep a stage-driven coherent peak out of the noise integral used for equipartition.

Account for exposure, sample size and unobserved motion

A camera typically reports the position averaged during an exposure of duration t_e. Its variance is the double integral of the covariance over the exposure interval divided by t_e^2. Evaluating the exponential local model and introducing $u = t_e/\tau$ gives a multiplicative factor $2(u - 1 + e^{-u})/u^2$. At $u = 1$ the reported physical-motion variance is only 0.736 of the instantaneous variance. Interpreting it as the full variance would overestimate stiffness by approximately 36 percent. Independent pixel-localisation noise adds additional variance with the opposite sign of bias.

Estimate that localisation noise with an effectively immobilised reference bead at a comparable brightness. Test the averaging correction by changing exposure duration at constant sampling frequency where possible. A camera that reports overlapping exposures has correlated measurement errors and needs a different covariance model. A long series of closely spaced frames also contains fewer independent thermal configurations than its frame count suggests. Compare estimates across several blocks substantially longer than the relaxation time; residual low-frequency drift can dominate even when high-frequency position precision looks excellent.

A separate check uses a known uniform liquid flow. At a steady small displacement x_s, force balance gives

$$\kappa x_s = \zeta u_{\mathrm{flow}}.$$

For the default drag and stiffness a flow of 10 micrometres per second gives an offset of 26.3 nm. Repeat with reversed flow and use half the displacement difference to cancel an unknown zero-force position. Near a wall, the local flow at the bead and wall-corrected drag may differ from a stage velocity and the unbounded Stokes value. Those effects alter the mechanical calibration even if the optical field is unchanged.

A known stage oscillation calibrates position in a fluid

Brownian fluctuations alone determine the product of several unknown calibration parameters. A weak sinusoidal translation of the sample cell adds a known mechanical signal at a chosen frequency. Work in laboratory coordinates with a fixed optical trap and with fluid moving uniformly at velocity $u(t)$ near the bead, well away from walls in the idealised model. The drag is now $-\zeta(\dot{x}-u)$. Superposing a stage displacement $A\cos\omega t$ on thermal fluctuations gives a deterministic response whose complex amplitude satisfies $(\kappa-i\omega\zeta)\widetilde{x}=-i\omega\zeta A$ in the negative-exponential convention. Therefore its amplitude ratio is

$$\frac{|\widetilde{x}|}{A}=\frac{\omega\tau}{\sqrt{1+\omega^2\tau^2}}.$$

At the trap corner the bead displacement amplitude is the known stage amplitude divided by the square root of two. A calibrated 80.0 nm stage oscillation there produces a predicted bead amplitude of 56.6 nm. If a position-sensitive receiver records a coherent peak amplitude of 0.170 V, its displacement gain is 3.01 mV/nm. At very low drive frequencies the fluid velocity is small for fixed stage amplitude and the trap holds the bead nearly stationary. At high enough frequency the bead follows the fluid displacement within the overdamped model, but finite hydrodynamic, optical and electronic bandwidths eventually matter.

After determining the gain, convert the surrounding voltage spectrum into spatial units and use the area of the stochastic fluctuation spectrum to find stiffness. Exclude the calibrated line and account for the finite spectral averaging window; a narrow drive that leaks into neighbouring bins can inflate the inferred thermal variance. Dividing measured stiffness by the independently determined corner rate then gives drag, which can be compared with the known viscosity and sphere radius. This avoids assuming free-space Stokes drag in the first step of a near-wall measurement.

There are two limiting systematic effects that a good fit to a single drive frequency cannot resolve. A frequency-dependent stage calibration changes the actual fluid displacement, and motion of the optics relative to the cell generates a direct detector background even with the bead immobilised. Repeat the drive above and below the corner, compare coherent phase as well as amplitude and acquire a trapped and immobilised reference. Correct conversion should explain the entire complex response with one spatial gain and one corner time. A single calibration number obtained at one frequency can hide a phase lag in the stage controller or in the optical position channel.

One last cross-check reverses the stage drive phase by half a period. The coherent bead response must reverse, while thermal background power remains unchanged. A component that fails this reversal could come from ambient vibration unrelated to the stage. Averaging the two traces with and without a sign change separates these contributions before either is used in an optical-force calibration.

Tutorial · Separate a unit-conversion error from excess noise

In a single-axis trap at 298 K a calibrated detector reports a variance of 2500 square nanometres after removal of localisation noise and exposure correction. An independent spectral fit gives a corner of 175 Hz. Infer stiffness, drag and the equivalent unbounded hydrodynamic radius for a liquid of viscosity 1.00 mPa s. Compare against a nominal 80 nm radius bead.

▼ **Infer the three quantities and test their consistency**

A nanometre squared is 10^{-18} square metres, so the variance is 2.50×10^{-15} square metres. Equipartition gives 1.65×10^{-6} N/m. Dividing by $2\pi f_c$ gives drag 1.50×10^{-9} N s/m and radius $\zeta/(6\pi\eta) = 79.5$ nm. The agreement supports the local overdamped description under the assumed temperature and viscosity. Forgetting the factor of 2π would predict a radius roughly six times larger and could be mistaken for a huge near-wall correction.

Now imagine that additional uncorrelated read noise had been left in the variance. This raises apparent variance, lowers the equipartition stiffness and would lower the inferred drag at the same fitted corner. Nearby-wall drag for a given radius instead normally raises hydrodynamic resistance for parallel translation. Their effects on the inferred radius thus differ in sign under the stated assumptions. Varying height and independently calibrating the measurement noise are appropriate diagnostics before attributing a discrepancy to bead dimensions.

Tutorial · Test a force-law inference with power and concentration controls

At fixed bead radius a trap is operated at two powers differing by a factor of two. Measured corner frequencies increase from 42 to 84 Hz, but rms displacement falls only from 110 to 87 nm. What does the harmonic model with fixed drag and temperature predict? Estimate a power-independent additive position-noise variance that could reconcile the two readings.

▼ **Use two variances to separate scale and offset**

The doubled corner gives doubled stiffness if drag is unchanged. Physical thermal variance should halve. The predicted rms at higher power would be 77.8 nm if the lower-power value were entirely physical. Let the low-power measured variance be $v + b$ and the high-power one $v/2 + b$, where b is a constant additive noise variance. Solving from 12,100 and 7,569 square nanometres gives $v = 9062$ and $b = 3038$ square nanometres, or approximately 55.1 nm rms noise.

This is a large inferred detector noise and demands an independent immobilised reference check. Power-dependent heating of the liquid changes temperature and viscosity, so an unchanged product inferred from only two quantities cannot prove that heating is absent. Test several powers, reverse their order to reveal slow drift and vary bead concentration to exclude overlapping particles in the tracking window. The inferred additive term should remain constant across those controls if it comes from the specified detector model.

Tolić-Nørrelykke and colleagues · position and force calibration using thermal and driven measurements

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.

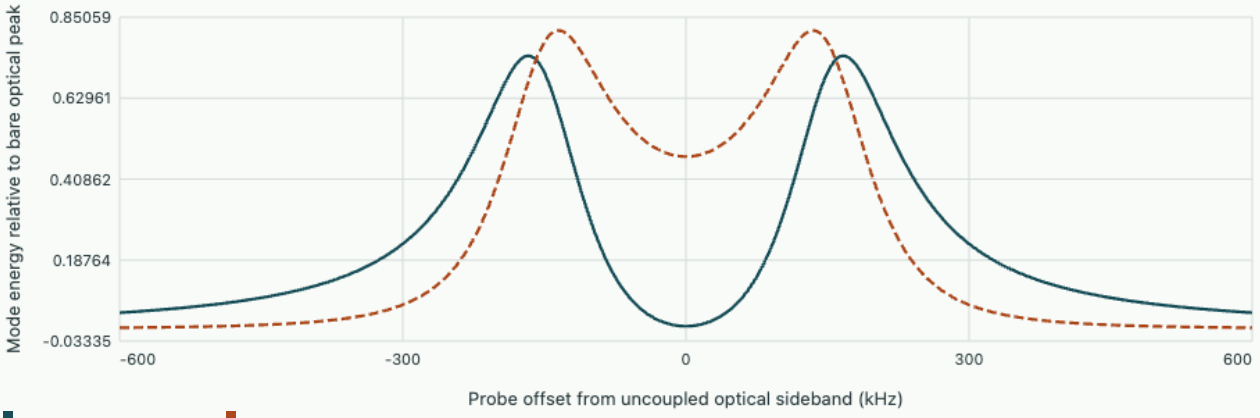
Open the live experiment

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

Scattering from a central potential and what angular data can recover

Analytical mechanics from constraints to invariants

A detector records angles, whereas the force law refers to separations that were never directly observed. Derive the forward map from incoming impact parameter to angle, distinguish the several trajectories a detector may count in one angular bin, and invert sufficiently complete signed deflection data under stated hypotheses.

Build the deflection integral from a radial turning point

Use relative mass μ, asymptotic relative speed v_∞, incoming energy $E = \mu v_\infty^2/2 > 0$ and impact parameter $b \geq 0$. With a central interaction that tends to zero at infinity, angular momentum is $\ell = \mu v_\infty b$. For a single repulsive encounter let r_0 be the closest approach. Energy gives

$$\dot r^2 = v_\infty^2[1 - U(r)/E - b^2/r^2];$$

r_0 is the outer turning root reached from infinity. Dividing $\dot\varphi = \ell/(\mu r^2)$ by the radial speed and integrating over the outgoing half trajectory gives

$$\Phi(b) = \int_{r_0}^{\infty} \frac{b\,dr}{r^2\sqrt{1 - U(r)/E - b^2/r^2}},$$

$$\chi(b) = \pi - 2\Phi(b).$$

The straight, free trajectory sweeps $\Phi = \pi/2$ between closest approach and infinity and gives zero deflection. For repulsive bend away from the centre this sign convention gives positive χ. A weak attraction can give negative signed deflection even though a detector reporting the angle between incoming and outgoing velocities may supply only its magnitude. If the path winds by more than a complete turn, angular data modulo 2π also lose winding information. The square-root singularity in the integral for an ordinary turning point is integrable. A substitution $r = r_0 + s^2$ makes the near-endpoint quadrature regular. A double turning point requires separate analysis because it can make the angular advance diverge.

Incoming trajectories with parameters between b and $b + db$ represent annular area $2\pi b\,db$ for an axially symmetric beam. At a detected polar angle θ the solid angle of a ring is $2\pi \sin\theta\,d\theta$. Conservation of incoming counts gives

$$\frac{d\sigma}{d\Omega} = \sum_j \frac{b_j}{\sin\theta}\left|\frac{db_j}{d\theta}\right|.$$

The sum includes all incoming impact parameters reaching that angular bin. Correcting measured counts requires incident flux, target number, angular acceptance and efficiency. The formula is a density in solid angle. Its apparent divergence at an isolated angle need not imply an infinite number in a finite acceptance bin. Integrate the density against the actual bin solid angle or use the preimage areas of the bin directly.

Solve a hard sphere and an inverse-square repulsion

A hard sphere of effective exclusion radius a specularly reverses the incoming normal velocity at the impact surface. The angle between incoming and outgoing directions is $\theta = 2\arccos(b/a)$ for $b < a$. Therefore $b = a\cos(\theta/2)$,

$$|db/d\theta| = (a/2)\sin(\theta/2).$$

The factors cancel to give $d\sigma/d\Omega = a^2/4$, independent of polar angle, and total cross section πa^2. Equal polar angular bins do not give equal counts: their solid angles differ by $\sin\theta$. This geometric distinction often accounts for a measured histogram that does not look uniform even for a constant differential cross section.

For $U(r) = K/r$ with $K > 0$, the orbit is a repulsive hyperbola. The analytic relation is

$$\tan(\theta/2) = K/(2Eb).$$

It can be derived from the energy and the angle through the deflection integral: put $u = b/r$ so that $\Phi = \int_0^{u_0} du/\sqrt{1 - (K/Eb)u - u^2}$. Complete the square and evaluate the arcsine at the root to get

$$\Phi = \pi/2 - \arctan[K/(2Eb)].$$

Define $d = K/(2E)$; then $b = d\cot(\theta/2)$,

$$d\sigma/d\Omega = d^2/[4\sin^4(\theta/2)].$$

High precision near the forward direction is physically limited by finite beam size, detector acceptance and any screening of the long-range interaction.

For $d = 2.0\,\mathrm{mm}$ in an abstract mechanical analogue, a deflection of 60° corresponds to $b = 3.46\,\mathrm{mm}$, and the differential cross section is $16\,\mathrm{mm}^2/\mathrm{sr}$. The cross section for deflections greater than 60° is the area $\pi b^2 = 37.7\,\mathrm{mm}^2$. The differential value and integrated area answer different laboratory questions. Doubling energy at fixed interaction halves d, quarters the cross section into a fixed angular interval and reduces the deflection at a fixed incoming b. Measuring only a total count without specifying angular threshold is inadequate for a long-range interaction.

Recover impact parameter from angular counting under monotonicity

Suppose the deflection decreases continuously from π at $b = 0$ to zero at large b. Integrating counts for all angles at least θ then gives precisely the incident disk $\pi b(\theta)^2$. The recovered map satisfies

$$b(\theta)^2 = 2\int_\theta^\pi (d\sigma/d\Omega)(\theta')\sin\theta'\,d\theta'.$$

Applying this formula to constant hard-sphere intensity gives $b^2 = a^2(1+\cos\theta)/2$, the geometrical result. It also exposes the information needed from a detector: a finite upper cutoff can leave the reconstruction with an unknown disk-area offset.

Failing monotonicity, the same cumulative counts give the total area of all incoming regions producing deflections above threshold, which can be a collection of annuli. No unique $b(\theta)$ follows from that one number. Near a smooth extremum $\theta(b) = \theta_r + c(b-b_r)^2/2 + \cdots$, two branches contribute inverse square-root factors proportional to $1/\sqrt{|\theta-\theta_r|}$. A finite bin contains finite counts because integrating that singularity gives a square root. Tagging the incoming impact parameter in a controlled beam or measuring trajectory time delay can separate branches invisible in a final-angle histogram.

Invert a signed deflection function at one energy

Assume smooth $U(r)$ with $\rho(r) = r\sqrt{1-U(r)/E}$ strictly increasing on the radial interval sampled by the trajectories. The turning condition is

$$\rho(r_0) = b.$$

Write $W(\rho) = \ln[r(\rho)/\rho]$ and transform the deflection integral to $\Phi = b\int_b^\infty [d\ln r/d\rho]\, d\rho/\sqrt{\rho^2-b^2}$. The term from $d\ln\rho/d\rho$ gives the free angle $\pi/2$; the remainder has $\chi = -2b\int_b^\infty W'(\rho)\, d\rho/\sqrt{\rho^2-b^2}$. This is an Abel-type map. Multiply by $1/\sqrt{b^2-s^2}$ and integrate from $b=s$ to infinity. Interchanging the integrals produces the inner integral

$$\int_s^\rho \frac{b\,db}{\sqrt{b^2-s^2}\sqrt{\rho^2-b^2}} = \frac{\pi}{2}.$$

Substitute $b^2 = s^2 + (\rho^2-s^2)\sin^2\alpha$ to check the identity. With $W(\infty)=0$ the inversion is $W(s) = \pi^{-1}\int_s^\infty \chi(b)\, db/\sqrt{b^2-s^2}$. Recover $r = se^{W(s)}$ and finally $U(r) = E[1-(s/r)^2]$ parametrically. This derivation explains the restrictions. A trajectory only samples radii outside its turning point, the signed angle must retain branch information, and the map to ρ must be single valued. Unknown large-b tails influence the integral even when only comparatively small radii are of interest.

Derive the weak-deflection formula from a straight incoming trajectory

When the deflection is small one can calculate the transverse impulse along the undeflected line before evaluating a nonlinear radial integral. Put the incoming velocity along x and the impact point at $y = b > 0$, so the unperturbed trajectory is $x = v_\infty t$, $r = \sqrt{x^2+b^2}$. The transverse force is $-U'(r)b/r$. Dividing its time integral by the incoming momentum μv_∞ gives the signed leading deflection

$$\chi(b) \simeq -\frac{b}{E}\int_b^\infty \frac{U'(r)\, dr}{\sqrt{r^2-b^2}}.$$

The factor follows from replacing dt by dx/v_∞, using symmetry between positive and negative x, and the identity $dx = r\,dr/\sqrt{r^2 - b^2}$. A decreasing positive potential gives a positive angle, consistent with repulsive scattering in the chapter's convention. The approximation neglects longitudinal momentum change at leading order; a finite speed change far from the interaction would also be inconsistent with elastic energy conservation. Near a turning barrier or for a large angle the original deflection integral must replace this straight path.

An analytically tractable finite-range model is $U(r) = U_0 \exp[-r^2/(2a^2)]$. Along the incoming straight line the transverse force is $(U_0 b/a^2)\exp[-(x^2 + b^2)/(2a^2)]$. The Gaussian x integral gives

$$\chi(b) \simeq \sqrt{\pi/2}(U_0/E)(b/a)\exp[-b^2/(2a^2)].$$

A central hit has zero transverse force by symmetry if the particle crosses through the weak smooth potential rather than being reflected. Distant particles see little interaction, so the deflection must peak at an intermediate impact parameter. Differentiation places the maximum at $b = a$, with height $\sqrt{\pi/(2e)}U_0/E$.

For $U_0/E = 0.10$ the peak is about $0.0760\,\mathrm{rad} = 4.36^\circ$, comfortably small for this leading calculation. Two impact-parameter families can contribute to one smaller nonzero angle. For example $b = 0.50a$ gives $0.0553\,\mathrm{rad}$; a trajectory with b between $1.5a$ and $1.7a$ produces a similar deflection on the declining branch. A detector counting more events near four degrees may be recording the merging of these branches, rather than a particularly large local force at the turning separation. Reconstructing $b(\theta)$ from monotonically accumulated disk areas would be invalid for this smooth repulsive bump even though every encounter is elastic.

Measure the energy dependence to distinguish the weak limit. For fixed b/a the leading angle scales as $1/E$. Doubling energy halves the predicted peak angle. A finite spread of incident energies therefore broadens the observed rainbow before detector angular resolution is considered. A position-tagged incident beam can furnish $\chi(b)$ point by point; a detector measuring final angles alone mixes weakly and more strongly penetrating paths. Adding incoming-position information changes what inverse problem is identifiable, without improving the detector's angular precision.

Limit a scattering inference using finite range and measurement bandwidth

The Abel reconstruction requires data to arbitrarily large b. A finite experiment measures up to $b_{\max}$, and the unmeasured tail contributes an integral to $W(s)$. For the example of a known asymptotic $\chi = C/b$, the tail from $b_{\max}$ to infinity evaluates to $C\,\arcsin(s/b_{\max})/(\pi s)$ for $s < b_{\max}$. Substitute $b = s/\cos\alpha$ to obtain this result. For $s \ll b_{\max}$, it approaches $C/(\pi b_{\max})$. Setting the tail to zero therefore introduces an offset across the reconstructed log-radius even at small s. Bounded angular noise and uncertainty in the tail model should be propagated separately.

As a numerical illustration take $C = 5.0\,\mathrm{mm}$,

$$b_{\max} = 100\,\mathrm{mm},$$
$$s = 20\,\mathrm{mm}.$$

The missing contribution is about 0.0160, corresponding to roughly a 1.6% multiplicative error in reconstructed radius if other terms were exact. At that same s, the leading whole weak-tail value $C/(2s) = 0.125$ would be biased appreciably by this omission. A model with a short exponential tail would give a quite different correction. Adding a physically motivated tail with an uncertainty range is more informative than extending a noisy measured curve by an unrestricted polynomial.

The lower integration endpoint has a square-root weight. Uniform samples in b can be wasteful near a selected reconstruction radius and coarse interpolation there can dominate the result. Substitute $b = s\cosh u$ to obtain

$$db/\sqrt{b^2 - s^2} = du.$$

The Abel integral becomes a straightforward integral of $\chi(s\cosh u)$ over u, with a finite upper limit if the data stop at $b_{\max}$. This substitution removes a numerical singularity but adds no information about missing impact parameters. Validate a reconstruction by inserting its inferred potential back into the original forward integral and predicting angular measurements withheld from the fit. Agreement with the very data used in the inversion alone can conceal an incorrect tail or a misassigned branch.

Convert centre-of-momentum angles into laboratory angles

An ordinary collision experiment often launches a particle of mass m_1 into a stationary particle of mass m_2. In Newtonian elastic scattering the centre-of-mass speed is $V = m_1 v_{\mathrm{in}}/(m_1 + m_2)$, while the projectile's speed in that frame is $u_1 = m_2 v_{\mathrm{in}}/(m_1 + m_2)$. Rotate that velocity by the centre-of-mass angle χ in the scattering plane, then add the centre velocity back. The outgoing laboratory velocity components are $V + u_1\cos\chi$ longitudinally and $u_1\sin\chi$ transversely. Consequently $\theta_{\mathrm{lab}} = \operatorname{atan2}(\sin\chi, m_1/m_2 + \cos\chi)$. The centre-of-mass solid-angle density derived from the impact parameter does not transfer unchanged to a lab angular histogram.

If $m_1 = m_2$, the identity $1 + \cos\chi = 2\cos^2(\chi/2)$ gives $\theta_{\mathrm{lab}} = \chi/2$ except when the projectile stops and loses a well-defined direction. A heavy projectile can be restricted to a narrow forward cone. Let $r = m_1/m_2 > 1$. Differentiating the laboratory angle gives

$$d\theta_{\mathrm{lab}}/d\chi = (1 + r\cos\chi)/(1 + 2r\cos\chi + r^2).$$

It vanishes at $\cos\chi = -1/r$. At that angle $\sin\theta_{\max} = 1/r$; hence $r = 2$ limits the projectile to $\theta_{\max} = 30^\circ$. Two centre-of-mass angles can generate the same nonzero laboratory angle below this maximum. This laboratory-frame fold may appear even for a monotone centre-of-mass repulsive potential.

Energy helps separate the branches. The outgoing laboratory projectile energy divided by its incident energy is $[m_1^2 + m_2^2 + 2m_1 m_2\cos\chi]/(m_1 + m_2)^2$. For $m_1 = 2m_2$, head-on reversal in the momentum frame gives energy fraction $1/9$, while forward scattering gives one. Both can land near the laboratory

beam direction. A detector reporting angle only cannot distinguish a barely deflected fast projectile from a strongly deflected slow one emerging along almost the same line. Recording time of flight to a downstream detector or measuring the recoil supplies the missing information. Instrumental and kinematic multiplicity should therefore be addressed before attributing a two-branch angular signal to an exotic force law.

Ask what a flight-time measurement adds to angle data

At a given impact parameter choose a large comparison sphere of radius R enclosing a short-range potential. The transit time from its incoming boundary to the outgoing boundary is $t_U(R,b) = 2v_\infty^{-1}\int_{r_0}^{R} dr/\sqrt{1-U/E-b^2/r^2}$. The free reference at the same b is $t_0 = 2\sqrt{R^2-b^2}/v_\infty$. Their difference is a classical time delay under a precisely specified pair of surfaces. It combines a different path length with the altered speed along the path. A repulsive barrier can cause either a shorter geometrical penetration or longer dwell near a turning point, so the sign cannot be inferred merely from the word repulsive.

For an impenetrable sphere of radius $a > b$, the particle moves at free speed on each straight segment. The radial turning point is replaced by specular contact at a, so $t_{\mathrm{hard}} = 2[\sqrt{R^2-b^2}-\sqrt{a^2-b^2}]/v_\infty$. The delay is $-2\sqrt{a^2-b^2}/v_\infty$. Using $a = 3.0\,\mathrm{mm}$, $b = 1.8\,\mathrm{mm}$, and $v_\infty = 2.0\,\mathrm{m/s}$, one finds $-2.4\,\mathrm{ms}$. The hard sphere takes less transit time between the two points where the respective trajectories cross the large sphere. Those points differ from the endpoints of the free line, so this negative delay implies no acceleration above the fixed free speed.

For a potential with a true long-range $1/r$ tail the simple free-reference difference can acquire a logarithmic dependence on the comparison radius. A reported time delay then needs its reference geometry and tail model, just as a measured forward angle needs its angular acceptance. In the inverse problem, combining deflection and delay can help separate long dwell near a nearly unstable circular orbit from a large geometrical bend at an ordinary turning point. A large angle alone contains insufficient information to assign the time spent close to the scattering centre.

Reconstruct a finite-depth potential well from two deflections

A circularly symmetric square well provides a forward problem in which the trajectory is straight in each region and changes direction at the potential boundary. Let $U(r) = -U_0 < 0$ for $r < a$ and zero outside. A particle with $b < a$ enters with speed v_∞ and travels inside at $v_{\mathrm{in}} = v_\infty\sqrt{1+U_0/E} \equiv n v_\infty$. Angular momentum conservation across the radial impulse gives

$$v_\infty \sin i = v_{\mathrm{in}} \sin r_i,$$

where i and r_i are angles of the velocity to the normal outside and inside. Thus $\sin i = b/a,\ \sin r_i = b/(na)$. At each crossing the direction bends toward the centre. The magnitude of net deflection is $2(i - r_i)$, with negative signed deflection in the convention used for attraction here.

Take $n = 1.25,\ b = 0.60a$. Then $i = 36.87^\circ$, $r_i = 28.69^\circ$, and the total deflection has magnitude 16.35°. As b approaches a from below, the magnitude approaches $2[\pi/2 - \arcsin(1/n)]$, about 73.7°.

For $b > a$ the trajectory does not touch the ideal sharp boundary and is undeflected. A smooth finite-width transition removes the ideal discontinuity at grazing incidence and can generate trajectories that linger near the outer part of the well. Examining a well with measured edge width is therefore different from trying to fit all grazing events with a perfectly sharp step.

If a is known, an incoming position tag and a measured deflection at one b determine $n = \sin i / \sin(i - |\chi|/2)$, hence $U_0/E = n^2 - 1$. The preceding numbers give

$$U_0/E = 0.5625.$$

For an independently known E this determines the well depth. If a is unknown too, one angle supplies only one equation for two parameters. Multiple impact parameters, an independent threshold for entering the well, or a measured time of flight can supply the additional scale. Near $b = 0$ the deflection is approximately $2(b/a)(1 - 1/n)$, so a series of very small impact parameters measures mainly that one combination; broad coverage is needed to resolve a and n individually.

The time inside the well is $2a \cos r_i/(nv_\infty)$, because the interior chord has offset b/n and speed nv_∞. At the same b, the time between intersections of the free path with the circle is $2a \cos i/v_\infty$. For $n = 1.25, b = 0.6a$, their ratio is $\cos 28.69^\circ / [1.25 \cos 36.87^\circ] \approx 0.877$. The faster interior motion crosses a longer chord in less time. For grazing free paths the reference chord becomes very short and the comparison can reverse. Reporting which part of the trajectory the timing equipment resolves is therefore as important as measuring the dwell accurately.

Recover incoming annuli from scattering angles

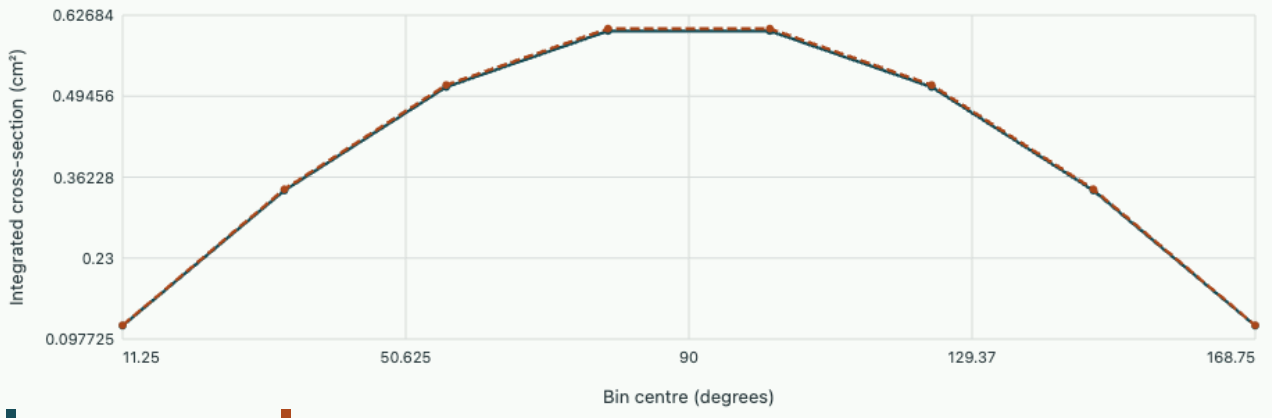
Bin the exactly solvable hard-sphere reflection in polar angle and compare the integrated cross-section with the incoming geometric area.

Open this experiment online

Try the calculation

Polar detector bins 8

Effective radius (cm) 1



Exact sum 3.1416 cm². Midpoint sum 3.1619 cm². Differential cross-section 0.25 cm²/sr.

Propagate uncertainties from a detector ring to an inferred force

Suppose N counts are recorded in an angular ring and the product of incident fluence, target count and calibrated efficiency is C. The estimate of bin-averaged cross section is $N/(C\Delta\Omega)$, with $\Delta\Omega = 2\pi[\cos\theta_{\mathrm{low}} - \cos\theta_{\mathrm{high}}]$ for full azimuth. Approximately independent counting events have fractional standard error $1/\sqrt{N}$, but the fluence and efficiency create correlated scale errors across angular bins. When integrating bins to reconstruct $b^2(\theta)$, a shared five-percent normalisation uncertainty does not shrink just because the number of angular bins grows. It propagates into an approximately 2.5% scale uncertainty in the recovered impact parameter.

For an ideal hard sphere, let an experiment integrate the area feeding angles greater than 90° and find $14.1 \pm 0.7\,\mathrm{mm}^2$. Since this is $\pi b(90^\circ)^2 = \pi a^2/2$, the inferred radius is $\sqrt{2(14.1)/\pi} = 3.00\,\mathrm{mm}$ with uncertainty approximately $0.075\,\mathrm{mm}$ for small errors. If the unobserved backward final ten degrees are simply omitted, an extra solid angle $\Delta\Omega = 2\pi[1 - \cos 10^\circ]$ is missed. Its small size for a smooth hard-sphere law can be estimated, but a general inverse analysis cannot assume that backward intensity is smooth without testing it.

Is multiplying each measured bin centre by the angular density there sufficient for a steep forward-scattering law?

▼ **Work through the solution**

A finite bin measures an integral against solid angle, not the value at its centre. For a rapidly varying $\sin^{-4}(\theta/2)$ density near zero, the lower-angle part of the bin can dominate its count. Use the exact preimage annulus $\pi[b(\theta_{\mathrm{low}})^2 - b(\theta_{\mathrm{high}})^2]$ or integrate the model numerically over the detector's angle-dependent efficiency. Repeat with narrower bins as a check. If the lower acceptance overlaps the direct

beam, separate the unscattered population and account for its instrumental angular spread before assigning a cross section to small deflections.

Work directly in counts for a finite annular detector

For the repulsive inverse-square example take $d = 2.0\,\mathrm{mm}$ and a full-azimuth ring that accepts polar angles from 55° to 65°. The expected incident area is $\pi d^2[\cot^2(27.5^\circ) - \cot^2(32.5^\circ)]$, about $15.4\,\mathrm{mm}^2$. Multiplying by an incident fluence of 100 particles per square millimetre and efficiency 0.80 predicts about 1230 counts. A Poisson fluctuation scale is then roughly 35 counts. The exact integral supplies a meaningful reference even when the plotting software reports a differential value at the centre of the bin.

Moving both angular boundaries one degree outward decreases the count because the repulsion sends fewer particles through larger angles. The change can be calculated by reevaluating the two cotangents, keeping the fluence and efficiency fixed. This sensitivity allows a calibration uncertainty in the angular zero to be propagated into a count prediction without treating it as a force uncertainty. If fluence is independently measured the dependence on d^2 can then estimate the interaction scale. If fluence is itself unknown, a single count measures only their product; independent calibration or measurements across energies are needed to separate them.

Tutorial on inverse scattering

An incoming uniform beam scatters from a hard disk in three-dimensional impact geometry with effective radius $3.0\,\mathrm{mm}$. What area feeds the polar bin from 60° to 90°?

▼ Work through the solution

The accepted impact parameters run from $a\cos 45^\circ$ to $a\cos 30^\circ$. The difference of their squared radii is $9(0.75 - 0.50) = 2.25\,\mathrm{mm}^2$, giving area $2.25\pi = 7.07\,\mathrm{mm}^2$. Alternatively the bin has solid angle $2\pi(\cos 60^\circ - \cos 90^\circ) = \pi$ steradians and differential cross section $a^2/4 = 2.25\,\mathrm{mm}^2/\mathrm{sr}$. The agreement checks both the Jacobian and the angular convention.

At large impact parameter a weak repulsion gives $\chi(b) \simeq C/b$ with C a length. Use the inversion to identify the leading tail of the potential.

▼ Work through the solution

Substitute $b = s/\cos\alpha$ into the Abel integral. One obtains

$$W(s) \simeq C/(2s).$$

Thus $r \simeq s + C/2$ and $U/E = 1 - e^{-2W} \simeq C/s \simeq C/r$. The potential tail is EC/r. For the exact repulsive Coulomb result $C = K/E$, matching the original K/r. This check uses the weak-tail expansion at large radii only; continuing it to small radii would discard higher deflection terms.

A measured angular bin near a rainbow doubles in width towards its allowed side. By what approximate factor do its counts change if background is negligible?

▼ **Work through the solution**

At a simple fold the angular density grows as the inverse square root of distance from the extremal angle. For a bin of width $\Delta\theta$ starting at the fold, the integral scales as $\sqrt{\Delta\theta}$ provided $\sin\theta_r$ is nonzero and other weights are locally smooth. Doubling the width gives a factor $\sqrt{2}$, unlike the factor two from a smooth background. Finite angular resolution rounds off the divergence and can obscure this scaling in very narrow bins.

Compare angular bins and conserved incoming area.

Fließbach's mechanics text treats trajectories in central fields.

Which part of a Bell correlation survives imperfect state preparation?

Computational quantum experiments

A source emits two spin-half systems in the singlet state. Two observers select measurement directions separately and record outcomes $+1$ or -1. Construct a Bell–CHSH statistic from four measurable correlations and determine how state preparation, angular alignment and finite sample size affect the prediction.

Relate a counted coincidence to an operator

Denote unit measurement directions by $\mathbf{a}, \mathbf{b}$. Outcome projectors are $(1 + s\,\boldsymbol{\sigma}\cdot\mathbf{a})/2$ and $(1 + t\,\boldsymbol{\sigma}\cdot\mathbf{b})/2$, for $s, t = \pm 1$. The singlet has zero individual spin expectation and tensor $\langle \sigma_i \otimes \sigma_j \rangle = -\delta_{ij}$. The diagonal entries follow because measurements along any common axis are exactly opposite and the state is rotationally invariant. Its joint probability is

$$p(s,t|\mathbf{a},\mathbf{b}) = \tfrac{1}{4}(1 - st\,\mathbf{a}\cdot\mathbf{b}),$$

$$E(\mathbf{a},\mathbf{b}) = -\mathbf{a}\cdot\mathbf{b}.$$

Summing over either outcome leaves one half for the other observer, independent of the remote angle. Covariance can change while the individual counting rates remain constant. At an angle of 60 degrees the probabilities of the two equal-outcome combinations sum to one quarter, and of opposite outcomes to three quarters. Their signed difference gives –0.5. The lab models spin-axis angles; linear polarisation analysers give a doubled angle in their effective two-state representation.

Derive a bound and choose a set of directions

If a trial had four predetermined local binary responses A, A', B, B', its combination $AB + AB' + A'B - A'B' = A(B + B') + A'(B - B')$ would have magnitude two, since exactly one of the parentheses vanishes. Averaging over a common distribution of such assignments preserves the bound. The quantum prediction combines expectation values measured on different subensembles with independently chosen setting pairs. Use coplanar angles $a = 0,\ a' = \pi/2,\ b = \theta,\ b' = -\theta$.

$$S = E(a,b) + E(a,b') + E(a',b) - E(a',b') = -2(\cos\theta + \sin\theta).$$

At $\theta = 45^\circ$ its magnitude is $2\sqrt{2}$. The derivative vanishes there, making small symmetric angular errors second order. If a fraction $1 - v$ of pairs is completely unpolarised, the density operator is $v|\Psi^-\rangle\langle\Psi^-| + (1 - v)I/4$. Every correlator gains a factor v. The best attainable magnitude exceeds two only when $v > 1/\sqrt{2}$. This is a statement about this measurement family and this noise model. A pair source and a scalar contrast alone do not define an experimental protocol.

Assess one set of finite counts

At $v = 0.80$, $\theta = 45^\circ$ the expected magnitude is about 2.263. Each trial product is a binary variable with variance $1 - E^2$. For N independent pairs per setting, the sample correlation variance is $(1 - E^2)/N$. Assuming independent subensembles, variance of the signed sum is the sum of the four variances. Here every squared correlation is 0.32. With 1000 trials per setting the approximate standard error of the statistic is

$$\sqrt{4(1 - 0.32)/1000} = 0.0522.$$

The excess 0.263 is roughly five standard errors under this idealised independent-trial model. It is not a substitute for a finite-sample hypothesis test that accounts for an actual protocol's setting choices, losses, timing and possible memory.

Change settings and preparation independently

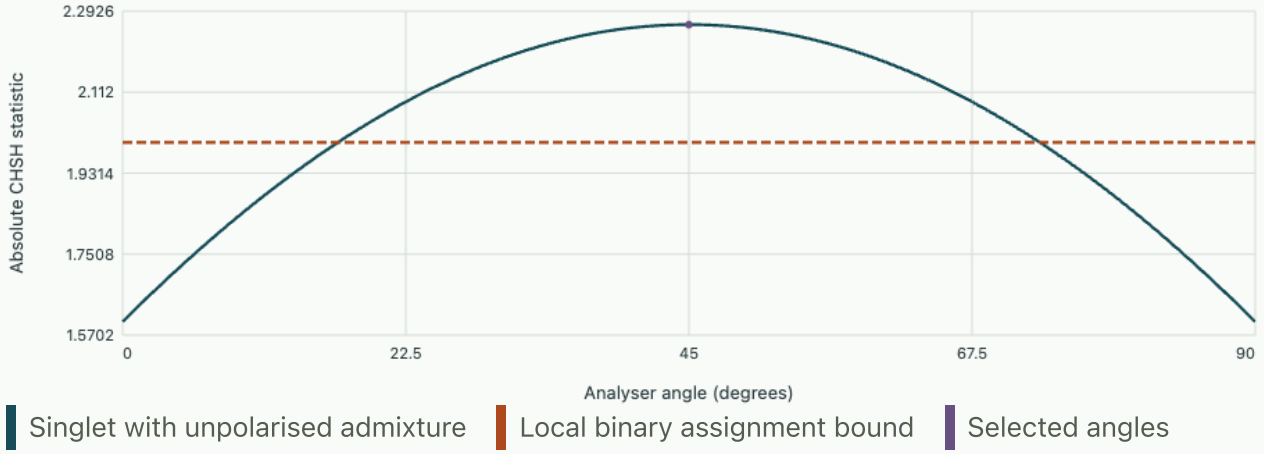
Inspect the predicted absolute statistic and the local assignment bound. Find the angles for which the chosen source can exceed two. Then vary visibility at a fixed angle. Small individual spin polarisation drifts could alter the marginal counts; unequal channel efficiencies could change a coincidence-selected sample. Monitoring all four joint count tables and both single-observer marginal distributions allows more checks than retaining only their final sum.

Open this experiment online

Try the calculation

Preparation visibility **0.8**

Symmetric analyser angle (degrees) **45**



Predicted |S| = 2.2627. With these angles the visibility required to reach two is 0.70711.

Tutorial

At visibility 0.70, can increasing the number of trials eventually make the expected maximum exceed two?

▼ **Work through the solution**

No. The largest mean value is $2\sqrt{2}(0.70) = 1.980$. More trials narrow the sampling distribution around that value. Changing the preparation or measurement family addresses a different question.

What is the predicted fractional loss at an angle δ away from the optimum, and how large is it at five degrees?

▼ **Work through the solution**

Write $\cos\theta + \sin\theta = \sqrt{2}\cos(\theta - \pi/4)$. The fractional loss is $1 - \cos\delta$, about 0.00381 at five degrees. Asymmetric offsets on different analyser axes need not enter the same way.

A paper gives one large value of $|S|$ but discards all pairs for which one detector was silent. Can the ideal bound derived above automatically be applied to these retained counts?

▼ **Work through the solution**

The derivation averaged all setting pairs over a common underlying ensemble. Setting- or state-dependent retention can alter those ensembles. One needs the trial definition, data handling and efficiency assumptions to interpret the reported statistic.

For state preparation and independent measurement on tensor factors, revisit the composite-state chapter.

Recover a transition frequency from two separated pulses

Computational quantum experiments

A two-level system is driven by two short pulses with a free evolution interval between them. Determine why a population measurement carries phase information. Use a scan at several delays to separate a constant detuning, intrinsic coherence loss and trial-to-trial frequency variation.

Follow the amplitudes through both pulses

Use basis $|g\rangle, |e\rangle$. Let the first pulse prepare $(|g\rangle + |e\rangle)/\sqrt{2}$ in the frame of the applied oscillator. A detuning Δ in radians per second accumulates a relative phase $\Delta\tau$ in free time τ. Immediately before the second pulse each state has population one half; measuring population here cannot reveal detuning. Choose a second pulse equivalent to projecting onto $(|g\rangle + e^{-i\phi}|e\rangle)/\sqrt{2}$. Then one output population is

$$P(\tau, \phi) = \tfrac{1}{2}[1 + \cos(\Delta\tau - \phi)].$$

The sign of the phase follows our definitions of detuning and the reference pulse. Shifting the reference phase makes it possible to check that sign. A measurement using only $\phi = 0$ cannot distinguish Δ from $-\Delta$. Coherence is the off-diagonal density-matrix element that the second pulse converts into an observable population difference.

Derive two envelopes with different origins

A homogeneous decay time T_2 multiplies the off-diagonal element by $e^{-\tau/T_2}$. If instead the detuning of each experimental shot is constant during that shot and drawn from a Gaussian distribution of mean $\bar{\Delta}$ and standard deviation σ_Δ, average the complex phase, not its absolute value. Completing the square in the Gaussian integral gives

$$\langle e^{i\Delta\tau} \rangle = e^{i\bar{\Delta}\tau} e^{-\sigma_\Delta^2 \tau^2/2}.$$

With both independent mechanisms the measured contrast is $C(\tau) = \exp[-\tau/T_2 - \sigma_\Delta^2 \tau^2/2]$. One mechanism gives an initial logarithmic slope, the other adds curvature. A single low contrast measured at just one delay cannot establish which one caused it.

Infer parameters from quadratures and a frequency scan

Record the $\phi = 0, \pi/2$ outputs, then form $X = 2P_0 - 1 = C\cos(\bar{\Delta}\tau)$ and $Y = 2P_{\pi/2} - 1 = C\sin(\bar{\Delta}\tau)$. Their length estimates contrast and their angle estimates accumulated phase modulo 2π. Suppose at $0.50\,\mathrm{ms}$ the readings are 0.500 and 0.850. Then $X = 0,\ Y = 0.700$, and the phase is $\pi/2$ modulo 2π. With an independent bound $|\bar{\Delta}/2\pi| < 1\,\mathrm{kHz}$, the implied cyclic detuning is $+500\,\mathrm{Hz}$. Longer delays resolve fine frequency differences at the cost of phase ambiguity and declining contrast. Use a shorter-delay record to unwrap a longer one.

At a contrast of 0.700 the relation $-\ln(0.700) = \tau/T_2 + \sigma_\Delta^2 \tau^2/2$ supplies one equation for two unknowns. Several delays, with independent error bars, are needed to identify the linear and quadratic terms. At very low signal the length $\sqrt{X^2 + Y^2}$ has a positive noise bias. Fitting the two directly measured quadratures can be preferable to fitting the length.

Explore two mechanisms

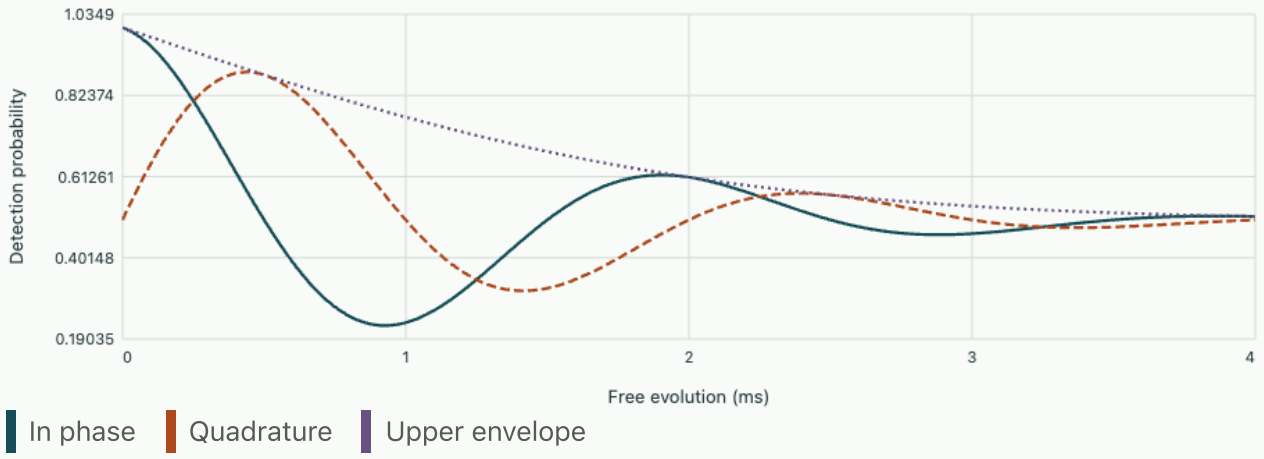
The graph holds the homogeneous T_2 at 2 ms. It lets you set signed cyclic detuning and shot-to-shot cyclic frequency spread. Compare both quadrature populations with the upper contrast envelope. First reverse the sign of the detuning, then increase the spread without changing the mean. Which effect interchanges peaks of one quadrature and which damps both?

Open this experiment online

Try the calculation

Mean cyclic detuning (Hz) **500**

Shot-to-shot rms spread (Hz) **80**



Contrast after 1 ms = 0.53455. The two probabilities at 0.50 ms are 0.5 and 0.87729.

Tutorial

What is the fringe spacing in a scan of cyclic detuning at fixed delay of 0.80 ms?

▼ **Work through the solution**

One full phase cycle needs $\delta f\, \tau = 1$. The spacing is

$$1/0.00080 = 1250\,\mathrm{Hz}.$$

The angular-frequency spacing is 2π times larger.

A fixed delay gives

$$P_0 = P_{\pi/2} = 0.5.$$

Does that show a phase of 45 degrees?

▼ Work through the solution

Both centred quadratures are zero, so the contrast is unresolved and the angle is undefined. A phase of 45 degrees with finite contrast gives equal, nonzero centred quadratures.

Explain how a refocusing pulse halfway through the evolution helps identify quasistatic detuning noise.

▼ Work through the solution

The pulse exchanges the phase-accumulating states, reversing phase winding for a detuning that stays constant through both halves. Random variations during the interval need not cancel. Recovering contrast in the refocused sequence supports a slowly varying contribution to the original envelope.

Compare a continuous drive in the transition chapter with the role of the second pulse here.

Transport coefficients from equilibrium fluctuations

Statistical mechanics from microscopic counting to material response

How can a fluid in equilibrium reveal the diffusion constant that would govern a spreading drop, or the viscosity that would resist an applied shear? Derive an exact connection between displacements and velocity correlations, generalise it to stress and heat currents and test how a finite record can bias each number. The assumptions behind convergence matter as much as the prefactor.

Integrate the velocity correlation before taking a long-time limit

Observe one Cartesian coordinate of a tagged particle in a homogeneous equilibrium fluid with zero mean drift. The displacement over time t is the integral of its velocity. Square that integral, average over the equilibrium ensemble and use stationarity to replace the two-time expectation by $C_v(t_1 - t_2)$. For a scalar component in a time-reversal-symmetric equilibrium state this correlation is even. Integrate over the two triangles of the time square:

$$\langle [x(t) - x(0)]^2 \rangle = 2 \int_0^t (t - s) C_v(s)\, ds,$$

$$\frac{1}{2} \frac{d}{dt} \langle \Delta x^2(t) \rangle = \int_0^t C_v(s)\, ds.$$

If the correlation is integrable and boundary effects remain absent, the long-time slope yields a diffusion coefficient $D = \int_0^\infty C_v(s) ds$. In d isotropic dimensions one may use $D = (1/d) \int \langle \mathbf{v}(s) \cdot \mathbf{v}(0) \rangle ds$ or the mean-square-displacement slope divided by $2d$. Using the full vector dot product without division by dimension gives an overestimate. Directional diffusion in an anisotropic material requires tensor components rather than a forced scalar average.

The derivation is kinematic and does not assume a Markovian motion. A fluid with velocity memory or an oscillating solid still obeys the finite-time relation. Those systems need not exhibit the same long-time proportionality to time. A bound oscillator can have persistent or alternating velocity correlations and a bounded displacement. An observation window in which the mean square grows over a fraction of an oscillation is insufficient to claim diffusion.

A solvable benchmark from ballistic to diffusive motion

For one coordinate take the underdamped Langevin equation $m\dot{v} = -\zeta v + \xi$, with zero-mean white noise satisfying $\langle \xi(t) \xi(t') \rangle = 2 k_B T \zeta\, \delta(t - t')$. In equilibrium $\langle v^2 \rangle = k_B T / m$. Condition on $v(0)$ and integrate for positive time. The later random force has no correlation with the initial velocity, giving $C_v = (k_B T / m) e^{-t/\tau_v}$, where $\tau_v = m/\zeta$. Its area and the resulting displacement are

$$D = \frac{k_B T}{\zeta},$$

$$\langle \Delta x^2(t) \rangle = 2D[t - \tau_v (1 - e^{-t/\tau_v})].$$

At $t \ll \tau_v$, the constant and linear terms inside the brackets cancel and displacement is ballistic:

$$\langle \Delta x^2 \rangle \simeq (k_B T/m) t^2.$$

At $t \gg \tau_v$ the slope approaches $2D$. Fitting a line through the origin at times only a few relaxation times overestimates the influence of the negative intercept in this long-time expression and underestimates D. Comparisons at multiple lags distinguish crossover curvature from a true anomalous power law.

For an inert sphere of radius 0.50 micrometres and density 2000 kg per cubic metre in water with $\eta = 1.00$ mPa s at 298 K, its bare mass is 1.05×10^{-15} kg and Stokes drag 9.42×10^{-9} kg/s. The estimate $\tau_v = 0.111$ microseconds is far faster than ordinary video tracking. $D = 0.437$ square micrometres per second and the one-coordinate rms displacement over 10 ms is 0.0935 micrometres. Hydrodynamic added mass and memory affect very short-time behaviour; the simple exponential is an analytical reference, not a claim that a water suspension remains Markovian down to arbitrary times.

Relate a weak imposed force to equilibrium motion

In the same soluble model a small constant force F turns the mean velocity into $F(1 - e^{-t/\tau_v})/\zeta$, so the mobility $b = \lim \langle v \rangle / F = 1/\zeta$. Together with equilibrium diffusion this gives the Einstein relation $D = k_B T b$. A broader equilibrium argument considers a dilute suspension in a weak conservative potential $U(x)$. Its current is

$$j = -D\,\partial_x n + bFn,$$

with $F = -\partial_x U$. The Boltzmann distribution $n \propto e^{-U/k_B T}$ must carry zero current; substitution gives the same relation without assuming exponential velocity relaxation.

Linear response connects the full weak-force impulse response to an equilibrium correlation. For a perturbation coupled as $-F(t)x$, classical equilibrium response of velocity has causal kernel $C_v(t)/(k_B T)$ for $t > 0$ under the usual microreversibility assumptions. Integrating the response kernel for a constant weak force gives

$$b = (1/k_B T) \int_0^\infty C_v(t) dt,$$

which recovers the kinematic diffusion relation. Strong forcing, an active bath or a nonequilibrium steady state need not obey this simple equilibrium identity.

A direct check compares independently measured weak sedimentation or calibrated drift with freely diffusing trajectories. For the previous bead a force 10 fN would produce a mean drift about 1.06 micrometres per second in the ideal Stokes regime. During a 10 ms frame its mean drift is about 0.011 micrometres, much smaller than rms Brownian motion, so many frames or many particles are needed. A single short jagged record can appear to move against the force. Use equal positive and negative applied forces to cancel constant stage drift in the estimated mobility.

Work through diffusion and mobility in the random-walk chapter

Stress and heat current carry their own normalisations

In a homogeneous isotropic laboratory fluid, define P_{xy} as a spatially averaged off-diagonal microscopic stress, with units Pa. Its equilibrium mean is zero in an isotropic state without imposed shear. A weak velocity gradient couples to momentum transport. The Green–Kubo expression for dynamic shear viscosity is

$$\eta = \frac{V}{k_B T} \int_0^\infty \langle P_{xy}(t) P_{xy}(0) \rangle \, dt.$$

If a program outputs an extensive stress moment $J_{xy} = V P_{xy}$, the prefactor is $1/(V k_B T)$. Multiplying an already extensive current by volume again is a common orders-of-magnitude error. Check dimensions: pressure squared times time times volume divided by energy gives pascal seconds. Kinetic and interaction contributions to microscopic stress must both be included according to the model; a calculation with ideal kinetic flux alone need not reproduce the viscosity of a dense fluid.

Similarly for a component of total heat current $J_{Q,x}$ with units energy times length per time, the thermal conductivity in an isotropic equilibrium single-component system is

$$\kappa = \frac{1}{V k_B T^2} \int_0^\infty \langle J_{Q,x}(t) J_{Q,x}(0) \rangle \, dt.$$

Here $k_B T^2$ means k_B times T^2. If a heat-flux density is used instead, the volume prefactor moves as for stress. Heat current must exclude energy carried by a net mass flow as appropriate to the chosen thermodynamic constraint. In mixtures several conserved species currents can couple, and the measured conductivity at zero mass flux can differ from one unprojected current coefficient. The meaning of the measured flux and the imposed condition must be specified before comparing coefficients.

For an illustrative volume 10^{-24} cubic metres at 300 K, assume one stress autocorrelation $C_P(t) = (2.0 \times 10^6 \ \mathrm{Pa})^2 e^{-t/(2.0 \ \mathrm{ps})}$. Its area is 8.0 square pascal seconds. Multiplication by $V/(k_B T)$ gives viscosity 1.93×10^{-3} Pa s. Doubling the simulated volume of an otherwise similar homogeneous fluid generally reduces the variance of its volume-averaged stress, keeping the physical transport coefficient intensive. Holding the same numerical stress variance while merely changing the prefactor would describe a different equilibrium fluctuation model.

Find a correlation integral's useful time window

An estimate at lag s averages over pairs of samples separated by that lag. In a record of duration T_{rec}, the number of available pairs decreases with lag. Overlapping time origins strongly correlate the estimates, so plotting one error bar per raw pair exaggerates precision. Integrating an estimated correlation to the maximum possible lag frequently accumulates late random noise rather than useful signal. Conversely stopping too early omits a physical tail.

For a positive single-exponential reference, integrating to three correlation times retains $1 - e^{-3} = 95.0$ percent of the area. Five times retain 99.3 percent. A correlation containing a small, much slower component can look essentially flat on the main decay scale while contributing appreciably to the area.

Divide a long run into blocks, compare running integrals and check several window lengths. Report the variation between independent blocks and any explicit tail model used beyond the measured interval.

At short lags, finite exposure averages correlated positions or velocities. Uncorrelated localisation errors in position increase nonzero-lag displacement variance by a constant $2s_x^2$ per coordinate. Numerically differentiating those same noisy positions creates large artificial velocity variance and negative correlations between adjacent increments. An apparent negative velocity lobe can therefore be physical backscattering or a measurement artefact. Comparing a raw displacement fit with the velocity integral after independently characterising detector noise is more informative than differentiating first and interpreting every feature as fluid memory.

A forced shear wave gives a second viscosity measurement

An equilibrium stress correlation estimates viscosity without imposing any flow. A separate driven measurement can test that coefficient against a directly observed penetration depth. Take a Newtonian fluid filling $y > 0$, driven by a laterally oscillating plate at $y = 0$ with speed $U_0 \cos \omega t$. Neglect boundaries at large y, compressibility and nonlinear convective effects in this spatially uniform tangential shear. The momentum equation reduces to

$$\partial_t u = \nu \partial_y^2 u,$$

with kinematic viscosity $\nu = \eta/\rho$.

Seek $u = \mathrm{Re}[U_0 e^{-\lambda y - i\omega t}]$ and choose the square root with positive real part. Then $\lambda = (1 - i)/\delta$,

$$\delta = \sqrt{2\nu/\omega}.$$

Both amplitude decay and phase lag carry information: at distance one penetration depth the amplitude is e^{-1} of the wall speed and the lag magnitude is one radian. A single complex velocity map can determine viscosity from either slope and use their agreement to test the simple model.

For a fluid of density 1000 kg per cubic metre and measured penetration depth 0.126 mm at 20 Hz, $\nu = \omega \delta^2 / 2 \simeq 1.00 \times 10^{-6}$ square metres per second and $\eta \simeq 1.00$ mPa s. A stress-fluctuation integral yielding a value twice as large would predict a penetration depth larger by $\sqrt{2}$, approximately 0.178 mm. Such a comparison can discriminate a missing component or volume factor in the equilibrium calculation if density, temperature and fluid composition match.

At high enough frequency the fluid may have a complex frequency-dependent viscosity rather than the zero-frequency coefficient from a Green–Kubo area. In linear response that generalisation is the one-sided Fourier transform of the equilibrium stress correlation with appropriate prefactors. A single-exponential stress memory gives a frequency-dependent response with a characteristic roll-off and phase lag. Comparing a high-frequency penetration measurement directly to a zero-frequency area without checking the timescale can therefore produce a genuine difference, not a statistical failure.

Remove a centre-of-mass mode only when the observable requires it

In a finite periodic particle simulation total momentum may be exactly conserved. If the initial centre of mass has a random velocity and no thermostat removes it, every particle's measured laboratory coordinate includes the common drift $V_{\mathrm{CM}}t$. Its mean-squared displacement gains a ballistic term $\langle V_{\mathrm{CM}}^2\rangle t^2$. In a canonical ensemble the centre-of-mass velocity variance in one direction is k_BT/M_{tot}. Though it falls with system size, at sufficiently long times its quadratic contribution exceeds the linear diffusive part.

For 1000 equal particles of mass m, an equilibrium centre-of-mass mode contributes one thousandth of the single-particle instantaneous velocity variance. In a long-time autocorrelation this small constant never integrates to a finite diffusion coefficient. Subtract centre-of-mass displacement or work in a zero-total-momentum ensemble before taking the long-time integral of an internal tracer's motion. In a real sample with externally imposed drift, decide whether the physical quantity sought is diffusion about that drift or absolute laboratory displacement; the subtraction has a physical meaning rather than merely a cosmetic improvement.

Conversely, indiscriminately subtracting the mean velocity of a tiny observation subregion separately at every time can remove genuine hydrodynamic long-wavelength motion. A strongly applied single-particle thermostat can alter current correlations and suppress transport even while maintaining the correct static temperature histogram. Check the thermostat strength and system size; compare an equilibrium-prepared weakly thermostatted trajectory with an energy-conserving production segment where appropriate. Maintaining the correct one-time distribution is not sufficient to preserve all dynamical correlation functions.

The distinction also matters when removing an estimated mean stress before integrating its fluctuations. The true shear mean may vanish by symmetry, but a short noisy record has a nonzero sample mean. Subtracting a slowly varying trend on the same timescale as a genuine stress tail can erase physical viscosity. Use independent samples, separately measured drifts and explicit sensitivity to the subtraction window. A persistent nonzero off-diagonal stress under nominal isotropic conditions may indicate incomplete equilibration, a deformed box or an unrelaxed solid rather than a fluid transport coefficient.

Heat and particle currents can share fluctuations

Consider two thermodynamic fluxes J_1, J_2 driven weakly by conjugate forces X_1, X_2, chosen so entropy production density is $J_1X_1 + J_2X_2$. Linear response writes $J_i = \sum_j L_{ij}X_j$. Near time-reversal-symmetric equilibrium, with flux parities chosen appropriately and no magnetic-field reversal complications, reciprocal relations give

$$L_{12} = L_{21}.$$

The symmetric part must be positive semidefinite so that entropy production is nonnegative for every small force. Thus $L_{11}, L_{22} \geq 0$ and $L_{12}^2 \leq L_{11}L_{22}$.

Suppose a thermal measurement enforces $J_1 = 0$ by allowing a compensating force X_1 to develop, as when a concentration or electrochemical gradient cancels a coupled particle flux. Solving the first row gives

$$X_1 = -L_{12}X_2/L_{11}.$$

Substitution into the second yields response coefficient $L_{22} - L_{21}L_{12}/L_{11}$. The measured constrained heat response can therefore be smaller than the coefficient at fixed $X_1 = 0$. The difference has a simple operational origin in the boundary condition enforced during the experiment.

Take consistent normalised coefficients $L_{11} = 4, L_{22} = 3, L_{12} = L_{21} = 2$. At zero first force the heat flux slope is three. At zero first flux it is $3 - 4/4 = 2$. Both are positive and consistent with the same equilibrium correlation matrix. Comparing a simulated heat-current autocorrelation corresponding to the unconstrained coefficient against a zero-particle-flux experiment requires the mixed-current correlation too. Apparent disagreement by a factor of one and a half can arise entirely from this distinction.

Green–Kubo cross-correlations must use properly defined conjugate fluxes and units. Redefining heat current by adding a constant multiple of particle current changes the matrix entries but must leave a physically defined zero-particle-flux heat measurement invariant when the forces are transformed consistently. This reference invariance is analogous to subtracting a per-particle energy reference in grand-canonical covariance. It provides a powerful test of a multicomponent transport implementation without relying on any empirical transport value.

A lattice trajectory provides an exactly known correlation integral

For a one-dimensional nearest-neighbour walk with steps of length a every interval Δt, suppose successive step directions have correlation $\langle s_{j+k}s_j\rangle = r^k$ with $|r| < 1$, and zero mean. Squaring the sum of n steps gives

$$a^2[n + 2\sum_{k=1}^{n-1}(n-k)r^k].$$

In the large-n limit divide by $2n\Delta t$ to obtain

$$D = a^2(1+r)/[2\Delta t(1-r)].$$

The uncorrelated random-walk value is recovered at $r = 0$.

With $a = 1$ micrometre and $\Delta t = 0.10$ s, uncorrelated diffusion is five square micrometres per second. Persistence $r = 0.5$ raises it to fifteen, while anticorrelation $r = -0.5$ lowers it to $5/3$. All three walks have the same step-size variance, so an estimate based only on one-frame displacement misses the entire correlation effect. The finite-lag autocorrelation sum and a long-time displacement fit should recover the same value on a long simulated record.

The discrete formula also clarifies the special zero-lag weight in numerical integration. An integral approximated from sampled symmetric correlations gives half weight to the zero lag on a one-sided

trapezoidal rule. A careless double count of the instantaneous self term predicts twice the uncorrelated diffusion. Benchmark a numerical estimator with this exact walk before applying it to molecular data with unknown late tails and noisy velocity derivatives.

Measure an off-diagonal diffusion coefficient after rotating the sample

In an anisotropic sheet a tracer's two-dimensional displacement covariance need not align with the camera axes. Let its long-time principal diffusion coefficients be D_1, D_2 and the first principal axis make angle θ with laboratory x. In matrix form the laboratory tensor is $R\,\mathrm{diag}(D_1, D_2)R^{\mathsf{T}}$, with R the planar rotation. At sufficiently large lag in an effectively unbounded drift-free region, the displacement covariance matrix tends to $2Dt$. An off-diagonal covariance is a measurable correlation between displacement components even if the noise is independent in the material principal directions.

Take $D_1 = 2.0$ and $D_2 = 0.50$ square micrometres per second with $\theta = 30^\circ$. Laboratory components are $D_{xx} = 1.625$, $D_{yy} = 0.875$ and $D_{xy} = 0.650$, in the same units. At 0.10 s the cross covariance is 0.130 square micrometres. Rotating the physical sheet by ninety degrees interchanges the two diagonal values and reverses the sign of the off-diagonal one. An instrument-fixed camera readout artefact would not generally follow this material rotation.

To infer the principal axes from data, fit each covariance component versus lag using a common drift subtraction and independent position calibration. Diagonalise the symmetric fitted diffusion matrix. Positive semidefiniteness requires

$$D_{xy}^2 \le D_{xx} D_{yy}.$$

A strong violation indicates incompatible estimates, unit conversion errors or statistical uncertainty, not a physically permissible strongly anisotropic diffusion tensor. Near-degenerate principal values make the orientation poorly determined even when the mean diffusion coefficient is precise. Report uncertainties on eigenvalues and orientation separately.

A finite field of view introduces a further bias. Fast particles along the high-diffusion direction leave the image sooner, and a tracking pipeline retaining only long trajectories can favour slow or fortuitously confined paths. Compare different fields of view and minimum track durations, or fit a measurement model incorporating censoring. A physical material anisotropy should persist after correcting these effects; apparent anisotropy aligned with the short direction of a rectangular observation window demands particular scrutiny.

A weak calibrated force along different axes supplies an independent check through the tensor Einstein relation under equilibrium conditions. If the diffusion tensor predicts off-axis mean drift, the mobility experiment should observe its sign and magnitude after reversing the force. A discrepancy could reveal active forcing, nonstationarity, an incorrect temperature or a mismatch between the mobility and displacement observables. It cannot be diagnosed from a scalar mean square alone.

Another rotational invariant is the trace of the diffusion tensor. In the example its value stays 2.50 square micrometres per second for every sample orientation, while the determinant stays one square micrometre

to the fourth per second squared. A comparison of two camera orientations therefore provides numerical checks independent of the fitted angle. A drift mistakenly left in the displacement covariance can violate the expected linear dependence on lag, making the inferred effective eigenvalues grow with the selected fitting window. Subtract the ensemble mean displacement at each lag and record that mean separately as a flow measurement. Conversely do not subtract an individual best-fit line from every short trajectory without modelling the resulting loss of long-time fluctuations; that procedure can bias both diffusion coefficients downward.

Choose an integration window for a transport coefficient

Integrate an oscillatory velocity correlation over a finite record. Compare a short apparent plateau with the independently computed infinite-time result.

Open this experiment online

Try the calculation

Correlation decay time τ (ps) 1

Oscillation frequency Ω (rad/ps) 1



With C(0) = 1 nm²/ps² the exact 1D diffusion coefficient is 0.5 nm²/ps. At 2τ the integral gives 0.58969 nm²/ps.

Tutorial · Invert diffusion measurements at two lags

In a one-dimensional diffusive window a measured displacement variance is 0.024 square micrometres at 10 ms and 0.064 at 50 ms. Assume long-time Brownian displacements plus independent frame-localisation variance per position measurement. Estimate D and localisation precision.

▼ **Fit slope and intercept separately**

The variance slope is $(0.064 - 0.024)/(0.050 - 0.010) = 1.00$ square micrometre per second, so $D = 0.500$ square micrometres per second. The intercept is 0.014 square micrometres, equal to twice the single-frame localisation variance. The implied rms localisation error is $\sqrt{0.007} = 0.0837$ micrometres. Forcing the fit through the origin at the shortest lag would give $D = 1.20$, more than twice the inferred value.

Acquire more lags to test the linear model. A static drift contributes a term growing as t^2 if its mean displacement is not removed, while finite exposure can lower the physically measured short-lag variance. Two points always define a line and cannot independently validate either noise model.

Tutorial · Estimate a hidden slow tail in viscosity

A stress correlation is modelled as $A[e^{-t/\tau} + 0.02e^{-t/(40\tau)}]$. What fraction of its true area comes from the low-amplitude tail? How much does a cutoff at 5τ capture?

▼ **Weight a correlation by its duration**

The fast and slow areas are $A\tau$ and $0.80A\tau$. The slow contribution is 44.4 percent of total. Integration to five fast times captures $A\tau[(1 - e^{-5}) + 0.80(1 - e^{-5/40})] \simeq 1.087A\tau$, or about 60.4 percent. Visual disappearance of the dominant peak therefore does not establish convergence of the transport coefficient.

The slow contribution must be supported by independent long blocks or a physically motivated model. A fitted tiny constant baseline can generate arbitrarily large apparent area if extrapolated indefinitely. Subtracting a sample mean estimated from the same short record can conversely suppress a genuine slow mode. Compare records much longer than the proposed tail time before trusting the extrapolation.

Tutorial · Is an algebraic tail compatible with ordinary diffusion?

Suppose a measured long-time velocity correlation remains positive and scales as $t^{-\alpha}$. Determine when its Green–Kubo area exists and infer the displacement growth for $0 < \alpha < 1$ and for $\alpha = 1$. Why would a finite periodic simulation require care?

▼ **Integrate the time-dependent slope before naming a diffusion coefficient**

For $\alpha > 1$ the tail integral converges. For $0 < \alpha < 1$, the running area grows as $t^{1-\alpha}$, so displacement variance grows as $t^{2-\alpha}$ under the stated persistent positive tail. At $\alpha = 1$, the running area grows logarithmically and the displacement variance is of order $t \log t$. Sign oscillations, finite-size termination or another late-time crossover can alter these conclusions.

A periodic box admits only discrete long-wavelength modes and may cut off a hydrodynamic tail. Repeating at larger sizes and comparing the range before signals traverse the box helps identify a bulk regime. Unwrap particle coordinates crossing periodic boundaries before calculating a displacement. A

wrapped coordinate can falsely saturate and imitate a confined particle even in a freely diffusing periodic fluid.

Cambridge Kinetic Theory · correlations, response and transport

Applied study

Infer parameters from noisy observations and use sensitivity to design a more informative experiment.

From predictions to parameter estimates

A forward problem takes material properties and predicts a response. In an inverse problem the responses have been observed and one or more material properties are unknown. Suppose a model predicts $g_i(p)$ at observation i, where p is a vector of parameters. Write y_i for the measured value and $s_i > 0$ for its estimated standard deviation. With independent Gaussian measurement errors of known variance, the maximum-likelihood fit minimises

$$\chi^2(p) = \sum_i r_i(p)^2,$$

$$r_i(p) = \frac{g_i(p) - y_i}{s_i}.$$

The weights put observations with different units or precision on a comparable scale. A small residual says that the chosen parameter fits the observations. To know whether it is the only plausible value, examine how the predictions change as the parameter moves.

Fit a spring and inspect the residuals

A spring in its elastic range has force $F = kx$. For equally precise force measurements at accurately known extensions, minimising the sum of squares gives $\sum_i x_i(kx_i - F_i) = 0$, hence $\hat{k} = \sum_i x_i F_i / \sum_i x_i^2$. At extensions of 1, 2 and 3 mm the measured forces are 1.1, 1.9 and 3.0 N. The fitted stiffness is

$$13.9/14 = 0.9929\,\mathrm{N/mm}.$$

Plotting the signed residuals against extension can reveal curvature or a preload that this model omits.

If an unknown preload b is allowed, fit $F = b + kx$. The columns of the design matrix are then 1 and x_i. Measurements at a single extension can determine $b + kx$ but cannot distinguish k from b, however many times the measurement is repeated. Multiple extensions add independent information. For a general linear model a QR or singular-value decomposition solves this least-squares problem without squaring the condition number by explicitly forming the normal equations.

Design a measurement for a nonlinear model

For a nonlinear response, approximate the residual near a trial point by $r(p + \Delta p) \simeq r(p) + J\Delta p$, with $J_{ij} = \partial r_i / \partial p_j$. A Gauss–Newton step minimises the norm of this linear expression. A damped step can also penalise the size of Δp when the local linear model gives a poor prediction. Small singular values of the appropriately scaled Jacobian identify parameter combinations that the data barely constrain.

Consider first-order decay $c(t, T) = c_0 e^{-k(T)t}$ with an unknown reference rate k_r and activation energy E. Write $k(T) = k_r \exp[-(E/R)(1/T - 1/T_r)]$. At $T = T_r$ the factor multiplying E vanishes. A complete decay curve at that temperature can estimate k_r, but carries no information about E. Taking measurements on either side of T_r changes that sensitivity. For a locally linear, full-rank, well-scaled problem with unit-variance normalised errors, $(J^{\mathsf{T}} J)^{-1}$ approximates parameter covariance. The diagonal estimates uncertainty under the assumed model and noise law.

Problem class

Choose a second spring measurement

A force sensor has an unknown constant offset. Compare measurements at 2 and 2 mm with measurements at 1 and 3 mm. For the latter design, suppose the measured forces are 1.3 N and 3.1 N. Recover the slope and offset. What remains unknown if each extension is sampled only once?

▼ Check rank before fitting the line

Two identical extensions give two identical rows $(1, 2)$ and cannot separate slope and offset. At distinct extensions the difference gives $k = (3.1 - 1.3)/(3 - 1) = 0.9$ N/mm, then

$$b = 1.3 - 0.9 = 0.4\,\mathrm{N}.$$

Two observations exactly fit two unknowns. No residual degrees of freedom remain to estimate measurement noise or check linearity from those data alone.

Know when a sample is informative

At one fixed temperature, $c = c_0 e^{-kt}$. Assume the same absolute concentration uncertainty at every sampling time. At what time is a single concentration measurement most sensitive to k? Evaluate the result for $k = 10^{-3}\ \mathrm{s}^{-1}$.

▼ Maximise the derivative with respect to the rate

The magnitude is $c_0 t e^{-kt}$. Its time derivative is $c_0 e^{-kt}(1 - kt)$, so it peaks at $t = 1/k = 1000$ s. The slope is small at very early times because little decay has occurred. At very late times very little material remains. Relative rather than absolute measurement errors would give a different sampling criterion.

ONESIMULATION LABORATORY

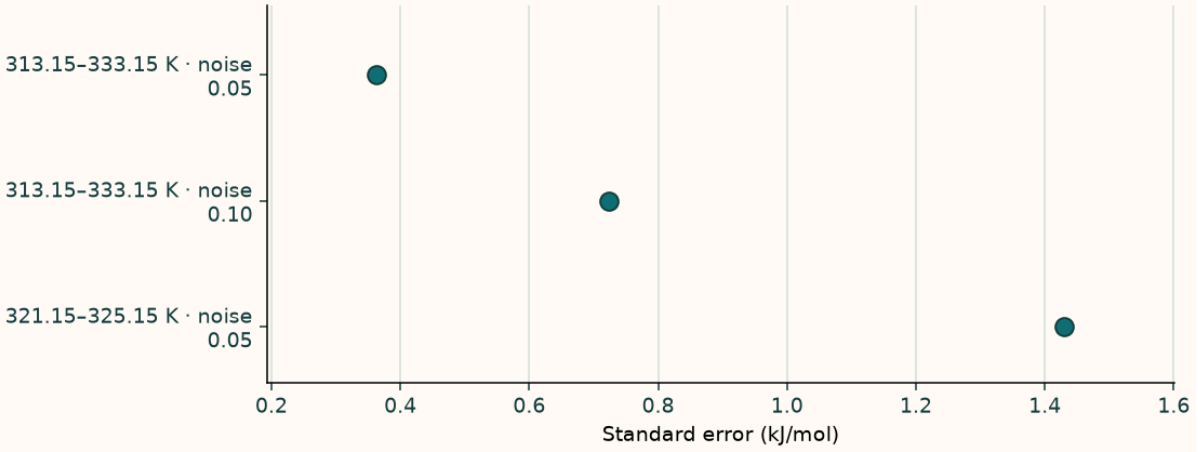
Find out what a temperature series identifies

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.

Fit k_r and E simultaneously to five simulated concentration histories with a fixed noise seed. The data-generating values are $k_r = 0.001$ s⁻¹ and $E = 115$ kJ/mol. Keep the same starting guess. Predict which change most increases the standard error of E: doubling `noise`, or moving the five `temperatures` closer to their common reference of 323.15 K.

OneSimulation least-squares fits. The last column is the local standard error of the fitted activation energy.

Temperature range and noise (mol/m³)	Fitted E (kJ/mol)	Standard error (kJ/mol)
313.15–333.15 K · noise 0.05	114.0367	0.3636
313.15–333.15 K · noise 0.10	113.0635	0.7228
321.15–325.15 K · noise 0.05	111.1489	1.4300



Narrowing the temperature range raises the fitted energy's standard error from 0.364 to 1.430 kJ/mol. With these runs that costs more precision than doubling the sensor noise.

In the wide case the five temperatures are spaced by 5 K. In the narrow case they are spaced by 1 K. Use the same 12 sampling times per experiment. Try a different random seed after checking the table. Compare the fitted error bar with the scatter of estimates across several seeds. Increasing the number of repeats probes noise; changing the temperature range also changes parameter identifiability. Inspect the model, residuals and fitting procedure.

Download experiment files (ZIP)

Extract the files together and open `arrhenius_parameter_estimation.ipynb` in a local OneSimulation environment.

Change parameters · Change the laboratory inputs

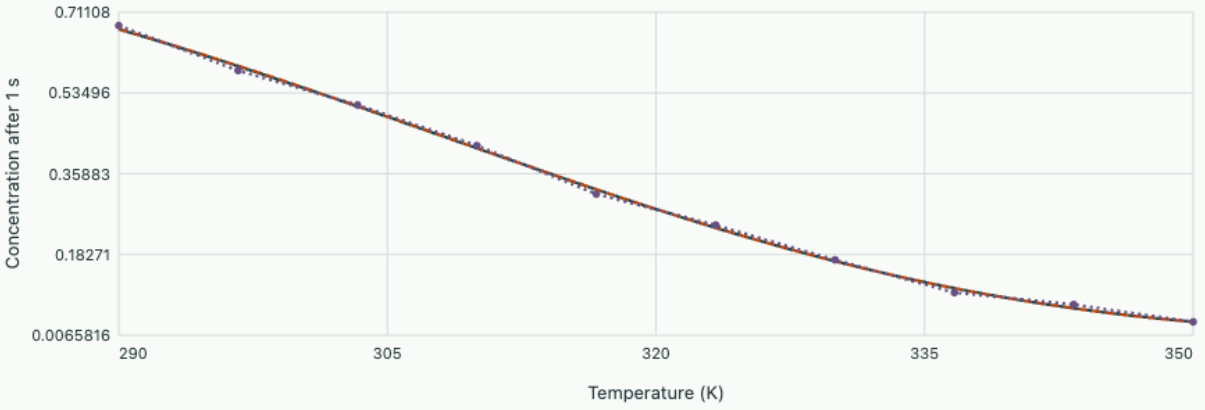
Change one input at a time and compare the recalculated curves with the argument above. The browser evaluates this small model locally; the accompanying numerical values state its assumptions and reference checks. Reset restores the reference inputs, and the data can be downloaded as CSV.

Open this calculation with adjustable inputs

Try the calculation

Temperature span around 320 K **60**

Concentration noise SD **0.01**



Fitted activation energy 30.041 kJ/mol, local standard error 0.59202 kJ/mol; true energy 30 kJ/mol. Prefactor and energy are fitted jointly. The fixed sinusoidal perturbations isolate noise level and temperature span.

Problem sheet 3 · Choose between repeating and spreading measurements

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.

Adjoint sensitivity

$$J^{\mathsf{T}} r$$

This operation maps a data-space vector back to parameter space. For a discretised physical model the chosen inner products determine the correct adjoint.

Derivation · Adjoint gradients through a forward solver

Forward operator

$$y = G(m) + \eta$$

The operator maps parameters to predicted observations. The noise and discrepancy model must use the same physical units as the readout.

Derivation · Forward maps, information and identifiability

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.

Derivation · Predicting a new measurement when the fitted parameters are correlated

Resolution matrix

$$R = G_{\text{reg}}^{\dagger} G$$

For a linear regularised inverse, a row describes how the estimated parameter mixes the actual components. A sharp-looking result may still have broad resolution kernels.

Derivation · A reconstruction audit from observations to a defensible claim

Tikhonov estimate

$$\arg\min_m \{\|Gm - y\|^2 + \alpha \|Lm\|^2\}$$

The regularisation operator selects which departures are costly. Compare values of alpha only after fixing how both terms are scaled.

Derivation · Regularisation, resolution and the price of stability

Whitened residual

$$r = \Sigma^{-1/2}(y - G(m))$$

For a known positive covariance this expresses mismatch in noise units and removes the modelled correlations between measurements.

Derivation · Weighted least squares and residual diagnostics

Further reading

- Work with the companion test problems and regularisation tools. Inspect the data coefficients and singular values before selecting a truncation or Tikhonov parameter. Per Christian Hansen · Discrete Inverse Problems

Compare specific topics with the university sources