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.

Contents

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

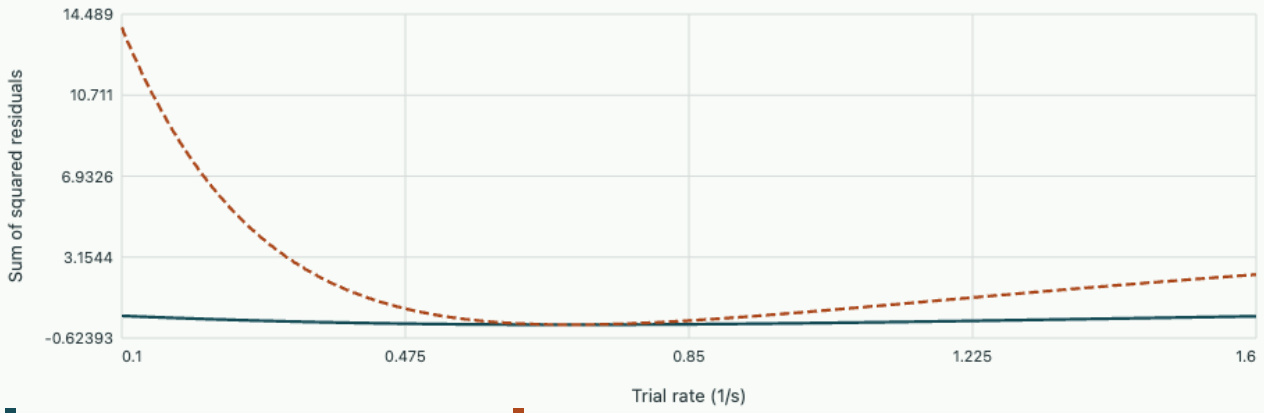
The calculation uses the four times above and adds a controlled perturbation to one late observation. It solves the two linear nuisance parameters afresh at each trial rate; a second curve evaluates the same rate with amplitude and background fixed to their original noiseless values. Comparing them makes 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.

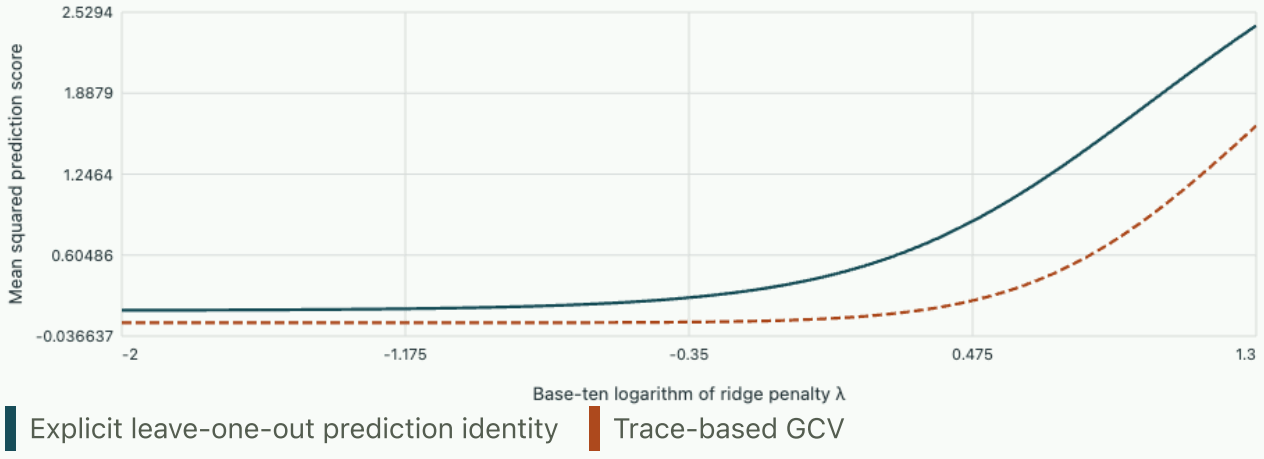
The OneSimulation model refits five separate reduced data sets for every penalty. The first CSV score therefore measures actual omitted-sensor predictions. Compare it point by point with the faster identity used in the browser.

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?

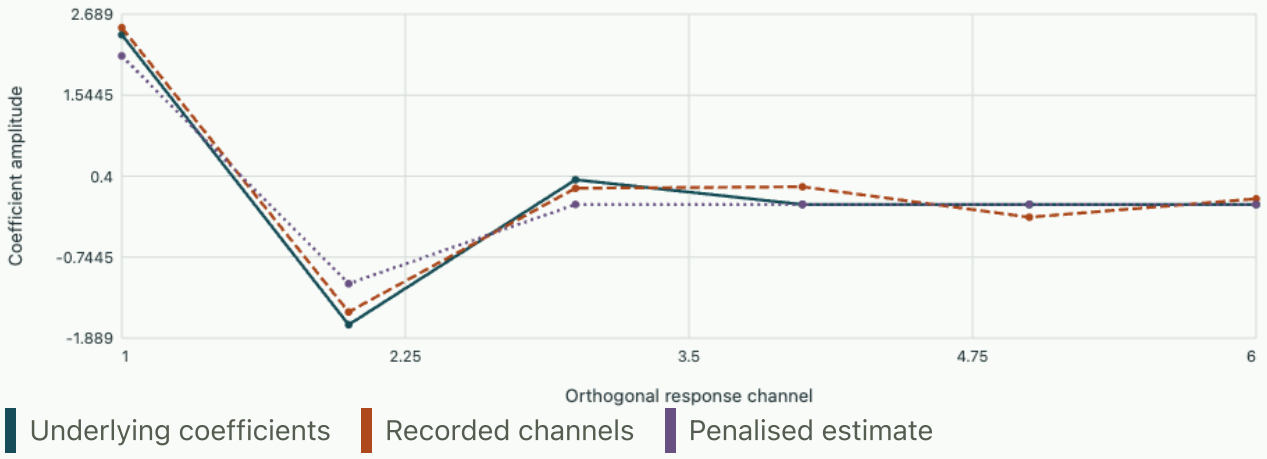
The six orthogonal channels have three true nonzero coefficients. Set the perturbation multiplier to one and start with zero threshold. Increase the threshold just far enough to remove the three noise-only channels. Does the weak third signal survive? Now set the perturbation to zero while keeping that 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

$$m^+ = (2.4, 0.6),$$
$$P^+ = \begin{pmatrix} 0.8 & 0.2 \\ 0.2 & 0.8 \end{pmatrix}.$$

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 $P^+ = P - Kh^TP$. 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

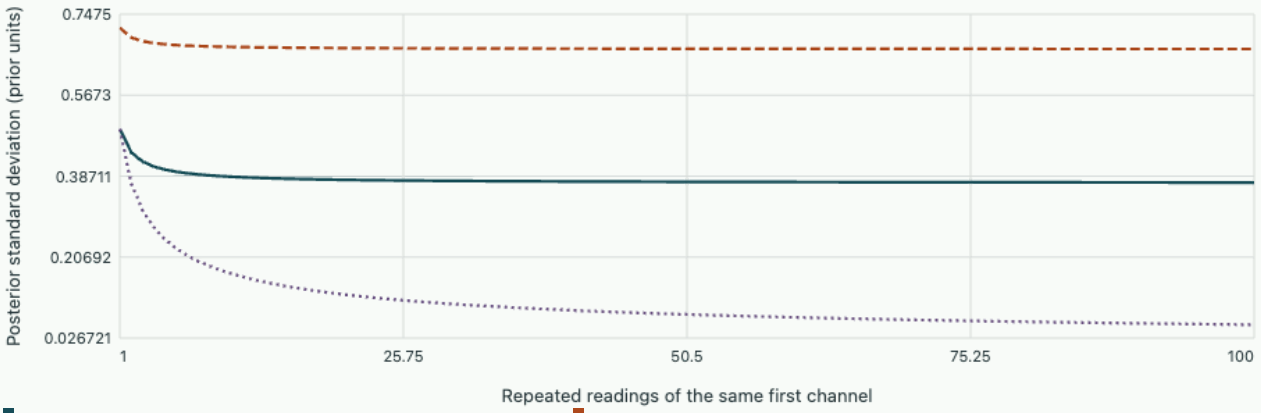
Two parameters have unit prior variance and adjustable correlation. Observe the first parameter repeatedly. Each reading has independent Gaussian noise of standard deviation 0.4 plus the same Gaussian calibration offset, drawn independently of the parameters and the individual noise. First set the calibration 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 $\begin{pmatrix} 0.04 & 0.16 \\ 0.16 & 0.64 \end{pmatrix}$ from the preceding covariance, giving $\begin{pmatrix} 0.76 & 0.04 \\ 0.04 & 0.16 \end{pmatrix}$. 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.

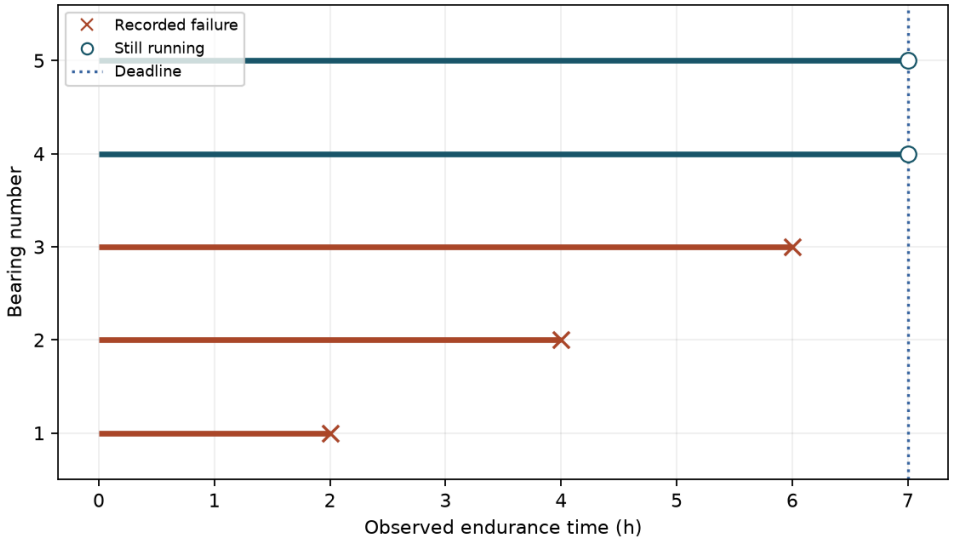
Audit a batch stopped after seven hours

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.

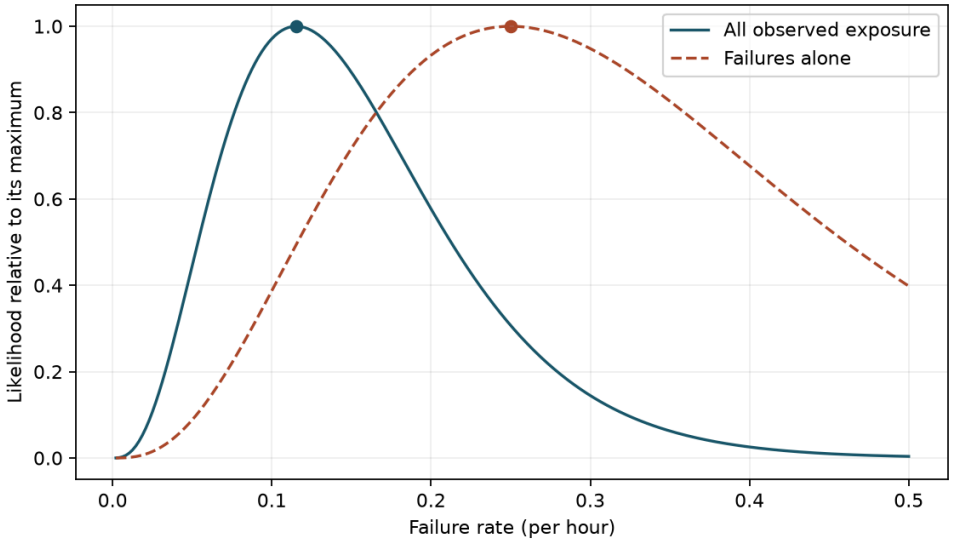
For a candidate rate of 0.20 per hour the log likelihood relative to the full-data maximum is

$$3\log(0.20/(3/26)) - 26(0.20 - 3/26) \approx -0.550.$$

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.

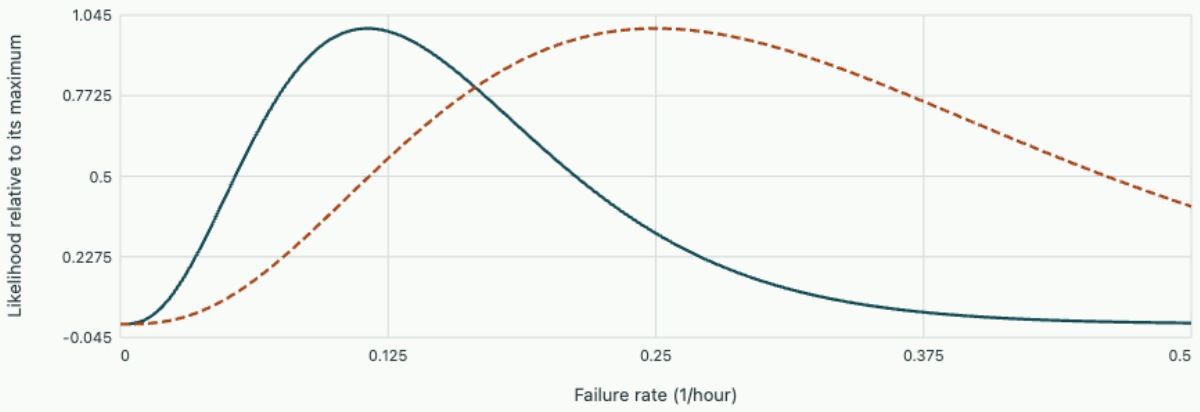
Change parameters · Explore this notebook model

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



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.

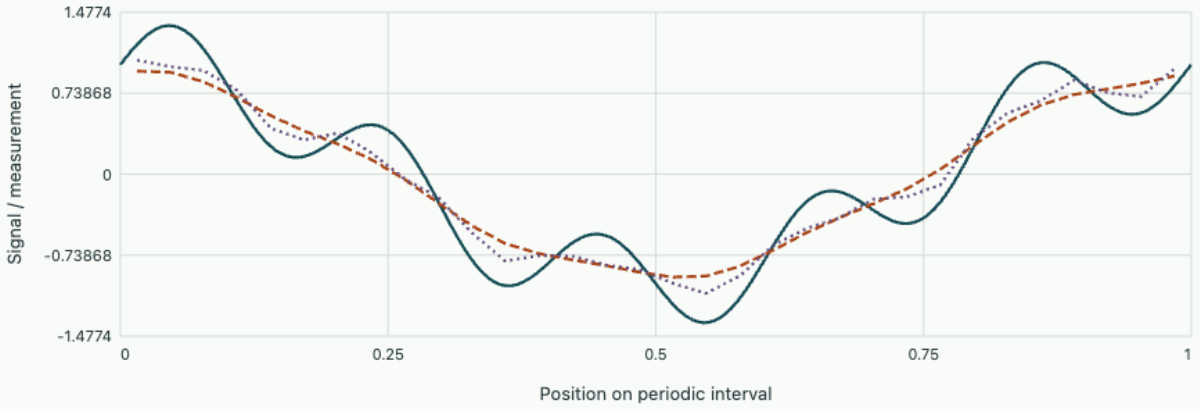
Download the executable notebook · Python source

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]

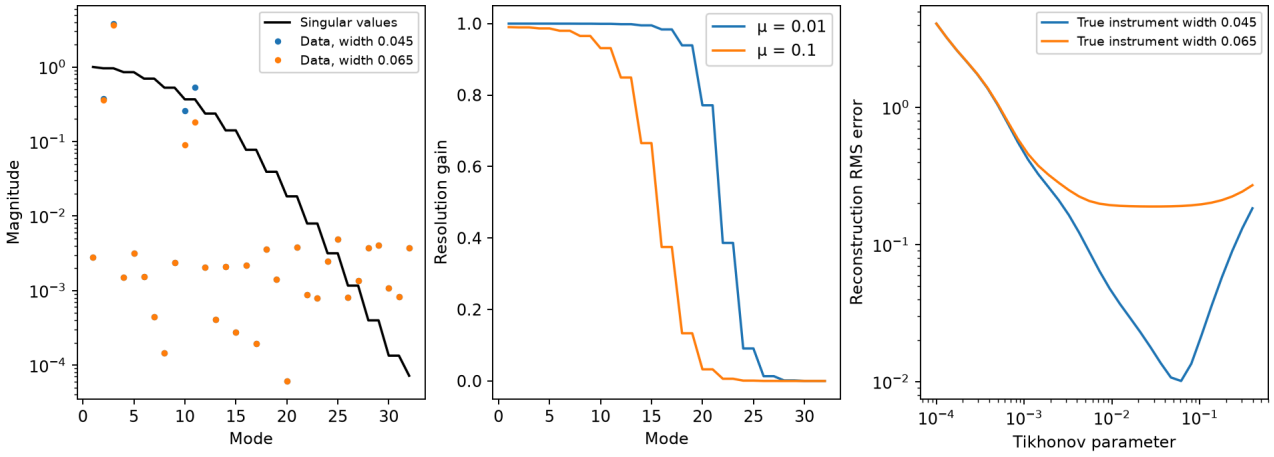
def signal(x):
    return np.cos(2*np.pi*x) + 0.4*np.sin(10*np.pi*x)

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 $M = A^TA + L^TL$ is positive definite. Factor $M = R^TR$. 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.

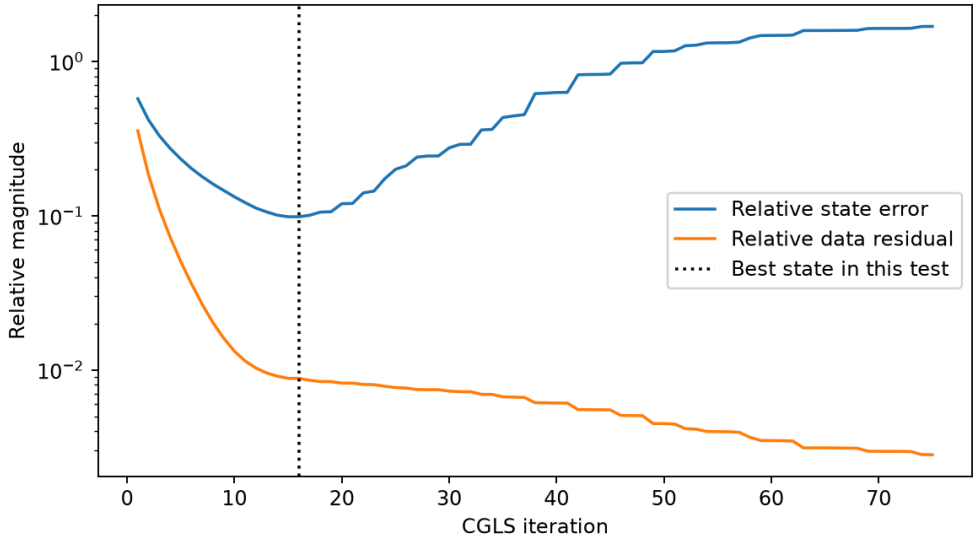
For $c_i > 0$ put $u_i = Ax_i/c_i$. With $x = \sum \xi_i x_i$, the component objective is $(c_i\xi_i - u_i^Tb)^2 + \mu^2 s_i^2\xi_i^2$. Minimisation gives

$$\xi_i = \frac{c_i}{c_i^2 + \mu^2 s_i^2}\, u_i^T b,$$

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

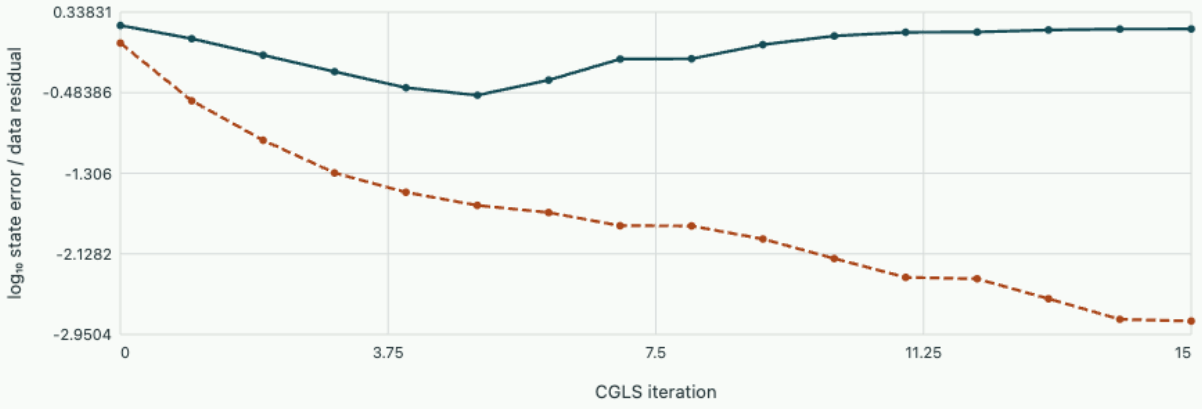
Change parameters · Early stopping and penalties that leave part of a field free

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.