Matrix computations and spectral accuracy

Trace how algebraic structure, representation and numerical precision affect the solution of linear problems. Verify each conclusion with an independent diagnostic.

Contents

1. **When stable eigenvalues allow large transient growth**
   Two damped states exchange energy in only one direction. Its eigenvalues predict ultimate decay, while a short-time response can grow by a large factor. Compute the matrix exponential, its gain and a resolvent-based sensitivity check.
2. **How a random sketch finds a dominant subspace**
   An instrument returns a matrix too large to factor densely but cheap enough to multiply by a few vectors. Build a low-rank approximation by probing its range, explain what a random draw can miss and design a residual check using fresh probes.
3. **Solve a singular network without inventing a ground current**
   A floating resistor network has unique voltage drops but no absolute voltage reference. Derive its nullspace, test charge balance, and distinguish adding a computational gauge from changing the physical boundary conditions.
4. **Refitting a sensor as observations arrive or disappear**
   A calibration line receives one new observation at a time. Update its least-squares factorisation using rotations, quantify the influence of a deleted observation and detect when removing a sample destroys identifiability.
5. **Which resistor can carry its current through another path?**
   A short circuit diagram leads to a matrix that describes both voltage differences and sensitivity to losing an edge. Change one conductance and follow where the current goes.
6. **How a computer can lose a sensor direction**
   Two sensors report almost the same quantity. Trace a perturbation through orthogonalisation, a Gram matrix and a rectangular SVD, then decide which computed digits the data support.
7. **Recovering a spectrum from three-term recurrences**

A large vibrating system supplies products with its stiffness operator. Derive the short recurrence implied by symmetry and use a residual and a moment test to interpret its small projected spectrum.

Matrix computations and spectral accuracy · 1. When stable eigenvalues allow large transient growth

Two damped states exchange energy in only one direction. Its eigenvalues predict ultimate decay, while a short-time response can grow by a large factor. Compute the matrix exponential, its gain and a resolvent-based sensitivity check.

Solve the evolution one component at a time

Consider dimensionless time t and state $x = (x_1, x_2)^T$ with $\dot{x} = Ax$, $A = \left(\begin{smallmatrix} -1 & K \\ 0 & -2 \end{smallmatrix}\right)$, for real $K \geq 0$. The second component decays independently; its transient drives the first. A physical example is a linearised pair of relaxation processes with one readout feeding another. Choose state units so the Euclidean norm represents the amplification of interest. If the coordinates measure unrelated quantities such as volts and kelvins, the unweighted norm changes if one simply changes units. A physical energy or covariance supplies a more appropriate weighted norm.

Starting from $x(0) = (u, v)$, integrate the second row to find

$$x_2 = e^{-2t} v.$$

The first row obeys $(e^t x_1)' = K e^{-t} v$, hence

$$e^{tA} = \begin{pmatrix} e^{-t} & K(e^{-t} - e^{-2t}) \\ 0 & e^{-2t} \end{pmatrix}.$$

Both eigenvalues are negative and the whole matrix exponential tends to zero. Yet for $v = 1, u = 0$ the driven component reaches $K/4$ at $t = \ln 2$, since $e^{-t} - e^{-2t}$ is maximal there. With $K = 12$, the state at this instant is $(3, 0.25)$ and its norm is 3.010. This is a lower bound on the maximum possible gain over unit input vectors and on the maximum over all times. The peak of the driven component need not coincide exactly with the peak of the total norm.

Find the input that grows most at a chosen time

For $M = e^{tA}$, the squared output norm is $x_0^T M^T M x_0$. Its maximum over unit vectors is the largest eigenvalue of $M^T M$, hence the gain is the largest singular value of M. For the triangular matrix above denote its entries by a, b, d on the first diagonal, off-diagonal and second diagonal. The characteristic polynomial of $M^T M$ has trace $s = a^2 + b^2 + d^2$ and determinant $a^2 d^2$. Its larger eigenvalue is $[s + \sqrt{s^2 - 4a^2 d^2}]/2$. This direct formula lets a simulation compare a specified initial condition with the optimal combination at each time. The optimal right singular vector varies with time; a plot of the optimal curve is an envelope of different preparations.

Another test asks whether any instantaneous squared norm can increase at $t = 0$. Its derivative is $x_0^T (A^T + A) x_0$. The largest eigenvalue of the symmetric part $H = (A + A^T)/2$ is $(-3 + \sqrt{1 + K^2})/2$. It is positive precisely when $K > \sqrt{8}$. If H is negative semidefinite every instantaneous state has a nonpositive norm derivative and the Euclidean semigroup is contractive. This

criterion concerns energy growth in a specified norm and is distinct from asymptotic stability. At $K = 12$ the eigenvector of H giving fastest initial growth is also not generally the input giving largest long-time response.

Link a small singular value to sensitivity of eigenvalues

For z away from either eigenvalue, direct triangular inversion gives an off-diagonal resolvent entry $K/[(z+1)(z+2)]$. Even at complex points visibly far from the eigenvalues the resolvent norm can be large. To see what that means, let $\sigma_{\min}(zI - A) = \varepsilon$ with unit right and left singular vectors v, u. Then $(zI - A)v = \varepsilon u$. The rank-one perturbation $E = \varepsilon u v^*$ makes $(A + E)v = zv$, while its spectral norm is precisely ε. Conversely any perturbation moving an eigenvalue to z must have norm at least $\sigma_{\min}$. Therefore a contour of $1/\|(zI - A)^{-1}\|$ maps the perturbation required for an eigenvalue there. This is the pseudospectrum construction. Physical perturbations may have additional sparsity or sign constraints, so unstructured contours describe the allowed full matrix perturbations of the stated norm.

Compare one preparation with the worst-case envelope

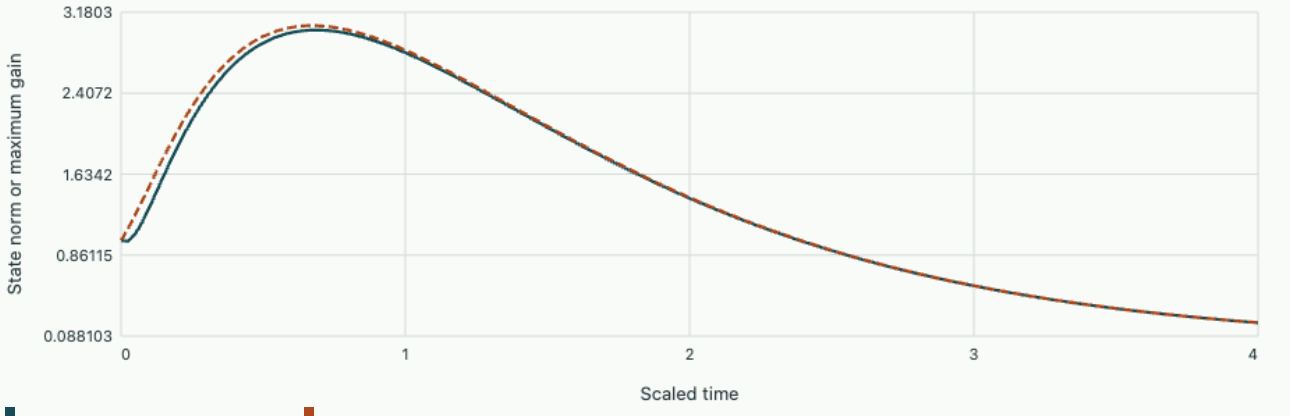
The experiment evaluates the exact triangular evolution and the largest singular value at each time. Change the coupling first, then rotate the unit initial input. At zero coupling both state components decay. At $K = 12$ with initial excitation only in the second coordinate the output norm exceeds three before decaying. Explain why the envelope can exceed every single displayed preparation at different times and why changing the input does not alter either eigenvalue.

Open this experiment online

Try the calculation

Coupling K **12**

Initial angle (degrees) **90**



At t = ln 2, selected norm 3.010, optimal gain 3.051. Both eigenvalues stay −1 and −2.

How a detector can miss the strongest transient

The apparatus often measures a scalar trace $y = c^T x$. With preparation $x(0) = (0, 1)^T$ and a detector aligned with the first coordinate, $y = K(e^{-t} - e^{-2t})$. A detector aligned with the second coordinate instead records e^{-2t}, whose graph decreases immediately. Rotating the detector leaves the evolution untouched. At $t = \ln 2$ and $K = 12$, two orthogonal recordings therefore give 3 and 0.25. Their sum of squares recovers 9.0625, so their joint norm reproduces the transient concealed in the second recording alone.

Given an input direction b and a detector c, the measurable input-to-output amplitude is $c^T e^{tA} b$. The optimal detector for that input is parallel to $e^{tA} b$ by Cauchy–Schwarz. Maximising also over the preparation recovers the singular-value experiment above. Thus a discrepancy between a large theoretical bound and a small experimental trace can arise from either the excitation or the readout. An experiment that rotates only one of them cannot by itself test the unrestricted bound.

Tutorial · verify three distinct amplification diagnostics

Find the exact growth at one time

With $K = 8$, compare the driven component for input $(0, 1)$ and the maximum Euclidean gain at $t = \ln 2$.

▼ **Work through the solution**

The matrix entries are

$$a = 0.5, b = 2, d = 0.25.$$

The specified input gives norm $\sqrt{4.0625} = 2.016$; the driven component itself is 2. The optimum follows from $s = 4.3125$ and $a^2 d^2 = 0.015625$, giving gain about 2.076. Selecting the unit right singular vector of the larger eigenvalue attains it. A direct check is to rotate a unit input around the circle and find the maximum by sampling increasingly fine angles.

Construct a structured perturbation placing an eigenvalue at zero

Keep both diagonal entries fixed and alter only A_{21} by δ. Determine the value needed for a zero eigenvalue and compare it with changing only a diagonal entry.

▼ **Work through the solution**

The determinant becomes $2 - K\delta$. Thus $\delta = 2/K$, or $1/6$ for $K = 12$. Changing only the first diagonal entry would require a shift of one, and changing only the second a shift of two. The smaller lower-left change closes a positive feedback loop. It need not be the smallest unrestricted spectral-

norm perturbation, which is obtained from the singular value of $-A$, but it explicitly demonstrates which physical feedback is implicated.

Decide which conclusion survives nondimensionalisation

Replace the coordinates by $y = (x_1/10, x_2)$. How do the eigenvalues, coupling and interpretation of a Euclidean gain change?

▼ Work through the solution

The similar matrix DAD^{-1}, with $D = \mathrm{diag}(0.1, 1)$, has diagonal entries $-1, -2$ and off-diagonal $K/10$. The dynamics represent the same trajectories in different units; the eigenvalues agree. The bare Euclidean norm in y now downweights the original first component, so it can show much less gain. To recover the original norm use $y^T D^{-T} D^{-1} y$. A statement about growth must name the norm and scales, especially when assessing whether a stable numerical model can amplify perturbations beyond a physical tolerance.

Further reading

- Oxford · Trefethen, Pseudospectra of linear operators
- Course investigation · Arnoldi and eigenvalue residuals

Matrix computations and spectral accuracy · 2. How a random sketch finds a dominant subspace

An instrument returns a matrix too large to factor densely but cheap enough to multiply by a few vectors. Build a low-rank approximation by probing its range, explain what a random draw can miss and design a residual check using fresh probes.

Start with the approximation that would be best if its subspace were known

Let A be an $m \times n$ real data matrix. It can represent a collection of spatial snapshots, one per column. A reduced rank-k model stores k spatial patterns and a small coefficient matrix. If a matrix Q has orthonormal columns spanning a proposed range, then the closest approximation with columns in that range in Frobenius norm is QQ^TA: each snapshot is independently orthogonally projected. Store $B = Q^TA$; its singular-value decomposition is small when Q has relatively few columns. The distinction between finding the subspace and factoring within it is important. A perfect small SVD cannot recover a spatial direction missing from Q.

If $A = U\Sigma V^T$ is exactly rank k, choose $n \times k$ test matrix Ω and form $Y = A\Omega = U_k\Sigma_k V_k^T\Omega$. If the last $k \times k$ factor is invertible, Y spans exactly the range of A. Independent Gaussian columns make singularity a zero-probability event in exact arithmetic, because the rotated coefficients are again a continuous Gaussian matrix. This statement does not imply every draw has a well-conditioned factor, nor that a single probe spans every near-rank-k matrix. Orthogonalising Y with a numerically reliable QR yields the candidate basis Q.

Derive the effect of gap and starting alignment

Consider rank reduction of $A = \mathrm{diag}(10, 1)$ with one unit probe $\omega = (\cos\theta, \sin\theta)$. Its image is $(10\cos\theta, \sin\theta)$, so the angle ϕ between the sampled direction and dominant left singular vector obeys $\tan\phi = \tan\theta/10$ whenever $\cos\theta \neq 0$. A typical mixed direction is pulled toward the dominant range. A nearly perpendicular probe can still give a poor direction even with a substantial singular gap. For $\theta = 45^\circ$, $\tan\phi = 0.1$. The residual $(I - qq^T)A$ has only one nonzero singular value because its left range is one-dimensional. With $q = (\cos\phi, \sin\phi)$, its norm squared is $100\sin^2\phi + \cos^2\phi$, giving $\sqrt{200/101} = 1.407$. The optimal rank-one error is the discarded singular value 1.

One power step forms $Y = (AA^T)A\Omega$, raising singular values to the third power in the sampled direction while the approximation is still formed for the original A. The angle ratio shrinks to $\tan\theta/1000$; the residual norm is then about 1.00005 for the same 45° probe. Additional steps require more passes over data. In finite precision reorthogonalise between multiplications so weak required directions are not erased by a large gap. If the target includes both leading and moderate directions, blindly powering a block without intermediate QR can make the very reduced subspace one wants numerically rank deficient.

Independently test an approximation after building it

Add p extra columns to Ω to reduce the chance of poorly exciting a dominant combination. An oversampled Q has $k + p$ columns; project to $B = Q^T A$, take its dominant k small left singular vectors and map them back through Q. For an exactly rank-k input the extra columns need appropriate numerical rank handling. In the nearly low-rank case they provide room to represent part of the tail and improve the final cut. Cost depends on access: computing $A^T Q$ or performing every additional power pass can dominate on streamed or distributed data, even though the final dense SVD is tiny.

Freeze the fitted approximation $\tilde{A}$ and consider an independent random sign vector g with entries equally likely ± 1. Write $R = A - \tilde{A}$. Since $\mathbb{E}[gg^T] = I$,

$$\mathbb{E}\|Rg\|^2 = \operatorname{tr}(R^T R) = \|R\|_F^2.$$

Multiple independent checks estimate residual energy and its sampling spread without explicitly assembling R. This estimates Frobenius norm squared; the maximum directional error is a different objective. A single small response does not certify a small spectral norm or guarantee missed directions are unimportant to a particular downstream measurement. Test the physically relevant load or measurement directions as well.

Background · Why must the test vector be independent of the sketch?

Build the sampled range Q by orthogonalising $A\Omega$. With exact arithmetic every sampled column already lies in the range of Q. Applying the residual to those same vectors gives

$$(I - QQ^T)A\Omega = 0.$$

For example, one column sampled from the two-dimensional identity spans only a line. It reconstructs its own training vector exactly while the unit vector perpendicular to it has error one. A fresh isotropic probe has, on average, nonzero energy in every missing direction. For a standard Gaussian vector g independent of Q, the mean squared diagnostic is $\mathbb{E}\|Rg\|^2 = \|R\|_F^2$, with $R = (I - QQ^T)A$. This identity follows by expanding $g^T R^T R g$ and using $\mathbb{E}[gg^T] = I$. One draw is still a noisy diagnostic, especially when only a single direction is missing.

Which starting direction defeats a random range probe?

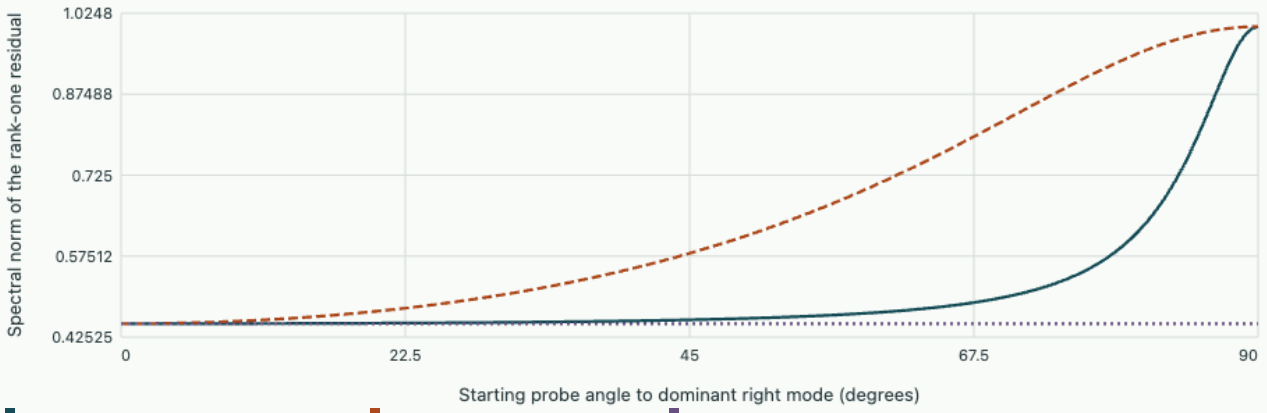
For a diagonal matrix with singular values 1 and r, rotate a test vector through the dominant right singular direction. Each power iteration multiplies the tangent of its output angle by r². Set r to 0.45, compare zero and two iterations, then inspect the exactly orthogonal probe at 90°. Explain why its residual stays at one.

Open this experiment online

Try the calculation

Singular-value ratio σ_2/σ_1 **0.45**

Power iterations **1**



Selected iteration count | Single range probe | Best possible rank-one residual

At 60°, residual 0.47105 against the best possible 0.45. A probe orthogonal to the leading mode gives residual one at every iteration count.

Tutorial · check an apparently good random approximation

Identify a failure when the test vector is reused

A matrix A is approximated by the orthogonal projection onto the range of $A\Omega$. You test residuals only on the original columns of Ω and find zeros. What does this certify?

▼ Work through the solution

By construction $QQ^TA\Omega = A\Omega$, up to roundoff if the full sampled range was kept. Thus the test is nearly tautological and does not measure error on independent directions. If Ω has a missing dominant right singular direction, all these residuals can vanish while the leading reconstruction is poor. Use fresh random probes and targeted physical snapshots held out during fitting. If a final rank truncation is applied within Q, the original probe residual need no longer vanish exactly but remains dependent on the fitted subspace.

Calculate a deliberately poor draw

Use $A = \mathrm{diag}(10, 1)$ and $\tan\theta = 100$. Compare the one-pass angle and residual with one power iteration.

▼ Work through the solution

Initially $\tan\phi = 10$, so $\sin^2\phi = 100/101$ and the residual norm squared is $(10000+1)/101$, giving about 9.95. With one power step $\tan\phi = 0.10$, giving 1.407 instead. The bad starting direction had a small but nonzero overlap, so powering can recover it. An exactly perpendicular deterministic probe remains perpendicular after every power step. Oversampling adds independent opportunities to see a missed direction.

Design a stopping criterion for snapshot energies

Independent residual probes give squared norms $1.0, 1.4, 0.8, 1.6, 1.2$ in scaled energy units, while independently known $\|A\|_F^2 = 100$. Estimate the residual energy fraction and its sampling uncertainty. Does it imply a 1.2% relative error on every snapshot?

▼ **Work through the solution**

The sample mean is 1.2 and sample standard deviation $\sqrt{0.10} = 0.316$, so the standard error is about 0.141 for independent probes. The estimated fraction of total squared Frobenius norm is 1.20%, with approximate standard error 0.14 percentage points. The relative Frobenius norm itself is about $\sqrt{0.012} = 0.110$. Individual snapshots may fare much better or worse; the tests estimate a global aggregate. Five samples give only a rough uncertainty estimate, particularly if residual energy is concentrated in a few directions.

Further reading

- Caltech · Halko, Martinsson and Tropp on randomised matrix decompositions
- Course investigation · singular values and low-rank error

Matrix computations and spectral accuracy · 3. Solve a singular network without inventing a ground current

A floating resistor network has unique voltage drops but no absolute voltage reference. Derive its nullspace, test charge balance, and distinguish adding a computational gauge from changing the physical boundary conditions.

Build the conductance matrix as an energy Hessian

Orient each of the m connections of a connected graph and form an incidence matrix B with one row per edge, containing $+1$ at its upstream node and -1 at its downstream node. For node voltages v, differences Bv determine currents $i_e = g_e(Bv)_e$ in edges of conductance $g_e > 0$. Conservation at nodes reads $B^T i = b$, where externally injected currents b are positive into the network. Eliminating edge currents yields $Lv = b,\ L = B^T G B$ with $G = \mathrm{diag}(g_e)$. Reversing an edge changes a row of B and its current sign together; L and physical voltage predictions are invariant.

The identity $v^T L v = \sum_e g_e (v_i - v_j)^2 \geq 0$ gives the nullspace directly. Equality forces equal voltage along each edge and hence constant voltage on a connected graph. Thus $\ker L = \mathrm{span}\{\mathbf{1}\}$ and its range is $\mathbf{1}^\perp$. A solution exists exactly when total injection is zero. On a graph with disconnected components the current sum must vanish separately on each component; one global cancellation can hide two incompatible islands. This is a compatibility condition on the physical data, distinct from the gauge freedom that persists for already compatible data.

Compare grounding, projection and a bordered solve

If b is compatible, choosing one node potential as zero removes one variable and one redundant conservation equation. The remaining principal submatrix is positive definite because a zero-energy vector extending with zero at the reference must vanish everywhere. An alternative chooses mean-zero voltage without favouring one node. Introduce a Lagrange multiplier λ and solve

$$\begin{pmatrix} L & \mathbf{1} \\ \mathbf{1}^T & 0 \end{pmatrix} \begin{pmatrix} v \\ \lambda \end{pmatrix} = \begin{pmatrix} b \\ 0 \end{pmatrix}.$$

The last equation fixes the gauge. Multiplying the first block by $\mathbf{1}^T$ gives

$$\lambda = (\sum b_i)/n.$$

For consistent data $\lambda = 0$, so the bordered solution satisfies the original equations exactly and is the Moore–Penrose minimum-norm solution. For inconsistent data the solver effectively removes a uniform current λ from every node. This may be sensible as a least-squares projection under an explicitly chosen equal-error model, but it is not the same physical network as a returning current concentrated at an actual external lead. Retain and interpret the multiplier.

Another formulation works directly in the zero-mean subspace with projector $P = I - \mathbf{1}\mathbf{1}^T/n$. Iterative methods can project the right side and reproject iterates to remove accumulated constant drift. A

preconditioner must preserve the intended subspace or be followed by the appropriate projection. The gauge-invariant observable $v_i - v_j$ is always preferable for comparing solvers with different references. Comparing raw voltages node by node can report a large error between mathematically equivalent physical solutions. A separate edge-current balance and dissipated power check validate more than one reference-dependent vector residual.

Verify a triangle by current conservation and work

Connect nodes 1–2 and 2–3 with 1 S each, and 1–3 with 0.5 S. Inject 1 A at node 1 and remove it at node 3. The direct branch has 2 Ω and the two-edge indirect branch also totals 2 Ω, so 0.5 A flows along each path. The terminal difference is 1 V. Taking $v_3 = 0$ gives $v = (1, 0.5, 0)$ V, while subtracting their mean yields $(0.5, 0, -0.5)$ V. Both produce Joule power $2(0.5^2/1) + (0.5^2/0.5) = 1$ W. The terminal injection work is $b^T v = 1$ W in either gauge because the sum of b vanishes.

A guard against concealed modelling error is to solve first with all conservation rows, reporting the incompatible component of b; only then introduce the intended reference. If measured injected currents are $1.00, 0, -0.97$ A and node 3 is grounded by deleting its equation, the computed branch currents imply that node 3 actually returns 1.00 A. The measured 0.03 A mismatch is thereby absorbed entirely by the chosen reference. The bordered projection instead distributes 0.01 A equally among the nodes. A calibrated return-lead model or a covariance-weighted correction can choose between such interpretations. A small residual after row deletion alone cannot tell whether the three measured currents were physically consistent.

What changes when a floating voltmeter is re-zeroed?

Connect three nodes by a triangle. The 1–3 and 2–3 conductances are each 1 S. Inject the selected current at node 1 and remove it at node 3. The graph sweeps a common reference-voltage shift. Use the oriented drop across branch 1–2 to predict its current, then compare the total injected power with the displayed dissipation.

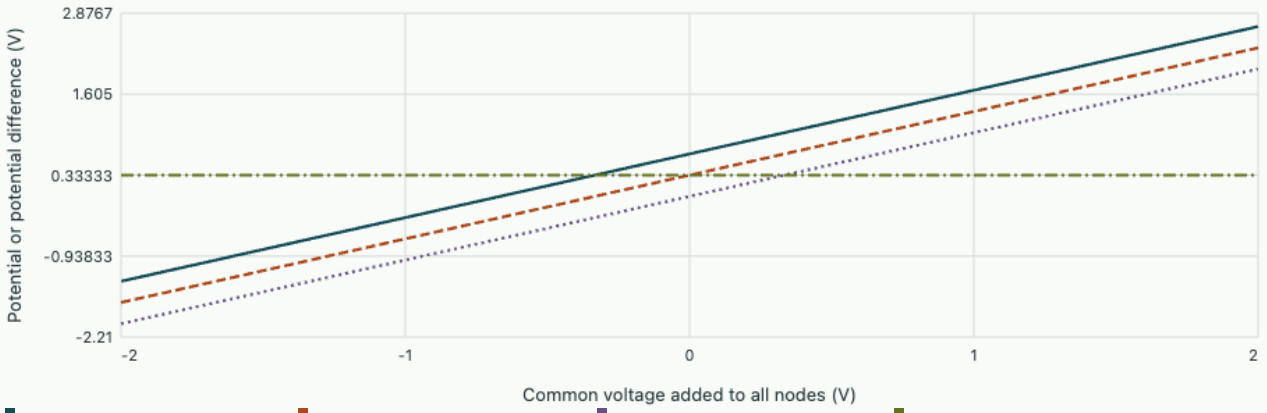
The downloadable OneSimulation circuit solves the three resistors and current source twice, choosing a different reference node. The two absolute voltage reports change. Their branch drops must agree.

Open this experiment online

Try the calculation

Conductance between nodes 1 and 2 (S) **1**

Current from node 1 to node 3 (A) **1**



Branch current I_{12} = 0.33333 A. Total Joule dissipation $I(V_1–V_3)$ = 0.66667 W at every gauge.

Download the OneSimulation model

In your OneSimulation environment, run the model with these inputs.

```
python networks-and-calibration.py network --conductance 1 --current 1
```

Tutorial · audit gauges, disconnected graphs and sensitivities

Find an effective resistance without choosing a reference

For connected L, point injection pair $b = I(e_a - e_b)$ and minimum-norm inverse L^+, derive the effective resistance and verify its sign.

▼ Work through the solution

The gauge-fixed solution is $v = L^+ b$, so terminal voltage divided by I is $R = (e_a - e_b)^T L^+ (e_a - e_b)$. Positive definiteness on the zero-sum subspace gives nonnegative R, strictly positive for distinct connected nodes. In the example it is 1 Ω, independently confirmed by the parallel branches. Summing injected power $I^2 R$ matches edge dissipation.

Determine how much one connection is worth

What is the first derivative of the triangle's effective resistance with respect to the direct conductance? Compare a prediction for raising it by 0.01 S with a direct circuit reduction.

▼ **Work through the solution**

For unit injection, differentiating $Lv = b$ on the fixed gauge subspace gives

$$dR = -v^T (dL) v.$$

A single conductance change has $dL = (e_i - e_j)(e_i - e_j)^T dg$, so

$$dR/dg = -(v_i - v_j)^2.$$

The direct drop is 1 V per ampere, giving -1 in compatible units. Prediction is $R \simeq 0.990\ \Omega$; exact parallel conductance becomes 1.01 S, so $R = 0.99010\ \Omega$. A positive conductance increment cannot increase effective resistance here.

Diagnose an island and an unobservable offset

Remove every link between two internally connected components carrying net injections $+0.10, -0.10$ A. Is one ground constraint enough? What if both component sums are corrected to zero?

▼ **Work through the solution**

The disconnected graph has two independent constant null vectors, one per component. Although the total current is zero, both component compatibility tests fail. No voltage gauge can make the missing inter-island return path appear. If both sums become zero, two independent gauge constraints still remain necessary for unique node voltages. Potential differences within either component are uniquely determined; differences across the two floating components remain arbitrary unless a physical reference or coupling is introduced.

Further reading

- ETH Zürich · numerical methods for CSE, sparse systems and positive definite forms
- Course investigation · constrained systems and Schur complements

Matrix computations and spectral accuracy · 4. Refitting a sensor as observations arrive or disappear

A calibration line receives one new observation at a time. Update its least-squares factorisation using rotations, quantify the influence of a deleted observation and detect when removing a sample destroys identifiability.

Rotate away a new observation while preserving the objective

Suppose A has full column rank p, observations b, and an orthogonal factorisation has already transformed the upper part of the problem into $Rx = z$ with upper-triangular R. Lower transformed rows carry residuals but no coefficients. Because the transformations are orthogonal,

$$\|Ax - b\|^2 = \|Rx - z\|^2 + s^2,$$

where s^2 is the accumulated minimum residual sum of squares. Appending row a^T and response y requires minimising $\|Rx - z\|^2 + (a^T x - y)^2 + s^2$. One can update the small augmented triangle instead of reloading every original sample. This is particularly useful when storing raw observations is expensive and fits need refreshing repeatedly.

Write the new row directly below $[R\ z]$. At column j, combine diagonal value r with bottom value q using a Givens rotation with $c = r/\operatorname{hypot}(r, q),\ s = q/\operatorname{hypot}(r, q)$. Multiplication of the two selected rows by $\left(\begin{smallmatrix} c & s \\ -s & c \end{smallmatrix}\right)$ gives a nonnegative diagonal $\sqrt{r^2 + q^2}$ and zero in the bottom column. Apply the same rotation to every remaining column and to the right side. If both inputs vanish choose the identity and investigate numerical rank separately. A robust hypotenuse evaluation avoids overflow when $r^2 + q^2$ itself exceeds the available range even though its square root is representable.

After sweeping all p columns the bottom row is zero in coefficients and has transformed response γ. The new triangular solve gives the updated fit and the minimum residual sum rises by γ^2. Each step preserves a squared norm, so this identity is a direct independent audit of online accumulation. Weighted least squares fits into the same calculation by scaling the row and response by the square root of its weight first. Dynamic reweighting of historical observations is a different operation and generally requires preserved information about individual rows.

Use one-dimensional corrections for interpretation

The rank-one update also has a useful algebraic form without explicitly computing a dense inverse. Set $G = A^T A$, $w = G^{-1} a$ evaluated by two triangular solves with R, and $h = a^T w$. After adding a point the normal matrix becomes $G + aa^T$. Subtracting the old normal equations from the new ones gives $(G + aa^T)\delta x = a(y - a^T x)$, so $\delta x = w(y - a^T x)/(1 + h)$, since $(G + aa^T)w = (1 + h)a$. The increment to residual sum of squares is $(y - a^T x)^2/(1 + h)$. This formula for diagnostics is consistent with the orthogonal update and shows explicitly how a surprising response interacts with geometric leverage.

As an example fit a line $y = c + dt$ at times $t = -1, 0, +1$ with readings $0, 1, 2$. The exact line is $c = d = 1$, with $G = \operatorname{diag}(3, 2)$. A new observation at $t = 2$ reports $y = 4$ rather than predicted 3.

Here $a = (1, 2)^T$, $w = (1/3, 1)^T$, $h = 7/3$. The corrected coefficients are $c = 1.1,\ d = 1.3$, and residual sum rises by $1/(1 + 7/3) = 0.30$. Substitution back into all four data gives residuals $0.2, -0.1, -0.4, 0.3$; their squared sum is indeed 0.30. The update has moved a previously exact historical fit because the new reading changed the estimated slope and intercept.

Distinguish a safe deletion from rank loss

For deleting a row already included in the fitted G, use that row's in-sample leverage $\ell = a^T G^{-1} a$. It lies between zero and one as a diagonal element of the orthogonal projection $A(A^T A)^{-1} A^T$. If $\ell < 1$, the remaining-data fit changes by $-G^{-1} a\, r/(1 - \ell)$, where $r = y - a^T x$ is the residual while the point was present. The leave-one-out prediction residual is $r/(1 - \ell)$. A point with small in-sample residual can therefore be highly influential when nearly essential to identification. If $\ell = 1$, deletion loses rank, and a formula dividing by $1 - \ell$ is warning of a structural change. Downdating a triangular factor should use a well-conditioned dedicated downdate or refactor with rank detection when close to this boundary. Squaring a poorly conditioned design into G just for convenience can lose accuracy.

Refit the line after one influential observation

Begin with three calibration pairs on y = x at x = –1, 0 and 1. Place a fourth sensor farther from the centre and choose its deviation from the old line. Predict its training residual and the error made when the entire fourth observation is omitted. Move it farther out with a fixed nonzero offset. Which residual approaches zero and which retains the full offset? Set the offset to zero and check whether a high-leverage point must also have a large residual.

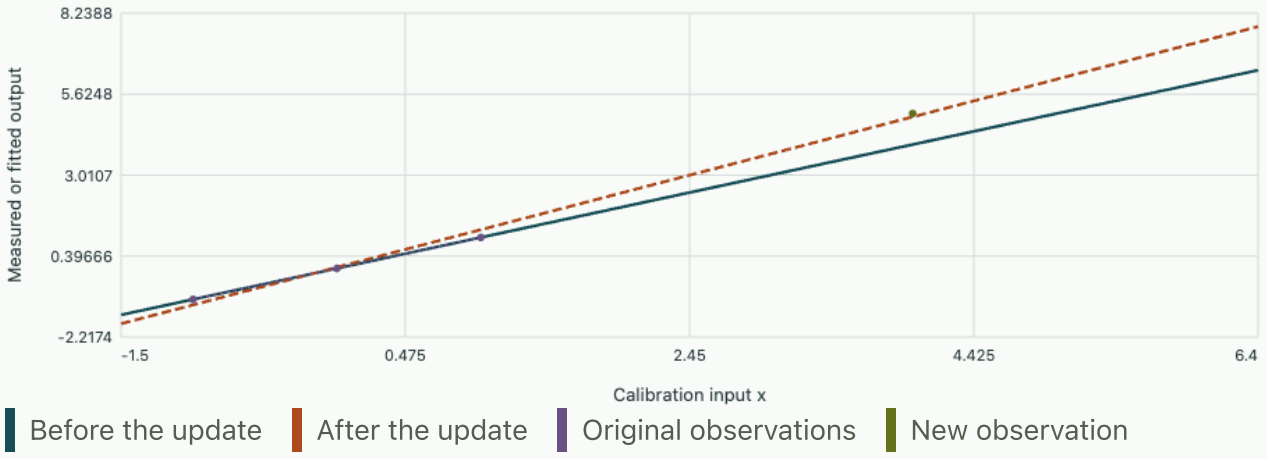
The OneSimulation model fits the full augmented observation matrix with its dense least-squares solver. Compare this independent refit with the one-row update on screen.

Open this experiment online

Try the calculation

New observation coordinate 4

New observation's offset from y = x 1



Updated slope 1.2143, intercept 0.035714. New row leverage 0.89286; deleted prediction residual 1 versus training residual 0.10714.

Download the OneSimulation model

In your OneSimulation environment, run the model with these inputs.

```
python networks-and-calibration.py line --coordinate 4 --offset 1
```

Tutorial · inspect an update from three angles

Perform a complete scalar QR update

Three readings of a constant have average 2 and residual sum 0.50. Add value 5 at equal weight. Find the updated average, residual increment and scalar triangle.

▼ **Work through the solution**

Before appending,

$$R = \sqrt{3},\ z = 2\sqrt{3}.$$

With appended row 1 the new diagonal is 2. The rotated right side is $(\sqrt{3}/2)(2\sqrt{3}) + (1/2)5 = 5.5$, giving average 2.75. The bottom response is $(-1/2)(2\sqrt{3}) + (\sqrt{3}/2)5 = 1.5\sqrt{3}$, contributing 6.75 to sum of squares and giving new total 7.25. The mean from the raw sum is also $(6 + 5)/4 = 2.75$.

Reverse the four-point line update

For the newly appended point in the line example, infer its in-sample leverage from the original append leverage, then recover the original leave-one-out residual.

▼ **Work through the solution**

Adding its own row changes leverage from h to $\ell = h/(1+h) = 0.70$, obtained by solving $(G + aa^T)^{-1}a = w/(1+h)$. Its current fitted value is $1.1 + 2(1.3) = 3.7$, giving residual 0.3. Dividing by $1 - \ell = 0.3$ recovers the predicted deletion residual 1.0. Confusing h with ℓ would incorrectly put a leverage larger than one inside a leave-one-out formula.

Detect a geometry where a zero residual is inevitable

Fit a line with many observations at $t = 0$ and exactly one at $t = 1$. Show why the latter's in-sample residual can be zero for any value and why its fit quality cannot validate the slope.

▼ **Work through the solution**

All $t = 0$ samples determine the intercept by their mean. The unique nonzero time then fixes the slope to hit its own response exactly, so its residual is always zero and its in-sample leverage is one. Removing it leaves only one time coordinate and the slope unidentifiable. An extra independent observation at a distinct time, or a justified physical slope prior, is needed to test predictive slope accuracy. Increased repetitions at $t = 0$ reduce intercept noise without repairing this missing diagnostic.

Further reading

- ETH Zürich · numerical methods for CSE, least squares and orthogonal factorisation
- Course investigation · orthogonal reduction for least squares
- Separate algebraic solver error from mesh error

Matrix computations and spectral accuracy · 5. Which resistor can carry its current through another path?

A short circuit diagram leads to a matrix that describes both voltage differences and sensitivity to losing an edge. Change one conductance and follow where the current goes.

Ground a node and derive the electrical leverage

Nodes A, B and C form a triangular loop. Node D connects only to C. Every edge is an ideal linear resistor. Orient an edge from node i to node j and give it an incidence vector b_e with a +1 at i and a –1 at j. Its current under nodal voltages v is $g_e b_e^T v$, where conductance g_e is measured in siemens. Kirchhoff's current law sums all outgoing currents into the net injection f. It gives

$$Lv = f,$$

$$L = \sum_e g_e b_e b_e^T.$$

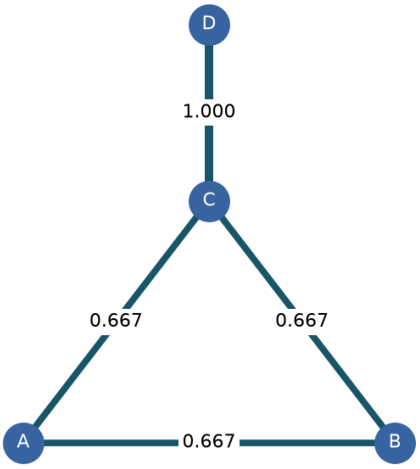
Adding a constant to every potential preserves all currents. Fix the potential of D at zero and remove its row and column. For a connected network the resulting grounded matrix is positive definite. Inject one ampere at one end of an edge and withdraw it at the other. With grounded coordinates the voltage difference per unit current is the effective resistance $R_e = b_e^T L_g^{-1} b_e$. The fraction of this test current that flows through the tested edge is $\ell_e = g_e R_e$. A bridge carries the full ampere. A parallel path reduces its share.

There is also a global check on the calculation. The sum of all the leverages is $\operatorname{tr}(L_g^{-1} \sum_e g_e b_e b_e^T) = n - 1$ on a connected graph with n nodes. Each leverage depends only on the graph and its conductances. The actual current in a specified operating condition also depends on the chosen source and sink.

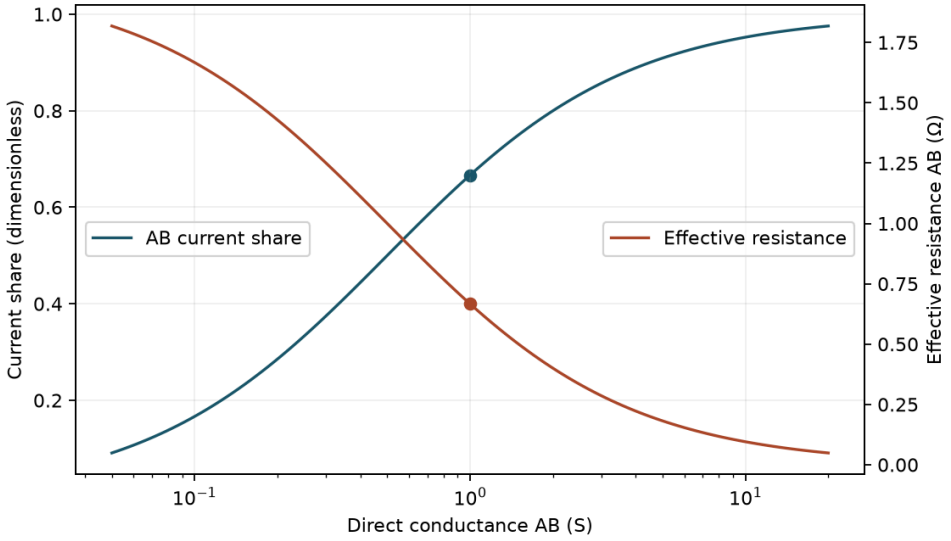
Solve the triangle by hand

With 1 S on every branch, the detour A–C–B consists of two 1 Ω resistors in series. Inject 1 A at A and extract it at B. The detour's 0.5 S combines with the direct 1 S, so the measured drop is $1/1.5 = 2/3$ V. The direct current is 2/3 A. The dead-end C–D branch carries zero current in this experiment. When tested between its own endpoints, however, C–D carries the entire test current. The four leverages are consequently $2/3, 2/3, 2/3, 1$ and sum to three.

Double A–B's conductance to 2 S. Its leverage rises to $2/2.5 = 0.8$, while the voltage needed for the same 1 A drops to 0.4 V. This is why the two traces in the experiment have opposite slopes. Rerun with a weaker bridge. Its resistance increases, yet its leverage stays equal to one.



Each label gives the share of a separate one-ampere test between that edge's endpoints. The pendant branch remains at unity for any positive conductance.



The current-share curve tends to one as the direct branch dominates the detour. The voltage per ampere falls from two ohms toward zero.

Exercise · remove the direct branch

How much does the terminal resistance rise when the 1 S direct edge is removed? What changes if the bridge is removed instead?

▼ Work through the solution

For A and B the sole remaining connection consists of two 1 Ω resistors, so resistance rises from 2/3 Ω to 2 Ω, an increase of 4/3 Ω. Removing C–D instead isolates D. Measurements within the triangle remain finite, but a test injecting current at D and withdrawing it at C no longer has a finite DC voltage solution.

Exercise · differentiate the measured voltage

Under a fixed one-ampere A-to-B test, find the sensitivity of the terminal resistance to the direct conductance at 1 S.

▼ **Work through the solution**

For variable g the resistance is $1/(g+0.5)$. Differentiation gives $-1/(g+0.5)^2 = -4/9$ Ω/S at the chosen point. The magnitude equals the square of the voltage drop per ampere across that branch. Differentiating $L_g v = f$ at fixed injection gives the same result for any branch, $\partial(f^T v)/\partial g_e = -(b_e^T v)^2$ for the unit test.

Change parameters · Explore this notebook model

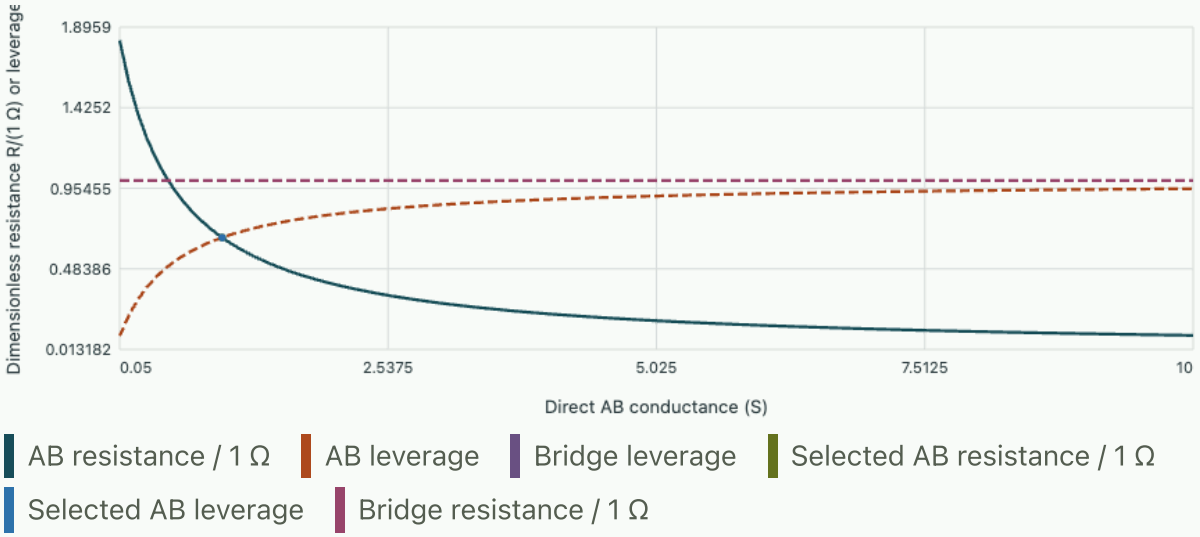
Change one input at a time and compare the recalculated curves with the argument above. The browser evaluates this small model locally; the accompanying numerical values state its assumptions and reference checks. Reset restores the reference inputs, and the data can be downloaded as CSV.

Open this calculation with adjustable inputs

Try the calculation

Direct AB conductance (S) 1

Bridge CD conductance (S) 1



Resistance is normalised by 1 Ω on this shared dimensionless axis. Selected AB resistance 0.66667 Ω, leverage 0.66667. Bridge resistance 1 Ω and bridge leverage 1; the four edge leverages sum to 3.

Open editable notebook

Download .ipynb

Matrix computations and spectral accuracy · 6. How a computer can lose a sensor direction

Two sensors report almost the same quantity. Trace a perturbation through orthogonalisation, a Gram matrix and a rectangular SVD, then decide which computed digits the data support.

Subtracting a projection in finite precision

Let A be a real m-by-n matrix, with $m \geq n$ and independent columns. The first unit vector is $q_1 = a_1/\|a_1\|_2$. To construct another, classical Gram–Schmidt computes all coefficients $r_{ij} = q_i^T a_j$ against the original column before subtracting the sum. Modified Gram–Schmidt instead updates a running residual v after each projection, computing the next coefficient from that residual.

$$v^{(0)} = a_j,$$

$$r_{ij} = q_i^T v^{(i-1)},$$

$$v^{(i)} = v^{(i-1)} - q_i r_{ij}.$$

In exact arithmetic the two formulas agree by orthogonality. In floating point, a component along an earlier q_i can survive subtraction. Dividing by the norm of a nearly cancelled column amplifies that component. For the three columns of the matrix with first row $(1, 1, 1)$ and lower block δI_3, the difference of two original columns has size proportional to δ. Computing both of the third column's coefficients before either subtraction carries the error of previously computed directions into both coefficients. Sequential projection removes some of that contamination. A second pass projects the surviving residual again. Monitor both $\|A - QR\|/\|A\|$ and $\|I - Q^T Q\|$. A small first diagnostic alone permits poor orthogonality.

Why a reflector gives a backward-error result

Given nonzero x, set $v = x + \operatorname{sign}(x_1)\|x\|e_1$, using positive sign at zero. Then $H = I - 2vv^T/(v^T v)$ is orthogonal and Hx has only its first entry nonzero. The sign ensures the construction of v_1 adds magnitudes. Assume normal-range floating-point arithmetic with unit roundoff u and $mu \ll 1$. A length-m dot product has absolute error bounded by $\gamma_m |v|^T |x|$, with $\gamma_m = mu/(1 - mu)$, by multiplying the factors $1 + \delta_i$, $|\delta_i| \leq u$, along the sum. Cauchy–Schwarz bounds this by $\gamma_m \|v\|\|x\|$.

Apply a stored, normalised reflector as a dot product and an axpy. Each column incurs a perturbation of order mu times its norm, including normalisation error. The exact reflector for the stored direction has norm one. Pull each local error backwards through the preceding exact orthogonal transformations and add. In Frobenius norm this yields

$$A + \Delta A = Q\widehat{R}, \qquad \|\Delta A\|_F \leq Cmnu\,\|A\|_F + O(u^2),$$

where C accommodates the implementation's fixed number of elementary operations. This explains the nature of the bound without claiming an implementation-independent sharp constant. A tiny singular value may still have large relative uncertainty under this absolute perturbation.

Turn a rectangular matrix into a bidiagonal one

Apply a left reflector to column 1 and a right reflector to row 1, starting its operation in column 2. Repeat on the trailing block. Each transformation preserves earlier zeros. This yields $A = UBV^T$ with orthogonal factors and upper bidiagonal B. As a concrete first step take rows $(3, 0), (4, 5), (0, 12)$. Multiplying the first two rows by the rotation with rows $(3/5, 4/5), (-4/5, 3/5)$ leaves rows $(5, 4), (0, 3), (0, 12)$. A second left rotation combines $3, 12$, giving diagonal entries $5, \sqrt{153}$ and superdiagonal 4. Two columns make the right step trivial here. A third column would require eliminating the new first row's tail.

For a bidiagonal matrix one can carry out shifted QR on $B^T B$ implicitly. If the first diagonal and off-diagonal entries of that tridiagonal matrix are a, b, select a right rotation from $(a - \mu, b)$. Applying it to B introduces a subdiagonal entry. A left rotation eliminates that entry and moves the disturbance to the next column pair. Repeat to the bottom and accumulate both sets of rotations. Vanishing superdiagonals split independent blocks. This obtains singular values without explicitly storing an already rounded Gram matrix. For two almost parallel sensor columns $(1, \delta, 0)^T, (1, -\delta, 0)^T$ the singular values are $\sqrt{2}, \sqrt{2}|\delta|$. The entries $1 \pm \delta^2$ of the Gram matrix can round to the same number.

OneSimulation experiment · Resolve the smaller singular value

Change the separation grid. Which lost singular value can a direct SVD still measure?

Download the executable notebook · Python source

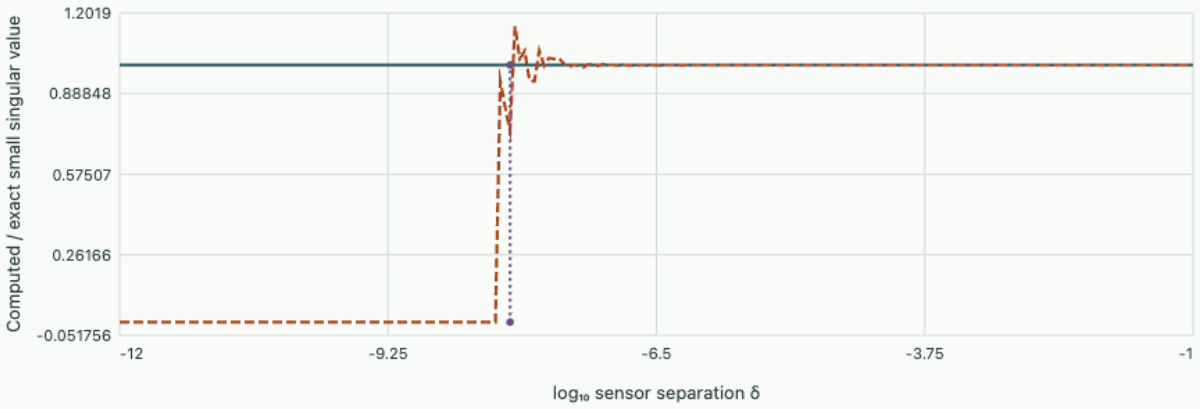
Adjust the experiment in the browser. Move the selected separation across the loss threshold and switch between binary32 and binary64 arithmetic. The Gram estimate evaluates the rounded difference $(1 + \delta^2) - (1 - \delta^2)$; the stable comparison uses the exact orthogonality of the two nonzero rows. It is an analytic benchmark, while the notebook below runs the rectangular SVD and OneSimulation eigenvalue solver. Locate the threshold near $\delta \sim \sqrt{u}$, then explain why changing arithmetic precision cannot recover information already lost in noisy measurements.

Open the live experiment

Try the calculation

Selected $\log_{10}\,\delta$ **-8**

Arithmetic bits **64**



Stable formula from orthogonal rows Rounded Gram-matrix difference

Selected separation

At δ = 1e-8, exact σmin = 1.4142e-8, condition number = 100000000. The 64-bit Gram calculation retains 0.74506 of the small singular value. The reference curve uses the analytic row-orthogonality formula, not an implemented SVD.

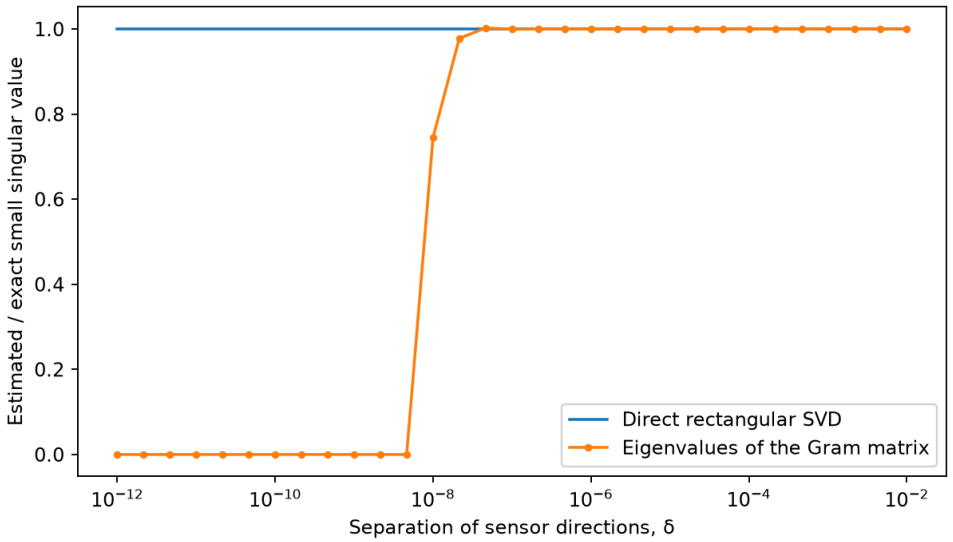
```
"""Measure loss of the weaker of two almost parallel sensor directions."""
import matplotlib.pyplot as plt
import numpy as np
from onesim.numerics.linear.matrices.dense_linear_algebra import symmetric_eigenvalues

separations = np.logspace(-12, -2, 31)
direct, normal = [], []
for delta in separations:
    matrix = np.array([[1.0, 1.0], [delta, -delta], [0.0, 0.0]])
    reference = np.sqrt(2.0) * delta
    direct.append(np.linalg.svd(matrix, compute_uv=False)[-1] / reference)
    smallest = symmetric_eigenvalues(matrix.T @ matrix)[0]
    normal.append(np.sqrt(max(smallest, 0.0)) / reference)

assert np.allclose(direct, 1.0)
assert normal[0] == 0.0
print(f"At delta={separations[0]:.0e}: direct/reference={direct[0]:.3f}, Gram/reference=
{normal[0]:.3f}")
fig, ax = plt.subplots(figsize=(7, 4), layout="constrained")
ax.semilogx(separations, direct, label="Direct rectangular SVD")
ax.semilogx(separations, normal, "o-", ms=3, label="Eigenvalues of the Gram matrix")
ax.set(xlabel="Separation of sensor directions, δ", ylabel="Estimated / exact small singular
value")
ax.legend()
fig.savefig("numerics-small-singular-value.png", dpi=170)
```

Output from the reference run

```
At delta=1e-12: direct/reference=1.000, Gram/reference=0.000
```



A Gram-matrix eigenvalue collapses as the separation of two sensor directions decreases.

Tutorial on orientation and detectability

Reconstruct the product for the bidiagonal example

▼ Work through the solution

The trace of B^TB is $25 + 16 + 153 = 194$, equal to the sum of squares of the six input entries. Its determinant is $25(153) = 3825$. Trace and determinant determine the two squared singular values independently.

Predict a failure threshold and compare two noise sources

▼ Work through the solution

When δ^2 is of order u, rounding an explicitly formed Gram matrix threatens its weaker eigenvalue. Perturbations in the original data of norm 10^{-5} already permit errors of that size in singular values by the variational characterisation. A stable algorithm cannot certify a sensor separation much smaller than instrument uncertainty.

Trefethen and Bau · QR, rounding error and numerical SVD

Matrix computations and spectral accuracy · 7. Recovering a spectrum from three-term recurrences

A large vibrating system supplies products with its stiffness operator. Derive the short recurrence implied by symmetry and use a residual and a moment test to interpret its small projected spectrum.

Let symmetry remove old directions

Assume $A = A^T$, choose $\|q_1\| = 1$, and initialise $\beta_0 = 0, q_0 = 0$. Arnoldi projects Aq_j against every available basis vector. For $i < j - 1$, Aq_i already lies in the span through q_{i+1}, so $q_i^T A q_j = (Aq_i)^T q_j = 0$. For $i = j - 1$ the product is the previously computed β_{j-1}. Thus the only new diagonal coefficient is $\alpha_j = q_j^T A q_j$.

$$w = Aq_j - \beta_{j-1} q_{j-1} - \alpha_j q_j,$$

$$\beta_j = \|w\|,$$

$$q_{j+1} = w / \beta_j.$$

A zero β_j signals an invariant subspace in exact arithmetic. Stacking the identities gives $AQ_k = Q_k T_k + \beta_k q_{k+1} e_k^T$ where T_k is real symmetric tridiagonal. A normalised eigenpair (θ, z) of T_k gives a Ritz vector $Q_k z$ and residual norm $\beta_k |z_k|$. In an eigenbasis of A, that residual is at least the distance from θ to the nearest eigenvalue. To bound an angle to a particular eigenvector also separate its eigenvalue from its neighbours.

Read the first two measurements

Use $A = \operatorname{diag}(1, 2, 4)$ and $q_1 = (1, 1, 1)^T / \sqrt{3}$. Direct averaging gives $\alpha_1 = 7/3$ and $\beta_1 = \sqrt{14}/3$. The second direction is $(-4, -1, 5)^T / \sqrt{42}$, giving $\alpha_2 = 59/21$. The projected pair captures only part of the original three-dimensional spectrum. The first quadrature node $7/3$ already gives the exact weighted average $q_1^T A q_1$, while its square misses $q_1^T A^2 q_1 = 7$ by $14/9$, exactly β_1^2.

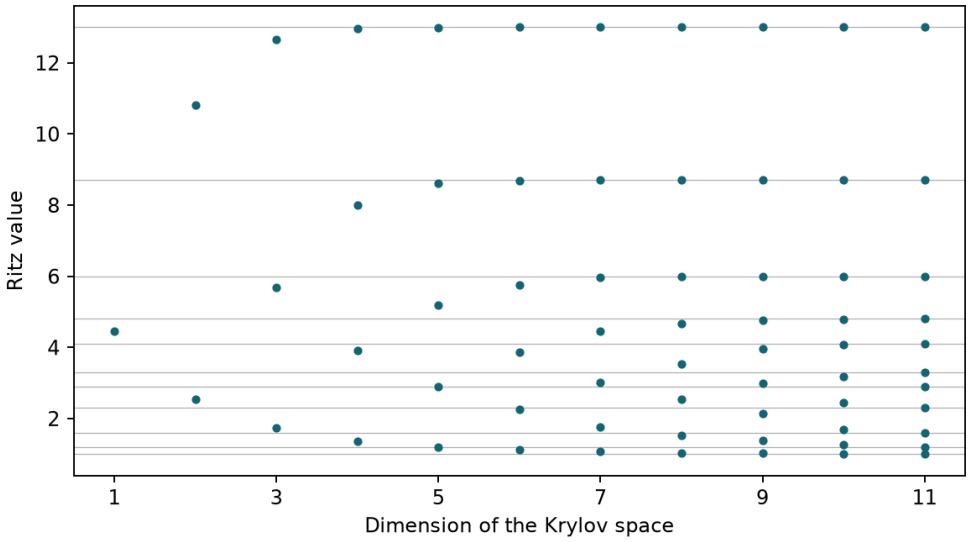
More generally, $q_1^T f(A) q_1$ integrates f against masses $|v_i^T q_1|^2$ at the eigenvalues. For degrees up to $2k - 1$, paths through the tridiagonal recurrence cannot step outside the first k directions and return to the first one in time. Hence

$$q_1^T p(A) q_1 = e_1^T p(T_k) e_1 = \sum_i |z_{1i}|^2 p(\theta_i).$$

This is the Gauss quadrature rule for that spectral measure. It also explains why a mode orthogonal to the starting vector remains invisible regardless of the iteration count.

Why can a computed tridiagonal matrix contain a duplicate mode?

A rounding perturbation can reintroduce a converged eigendirection into later vectors. The three-term recurrence removes only the last two components explicitly. The reintroduced component can grow until a second Ritz vector reports the old mode. Monitor the Gram matrix of stored basis vectors and residuals in the original space. Selective or full reorthogonalisation trades extra products and storage for independent directions.



Each column shows the projected eigenvalues at one subspace dimension. Horizontal lines mark the exact spectrum. The isolated upper mode settles early. The starting vector has nonzero overlap with every mode.

Tutorial on invisible modes and misleading accuracy

Start from a different vector for the same operator

▼ **Work through the solution**

With $q_1 = (1, 0, 0)^T$, the first residual vanishes. The one-node quadrature is exact for every polynomial for this particular observation, despite both eigenvalues 2 and 4 remaining undetected.

Derive the quadrature weights for a two-mode experiment

▼ **Work through the solution**

With equal projections on two distinct eigenvectors, the measure has two masses of one half. The first recurrence coefficient is their arithmetic mean and the first off-diagonal coefficient is half their absolute

difference. The second diagonal equals the first. Diagonalising that two-by-two matrix returns both eigenvalues and weights without prior knowledge of the modes.

Trefethen and Bau · Krylov projections and spectral quadrature

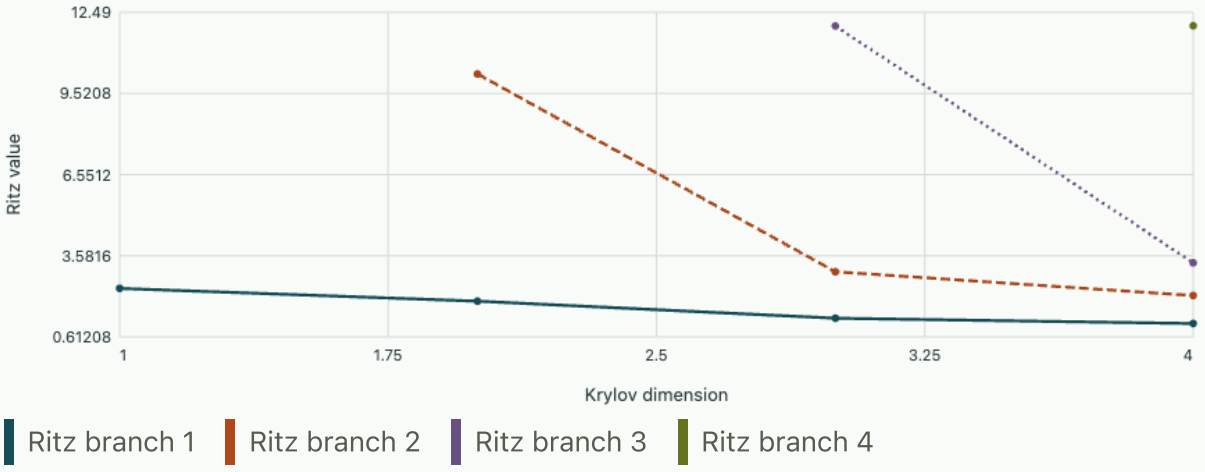
Change parameters · Recovering a spectrum from three-term recurrences

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

Separated largest eigenvalue	12
Starting overlap with highest mode	0.4
Krylov dimension	4



Exact spectrum: 1, 1.4, 2, 2.6, 3.4, 12. Full reorthogonalisation suppresses duplicate directions. Zero initial highest-mode overlap makes that eigenvalue invisible.