MATRIX CONCEPTS

Digital filter design — state-space, stability, and unitary or Hermitian structure

September 2026 • Prepared for Brendan Lynskey

What this report is about, in plain terms. Three matrix ideas do most of the organising work in digital filter design, and each answers a question that comes up on the job rather than only in a textbook. The first is stability, which turns out to be a statement about the eigenvalues of the filter's state matrix A: keep them inside the unit circle and the filter is safe. The second is the choice of optimal coefficients, which is governed by Hermitian structure — Wiener filtering and the eigenfilter method both reduce to an eigenvalue problem on a positive-semidefinite matrix. The third is lossless splitting and rebuilding, the business of dividing a signal into subbands and putting it back together without damage, which is governed by paraunitary structure: a unitary matrix generalised so that it holds at every frequency rather than at one. Each section below takes one of those threads and follows it from the matrix, through what its structure guarantees, to a small worked example. This is a companion article to 'Matrix Methods in Network Parameters'.

1 Introduction: three matrix structures in one filter

You feed samples x[n] in and samples y[n] come out. If the filter remembers anything at all — any recursive structure, and even a plain FIR when you look at it as a delay line — then it carries an internal state, and matrices are the tidy way to keep track of that state. The three questions you actually find yourself asking are all matrix questions in disguise: will this thing stay stable when I quantise it, what coefficients are the best ones for the job, and can I split the signal into bands and still put it back together exactly?

The three sections that follow answer those in turn, and the closing section maps the results back onto the continuous-time networks of the companion article, where the same ideas appear wearing analogue clothes.

A dictionary for the analogue-to-digital boundary. In continuous time, treated in §5 of the companion, a network is stable when its poles lie in the open left half-plane, and the boundary is the jω-axis — which is also the axis along which the frequency response is evaluated, since $s = j\omega$ there. In discrete time, treated in §2 below, a filter is stable when its poles lie strictly inside the unit circle, and the boundary is the circle $|z| = 1$ — which is again also the evaluation contour, since $z = e^{j\omega}$ there. It is the same idea carried across by a conformal map between s and z. Neither of these is the circle that appears in §4 of the companion, where the eigenvalues of a lossless scattering matrix at one frequency also have unit modulus. That is a third setting in which $|\cdot| = 1$ shows up, and it means something quite different: how the ports of a network exchange energy, not whether anything is stable.

2 The state matrix A: eigenvalues and stability

Keep the samples the filter still needs in a state vector **q**. On each tick the new state is the old state pushed through a matrix plus a contribution from the fresh input, **q**[n+1] = A**q**[n] + **b**x[n], and the output is a weighted combination of the state plus perhaps a direct path from the input. Cut the input off and the state simply evolves as A^n applied to whatever it started with, so the long-run behaviour is decided entirely by the eigenvalues of A. A mode whose eigenvalue satisfies $|\lambda| < 1$ fades away, one with $|\lambda| = 1$ hums on forever, and one with $|\lambda| > 1$ grows without bound. Stability therefore means that every eigenvalue lies strictly inside the unit circle, which is the same thing as saying that every pole of H(z) does. The textbook complication of a defective matrix, where repeated eigenvalues bring a factor of n^m along with λ^n, changes the details of the transient but not the location of the fence.

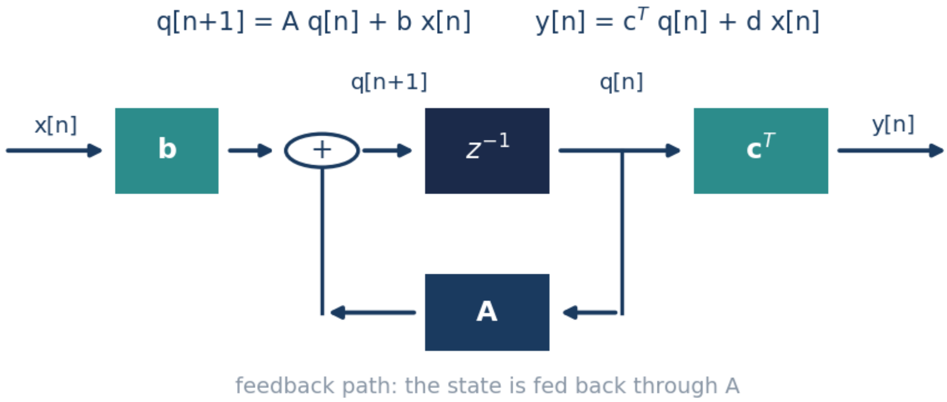

Figure 1 — The state-space form. The summing junction builds **q**[n+1] = A**q**[n] + **b**x[n]; one delay produces **q**[n], which both drives the output tap and returns through A.

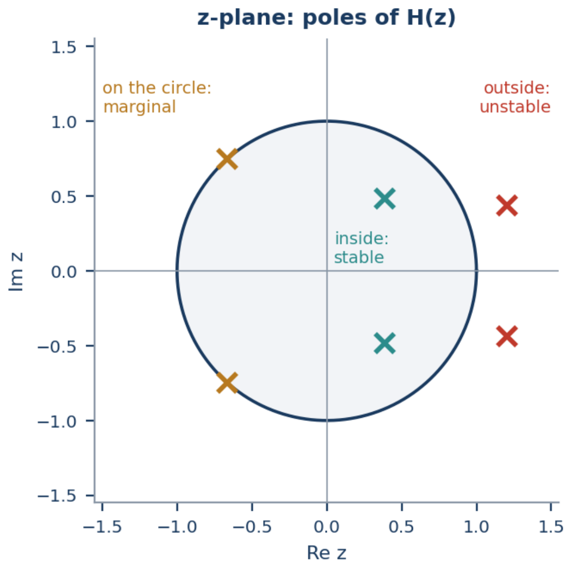

Figure 2 — The z-plane stability map. Inside the circle is stable, on it is marginal, outside it is unstable. This circle is not the one on which the eigenvalues of a lossless S-matrix sit in the companion article.

Two reassurances. The first concerns hidden states. A realisation can contain internal modes that never reach the output, and they cancel out of H(z), so an eigenvalue of A need not be a pole; textbooks call such a realisation non-minimal. In practice the direct, cascade and lattice forms you would actually build are minimal, so eigenvalue and pole coincide and there is no mystery to chase. The second concerns FIR filters, which cannot blow up at all: their state matrix is just a shift register, every eigenvalue is exactly zero, and only the presence of feedback makes instability possible in the first place.

A useful picture for a single real pole is a leaky bucket. At each sample you keep a fraction r of what was already there and pour the fresh input on top. With r = 0.9 the bucket remembers roughly the last ten samples; push r towards 1 and it remembers for a very long time, which is exactly why the response becomes sharply peaked — a filter that averages over many cycles of its favourite tone rejects everything else. With r greater than 1 you have the public-address system howling, each lap round the loop returning louder than the last. For a complex pole the angle θ chooses which tone is the favourite one, and the radius r still decides how long the memory lasts.

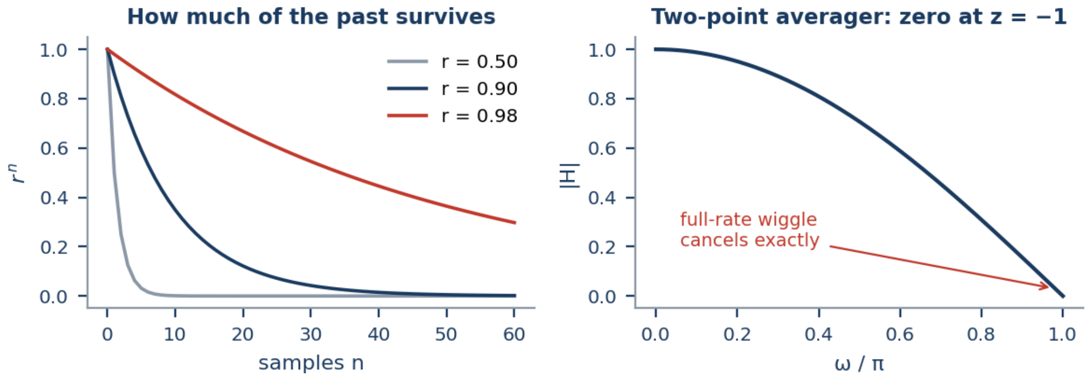

Figure 3 — On the left, how much of an input survives after n samples for three values of the keep-fraction r. On the right, the simplest FIR of all, a two-point averager, whose single zero at z = −1 removes the fastest wiggle the sample rate can carry.

2.1 Why tones are the natural language

Here is the part worth pausing on. Feed any linear time-invariant filter a pure exponential z^n and what comes out is the same exponential, scaled by a single complex number H(z). Delays and weighted sums cannot manufacture a new shape out of an exponential; all they can do is scale it. That is precisely what an eigenvector is for a matrix, and it is why exponentials rather than, say, square waves are the natural basis for describing filters. Each natural mode of the filter is such a tone, λ_k^n, whose angle sets the pitch and whose radius sets the rate at which it swells or decays.

Put a pole pair at radius 0.9 and angle ±0.8 radians per sample and you get a resonator: the magnitude response humps up near ω = 0.8 to about 17 dB. Nudge the radius towards 1 and the hump becomes taller and narrower, which is the continuous-time story about pole distance from the jω-axis wearing digital clothes. Pull the radius back towards 0 and the response flattens out. The two panels of Figure 4 are the same statement told twice, once in frequency and once in time.

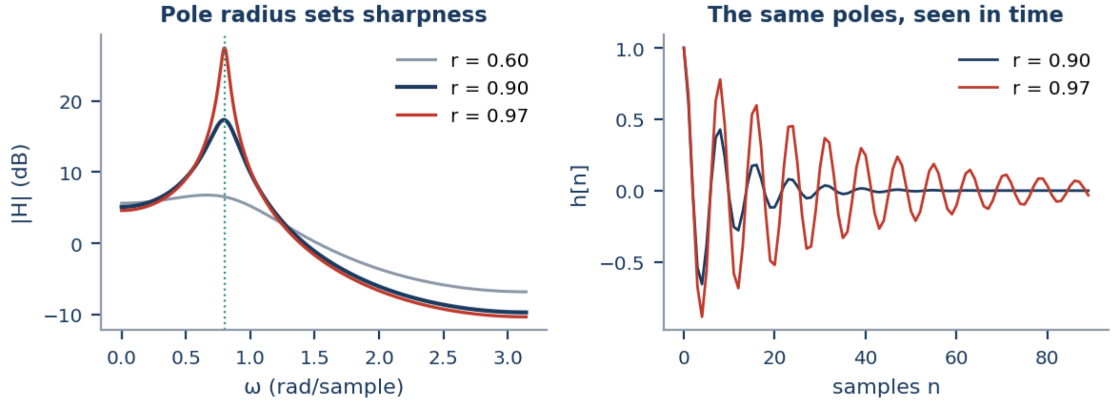

Figure 4 — A resonator at angle 0.8 for three pole radii. Moving the pole towards the unit circle sharpens the peak and lengthens the ringing by the same amount, because these are two descriptions of one behaviour.

2.2 The shortest possible EQ lesson

Take $y[n] = (x[n] + x[n-1])/2$, the average of the current sample and the last one. A constant input sails through at unit gain, because averaging two equal numbers changes nothing. An input alternating between +1 and −1, the fastest wiggle the sample rate can represent, averages to zero every time and is killed outright. That is a zero sitting at $z = -1$ doing its job, and every larger FIR design is the same trick scaled up: place the $N-1$ zeros so that they fall on the parts of the spectrum you want removed and avoid the parts you want kept. There is no feedback anywhere in that description, which is why an FIR has nothing to become unstable with.

3 Hermitian matrices: choosing optimal coefficients

A different matrix appears as soon as you stop asking whether a filter is stable and start asking whether its coefficients are any good. Both classical answers — Wiener filtering and the eigenfilter method — reduce to the same geometric picture, in which the quantity you want to minimise is a bowl-shaped quadratic form and the answer is a particular direction in coefficient space.

For a length-N FIR operating on a stationary input with autocorrelation $r[k]$, the optimal coefficients in the least-squares sense solve the Wiener–Hopf equations $R\mathbf{w} = \mathbf{p}$, where R is the autocorrelation matrix with entries $R_{ij} = r[i-j]$ and $\mathbf{p}$ is the cross-correlation with the desired signal. R is Toeplitz because the input is stationary, and it is Hermitian positive semidefinite by construction, since $\mathbf{w}^H R \mathbf{w} = E|\mathbf{w}^H \mathbf{x}|^2$ is an expected squared magnitude and cannot be negative. That is the same positive-semidefinite structure that expresses passivity in the companion article, arrived at from a completely different direction. The Toeplitz structure is what Levinson–Durbin exploits to solve the system in $O(N^2)$ rather than $O(N^3)$ operations, and the solution has a clean interpretation: the residual error is orthogonal to the data used to predict it, and a lattice implementation peels off one layer of predictability at a time.

Where 'real' matters and where it does not. The Hermitian condition $R \geq 0$ holds for complex IQ data just as it does for real samples; nothing about reality is needed to establish it. What reality buys you is the extra structure of a real symmetric R, which is a special case rather than the general one — and the Toeplitz property comes from stationarity, not from the signal being real. Levinson–Durbin has a standard complex Hermitian form and works perfectly well on baseband IQ. The one thing that genuinely does require real signals is the negative-frequency mirror relation, so an argument that leans on $H(-e^{j\omega})$ being the conjugate of $H(e^{j\omega})$ should not be applied to a complex-baseband design. The same caution appears in §3.3 of the companion for exactly the same reason.

The eigenfilter method makes the geometry explicit. Build a matrix Q that measures how much energy a candidate set of taps leaks into the stopband, normalised to unit total tap energy. Then $\mathbf{w}^H Q \mathbf{w}$ is a bowl over coefficient space, and the eigenvectors of Q are the tap shapes ranked from leakiest to tightest. The best design is the eigenvector belonging to the smallest eigenvalue, and the normalisation constraint is there only to rule out the trivial answer of setting every tap to zero. For a three-tap example the ranking is easy to see by eye: the smooth shape [1, 2, 1] leaks least, and the alternating shape [1, −2, 1] leaks most, because the latter is doing its best to emphasise exactly the high-frequency content the stopband is meant to suppress.

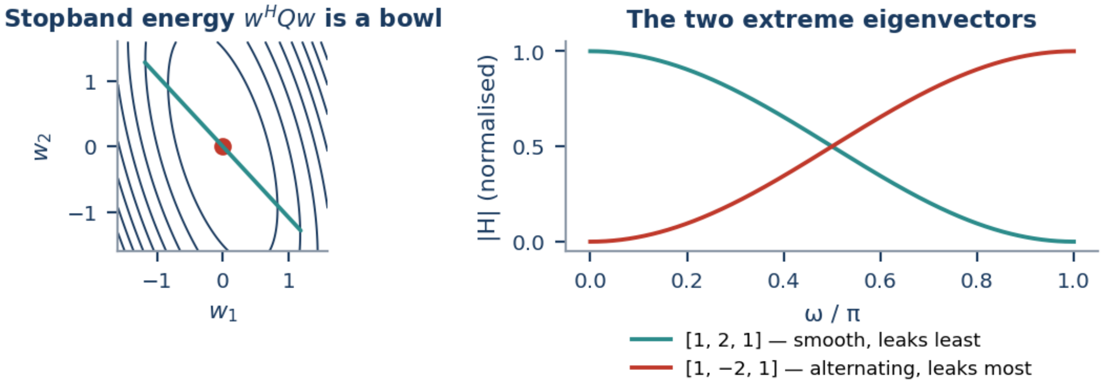

Figure 5 — On the left, stopband energy as a quadratic bowl over coefficient space, with the shallowest direction corresponding to the smallest eigenvector of Q. On the right, the responses of the two extreme three-tap shapes.

4 Paraunitary matrices: lossless filter banks

Splitting a signal into subbands for audio or image coding, processing each band and then rebuilding the original, only works if the splitter conserves energy at every frequency rather than merely on average. Written in terms of the polyphase matrix of the analysis bank, the requirement is that $\tilde{H}_p(z)\, H_p(z) = cI$, where the tilde denotes the paraconjugate — the transpose of H with z replaced by 1/z* and the coefficients conjugated. On the unit circle that condition says nothing more exotic than that the bank performs a rotation at each frequency, which is the same losslessness you met as unitarity in the companion article, now allowed to vary with frequency instead of holding at a single one. The Daubechies wavelets are constructed on exactly this principle.

A rotation, once per frequency. A unitary matrix redistributes energy among its outputs without creating or destroying any. A paraunitary bank does that independently at every point on the unit circle: rotate the signal apart into bands, do whatever you came to do, rotate it back, and in exact arithmetic nothing has been lost. Bit-exact reconstruction on a real machine needs integer lifting on top of this, because rounding breaks the perfect undo even when the mathematics is flawless. Notice which circle is in play: §4.4 of the companion rotates in the plane of the eigenvalues of S at one fixed frequency, whereas here the rotations are indexed by frequency and ride around the z-plane circle.

Perfect reconstruction with no amplitude distortion ⇔ the polyphase matrix is paraunitary

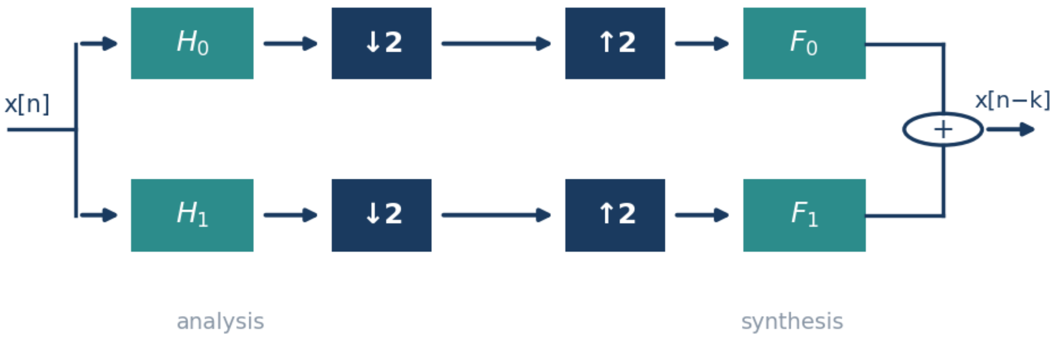

Figure 6 — A two-channel analysis and synthesis bank. Perfect reconstruction with no amplitude distortion is equivalent to the polyphase matrix being paraunitary.

5 Practical relevance, and the map back to continuous time

These structures are not confined to filter design proper. The vector-fitted macromodels of §6 of the companion are realised as state-space systems, and whether such a model is usable comes down to the eigenvalue test of §2 applied to its A matrix. The equaliser taps in a SerDes receiver minimise a quadratic error built from the channel's Hermitian positive-semidefinite autocorrelation, which is the Wiener problem of §3 in different packaging. The two common equaliser architectures divide the labour in a way worth remembering: a feed-forward equaliser is an FIR that pre-shapes the signal and unavoidably boosts noise along with it, while a decision-feedback equaliser operates on already-decided bits and can therefore cancel post-cursor interference without amplifying noise at all, at the cost of being unable to touch pre-cursor interference and of propagating its own errors when it decides wrongly.

Two practical hazards fall outside the linear theory above and deserve naming. Quantising the coefficients of a filter moves its poles, sometimes far enough to cross the unit circle, which is why a stability check belongs after quantisation as well as before it and why cascaded second-order sections are preferred to a single high-order direct form. And overflow and limit cycles are nonlinear finite-wordlength effects that the eigenvalue analysis of §2 simply does not model, because that analysis assumes arithmetic of unlimited precision.

Idea	Continuous-time networks	Discrete-time filters
Stability boundary	The jω-axis, Re(s) = 0	The circle |z| = 1
Stable region	The open left half-plane	The open unit disc
What closeness to the boundary buys	Small σ gives high $Q \approx \omega_0/2\sigma$	Radius r near 1 gives a slow envelope r^n and a peak at ω ≈ θ
Lossless mixing	A unitary S(f), whose eigenvalues have unit modulus at that frequency	A paraunitary $H_p(e^{j\omega})$, unitary at every ω on the circle
Optimal design	Not treated here	Hermitian positive-semidefinite R and Q
What a violation looks like	The model creates energy, and the transient diverges	The recursion is unstable, or settles into a limit cycle

Table 1 — The cross-map between the two articles. The confusion worth guarding against is mixing up the eigenvalues of a scattering matrix, which describe how ports exchange energy at one frequency, with the poles of a transfer function, which describe how the response varies with frequency.

6 Summary

Stability is a statement about the eigenvalues of the state matrix in the z-plane: every $|\lambda_k|$ must be less than 1. An FIR has all its eigenvalues at zero and is stable by construction, so the whole game is played with recursive structures. Optimal coefficients are a statement about Hermitian positive-semidefinite matrices: Wiener–Hopf solved in $O(N^2)$ by Levinson–Durbin, or the eigenfilter's smallest eigenvector, both resting on the spectral theorem in the same way that the passivity test of the companion article does. Lossless analysis and synthesis is a statement about paraunitary matrices, which are unitary matrices made frequency-dependent, and bit-exact reversibility on finite-precision hardware needs lifting on top.

Glossary

BIBO stability. Every bounded input produces a bounded output, which for a minimal realisation is equivalent to every eigenvalue of A lying inside the unit circle.

Minimal realisation. One that is both controllable and observable, so that the eigenvalues of A are exactly the poles of H(z) with none cancelling out.

Paraconjugate and paraunitary. The paraconjugate is $H\sim(z) = H^T(1/z^*)^*$; a matrix is paraunitary when $H\sim H = cI$, which on the unit circle reduces to being a scaled unitary matrix at each frequency.

Rayleigh quotient. The ratio $\mathbf{w}^H Q\mathbf{w} / \mathbf{w}^H\mathbf{w}$, whose stationary points are the eigenvectors of Q. This is what makes the eigenfilter method work.

State matrix A. The matrix in $\mathbf{q}[n+1] = A\mathbf{q}[n] + \mathbf{b}x[n]$; for a minimal realisation its eigenvalues are the poles and they decide stability.

Toeplitz matrix. One whose entries depend only on $i - j$. Stationarity of the input produces this structure, and it is what Levinson–Durbin exploits.

References

[1] Oppenheim and Schafer, Discrete-Time Signal Processing, 3rd ed., Pearson, 2009.

[2] Vaidyanathan, Multirate Systems and Filter Banks, Prentice Hall, 1993.

[3] Proakis and Manolakis, Digital Signal Processing, 4th ed., Pearson, 2006.

[4] Haykin, Adaptive Filter Theory, 5th ed., Pearson, 2013.

[5] Horn and Johnson, Matrix Analysis, 2nd ed., Cambridge, 2012.

[6] Shankar, Basic Training in Mathematics, Springer, 1995.