Matrix Computation II: Spectral Radius, Stability, and Conditioning

N-20260312

§1.1 The long-term question

A norm measures one-step amplification. Spectral radius measures long-term exponential behavior. When a matrix is applied repeatedly,

$$x,\ Ax,\ A^2x,\dots,$$

which directions survive is controlled by eigenvalues.

§1.2 Spectral radius

For $A \in \mathbb{C}^{n\times n}$ with eigenvalues $\lambda_1,\dots,\lambda_n$, define

$$\rho(A) = \max_j |\lambda_j|.$$

For any operator norm,

$$\rho(A) \le \|A\|.$$

Indeed, if $Ax = \lambda x$, $x \ne 0$, then

$$|\lambda|\|x\| = \|Ax\| \le \|A\|\|x\|.$$

For normal matrices,

$$\rho(A) = \|A\|_2.$$

§1.3 Spectral radius is not a norm

The nonzero nilpotent matrix

$$N = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$$

has $\rho(N) = 0$. Thus positive definiteness fails.

Let

$$N_1 = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \qquad N_2 = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}.$$

Then $\rho(N_1) = \rho(N_2) = 0$, but

$$\rho(N_1 + N_2) = 1.$$

So the triangle inequality fails.

Finally, take

$$E = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \qquad F = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}.$$

Then $\rho(E) = \rho(F) = 1$, while $\rho(EF) = (1+\sqrt{5})/2 > 1$. So submultiplicativity fails.

§1.4 Approximating spectral radius

For every $\varepsilon > 0$, there exists a matrix norm such that

$$\|A\| \le \rho(A) + \varepsilon.$$

The proof uses Jordan form. Write $P^{-1}AP = J$. On each Jordan block, conjugate by

$$D = \operatorname{diag}(1, \delta, \delta^2, \ldots),$$

which shrinks the superdiagonal entries from 1 to δ. Choosing δ small makes the norm of the block close to the modulus of its eigenvalue.

§1.5 Convergent matrices

A matrix is convergent if

$$A^k \to 0.$$

The criterion is

$$A^k \to 0 \quad \Longleftrightarrow \quad \rho(A) < 1.$$

Necessity follows because eigenvalues of A^k are λ^k. Sufficiency follows by choosing a norm with $\|A\| < 1$, using the approximation theorem, and then

$$\|A^k\| \le \|A\|^k \to 0.$$

§1.6 Jordan boundary behavior

If $J = \lambda I + N$, $N^m = 0$, then

$$J^k = \sum_{j=0}^{m-1} \binom{k}{j} \lambda^{k-j} N^j.$$

Polynomial factors cannot defeat exponential decay when $|\lambda| < 1$, but they matter when $|\lambda| = 1$. This is why the strict inequality $\rho(A) < 1$ is essential.

§1.7 Neumann series

If $\rho(A) < 1$, then

$$\sum_{k=0}^{\infty} A^k = (I - A)^{-1}.$$

This follows from

$$(I - A)(I + A + \cdots + A^m) = I - A^{m+1}.$$

Neumann series turns a static inverse into repeated feedback accumulation.

§1.8 Matrix power series

If $\sum a_k z^k$ has scalar radius of convergence R, then

$$\sum a_k A^k$$

converges when $\rho(A) < R$ and diverges when $\rho(A) > R$. Eigenvalues decide whether the matrix lies inside the convergence disk.

§1.9 Condition number

For invertible A, define

$$\kappa(A) = \|A\|\|A^{-1}\|.$$

In the Euclidean norm,

$$\kappa_2(A) = \frac{\sigma_{\max}(A)}{\sigma_{\min}(A)}.$$

If A is normal and invertible, then

$$\kappa_2(A) = \frac{\max|\lambda_j|}{\min|\lambda_j|}.$$

A large condition number means solving $Ax = b$ is intrinsically sensitive.

§1.10 Perturbation bound

If $(A + \delta A)\widehat{x} = b + \delta b$, then a standard estimate is

$$\frac{\|x - \widehat{x}\|}{\|x\|} \le \frac{\kappa(A)}{1 - \kappa(A)\frac{\|\delta A\|}{\|A\|}} \left(\frac{\|\delta A\|}{\|A\|} + \frac{\|\delta b\|}{\|b\|} \right),$$

provided

$$\kappa(A)\frac{\|\delta A\|}{\|A\|} < 1.$$

The denominator is part of the theorem, not a decoration.

§1.11 Ill-conditioned example

Let

$$A_\varepsilon = \begin{bmatrix} 1 & -1 \\ -1 & 1+\varepsilon \end{bmatrix}.$$

The eigenvalues are

$$\lambda_\pm = \frac{2 + \varepsilon \pm \sqrt{4 + \varepsilon^2}}{2}.$$

For small ε,

$$\lambda_+ \approx 2, \qquad \lambda_- \approx \varepsilon/2,$$

so

$$\kappa_2(A_\varepsilon) \approx \frac{4}{\varepsilon}.$$

The matrix is invertible for $\varepsilon > 0$, but nearly singular when ε is small.

§1.12 Numerical and ML interpretation

Repeated multiplication appears in recurrent systems, graph propagation, linearized training dynamics, and iterative solvers. Spectral radius tells whether powers decay or grow asymptotically. Condition number tells whether solving or inversion amplifies perturbations.

Mixed precision increases the relative size of rounding perturbations. If the problem is ill-conditioned, those perturbations can become wrong directions. The safe statement is not that condition number explains all training failures, but that it reveals a basic numerical instability channel.

§1.13 Power growth, inversion, and sensitivity

quantity	meaning	where it appears
$\rho(A)$	asymptotic eigenvalue growth	A^k and stability
$\|A\|$	one-step amplification	single perturbation bound
$A^k \to 0$	$\rho(A) < 1$	decaying iteration
$(I-A)^{-1}$	$\sum_{k\geq 0} A^k$	Neumann series
$\kappa(A)$	$\|A\|\|A^{-1}\|$	sensitivity of solving

The distinction between $\rho(A)$ and $\|A\|$ is the main computational point here. A norm controls what can happen in one step, while the spectral radius controls the eventual behavior of powers. Jordan blocks explain why the boundary case $\rho(A) = 1$ is delicate: polynomial growth can remain even when all eigenvalues have modulus one.

For solving $Ax = b$, the condition number measures how much relative error can be magnified by inversion. The ill-conditioned example is therefore not a numerical accident; it shows that a small perturbation in b can point in a direction where A^{-1} is large. This is the same instability channel that appears in iterative dynamics and mixed-precision computations.