

Matrix Computation III: Canonical Forms, Localization, and Iteration

N-20260318

§1.1 Why canonical forms matter

Canonical forms explain what a matrix cannot hide under a change of basis. For computation they answer: how do powers behave, where can eigenvalues be, and which direction dominates an iteration?

§1.2 Polynomial matrices

A λ -matrix is a matrix whose entries lie in $\mathbb{C}[\lambda]$. Its rank is the largest order of a minor that is not the zero polynomial. A square λ -matrix is nonsingular if its determinant is not the zero polynomial.

The elementary operations are row or column exchange, multiplication of a row or column by a nonzero constant, and adding $\varphi(\lambda)$ times one row or column to another.

§1.3 Smith normal form

Every rank r polynomial matrix is equivalent to

$$\text{diag}(d_1(\lambda), \dots, d_r(\lambda), 0, \dots, 0),$$

where each d_i is monic and

$$d_1 \mid d_2 \mid \dots \mid d_r.$$

This is the Smith normal form.

Let $D_k(\lambda)$ be the monic greatest common divisor of all nonzero k -minors. The invariant factors are

$$d_1 = D_1, \quad d_k = \frac{D_k}{D_{k-1}}.$$

Factoring the nonconstant invariant factors gives elementary divisors.

§1.4 Jordan data from elementary divisors

For $A \in \mathbb{C}^{n \times n}$, apply Smith form to

$$\lambda I - A.$$

An elementary divisor

$$(\lambda - \lambda_0)^s$$

corresponds to a Jordan block of size s for eigenvalue λ_0 .

Thus Smith form is an algebraic route to Jordan structure.

§1.5 Jordan chains

A Jordan block is

$$J_s(\lambda) = \begin{bmatrix} \lambda & 1 & & 0 \\ & \lambda & \ddots & \\ & & \ddots & 1 \\ 0 & & & \lambda \end{bmatrix}.$$

If $A = PJP^{-1}$, the columns in one Jordan chain satisfy

$$Ap_1 = \lambda p_1, \quad Ap_i = \lambda p_i + p_{i-1}.$$

The later vectors are generalized eigenvectors.

§1.6 Computing powers

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

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

For $A = PJP^{-1}$,

$$A^k = PJ^kP^{-1}.$$

Jordan form explains polynomial factors in matrix powers.

§1.7 Gershgorin discs

For $A = (a_{ij})$, define

$$R_i = \sum_{j \neq i} |a_{ij}|.$$

Every eigenvalue lies in

$$\bigcup_i \{z \in \mathbb{C} : |z - a_{ii}| \leq R_i\}.$$

Proof: choose i with $|x_i| = \max_j |x_j|$ for an eigenvector x . From $Ax = \lambda x$,

$$(\lambda - a_{ii})x_i = \sum_{j \neq i} a_{ij}x_j,$$

so $|\lambda - a_{ii}| \leq R_i$.

§1.8 Gershgorin figure

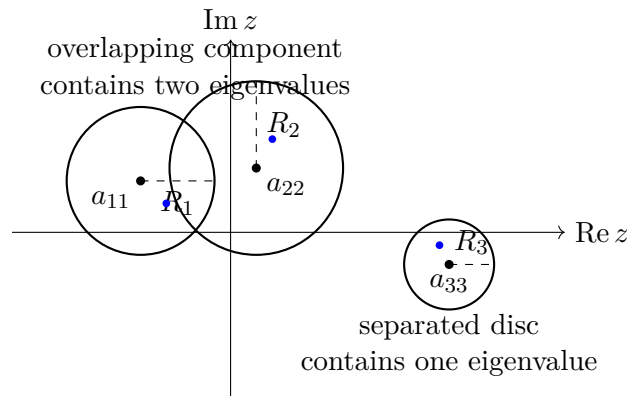


Figure 1.1: Gershgorin discs localize eigenvalues. A separated component containing one disc contains exactly one eigenvalue.

§1.9 Separated components and scaling

If a connected component of the union of Gershgorin discs consists of exactly m discs and is separated from the others, then it contains exactly m eigenvalues.

Column Gershgorin discs can be obtained by applying the theorem to A^T . Diagonal similarity

$$D^{-1}AD$$

does not change eigenvalues but changes radii, since

$$(D^{-1}AD)_{ij} = d_i^{-1}a_{ij}d_j.$$

This is a practical way to improve localization.

§1.10 Power iteration

If

$$|\lambda_1| > |\lambda_2| \geq \cdots,$$

and the initial vector has a nonzero component in the dominant eigenvector direction, then

$$u_{k+1} = \frac{Au_k}{\|Au_k\|}$$

converges toward the dominant eigenvector direction. The convergence rate is governed by

$$\left| \frac{\lambda_2}{\lambda_1} \right|.$$

§1.11 Inverse and shifted inverse iteration

Inverse iteration applies power iteration to A^{-1} , so it finds eigenvalues of small modulus. Shifted inverse iteration applies inverse iteration to

$$A - pI.$$

Eigenvalues of $(A - pI)^{-1}$ are

$$\frac{1}{\lambda_j - p}.$$

Thus the eigenvalue closest to p becomes dominant.

§1.12 Numerical traps

Shifted inverse iteration requires solving systems with $A - pI$, which may be nearly singular. This is why the method can be powerful and unstable at the same time. Pivoting and residual checks are essential.

Gershgorin discs are cheap localization, not exact computation. Overlap does not mean eigenvalues are evenly spread through the overlap.

§1.13 ML-facing interpretation

Power iteration appears in PageRank, spectral normalization, graph propagation, and monitoring dominant eigenvalue behavior. State-space and recurrent models depend on polynomial and spectral structure: the minimal polynomial, Jordan blocks, and spectral radius all describe memory and long-run dynamics.

§1.14 Choosing an eigenvalue tool

Tool	What it exposes	When to use it
Smith form	invariant factors	module or integer-structure questions
Jordan form	eigenvalue chains	powers, nilpotent parts, generalized eigenvectors
Gershgorin	discs containing eigenvalues	quick localization before computation
Power iteration	dominant eigenpair	one eigenvalue separated in modulus
Shifted inverse iteration	eigenvalue near a target	refining a chosen spectral region

The order matters in practice. Gershgorin gives a cheap first picture of where eigenvalues can be and can sometimes separate clusters. Power iteration is then useful only if the dominant eigenvalue is separated and the starting vector has a component in its eigendirection. Shifted inverse iteration reverses the question: after a target μ has been chosen, solving $(A - \mu I)y = x$ amplifies the eigendirection whose eigenvalue is closest to μ . The numerical traps in the note are exactly the failure modes of these assumptions: poor separation, defective eigenspaces, or a shift that makes the linear solve unstable.