Jordan’s Normal Form

Our objective is to demonstrate that for any given complex n-by-n matrix B there exists at least one invertible matrix C that transforms B by *Similarity* into a diagonal sum

$$C^{-1}BC = \begin{bmatrix} \beta_1 I_1 + J_1 & O & O & \dots & O \\ O & \beta_2 I_2 + J_2 & O & \dots & O \\ O & O & \beta_3 I_3 + J_3 & \dots & O \\ \dots & \dots & \dots & \dots & \dots \\ O & O & O & \dots & \beta_L I_L + J_L \end{bmatrix}$$

of *Jordan Blocks* each of the form ßI + J , where ß is an eigenvalue of B and J is obtained from the identity matrix I either by deleting its first row and appending a last row of zeros, or equivalently by deleting its last column and prepending a first column of zeros. For example, here is a 4-by-4 Jordan Block:

$$ßI + J = \begin{bmatrix} \beta & 1 & 0 & 0 \\ 0 & \beta & 1 & 0 \\ 0 & 0 & \beta & 1 \\ 0 & 0 & 0 & \beta \end{bmatrix} .$$

Such a block has one repeated eigenvalue and only one eigenvector regardless of its dimension. Every eigenvalue $ß_j$ of B appears in at least one Jordan Block, and these blocks can appear in any order, and their various dimensions add up to the dimension n of B . We’ll see that B determines its Jordan blocks completely except for the order in which they appear. Since every matrix $Z^{-1}BZ$ *Similar* to B has the same blocks, they tell us all that can be known about the geometrical effect of a linear operator whose matrix, in an unknown coordinate system, is B . For instance they show how B decomposes the vector space into an *Irreducible* sum of *Nested Invariant Subspaces*, as will be explained later.

An important application of Jordan’s Normal Form is the extension of the definitions of scalar functions $f(ß)$ of a scalar argument ß to define matrices $f(B)$. However, we shall find that $f(B)$ is easier to find from a *Pennants* form of B , or from a triangular *Schur* form.

Jordan’s canonical form under similarity is hard to discover because it can be a discontinuous function of its data B . For example, no matter how tiny the nonzero number μ may be, Jordan’s Normal Form of

$$\begin{bmatrix} \beta & 1 & 0 & 0 \\ 0 & \beta & 1 & 0 \\ 0 & 0 & \beta & 1 \\ \mu & 0 & 0 & \beta \end{bmatrix}$$

must be diagonal with four 1-by-1 Jordan blocks; do you see why? And do you see why Jordan’s Normal Form of

$$\begin{bmatrix} \beta & 1 & 0 & 0 \\ 0 & \beta & \mu & 0 \\ 0 & 0 & \beta & 1 \\ 0 & 0 & 0 & \beta \end{bmatrix}$$

is the same for all $\mu \neq 0$? Irreducible invariant subspaces are not determined uniquely if $\mu = 0$.

Discovering the Jordan blocks takes several steps each intended to simplify the problem. The first step identifies the eigenvalues $ß_j$ of B as the zeros (generally complex numbers) of its *Characteristic Polynomial* $\det(\lambda I - B) = \lambda^n - \mathrm{Trace}(B)\lambda^{n-1} + \dots + (-1)^n \det(B) = \prod_j (\lambda - ß_j)$.

The Cayley-Hamilton Theorem:

Every square matrix satisfies its own Characteristic Equation;

i.e., $f(B) = O$ when $f(\lambda) := \det(\lambda I - B) = \sum_{0 \le j \le n} f_j \lambda^j$ is the characteristic polynomial of B. This theorem is stated with an incorrect proof or none in many texts on linear algebra, which is reason enough to present a correct proof here:

Let the *Classical Adjoint* or *Adjugate* of $\lambda I - B$ be $A(\lambda) := \mathrm{Adj}(\lambda I - B)$. It is known to satisfy $A(\lambda)(\lambda I - B) = (\lambda I - B)A(\lambda) = f(\lambda)I$. At first sight, we could replace the scalar λ by the matrix B in the last equation to get $f(B) = (BI - B)A(B) = O$, which is what the theorem claims. But this is not a proof. How do we know that a matrix identity, valid for all *scalar* values of a variable λ, remains valid after λ is replaced by a *matrix* ? It's not so in general, as the next examples show: Set

$$P := \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}, \quad Q := \begin{bmatrix} 1 & 1 \\ -1 & -1 \end{bmatrix}, \quad R := \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \quad \text{and} \quad S := -\begin{bmatrix} 0 & 0 \\ 4 & 4 \end{bmatrix} ;$$

then $P\lambda Q = O$ for all scalars λ but $PRQ = P \ne O$, and $(Q - \lambda I)(Q + \lambda I) = -\lambda^2 I$ for all scalars λ but $(Q - PI)(Q + PI) = S \ne -P^2 I$. These counter-examples reveal a flaw in the simple-minded substitution of B for λ above. A correct proof must be more complicated:

Each element of adjugate $A(\lambda)$ is a polynomial in λ of degree at most $n-1$; it must have the form $A(\lambda) = \sum_{0 \le j < n} A_j \lambda^j$ in which every coefficient A_j is an n-by-n matrix. In fact every A_j is a polynomial in B computable from the identity $A(\lambda)(\lambda I - B) = (\lambda I - B)A(\lambda) = f(\lambda)I$,

$$\textit{i.e.} \quad (\lambda I - B) \sum_{0 \le j < n} A_j \lambda^j = \sum_{0 \le j \le n} f_j \lambda^j I ,$$

by matching the coefficients of successive powers of λ. Begin with the coefficient of λ^n; $A_{n-1} = f_n I = I$. Then for λ^{n-1} find that $A_{n-2} - BA_{n-1} = f_{n-1} I$, so $A_{n-2} = f_{n-1} I + B$. And in general $A_{j-1} = f_j I + BA_j$ for $j = n, n\text{-}1, \ldots, 3, 2, 1, 0$ in turn, starting from $A_n := O$ and ending at $A_{-1} := O$ to meet end-conditions in the sums. This confirms by reverse induction that every A_j is a polynomial in B with coefficients drawn from the numbers f_j, and therefore $BA_j = A_j B$ just as $\lambda A_j = A_j \lambda$, justifying simple-minded substitution. Alternatively observe that

$$O = A_{-1} = f_0 I + B(f_1 I + B(f_2 I + \ldots + B(f_{n-1} I + B)\ldots)) = f(B) ,$$

which is what the Cayley-Hamilton theorem claims. End of proof.

Triangular Forms *Similar* to B

Two other forms are almost as useful as Jordan's and far easier to exhibit. First is Schur's decomposition $B = QUQ^*$ in which $Q^* = Q^{-1}$ and U is upper-triangular with the eigenvalues of B on its diagonal. This unitary similarity has many uses and is relatively easy to compute with fair accuracy (QUQ^* is almost exactly B); its existence will be demonstrated below.

The second form to which every square matrix B can be reduced by complex similarity is a diagonal sum of triangular matrices of which each has only one eigenvalue, and this eigenvalue is distinct from the eigenvalue of every other triangle in that sum. Though still a continuous function of B, this similarity is more difficult to compute than Schur's, as we shall see later.

Schur's triangularization will be shown to exist through a process of *deflation* ; as each eigenvalue of B is chosen its eigenvector will be used to reduce by 1 the dimension of the matrix from which the next eigenvalue of B will be chosen. Here is how deflation works:

Choose any eigenvalue $ß_1$ of B and find eigenvector v_1 as a nonzero solution of the singular homogeneous linear system $(ß_1I - B)v_1 = o$. Then embed v_1 in a new basis $V := [v_1, v_2, \ldots]$ of the vector space as its first basis vector. B is the matrix of a linear operator whose matrix in the new basis is $V^{-1}BV = \begin{bmatrix} \beta_1 & b^T \\ o & \overline{B} \end{bmatrix}$ because $Bv_1 = ß_1v_1$ so $V^{-1}Bv_1 = \begin{bmatrix} ß_1 \\ o \end{bmatrix}$. Here $\overline{B}$ is a square matrix whose dimension is 1 less than B 's . The eigenvalues of B are $ß_1$ and the eigenvalues of $\overline{B}$ because $\det(\lambda I - B) = \det(\lambda I - V^{-1}BV) = (\lambda - ß_1)\det(\lambda I - \overline{B})$. What was just done to B can now be done to $\overline{B}$: Choose any eigenvalue $ß_2$ of $\overline{B}$ (and of B) and solve $\overline{B}\,\bar{v}_2 = ß_2\bar{v}_2$ for a nonzero eigenvector $\bar{v}_2$ of $\overline{B}$ (not of B) and then form a new basis $\overline{V} := [\bar{v}_2, \bar{v}_3, \ldots]$ for the space upon which $\overline{B}$ (not B) acts; the first column of $\overline{V}^{-1}\overline{B}\,\overline{V}$ is

$$\overline{V}^{-1}\overline{B}\,\bar{v}_2 = \begin{bmatrix} \beta_2 \\ o \end{bmatrix} . \text{ Set } W := \begin{bmatrix} 1 & o^T \\ o & \overline{V} \end{bmatrix} \text{ to find } (VW)^{-1}B(VW) = \begin{bmatrix} \beta_1 & b^T\overline{V} \\ o & \overline{V}^{-1}\overline{B}\,\overline{V} \end{bmatrix} = \begin{bmatrix} \beta_1 & \ldots & \ldots \\ 0 & \beta_2 & \ldots \\ o & o & \ldots \end{bmatrix} .$$

Repeating the process ultimately delivers an upper-triangular $U = Q^{-1}BQ$ with its eigenvalues on its diagonal in the order in which they were chosen as eigenvalues of B .

Exercise: Use this U to deduce the Cayley-Hamilton Theorem from the factorization $\det(\lambda I - B) = \prod_j (\lambda - ß_j)$. (Because the theorem's proof given earlier required no knowledge of eigenvalues, it works also for a scalar field, like the Rational field, whose matrices B may "lack" eigenvalues because the field is not algebraically closed.)

Schur's triangularization is a special case of deflation performed by *Unitary Similarities*. The given matrix B is regarded as a linear operator that maps a *Unitary Space* to itself; the space is endowed with a length $\|v\| := \sqrt{(v^*v)}$ defined as the root-sum-squares of the magnitudes of the elements of vector v . Only *orthonormal* bases are used for this space; every change from one such basis to another is represented by a *Unitary Matrix* whose inverse equals its complex conjugate transpose. When eigenvector v_1 is found it is divided by its length to *normalize* it so that $\|v_1\| = 1$, and then it is embedded in an orthonormal basis $V := [v_1, v_2, v_3, \ldots]$ so that $V^{-1} = V^*$. There are many ways to construct such a V . One computes subsequent columns $v_2, v_3, \ldots$ by applying Gram-Schmidt orthogonalization to the columns of $[v_1, I]$ and discarding a resulting column of zeros. Another computes the elementary orthogonal reflector $V = V^* = V^{-1} = I - 2uu^*/u^*u$ that swaps v_1 with the first column of the identity I . Likewise for $\overline{V}$, so W above is unitary too, and so is the product Q of unitary matrices. Thus we obtain Schur's triangularization $U = Q^*BQ$ with $Q^* = Q^{-1}$.

Below we'll need a special Schur triangularization in which repeated eigenvalues of B appear in adjacent locations on the diagonal of U . If $ß_1$ is a repeated eigenvalue of B it is also an eigenvalue of $\overline{B}$ and can be chosen for $ß_2$ above. Thus can the needed ordering of B 's eigenvalues be imposed upon all diagonal elements of upper triangle U , at least in principle.

Exercise: Use Schur’s triangularization to deduce that every Hermitian matrix $A = A^*$ is unitarily similar to a real diagonal matrix, so the eigenvectors of A can be chosen to form a complex orthonormal basis.

A real Schur decomposition of a real square matrix $B = QUQ^T$ exists in which $Q^T = Q^{-1}$ is real orthogonal and U is real block-upper-triangular with 1-by-1 and 2-by-2 blocks on its diagonal; each 1-by-1 block is a real eigenvalue of B, and each 2-by-2 block is a real matrix with two complex-conjugate eigenvalues of B. The existence is proved by choosing, for any complex eigenvalue $ß_1$, a pair of orthogonal vectors that span an invariant subspace of B belonging to this complex eigenvalue and its conjugate, and embedding the pair as the first two vectors of a new basis for the space. This change of basis deflates the eigenproblem of B to a real matrix $\overline{B}$ of dimension 2 less than B ’s. The deflation and subsequent real block-triangularization is otherwise very much like the foregoing complex triangularization.

Exercise: Use Schur’s real triangularization to deduce that every real symmetric matrix $A = A^T$ is orthogonally similar to a real diagonal matrix, so the eigenvectors of A can be chosen to form a real orthonormal basis.

Next we shall show how Schur’s $U = Q^*BQ$ can be reduced by further similarities (changes of basis) to a diagonal sum of *Pennants*, each an upper-triangular matrix with one eigenvalue of B repeated as often as the triangle’s dimension. First we need a …

Lemma £ : Suppose square matrices F and P of perhaps different dimensions have no eigenvalue in common†. Define a linear operator $£(Z) := ZF - PZ$ mapping the vector space of matrices Z of appropriate dimensions to itself; then $£(Z) = O$ only when $Z = O$, so this linear operator £ is invertible.

Proof: Suppose $£(Z) = O$. Then $PZ = ZF$ and hence $P^2Z = PZF = ZF^2$, and similarly $P^kZ = ZF^k$ for $k = 0, 1, 2, 3, \dots$. Now consider the characteristic polynomial of F, namely $\Phi(\lambda) := \det(\lambda I - F) = \prod_{1\le j\le n} (\lambda - \varphi_j)$ where $\varphi_1, \varphi_2, \dots, \varphi_n$ are all the eigenvalues (perhaps not all distinct) of F. The Cayley-Hamilton theorem says $\Phi(F) = O$; but all the factors of $\Phi(P) = \prod_{1\le j\le n} (P - \varphi_j I)$ are nonsingular so it is too. Expand $\Phi(\dots)$ in powers of its argument to see term-by-term that $\Phi(P)Z = Z\Phi(F) = O$, so $Z = O$, as claimed. End of proof.

Therefore the equation $£(X) = XF - PX = Y$ can be solved for $X = £^{-1}(Y)$ in terms of F, P and Y of the right dimensions so long as F and P have no eigenvalue in common†, *i.e.* so long as $\mathrm{GCD}\big(\det(\lambda I - F), \det(\lambda I - P)\big) = 1$. Entirely rational closed-form formulas for X do exist:

Let $Y_k := P^{k-1}Y + P^{k-2}YF + \dots + PYF^{k-2} + YF^{k-1}$ for every integer $k \ge 0$. Evidently $Y_0 = O$ and $Y_1 = Y$ and $Y_k = XF^k - P^kX$. For each k substitute Y_k for λ^k in $\Phi(\lambda)$ to get $¥ = X\Phi(F) - \Phi(P)X = -\Phi(P)X$, whereupon $X := -\Phi(P)^{-1}¥$. This is a rational formula requiring no knowledge of eigenvalues, but rarely useful for numerical computation. Another way begins by reducing F and P by similarity each to one of the upper-triangular forms above, say $U := Q^{-1}FQ$ and $R := D^{-1}PD$. Then solve $ZU - RZ = D^{-1}YQ$ for Z element by element starting in its lower left corner and working up to the right by diagonals. Finally set $X := DZQ^{-1}$.

---

† Sometimes the phrase “ disjoint spectra ” is said instead of “ no eigenvalue in common.” First Physicists and Chemists said “spectrum” for the set of eigenvalues of a linear operator that Quantum Mechanics associates with an atom or molecule radiating light in colors characterized by its spectrum. Now Mathematicians say it too.

Block-Diagonalizing a Block-Triangular Matrix

Lemma £ above helps us construct a triangular similarity that block-diagonalizes a given block-triangular matrix: $\begin{bmatrix} P & Y \\ O & F \end{bmatrix}$ is similar to $\begin{bmatrix} I & X \\ O & I \end{bmatrix}^{-1}\begin{bmatrix} P & Y \\ O & F \end{bmatrix}\begin{bmatrix} I & X \\ O & I \end{bmatrix} = \begin{bmatrix} I & -X \\ O & I \end{bmatrix}\begin{bmatrix} P & Y \\ O & F \end{bmatrix}\begin{bmatrix} I & X \\ O & I \end{bmatrix} = \begin{bmatrix} P & O \\ O & F \end{bmatrix}$ whenever matrix X satisfies $XF - PX = Y$, and such an X exists when square matrices P and F have *disjoint spectra* (no common eigenvalue). Repeat this process upon P and F so long as they too are block-triangular and have on their diagonals square blocks whose spectra are disjoint. Such configurations come from the special Schur triangularizations mentioned above.

Therefore, after any square matrix B has been triangularized by a similarity $Q^{-1}BQ = U$ in such a way that equal eigenvalues of B are consecutive on the diagonal of U, similarities can further reduce the triangle U ultimately to a diagonal sum of *Pennants* like this:

$$(QK)^{-1}B(QK) = K^{-1}UK = \begin{bmatrix} \beta_1 I_1 + N_1 & O & O & \dots & O \\ O & \beta_2 I_2 + N_2 & O & \dots & O \\ O & O & \beta_3 I_3 + N_3 & \dots & O \\ \dots & \dots & \dots & \dots & \dots \\ O & O & O & \dots & \beta_k I_k + N_k \end{bmatrix} ;$$

here $ß_1, ß_2, ß_3, \dots, ß_k$ are the *distinct* eigenvalues of B. Each pennant $ßI + N$ is a triangle with only one eigenvalue on its diagonal, like this 4-by-4 example:

$$ßI + N = \begin{bmatrix} \beta & ? & ? & ? \\ 0 & \beta & ? & ? \\ 0 & 0 & \beta & ? \\ 0 & 0 & 0 & \beta \end{bmatrix} .$$

Each such pennant is copied from corresponding elements of U. Each strictly upper-triangular N has zeros on its diagonal and is therefore *Nilpotent*, meaning $N^m = O$ for some positive integer $m \le \dim(N)$. Upper-triangular matrix K is a pennant too, with 1’s on its diagonal and nonzero superdiagonal blocks only where $K^{-1}UK$ puts zero blocks above the diagonal.

The foregoing diagonal sum of pennants can be shown to be a continuous function of B in the following sense: As the elements of B all vary continuously but not too far, the elements of QK and of each pennant in the diagonal sum also vary continuously; however the eigenvalues on each pennant’s diagonal can become no longer all equal, though different pennants’ spectra remain disjoint. Eigenvalues can vary abruptly; an eigenvalue of multiplicity m can split into m eigenvalues that spread apart as fast as the m^{th} root of the perturbations in B, as in the first example above with a tiny perturbation μ. Worse, the elements of QK or $(QK)^{-1}$ can be so gargantuan that roundoff committed during the pennants’ numerical computation gets amplified enough to obliterate some of the data in B. A few computer programs (not MATLAB) try to avoid this obliteration by locating tight clusters of B’s eigenvalues, choosing one cluster per approximated pennant, in such a way that the elements of QK and $(QK)^{-1}$ never become intolerably big. This is a tough task. Several years ago Prof. Ming Gu proved a conjecture of Prof. J.W. Demmel (they are both here at UCB) to the effect that attempts to avoid obliterating the data must occasionally consume a lot of time trying to choose suitable clusters. Specifically, for all dimensions sufficiently large there are rare matrices B for which choosing clusters consumes time that grows like an exponential function of B’s dimension though the time required to compute all approximate pennants would grow like the cube of dimension (comparable to several matrix multiplications or inversions) if a good choice of clusters were known in advance. Therefore our discussion of pennants and of Jordan’s Normal Form is entirely theoretical, not a recipe for infallible numerical computation with arithmetic operations rounded to finite precision.

Functions of Matrices and Sylvester’s Interpolation Formula

The diagonal sum of pennants helps us understand the extension of a scalar function $f(\lambda)$ of a scalar argument λ to matrix arguments. We have already seen how a polynomial $f(\ldots)$ can be extended to a matrix argument; it happened in the Cayley-Hamilton theorem. Analogously, $\exp(\lambda) = \sum_{n\ge0} \lambda^n/n!$ can be extended to define $\exp(B) := \sum_{n\ge0} B^n/n!$ for all square matrices B since the infinite series converges absolutely and, ultimately, very quickly.

Exercise: The economist M. Keynes said “Ultimately we are all dead.” Roughly how many terms of the series for $\exp(-1000)$ must be added up until every subsequent term is tinier than the sum of the series? How much bigger than its sum is the biggest term in the series? One way to answer these questions is to use a computer to generate those terms and the value of $\exp(-1000)$. A better way uses Stirling’s asymptotic approximation

$$n! \approx \sqrt{(2\pi n)}\,\exp(\,n\cdot\log(n) - n + 1/(12n) + \ldots\,) \quad \text{for big } n .$$

Answering these questions will help you appreciate why computers don’t compute $\exp(\ldots)$ just from its series .

Exercise: Why bother to compute $\exp(\ldots)$ for matrices? Confirm from the series that, for any scalar variable τ and *constant* square matrix B , the derivative $d\exp(\tau B)/d\tau = B\cdot\exp(\tau B) = \exp(\tau B)B$. Then show that the solution $y(\tau)$ of $dy/d\tau = By + c(\tau)$ is the vector function $y(\tau) = \exp(\tau B)\big(y(0) + \int_o^\tau \exp(-\theta B)c(\theta)d\theta\,\big)$.

Exercise: We say that matrix Y is a square root of X when $Y^2 = X$. One of $\begin{bmatrix}0&1\\0&0\end{bmatrix}$ and $\begin{bmatrix}0&0&1\\0&0&0\\0&0&0\end{bmatrix}$ has square roots and the other doesn’t; say which and why. Explain why every n-by-n Hermitian Positive Definite $X = X^*$ has at least 2^n square roots Y , and if more than 2^n then infinitely many. Not every such $Y = \sqrt{X}$ since …

If an extension of $f(\ldots)$ to square matrix arguments exists, it is expected to have certain properties, among them $Bf(B) = f(B)B$, $C^{-1}f(B)C = f(C^{-1}BC)$ for all invertible C , and $f(\begin{bmatrix}B_1&O\\O&B_2\end{bmatrix}) = \begin{bmatrix}f(B_1)&O\\O&f(B_2)\end{bmatrix}$ whenever $f(B_j)$ is defined for both square submatrices B_j , since these equations are certainly satisfied when $f(\ldots)$ is a polynomial or an absolutely convergent power series, as you should verify. P.A.M. Dirac summed up these expectations in one line:

$$Zf(B) = f(B)Z \quad \text{for all matrices } Z \text{ that satisfy } ZB = BZ .$$

These expectations are met by *Sylvester’s Interpolation Formula* which expresses $f(B)$ as a polynomial $\Phi(B)$ with coefficients that depend upon B and $f(\ldots)$ as follows:

Suppose polynomials $\Phi(\lambda)$ and $\Psi(\lambda)$ have these properties …
- $\Psi(B) = O$, and
- $f(\lambda) = \Phi(\lambda) + \Psi(\lambda)\cdot\mu(\lambda)$ for a function $\mu(\lambda)$ sufficiently differentiable at all zeros of $\Psi(\ldots)$.

Then Sylvester’s Interpolation Formula defines $f(B) := \Phi(B)$. How does this work?

First $\Psi(\ldots)$ has to be chosen, and then it has to be used to determine $\Phi(\ldots)$. One candidate is $\Psi(\lambda) = \det(\lambda I - B)$ because the Cayley-Hamilton theorem says this $\Psi(B) = O$, but a better one may be the monic polynomial $\Psi(\lambda)$ of minimum degree that satisfies $\Psi(B) = O$; here a monic polynomial of degree d has 1 as the coefficient of λ^d . This $\Psi(\lambda)$ of minimum degree divides evenly into every polynomial $p(\lambda)$ satisfying $p(B) = O$, as would $p(\lambda) = \det(\lambda I - B)$, because $p(\lambda)$ divided by $\Psi(\lambda)$ yields a quotient polynomial $q(\lambda)$ and a remainder polynomial $r(\lambda) = p(\lambda) - \Psi(\lambda)q(\lambda)$ of degree less than $\Psi(\lambda)$ ’s ; but then $r(B) = O$ and so $r(\lambda) = 0$ since no nonzero polynomial $r(\ldots)$ of degree less than $\Psi(\ldots)$ ’s can satisfy $r(B) = O$.

This $\Psi(\dots)$, called " the *Minimum Polynomial* of B " , is determined uniquely by B ; …
Exercise: Explain why. Then use the pennants form of B to show that its every eigenvalue ß is a zero of $\Psi(\dots)$ with the same multiplicity m as the maximum index m at which $N^{m-1} \ne O$ for the pennant ßI + N . Use Jordan's Normal Form to show that the distinct zeros, if any, of the polynomial $\det(\lambda I - B)/\Psi(\lambda)$ are those distinct eigenvalues of B each with two or more eigenvectors; if they exist Sylvester would call B " *Derogatory*."

For Sylvester's formula to work $\Psi(\dots)$ does not have to be the minimum polynomial of B ; any polynomial multiple of that minimum polynomial is acceptable. After $\Psi(\dots)$ has been chosen it determines uniquely a polynomial $\Phi(\dots)$ of minimum degree (less than $\Psi(\dots)$'s) that *Interpolates* (matches) $f(\dots)$ and perhaps its first few derivatives $f'(\dots)$, $f''(\dots)$, … (if they all have finite values) at every zero ß of $\Psi(\dots)$ in the following sense:

if ß is a zero of $\Psi(\dots)$ of multiplicity m ,

which means $\Psi(ß) = \Psi'(ß) = \Psi''(ß) = \dots = \Psi^{(m-1)}(ß) = 0 \ne \Psi^{(m)}(ß)$,

then $\Phi(ß) = f(ß)$, $\Phi'(ß) = f'(ß)$, $\Phi''(ß) = f''(ß)$, …, and $\Phi^{(m-1)}(ß) = f^{(m-1)}(ß)$.

Why must $\Phi(\dots)$ satisfy these equations? How do they determine $\Phi(\dots)$? Why *uniquely* ?
$\Phi(\dots)$ must satisfy the last m equations at every m-tuple zero ß of $\Psi(\dots)$ because $f(\lambda) = \Phi(\lambda) + \Psi(\lambda)\cdot\mu(\lambda)$ was assumed; it implies that the Taylor Series expansion of $f(\lambda)-\Phi(\lambda)$ in powers of $(\lambda-ß)$ begins at $(\lambda-ß)^m$. If the aggregate of those equations for all zeros of $\Psi(\dots)$ determines $\Phi(\lambda)$, it is determined uniquely because the difference between two such polynomials is a polynomial of degree less than $\Psi(\dots)$'s and yet, having all the zeros of $\Psi(\dots)$ with at least the same multiplicities, this difference would have degree no less than $\Psi(\dots)$'s if it did not vanish. $\Phi(\dots)$ is now determined by the aggregate of those equations because of their linearity in the desired coefficients of $\Phi(\dots)$; the number of linear equations is the same as the number of coefficients, namely the degree of $\Psi(\dots)$, and the equations' solution is unique if it exists, so it exists and Sylvester's $f(B) := \Phi(B)$.

Exercise: Prove that if all zeros ßj of $\Psi(\dots)$ are distinct $\Phi(\lambda) = \sum_j f(ß_j)\Psi(\lambda)/((\lambda-ß_j)\Phi'(ß_j))$; this is *Lagrange's Interpolating Polynomial*. Repeated zeros of $\Psi(\dots)$ lead to a different $\Phi(\dots)$ named after Hermite.

In short, when $f(\dots)$ and enough of its derivatives take finite values on the spectrum of B , so that Sylvester's Interpolation Formula $f(B) := \Phi(B)$ does exist, it provides a polynomial that *defines* $f(B)$ but does not provide an always good recipe for computing it.

Analogous situations arise for functions like exp(…), sin(…), cos(…), … which are defined well by infinite series or by differential equations both of which are impractical ways to compute those functions numerically at big arguments; recall the Exercise above about exp(–1000) . For a matrix B of big dimension n , the n–1 matrix multiplications required for an explicit computation of the polynomial $\Phi(B)$ would generally consume time roughly proportional to n^4 , far greater than the time (roughly proportional to n^3) that will be needed to compute $f(B)$ from its pennants form when it can be computed, which is almost always. Worse, the coefficients of a polynomial $\Phi(\dots)$ can be huge compared with its value, so the explicit computation of $\Phi(B)$ can turn into a mess of rounding errors left behind after cancellation. For instance consider the Tchebyshev polynomials $T_k(\lambda) := \cos(k\cdot\arccos(\lambda))$. They *are* polynomials because they satisfy the recurrence $T_0(\lambda) = 1$, $T_1(\lambda) = \lambda$ and $T_{k+1}(\lambda) = 2\lambda T_k(\lambda) - T_{k-1}(\lambda)$ for k = 1, 2, 3, … in turn. Although $-1 \le T_k(\lambda) \le 1$ when $-1 \le \lambda \le 1$, coefficients of $T_k(\lambda)$ grow exponentially with k ; confirm for $k \ge 2$ that $T_k(\lambda) = 2^{k-1}\lambda^k - 2^{k-3}k\lambda^{k-2} + \dots$. This formula is a numerically unstable way to compute values of $T_k(\lambda)$ when λ is almost ± 1 ; use the recurrence instead.

There's a faster way than Sylvester's Interpolation Formula to compute $f(B)$ when a pennants form of B like $(QK)^{-1}B(QK)$ above is available: $f(B) = (QK)\, f((QK)^{-1}B(QK))\, (QK)^{-1}$. It is derived from Sylvester's polynomial thus: $f(B) = \Phi(B) = (QK)\, \Phi((QK)^{-1}B(QK))\, (QK)^{-1}$, in which inner factor $\Phi((QK)^{-1}B(QK))$ could be obtained from the pennants form by replacing its every pennant $ßI+N$ by $\Phi(ßI+N)$. However, $\Phi(\dots)$ need not be computed at all. Since $\Psi((QK)^{-1}B(QK)) = (QK)^{-1}\Psi(B)(QK) = O$, every pennant's $\Psi(ßI+N) = O$; consequently $\Phi(ßI+N) = f(ßI+N)$, which can be computed faster than $\Phi(ßI+N)$ can as follows:

Exercise: Show that $f(ßI + N) = f(ß)I + f'(ß)N + f''(ß)N^2/2 + \dots + f^{(m-1)}(ß)N^{m-1}/(m-1)!$ when $N^m = O$. Use this formula and the pennants form of B to confirm that $\exp(B) = e^\tau \exp(B-\tau I)$ for every scalar τ. (*Warning*: $\exp(B+C) = \exp(B)\cdot\exp(C)$ only if $BC = CB$; otherwise the *Campbell-Baker-Hausdorff* series must be used.)

Fortunately, Sylvester's definition of $f(B) := \Phi(B)$ as a polynomial does not compel us to compute the polynomial, which could have high degree and huge coefficients causing loss of accuracy to roundoff. Instead, the pennants form permits a more direct computation:

$$f(B) = (QK)\begin{bmatrix} f(\beta_1 I_1 + N_1) & O & O & \dots & O \\ O & f(\beta_2 I_2 + N_2) & O & \dots & O \\ O & O & f(\beta_3 I_3 + N_3) & \dots & O \\ \dots & \dots & \dots & \dots & \dots \\ O & O & O & \dots & f(\beta_k I_k + N_k) \end{bmatrix}(QK)^{-1}$$

in which each pennant $f(ßI+N) = f(ß)I + f'(ß)N + f''(ß)N^2/2 + \dots + f^{(m-1)}(ß)N^{m-1}/(m-1)!$ when $N^m = O$. Most often $m = 1$ or 2.

Actually, we needn't compute B's pennants form to compute $f(B)$. The pennants form's purpose is to help us understand $f(B)$, after which we can compute it in better but unobvious ways developed here at UCB by Prof. B.N. Parlett and his students. An outline follows.

Consider first an approximate pennant $P \approx ßI + N$; here P is upper-triangular with all diagonal elements nearly equal, so $P-ßI \approx N$ is nearly nilpotent. We can't be sure the Taylor series

$$f(P) = f(ß)I + f'(ß)(P-ßI) + f''(ß)(P-ßI)^2/2 + \dots + f^{(m-1)}(ß)(P-ßI)^{m-1}/(m-1)! + \dots$$

will converge rapidly when $P-ßI \approx N$ is nearly nilpotent since its elements may be huge. Still, if the dimension m of P is small, a polynomial close to the first m terms of this series can be used in Sylvester's Interpolation Formula to compute $f(P)$ as follows: Let $\Psi(\lambda) := \det(\lambda I-P) = (\lambda-ß_1)(\lambda-ß_2)(\lambda-ß_3)(\dots)(\lambda-ß_m)$ and express

$$f(\lambda) = f(ß_1) + \big(\Delta f(ß_1, ß_2) + \big(\Delta^2 f(ß_1, ß_2, ß_3) + \dots + \Delta^{m-1} f(ß_1, ß_2, \dots, ß_m)(\lambda-ß_{m-1})\dots\big)(\lambda-ß_2)\big)(\lambda-ß_1) + \\ + \Delta^m f(ß_1, ß_2, \dots, ß_m, \lambda)\,\Psi(\lambda)$$

in terms of its *Divided Differences* $\Delta^{k-1} f(ß_1, ß_2, \dots, ß_k)$ using *Newton's Divided Difference Formula*; these are explained (usually in a different notation) in texts like Conte & de Boor *Elementary Numerical Analysis* 3rd ed. (1980) ch. 2. Like derivatives, divided differences can be computed from the function $f(\dots)$ and scalar values of its argument. For instance $\Delta f(ß_1, ß_2) = (f(ß_1)-f(ß_2))/(ß_1-ß_2)$ if $ß_1 \ne ß_2$; otherwise $\Delta f(ß, ß) = f'(ß)$. Then

$$f(P) = f(ß_1)I + \big(\Delta f(ß_1, ß_2) + \big(\Delta^2 f(ß_1, ß_2, ß_3) + \dots + \Delta^{m-1} f(ß_1, ß_2, \dots, ß_m)(P-ß_{m-1}I)\dots\big)(P-ß_2 I)\big)(P-ß_1 I)$$

because $\Psi(P) = O$. Thus $f(P)$ can be computed quickly from a polynomial for small approximate pennants P.

Next consider that $f(\begin{bmatrix} P & Y \\ O & F \end{bmatrix}) = \begin{bmatrix} f(P) & X \\ O & f(F) \end{bmatrix}$ has to satisfy $\begin{bmatrix} P & Y \\ O & F \end{bmatrix} f(\begin{bmatrix} P & Y \\ O & F \end{bmatrix}) = f(\begin{bmatrix} P & Y \\ O & F \end{bmatrix})\begin{bmatrix} P & Y \\ O & F \end{bmatrix}$, so X must satisfy an equation $XF-PX = Yf(F)-f(P)Y$ whose solution exists when the spectra of P and F are disjoint, and X can be computed quickly when P and F are upper-triangular; see after Lemma £. Thus Schur's decomposition $U = Q^{-1}BQ$ provides an upper triangle from which $f(B) = Qf(U)Q^{-1}$ can be computed directly block by block, starting at near-pennant diagonal blocks and working to the upper right, without computing U's pennants form.

So, $f(B)$ can be computed without reducing B to a diagonal sum of pennants by similarities. Can we compute $f(B)$ without knowing Schur's triangle $U = Q^{-1}BQ$ nor B 's eigenvalues?

One might surmise so at first. Certainly $f(B)$ can be computed whenever $f(\lambda) = p(\lambda)/q(\lambda)$ is a *Rational* function, a ratio of polynomials, whose denominator polynomial $q(\lambda)$ has no zero coincident with an eigenvalue of B . (Otherwise $q(B)^{-1}$ would not exist; do you see why?) Moreover $f(B) = p(B)q(B)^{-1} = \Phi(B)$ wherein all the coefficients of Sylvester's polynomial $\Phi(\dots)$ can be determined by finitely many rational arithmetic operations upon the elements of B without first computing its eigenvalues. An inelegant way to do so is suggested by the …

Exercise: Use the Cayley-Hamilton theorem to express $q(B)^{-1}$, if it exists, as a polynomial in B .

Exercise: Show that, in principle, B 's minimum polynomial can be determined by finitely many comparisons and rational arithmetic operations upon B 's elements. Hint: $I, B, B^2, B^3, \dots$ can't all be linearly independent.

The situation is different for non-rational functions. Any *analytic* function $f(\dots)$ can be extended to $f(B)$ provided no eigenvalue of B falls upon a *singularity* of $f(\dots)$. A function $f(\dots)$ is *analytic* if its domain in the complex λ-plane consists of a region at every interior point ß of which the Taylor series

$$f(ß) + f'(ß)(\lambda-ß) + f''(ß)(\lambda-ß)^2/2 + \dots + f^{(m)}(ß)(\lambda-ß)^m/m! + \dots$$

converges to $f(\lambda)$ for all sufficiently small $|\lambda-ß| > 0$. The *singularities* of $f(\dots)$ are the boundary-points of its domain. For example, polynomials and $\exp(\dots)$ and $\sin(\dots)$ are all analytic with no finite singularity; rational functions are analytic with their singularities at their *poles* (where they take infinite values); $\cot(\dots)$ is analytic with singularities at all integer multiples of π ; analytic functions $\ln(\dots)$ and $\sqrt{(\dots)}$ are singular at 0 and along the negative real axis across which they jump discontinuously. Every textbook about analytic functions of a complex variable explains *Cauchy's Integral Formula*

$$f(ß) = \int_{Ç} (f(\lambda)/(\lambda-ß))\, d\lambda /(2\pi\imath)$$

in which $\imath = \sqrt{(-1)}$ and the path of integration runs counter-clockwise around a closed curve $Ç$ lying in the domain of $f(\dots)$ and surrounding ß but no singularity of $f(\dots)$. (The integral is zero if ß lies outside $Ç$.) This formula can be used to extend $f(\dots)$ to a matrix argument:

$$f(B) := \int_{Ç} (f(\lambda)(\lambda I-B)^{-1})\, d\lambda /(2\pi\imath)$$

provided $Ç$ surrounds the spectrum of B but no singularity of $f(\dots)$. Although no eigenvalue of B appears explicitly in this integral formula, it delivers the same result as did Sylvester's Interpolation Formula. To see why, let $\Psi(\dots)$ be either B 's minimum polynomial or its characteristic polynomial and consider $\Theta(ß, \lambda) := (1 - \Psi(ß)/\Psi(\lambda))/(\lambda-ß)$. Despite first appearances, $\Theta(ß, \lambda)$ *is* a polynomial in ß of degree one less than Ψ 's with coefficients that are rational functions of λ . (Do you see why?) Then we find that Sylvester's Polynomial

$$\Phi(ß) := \int_{Ç} f(\lambda)\Theta(ß, \lambda)\, d\lambda /(2\pi\imath) = f(ß) - \Psi(ß) \int_{Ç} (f(\lambda)/((\lambda-ß)\Psi(\lambda)))\, d\lambda /(2\pi\imath)$$

has coefficients determined by $f(\dots)$, by $\Psi(\dots)$ and by integration around any closed curve $Ç$ inside which lies the spectrum of B and no singularity of $f(\dots)$. Then $f(B) = \Phi(B)$ since $\Psi(B) = O$, and yet no eigenvalue of B appears explicitly in the integral that defined $\Phi(\dots)$.

But it's a swindle. Attempts to simplify the integral that defined $\Phi(\dots)$ by expressing it in a "closed form" in terms of elementary transcendental functions like $\ln(\dots)$ and $\arctan(\dots)$ etc. bring the eigenvalues of B back. They persist unless $f(\dots)$ is rational or sometimes algebraic.

Example: exp(Bτ)
Solutions $\eta(\tau)$ of a 2nd-order linear homogeneous differential equation $\eta'' - 2\alpha\eta' - \gamma\eta = 0$ (here $\eta' = d\eta/d\tau$ and $\eta''(\tau) = d^2\eta/d\tau^2$) with constant coefficients α and γ can be obtained from an equivalent first-order homogeneous system of linear differential equations $y' = By$ with $y = \begin{bmatrix}\eta'\\ \eta\end{bmatrix}$ and a constant coefficient matrix $B = \begin{bmatrix}2\alpha & \gamma\\ 1 & 0\end{bmatrix}$. Solutions are $y(\tau) = \exp(B\tau)y_o$ for any constant $y_o = y(0)$. To compute $\exp(B\tau)$ we get the eigenvalues $ß_j = \alpha \pm \sqrt{(\alpha^2 + \gamma)}$ of B as the zeros of its characteristic polynomial $\det(\lambda I - B) = \lambda^2 - 2\alpha\lambda - \gamma = (\lambda - ß_1)(\lambda - ß_2)$.

$ß_1 \ne ß_2$ if $\alpha^2 + \gamma \ne 0$, and then the Lagrange interpolating polynomial that matches $\exp(\lambda\tau)$ at $\lambda = ß_1$ and at $\lambda = ß_2$ is $\Phi(\lambda) := ((\lambda–ß_2)\exp(ß_1\tau) - (\lambda–ß_1)\exp(ß_2\tau))/(ß_1 - ß_2)$, and then $\exp(B\tau) = \Phi(B)$. Note how this degenerates towards 0/0 as $ß_1$ and $ß_2$ approach equality.

$ß_1 = ß_2 = \alpha$ if $\alpha^2 + \gamma = 0$, and then the Hermite interpolating polynomial that matches $\exp(\lambda\tau)$ and its derivative $\tau\cdot\exp(\lambda\tau)$ at $\lambda = \alpha$ is $\Phi(\lambda) := (1 + \tau(\lambda–\alpha))\exp(\alpha\tau)$, and then $\exp(B\tau) = \Phi(B)$ again. Many texts provide this and the former formula for $\Phi(\dots)$.

Although $\exp(B\tau)$ is a continuous function of B , the formula for $\Phi(B)$ jumps from one form to another as $\alpha^2 + \gamma$ passes through zero. This jump, reflecting the discontinuity of Jordan's Normal Form of B rather than any aspect of the differential equation, is an artifact of algebra removable by using the pennants form of B instead of Jordan's, and Newton's interpolating polynomial (twice) instead of Lagrange's. Set $\delta := (ß_1–ß_2)/2 = \sqrt{(\alpha^2 + \gamma)}$ and find

$$\Phi(\lambda) = \exp(\alpha\tau)\big(\cosh(\delta\tau) + (\lambda–\alpha)\sinh(\delta\tau)/\delta \big) = \exp(\alpha\tau)\big(\cos(\imath\delta\tau) - (\lambda–\alpha)\sin(\imath\delta\tau)\imath/\delta \big)$$

wherein it is understood that $\sinh(\delta\tau)/\delta \longrightarrow \tau$ and $-\sin(\imath\delta\tau)\imath/\delta \longrightarrow \tau$ as $\delta \longrightarrow 0$. Despite appearances the two forms of $\Phi(\lambda)$ are really the same because of identities $\cos(\imath\mu) = \cosh(\mu)$ and $\sin(\imath\mu) = \sinh(\mu)\imath$; use whichever is more convenient in the light of the sign of $\alpha^2 + \gamma$.

… … …

Example: Nonexistent and Nonunique √(…)
Nothing like Sylvester's Polynomial Interpolation Formula can cope with extensions of some functions $f(\dots)$ to some square matrices B . The difficulties will be illustrated by using $\sqrt{(\dots)}$ as an example. We may say that Y is a square root of B when $Y^2 = B$, but not every such square root Y , if any exist, is a polynomial in B as Sylvester's formula requires of $\sqrt{B}$. Existence is problematical when 0 is an eigenvalue of B because 0 is also a singularity of $\sqrt{(\dots)}$; its derivative is infinite there. Therefore it is not surprising that $\begin{bmatrix}0 & 1\\ 0 & 0\end{bmatrix}$ has no square roots; but it is surprising that $B = \begin{bmatrix}0 & 1 & 0\\ 0 & 0 & 0\\ 0 & 0 & 0\end{bmatrix}$ has infinitely many square roots $Y = \begin{bmatrix}0 & \xi & \eta\\ 0 & 0 & 0\\ 0 & 1/\eta & 0\end{bmatrix}$, as η and ξ run through all scalar values except $\eta = 0$, yet $\sqrt{B}$ does not exist. A derogatory B , with more than one independent eigenvector for some eigenvalue, may have infinitely many square roots; for instance, $\pm\sqrt{I} = \pm I$ but the 2-by-2 square roots Y of I include also all 2-by-2 matrices with eigenvalues +1 and –1 , namely $Y = \pm\begin{bmatrix}\sqrt{1-\xi\eta} & \eta\\ \xi & -\sqrt{1-\xi\eta}\end{bmatrix}$.

Jordan’s Normal Form of Nilpotent Matrices

These notes began by displaying Jordan’s Normal Form of B ,

$$C^{-1}BC = \begin{bmatrix} \beta_1 I_1 + J_1 & O & O & \dots & O \\ O & \beta_2 I_2 + J_2 & O & \dots & O \\ O & O & \beta_3 I_3 + J_3 & \dots & O \\ \dots & \dots & \dots & \dots & \dots \\ O & O & O & \dots & \beta_L I_L + J_L \end{bmatrix},$$

and now its existence will be proved. This form, the canonical form under *Similarity*, is important because all matrices Similar to B have the same Jordan Normal Form except that its Jordan blocks ßI + J may appear ordered differently along the diagonal. The proof is complicated because the number of blocks and their dimensions depend discontinuously upon B whenever any block is 2-by-2 or bigger, or whenever any two blocks share an eigenvalue ß ; in short, the Jordan Normal Form is discontinuous whenever it is interesting.

These interesting cases are rare because they require that some eigenvalue be repeated, which occurs just when the characteristic polynomial $\det(\lambda I–B)$ and its derivative $\mathrm{Trace}(\mathrm{Adj}(\lambda I–B))$ have at least one zero in common, and therefore have a nontrivial Greatest Common Divisor, which occurs just when these polynomials’ coefficients satisfy a polynomial equation illustrated by the following example:

$\lambda^4 + \omega\lambda^3 + \xi\lambda^2 + \eta\lambda + \zeta$ and its derivative $4\lambda^3 + 3\omega\lambda^2 + 2\xi\lambda + \eta$ vanish together for some λ just when

$$\det\left(\begin{bmatrix} 1 & \omega & \xi & \eta & \zeta & 0 & 0 \\ 0 & 1 & \omega & \xi & \eta & \zeta & 0 \\ 0 & 0 & 1 & \omega & \xi & \eta & \zeta \\ 0 & 0 & 0 & 4 & 3\omega & 2\xi & \eta \\ 0 & 0 & 4 & 3\omega & 2\xi & \eta & 0 \\ 0 & 4 & 3\omega & 2\xi & \eta & 0 & 0 \\ 4 & 3\omega & 2\xi & \eta & 0 & 0 & 0 \end{bmatrix}\right) = 0 .$$

Thus the elements of interesting matrices B , the ones with repeated eigenvalues, satisfy one complicated polynomial equation which places B on a convoluted hypersurface in the space of all matrices of B ’s dimensions. This hypersurface intersects itself in many ways. For example, in the 4-dimensional space of 2-by-2 matrices $\begin{bmatrix} \alpha & \beta \\ \gamma & \delta \end{bmatrix}$ the interesting ones, with a double eigenvalue, lie on a 3-dimensional circular-conical cylinder whose equation is $(\alpha - \delta)^2 + 4\beta\gamma = 0$; its self-intersections form two 2-dimensional planes whereon $\alpha = \delta$ and either $\beta = 0$ or $\gamma = 0$, and these two intersect along a line whereon $\alpha = \delta$ and both $\beta = 0$ and $\gamma = 0$. Only on that line do matrices have two eigenvectors for one eigenvalue. The self-intersections fall into many families of shapes some of which correspond to different special cases in the proof of the Jordan Normal Form’s existence, complicating it.

Our proof will be computational to minimize abstractness and to indicate how Jordan’s Normal Form can be obtained, at least in principle, from arithmetic performed exactly. (Rounded arithmetic operations are problematical for reasons discussed after the pennants form above.) The first step reduces B by Similarity to a pennants form, a diagonal sum of blocks each like ßI + N where ß is an eigenvalue of B and N is strictly upper-triangular and therefore *nilpotent* ; $N^m = O$ for some m . Recall the pennants form $(QK)^{-1}B(QK)$ displayed above.

The next step seeks for each pennant $ßI + N$ an invertible matrix A such that $A^{-1}(ßI + N)A$ is a Similarity exhibiting Jordan’s Normal Form of the chosen $ßI + N$. If all searches succeed, the diagonal sum $\hat{A}$ of the A ’s will provide a Similarity $\hat{A}^{-1}(QK)^{-1}B(QK)\hat{A}$ that exhibits Jordan’s Normal Form of B ’s pennants form and hence of B (with $C = QK\hat{A}$ above).

Two more steps will simplify each search. First observe that $A^{-1}(ßI + N)A = ßI + A^{-1}NA$ is Jordan’s Normal Form of $ßI + N$ just when $A^{-1}NA$ is Jordan’s Normal Form of N, so no generality is lost by seeking Jordan’s Normal Form of only nilpotent matrices. Second, the search will exploit a proof by induction starting with the hypothesis that, for each nilpotent matrix $\tilde{N}$ of dimensions smaller than N ’s, a Similarity can be found that transforms $\tilde{N}$ into its Jordan Normal Form. This hypothesis is valid for all 1-by-1 nilpotent matrices since there is only one of them, namely $[0]$, and the 1-by-1 identity $[1]$ effects the desired Similarity.

Exercise: Although nilpotent $\tilde{N}$ need not be in its pennants form, this form is strictly upper-triangular; why?

Since N is nilpotent, $N^{m-1} \neq O = N^m$ for some positive integer m. It is important to our search. If $m = 1$ then $N = O$ is already in Jordan’s Normal Form and our search is over. Let us suppose $m \ge 2$ and continue searching. Since $N^{m-1} \neq O$ we can find a column vector v such that $N^{m-1}v \neq o$ and then construct matrix $V := [N^{m-1}v, N^{m-2}v, \dots Nv, v]$. Its columns must be linearly independent for the following reason: If $\sum_{k\le j<m} \mu_j N^j v = o$ for some $k \ge 0$ then $N^{m-k-1}\sum_{k\le j<m} \mu_j N^j v = \mu_k N^k v = o$ and so $\mu_k = 0$; setting $k = 0, 1, 2, \dots, m-1$ in turn implies every $\mu_j = 0$. Therefore V can be embedded in a square matrix $[V, \overline{V}]$ all of whose columns are linearly independent; $\overline{V}$ is empty if $m = \text{dimension}(N)$. Anyway $[V, \overline{V}]^{-1}$ exists.

V was designed to satisfy $NV = VJ$ where J is the m-by-m nilpotent Jordan block; if $m = 4$

$J = \begin{bmatrix} 0&1&0&0\\ 0&0&1&0\\ 0&0&0&1\\ 0&0&0&0 \end{bmatrix}$, for example. Therefore $[V, \overline{V}]^{-1}N[V, \overline{V}] = \begin{bmatrix} J & ?\\ O & \tilde{N} \end{bmatrix}$ in which $\tilde{N}$ is empty if $m = \text{dimension}(N)$. If $\tilde{N}$ is empty we have found Jordan’s Normal Form of N so our search is over. Let us suppose $2 \le m < \text{dimension}(N)$ and continue searching.

Now $\tilde{N}$ is not empty but still $\begin{bmatrix} J & ?\\ O & \tilde{N} \end{bmatrix}^m = \begin{bmatrix} J^m & ?\\ O & \tilde{N}^m \end{bmatrix} = [V, \overline{V}]^{-1}N^m[V, \overline{V}] = O$. This tells us $J^m = O$ (which we knew already) and $\tilde{N}^m = O$. Our induction hypothesis says that this nilpotent $\tilde{N}$ can be transformed by a Similarity to its Jordan Normal Form $\overline{A}^{-1}\tilde{N}\overline{A}$, say. Substituting $\overline{V}\,\overline{A}$ for $\overline{V}$ and renaming it $\overline{V}$ replaces $\tilde{N}$ by its Jordan Normal Form in a Similarity

$$[V, \overline{V}]^{-1}N[V, \overline{V}] = \begin{bmatrix} J & R\\ O & \tilde{N} \end{bmatrix}$$

that defines R and exhibits $\tilde{N}$ already as a diagonal sum of Jordan blocks $J_1, J_2, J_3, \dots$ each like J but of perhaps diverse dimensions. Later we shall see how very special R is.

To complete our search we must find one more Similarity, one that gets rid of R.

The desired Similarity takes the form $\begin{bmatrix} I & Z \\ O & I \end{bmatrix}^{-1} \begin{bmatrix} J & R \\ O & \tilde{N} \end{bmatrix} \begin{bmatrix} I & Z \\ O & I \end{bmatrix} = \begin{bmatrix} J & O \\ O & \tilde{N} \end{bmatrix}$. Since $\begin{bmatrix} I & Z \\ O & I \end{bmatrix}^{-1} = \begin{bmatrix} I & -Z \\ O & I \end{bmatrix}$, this Similarity gets rid of R if and only if $Z\tilde{N} - JZ = R$. We have to compute a solution Z of this linear equation in order to compute the desired Similarity. The equation cannot be solved with the aid of Lemma £ above because the spectra of $\tilde{N}$ and J are not disjoint; they are the same. Consequently the linear operator that maps Z to $Z\tilde{N} - JZ$ is singular; $Z\tilde{N} - JZ = R$ can be solved for Z (nonuniquely) if and only if this equation is consistent. To demonstrate its consistency we shall apply one of *Fredholm’s Alternatives* :

The equation $Fz = r$ has at least one solution z if and only if
$w^T r = 0$ whenever $w^T F = o^T$.
(See class notes titled “ The Reduced Row-Echelon Form is Unique”.)

Instead of $w^T r$ we must use Trace(WR) because this runs through all linear combinations of the elements of R as W runs through all matrices of the same shape as R^T . (Do you agree?) Instead of $w^T Fz$ we must use $\mathrm{Trace}(W(Z\tilde{N} - JZ)) = \mathrm{Trace}((\tilde{N}W - WJ)Z)$; the last equation follows from $\mathrm{Trace}((WZ)\tilde{N}) = \mathrm{Trace}(\tilde{N}(WZ))$. As $w^T Fz = 0$ for all z just when $w^T F = o^T$, so does $\mathrm{Trace}(W(Z\tilde{N} - JZ)) = 0$ for all Z just when $\tilde{N}W - WJ = O$. Therefore, instead of $w^T F = o^T$ we must use $\tilde{N}W - WJ = O$. In short, …

To find a Similarity that gets rid of R and exhibits Jordan’s Normal Form of N ,
we must solve $Z\tilde{N} - JZ = R$ for Z , which can be done if and only if
$\mathrm{Trace}(WR) = 0$ whenever $\tilde{N}W = WJ$.

This last line is all that remains for us to prove; we must deduce that Trace(WR) = 0 for *all* solutions W of $\tilde{N}W = WJ$ from a property of R . What property of R do we need? Set

$$S := J^{m-1}R + J^{m-2}R\tilde{N} + J^{m-3}R\tilde{N}^2 + \dots + J^2R\tilde{N}^{m-3} + JR\tilde{N}^{m-2} + R\tilde{N}^{m-1} .$$

S = O is the property we need. To prove S = O confirm first (by induction for m = 1, 2, 3, … in turn) that $\begin{bmatrix} J & R \\ O & \tilde{N} \end{bmatrix}^m = \begin{bmatrix} J^m & S \\ O & \tilde{N}^m \end{bmatrix}$, and then recall that $\begin{bmatrix} J & R \\ O & \tilde{N} \end{bmatrix}^m = [V, \overline{V}]^{-1}N^m[V, \overline{V}] = O$.

From S = O we shall deduce that Trace(WR) = 0 whenever $\tilde{N}W = WJ$ via a sequence of steps that simplify the problem ultimately to a nilpotent N with only two Jordan blocks.

$\tilde{N}$ is a diagonal sum of nilpotent Jordan blocks $J_1, J_2, J_3, \dots$ that induce partitions of R, S and W into blocks of corresponding sizes as follows: split $R = [R_1, R_2, R_3, \dots]$ so that

$\begin{bmatrix} J & R \\ O & \tilde{N} \end{bmatrix} = \begin{bmatrix} J & R_1 & R_2 & R_3 & \dots \\ O & J_1 & O & O & \dots \\ O & O & J_2 & O & \dots \\ O & O & O & J_3 & \dots \\ \dots & \dots & \dots & \dots & \dots \end{bmatrix}$ and then $S = [S_1, S_2, S_3, \dots]$ and $W = \begin{bmatrix} W_1 \\ W_2 \\ W_3 \\ \dots \end{bmatrix}$ compatibly. Every

$S_k := J^{m-1}R_k + J^{m-2}R_kJ_k + J^{m-3}R_kJ_k^2 + \dots + J^2R_kJ_k^{m-3} + JR_kJ_k^{m-2} + R_kJ_k^{m-1} = O$, and every W_k satisfies $J_kW_k = W_kJ$ just when $\tilde{N}W = WJ$, so that $\mathrm{Trace}(WR) = \sum_k \mathrm{Trace}(W_kR_k) = 0$ for *all* such W just when every $\mathrm{Trace}(W_kR_k) = 0$ for every k = 1, 2, 3, … . (You agree?)

Whatever is done to cope with each Jordan block J_k and its associated R_k, S_k and W_k can be done independently of the others; nothing is lost by pretending that $\tilde{N}$ is just one Jordan block thereby dropping the subscripts k to simplify the process. In short, …

There is a Similarity that exhibits Jordan's Normal Form of N if and only if
$$\mathrm{Trace}(WR) = 0 \text{ whenever } \tilde{N}W = WJ$$
where J and $\tilde{N}$ are Jordan blocks, $J^m = O \ne J^{m-1}$, $\tilde{N}^m = O$, and
$$S := \sum_{1\le j\le m} J^{m-j}R\tilde{N}^{j-1} = O .$$

Next let n be the dimension of $\tilde{N}$; from $\tilde{N}^m = O$ follows $n \le m$. (Recall that m is J's dimension.) Let n-by-m matrix $W_o := [O, I]$ have $m-n$ columns of zeros followed by the n-by-n identity matrix. Observe that $\tilde{N}W_o = [O, \tilde{N}] = W_oJ$. Moreover, …

As $\pi(\dots)$ runs through all polynomials of degree less than n,
$W := \pi(\tilde{N})W_o = W_o\pi(J)$ runs through all solutions of $\tilde{N}W = WJ$.

To see why this is so, define function $\mathrm{urc}(X)$ to be the element in the Upper-Right Corner of its matrix argument X. Thus $\mathrm{urc}(W) = \omega_{1m}$. More generally the element of W in row i and column j is $\omega_{ij} = \mathrm{urc}(\tilde{N}^{i-1}WJ^{m-j})$ because pre-multiplication by $\tilde{N}^{i-1}$ shifts row i up to row 1, and post-multiplication by J^{m-j} shifts column j right to column m. If $\tilde{N}W = WJ$ then $\omega_{ij} = \mathrm{urc}(\tilde{N}^{i-1}WJ^{m-j}) = \mathrm{urc}(\tilde{N}^{m-1+i-j}W) =: \pi_{j-i-m+n}$, say. Here $\pi_k = 0$ if $k < 0$ (since then $\tilde{N}^{m-1+i-j} = \tilde{N}^{n-1-k} = O$), and $k = j-i-m+n < m-0-m+n = n$. Set $\pi(\lambda) := \sum_{0\le k<n} \pi_k\lambda^k$ and observe that the element in row i and column j of $\pi(\tilde{N})W_o = \sum_{0\le k<n} \pi_k\tilde{N}^kW_o$ is
$$\begin{aligned}\mathrm{urc}(\tilde{N}^{i-1}\textstyle\sum_{0\le k<n} \pi_k\tilde{N}^kW_oJ^{m-j}) &= \mathrm{urc}(\textstyle\sum_{0\le k<n} \pi_k\tilde{N}^{k+i-1+m-j}W_o) && \dots \text{ since } W_oJ = \tilde{N}W_o \\ &= \mathrm{urc}(\textstyle\sum_{0\le k<n} \pi_k\tilde{N}^{k+i-1+m-j}) && \dots \text{ since } W_o := [O, I] \\ &= \mathrm{urc}(\pi_{j-i-m+n}\tilde{N}^{n-1}) = \pi_{j-i-m+n} = \omega_{ij} ;\end{aligned}$$
therefore every solution W of $\tilde{N}W = WJ$ has the form $W = \pi(\tilde{N})W_o$ as asserted above.

Next we shall demonstrate how every such W satisfies $\mathrm{Trace}(WR) = 0$, completing the proof that there is a Similarity that exhibits the Jordan Normal Form. $\mathrm{Trace}(WR) = \mathrm{Trace}(RW)$, and $W = \sum_{0\le k<n} \pi_k\tilde{N}^kW_o$, so we need only demonstrate $\mathrm{Trace}(R\tilde{N}^kW_o) = O$ for all $k \ge 0$:
$$\begin{aligned}\mathrm{Trace}(R\tilde{N}^kW_o) &= \textstyle\sum_{1\le j\le m} \mathrm{urc}(J^{j-1} R\tilde{N}^kW_o J^{m-j}) && \dots \text{ since } R\tilde{N}^kW_o \text{ is m-by-m} \\ &= \textstyle\sum_{1\le j\le m} \mathrm{urc}(J^{j-1} R\tilde{N}^{m-j+k}W_o) && \dots \text{ since } W_oJ = \tilde{N}W_o \\ &= \mathrm{urc}(\textstyle\sum_{1\le j\le m} J^{j-1} R\tilde{N}^{m-j+k}W_o) = \mathrm{urc}(S\tilde{N}^kW_o) = O . && \text{End of proof.}\end{aligned}$$

The foregoing proof was devoted mostly to proving that the linear system $Z\tilde{N} - JZ = R$ has at least one solution Z, not to computing it. Elements of Z can be computed in order starting at the lower left and working up to the right by diagonals. Computation is tricky because the system $Z\tilde{N} - JZ = R$ is both over- and under-determined. It is over-determined because no solution Z exists unless R is constrained by $S = O$, which means each diagonal of R that ends in its last row must sum to zero. The system is under-determined because it does not determine its solution Z uniquely; another solution is $Z + Z_o\pi(\tilde{N})$ where Z_o is obtained from the n-by-n unit matrix by appending $m-n$ rows of zeros, and $\pi(\dots)$ is any polynomial of degree less than n. Jordan's Normal Form of B can vary discontinuously, and the Similarity $C^{-1}\dots C$ that exhibits it violently discontinuously, as B varies.

Nested Irreducible Invariant Subspaces

We know every linear operator **B** that maps a vector space to itself is represented by a matrix

$$\mathbf{C}^{-1}\mathbf{B}\mathbf{C} = \begin{bmatrix} \beta_1 I_1 + J_1 & O & O & \dots & O \\ O & \beta_2 I_2 + J_2 & O & \dots & O \\ O & O & \beta_3 I_3 + J_3 & \dots & O \\ \dots & \dots & \dots & \dots & \dots \\ O & O & O & \dots & \beta_L I_L + J_L \end{bmatrix}$$

in Jordan's Normal Form if an appropriate basis **C** is chosen. Among the basis vectors must be eigenvectors $\mathbf{c}_j \neq o$ that satisfy $\mathbf{B}\mathbf{c}_j = ß_j\mathbf{c}_j$, but if the Jordan Normal Form is not purely diagonal these eigenvectors are too few to span the space; then extra vectors have to be found to fill out the basis. These extra vectors needed for " an appropriate basis " are called " principal vectors " or " generalized eigenvectors." For every J_j of dimension greater than 1 there is a principal vector $\mathbf{d}_j$ that satisfies $\mathbf{B}\mathbf{d}_j = ß_j\mathbf{d}_j + \mathbf{c}_j$. For every J_j of dimension greater than 2 there is a principal vector $\mathbf{e}_j$ that satisfies $\mathbf{B}\mathbf{e}_j = ß_j\mathbf{e}_j + \mathbf{d}_j$. And so on. For example, if

$$B = ßI + J = \begin{bmatrix} \beta & 1 & 0 & 0 \\ 0 & \beta & 1 & 0 \\ 0 & 0 & \beta & 1 \\ 0 & 0 & 0 & \beta \end{bmatrix} \text{ , then } c = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} \text{ , } d = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} \text{ , } e = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} \text{ , } f = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} \text{ , and } Bf = ßf + e \text{ ,}$$

$Be = ße + d$, $Bd = ßd + c$, and finally $Bc = ßc$.

Unlike eigenvectors, whose directions are determined uniquely except when different Jordan blocks have the same eigenvalue, principal vectors are intrinsically nonunique; for example, d+c is a principal vector as well as d . In fact, for any polynomial $\pi(\dots)$ with $\pi(0) = 1$, the columns of $P := \pi(J)$ after the first provide another set of principal vectors for $ßI + J$; do you see why? Their intrinsic nonuniqueness makes principal vectors sometimes more difficult to handle theoretically than the subspaces they span.

Partition the basis **C** into blocks of basis vectors corresponding to the Jordan blocks of $\mathbf{C}^{-1}\mathbf{B}\mathbf{C}$ thus: $\mathbf{C} = [\mathbf{C}_1, \mathbf{C}_2, \mathbf{C}_3, \dots, \mathbf{C}_L]$, so that $\mathbf{B}\mathbf{C}_j = \mathbf{C}_j(ß_jI + J_j)$. This shows that the range of $\mathbf{C}_j$ is a subspace mapped to itself by **B** ; it is called an *Invariant Subspace* <u>of</u> **B** although strictly speaking it is a subspace of the vector space. **B** decomposes this vector space into an *Irreducible* sum of *Invariant Subspaces* : Each such subspace, spanned by those columns $\mathbf{C}_j$ of **C** that correspond to a Jordan block $(ß_jI_j + J_j)$, is mapped to itself by **B** , has zero intersection with all those other invariant subspaces, and is irreducible (cannot itself be decomposed into a *sum* of two invariant subspaces). **B** 's effect upon each invariant subspace is revealed completely by its corresponding Jordan block, which reveals how this invariant subspace may contain a nested sequence of invariant sub-subspaces as follows:

For simplicity, suppose ß appears in just one Jordan block $ßI + J$ and its dimension is m , so $J^m = O \neq J^{m-1}$. Then the invariant subspace corresponding to this block is determined uniquely as the m-dimensional null-space of $(\mathbf{B} - ß\mathbf{I})^m$. Within this invariant subspace, if $m > 1$, is another invariant sub-subspace, the (m–1)-dimensional nullspace of $(\mathbf{B} - ß\mathbf{I})^{m-1}$. And so on; the innermost nonzero invariant subspace in the nest is the nullspace of $\mathbf{B} - ß\mathbf{I}$.

Thus, when no eigenvalue of **B** appears in more than one Jordan block, all of **B** 's invariant subspaces, including the nested ones, are determined uniquely; and this proves that Jordan's Normal Form is unique except for the ordering of Jordan blocks along the diagonal. Agreed?

But the *Derogatory* case, when some eigenvalue ß of **B** appears in more than one Jordan block, is not so simple. In this case **B** determines uniquely the invariant subspace associated with ß ; it is the nullspace of $(\mathbf{B} - ß\mathbf{I})^k$ for all sufficiently large k . Its further decomposition into a sum of *irreducible* invariant subspaces is an accident of a computational process not determined uniquely by **B** . However, **B** does determine the dimensions of its Jordan blocks uniquely as follows: (Check this by working with the Jordan Normal Form instead of **B** .)

For m = 1, 2, 3, … let $n_m(ß)$ be the number of Jordan blocks of dimension m with ß on their diagonals. Finitely many $n_m(ß) \neq 0$. And $n_1(ß) + n_2(ß) + n_3(ß) + \ldots = \text{Nullity}(\mathbf{B} - ß\mathbf{I})$. Also $n_1(ß) + 2n_2(ß) + 2n_3(ß) + \ldots = \text{Nullity}((\mathbf{B} - ß\mathbf{I})^2)$, from which follows that $n_1(ß) = 2\,\text{Nullity}(\mathbf{B} - ß\mathbf{I}) - \text{Nullity}((\mathbf{B} - ß\mathbf{I})^2)$. In a similar fashion for k = 1, 2, 3, …,

$$n_1(ß) + 2n_2(ß) + 3n_3(ß) + \ldots + (k-1)n_{k-1}(ß) + kn_k(ß) + kn_{k+1}(ß) + \ldots = \text{Nullity}((\mathbf{B} - ß\mathbf{I})^k) ,$$

which implies $n_m(ß) = -\text{Nullity}((\mathbf{B} - ß\mathbf{I})^{m-1}) + 2\cdot\text{Nullity}((\mathbf{B} - ß\mathbf{I})^m) - \text{Nullity}((\mathbf{B} - ß\mathbf{I})^{m+1})$ for all m > 0 . Thus **B** determines the numbers of its Jordan blocks of all dimensions.

Exercise: For any nontrivial projector P , satisfying $O \neq P = P^2 \neq I$, show that its Jordan blocks are all 1-by-1 .

Exercise: First show that every complex square matrix B is Similar to its transpose by showing that they have the same Jordan Normal Form. Then show that B must be a product of two complex symmetric matrices; *i.e.*, $B = QR^{-1}$ where complex $Q = Q^T$ and $R = R^T$. Hint: Reverse the order of invariant subspaces' basis vectors.

The Real Jordan Normal Form

For any given real n-by-n matrix B there exists at least one real invertible matrix C that transforms B by *Similarity* into a diagonal sum

$$C^{-1}BC = \begin{bmatrix} E_1 + K_1 & O & O & \ldots & O \\ O & E_2 + K_2 & O & \ldots & O \\ O & O & E_3 + K_3 & \ldots & O \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ O & O & O & \ldots & E_L + K_L \end{bmatrix}$$

of *Real Jordan Blocks* each of the form E + K where either

E + K = ßI + J with a real eigenvalue ß of B , like this 6-by-6 example: $\begin{bmatrix} \beta & 1 & 0 & 0 & 0 & 0 \\ 0 & \beta & 1 & 0 & 0 & 0 \\ 0 & 0 & \beta & 1 & 0 & 0 \\ 0 & 0 & 0 & \beta & 1 & 0 \\ 0 & 0 & 0 & 0 & \beta & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$,

or

E + K = $(ßI + \mu S) + J^2$ for a pair of complex conjugate eigenvalues $ß \pm \imath\mu$ of B , $\mu \neq 0$, and $S = -S^T$ is a diagonal sum of 2-by-2 skew-symmetric matrices that satisfies $S^2 = -I$.

Here is a 6-by-6 example:

$$E + K = (ßI + \mu S) + J^2 = \begin{bmatrix} \beta & \mu & 1 & 0 & 0 & 0 \\ -\mu & \beta & 0 & 1 & 0 & 0 \\ 0 & 0 & \beta & \mu & 1 & 0 \\ 0 & 0 & -\mu & \beta & 0 & 1 \\ 0 & 0 & 0 & 0 & \beta & \mu \\ 0 & 0 & 0 & 0 & -\mu & \beta \end{bmatrix} .$$

Such a block has two equally repeated complex conjugate eigenvalues $ß \pm \imath\mu$ and only two complex conjugate eigenvectors regardless of its dimension, which is always even.

Every eigenvalue of B appears in at least one Jordan Block, and these blocks can appear in any order, and their various dimensions add up to the dimension of B , which determines its Jordan blocks completely except for the order in which they appear and the signs of their imaginary parts μ . The proof that such a real Jordan Normal Form exists is more complicated than the proof for the complex case but no more illuminating, so it is not presented here.

Exercise: First show that every real square matrix B is Similar to its transpose by showing that they have the same Real Jordan Normal Form. Then show that B must be a product of two real symmetric matrices of which at least one is invertible; *i.e.*, $B = HY^{-1}$ where real $H = H^T$ and $Y = Y^T$. Hint: See the complex case.

A Rational Canonical Form

Any monic polynomial $\Psi(\lambda) = \lambda^n - \mu_1\lambda^{n-1} - \mu_2\lambda^{n-2} - \dots - \mu_{n-1}\lambda - \mu_n$ is the Characteristic

Polynomial of its *Companion Matrix* $Y = \begin{bmatrix} 0 & 1 & 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & 1 & 0 & \dots & 0 & 0 \\ 0 & 0 & 0 & 1 & \dots & 0 & 0 \\ 0 & 0 & 0 & 0 & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & 0 & \dots & 0 & 1 \\ \mu_n & \mu_{n-1} & \mu_{n-2} & \mu_{n-3} & \dots & \mu_2 & \mu_1 \end{bmatrix}$. In fact,

$\Psi(\dots)$ is the *Minimum* polynomial of Y because $\text{Nullity}(\lambda I - Y) \le 1$. (Do you see why?)

Conversely, for any given square matrix B there exists at least one invertible matrix C that

transforms B by *Similarity* into a diagonal sum $C^{-1}BC = \begin{bmatrix} Y_1 & O & O & \dots & O \\ O & Y_2 & O & \dots & O \\ O & O & Y_3 & \dots & O \\ \dots & \dots & \dots & \dots & \dots \\ O & O & O & \dots & Y_L \end{bmatrix}$ of companion

matrices; Y_1 is the companion of the minimum polynomial of B , and every Y_{j+1} is the companion of a divisor of the minimum polynomial of Y_j . Although C may be computed from the elements of B in finitely many rational arithmetic operations, the diagonal sum and C are discontinuous functions of the elements of B ; worse, C is violently discontinuous.

In the 1940s an algorithm called Danilewski's Method used to be recommended as a fast way to compute C and the diagonal sum $C^{-1}BC$, but the method is unreliable for any but a small-dimensioned matrix B whenever it is nearly derogatory; even when Y_1 is not bad, subsequent companion matrices Y_{j+1} need not be companions of divisors of minimum polynomials of previous Y_j s . Hardly any current texts describe the method, so a succinct description is presented here. The method consists of a finite sequence of elementary rational Similarities; each is an elementary column operation followed by its inverse operation applied to rows. For j = 1, 2, 3, … in turn, …

In row j find the after-diagonal element of largest magnitude, and swap that element’s column with column j+1 ; then swap the correspondingly numbered rows.

Divide column j+1 by its element in row j and then multiply row j+1 by that number, unless it is zero in which case stop before the division. Otherwise the element in column j+1 of row j is now 1 .

Subtract multiples of column j+1 from all *other* columns to annihilate their elements in row j ; then add the same multiples of all other rows to row j+1 .

The process stops either because the Similarities have reduced B to the form of a companion matrix Y , or because they have reduced B to the form $\begin{bmatrix} \bar{Y} & O \\ R & \bar{B} \end{bmatrix}$ in which $\bar{Y}$ is a companion matrix but $\bar{B}$ is probably not, and R is probably nonzero. Skipping the last row of $[\bar{Y}, O]$ and resuming the process from the first row of $[R, \bar{B}]$ may (or may not) reduce $\bar{B}$ to companion form, but will probably not annihilate R . Lemma £ *et seq.* offers a way in principle to get rid of R when $\bar{Y}$ and $\bar{B}$ turn out to have disjoint spectra; otherwise deem Danilewski’s method to have failed because of an unlucky initial ordering of the rows and columns of B . The method may succeed if restarted after a shuffle of B ’s rows and corresponding columns has put a different diagonal element into the upper left-hand corner of B . Even if the method would succeed in exact rational arithmetic, it may give poor results in rounded arithmetic if the multiples of column j+1 subtracted from *previous* columns have to be huge multiples that amplify the effect of roundoff. Nobody knows a good way to compute the rational canonical form in rounded arithmetic without computing Jordan’s Canonical Form first.

The Souriau/Frame/Faddeev Method

For a square matrix B of integers, can its characteristic polynomial

$$f(\lambda) := \det(\lambda I - B) = \sum_{0\le j\le n} f_j \lambda^j$$

be computed from the elements of B by exclusively integer arithmetic? If by “ exclusively integer arithmetic ” is meant only adds, subtracts and multiplies, no divides, then all known methods require work that grows exponentially with the dimension of B . But if “ exclusively integer arithmetic ” lets divisions by small integers produce integer quotients, there is a faster method traceable to U.J.J. Leverrier in the mid-nineteenth century and subsequently improved independently by J.M. Souriau, J.S. Frame and D.K. Faddeev a century later. This method requires a number of arithmetic operations of order $\text{dimension}(B)^4$. Approximate methods that compute eigenvalues first go faster for strictly numerical matrices of large dimensions, taking a number of arithmetic operations of order $\text{dimension}(B)^3$. Therefore this method serves mainly for computations (perhaps symbolic) that cannot tolerate roundoff. Here is how it goes:

Set $B_1 := B$ and for j = 1, 2, 3, …, n := dimension(B) in turn compute

$f_j := \text{Trace}(B_j)/j$; … the division goes evenly.

$B_{j+1} := B\cdot(B_j - f_j I)$. (Don’t bother to compute B_{n+1} .)

As a check on the computation, expect $B_n = f_n I$. This vanishes when B is singular, and then its Adjugate $\text{Adj}(B) = (-1)^{n-1} (B_{n-1} - f_{n-1} I)$.

Exercise: Validate the foregoing claims by observing the form B_j takes if B is the companion matrix of $f(\ldots)$.

I doubt the existence of an analogous method to compute B ’s minimum polynomial.