also A is invertible. This proves

(13.4) Proposition: Any diagonally dominant matrix is invertible.

In particular, the first of the three matrices in (13.3) we now know to be invertible. As it turns out, the other two are also invertible; thus, diagonal dominance is only sufficient but not necessary for invertibility. Equivalently, a noninvertible matrix cannot be diagonally dominant.

In particular, for $(A - \mu\,\mathrm{id})$ to be *not* invertible, it must fail to be diagonally dominant, i.e.,

$$\exists i \quad |A(i,i) - \mu| \le \sum_{j \ne i} |A(i,j)|. \tag{13.5}$$

This gives the famous

(13.6) **Gershgorin Circle Theorem**: The spectrum of $A \in \mathbb{C}^{n \times n}$ is contained in the union of circles

$$B_{r_i}(A(i,i)) := \{z \in \mathbb{C} : |A(i,i) - z| \le r_i := \sum_{j \ne i} |A(i,j)|\}, \quad i = 1{:}n.$$

For the three matrices in (13.3), this says that

$$\mathrm{spec}(\begin{bmatrix} 2 & -1 \\ 2 & 3 \end{bmatrix}) \subset B_1(2) \cup B_2(3) \quad \mathrm{spec}(\begin{bmatrix} -2 & -1 \\ 3 & 3 \end{bmatrix}) \subset B_1(-2) \cup B_3(3) \quad \mathrm{spec}(\begin{bmatrix} -2 & -1 \\ 4 & 3 \end{bmatrix}) \subset B_1(-2) \cup B_4(3).$$

More than that, according to a *refinement of the Gershgorin Circle Theorem*, the second matrix must have one eigenvalue in the ball $B_1(-2)$ and another one in the ball $B_3(3)$, since these two balls have an empty intersection. By the same refinement, if the third matrix has only one eigenvalue, then it would necessarily have to be the point -1, i.e., the sole point common to the two balls $B_1(-2)$ and $B_4(3)$.

The trace of a linear map

Recall that the *trace* of a square matrix A is given by

$$\mathrm{trace}(A) = \sum_j A(j,j).$$

Further, as already observed in (6.27), if the product of the two matrices B and C is square, then

$$\mathrm{trace}(BC) = \sum_j \sum_k B(j,k)C(k,j) = \sum_{jk} B(j,k)C(k,j) = \mathrm{trace}(CB). \tag{13.7}$$

Hence, if $A = V\widehat{A}V^{-1}$, then

$$\mathrm{trace}(A) = \mathrm{trace}(V(\widehat{A}V^{-1})) = \mathrm{trace}(\widehat{A}V^{-1}V) = \mathrm{trace}\,\widehat{A}.$$

This proves

(13.8) Proposition: Any two similar matrices have the same trace.

This permits the definition of the **trace** of an arbitrary linear map A on an arbitrary finite-dimensional vector space X as the trace of the matrices similar to it. In particular, $\mathrm{trace}(A)$ equals the sum of the diagonal entries of any Schur form for A, i.e., $\mathrm{trace}(A)$ is the sum of the eigenvalues of A, however with some of these eigenvalues possibly repeated.

For example, $\mathrm{trace}(\mathrm{id}_n) = n$, while $\mathrm{spec}(\mathrm{id}_n) = \{1\}$.

Offhand, such eigenvalue *multiplicity* seems to depend on the particular Schur form (or any other triangular matrix representation) for A. But, since all of these matrices have the same trace, you will not be surprised to learn that all these triangular matrix representations for A have each eigenvalue appear on its diagonal with exactly the same multiplicity. This multiplicity of $\mu \in \mathrm{spec}(A)$ is denoted

$$\#_a\mu$$

and is called the **algebraic multiplicity** of the eigenvalue, and is readily identified as the dimension of $\cup_r \mathrm{null}(A - \mu\,\mathrm{id})^r$. Further, the polynomial

$$\chi_A := \prod_{\mu \in \mathrm{spec}(A)} (\cdot - \mu)^{\#_a\mu}$$

is the much-studied **characteristic polynomial** for A.

It would not take much work to validate all these claims directly. But I prefer to obtain them along more traditional lines, namely via determinants.

Determinants

The **determinant** is, by definition, the unique multilinear alternating form

$$\det : [a_1, \ldots, a_n] \to \mathbb{F}$$

for which

$$\det(\mathrm{id}_n) = 1. \tag{13.9}$$

Here, **multilinear** means that det is *linear* in each of its n arguments, i.e.,

$$\det[\ldots, a + \alpha b, \ldots] = \det[\ldots, a, \ldots] + \alpha \det[\ldots, b, \ldots]. \tag{13.10}$$

(Here and below, the various ellipses ... indicate the other arguments, the ones that are kept fixed.) Further, **alternating** means that the interchange of two arguments reverses the sign, i.e.,

$$\det[\ldots, a, \ldots, b, \ldots] = -\det[\ldots, b, \ldots, a, \ldots].$$

In particular, $\det A = 0$ in case two columns of A are the same, i.e.,

$$\det[\ldots, b, \ldots, b, \ldots] = 0.$$

Combining this last with (13.10), we find that

$$\det[\ldots, a + \alpha b, \ldots, b, \ldots] = \det[\ldots, a, \ldots, b, \ldots],$$

i.e., *addition of a scalar multiple of one argument to a different argument does not change the determinant.*

In particular, *if* $A = [a_1, a_2, \ldots, a_n]$ *is not invertible, then* $\det A = 0$ since then there must be some column a_j of A writable as a linear combination of other columns, i.e.,

$$\det A = \det[\ldots, a_j, \ldots] = \det[\ldots, 0, \ldots] = 0,$$

the last equality by the multilinearity of the determinant.

Conversely, *if* A *is invertible, then* $\det A \neq 0$, and this follows from the fundamental determinantal identity

$$\det(AB) = \det(A)\det(B) \tag{13.11}$$

since, for an invertible A,

$$1 = \det(\mathrm{id}_n) = \det(AA^{-1}) = \det(A)\det(A^{-1}),$$

the first equality by (13.9).

(13.12) Theorem: For all $A \in \mathbb{C}^{n\times n}$, $\mathrm{spec}(A) = \{\mu \in \mathbb{C} : \det(A - \mu\,\mathrm{id}) = 0\}$.

Of course, this theorem is quite useless unless we have in hand an explicit formula for the determinant. Here is the standard formula:

$$\det[a_1, a_2, \ldots, a_n] = \sum_{\mathbf{i}\in \mathrm{S}_\mathbf{n}} (-)^{\mathbf{i}} \prod_j a_j(\mathbf{i}(j)) \tag{13.13}$$

in which the sum is over all permutations of order n, i.e., all 1-1 (hence invertible) maps $\mathbf{i} : \{1, \ldots, n\} \to \{1, \ldots, n\}$, and the number $(-)^{\mathbf{i}}$ is 1 or -1 depending on the **parity** of the number of interchanges it takes to bring the sequence $\mathbf{i}$ back into increasing order.

For $n = 1$, we get the trivial fact that, for any scalar a, $\mathrm{spec}([a]) = \{a\}$.

For $n = 2$, (13.12) implies that

$$\mathrm{spec}(\begin{bmatrix} a & b \\ c & d \end{bmatrix}) = \{\mu \in \mathbb{C} : (a-\mu)(d-\mu) = bc\}.$$

For $n = 3$, we get

$$\mathrm{spec}(\begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix}) = \{\mu \in \mathbb{C} : p(\mu) = 0\},$$

with

$$p(\mu) := (a-\mu)(e-\mu)(i-\mu) + bfg + chd - c(e-\mu)g - (a-\mu)fh - bd(i-\mu).$$

For $n = 4$, (13.13) already involves 24 summands, and, for general n, we have $n! = 1 \cdot 2 \cdots n$ summands. Thus, even with this formula in hand, the theorem is mostly only of theoretical interest since already for modest n, the number of summands involved becomes too large for any practical computation.

In fact, the determinant $\det A$ of a given matrix A is usually computed with the aid of some factorization of A, relying on the fundamental identity (13.11) and on the following

(13.14) Lemma: The determinant of any triangular matrix is just the product of its diagonal entries.

Proof: This observation follows at once from (13.13) since any permutation $\mathbf{i}$ other than the identity $(1, 2, \ldots, n)$ must have $\mathbf{i}(k) < k$ for some k, hence the corresponding product $\prod_j a_j(\mathbf{i}(j))$ in (13.13) will be zero for any lower triangular matrix. Since any such $\mathbf{i}$ must also have $\mathbf{i}(h) > h$ for some h, the corresponding product will also vanish for any upper triangular matrix. Thus, in either case, only the product $\prod_j a_j(j)$ is possibly nonzero. □

So, with $A = PLU$ as constructed by Gauss-elimination, with L unit lower triangular and U upper triangular, and P a permutation matrix, we have

$$\det A = (-)^P \prod_j U(j,j),$$

with the number $(-)^P$ equal to 1 or -1 depending on the parity of the permutation carried out by P, i.e., whether the number of row interchanges made during Gauss elimination is even or odd.

Formula (13.13) is often taken as the definition of $\det A$. It is a simple consequence of the fundamental identity (13.11), and the latter follows readily from the multilinearity and alternation property of the determinant. For these and other details, see the chapter 'More on determinants'.

Annihilating polynomials

The nontrivial polynomial p is called **annihilating for** $A \in L(X)$ if $p(A) = 0$.

For example, A is *nilpotent* exactly when, for some k, the monomial $()^k$ annihilates A, i.e., $A^k = 0$. As another example, A is a linear projector (or, idempotent) exactly when the polynomial $p : t \mapsto t(t-1)$ annihilates A, i.e., $A^2 = A$.

Annihilating polynomials are of interest because of the following version of the **Spectral Mapping Theorem**:

(13.15) Theorem: For any polynomial p and any linear map $A \in L(X)$ with $\mathbb{F} = \mathbb{C}$,

$$\operatorname{spec}(p(A)) = p(\operatorname{spec}(A)) := \{p(\mu) : \mu \in \operatorname{spec}(A)\}.$$

Proof: If $\mu \in \operatorname{spec}(A)$, then, for some nonzero x, $Ax = \mu x$, therefore also $p(A)x = p(\mu)x$, hence $p(\mu) \in \operatorname{spec}(p(A))$. In other words, $p(\operatorname{spec}(A)) \subset \operatorname{spec}(p(A))$.

Conversely, if $\nu \in \operatorname{spec}(p(A))$, then $p(A) - \nu\,\mathrm{id}$ fails to be 1-1. However, assuming without loss of generality that p is a monic polynomial of degree r, we have $p(t) - \nu = (t - \mu_1)\cdots(t - \mu_r)$ for some scalars $\mu_1, \ldots, \mu_r$, therefore

$$p(A) - \nu\,\mathrm{id} = (A - \mu_1\,\mathrm{id})\cdots(A - \mu_r\,\mathrm{id}),$$

and, since the left-hand side is not 1-1, at least one of the factors on the right must fail to be 1-1. This says that some $\mu_j \in \operatorname{spec}(A)$, while $p(\mu_j) - \nu = 0$. In other words, $\operatorname{spec}(p(A)) \subset p(\operatorname{spec}(A))$. □

In particular, if p annihilates A, then $p(A) = 0$, hence $\operatorname{spec}(p(A)) = \{0\}$, therefore $\operatorname{spec}(A) \subset \{\mu \in \mathbb{C} : p(\mu) = 0\}$.

For example, 0 is the only eigenvalue of a nilpotent linear map. The only possible eigenvalues of an idempotent map are the scalars 0 and 1.

The best-known annihilating polynomial for a given $A \in \mathbb{F}^{n\times n}$ is its *characteristic polynomial*, i.e., the polynomial

$$\chi_A : t \mapsto \det(t\,\mathrm{id}_n - A).$$

To be sure, by (10.25), we can write any such A as the product $A = V\widehat{A}V^{-1}$ with $\widehat{A}$ upper triangular. Correspondingly,

$$\chi_A(t) = \det V \det(t\,\mathrm{id}_n - \widehat{A})(\det V)^{-1} = \det(t\,\mathrm{id}_n - \widehat{A}) = \chi_{\widehat{A}}(t) = \prod_j (t - \widehat{A}(j,j)),$$

the last equation by (13.14)Lemma. Consequently, $\chi_A(A) = V\chi_A(\widehat{A})V^{-1}$, with

$$\chi_A(\widehat{A}) = (\widehat{A} - \mu_1\,\mathrm{id})\cdots(\widehat{A} - \mu_n\,\mathrm{id}), \qquad \mu_j := \widehat{A}(j,j), \quad j = 1{:}n,$$

and this, we claim, is necessarily the zero map, for the following reason: The factor $(\widehat{A} - \mu_j \,\mathrm{id})$ is upper triangular, with the jth diagonal entry equal to zero. This implies that, for each i, $(\widehat{A} - \mu_j \,\mathrm{id})$ maps

$$T_i := \mathrm{ran}[e_1, \ldots, e_i]$$

into itself, but maps T_j into T_{j-1}. Therefore

$$\begin{aligned} \mathrm{ran}\,\chi_A(\widehat{A}) = \chi_A(\widehat{A})T_n &= (\widehat{A} - \mu_1 \,\mathrm{id}) \cdots (\widehat{A} - \mu_n \,\mathrm{id})T_n \\ &\subset (\widehat{A} - \mu_1 \,\mathrm{id}) \cdots (\widehat{A} - \mu_{n-1} \,\mathrm{id})T_{n-1} \\ &\subset (\widehat{A} - \mu_1 \,\mathrm{id}) \cdots (\widehat{A} - \mu_{n-2} \,\mathrm{id})T_{n-2} \\ &\cdots \\ &\subset (\widehat{A} - \mu_1 \,\mathrm{id})T_1 \subset T_0 = \{0\}, \end{aligned}$$

or, $\chi_A(\widehat{A}) = 0$, therefore also $\chi_A(A) = 0$. This is known as the **Cayley-Hamilton Theorem**.

Note that the collection $I_A := \{p \in \Pi : p(A) = 0\}$ of all polynomials that annihilate a given linear map A is an **ideal**, meaning that it is a linear subspace that is also closed under multiplication by polynomials: if $p \in I_A$ and $\in \Pi$, then their product $qp : t \mapsto q(t)p(t)$ is also in Since I_A is not empty, it contains a monic polynomial of minimal degree. This polynomial is called the **minimal polynomial for** A and is denoted by p_A. Using the Euclidean algorithm (see Backgrounder), it is easy to see p_A must be a factor of every $p \in I_A$; in technical terms, I_A is a **principal** ideal.

In exactly the same way, the collection $I_{A,x} := \{p \in \Pi : p(A)x = 0\}$ is seen to be a principal ideal, with $p_{A,x}$ the unique monic polynomial of smallest degree in it. Since $I_A \subset I_{A,x}$, it follows that $p_{A,x}$ must be a factor for any $p \in I_A$ and, in particular, for χ_A.

13.1 (a) Prove: *If the minimal annihilating polynomial $p = p_{A,x}$ of the linear map $A \in L(X)$ at some $x \in X\backslash 0$ has degree equal to* $\dim X$, *then* $p_{A,x}(A) = 0$. (b) Prove that the spectrum of the companion matrix (see H.P. ([redacted]"hwcompanion[redacted])) of the monic polynomial p equals the zero set of p.

13.2 make one about the coefs of char.pol. being symmetric functions of evs, and one about the ith coeff. being the sum of the $n-i$th principal minors. all of these, including the trace, are invariant under similarity.

The multiplicity of an eigenvalue

Since χ_A is of exact degree n in case $A \in \mathbb{C}^n$, χ_A has exactly n zeros *counting multiplicities.* This means that

$$\chi_A(t) = (t - \mu_1) \cdots (t - \mu_n)$$

for a certain n-sequence μ. Further,

$$\mathrm{spec}(A) = \{\mu_j : j = 1{:}n\},$$

and this set may well contain only one number, as it does when $A = 0$ or $A = \mathrm{id}$. However, it is customary to associate with each eigenvalue, μ, its **algebraic multiplicity**, $\#_a\mu$, which, by definition, is its multiplicity as a zero of the characteristic polynomial, or, equivalently as we saw, as a diagonal entry of any triangular matrix similar to A. For example, the matrix id_n has only the eigenvalue 1, but it has algebraic multiplicity n. In this way, each $A \in \mathbb{C}^{n\times n}$ has n eigenvalues *counting algebraic multiplicity.*

I have been saying '*algebraic* multiplicity' rather than just 'multiplicity', since there is a second way of counting eigenvalues, and that is by geometric multiplicity. The **geometric multiplicity**, $\#_g\mu$, of the eigenvalue μ for A is, by definition, the dimension of the space of corresponding eigenvectors, i.e.,

$$\#_g\mu := \dim \mathrm{null}(A - \mu \,\mathrm{id}).$$

For the sole eigenvalue, 1, of id_n, algebraic and geometric multiplicity coincide. In contrast, the sole eigenvalue, 0, of $\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$, has algebraic multiplicity 2 but its geometric multiplicity is only 1.

An eigenvalue is (algebraically or geometrically) **simple** if it has (algebraic or geometric) multiplicity 1.

13.16 Proposition: For any eigenvalue, the algebraic multiplicity is no smaller than the geometric multiplicity, with equality if and only if the eigenvalue is not defective.

13.3

(i) Prove that *the multiplicity with which an eigenvalue μ of A appears as a diagonal entry of a triangular matrix T similar to A is the same for all such triangular matrices.* (Hint: Prove that it equals the multiplicity of μ as a zero of the **characteristic polynomial** $\chi_A : t \mapsto \det(t\,\mathrm{id}_n - A)$ of A; feel free to use what we proved about determinants, like: $\det(AB) = \det(A)\det(B)$, and $\det(T) = \prod_j T(j,j)$.)

(ii) Determine the algebraic and geometric multiplicities for all the eigenvalues of the following matrix. (Read off the eigenvalues; use elimination to determine geometric multiplicities.)

$$A := \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 2 & 1 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 3 & 0 & 1 \\ 0 & 0 & 0 & 0 & 3 & 1 \\ 0 & 0 & 0 & 0 & 0 & 3 \end{bmatrix}$$

Perron-Frobenius

We call the matrix A **positive** (**nonnegative**) and write $A > 0$ ($A \ge 0$) in case all its entries are positive (nonnegative). A positive (nonnegative) matrix A of order n maps the **positive orthant**

$$\mathbb{R}^n_+ := \{y \in \mathbb{R}^n : y \ge 0\}$$

into its interior (into itself). Thus the (scaled) power method, started with a nonnegative vector, would converge to a nonnegative vector if it converges. This suggests that the absolutely largest eigenvalue for a nonnegative matrix is nonnegative, with a corresponding nonnegative eigenvector. The Perron-Frobenius theorem makes this intuition precise.

Since A maps $\mathbb{R}^n_+$ into itself, it makes sense to consider, for given $y \in \mathbb{R}^n_+\backslash 0$, scalars α for which $Ay \ge \alpha y$ (in the sense that $(Ay)_j \ge \alpha y_j$, all j), i.e., for which $Ay - \alpha y \ge 0$. The largest such scalar is the nonnegative number

$$r(y) := \min\{(Ay)_j/y_j : y_j > 0\}, \quad y \in \mathbb{R}^n_+\backslash 0.$$

The basic observation is that

$$Ay - \alpha y > 0 \quad \Longrightarrow \quad r(y) > \alpha. \tag{13.17}$$

The function r so defined is scale-invariant, i.e., $r(\alpha y) = r(y)$ for all $\alpha > 0$, hence r takes on all its values already on the set $S_+ := \{y \ge 0 : \|y\| = 1\}$. At this point, we need, once again, a result that goes beyond the scope of these notes, namely the fact that S_+ is compact, while r is continuous at any $y > 0$ and upper semicontinuous at any $y \ge 0$, hence r takes on its supremum over $\mathbb{R}^n_+\backslash 0$ at some point in S_+. I.e., there exists $x \in S_+$ for which

$$\mu := r(x) = \sup r(S_+) = \sup r(\mathbb{R}^n_+\backslash 0).$$

Assume now, in addition to $A \ge 0$, that also $p(A) > 0$ for some polynomial p.

Claim: $Ax = \mu x$.

Proof: Assume that $Ax \ne \mu x$. Since $\mu = r(x)$, we have $Ax - \mu x \ge 0$, therefore $A(p(A)x) - \mu p(A)x = p(A)(Ax - \mu x) > 0$, hence, by (13.17), $r(p(A)x) > \mu = \sup r(S_+)$, a contradiction. □

Claim: $x > 0$.

Proof: Since $0 \neq x \geq 0$ and $p(A) > 0$, we have $p(\mu)x = p(A)x > 0$, hence $x > 0$. □

Consequence: *x is the unique maximizer for r.*

Proof: If also $r(y) = \mu$ for some $y \in S_+$, then by the same argument $Ay = \mu y$, therefore $Az = \mu z$ for all $z = x + \alpha(y - x)$, and each of these z must be positive if it is nonnegative, and this is possible only if $y - x = 0$. □

Consequence: *For any eigenvalue ν of any matrix B with eigenvector y, if $|B| \leq A$, then $|\nu| \leq \mu$, with equality only if $|\,y/\|y\|\,| = x$ and $|B| = A$.* (More precisely, equality implies that $B = \exp(\mathrm{i}\varphi)DAD^{-1}$, with $D := \operatorname{diag}(\ldots, y_j/|y_j|, \ldots)$ and $\exp(\mathrm{i}\varphi) := \nu/|\nu|$.)

Proof: Observe that

$$|\nu||y| = |By| \leq |B|\,|y| \leq A|y|, \tag{13.18}$$

hence $|\nu| \leq r(|y|) \leq \mu$. If now there is equality, then, by the uniqueness of the minimizer x (and assuming without loss that $\|y\| = 1$), we must have $|y| = x$ and equality throughout (13.18), and this implies $|B| = A$. More precisely, $D := \operatorname{diag}(\ldots, y_j/|y_j|, \ldots)$ is then well defined and satisfies $y = D|y|$, hence $C|y| = \mu|y| = A|y|$, with $C := \exp(-\mathrm{i}\varphi)D^{-1}BD \leq A$ and $\nu =: \mu\exp(\mathrm{i}\varphi)$, therefore $C = A$. □

Consequences. By choosing $B = A$, we get that $\mu = \rho(A) := \max\{|\nu| : \nu \in \sigma(A)\}$, *and that μ has geometric multiplicity 1 (as an eigenvalue of A).*

We also get that $\rho(A)$ *is strictly monotone in the entries of A*, i.e., that $\rho(\widehat{A}) > \rho(A)$ in case $\widehat{A} \geq A \neq \widehat{A}$ (using the fact that $p(A) > 0$ and $\widehat{A} \geq A$ implies that also $q(\widehat{A}) > 0$ for some polynomial q; see below).

As a consequence, we find *computable* upper and lower bounds for the spectral radius of A:

Claim:

$$\forall\{y > 0\}\;\; r(y) \leq \rho(A) \leq R(y) := \max_j (Ay)_j/y_j,$$

with equality in one or the other if and only if there is equality throughout if and only if $y = \alpha x$ (for some positive α). In particular, $\rho(A)$ is the only eigenvalue of A with positive eigenvector.

Proof: Assume without loss that $\|y\| = 1$. We already know that for any such $y > 0$, $r(y) \leq \rho(A)$ with equality if and only if $y = x$. For the other inequality, observe that $R(y) = \|D^{-1}ADe\|_\infty$ with $D := \operatorname{diag}(\ldots, y_j, \ldots)$ and $e := (1, \ldots, 1)$. Since $D^{-1}AD \geq 0$, it takes on its max-norm at e, hence

$$R(y) = \|D^{-1}AD\|_\infty \geq \rho(D^{-1}AD) = \rho(A).$$

Now assume that $r(y) = R(y)$. Then $Ay = r(y)y$, hence $r(y) \leq r(x) = \rho(A) \leq R(y) = r(y)$, therefore equality must hold throughout and, in particular, $y = x$.

If, on the other hand, $r(y) < R(y)$, then we can find $\widehat{A} \neq A \leq \widehat{A}$ so that $\widehat{A}y = R(y)y$ (indeed, then $z := R(y)y - Ay$ is nonnegative but not 0, hence $\hat{A} := A + y_1^{-1}[z]{e_1}^{\mathrm{t}}$ does the job) therefore $r_{\widehat{A}}(y) = R(y) = R_{\widehat{A}}(y)$, hence $R(y) = \rho(\widehat{A}) > \rho(A)$. □

Claim: *μ has simple algebraic multiplicity.*

Proof: Since we already know that μ has simple geometric multiplicity, it suffices to show that μ is not a defective eigenvalue, i.e., that $\operatorname{null}(A - \mu\,\mathrm{id}) \cap \operatorname{ran}(A - \mu\,\mathrm{id}) = \{0\}$. So assume to the contrary that $Ay - \mu y$ is an eigenvector of A belonging to μ. Then, by the simple geometric multiplicity of μ, we may assume without loss that $Ay - \mu y = x$, or $Ay = \mu y + x$, therefore, for all k, $A^k y = \mu^k y + k\mu^{k-1}x$, hence, finally,

$$(A/\mu)^k y = y + k(x/\mu).$$

Hence, for large enough k, $z := (A/\mu)^k y$ has all its entries positive, and $Az = Ay + kx = \mu y + (k+1)x = \mu(z + x/\mu) > \mu z$, therefore $r(z) > \mu$, a contradiction. □

The collection of these claims/consequences constitutes the **Perron-Frobenius Theorem**. Oskar Perron proved all this under the assumption that $A > 0$ (i.e., $p(t) = t$). Frobenius extended it to all $A \ge 0$ that are **irreducible**. While this term has some algebraic and geometric meaning (see below), its most convenient definition for the present purpose is that $p(A) > 0$ for some polynomial p. In the contrary case, A is called **reducible**, and not(iv) below best motivates such a definition. Here are some equivalent statements:

Claim: *Let $A \ge 0$. Then the following are equivalent:*

(i) *$p(A) > 0$ for some polynomial p.*

(ii) *For all (i, j), there exists $k = k(i, j)$ so that $A^k(i, j) > 0$.*

(iii) *No proper A-invariant subspace is spanned by unit-vectors.*

(iv) *For no permutation matrix P is*

$$PAP^{-1} = \begin{bmatrix} B & C \\ 0 & D \end{bmatrix} \tag{13.19}$$

with B, D square matrices of positive order.

(v) *The directed graph for A is strongly connected.*

Proof: (ii)$\Longrightarrow$(i) since then $p(A) := \sum_{i,j} A^{k(i,j)} > 0$.

If (ii) does not hold, then there exists (i, j) so that $A^k(i, j) = 0$ for all k. But then also $p(A)(i, j) = 0$ for all polynomials p; in other words, (i)$\Longrightarrow$(ii). Further, it says that the set $J := J(j) := \{r : \exists\{k\}\ A^k(r, j) \ne 0\}$ is a proper subset of $\{1, \ldots, n\}$ (since it doesn't contain i), but neither is it empty (since it contains j, as $A^0(j, j) \ne 0$). Since $A^{k+\ell}(r, j) = \sum_m A^k(r, m)A^\ell(m, j)$, it follows that $J(m) \subset J(j)$ for all $m \in J(j)$. This implies, in particular, that $A(r, m) = 0$ for all $r \notin J(j), m \in J(j)$, hence that $\operatorname{span}(e_m)_{m\in J(j)}$ is a proper A-invariant subspace, thus implying not(iii). It also implies not(iv), since it shows that the columns $A(:, m), m \in J(j)$, have zero entries in rows $r, r \notin J(j)$, i.e., that (13.19) holds for the permutation $P = [(e_m)_{m\in J(j)}, (e_r)_{r\notin J(j)}]$, with both B and D of order $< n$.

Conversely, if e.g., (iii) does not hold, and $\operatorname{span}(e_m)_{m\in J(j)}$ is that proper A-invariant subspace, then it is also invariant under any $p(A)$, hence also $p(A)(r, m) = 0$ for every $r \notin J(j)$, $m \in J(j)$, i.e., (i) does not hold.

The final characterization is explicitly that given by Frobenius, – except that he did not formulate it in terms of graphs; that was done much later, by Rosenblatt (1957) and Varga (1962). Frobenius (???) observed that, since

$$A^k(i, j) = \sum_{j_1} \cdots \sum_{j_{k-1}} A(i, j_1) \cdots A(j_{k-1}, j),$$

therefore, for $i \ne j$, $A^k(i, j) \ne 0$ if and only if there exists some sequence $i =: i_0, i_1, \ldots, i_{k-1}, i_k := j$ so that $A(i_r, i_{r+1}) \ne 0$ for all r. Now, the **directed graph** of A is the graph with n vertices in which the directed edge (i, j) is present iff $A(i, j) \ne 0$. Such a graph is called **strongly connected** in case it contains, for each $i \ne j$, a path connecting vertex i with vertex j, and this, as we just observed, is equivalent to having $A^k(i, j) \ne 0$ for some $k > 0$. In short, (ii) and (v) are equivalent. □

There are various refinements of this last claim available. For example, in testing whether the directed graph of A is strongly connected, we only need to check paths involving distinct vertices, and such paths involve at most n vertices. Hence, in condition (ii), we need to check only for $k < n$. But, with that restriction, (ii) is equivalent to having $\mathrm{id}_n + A + \cdots + A^{n-1} > 0$ and, given that $A \ge 0$, this, in turn, is equivalent to having $(\mathrm{id}_n + A)^{n-1} > 0$, i.e., to having (i) hold for quite specific polynomials.

13.4 T/F

() If the sum $A + B$ of two matrices is defined, then $\det(A + B) = \det(A) + \det(B)$.

