vertex i to vertex j is the sum of the probabilities that we would have gone from i to some k in the first step and thence to j in the second step, i.e., the number

$$\sum_k M_{i,k} M_{k,j} = M^2_{i,j}.$$

More generally, the probability that we have gone after m steps from vertex i to vertex j is the number $M^m_{i,j}$, i.e., the (i,j)-entry of the mth power of the matrix M.

A study of the powers of such a stochastic matrix reveals that, for large m, all the rows of M^m look more and more alike. Precisely, for each row i,

$$\lim_{m\to\infty} M^m_{i:} = x_\infty$$

for a certain (i-independent) vector x_∞ with nonnegative entries that sum to one; this is part of the so-called Perron-Frobenius Theory. In terms of the random walk, this means that, for large m, the probability that we will be at vertex j after m steps is more or less independent of the vertex we started off from. One can find this limiting probability distribution x_∞ as a properly scaled eigenvector of the transpose M^{t} of M belonging to the eigenvalue 1.

As the simple example $M = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ shows, the last paragraph isn't quite correct. Look for the discussion of the Perron-Frobenius theorem later in these notes (see pages 134ff).

polynomials in a map: Once we know the powers A^k of A, we can also construct polynomials in A, in the following way. If p is the polynomial

$$p : t \mapsto c_0 + c_1 t + c_2 t^2 + \cdots + c_k t^k,$$

then we define the linear map $p(A)$ to be what we get when we substitute A for t:

$$p(A) := c_0 \,\mathrm{id} + c_1 A + c_2 A^2 + \cdots + c_k A^k.$$

We can even consider power series. The most important example is the **matrix exponential**:

$$(10.3) \qquad \exp(A) := \mathrm{id} + A + A^2/2 + A^3/6 + \cdots + A^k/k! + \cdots.$$

The matrix exponential is used in solving the first-order system

$$(10.4) \qquad Dy(t) = Ay(t) \text{ for } t > 0, \quad y(0) = b$$

of constant-coefficient ordinary differential equations. Here A is a square matrix, of order n say, and $y(t)$ is an n-vector that depends on t. Further,

$$Dy(t) := \lim_{h\to 0} (y(t+h) - y(t))/h$$

is the first derivative at t of the vector-valued function y. One verifies that the particular function

$$y(t) := \exp(tA)b, \quad t \ge 0,$$

solves the differential equation (10.4). Practical application does require efficient ways for evaluating the power series

$$\exp((tA)) := \mathrm{id} + tA + (tA)^2/2 + (tA)^3/6 + \cdots + (tA)^k/k! + \cdots,$$

hence for computing the powers of tA.

Eigenvalues and eigenvectors

The calculation of $A^k x$ is simplest if A maps x to a scalar multiple of itself, i.e., if

$$Ax = \mu x = x\mu$$

for some scalar μ. For, in that case, $A^2 x = A(Ax) = A(x\mu) = Ax\mu = x\,\mu^2$ and, more generally,

$$Ax = x\mu \quad \Longrightarrow \quad A^k x = x\,\mu^k, \quad k = 0, 1, 2, \ldots. \tag{10.5}$$

If $x = 0$, this will be so for any scalar μ. If $x \neq 0$, then this will be true for at most one scalar μ. That scalar is called an *eigenvalue for* A with associated *eigenvector* x.

(10.6) Definition: Let $A \in L(X)$. Any scalar μ for which there is a *nontrivial* vector $x \in X$ so that $Ax = x\mu$ is called an **eigenvalue** of A, with (μ, x) the corresponding **eigenpair**. The collection of all eigenvalues of A is called the **spectrum** of A and is denoted $\mathrm{spec}(A)$.
Thus

$$\mathrm{spec}(A) = \{\mu \in \mathbb{F} : \ A - \mu\,\mathrm{id} \ \text{ is not invertible}\}.$$

All the elements of $\mathrm{null}(A - \mu\,\mathrm{id})\backslash 0$ are called the **eigenvector**s of A associated with μ. The number

$$\rho(A) := \max|\,\mathrm{spec}(A)| = \max\{|\mu| : \mu \in \mathrm{spec}(A)\}$$

is called the **spectral radius of** A.

In the best of circumstances, there is an entire basis $V = [v_1, v_2, \ldots, v_n]$ for $X = \mathrm{dom}\,A$ consisting of eigenvectors for A. In this case, it is very easy to compute $A^k x$ for any $x \in X$. For, in this situation, $Av_j = v_j \mu_j$, $j = 1{:}n$, hence

$$AV = [Av_1, \ldots, Av_n] = [v_1\mu_1, \ldots, v_n\mu_n] = V\mathrm{M},$$

with M the *diagonal* matrix

$$\mathrm{M} := \mathrm{diag}(\mu_1, \mu_2, \ldots, \mu_n).$$

Therefore, for any k,

$$A^k V = V\mathrm{M}^k = V\,\mathrm{diag}(\mu_1^k, \ldots, \mu_n^k).$$

Also, since V is a basis for X, any $x \in X$ can be written (uniquely) as $x = Va$ for some n-vector a and thus

$$A^k x = A^k V a = V\mathrm{M}^k a = v_1\mu_1^k a_1 + v_2\mu_2^k a_2 + \cdots + v_n\mu_n^k a_n$$

for any k. For example, for such a matrix and for any t,

$$\exp(tA) = V\exp(t\mathrm{M})V^{-1} = V\,\mathrm{diag}(\ldots, \exp(t\mu_j), \ldots)V^{-1}.$$

To be sure, if A is not 1-1, then at least one of the μ_j must be zero, but this doesn't change the fact that M is a diagonal matrix.

(10.7) Example: The matrix $A := \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$ maps the 2-vector $x := (1, 1)$ to $3x$ and the 2-vector $y := (1, -1)$ to itself. Hence, $A[x, y] = [3x, y] = [x, y]\,\mathrm{diag}(3, 1)$ or

$$A = V\,\mathrm{diag}(3, 1)V^{-1}, \quad \text{with} \quad V := [x, y] = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}.$$

Elimination gives

$$[V,\ \mathrm{id}] = \begin{bmatrix} 1 & 1 & 1 & 0 \\ 1 & -1 & 0 & 1 \end{bmatrix} \to \begin{bmatrix} \mathbf{1} & 1 & 1 & 0 \\ 0 & -2 & -1 & 1 \end{bmatrix} \to \begin{bmatrix} 1 & 0 & 1/2 & 1/2 \\ 0 & \mathbf{-2} & -1 & 1 \end{bmatrix} \to \begin{bmatrix} 1 & 0 & 1/2 & 1/2 \\ 0 & 1 & 1/2 & -1/2 \end{bmatrix},$$

hence

$$V^{-1} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} /2.$$

It follows that, for any k,

$$A^k = V\,\mathrm{diag}(3^k, 1)V^{-1} = \begin{bmatrix} 3^k & 1 \\ 3^k & -1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} /2 = \begin{bmatrix} 3^k+1 & 3^k-1 \\ 3^k-1 & 3^k+1 \end{bmatrix} /2.$$

In particular,

$$A^{-1} = \begin{bmatrix} 1/3+1 & 1/3-1 \\ 1/3-1 & 1/3+1 \end{bmatrix} /2 = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} /3.$$

Also,

$$\exp(tA) = V\,\mathrm{diag}(\mathrm{e}^{3t}, \mathrm{e}^{t})V^{-1} = \begin{bmatrix} \mathrm{e}^{3t}+\mathrm{e}^{t} & \mathrm{e}^{3t}-\mathrm{e}^{t} \\ \mathrm{e}^{3t}-\mathrm{e}^{t} & \mathrm{e}^{3t}+\mathrm{e}^{t} \end{bmatrix}.$$

□

10.1 Let $A = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}$. (i) Find a basis V and a diagonal matrix M so that $A = V\mathrm{M}V^{-1}$. (ii) Determine the matrix $\exp(A)$.

10.2 Let $A = \begin{bmatrix} 4 & 1 & -1 \\ 2 & 5 & -2 \\ 1 & 1 & 2 \end{bmatrix}$.

Use elimination to determine *all* eigenvectors for this A belonging to the eigenvalue 3, and all eigenvectors belonging to the eigenvalue 5. (It is sufficient to give a basis for $\mathrm{null}(A - 3\,\mathrm{id})$ and for $\mathrm{null}(A - 5\,\mathrm{id})$.)

10.3

(a) Prove that the matrix $A = \begin{bmatrix} 4 & 1 & -1 \\ 2 & 5 & -2 \\ 1 & 1 & 2 \end{bmatrix}$ maps the vector space $Y := \mathrm{ran}\,V$ with $V := \begin{bmatrix} 0 & 2 \\ 3 & 1 \\ 1 & 1 \end{bmatrix}$ into itself, hence the **restriction** of A to Y, i.e.,

$$A_{|Y} := B : Y \to Y : y \mapsto Ay$$

is a well-defined linear map. (You will have to verify that $\mathrm{ran}\,AV \subseteq \mathrm{ran}\,V$; looking at the rref of $[V\ AV]$ should help.)

(b) Determine the matrix representation of B with respect to the basis V for $\mathrm{dom}\,B = Y$, i.e., compute the matrix $V^{-1}BV$. (Hint: (5.4)Example tells you how to read off this matrix from the calculations in (a).)

(c) Determine the spectrum of the linear map $B = A_{|Y}$ defined in (a). (Your answer in (b) could be helpful here since similar maps have the same spectrum.)

10.4 Prove that 0 is the only eigenvalue of the matrix $A = \begin{bmatrix} 0 & 1 & 2 \\ 0 & 0 & 3 \\ 0 & 0 & 0 \end{bmatrix}$ and that, up to scalar multiples, e_1 is the only eigenvector for A.

10.5 Let $\mu \in \mathrm{spec}(A)$ (hence $Ax = \mu x$ for some $x \neq 0$). Prove:

(i) For any scalar α, $\alpha\mu \in \mathrm{spec}(\alpha A)$.

(ii) For any scalar α, $\mu + \alpha \in \mathrm{spec}(A + \alpha\,\mathrm{id})$.

(iii) For any natural number k, $\mu^k \in \mathrm{spec}(A^k)$.

(iv) If A is invertible, then $\mu \neq 0$ and $\mu^{-1} \in \mathrm{spec}(A^{-1})$.

(v) If A is a matrix, then $\mu \in \mathrm{spec}(A^{\mathrm{t}})$ and $\overline{\mu} \in \mathrm{spec}(A^{\mathrm{c}})$.

Diagona(liza)bility

Definition: A linear map $A \in L(X)$ is called **diagona(liza)ble** if it has an **eigenbasis**, i.e., if there is a basis for its domain X consisting entirely of eigenvectors for A.

(10.8) Lemma: If V_μ is a basis for $\mathrm{null}(A - \mu\,\mathrm{id})$, then $[V_\mu : \mu \in \mathrm{spec}(A)]$ is 1-1.

Proof: Note that, for any $\mu \in \mathrm{spec}(A)$ and any ν,

$$(A - \nu\,\mathrm{id})V_\mu = (\mu - \nu)V_\mu,$$

and, in particular, $(A - \mu\,\mathrm{id})V_\mu = 0$. Hence, if $\sum_\mu V_\mu a_\mu = 0$, then, for each $\mu \in \mathrm{spec}(A)$, after applying to both sides of this equation the product of all $(A - \nu\,\mathrm{id})$ with $\nu \in \mathrm{spec}(A)\backslash\mu$, we are left with the equation $(\prod_{\nu\neq\mu}(\mu - \nu))V_\mu a_\mu = 0$, and this implies that $a_\mu = 0$ since V_μ is 1-1 by assumption. In short, $[V_\mu : \mu \in \mathrm{spec}(A)]a = 0$ implies $a = 0$. □

(10.9) Corollary: $\#\mathrm{spec}(A) \le \dim\mathrm{dom}\,A$, with equality only if A is diagonalizable.

(10.10) Proposition: A linear map $A \in L(X)$ is diagonalizable if and only if

$$\dim X = \sum_{\mu\in\mathrm{spec}(A)} \dim\mathrm{null}(A - \mu\,\mathrm{id}). \qquad (10.11)$$

Proof: By (10.8)Lemma, (10.11) implies that $\mathrm{dom}\,A$ has a basis consisting of eigenvectors for A.

Conversely, if V is a basis for $X = \mathrm{dom}\,A$ consisting entirely of eigenvectors for A, then $A = V\mathrm{M}V^{-1}$ for some diagonal matrix $\mathrm{M} =: \mathrm{diag}(\mu_1, \ldots, \mu_n)$, hence, for any scalar μ, $(A - \mu\,\mathrm{id}) = V(\mathrm{M} - \mu\,\mathrm{id})V^{-1}$. In particular, $\mathrm{null}(A - \mu\,\mathrm{id}) = \mathrm{ran}[v_j : \mu = \mu_j]$, hence $\sum_{\mu\in\mathrm{spec}(A)} \dim\mathrm{null}(A - \mu\,\mathrm{id}) = \sum_{\mu\in\mathrm{spec}(A)} \#\{j : \mu_j = \mu\} = n = \#V = \dim X$. □

(10.10)Proposition readily identifies a circumstance under which A is *not* diagonable, namely when $\mathrm{null}(A-\mu\,\mathrm{id})\cap\mathrm{ran}(A-\mu\,\mathrm{id}) \neq \{0\}$ for some μ. For, with V_ν a basis for $\mathrm{null}(A-\nu\,\mathrm{id})$ for any $\nu \in \mathrm{spec}(A)$, we compute $AV_\nu = \nu V_\nu$, hence $(A-\mu\,\mathrm{id})V_\nu = (\nu-\mu)V_\nu$ and therefore, for any $\nu \neq \mu$, $V_\nu = (A-\mu\,\mathrm{id})V_\nu/(\nu-\mu) \subset \mathrm{ran}(A - \mu\,\mathrm{id})$. This places all the columns of the 1-1 map $V_{\backslash\mu} := [V_\nu : \nu \neq \mu]$ in $\mathrm{ran}(A - \mu\,\mathrm{id})$ while, by (10.8)Lemma, $\mathrm{ran}\,V_\mu \cap \mathrm{ran}\,V_{\backslash\mu}$ is trivial. Hence, if $\mathrm{ran}\,V_\mu = \mathrm{null}(A - \mu\,\mathrm{id})$ has nontrivial intersection with $\mathrm{ran}(A - \mu\,\mathrm{id})$, then $\mathrm{ran}\,V_{\backslash\mu}$ cannot be all of $\mathrm{ran}(A - \mu\,\mathrm{id})$, and therefore

$$\sum_{\nu\neq\mu} \dim\mathrm{null}(A - \nu\,\mathrm{id}) = \#V_{\backslash\mu} \quad < \quad \dim\mathrm{ran}(A - \mu\,\mathrm{id}) = \dim X - \dim\mathrm{null}(A - \mu\,\mathrm{id}),$$

hence, by (10.10)Proposition, such A is not diagonable.

This has motivated the following

Definition: The scalar μ is a **defective eigenvalue** of A if

$$\operatorname{null}(A - \mu\,\mathrm{id}) \cap \operatorname{ran}(A - \mu\,\mathrm{id}) \neq \{0\}.$$

Any such μ certainly is an eigenvalue (since, in particular, $\operatorname{null}(A-\mu\,\mathrm{id}) \neq \{0\}$), but I don't care for such *negative labeling*; if it were up to me, I would call such μ an **interesting eigenvalue**, since the existence of such eigenvalues makes for a richer theory. Note that, by (4.18)Proposition, μ is a defective eigenvalue for A iff, for any bases V and W for $\operatorname{ran}(A - \mu\,\mathrm{id})$ and $\operatorname{null}(A - \mu\,\mathrm{id})$ respectively, $[V, W]$ is not 1-1.

(10.12) Corollary: If A has a defective eigenvalue, then A is not diagonable.

10.6 Prove: if $A \in L(X)$ is diagonalizable and $\#\operatorname{spec}(A) = 1$, then $A = \mu\,\mathrm{id}_X$ for some $\mu \in \mathbb{F}$.

10.7 What is a simplest matrix A with $\operatorname{spec}(A) = \{1, 2, 3\}$?

10.8 For each of the following matrices $A \in \mathbb{F}^{2\times 2}$, determine whether or not 0 is a defective eigenvalue (give a reason for your answer). (a) $A = 0$. (b) $A = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}$. (c) $A = \begin{bmatrix} -2 & -1 \\ 4 & 2 \end{bmatrix}$. (d) $A = \mathrm{id}_2$.

10.9 Prove that any linear projector P on a finite-dimensional vector space X is diagonalizable. (Hint: Show that, for any basis U for $\operatorname{ran} P$ and any basis W for $\operatorname{null} P$, $V := [U, W]$ is a basis for X, and that all the columns of V are eigenvectors for P. All of this should follow from the fact that $P^2 = P$.)

Similarity is an equivalence relation

Definition: We say that $A \in L(X)$ and $B \in L(Y)$ are **similar to each other** and write

$$A \sim B$$

in case there is an invertible $V \in L(Y, X)$ so that

$$A = VBV^{-1}.$$

In particular, *a linear map is diagonable if and only if it is similar to a diagonal matrix.*

In trying to decide whether or not a given linear map A is diagonable, it is sufficient to decide this question for any convenient linear map B similar to A. For, if such a B is diagonable, i.e., similar to a diagonal matrix, then A is similar to that very same diagonal matrix. This follows from the fact that similarity is an equivalence relation:

(10.13) Proposition: Similarity is an **equivalence relation**. Specifically,

(i) $A \sim A$ (**reflexive**);

(ii) $A \sim B$ implies $B \sim A$ (**symmetric**);

(iii) $A \sim B$ and $B \sim C$ implies $A \sim C$ (**transitive**).

Proof: Certainly, $A \sim A$, since $A = \mathrm{id} A \,\mathrm{id}$. Also, if $A = VBV^{-1}$ for some invertible V, then also $W := V^{-1}$ is invertible, and $B = V^{-1}AV = WAW^{-1}$. Finally, if $A = VBV^{-1}$ and $B = WCW^{-1}$, then $U := VW$ is also invertible, and $A = V(WCW^{-1})V^{-1} = UCU^{-1}$. □

Now, any linear map $A \in L(X)$ on a *finite-dimensional* vector space X is similar (in many ways if X is not trivial) to a *matrix*. Indeed, for any basis V for X, $\widehat{A} := V^{-1}AV$ is a matrix similar to A. The map $\widehat{A}$ so defined is indeed a matrix since both its domain and its target is a coordinate space (the same one, in fact; hence $\widehat{A}$ is a *square* matrix). We conclude that, in looking for ways to decide whether or not a linear map is diagonable, it is sufficient to know how to do this for square *matrices*.

Are all square matrices diagonable?

By (10.12)Corollary, this will be so only if all square matrices have only nondefective eigenvalues.

(10.14) Example: The simplest example of a matrix with a defective eigenvalue is provided by the matrix

$$N := \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} = [0, e_1].$$

Since this matrix is triangular, so is $N - \mu\,\mathrm{id}$ for any μ, and as $-\mu$ are the latter's diagonal entries, $N - \mu\,\mathrm{id}$ fails to be invertible if and only if $\mu = 0$. Hence, $\mathrm{spec}(N) = \{0\}$. Yet $\mathrm{null}\, N = \mathrm{ran}[e_1] = \mathrm{ran}\, N$, hence the only eigenvalue of N is defective, and N fails to be diagonable, by (10.12)Corollary.

Of course, for this simple matrix, one can see directly that it cannot be diagonable, since, if it were, then some basis V for $\mathbb{R}^2$ would consist entirely of eigenvectors for the sole eigenvalue, 0, for N, hence, for this basis, $NV = 0$, therefore $N = 0$, contrary to fact. □

We will see shortly that, on a finite-dimensional vector space over the complex scalars, almost all linear maps are diagonable, and all linear maps are almost diagonable.

Does every square matrix have an eigenvalue?

Since an eigenvalue for A is any *scalar* μ for which $\mathrm{null}(A - \mu\,\mathrm{id})$ is not trivial, the answer necessarily depends on what we mean by a scalar.

If we only allow *real* scalars, i.e., if $\mathbb{F} = \mathbb{R}$, then not every matrix has eigenvalues. The simplest example is a rotation of the plane, e.g., the matrix

$$A := \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} = [e_2, -e_1].$$

This linear map rotates every $x \in \mathbb{R}^2$ 90 degrees counter-clockwise, hence the only vector x mapped by it to a scalar multiple of itself is the zero vector. In other words, this linear map has no eigenvectors, hence no eigenvalues.

The situation is different when we also allow *complex* scalars, i.e., when $\mathbb{F} = \mathbb{C}$, and this is the reason why we considered complex scalars all along in these notes. Now every (square) matrix has eigenvalues, as follows from the following simple argument.

(10.15) Theorem: Any linear map A on some nontrivial finite-dimensional vector space X over the *complex* scalar field $\mathbb{F} = \mathbb{C}$ has eigenvalues.

Proof: Let $n := \dim X$, pick any $x \in X\backslash 0$ and consider the column map

$$V := [x, Ax, A^2x, \ldots, A^n x].$$

Since $\#V > \dim \operatorname{tar} V$, V cannot be 1-1. This implies that some column of V is free. Let $A^d x$ be the first free column, i.e., the first column that is in the range of the columns preceding it. Then $\operatorname{null} V$ contains exactly one vector of the form

$$a = (a_0, a_1, \ldots, a_{d-1}, 1, 0, \cdots, 0),$$

and this is the vector we choose. Then, writing the equation $Va = 0$ out in full, we get

$$a_0 x + a_1 A x + \cdots + a_{d-1} A^{d-1} x + A^d x \; = \; 0. \tag{10.16}$$

Now here comes the trick: Consider the *polynomial*

$$p : t \mapsto a_0 + a_1 t + \cdots + a_{d-1} t^{d-1} + t^d. \tag{10.17}$$

Then, substituting for t our map A, we get the linear map

$$p(A) := a_0 \operatorname{id} + a_1 A + \cdots + a_{d-1} A^{d-1} + A^d.$$

With this, (10.16) can be written, very concisely,

$$p(A)x = 0.$$

This is not just notational convenience. Since $a_d = 1$, p isn't the zero polynomial, and since $x \neq 0$, d must be greater than 0, i.e., p cannot be just a constant polynomial. Thus, by the *Fundamental Theorem of Algebra*, p has zeros. More precisely,

$$p(t) = (t - z_1)(t - z_2) \cdots (t - z_d)$$

for certain (possibly *complex*) scalars $z_1, \ldots, z_d$. This implies (see (10.19)Lemma below) that

$$p(A) = (A - z_1 \operatorname{id})(A - z_2 \operatorname{id}) \cdots (A - z_d \operatorname{id}).$$

Now, $p(A)$ is not 1-1 since it maps the nonzero vector x to zero. Therefore, *not all the maps* $(A - z_j \operatorname{id})$, $j = 1{:}d$, *can be 1-1*. In other words, for some j, $(A - z_j \operatorname{id})$ fails to be 1-1, i.e., has a nontrivial nullspace, and that makes z_j an eigenvalue for A. □

(10.18) Example: Let's try this out on our earlier example, the rotation matrix

$$A := [e_2, -e_1].$$

Choosing $x = e_1$, we have

$$[x, Ax, A^2 x] = [e_1, e_2, -e_1],$$

hence the first free column is $A^2 x = -e_1$, and, by inspection,

$$x + A^2 x = 0.$$

Thus the polynomial of interest is

$$p : t \mapsto 1 + t^2 = (t - \mathrm{i})(t + \mathrm{i}),$$

with

$$\mathrm{i} := \sqrt{-1}$$

the *imaginary unit* (see the Backgrounder on complex numbers). In fact, we conclude that, with $y := (A + \mathrm{i} \operatorname{id})x$, $(A - \mathrm{i} \operatorname{id})y = p(A)x = 0$, while $y = Ae_1 + \mathrm{i}e_1 = e_2 + \mathrm{i}e_1 \neq 0$, showing that $(\mathrm{i}, e_2 + \mathrm{i}e_1)$ is an eigenpair for this A.

Polynomials in a linear map

The proof of (10.15)Theorem uses in an essential way the following fact.

(10.19) Lemma: If r is the product of the polynomials p and q, i.e., $r(t) = p(t)q(t)$ for all t, then, for any linear map $A \in L(X)$,

$$r(A) = p(A)q(A) = q(A)p(A).$$

Proof: If you wanted to check that $r(t) = p(t)q(t)$ for the polynomials r, p, q, then you would multiply p and q term by term, collect like terms and then compare coefficients with those of r. For example, if $p(t) = t^2 + t + 1$ and $q(t) = t - 1$, then

$$p(t)q(t) = (t^2 + t + 1)(t - 1) = t^2(t - 1) + t(t - 1) + (t - 1) = t^3 - t^2 + t^2 - t + t - 1 = t^3 - 1,$$

i.e., the product of these two polynomials is the polynomial r given by $r(t) = t^3 - 1$. The only facts you use are: (i) free reordering of terms (commutativity of addition), and (ii) things like $tt = t^2$, i.e., the fact that

$$t^i t^j = t^{i+j}.$$

Both of these facts hold if we replace t by A. □

Here is a further use of this lemma. We now prove that the polynomial p constructed in the proof of (10.15) has the property that every one of its roots is an eigenvalue for A. This is due to the fact that we constructed it in the form (10.17) with d the *smallest* integer for which $A^d x \in \operatorname{ran}[x, Ax, \ldots, A^{d-1}x]$. Thus, with μ any zero of p, we can write

$$p(t) = (t - \mu)q(t) \qquad (10.20)$$

for some polynomial q necessarily of the form

$$q(t) = b_0 + b_1 t + \cdots + b_{d-2}t^{d-2} + t^{d-1}.$$

The crucial point here is that q is of degree $< d$. This implies that $q(A)x \neq 0$ since, otherwise, $(b_0, b_1, \ldots, 1)$ would be a nontrivial vector in $\operatorname{null}[x, Ax, \ldots, A^{d-1}x]$ and this would contradict the choice of d as the index for which $A^d x$ is the *first* free column in $[x, Ax, A^2, \ldots]$. Since

$$0 = p(A)x = (A - \mu\,\mathrm{id})q(A)x,$$

it follows that μ is an eigenvalue for A with associated eigenvector $q(A)x$.

This is exactly how we got an eigenvector for the eigenvalue i in (10.18)Example.

(10.21) Example: As another example, consider again the matrix $A = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$ from (10.7)Example. We choose $x = e_1$ and consider

$$[x, Ax, \ldots, A^n x] = [e_1, Ae_1, A(Ae_1)] = \begin{bmatrix} 1 & 2 & 5 \\ 0 & 1 & 4 \end{bmatrix}.$$

Since $[e_1, Ae_1, A^2 e_1]$ is in row echelon form, we conclude that the first two columns are bound. Elimination gives the rref

$$\begin{bmatrix} 1 & 0 & -3 \\ 0 & 1 & 4 \end{bmatrix},$$

hence $(3,-4,1) \in \operatorname{null}[e_1, Ae_1, A^2e_1]$. Correspondingly, $p(A)e_1 = 0$, with

$$p(t) = 3 - 4t + t^2 = (t-3)(t-1).$$

Consequently, $\mu = 3$ is an eigenvalue for A, with corresponding eigenvector

$$(A - \mathrm{id})e_1 = (1,1);$$

also, $\mu = 1$ is an eigenvalue for A, with corresponding eigenvector

$$(A - 3\,\mathrm{id})e_1 = (-1,1).$$

Note that the resulting basis $\begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}$ for $\mathbb{F}^2$ consisting of eigenvectors for A differs in some detail from the one we found in (10.7)Example. After all, if v is an eigenvector, then so is αv for any scalar α. □

Here is some standard language concerning the items in our discussion so far. One calls any nontrivial polynomial r for which $r(A)x = 0$ an **annihilating polynomial for** A **at** x. We may assume without loss of generality that this polynomial is **monic**, i.e., its highest nonzero coefficient is 1, since we can always achieve this by dividing the polynomial by its highest nonzero coefficient without changing the fact that it is an annihilating polynomial for A at x. If such a polynomial is of exact degree k, say, then it has the form

$$r(t) = b_0 + b_1 t + \cdots + b_{k-1}t^{k-1} + t^k.$$

Since $r(A)x = 0$, we conclude that

$$b_0 x + b_1 Ax + \cdots + b_{k-1}A^{k-1}x + A^k x = 0.$$

In particular, $A^k x$ is in $\operatorname{ran}[x, Ax, \ldots, A^{k-1}x]$, i.e., the column $A^k x$ of $[x, Ax, A^2x, \ldots]$ is free. This implies that $k \ge d$, with d the degree of the polynomial p constructed in the proof of (10.15)Theorem. For, there we chose d as the *smallest* index for which $A^d x$ is a free column of $[x, Ax, A^2, \ldots]$. In particular, all prior columns must be bound. This makes p the unique monic polynomial of smallest degree for which $p(A)x = 0$.

Here, for the record, is a formal account of what we have proved.

(10.22) Proposition: For every $A \in L(X)$ with $\dim X < \infty$ and every $x \in X\backslash 0$, there is a unique monic polynomial p of smallest degree for which $p(A)x = 0$. This polynomial is called the **minimal polynomial for** A **at** x and is denoted

$$p_{A,x}.$$

It can be constructed in the form

$$p_{A,x}(t) = a_0 + a_1 t + \cdots + a_{d-1}t^{d-1} + t^d,$$

with d the smallest index for which $A^d x$ is a free column of $[x, Ax, A^2x, \ldots]$. Moreover, $(a_0, a_1, \ldots, 1)$ is the unique vector in $\operatorname{null}[x, Ax, \ldots, A^d x]$ with its last entry equal to 1.

Assuming that X is a vector space over $\mathbb{F} = \mathbb{C}$, every zero μ of $p_{A,x}$ is an eigenvalue of A, with associated eigenvector $q(A)x$, where $p_{A,x}(t) =: (t-\mu)q(t)$. (See the Backgrounder on Horner's method for the standard way to compute q from $p_{A,x}$ and μ.)

For *example*, consider the *permutation matrix* $P = [e_2, e_3, e_1]$ and take $x = e_1$. Then

$$[x, Px, P^2x, P^3x] = [e_1, e_2, e_3, e_1].$$

Hence, P^3x is the first free column here. The element in the nullspace corresponding to it is the vector $(-1,0,0,1)$. Hence, the minimal polynomial for P at $x = e_1$ is of degree 3; it is the polynomial $p(t) = t^3 - 1$. It has the zero $\mu = 1$, which therefore is an eigenvalue of P. A corresponding eigenvector is obtainable in the form $q(P)e_1$ with $q(t) := p(t)/(t-1) = t^2 + t + 1$, hence the eigenvector is $e_3 + e_2 + e_1$.

10.10 Use Elimination as in (10.21) to determine all the eigenvalues and, for each eigenvalue, a corresponding eigenvector, for each of the following matrices: (i) $\begin{bmatrix} 7 & -4 \\ 5 & -2 \end{bmatrix}$; (b) $[0, e_1, e_2] \in \mathbb{R}^{3\times 3}$ (try $x = e_3$); (iii) $\begin{bmatrix} -1 & 1 & -3 \\ 20 & 5 & 10 \\ 2 & -2 & 6 \end{bmatrix}$.

10.11

(a) Prove: If p is any nontrivial polynomial and A is any square matrix for which $p(A) = 0$, then $\operatorname{spec}(A) \subseteq \{\mu \in \mathbb{C} : p(\mu) = 0\}$. (Hint: prove first that, for any eigenvector x for A with eigenvalue μ and any polynomial p, $p(A)x = p(\mu)x$.)

(b) What can you conclude about $\operatorname{spec}(A)$ in case you know that A is *idempotent*, i.e., a linear projector, i.e., $A^2 = A$?

(c) What can you conclude about $\operatorname{spec}(A)$ in case you know that A is **nilpotent**, i.e., $A^q = 0$ for some integer q?

(d) What is the spectrum of the linear map $D : \Pi_k \to \Pi_k$ of differentiation, as a map on polynomials of degree $\le k$?

10.12 Use the minimal polynomial at e_1 to determine the spectrum of the following matrices: (i) $[e_2, 0]$; (ii) $[e_2, e_3, e_1]$; (iii) $[e_2, e_2]$; (iv) $[e_2, e_1, 2e_3]$.

10.13 Prove: (i) for any $A, B \in L(X)$, $\operatorname{null} A \cap \operatorname{null} B \subset \operatorname{null}(A + B)$. (ii) for any $A, B \in L(X)$ with $AB = BA$, $\operatorname{null} A + \operatorname{null} B \subset \operatorname{null}(AB)$. (iii) If d is the greatest common divisor of the nontrivial polynomials $p_1, \ldots, p_r$ and m is their smallest common multiple, then, for any $A \in L(X)$, $\operatorname{null} d(A) = \cap_j p_j(A)$ and $\operatorname{null} m(A) = \sum_j \operatorname{null} p_j(A)$.

10.14 A subset F of the vector space $X := C^{(1)}(\mathbb{R})$ of continuously differentiable functions is called D-**invariant** if the derivative Df of any $f \in F$ is again in F.

Prove: Any finite-dimensional D-invariant linear subspace Y of $C^{(1)}(\mathbb{R})$ is necessarily the nullspace of a constant-coefficient ordinary differential operator.

Every complex (square) matrix is similar to an upper triangular matrix

While having in hand a diagonal matrix similar to a given $A \in L(X)$ is very nice indeed, for most purposes it is sufficient to have in hand an *upper triangular* matrix similar to A. There are several reasons for this.

One reason is that, as soon as we have an upper triangular matrix similar to A, then we can easily manufacture from this a matrix similar to A and with off-diagonal elements as small as we please (except that, in general, we can't make them all zero).

A more fundamental reason is that, once we have an upper triangular matrix similar to A, then we know the entire spectrum of A, since, by (3.19)Proposition, a triangular matrix is noninvertible iff some diagonal entry is zero, hence the spectrum of an upper triangular matrix is the set of its diagonal entries, and similar matrices have the same spectrum. Here are the various facts.

(10.23) Proposition: If A and $\widehat{A}$ are similar, then $\operatorname{spec}(A) = \operatorname{spec}(\widehat{A})$.

Proof: If $\widehat{A} = V^{-1}AV$ for some invertible V, then, for any scalar μ, $\widehat{A} - \mu\,\mathrm{id} = V^{-1}(A - \mu\,\mathrm{id})V$. In particular, $\widehat{A} - \mu\,\mathrm{id}$ is not invertible (i.e., $\mu \in \operatorname{spec}(\widehat{A})$) if and only if $A - \mu\,\mathrm{id}$ is not invertible (i.e., $\mu \in \operatorname{spec}(A)$). □

(10.24) Corollary: If $A \in L(X)$ is similar to a triangular matrix $\widehat{A}$, then μ is an eigenvalue for A if and only if $\mu = \widehat{A}_{j,j}$ for some j. In a formula,

$$\operatorname{spec}(A) = \{\widehat{A}_{j,j} : \text{ all } j\}.$$

More precisely, if $\widehat{A} = V^{-1}AV$ is upper triangular and j is the smallest index for which $\mu = \widehat{A}_{j,j}$, then there is an eigenvector for A belonging to μ available in the form $w = Va$, with a the element in the standard basis for $\operatorname{null}(\widehat{A} - \mu \operatorname{id})$ associated with the (free) jth column, i.e., $a \in \operatorname{null}(\widehat{A} - \mu \operatorname{id})$, $a_j = 1$, and all other entries corresponding to free columns are 0; cf. (3.9).

The now-standard algorithm for computing the eigenvalues of a given matrix A is the **QR method**. It generates a sequence $B_1, B_2, B_3, \ldots$ of matrices all similar to A that converges to an upper triangular matrix. To the extent that the lower triangular entries of B_k are small (compared to $\|B_k\|$, say), the diagonal entries of B_k are close to eigenvalues of B_k, hence of A. The actual version of the QR method used in `MATLAB` is quite sophisticated, as much care has gone into making the algorithm reliable in the presence of round-off as well as fast.

The `MATLAB` command `eig(`A`)` gives you the list of eigenvalues of A. The more elaborate command `[V,M]=eig(`A`)` gives you, in `V`, a list of corresponding ‘eigenvectors’, in the sense that, approximately, $A\mathtt{V}(:,j) = \mathtt{V}(:,j)\mathtt{M}(j,j)$, all j.

□

(10.25) Theorem: Every complex (square) matrix is similar to an upper triangular matrix.

Proof: The proof is by induction on the order, n, of the given matrix A.

If $n = 1$, then A is a 1×1-matrix, hence trivially upper triangular. Assume that we have proved the theorem for all matrices of order $n-1$ and let A be of order n. Since the scalar field is $\mathbb{C}$, we know that A has an eigenvector, u_1, say, with corresponding eigenvalue, μ_1 say. Extend u_1 to a basis $U = [u_1, u_2, \ldots, u_n]$ for $\mathbb{C}^n$. Then

$$AU = [Au_1, \cdots, Au_n] = [u_1\mu_1, Au_2, \cdots, Au_n].$$

We want to compute $U^{-1}AU$. For this, observe that $U^{-1}u_1 = U^{-1}Ue_1 = e_1$. Therefore,

$$U^{-1}AU = [e_1\mu_1, U^{-1}Au_2, \cdots, U^{-1}Au_n].$$

Writing this out in detail, we have

$$U^{-1}AU = \widehat{A} := \begin{bmatrix} \mu_1 & \times & \cdots & \times \\ 0 & \times & \cdots & \times \\ \vdots & \vdots & \ldots & \vdots \\ 0 & \times & \cdots & \times \end{bmatrix} =: \begin{bmatrix} \mu_1 & C \\ 0 & A_1 \end{bmatrix}.$$

Here, C is some $1 \times (n-1)$ matrix of no further interest, A_1 is a matrix of order $n-1$, hence, by induction hypothesis, there is some invertible W so that $\widehat{A}_1 := W^{-1}A_1W$ is upper triangular. We compute

$$\operatorname{diag}(1, W^{-1})\widehat{A}\operatorname{diag}(1, W) = \begin{bmatrix} 1 & 0 \\ 0 & W^{-1} \end{bmatrix} \begin{bmatrix} \mu_1 & C \\ 0 & A_1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & W \end{bmatrix} = \begin{bmatrix} \mu_1 & CW \\ 0 & W^{-1}A_1W \end{bmatrix}.$$

The computation uses the fact that multiplication from the left (right) by a block-diagonal matrix multiplies the corresponding rows (columns) from the left (right) by the corresponding diagonal blocks. Since $\operatorname{diag}(1, W^{-1})\operatorname{diag}(1, W) = \operatorname{diag}(1, \operatorname{id}_{n-1}) = \operatorname{id}_n$, this shows that $\widehat{A}$ is similar to an upper triangular matrix. Since A is similar to $\widehat{A}$, this finishes the proof. □

Various refinements in this proof are possible (as we will show later, in the discussion of the Schur form), to give more precise information about possible upper triangular matrices similar to a given A. For the present, though, this is sufficient for our needs since it allows us to prove the following:

(10.26) Corollary: Every complex (square) matrix is similar to an 'almost diagonal' matrix. Precisely, for every complex matrix A and every $\varepsilon > 0$, there exists an upper triangular matrix B_ε similar to A whose off-diagonal entries are all $< \varepsilon$ in absolute value.

Proof: By (10.25)Theorem, we know that any such A is similar to an upper triangular matrix. Since similarity is transitive (see (10.13)Proposition), it is therefore sufficient to prove this Corollary in case A is upper triangular, of order n, say.

The proof in this case is a simple trick: Consider the matrix

$$B := W^{-1}AW,$$

with

$$W := \mathrm{diag}(\delta^1, \delta^2, \ldots, \delta^n),$$

and the *scalar* δ to be set in a moment. W is indeed invertible as long as $\delta \neq 0$, since then

$$W^{-1} = \mathrm{diag}(\delta^{-1}, \delta^{-2}, \ldots, \delta^{-n}).$$

Now, multiplying a matrix by a diagonal matrix from the *left* (*right*) multiplies the *rows* (*columns*) of that matrix by the diagonal entries of the diagonal matrix. Therefore,

$$B_{i,j} = (W^{-1}AW)_{i,j} = A_{i,j}\delta^{j-i}, \quad \text{all } i, j.$$

In particular, B is again upper triangular, and its diagonal entries are those of A. However, all its possibly nontrivial off-diagonal entries lie above the diagonal, i.e., are entries $B_{i,j}$ with $j > i$, hence are the corresponding entries of A multiplied with some *positive* power of the scalar δ. Thus, if

$$c := \max_{i<j} |A_{i,j}|$$

and we choose $\delta := \min\{\varepsilon/c, 1\}$, then, we can be certain that

$$|B_{i,j}| \leq \varepsilon, \qquad \text{all } i \neq j,$$

regardless of how small we choose that positive ε. □

10.15 T/F

(a) The only diagonalizable matrix A having just one factorization $A = VMV^{-1}$ with M diagonal is the empty matrix.

(b) If A is the linear map of multiplication by a scalar, then any basis for its domain is an eigenbasis for A.

(c) A triangular matrix of order n is diagonalizable if and only if it has n different diagonal entries.

(d) Any (square) triangular matrix is diagonalizable.

(e) Any matrix of order 1 is diagonalizable.

(f) A matrix of order n has n eigenvalues.

(g) Similar linear maps have the same spectrum.

(h) The linear map of differentiation on Π_k is nilpotent.

(i) The identity map is idempotent.

(j) If the matrix A has 3 eigenvalues, then it must have at least 3 columns.

(k) If $\mathrm{null}(A - \mu\,\mathrm{id})$ is not trivial, then every one of its elements is an eigenvector for A belonging to the eigenvalue μ.

