11. Convergence of the power sequence

Convergence of sequences in a normed vector space

Our discussion of the power sequence $A^0, A^1, A^2, \cdots$ of a linear map naturally involves the *convergence* of such a sequence.

Convergence of a vector sequence or a map sequence is most conveniently described with the aid of a norm, as introduced earlier, starting at page 78.

Suppose $z_1, z_2, z_3, \cdots$ is an infinite sequence of n-vectors. Recall our agreement to refer to the jth entry of the kth term z_k in such a vector sequence by $z_k(j)$. We say that this sequence **converges to the n-vector** z_∞ and write

$$z_\infty = \lim_{k\to\infty} z_k,$$

in case

$$\lim_{k\to\infty} \|z_\infty - z_k\| = 0.$$

It is not hard to see that

$$z_\infty = \lim_{k\to\infty} z_k \quad\iff\quad \forall\{i\}\ z_\infty(i) = \lim_{k\to\infty} z_k(i).$$

Note that $z_\infty = \lim_{k\to\infty} z_k$ if and only if, for every $\varepsilon > 0$, there is some k_0 so that, for all $k > k_0$, $\|z_\infty - z_k\| < \varepsilon$. This says that, for any given $\varepsilon > 0$ however small, all the terms in the sequence from a certain point on lie in the "ball"

$$B_\varepsilon(z_\infty) := \{y \in \mathbb{F}^n : \|y - z_\infty\| < \varepsilon\}$$

whose center is z_∞ and whose radius is ε.

(11.1) Lemma: A convergent sequence is necessarily bounded. More explicitly, if the sequence (x_k) of n-vectors converges, then $\sup_k \|x_k\| < \infty$, i.e., there is some c so that, for all k, $\|x_k\| \le c$.

The proof is a verbatim repeat of the proof of this assertion for scalar sequences, as given in the Backgrounder on scalar sequences.

Analogously, we say that the sequence $A_1, A_2, A_3, \cdots$ of matrices **converges** to the matrix B and write

$$\lim_{k\to\infty} A_k = B,$$

in case

$$\lim_{k\to\infty} \|B - A_k\|_\infty = 0.$$

As in the case of vector sequences, a convergent sequence of matrices is necessarily bounded.

Here, for convenience, we have used the map norm associated with the max-norm since we have the simple and explicit formula (7.16) for it. Yet we know from (7.24)Proposition that any two norms on any finite-dimensional normed vector space are equivalent. In particular, if $\|\ \|'$ is any norm on $L(\mathbb{F}^n) = \mathbb{F}^{n\times n}$, then there is a positive constant c so that

$$\|A\|_\infty / c \le \|A\|' \le c\|A\|_\infty, \quad \forall A \in \mathbb{F}^{n\times n}.$$

This implies that $\lim_{k\to\infty} \|B - A_k\|_\infty = 0$ if and only if

$$\lim_{k\to\infty} \|B - A_k\|' = 0,$$

showing that our definition of what it means for A_k to converge to B is independent of the particular matrix norm we use. We might even have chosen the matrix norm

$$\|A\|' := \max_{i,j} |A(i,j)| = \max_{x \neq 0} \frac{\|Ax\|_\infty}{\|x\|_1},$$

and so explicitly confirmed that convergence of matrices is entry-wise, i.e., $\lim_{k\to\infty} A_k = B$ if and only if

$$\lim_{k\to\infty} A_k(i,j) = B(i,j), \quad \forall i,j.$$

Note that, in this chapter, I am using `MATLAB`'s way of writing matrix entries, writing $A_k(i,j)$ instead of $(A_k)_{i,j}$ for the (i,j)-entry of A_k, in order to keep the number of subscripts down.

11.1 For each of the following matrices A, work out A^k for arbitrary $k \in \mathbb{N}$ and, from that, determine directly whether or not the power sequence $A^0, A^1, A^2, \ldots$ converges; if it does, also determine that limit. (i) $A := \alpha\,\mathrm{id}_X$; (ii) $A := \begin{bmatrix} 1/2 & 2^{10} \\ 0 & 1/2 \end{bmatrix}$; (iii) $A := [-e_1, e_2]$; (iv) $A = \begin{bmatrix} a & b \\ 0 & c \end{bmatrix}$.

Three interesting properties of the power sequence of a linear map

We have already most of the tools in hand needed to analyze the following three interesting properties that the **power sequence of** A, i.e., the sequence

$$A^0, A^1, A^2, \cdots \tag{11.2}$$

may have.

Let $A \in L(X)$ with $\dim X < \infty$. Then, for any basis V of X,

$$\widehat{A} := V^{-1}AV$$

is a matrix similar to A, and, for any k,

$$A^k = V\widehat{A}^k V^{-1}.$$

Thus, if we understand the sequence (11.2) for any square *matrix* A, then we understand (11.2) for any $A \in L(X)$ with $\dim X < \infty$.

For this reason, we state here the three interesting properties only for a *matrix* A.

We call the matrix A **power-bounded** in case its power sequence is bounded, i.e., $\sup_k \|A^k\|_\infty < \infty$, i.e., there is a constant c so that, for all k, $\|A^k\|_\infty \le c$.

We call the matrix A **convergent** in case its power sequence converges, i.e., in case, for some matrix B, $B = \lim_{k\to\infty} A^k$.

We call the matrix A **convergent to** 0 in case

$$\lim_{k\to\infty} A^k = 0.$$

See the Backgrounder on the convergence of scalar sequences and, in particular, on the scalar sequence $(\zeta^0, \zeta^1, \zeta^2, \cdots)$.

The first property is fundamental in the study of evolutionary (i.e., time-dependent) processes, such as weather or fluid flow. In the *simplest* approximation, the state of the system (be it the weather or waves on the ocean or whatever) at time t is described by some vector $y(t)$, and the state $y(t+\Delta t)$ at some slightly later time $t + \Delta t$ is computed as

$$y(t+\Delta t) = Ay(t),$$

with A some time-independent matrix. Such a process is called **stable** if $\|y(t)\|$ remains bounded for all time regardless of the initial state, $y(0)$, of the system. Since $y(k\Delta t) = A^k y(0)$, the requirement of stability is equivalent to the power boundedness of A.

The third property is fundamental in the study of iterative processes, as discussed earlier.

The second property is in between the other two. In other words, we have listed the three properties here in the order of increasing strength: if A is convergent to 0, then it is, in particular, convergent. Again, if A is convergent, then it is, in particular, power-bounded.

Suppose now that x is an eigenvector for A, with corresponding eigenvalue μ. Then $Ax = \mu x$, hence $A^k x = \mu^k x$ for $k = 1, 2, 3, \ldots$. Suppose A is powerbounded. Then, in particular, for some c, we should have $c\|x\|_\infty \ge \|A^k\|_\infty \|x\|_\infty \ge \|A^k x\|_\infty = \|\mu^k x\|_\infty = |\mu|^k \|x\|_\infty$. Since $\|x\|_\infty \ne 0$, this implies that the scalar sequence $(|\mu|^k : k = 1, 2, 3, \ldots)$ must be bounded, hence $|\mu| \le 1$. Since we took an arbitrary eigenvector, we conclude that

$$A \text{ powerbounded} \quad \Longrightarrow \quad \rho(A) \le 1. \tag{11.3}$$

Actually, more is true. Suppose that μ is a *defective* eigenvalue for A, which, to recall, means that

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

In other words, there exists an eigenvector for A belonging to μ of the form $x = (A - \mu \operatorname{id})y$. This implies that

$$Ay = x + \mu y.$$

Therefore

$$A^2 y = Ax + \mu Ay = \mu x + \mu(x + \mu y) = 2\mu x + \mu^2 y.$$

Therefore

$$A^3 y = 2\mu Ax + \mu^2 Ay = 2\mu^2 x + \mu^2(x + \mu y) = 3\mu^2 x + \mu^3 y.$$

By now, the pattern is clear:

$$A^k y = k\mu^{k-1} x + \mu^k y.$$

This also makes clear the difficulty: If $|\mu| = 1$, then

$$\|A^k\|_\infty \|y\|_\infty \ge \|A^k y\|_\infty \ge k\|x\|_\infty - \|y\|_\infty.$$

This shows that A cannot be powerbounded.

We conclude:

(11.4) Proposition: If the matrix A is powerbounded, then, for all $\mu \in \operatorname{spec}(A)$, $|\mu| \le 1$, with equality only if μ is a nondefective eigenvalue for A.

Now we consider the case that A is convergent (hence, in particular, powerbounded). If A is convergent, then, for any eigenvector x with associated eigenvalue μ, the sequence $(\mu^k x : k = 0, 1, 2, \ldots)$ must converge. Since x stays fixed, this implies that the scalar sequence $(\mu^k : k = 0, 1, 2, \ldots)$ must converge. This, to recall, implies that $|\mu| \le 1$ with equality only if $\mu = 1$.

Finally, if A is convergent to 0, then, for any eigenvector x with associated eigenvalue μ, the sequence $(\mu^k x)$ must converge to 0. Since x stays fixed (and is nonzero), this implies that the scalar sequence (μ^k) must converge to 0. This, to recall, implies that $|\mu| < 1$.

Remarkably, these simple necessary conditions just derived, for powerboundedness, convergence, and convergence to 0, are also sufficient; see (11.10)Theorem.

For the proof, we need one more piece of information, namely a better understanding of the distinction between defective and nondefective eigenvalues.

11.2 For each of the following four matrices, determine whether or not it is (a) powerbounded, (b) convergent, (c) convergent to zero. (i) id_n; (ii) $[1, 1; 0, 1]$; (iii) $[8/9, 10^{10}; 0, 8/9]$; (iv) $-\operatorname{id}_n$.

Splitting off the nondefective eigenvalues

Recall that the scalar μ is called a *defective* eigenvalue for $A \in L(X)$ in case

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

(11.5) Proposition: If M is a set of nondefective eigenvalues of $A \in L(X)$, for some finite-dimensional vector space X, then X has a basis $U = [V, W]$, with V consisting entirely of eigenvectors of A belonging to these nondefective eigenvalues, and W any basis for the subspace $Z := \operatorname{ran} p(A)$, with $p(t) := \prod_{\mu \in M}(t - \mu)$.

Further, Z is **A-invariant**, meaning that $A(Z) \subset Z$, hence $A|_Z : Z \to Z : z \mapsto Az$ is a well-defined map on Z, and $\operatorname{spec}(A|_Z) = \operatorname{spec}(A)\backslash M$.

Proof: Since $Ap(A) = p(A)A$, we have $AZ = A(\operatorname{ran} p(A)) = \operatorname{ran} Ap(A) = p(A) \operatorname{ran} A \subset \operatorname{ran} p(A) = Z$, showing Z to be A-invariant. This implies that $A|_Z : Z \to Z : z \mapsto Az$ is a well-defined linear map on Z.

We claim that X is the direct sum of $\operatorname{null} p(A)$ and $\operatorname{ran} p(A)$, i.e.,

(11.6) $$X = \operatorname{null} p(A) \dot{+} \operatorname{ran} p(A).$$

Since, by (4.15)Dimension Formula, $\dim X = \dim \operatorname{null} p(A) + \dim \operatorname{ran} p(A)$, it is, by (4.26)Proposition, sufficient to prove that

(11.7) $$\operatorname{null} p(A) \cap \operatorname{ran} p(A) = \{0\}.$$

For its proof, let

$$p_\mu : t \mapsto p(t)/(t - \mu), \qquad \mu \in M,$$

and recall from (5.6) that

$$(p_\mu / p_\mu(\mu) : \mu \in M)$$

is a Lagrange basis for the polynomials of degree $< \#M$. In particular,

$$1 = \sum_{\mu \in M} p_\mu / p_\mu(\mu).$$

Hence, with (10.19)Lemma, $\mathrm{id} = \sum_{\mu \in M} p_\mu(A)/p_\mu(\mu)$ and so, for any $x \in X$,

$$x = \sum_{\mu \in M} x_\mu,$$

with

$$x_\mu := p_\mu(A)x/p_\mu(\mu)$$

in $\operatorname{null}(A - \mu\,\mathrm{id})$ in case $x \in \operatorname{null} p(A)$ (since $(A - \mu\,\mathrm{id})x_\mu = p(A)x/p_\mu(\mu)$), but also in $\operatorname{ran}(A - \mu\,\mathrm{id})$ in case also $x \in \operatorname{ran} p(A) \subset \operatorname{ran}(A - \mu\,\mathrm{id})$, hence then $x_\mu = 0$ since we assumed that each $\mu \in M$ is not defective. This shows (11.7), hence (11.6).

More than that, we just saw that $x \in \operatorname{null} p(A)$ implies that $x = \sum_\mu x_\mu$ with $x_\mu \in \operatorname{null}(A - \mu\,\mathrm{id})$, all $\mu \in M$, hence, $\operatorname{null} p(A) \subset \operatorname{ran} V$, with

$$V := [V_\mu : \mu \in M]$$

and V_μ a basis for $\operatorname{null}(A - \mu\,\mathrm{id})$, all μ. On the other hand, each column of V is in $\operatorname{null} p(A)$, hence also $\operatorname{ran} V \subset \operatorname{null} p(A)$, therefore V is onto $\operatorname{null} p(A)$ and, since it is 1-1 by (10.8)Lemma, it is a basis for $\operatorname{null} p(A)$. Therefore, by (11.6), $U := [V, W]$ is a basis for X for any basis W for $Z = \operatorname{ran} p(A)$.

Finally, let $\nu \in \operatorname{spec}(A)$. If ν were in both M and $\operatorname{spec}(A|_Z)$, then $Ax = \nu x$ for some $x \in Z \backslash 0$, yet also $p(A)x = 0$, hence $0 \neq x \in \operatorname{null} p(A) \cap \operatorname{ran} p(A)$, contradicting (11.7). Thus, if $\nu \in M$, then $\nu \notin \operatorname{spec}(A|_Z)$. If, on the other hand, $\nu \notin M$, then, with x any eigenvector for ν, we have $p(A)x = \alpha x$ with

$$\alpha := \prod_{\mu \in M} (\nu - \mu) \neq 0,$$

and so, $x = \alpha^{-1} p(A)x \in \operatorname{ran} p(A) = Z$, hence $\nu \in \operatorname{spec}(A|_Z)$. This proves that $\operatorname{spec}(A|_Z) = \operatorname{spec}(A)\backslash M$. □

It follows that the matrix representation for A with respect to this basis $U = [V, W]$ has the simple form

$$U^{-1}AU = \begin{bmatrix} \mathrm{M} & 0 \\ 0 & \widehat{B} \end{bmatrix} := \operatorname{diag}(\mu_1, \ldots, \mu_r, \widehat{B}),$$

with $\mu_1, \ldots, \mu_r$ a sequence taken from M, and $\widehat{B}$ some square matrix, namely $\widehat{B} = W^{-1}AW$.

(11.8) Theorem: Let $A \in L(X)$, with X a finite-dimensional vector space.
(i) If A is diagonable, then all its eigenvalues are nondefective.
(ii) If $\mathbb{F} = \mathbb{C}$ and all of A's eigenvalues are nondefective, then A is diagonable.

Proof: (i) This is just a restatement of (10.12)Corollary.

(ii) If none of the eigenvalues of A is defective, then we can choose $M = \operatorname{spec}(A)$ in (11.5)Proposition, leaving $A|_Z$ as a linear map with an empty spectrum. Hence, if also $\mathbb{F} = \mathbb{C}$, then we know from (10.15)Theorem that $\operatorname{ran} W = \operatorname{dom} A|_Z$ must be trivial, hence V is a basis for X. □

Here is a simple *example*. Let $A = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$. Then A maps $x := (1, 1)$ to $(3, 3) = 3x$. Hence, $\mu := 3 \in \operatorname{spec}(A)$. We compute

$$\operatorname{ran}(A - \mu\,\mathrm{id}) = \operatorname{ran} \begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix} = \operatorname{ran} \begin{bmatrix} -1 \\ 1 \end{bmatrix},$$

since the first column of $(A - \mu\,\mathrm{id})$ is bound and the second is free. This also implies that $\operatorname{null}(A - \mu\,\mathrm{id})$ is one-dimensional, with $V := \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ a basis for it.

It follows, by inspection, that $\operatorname{null}(A - \mu\,\mathrm{id}) \cap \operatorname{ran}(A - \mu\,\mathrm{id}) = \{0\}$ since the only vector of the form $(1, 1)\alpha$ *and* of the form $(-1, 1)\beta$ is the zero vector. Equivalently, the matrix $U := \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}$ is 1-1, hence a basis for $\mathbb{R}^2$. Consequently, 3 is a nondefective eigenvalue for A.

Now, what about $A|_Z$, with $Z = \operatorname{ran}(A - \mu\,\mathrm{id})$? In this case, things are very simple since Z is one-dimensional. Since $A(Z) \subset Z$, A must map any $z \in Z$ to a scalar multiple of itself! In particular, since $z = (-1, 1) \in \operatorname{ran}(A - \mu\,\mathrm{id})$, A must map this z into a scalar multiple of itself, and this is readily confirmed by the calculation that A maps z to $-(2, 1) + (1, 2) = z$, i.e., to itself. This shows that z is an eigenvector for A belonging to the eigenvalue 1.

Altogether therefore,

$$AU = [Ax, Az] = [3x, z] = U\,\mathrm{diag}(3,1),$$

showing that A is actually diagonable.

This simple example runs rather differently when we change A to $A := \begin{bmatrix} 2 & 1 \\ 0 & 2 \end{bmatrix}$. Since A is upper triangular, its sole eigenvalue is $\mu = 2$. But $(A - \mu\,\mathrm{id}) = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$, and we saw earlier that its range and nullspace have the nontrivial vector e_1 in common. Hence, 2 is a defective eigenvalue for this matrix A.

(11.9) Example: Let $A := [x][y]^{\mathrm{t}}$ with $x, y \in \mathbb{R}^n\backslash 0$. Then $\mathrm{rank}\,A = 1$, hence $\mathrm{ran}\,A = \mathrm{ran}[x]$ is one-dimensional, therefore x is an eigenvector for A. Since $Az = x\ (y^{\mathrm{t}}z)$, we have, in particular,

$$Ax = x\ (y^{\mathrm{t}}x),$$

hence x is an eigenvector for A belonging to the eigenvalue $\mu := y^{\mathrm{t}}x$.

Since A is of rank 1, $\dim\mathrm{null}\,A = n-1$. Let V be a basis for $\mathrm{null}\,A$, i.e., $V \in L(\mathbb{R}^{n-1}, \mathrm{null}\,A)$ invertible. Then $U := [V, x]$ is 1-1 (hence a basis for $\mathbb{R}^n$) if and only if $x \notin \mathrm{ran}\,V$, i.e., if and only if $x \notin \mathrm{null}\,A$.

case $x \notin \mathrm{ran}\,V$: Then $U = [V, x]$ is a basis for $\mathbb{R}^n$. Consider the representation $\widehat{A} = U^{-1}AU$ for A with respect to this basis: With $V =: [v_1, v_2, \ldots, v_{n-1}]$, we have $Au_j = Av_j = 0$ for $j = 1{:}n{-}1$, therefore

$$\widehat{A}e_j = 0, \qquad j = 1{:}n{-}1.$$

Further, we have $Ax = x\ (y^{\mathrm{t}}x)$, therefore

$$\widehat{A}e_n = U^{-1}AUe_n = U^{-1}Ax = (y^{\mathrm{t}}x)e_n,$$

(recall that, for any $z \in \mathbb{R}^n$, $U^{-1}z$ provides the coordinates of z with respect to the basis U, i.e., $U(U^{-1}z) = z$). Hence, altogether,

$$\widehat{A} = [0, \ldots, 0, (y^{\mathrm{t}}x)e_n].$$

In particular, A is diagonable, with eigenvalues 0 and $y^{\mathrm{t}}x$.

case $x \in \mathrm{ran}\,V$: Then $U = [V, x]$ is not a basis for $\mathbb{R}^n$. Worse than that, A is now not diagonable. This is due to the fact that, in this case, the eigenvalue 0 for A is *defective*: For, $x \neq 0$ while $Ax = 0$, hence

$$\{0\} \neq \mathrm{ran}(A - 0\,\mathrm{id}) = \mathrm{ran}\,A = \mathrm{ran}[x] \subset \mathrm{null}\,A = \mathrm{null}(A - 0\,\mathrm{id}).$$

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

It is hard to tell just by looking at a matrix whether or not it is diagonable. There is one exception: If A is hermitian, i.e., equal to its conjugate transpose, then it is not only diagonable, but has an orthonormal basis of eigenvectors, as is shown in the next chapter.

11.3 Prove: *If* $A = \begin{bmatrix} B & C \\ 0 & D \end{bmatrix}$, *with B and D square matrices, then* $\mathrm{spec}(A) = \mathrm{spec}(B) \cup \mathrm{spec}(D)$. (Hint: Prove first that such a matrix A is invertible if and only if both B and D are invertible.)

11.4 Determine the spectrum of the matrix $A := \begin{bmatrix} 1 & 2 & 3 & 4 \\ 2 & 4 & 5 & 6 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 2 & 1 \end{bmatrix}$.

11.5 (a) Determine the spectrum of the matrix $A := \begin{bmatrix} 1 & 2 & a \\ 2 & 1 & b \\ 0 & 0 & 3 \end{bmatrix}$. (b) For which choices of a and b is A not diagonable?

Three interesting properties of the power sequence of a linear map: The sequel

(11.10) Theorem: Let $A \in \mathbb{C}^{n\times n}$. Then:

(i) A is powerbounded iff, for all $\mu \in \operatorname{spec}(A)$, $|\mu| \le 1$, with $|\mu| = 1$ only if μ is not defective.

(ii) A is convergent iff, for all $\mu \in \operatorname{spec}(A)$, $|\mu| \le 1$, with $|\mu| = 1$ only if μ is not defective and $\mu = 1$.

(iii) A is convergent to 0 iff $\rho(A) < 1$.

Proof: We only have to prove the implications '$\Longleftarrow$', since we proved all the implications '$\Longrightarrow$' in an earlier section (see pages 112ff).

We begin with (iii). Since A is a matrix over the complex scalars, we know from (10.26)Corollary that, for any $\varepsilon > 0$, we can find an upper triangular matrix B_ε similar to A and with all off-diagonal entries less than ε in absolute value. This means, in particular, that $A = VB_\varepsilon V^{-1}$ for some (invertible) matrix V, hence, for any k, $A^k = V(B_\varepsilon)^k V^{-1}$, therefore,

$$\|A^k\|_\infty \le \|V\|_\infty \|B_\varepsilon\|_\infty^k \|V^{-1}\|_\infty.$$

We compute

$$\|B_\varepsilon\|_\infty = \max_i \sum_j |B_\varepsilon(i,j)| \le \max_i |B_\varepsilon(i,i)| + (n-1)\varepsilon,$$

since each of those sums involves $n-1$ off-diagonal entries and each such entry is less than ε in absolute value. Further, B_ε is upper triangular and similar to A, hence

$$\max_i |B_\varepsilon(i,i)| = \max\{|\mu| : \mu \in \operatorname{spec}(A)\} = \rho(A).$$

By assumption, $\rho(A) < 1$. This makes it possible to choose ε positive yet so small that $\rho(A) + (n-1)\varepsilon < 1$. With this choice, $\|B_\varepsilon\|_\infty < 1$, hence $\lim_{k\to\infty} \|B_\varepsilon\|_\infty^k = 0$. Therefore, since $\|V\|_\infty$ and $\|V^{-1}\|_\infty$ stay fixed throughout, also $\|A^k\|_\infty \to 0$ as $k \to \infty$. In other words, A is convergent to 0.

With this, we are ready also to handle (i) and (ii). Both assume that all eigenvalues of A of modulus 1 are nondefective. By (11.5)Proposition, this implies the existence of a basis $U = [V, W]$ for $\mathbb{C}^n$ so that V consists of eigenvectors of A belonging to eigenvalues of modulus 1, while $Z := \operatorname{ran} W$ is A-invariant and $\operatorname{spec}(A|_Z)$ has only eigenvalues of modulus < 1. In particular, $AV = V\mathrm{M}$ for some diagonal matrix M with all diagonal entries of modulus 1, and $AW = WB$ for some matrix B with $\operatorname{spec}(B) = \operatorname{spec}(A|_Z)$, hence $\rho(B) < 1$. Consequently, for any k,

$$A^k U = A^k[V, W] = [A^k V, A^k W] = [V\mathrm{M}^k, WB^k] = U \operatorname{diag}(\mathrm{M}^k, B^k).$$

In other words,

$$A^k = U \operatorname{diag}(\mathrm{M}^k, B^k) U^{-1}.$$

Therefore, $\|A^k\|_\infty \le \|U\|_\infty \max\{\|\mathrm{M}\|_\infty^k, \|B^k\|_\infty\} \|U^{-1}\|_\infty$, and this last expression converges since $\|\mathrm{M}\|_\infty = 1$ while $\|B^k\|_\infty \to 0$, by (iii). Since any convergent sequence is bounded, this implies that also the sequence $(\|A^k\|_\infty)$ must be bounded, hence we have finished the proof of (i).

Assume now, in addition, as in (ii) that all eigenvalues of A of modulus 1 are actually equal to 1. Then $\mathrm{M} = \mathrm{id}$, and so, $\lim_{k\to\infty} A^k = C := U \operatorname{diag}(\mathrm{M}, 0) U^{-1}$ since $A^k - C = U \operatorname{diag}(0, B^k) U^{-1}$, hence

$$\|A^k - C\|_\infty \le \|U\|_\infty \|B^k\|_\infty \|U^{-1}\|_\infty \le \mathrm{const} \|B^k\|_\infty \to 0$$

as $k \to \infty$. □

(11.11) Example: Here is a concrete example, chosen for its simplicity.

Let $A = \begin{bmatrix} 1 & 1 \\ 0 & \alpha \end{bmatrix}$. Then $\mathrm{spec}(A) = \{1, \alpha\}$. In particular, A is diagonable if $\alpha \neq 1$ (by (10.9)Corollary) since then A has two eigenvalues. On the other hand, if $\alpha = 1$, then A is not diagonable since it then looks like $\mathrm{id}_2 + N$, with $N := [0, e_1]$ the simplest example of a non-diagonable matrix. Also, in the latter case, the sole eigenvalue, 1, is certainly defective since e_1 is both in $\mathrm{null}(A - \mathrm{id})$ and in $\mathrm{ran}(A - \mathrm{id})$.

Also,

$$A^k = \begin{bmatrix} 1 & 1 + \alpha + \cdots + \alpha^{k-1} \\ 0 & \alpha^k \end{bmatrix} = \begin{cases} \begin{bmatrix} 1 & \frac{1-\alpha^k}{1-\alpha} \\ 0 & \alpha^k \end{bmatrix} & \text{if } \alpha \neq 1; \\ \begin{bmatrix} 1 & k \\ 0 & 1 \end{bmatrix} & \text{otherwise.} \end{cases}$$

We see that A is powerbounded whenever $|\alpha| \le 1$ *except* when $\alpha = 1$, i.e., except when there is a defective absolutely largest eigenvalue.

Further, A is convergent iff $|\alpha| < 1$, i.e., if, in addition, the sole eigenvalue of size 1 is equal to 1 and is nondefective. □

The power method

The simple background for the success of the **power method** is the following corollary to (11.10)Theorem (ii).

(11.12) Proposition: If A has just one eigenvalue μ of absolute value $\rho(A)$ and μ is nondefective, then, for almost any x and almost any y, the sequence

$$A^k x / (y^c A^k x), \qquad k = 1, 2, \ldots$$

converges to an eigenvector of A belonging to that absolutely maximal eigenvalue μ. In particular, for almost any vector y, the ratio

$$y^c A^{k+1} x / y^c A^k x$$

converges to μ.

Proof: By assumption, there is (by (11.5)Proposition) a basis $U := [V, W]$, with V a basis for the space $\mathrm{null}(A - \mu\,\mathrm{id})$ of all eigenvectors of A belonging to that absolutely largest eigenvalue μ of A, and $B := A|_{\mathrm{ran}\, W}$ having all its eigenvalues $< |\mu|$ in absolute value. This implies that $\rho(B/\mu) < 1$. Therefore, for any $x =: [V, W](a, b)$,

$$A^k x = \mu^k V a + B^k W b = \mu^k \left(V a + (B/\mu)^k W b \right)$$

and $(B/\mu)^k W b \to 0$ as $k \to \infty$. Consequently, for any y,

$$\frac{y^c A^{k+1} x}{y^c A^k x} = \frac{\mu^{k+1} (y^c V a + y^c (B/\mu)^{k+1} W b)}{\mu^k (y^c V a + y^c (B/\mu)^k W b)} = \mu \frac{y^c V a + y^c (B/\mu)^{k+1} W b}{y^c V a + y^c (B/\mu)^k W b} \to \mu$$

provided $y^c V a \neq 0$. □

Note that the speed with which $y^c A^{k+1} x / y^c A^k x$ converges to μ depends on the speed with which $(B/\mu)^k W b \to 0$ as $k \to \infty$, hence, ultimately, on $\rho(B/\mu)$.

In the **scaled power method**, one would, instead, consider the sequence

$$x_{k+1} := A(x_k / \|x_k\|), \quad k = 0, 1, \ldots,$$

or, more simply, the sequence

$$x_{k+1} := A(x_k/y^{\mathrm{t}}x_k), \quad k = 0, 1, \ldots.$$

The power method is at the heart of good algorithms for the calculation of eigenvalues. In particular, the standard algorithm, i.e., the QR method, can be interpreted as a (very sophisticated) variant of the power method.

11.6 Using `MATLAB` if really necessary, try out the Power method on the following matrices A, starting at the specified vector x, and discuss success or failure. (Note: You can always use `eig(A)` to find out what the absolutely largest eigenvalue of A is (as well as some eigenvector for it), hence can tell whether or not the power method is working for you. If it isn't, identify the source of failure.) (a) $A = \begin{bmatrix} 0 & .2 & .2 & .3 \\ .2 & 0 & .2 & .3 \\ .5 & .4 & 0 & .4 \\ .3 & .4 & .6 & 0 \end{bmatrix}$, $x = (1, 1, 1, 1)$; (b) $A = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$, $x = (1, -1)$; (c) $A = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$, $x = e_1$; (d) $A = \begin{bmatrix} 4 & 1 & -1 \\ 2 & 5 & -2 \\ 1 & 1 & 2 \end{bmatrix}$, $x = (1, -2, -1)$.

11.7 T/F

(a) If the matrix A of order n has n eigenvalues, then none of its eigenvalues is defective.

(b) If, for some sequence $(x_n : n \in \mathbb{N})$ of m-vectors, $\lim_{n\to\infty} \|x_n\|_2 = 0$, then $\lim_{n\to\infty} \|x_n\| = 0$ for any norm $\|\cdot\|$ on $\mathbb{F}^m$.

(c) If all the eigenvalues of A are < 1, then $\lim_{k\to\infty} A^k \to 0$.

(d) If all the eigenvalues of A are ≤ 1 in absolute value, then A is power-bounded.

(e) If $p(A)x = 0$ for some polynomial p, $A \in L(X)$ and $x \in X\backslash\{0\}$, then every eigenvalue of A is a zero of p.

12. Canonical forms

Canonical forms exhibit essential aspects of a linear map. Of the three discussed in this chapter, only the Schur form has practical significance. But the mathematics leading up to the other two is too beautiful to be left out.

The only result from this chapter used later in these notes is the spectral theorem for hermitian matrices.

The Schur form

The discussion of the powers A^k of A used crucially the fact that any square matrix is similar to an upper triangular matrix. The argument we gave there for this fact is due to I. Schur, who used a refinement of it to show that the basis V for which $V^{-1}AV$ is upper triangular can even be chosen to be *unitary* or *orthonormal*, i.e., so that

$$V^{c}V = \text{id}.$$

(12.1) Schur's theorem: Every $A \in L(\mathbb{C}^n)$ is **unitarily similar** to an upper triangular matrix, i.e., there exists a unitary basis U for $\mathbb{C}^n$ so that $\widehat{A} := U^{-1}AU = U^{c}AU$ is upper triangular.

Proof: Simply repeat the proof of (10.25)Theorem, with the following modifications: Normalize the eigenvector u_1, i.e., make it have (Euclidean) length 1, then extend it to an o.n. basis for $\mathbb{C}^n$ (as can always be done by applying Gram-Schmidt to an arbitrary basis $[u_1, \ldots]$ for $\mathbb{C}^n$). Also, observe that unitary similarity is also an equivalence relation since the product of unitary matrices is again unitary. Finally, if W is unitary, then so is $\text{diag}(1, W)$. □

Here is one of the many consequences of Schur's theorem. It concerns **hermitian** matrices, i.e., matrices A for which $A^{c} = A$. By Schur's theorem, such a matrix, like any other matrix, is unitarily similar to an upper triangular matrix, i.e., for some unitary matrix U, $\widehat{A} := U^{c}AU$ is upper triangular. On the other hand, for any matrix A and any unitary matrix U,

$$(U^{c}AU)^{c} = U^{c}(A^{c})U.$$

In other words: if $\widehat{A}$ is the matrix representation for A with respect to a *unitary* basis, then $\widehat{A}^{c}$ is the matrix representation for A^{c} with respect to the very same basis. For our hermitian matrix A with its upper triangular matrix representation $\widehat{A} = U^{c}AU$ with respect to the unitary basis U, this means that also $\widehat{A}^{c} = \widehat{A}$, i.e., that the *upper* triangular matrix $\widehat{A}$ is also *lower triangular* and that its diagonal entries are all real. This proves the hard part of the following remarkable

(12.2) Corollary: A matrix $A \in \mathbb{C}^n$ is hermitian if and only it is unitarily similar to a real diagonal matrix.

Proof: We still have to prove that if $\widehat{A} := U^{c}AU$ is real and diagonal for some unitary U, then A is necessarily hermitian. But that follows at once from the fact that then $\widehat{A}^{c} = \widehat{A}$, therefore $A^{c} = (U\widehat{A}U^{c})^{c} = U\widehat{A}^{c}U^{c} = U\widehat{A}U^{c} = A$. □

A slightly more involved argument makes it possible to characterize all those matrices that are unitarily similar to a diagonal matrix (real or not). Such a matrix has enough eigenvectors to make up an entire orthonormal basis from them. Here are the details.

Start with the observation that diagonal matrices commute with one another. Thus, if $\widehat{A} := U^{c}AU$ is diagonal, then

$$A^{c}A = (U\widehat{A}^{c}U^{c})(U\widehat{A}U^{c}) = U\widehat{A}^{c}\widehat{A}U^{c} = U\widehat{A}\widehat{A}^{c}U^{c} = (U\widehat{A}U^{c})(U\widehat{A}^{c}U^{c}) = AA^{c},$$

