hence having $A^cA = AA^c$ is a necessary condition for A to be unitarily similar to a diagonal matrix. Remarkably, this condition is sufficient as well. Note that this condition can be directly tested by computing the two products and comparing them. It constitutes the only criterion for the diagona(liza)bility of a matrix available that can be tested for by finitely many calculations. Not surprisingly, matrices with this property are very convenient and have, correspondingly, been given a very positive label. They are called **normal**. (Another label might have been **boring**.)

One way to prove that normal matrices are unitarily similar to a diagonal matrix is by way of a refinement of Schur's theorem: It is possible to find a unitary basis that simultaneously upper-triangularizes two matrices A and B provided A and B **commute**, i.e., provided $AB = BA$. This is due to the fact that commuting matrices have some eigenvector in common.

Assuming this refinement of Schur's theorem (cf. (12.5)Theorem below), one would obtain, for a given normal matrix A, a unitary basis U so that both U^cAU and U^cA^cU are upper triangular. Since one of these is the conjugate transpose of the other, they must both be diagonal. This finishes the proof of

(12.3) Theorem: A matrix $A \in \mathbb{C}^n$ is unitarily similar to a diagonal matrix if and only if $AA^c = A^cA$.

Now for the proof of the refined Schur's theorem. Since the proof of Schur's theorem rests on eigenvectors, it is not surprising that a proof of its refinement rests on the following

(12.4) Lemma: If $A, B \in \mathbb{C}^n$ commute, then there exists a vector that is eigenvector for both of them.

Proof: Let x be an eigenvector for A, $Ax = x\mu$ say, and let $p = p_{B,x}$ be the minimal annihilating polynomial for B at x. Since $x \neq 0$, p has zeros. Let ν be one such and set $p =: (\cdot - \nu)q$. Since $\mathbb{F} = \mathbb{C}$, we know that $v := q(B)x$ is an eigenvector for B (for the eigenvalue ν). But then, since $AB = BA$, we also have $Aq(B) = q(B)A$, therefore

$$Av = Aq(B)x = q(B)Ax = q(B)x\mu = v\mu,$$

showing that our eigenvector v for B is also an eigenvector for A. □

(12.5) Schur's refined theorem: For every $A, B \in L(\mathbb{C}^n)$ that commute, there exists a unitary basis U for $\mathbb{C}^n$ so that both U^cAU and U^cBU are upper triangular.

Proof: This is a further refinement of the proof of (10.25)Theorem. The essential step in that proof was to come up with some eigenvector for A which was then extended to a basis, well, to an o.n. basis U for the proof of Schur's Theorem. Therefore, to have U simultaneously upper-triangularize both A and B, all that's needed is (i) to observe that, by (12.4)Lemma, we may choose u_1 to be a (normalized) eigenvector of A *and* B since, by assumption, $AB = BA$; and (ii) verify that the submatrices A_1 and B_1 obtained in the first step again commute (making it possible to apply the induction hypothesis to them). Here is the verification of this latter fact:

Assuming the eigenvalue of B corresponding to the eigenvector u_1 to be ν, we have

$$U^cAU = \begin{bmatrix} \mu & C \\ 0 & A_1 \end{bmatrix} \qquad U^cBU = \begin{bmatrix} \nu & D \\ 0 & B_1 \end{bmatrix}.$$

Therefore

$$\begin{bmatrix} \mu\nu & \mu D + CB_1 \\ 0 & A_1 B_1 \end{bmatrix} = \begin{bmatrix} \mu & C \\ 0 & A_1 \end{bmatrix} \begin{bmatrix} \nu & D \\ 0 & B_1 \end{bmatrix}$$
$$= U^{\rm c} A U U^{\rm c} B U = U^{\rm c} A B U = U^{\rm c} B A U = U^{\rm c} B U \, U^{\rm c} A U = \begin{bmatrix} \nu\mu & \nu C + D A_1 \\ 0 & B_1 A_1 \end{bmatrix},$$

hence also A_1 and B_1 commute. □

The primary decomposition

The following analysis goes back to Frobenius and could be viewed as a first step toward a finest A-invariant direct sum decomposition, aka the Jordan form, though the Jordan form is deduced in the next section without any reference to this section. We give the analysis here in the more general situation when the scalar field $\mathbb{F}$ may not be algebraically closed.

The 'primary decomposition' refers to the following facts (taken for granted here). The ring Π of (univariate) polynomials over the field $\mathbb{F}$ is a **unique factorization domain**. This means that each monic polynomial can be written in exactly one way (up to order of the factors) as a product of **irreducible** polynomials, i.e., monic polynomials that have no proper factors. Here, p is called a **proper factor of** q if (i) $0 < \deg p < \deg q$, and (ii) $q = hp$ for some polynomial h.

If $\mathbb{F} = \mathbb{C}$ (or any other algebraically closed field), then each such irreducible polynomial is a monic linear polynomial, i.e., of the form $(\cdot - \mu)$ for some scalar μ. Otherwise, irreducible polynomials may well be of higher than first degree. In particular, if $\mathbb{F} = \mathbb{R}$, then an irreducible polynomial may be of second degree, like the polynomial $()^2 + 1$, but no irreducible polynomial would be of higher degree than that.

The irreducible polynomials are the 'primes' in the 'ring' Π, hence the above-mentioned unique factorization is one into powers of 'primes', or a **prime factorization**.

To obtain the 'primary decomposition' of the linear space X with respect to the linear map $A \in L(X)$, it is convenient to start with the set

$$\mathcal{N}_A := \{p \in \Pi : \operatorname{null} p(A) \ne \{0\}\}$$

of all polynomials p for which $p(A)$ fails to be invertible. This set is not trivial, meaning that it contains more than just the zero polynomial, if, as we continue to assume, $\dim X < \infty$, since then

$$p_{A,x} \in \mathcal{N}_A, \quad \forall x \in X, \tag{12.6}$$

with $p_{A,x}$ the *minimal polynomial for A at x*, which, to recall, is the monic polynomial p of smallest degree for which $p(A)x = 0$.

Call an element of $\mathcal{N}_A$ **minimal** if it is monic and none of its proper factors is in $\mathcal{N}_A$, and let

$$\mathcal{Q}_A$$

be the collection of all minimal elements of $\mathcal{N}_A$.

The set $\mathcal{Q}_A$ is not empty since $\mathcal{N}_A$ is not empty and is closed under multiplication by a scalar, hence contains a monic polynomial of smallest degree. Any $q \in \mathcal{Q}_A$ is necessarily **irreducible**, since, otherwise, it would be the product of certain polynomials p with $p(A)$ 1-1, hence also $q(A)$ would be 1-1.

For every $q \in \mathcal{Q}_A$ and every $x \in \operatorname{null} q(A) \backslash 0$, necessarily $p_{A,x} = q$, by the minimality of $p_{A,x}$. This implies that

$$p, q \in \mathcal{Q}_A \text{ and } \operatorname{null} p(A) \cap \operatorname{null} q(A) \ne \{0\} \implies p = q. \tag{12.7}$$

(12.8) Lemma: Let p be a product of elements of $\mathcal{Q}_A$,

$$p =: \prod_{q\in\mathcal{Q}'_A} q(A)^{d_q}$$

say, with $d_q \in \mathbb{N}$ and $\mathcal{Q}'_A$ a finite subset of $\mathcal{Q}_A$. Then,

$$X_p := \operatorname{null} p(A) = \dot{+}_{q\in\mathcal{Q}'_A} \operatorname{null} q(A)^{d_q}, \tag{12.9}$$

i.e., $X_p = \operatorname{null} p(A)$ is the direct sum of the spaces $Y_q := \operatorname{null} q(A)^{d_q}$. In other words (by (4.26)Proposition), with V_q a basis for Y_q,

$$V_p := [V_q : q \in \mathcal{Q}'_A]$$

is a basis for X_p.

Proof: There is nothing to prove if $\mathcal{Q}'_A$ has just one element. So, assume that $\#\mathcal{Q}'_A > 1$, and consider the set

$$\mathcal{I} := \sum_{q\in\mathcal{Q}'_A} (p/q^{d_q})\Pi := \{\sum_{q\in\mathcal{Q}'_A} (p/q^{d_q})p_q : p_q \in \Pi\}$$

of all polynomials writable as a weighted sum of the polynomials

$$p/q^{d_q} = \prod_{g\in\mathcal{Q}'_A\backslash q} g^{d_g}$$

for $q \in \mathcal{Q}'_A$, with polynomial (rather than just scalar) weights. This set is a polynomial **ideal**, meaning that it is closed under addition, as well as under multiplication by polynomials. More than that, let q^* be the monic polynomial of smallest degree in $\mathcal{I}$. By Euclid's algorithm, for every $q \in \mathcal{I}$, there exist polynomials g and h with $q = hq^* + g$, hence $g = q - hq^* \in \mathcal{I}$, yet $\deg g < \deg q^*$, hence, by the minimality of q^*, $g = 0$. In other words, the monic polynomial q^* is a factor of every $q \in \mathcal{I}$, in particular of every p/q^{d_q} with $q \in \mathcal{Q}'_A$. But these polynomials have no proper factor in common. Therefore, q^* is necessarily the monic polynomial of degree 0, i.e., $q^* = ()^0$.

It follows that

$$()^0 = \sum_{q\in\mathcal{Q}'_A} (p/q^{d_q})h_q$$

for certain polynomials h_q. This implies that, for the corresponding linear maps

$$P_q : X_p \to X_p : y \mapsto (p/q^{d_q})(A)h_q(A)y, \quad q \in \mathcal{Q}'_A,$$

we have

$$\mathrm{id}_{X_p} = \sum_q P_q. \tag{12.10}$$

Also, for $q \neq g$, $P_qP_g = s(A)p(A) = 0$ for some $s \in \Pi$. Therefore also

$$P_q = P_q\,\mathrm{id}_{X_p} = P_q(\sum_g P_g) = \sum_g P_qP_g = P_qP_q.$$

This shows that each P_q is a linear projector, and, by (5.11), that X_p is the direct sum of the ranges of the P_q. It remains to show that

$$\operatorname{ran} P_q = Y_q = \operatorname{null} q(A)^{d_q}. \tag{12.11}$$

It is immediate that $\operatorname{ran} P_q \subset Y_q \subset X_p$. With that, $Y_q \subset \operatorname{null} P_g$ for all $g \in \mathcal{Q}'_A\backslash q$, and this implies (12.11), by (12.10). □

Now let $p = p_A$ be the **minimal (annihilating) polynomial for** A, meaning the monic polynomial p of smallest degree for which $p(A) = 0$.

To be sure, there is such a polynomial since X is finite-dimensional, hence so is $L(X)$ (by (4.24)Corollary), therefore $[A^r : r = 0{:}\dim L(X)]$ must fail to be 1-1, i.e., there must be some a for which

$$p(A) := \sum_{j \le \dim L(X)} a_j A^j = 0,$$

yet $a_j \neq 0$ for some $j > 0$, hence the set of all annihilating polynomials of positive degree is not empty, therefore must have an element of minimal degree, and it will remain annihilating and of that degree if we divide it by its leading coefficient.

By the minimality of p_A, every proper factor of p_A is necessarily in $\mathcal{N}_A$. Hence p_A is of the form

$$p_A = \prod_{q \in \mathcal{Q}'_A} q^{d_q}$$

for some $\mathcal{Q}'_A \subset \mathcal{Q}_A$. (In fact, it is immediate from (12.8)Lemma that necessarily $\mathcal{Q}'_A = \mathcal{Q}_A$, but we don't need that here.) This gives, with (12.8)Lemma, the primary decomposition for X wrto A:

$$X = \dot{+}_q \operatorname{null} q(A)^{d_q}. \tag{12.12}$$

Necessarily,

$$\operatorname{null} q(A)^{d_q} = \cup_r \operatorname{null} q(A)^r,$$

with d_q the smallest natural number for which this equality holds. Indeed, from (12.12), every $x \in X$ is uniquely writable as $x = \sum_g x_g$ with $x_g \in \operatorname{null} g(A)^{d_g}$, all $g \in \mathcal{Q}'_A$, and, since each $\operatorname{null} g(A)^{d_g}$ is A-invariant, we therefore have $q(A)^r x = \sum_g q(A)^r x_g = 0$ if and only if $q(A)^r x_g = 0$ for all $g \in \mathcal{Q}'_A$. However, as we saw before, for each $g \in \mathcal{Q}_A \backslash q$, $q(A)$ is 1-1 on $\operatorname{null} g(A)^{d_g}$, hence $q(A)^r x_g = 0$ if and only if $x_g = 0$. Therefore, altogether, $\operatorname{null} q(A)^{d_q} \supset \operatorname{null} q(A)^r$ for any r. This proves that

$$\operatorname{null} q(A)^{d_q} \supset \cup_r \operatorname{null} q(A)^r,$$

while the converse inclusion is trivial. If now $\operatorname{null} q(A)^r = \operatorname{null} q(A)^{d_q}$ for some $r < d_q$, then already $p := p_A / q^{d_q - r}$ would annihilate X, contradicting p_A's minimality.

If $\mathbb{F} = \mathbb{C}$, then each q is of the form $\cdot - \mu_q$ for some scalar μ_q and, correspondingly,

$$X = \dot{+}_q \operatorname{null}(A - \mu_q \operatorname{id})^{d_q}.$$

In particular, $A - \mu_q \operatorname{id}$ is nilpotent on

$$Y_q := \operatorname{null}(A - \mu_q \operatorname{id})^{d_q},$$

with degree of nilpotency equal to d_q. Since

$$A = \mu_q \operatorname{id} + (A - \mu_q \operatorname{id}),$$

it follows that

$$\exp(tA) = \exp(t\mu_q \operatorname{id}) \exp(t(A - \mu_q \operatorname{id})) = \exp(t\mu_q) \sum_{r < d_q} t^r (A - \mu_q \operatorname{id})^r / r! \quad \text{on } Y_q, \tag{12.13}$$

thus providing a very helpful detailed description of the solution $y : t \mapsto \exp(tA)c$ to the first-order ODE $y'(t) = Ay(t)$, $y(0) = c$, introduced in (10.4).

The Jordan form

The Jordan form is the result of the search for the 'simplest' matrix representation for $A \in \mathbb{F}^{n\times n}$. It starts off from the following observation.

Suppose $X := \mathbb{F}^n$ is the direct sum

$$X = Y_1 \dotplus Y_2 \dotplus \cdots \dotplus Y_r \tag{12.14}$$

of r linear subspaces, each of which is A-invariant. Then, with V_j a basis for Y_j, we have $AV_j \subset \operatorname{ran} V_j$, all j. This implies that the coordinate vector of any column of AV_j with respect to the basis $V := [V_1, \ldots, V_r]$ for X has nonzero entries only corresponding to columns of V_j, and these possibly nonzero entries can be found as the corresponding column in the matrix $V_j^{-1}AV_j$. Consequently, the matrix representation $\widehat{A} = V^{-1}AV$ for A with respect to the basis V is block-diagonal, i.e., of the form

$$\widehat{A} = \operatorname{diag}(V_j^{-1}AV_j : j = 1{:}r) = \begin{bmatrix} V_1^{-1}AV_1 & & \\ & \ddots & \\ & & V_r^{-1}AV_r \end{bmatrix}.$$

The smaller we can make the A-invariant summands Y_j, the simpler and more helpful is our overall description $\widehat{A}$ of the linear map A. Of course, the smallest possible A-invariant subspace of X is the trivial subspace, but it would not contribute any columns to V, hence we will assume from now on that our A-invariant direct sum decomposition (12.14) is **proper**, meaning that none of its summands Y_j is trivial.

With that, each Y_j has dimension ≥ 1, hence is as small as possible if it is 1-dimensional, $Y_j = \operatorname{ran}[v_j]$ say, for some nonzero v_j. In this case, A-invariance says that Av_j must be a scalar multiple of v_j, hence v_j is an eigenvector for A, and the sole entry of the matrix $[v_j]^{-1}A[v_j]$ is the corresponding eigenvalue for A.

Thus, at best, each Y_j is 1-dimensional, hence V consists entirely of eigenvectors for A, i.e., A is diagonable. Since we know that not every matrix is diagonable, we know that this best situation cannot always be attained. But we can try to make each Y_j as small as possible, in the following way.

(12.15) Jordan Algorithm:
input: $X = \mathbb{F}^n$, $A \in L(X)$.
$\mathcal{Y} \leftarrow \{X\}$
while $\exists Z_1 \dotplus Z_2 \in \mathcal{Y}$ with both Z_j nontrivial and A-invariant, **do**:
  replace $Z_1 \dotplus Z_2$ in $\mathcal{Y}$ by Z_1 and Z_2.
endwhile
output: the proper A-invariant direct sum decomposition $X = \dotplus_{Y\in\mathcal{Y}} Y$.

At all times, the elements of $\mathcal{Y}$ form a proper direct sum decomposition for X. Hence

$$\#\mathcal{Y} \leq \sum_{Y\in\mathcal{Y}} \dim Y = \dim X = n.$$

Since each pass through the **while**-loop increases $\#\mathcal{Y}$ by 1, the algorithm must terminate after at most $n-1$ steps.

Now consider any particular Y in the collection $\mathcal{Y}$ output by the algorithm. It is a nontrivial A-invariant subspace. Hence, with the assumption that $\mathbb{F} = \mathbb{C}$, we know that $Y \to Y : y \mapsto Ay$ is a linear map with some eigenvalue, μ say. This implies that the linear map

$$B : Y \to Y : y \mapsto (A - \mu\,\mathrm{id})y$$

is well-defined and has a nontrivial nullspace.

Claim 1: B is nilpotent, i.e., $B^q = 0$ for some q.

Proof: All the terms in the increasing sequence

$$\operatorname{null} B \subset \operatorname{null} B^2 \subset \operatorname{null} B^3 \subset \cdots$$

have dimension $\le \dim Y < \infty$, hence the sequence must be eventually constant, i.e., there must be some q with

$$\operatorname{null} B^q = \operatorname{null} B^{q+k}, \quad k = 1, 2, \ldots .$$

Set $Z_1 := \operatorname{ran} B^q$, $Z_2 := \operatorname{null} B^q$, and let $z \in Z_1 \cap Z_2$. Then $z = B^q y$ for some $y \in Y$ and $0 = B^q z = B^{q+q} y$, hence $y \in \operatorname{null} B^{q+q} = \operatorname{null} B^q$, therefore already $z = B^q y = 0$. This shows that $Z_1 \cap Z_2 = \{0\}$, while, by (4.15)Dimension formula, $\dim Z_1 + \dim Z_2 = \dim \operatorname{dom} B^q = \dim Y$, therefore, by (4.26)Proposition(iv), $Y = Z_1 \dot{+} Z_2$.

Also, each Z_j is B-invariant, hence also invariant under $A = B + \mu \operatorname{id}$.

Therefore, by construction of $Y \in \mathcal{Y}$, one of these two A-invariant summands must be trivial. Since we started off with $\operatorname{null} B \ne \{0\}$, this implies, finally, that $Z_1 = \operatorname{ran} B^q$ must be trivial, i.e., $B^q = 0$. □

Claim 2: Y has a basis of the form $[B^{q-1}y, B^{q-2}y, \ldots, y]$.

Proof: Choose the q in Claim 1 as small as possible. Then there exists $y \in Y$ with $B^{q-1}y \ne 0$, therefore there is also $z \in Y$ with $z^{\mathrm{t}} B^{q-1} y \ne 0$.

Let

$$\Lambda := [\lambda_i := z^{\mathrm{t}} B^{i-1} : i = 1{:}q], \quad V := [v_j := B^{q-j} y : j = 1{:}q].$$

Then their Gramian

$$\Lambda^{\mathrm{t}} V = (\ z^{\mathrm{t}} B^{i-1+q-j} y\ : (i,j) \in \underline{q} \times \underline{q})$$

is upper triangular with all diagonal entries equal to $z^{\mathrm{t}} B^{q-1} y \ne 0$, hence is invertible.

This implies, by (5.8), that Y is the direct sum $Z_1 \dot{+} Z_2$, with

$$Z_1 := \operatorname{ran} V, \quad Z_2 = \operatorname{null} \Lambda^{\mathrm{t}}.$$

Further, we have $Bv_j = v_{j-1}$, hence $BV = [0, v_1, \ldots, v_{q-1}]$ and therefore $B(Z_1) \subset Z_1$. Analogously, having $z \in Z_2 = \operatorname{null} \Lambda^{\mathrm{t}}$ implies that Bz is already in $\operatorname{null}[\lambda_2, \ldots \lambda_q]^{\mathrm{t}} \subset Z_2$. Hence, both Z_1 and Z_2 are B-invariant, hence also A-invariant.

Therefore, by construction of Y, one of these two A-invariant summands must be trivial. Since Z_1 contains the nonzero vector y, it follows that $Y = Z_1$. In particular, V is a basis for Y. □

Claim 3: The matrix representation for $A|_Y$ with respect to the basis $V_Y := [B^{q-j}y : j = 1{:}q]$ for Y has the simple form

(12.16)
$$V_Y^{-1}(A|_Y)V_Y = \begin{bmatrix} \mu & 1 & 0 & \cdots & 0 & 0 \\ 0 & \mu & 1 & \cdots & 0 & 0 \\ 0 & 0 & \mu & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \mu & 1 \\ 0 & 0 & 0 & \cdots & 0 & \mu \end{bmatrix} =: J(\mu, q)$$

Proof: As before, set $v_j := B^{q-j}y$, for $j = 1{:}q$. Then

$$Av_j = (B + \mu \operatorname{id})v_j = Bv_j + \mu v_j = v_{j-1} + \mu v_j$$

with $v_0 := 0$. □

It follows that the matrix representation for A with respect to the basis

$$V := [V_Y : Y \in \mathcal{Y}]$$

for $X = \mathbb{F}^n$ is block-diagonal, with each diagonal block a **Jordan block**, $J(\mu, q)$, i.e., of the form (12.16) for some scalar μ and some natural number q. Any such matrix representation for A is called a **Jordan (canonical) form** for A.

There is no reason to believe that such a Jordan form is unique. After all, it depends on the particular order we choose for the elements of $\mathcal{Y}$ when we make up the basis $V = [V_Y : Y \in \mathcal{Y}]$. More than that, there is, in general, nothing unique about $\mathcal{Y}$. For example, if $A = 0$ or, more generally $A = \alpha\,\mathrm{id}$, then any direct sum decomposition for X is A-invariant, hence V can be any basis for X whatsoever for this particular A.

Nevertheless, the Jordan form is canonical in the following sense.

(12.17) Proposition: Let $\widehat{A} =: \mathrm{diag}(J(\mu_Y, \dim Y) : Y \in \mathcal{Y})$ be a Jordan canonical form for $A \in \mathbb{F}^{n\times n}$. Then

(i) $\mathrm{spec}(A) = \{\widehat{A}(j,j) : j = 1{:}n\} = \cup_{Y\in\mathcal{Y}} \mathrm{spec}(A|_Y)$.

(ii) For each $\mu \in \mathrm{spec}(A)$ and each q,

$$n_q := \dim \mathrm{null}(A - \mu\,\mathrm{id})^q = \sum_{\mu_Y = \mu} \min(q, \dim Y), \tag{12.18}$$

hence $\Delta n_q := n_{q+1} - n_q$ equals the number of blocks for μ of order $> q$, giving the decomposition-independent number $-\Delta^2 n_{q-1}$ for the number of Jordan blocks of order q for μ.

In particular, the Jordan form is unique up to an ordering of its blocks.

While the Jordan form is mathematically quite striking, it is of no practical relevance since it does not depend continuously on the entries of A, hence cannot be determined reliably by numerical calculations.

13. Localization of eigenvalues

In this short chapter, we discuss briefly the standard techniques for 'localizing' the spectrum of a given linear map A. Such techniques specify regions in the complex plane that must contain parts or all of the spectrum of A. To give a simple example, we proved (in (12.2)Corollary) that all the eigenvalues of a hermitian matrix must be real, i.e., that $\operatorname{spec}(A) \subset \mathbb{R}$ in case $A^c = A$.

Since $\mu \in \operatorname{spec}(A)$ iff $(A - \mu\,\mathrm{id})$ is not invertible, it is not surprising that many localization theorems derive from a test for invertibility.

Gershgorin's circles

Let μ be an eigenvalue for A with corresponding eigenvector x. Without loss of generality, we may assume that $\|x\| = 1$ in whatever vector norm on $X = \operatorname{dom} A$ we are interested in at the moment. Then

$$|\mu| = |\mu|\|x\| = \|\mu x\| = \|Ax\| \le \|A\|\|x\| = \|A\|,$$

with $\|A\|$ the corresponding map norm for A. This proves that the spectrum of A must lie in the ball $B_{\|A\|}$ of radius $\|A\|$ centered at the origin. In other words,

$$\rho(A) \le \|A\| \tag{13.1}$$

for any map norm $\|\cdot\|$.

For example, no eigenvalue of $A = \begin{bmatrix} 1 & 2 \\ -2 & -1 \end{bmatrix}$ can be bigger than 3 in absolute value since $\|A\|_\infty = 3$.

A more refined containment set is obtained by the following more refined analysis.

If $E \in \mathbb{F}^{n\times n}$ has map-norm < 1, then $A := \mathrm{id}_n - E$ is surely 1-1 since then

$$\|Ax\| = \|x - Ex\| \ge \|x\| - \|Ex\| \ge \|x\| - \|E\|\|x\| = \|x\|(1 - \|E\|)$$

with the factor $(1 - \|E\|)$ *positive*, hence $Ax = 0$ implies that $\|x\| = 0$.

Now consider a **diagonally dominant** A, i.e., a matrix A with the property that

$$\forall i \quad |A(i,i)| > \sum_{j\ne i} |A(i,j)|. \tag{13.2}$$

For example, of the three matrices

$$\begin{bmatrix} 2 & -1 \\ 2 & 3 \end{bmatrix}, \quad \begin{bmatrix} -2 & -1 \\ 3 & 3 \end{bmatrix}, \quad \begin{bmatrix} -2 & -1 \\ 4 & 3 \end{bmatrix}, \tag{13.3}$$

only the first is diagonally dominant. Setting

$$D := \operatorname{diag} A = \operatorname{diag}(\ldots, A(i,i), \ldots),$$

we notice that (i) D is invertible (since all its diagonal entries are nonzero); and (ii) the matrix

$$E := -D^{-1}(A - D) : (i,j) \mapsto \begin{cases} -A(i,j)/A(i,i) & \text{if } i \ne j; \\ 0 & \text{otherwise} \end{cases}$$

has norm

$$\|E\|_\infty = \max_i \sum_{j\ne i} |A(i,j)/A(i,i)| < 1,$$

by the assumed diagonal dominance of A. This implies that the matrix $\mathrm{id} - E$ is invertible, therefore, since

$$\mathrm{id}_n - E = \mathrm{id}_n + D^{-1}(A - D) = D^{-1}A,$$