Chapter V. The Diagonalization Problem.

V.1 The Characteristic Polynomial.

The **characteristic polynomial** $p_A(x)$ of an $n \times n$ matrix is defined to be

$$p_A(x) = \det(A - xI) \qquad (x \text{ an indeterminate})$$

This is a polynomial in $\mathbb{K}[x]$. In fact $\det(A - xI)$ is a polynomial combination of the entries in $(A - xI)$, so it follows that $p_A(x)$ does determine a polynomial in the single unknown x; furthermore $\deg(p_A) = n$. Given a linear operator $T : V \to V$ on a finite dimensional space V and a basis $\mathfrak{X}$ we have

$$[T - xI]_{\mathfrak{X}\mathfrak{X}} = [T]_{\mathfrak{X}\mathfrak{X}} - x\, I_{n\times n} \qquad (n = \dim(V))$$

so we may define a characteristic polynomial for T in the obvious way.

$$p_T(x) = \det(T - xI) = \det\big([T]_{\mathfrak{X}\mathfrak{X}} - xI_{n\times n}\big) \qquad (x \text{ an indeterminate})$$

The discussions for operators and matrices are so similar that nothing is lost if we focus on matrices for the time being.

Next observe what happens if we write out the characteristic polynomial p_A,

$$p_A(x) = \det(A - xI) = c_0(A) + c_1(A)x + \ldots + c_n(A)x^n \tag{37}$$

In this formula the coefficients $c_i(A)$ are scalar-valued functions from $\mathrm{M}(n, \mathbb{K}) \to \mathbb{K}$.

1.1. Lemma. *Each coefficient $c_k(A)$ in (37) is a similarity invariant on matrix space*

$$c_k\big(SAS^{-1}\big) = c_k(A) \qquad \textit{for all } A \in \mathrm{M}(n, \mathbb{K}), S \in \mathrm{GL}(n, \mathbb{K})$$

Furthermore, if we identify $\mathrm{M}(n, \mathbb{K})$ with n^2-dimensional coordinate space $\mathbb{K}^{n^2}$ via the correspondence $A \mapsto (a_{11}, \ldots, a_{1n}; \ldots; a_{n_1}, \ldots, a_{nn})$, each coefficient $c_i(A)$ is a polynomial function of the matrix entries: there is a polynomial $F_i \in \mathbb{K}[\mathbf{x}] = \mathbb{K}[x_1, \ldots, x_{n^2}]$ such that $c_i(A) = F_i\big(a_{11}, a_{12}, \ldots, a_{nn}\big)$.

Proof: We have

$$\begin{aligned} \det\big(S(A - xI)S^{-1}\big) &= \det(SAS^{-1} - x\,SS^{-1}) \;=\; \det(SAS^{-1} - x\,I) \\ &= c_0\big(SAS^{-1}\big) + c_1\big(SAS^{-1}\big)x + \ldots + c_n\big(SAS^{-1}\big)x^n \quad , \end{aligned}$$

while at the same time

$$\begin{aligned} \det\big(S(A - xI)S^{-1}\big) &= \det(S)\cdot\det(A - xI)\cdot\det(S^{-1}) \\ &= \det(A - xI) \;=\; c_0(A) + c_1(A)x + \ldots + c_n(A)x^n \end{aligned}$$

for all $x \in \mathbb{K}$. Since these are the same polynomial in $\mathbb{K}[x]$ the coefficients must agree, hence $c_i\big(SAS^{-1}\big) = c_i(A)$.

The polynomial nature of the coefficients as functions of $A \in \mathbb{K}^{n^2}$ follows because $\det(A - xI)$ is a polynomial combination of entries $(A - xI)_{ij}$; the coefficients $c_k(A)$ are then polynomial functions of the a_{ij} when like powers of the unknown "x" are gathered together. $\square$

It is interesting to examine how the coefficients $c_k(A)$ are obtained from entries in A. Starting from the original definition of the determinant in Chapter IV,

$$\det(B) = \sum_{\sigma\in S_n} sgn(\sigma)\cdot\Big(\prod_{i=1}^{n} b_{i,\sigma(i)}\Big)\ ,$$

if we take $B = A - xI$ we have

$$B\ =\ A - xI\ =\ \begin{pmatrix} (a_{11}-x) & a_{12} & . & a_{1n} \\ a_{12} & (a_{22}-x) & . & a_{2n} \\ \vdots & & \ddots & \vdots \\ a_{n1} & . & . & (a_{nn}-x) \end{pmatrix}$$

It is clear that the only template yielding a product $b_{1,\sigma(1)}...b_{n\sigma(n)}$ involving x^n is the diagonal template corresponding to the trivial permutation $\sigma = e$; furthermore, in expanding the product $\prod_i (a_{ii} - x)$ we must take the "$-x$" instead of "a_{ii}" from each factor to get the power x^n. Thus $c_n(A) \equiv (-1)^n$ is constant on matrix space (and certainly a similarity invariant).

We claim that

$$\begin{aligned} \det(A - xI) &= (-1)^n x^n + (\text{terms of lower degree}) \\ &= (-1)^n x^n + (-1)^{n-1}\mathrm{Tr}(A)\, x^{n-1} + \ldots + \det(A)\cdot 1 \end{aligned} \tag{38}$$

To get the coefficient of x^{n-1} observe that a product $\prod_i b_{i,\sigma(i)}$ involving x^{n-1} must come from a template having $(n-1)$ marked spots on the diagonal, but then *all* marked spots must lie on the diagonal and we are again dealing with the diagonal template (for $\sigma = e$). In expanding the product $\prod_i (a_{ii} - x)$ we must now select the "$-x$" from $n-1$ factors and the "a_{ii}" from just one. Thus

$$c_{n-1}(A) = (-1)^{n-1}\cdot\sum_{i=1}^{n} a_{ii} = (-1)^{n-1}\mathrm{Tr}(A)$$

as in (38). Determining the other coefficients is tricky business, except for the constant term which is

$$a_0(A) = \det(A)$$

This follows because *every* template yields a product that contributes to this constant term. However if a template marks a spot on the diagonal we must select the "a_{ii}" term rather than the "x" from that diagonal entry $(a_{ii} - x)$. It follows that the constant term in (38) is:

$$\sum_{\sigma\in S_n} sgn(\sigma)\cdot\prod_{i=1}^{n} a_{i,\sigma(i)} = \det(A)$$

as claimed. We leave discussion of other terms in the expansion (38) for more advanced courses.

Factoring Polynomials. It is well known that if a nonconstant polynomial $f(x)$ in $\mathbb{K}[x]$ has a root $\alpha \in \mathbb{K}$, so $f(\alpha) = \sum_{i=0}^{n} c_i\alpha^i = 0$, then we can factor $f(x) = (x-\alpha)\cdot g(x)$ by long division of polynomials, with $\deg(g) = \deg(f) - 1$. In fact, applying the Euclidean algorithm for division with remainder in $\mathbb{K}[x]$: if $P, Q \in \mathbb{K}[x]$ and $\deg(Q) \geq 1$ we can always write

$$P(x) = A(x)Q(x) + R(x) \qquad \text{(with remainder } R \equiv 0 \text{ or } \deg(R) < \deg(Q))$$

Taking P to be any nonconstant polynomial in $\mathbb{K}[x]$ and $Q = (x - \alpha)$, we get $f(x) = A(x) \cdot (x - \alpha) + R(x)$ where $R(x)$ is either the zero polynomial, or $R(x)$ is nonzero with $\deg(R) < \deg(x - \alpha) = 1$ – i.e. $R(x)$ is then a nonzero *constant* polynomial $R = c\mathbb{1}$. If $\alpha \in \mathbb{K}$ is a root of f, replacing x by α everywhere yields the identity

$$0 = f(\alpha) = A(\alpha)\cdot(\alpha - \alpha) + R(\alpha) = 0 + R(\alpha) = R(\alpha)$$

Since $R = c\mathbb{1}$, this forces $R(x) \equiv 0$ and $f(x) = A(x)(x - \alpha)$ with no remainder – i.e. $(x - \alpha)$ divides $f(x)$ exactly.

If α_1 is a root of f we may split $f(x) = (x - \alpha_1) \cdot g_1(x)$. If we can find a root α_2 of $g_1(x)$ in $\mathbb{K}$ we can continue this process, obtaining $f(x) = (x - \alpha_1)(x - \alpha_2) \cdot g_2(x)$. Pushing this as far as possible we arrive at a factorization

$$f(x) = \prod_{i=1}^{s} (x - \alpha_i) \cdot g(x)$$

in which $g(x)$ has *no* roots in $\mathbb{K}$. We say that f **splits completely over** $\mathbb{K}$ if g reduces to a constant polynomial, so that $f(x) = c \prod_{i=1}^{n} (x - \alpha_i)$. There may be repeated factors, and if we gather together all factors of the same type this becomes

$$f(x) = c \prod_{j=1}^{r} (x - \alpha_i)^{m_j} \qquad (\alpha_i \in \mathbb{K})$$

The roots $\{\alpha_1, \ldots, \alpha_r\}$ are now *distinct* and the exponents $m_i \geq 1$ are their **multiplicities** as roots of $f(x)$; the constant c out front is the coefficient of the leading term $c_n x^n$ in $f(x)$.

1.2. Corollary. *A nonconstant polynomial $f(x) \in \mathbb{K}[x]$ can have at most $n = \deg(f)$ roots in any field of coefficients $\mathbb{K}$. More generally the sum of the multiplicities of the roots in $\mathbb{K}$ is at most n.*

Proof: If $f, g \neq 0$ in $\mathbb{K}[x]$ (so they have well defined degrees) we know that

$$\deg(f(x) + g(x)) = \deg(f(x)) + \deg(g(x))$$

But, $\deg\big(\prod_{i=1}^{r} (x - \alpha_i)^{m_i}\big) = \sum_{i=1}^{r} m_i$, so

$$r = \#(\text{distinct roots}) \leq (m_1 + \ldots + m_r) + \deg(g) = \deg(f) \quad \square$$

1.3. Exercise. If $f(x), h(x)$ are nonzero polynomials over any field, explain why the "degree formula"

$$\deg(f(x)h(x)) = \deg(f(x)) + \deg(h(x))$$

is valid.

1.4. Exercise. Verify that if $f(x) = \prod_{i=1}^{r} (x - \alpha_i) \cdot g(x)$ and $g(x)$ has no roots in $\mathbb{K}$, then the roots of f in $\mathbb{K}$ are $\{\alpha_1, ..., \alpha_r\}$.
Note: Repetitions are allowed; $f(x)$ might even have the form $(x - \alpha)^r \cdot g(x)$.)

1.5. Definition. *The distinct roots $\{\alpha_1, \ldots, \alpha_r\}$ in $\mathbb{K}$ of a nonconstant polynomial and their multiplicities are uniquely determined, and the set of roots is called the* **spectrum** *of the polynomial f and is denoted by $\mathrm{sp}_{\mathbb{K}}(f)$.*

Over the field $K = \mathbb{C}$ of complex numbers we have:

1.6. Theorem (Fundamental Theorem of Algebra). *If f is a nonconstant polynomial in $\mathbb{C}[x]$ then f has a root $\alpha \in \mathbb{C}$, so that $f(\alpha) = 0$.*

1.7. Corollary. Every non constant $f \in \mathbb{C}[x]$ splits completely over $\mathbb{C}$, with

$$f(x) = c \cdot \prod_{i=1}^{r} (x - \alpha_i)^{m_i} \qquad \text{where } m_1 + ... + m_r = n$$

Proof: Since f has a root we may factor $f = (x - \alpha_1) \cdot g_1(x)$. Unless $g_1(x)$ is a constant it also has a root, allowing us to write $f = (x - \alpha_1)(x - \alpha_2)g_2(x)$. Continue recursively. $\square$

Over $\mathbb{K} = \mathbb{R}$ or $\mathbb{Q}$, things get more complicated and $f(x)$ might not have any roots at all in $\mathbb{K}$. For example if $f(x) = x^2 + 1$ over $\mathbb{R}$, or $f(x) = x^2 - 2$ over $\mathbb{Q}$ since $\mathbb{Q}$ does not contain any element α such that $\alpha^2 = 2$ (there is no "square root of 2" in $\mathbb{Q}$). Nevertheless since $\mathbb{R} \subseteq \mathbb{C}$ we may regard any $f \in \mathbb{R}[x]$ as a complex polynomial that happens to have all real coefficients. All real roots α remain roots $\alpha + i0$ in $\mathbb{C}$ (lying on the real axis), but enough new roots appear in the larger field to split f completely as

$$f(x) = c \cdot \prod (x - \alpha_i) \qquad \text{with } \alpha_i \in \mathbb{C}$$

It is important to realize that the new non-real roots enter in "conjugate pairs."

1.8. Lemma. *If $f(x)$ is nonconstant in $\mathbb{R}[x]$ and $z = x + iy$ is a complex root when we identify $\mathbb{R} \subseteq \mathbb{C}$ and $\mathbb{R}[x] \subseteq \mathbb{C}[x]$, then the complex conjugate $\overline{z} = x - iy$ is also a root.*

Proof: There is nothing to prove if z is real ($y = 0$). Otherwise, recall that conjugation

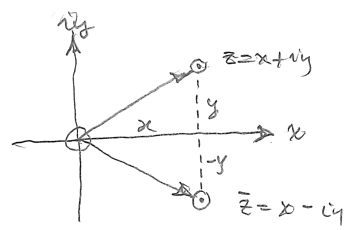

Figure 5.1. Non-real roots of a polynomial with real coefficients come in conjugate pairs $z = x + iy$ and $\overline{z} = x - iy$, mirror images of each other under reflection across the x-axis.

has the following algebraic properties.

$$\overline{z_1 + z_2} = \overline{z_1} + \overline{z_2} \qquad \text{and} \qquad \overline{z_1 z_2} = \overline{z_1}\,\overline{z_2}$$

Then

$$\overline{z^n} = \overline{z}^n \qquad \text{for all } n \in \mathbb{Z} \text{ and } z \in \mathbb{C}$$

Hence if $0 = f(z) = \sum_{j=0} c_j z^j$ with c_j real we have

$$\overline{(c_j z^j)} = \overline{c_j}\,\overline{(z^j)} = c_j \overline{z}^j$$

and

$$0 = \overline{0} = \overline{f(z)} = \sum_{j=0} \overline{(c_j z^j)} = \sum_{j=0} c_j \big(\overline{z}\big)^j = f(\overline{z})$$

Hence, $\overline{z}$ is also a root in $\mathbb{C}$. $\square$

The real roots of $f \in \mathbb{R}[x]$ are not subject to any constraints; in fact, all the roots might be real. The number of distinct non-real roots is always *even.*

1.9. Example. If $f \in \mathbb{K}[x]$ is quadratic,

$$f(x) = ax^2 + bx + c \qquad \text{with } a \neq 0,$$

the quadratic formula continues to apply for all fields except those of "*characteristic* 2," in which $2 = 1 + 1$ is equal to 0 (for insatnce $\mathbb{K} = \mathbb{Z}_2$). Except for this, the roots are given by:

$$\text{Quadratic Formula:} \qquad z_{\pm} = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

If the $\sqrt{\ldots}$ fails to exist in $\mathbb{K}$ the proper conclusion is that $f(x)$ has no roots in $\mathbb{K}$. If $\mathbb{K} = \mathbb{Q}$ or $\mathbb{R}$ this formula gives the correct roots in $\mathbb{C}$ even if there are no roots in $\mathbb{K}$.

Discussion: Complete the square. Adding/subtracting a suitable constant d we may write

$$\begin{aligned} ax^2 + bx + c &= a\Big(x^2 + \frac{b}{a}x + \frac{c}{a}\Big) = a\Big[\Big(x^2 + \frac{b}{a}x + d\Big) + \Big(\frac{c}{a} - d\Big)\Big] \\ &= a \cdot \Big(x^2 + \frac{b}{a}x + d\Big) + (c - ad) \end{aligned}$$

To make $x^2 + (b/a)x + d$ a "perfect square" of the form $(x+k)^2 = (x^2 + 2kx + k^2)$, we must take $k = b/(2a)$ and $d = k^2 = (b^2/4a^2)$. Then $c - ad = c - (b^2/4a^2)$, so that

$$0 = ax^2 + bx + c = a\Big(x + \frac{b}{2a}\Big)^2 + \Big(\frac{c}{a} - \frac{b^2}{4a^2}\Big) = a \cdot \Big(x + \frac{b}{2a}\Big)^2 + \Big(\frac{4ac - b^2}{4a}\Big)$$

This happens if and only if

$$a\Big(x + \frac{b}{2a}\Big)^2 = \Big(\frac{b^2 - 4ac}{4a}\Big)$$

if and only if

$$\Big(x + \frac{b}{2a}\Big)^2 = \frac{b^2 - 4ac}{4a^2}$$

if and only if

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} \qquad \Box$$

1.10. Example. Here are some examples of factorization of polynomials.

1. $x^2 - 1 = (x-1)(x+1)$ splits over $\mathbb{R}$, with two roots $+1, -1$ each of multiplicity one. On the other hand $x^2 + 1$ has no roots and does not split over $\mathbb{R}$, but it does split over $\mathbb{C}$, with $x^2 + 1 = (x - i)(x + i)$.

2. $x^2 + 2x + 1$ splits over $\mathbb{R}$ as $(x-1)^2$, but there is just one root, of multiplicity 2;

3. $x^3 - x^2 + x - 1$ has a root $x = 1$ in $\mathbb{R}$. Long division yields a quadratic,

   $$x^3 - x^2 + x - 1 = (x-1)(x^2+1)$$

   Over $\mathbb{R}$, there is just one root $\lambda_1 = 1$ with multiplicity $m(\lambda_1) = 1$; over $\mathbb{C}$ we get $x^2 + 1 = (x+i)(x-i)$ so there are two more roots roots $\lambda_2 = i$, $\lambda_3 = -i$ in the larger field $\mathbb{C}$.

4. $x^4 - 1 = (x^2-1)(x^2+1) = (x+1)(x-1)(x+i)(x-i)$.

5. $x^3 + x + 1$ has just one real root λ_1 because it is a strictly increasing function of $x \in \mathbb{R}$, and since it goes to $\pm\infty$ as $x \to \pm\infty$ it must cross the x-axis somewhere. But λ_1 is not so easy to write as an explicit algebraic expression involving sums, products, quotients, and cube roots. Such formulas exist, but are algorithms with possible branch points rather than simple expressions like the quadratic formula. A numerical estimate yields the real root $\lambda_1 = -0.6823 + i0$. There is a conjugate pair of complex roots $\lambda_2 = 0.3412 + 1.615\, i$ and $\lambda_3 = 0.3412 - 1.615\, i$, which could be found by (numerically) long dividing $f(x)$ by $(x - \lambda_1)$ and applying the quadratic formula to find the complex roots of the resulting quadratic.

V.2. Finding Eigenvalues.

If V is a finite dimensional vector space we say $\lambda \in \mathbb{K}$ is an **eigenvalue** for a linear operator $T : V \to V$ if there is $v \neq 0$ such that $T(v) = \lambda v$. For any $\lambda \in \mathbb{K}$ the **λ-eigenspace** is $E_\lambda = \{v \in V : (T - \lambda I)v = 0\}$. This vector subspace is nontrivial if and only if λ is an eigenvalue. The set of distinct eigenvalues is called the **spectrum** $\mathrm{sp}_{\mathbb{K}}(T)$ of the operator. When $\lambda = 0$ the eigenspace $E_{\lambda=0}(T)$ is just $\ker(T) = \{v \in V : T(v) = 0\}$ and when $\lambda = 1$ we get the subspace of *fixed vectors* $E_{\lambda=1}(T) = \{v : T(v) = v\}$.

The connection with determinants now emerges: $\lambda \in \mathbb{K}$ is an eigenvalue if and only if

$$\ker(T - \lambda I) \neq (0) \;\Leftrightarrow\; (T - \lambda I) \text{ is singular } \;\Leftrightarrow\; \det(T - \lambda I) = 0$$

Thus the eigenvalues are the roots in $\mathbb{K}$ of the characteristic polynomial $p_T \in \mathbb{K}[x]$.

2.1. Definition. *If $T : V \to V$ is a linear operator on a finite dimensional vector space then $\mathrm{sp}_{\mathbb{K}}(T)$ is the set of distinct roots in $\mathbb{K}$ of the characteristic polynomial $p_T(x) = \det(T - xI)$. We define the* **geometric multiplicity** *of an eigenvalue to be $\dim(E_\lambda)$; its* **algebraic multiplicity** *is the multiplicity of λ as a root of the characteristic polynomial, so that $p_T(x) = (x - \lambda)^m \cdot g(x)$ and $g(x)$ does not have λ as a root.*

2.2. Lemma. *Over any field $\mathbb{K}$,*

$$(\textit{algebraic multiplicity of } \lambda) \geq (\textit{geometric multiplicity})$$

Over $\mathbb{K} = \mathbb{C}$, the sum of the algebraic multiplicities of the (distinct) eigenvalues in $\mathrm{sp}_{\mathbb{C}}(T) = \{\lambda_1, ..., \lambda_r\}$ is $m(\lambda_1) + \ldots + m(\lambda_r) = n = \dim_{\mathbb{C}}(V)$.

Proof: Every eigenspace E_λ is T-invariant because $(T - \lambda I)T(v) = T(T - \lambda I)v = 0$ for $v \in E_\lambda$. This eigenspace has a basis of eigenvectors $\mathfrak{X}_\lambda = \{e_1, \ldots, e_d\}$, with respect to which

$$[T]_{\mathfrak{X}_\lambda} = \begin{pmatrix} \lambda & & & 0 \\ & \lambda & & \\ & & \ddots & \\ 0 & & & \lambda \end{pmatrix}$$

(diagonal). Extending $\mathfrak{X}_\lambda$ to a basis $\mathfrak{X} = \{e_1, ..., e_d, e_{d+1}, ..., e_n\}$ for all of V, we get

$$[T]_{\mathfrak{X}\mathfrak{X}} = \left(\begin{array}{ccc|c} \lambda & & & \\ & \ddots & & * \\ 0 & & \lambda & \\ \hline & 0 & & B \end{array}\right)$$

which implies that

$$[T - x\,I]_{\mathfrak{X}\mathfrak{X}} = \left(\begin{array}{ccc|c} \lambda - x & & & \\ & \ddots & & * \\ 0 & & \lambda - x & \\ \hline & 0 & & B - \lambda I \end{array}\right)$$

2.3. Lemma. *If A is of the form*

$$A = \left(\begin{array}{c|c} B & D \\ \hline 0 & C \end{array}\right)$$

where B and C are two square matrices, then $\det(A) = \det(B) \cdot \det(C)$.

Proof: If B is $m \times m$, a sequence of Type II and III row operations on rows $R_1, \ldots, R_m$ puts this block in upper triangular form; similar operations on rows $R_{m+1}, \ldots, R_n$ puts block C in upper triangular form without affecting any of the earlier rows. The net result is an echelon form $A' = [B', *; 0, C']$ for which $\det A' = \det(B') \cdot \det(C')$. Each of the determinants $\det(A'), \ldots, \det(C')$ differs from its counterpart by a $\pm$ sign; furthermore, the same moves that put B and C in upper triangular form also put A in upper triangular form when applied to the whole $n \times n$ matrix. We leave the reader to check that the sign changes cancel and yield $\det(A) = \det(B) \cdot \det(C)$. $\square$

This can also be seen by noting that if a template contributes to $\det(A)$, every column passing through block B must be marked at a spot in B; otherwise it would marked at a spot below B, whose entry is $= 0$. Likewise for the rows that meet block C, so a template contributes $\Leftrightarrow$ it has the form in Figure 5.2.

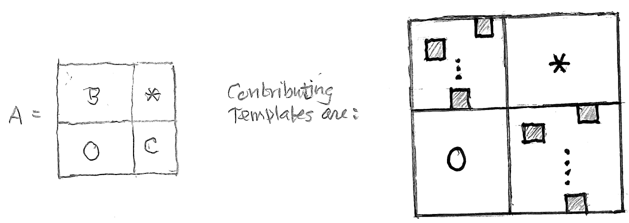

Figure 5.2. If A is a block upper-triangular square matrix, then $\det(A) = \det(B)\cdot\det(C)$ and the only templates that contribute to $\det(A)$ are those whose marked spots lie entirely within the blocks B and C.

Applying Lemma 2.3 we can complete the proof of Lemma 2.2. We now see that

$$p_T(x) = \det(T - x\,I) = (\lambda - x)^m \cdot Q(x) \qquad \text{where } Q(x) = \det(B - x\,I)$$

Obviously $\deg(Q(x)) = n - m$ and $p_T(x)$ has λ as a root of multiplicity at least m, so (*algebraic multiplicity of* λ) $\geq m = \dim(E_\lambda)$ as claimed. $\square$

It might still be possible for $(x - \lambda)$ to divide $Q(x)$, making the algebraic multiplicity larger than $\dim\big(E_\lambda\big)$. A good example is $A = \begin{pmatrix} \lambda & 1 \\ 0 & \lambda \end{pmatrix}$. The operator $L_A : \mathbb{R}^2 \to \mathbb{R}^2$ has $\dim\big(E_{\lambda=1}\big) = 1$, but $p_T(\lambda) = (\lambda - x)^2$ so the algebraic multiplicity is 2.

The following example illustrates the complete diagonalization process.

2.4. Example. Let $T = L_A : \mathbb{R}^3 \to \mathbb{R}^3$ with

$$A = \begin{pmatrix} 4 & 0 & 1 \\ 2 & 3 & 2 \\ 1 & 0 & 4 \end{pmatrix}$$

If $\mathfrak{X} = \{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ is the standard basis in $\mathbb{R}^3$ we have $[T]_{\mathfrak{X}\mathfrak{X}} = [L_A]_{\mathfrak{X}\mathfrak{X}} = A$ as in Exercise 4.13 of Chapter II, so

$$\begin{aligned} p_A(x) &= \det(A - x\,I) = \det \begin{pmatrix} 4-x & 0 & 1 \\ 2 & 3-x & 2 \\ 1 & 0 & 4-x \end{pmatrix} \\ &= \big[(4-x)(3-x)(4-x) + 0 + 0\big] - \big[(3-x) + 0 + 0\big] \\ &= -x^3 + 11x^2 - 39x + 45 \end{aligned}$$

We are looking for roots of a cubic equation. If you can guess a root α, then long divide by $x - \alpha$ to get $p_T(x) = (x - \alpha)\cdot(\textit{quadratic})$; otherwise you will have to use a numerical root-finding program. Trial and error reveals that $x = 3$ is a root and long division by $(x-3)$ yields

$$\begin{array}{r|rrrr} & -x^2 & +8x & -15 & \\ \hline x-3) & -x^3 & 11x^2 & -39x & +45 \\ & -x^3 & +3x^2 & & \\ \hline & & 8x^2 & -39x & +45 \\ & & 8x^2 & -24x & \\ \hline & & & -15x & +45 \\ & & & -15x & +45 \\ \hline & & & & 0 \end{array}$$

Then

$$\begin{aligned} -x^3 + 11x^2 - 39x + 45 &= (x-3)(-x^2 + 8x - 15) \\ &= -(x-3)(x-5)(x-3) = -(x-3)^2(x-5) \quad , \end{aligned}$$

so $\mathrm{sp}(A) = \{3, 5\}$ with algebraic multiplicities $m_{\lambda=3} = 2$, $m_{\lambda=5} = 1$. To determine the eigenspaces and geometric multiplicities we must solve systems of equations.

Eigenvalue $\lambda_1 = 3$: We must solve the matrix equation $(A - 3I)X = 0$. Row operations on $[A - 3I\,|\,0]$ yield

$$[A - 3I\,|\,0] = \left(\begin{array}{ccc|c} 1 & 0 & 1 & 0 \\ 2 & 0 & 2 & 0 \\ 1 & 0 & 1 & 0 \end{array}\right) \to \left(\begin{array}{ccc|c} 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right)$$

Solutions: x_2, x_3 are free variables and $x_1 = -x_3$, so

$$X = \begin{pmatrix} -x_3 \\ x_2 \\ x_3 \end{pmatrix} \qquad \text{and} \qquad E_{\lambda=3} = \ker(A - 3I) = \mathbb{R}\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} + \mathbb{R}\begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}$$

Thus $\lambda = 3$ has geometric multiplicity $\dim\big(E_{\lambda=5}\big) = 2$.

Eigenvalue $\lambda_2 = 5$: Solve matrix equation $(A - 5I)X = 0$. Row operations on $[A - 5I\,|\,0]$ yield

$$[A - 5I\,|\,0] = \left(\begin{array}{ccc|c} -1 & 0 & 1 & 0 \\ 2 & -2 & 2 & 0 \\ 1 & 0 & -1 & 0 \end{array}\right) \to \left(\begin{array}{ccc|c} 1 & 0 & -1 & 0 \\ 0 & 1 & -2 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right)$$

Solutions: x_3 is a free variable; $x_2 = 2x_3$, $x_1 = x_3$. So

$$X = \begin{pmatrix} x_3 \\ 2x_3 \\ x_3 \end{pmatrix} \quad \text{and} \quad E_{\lambda=5} = \ker(A - 5I) = \mathbb{R}\begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}$$

Thus $\lambda = 5$ has geometric multiplicity $\dim\big(E_{\lambda=5}\big) = 1$.

We showed earlier that the span $M = \sum_{\lambda \in \mathrm{sp}(T)} E_\lambda(T)$ of the eigenspaces of a linear operator is actually a *direct* sum $M = E_{\lambda_1} \oplus \ldots \oplus E_{\lambda_r}$. In the present situation $M = E_{\lambda=3} \oplus E_{\lambda=5}$ is all of V since the dimension add up to $\dim(V) = 3$. Taking a basis $\mathfrak{Y} = \{\mathbf{f}_1, .., \mathbf{f}_3\}$ that runs first through $E_{\lambda=3} = \mathbb{R}\mathbf{f}_1 \oplus \mathbb{R}\mathbf{f}_2$, and then through $E_{\lambda=5} = \mathbb{R}\mathbf{f}_3$, we obtain a diagonal matrix

$$[T]_{\mathfrak{Y}\mathfrak{Y}} = \left(\begin{array}{cc|c} 3 & 0 & 0 \\ 0 & 3 & 0 \\ \hline 0 & 0 & 5 \end{array}\right)$$

Once we have found the diagonalizing basis

$$\mathfrak{Y} = \{\mathbf{f}_1 = (0,1,0), \mathbf{f}_2 = (-1,0,1), \mathbf{f}_3 = (1,2,1)\}$$

we determine an invertible matrix Q such that $QAQ^{-1} = [T]_{\mathfrak{Y}\mathfrak{Y}} = \mathrm{diag}(3,3,5)$. To find Q recall that

$$[T]_{\mathfrak{Y}\mathfrak{Y}} = [\mathrm{id}]_{\mathfrak{Y}\mathfrak{X}} \cdot [T]_{\mathfrak{X}\mathfrak{X}} \cdot [\mathrm{id}]_{\mathfrak{X}\mathfrak{Y}} = [\mathrm{id}]_{\mathfrak{Y}\mathfrak{X}} \cdot A \cdot [\mathrm{id}]_{\mathfrak{X}\mathfrak{Y}} = QAQ^{-1}$$

Here $[\mathrm{id}]_{\mathfrak{X}\mathfrak{Y}} = Q^{-1}$ and $[\mathrm{id}]_{\mathfrak{Y}\mathfrak{X}} = [\mathrm{id}]_{\mathfrak{X}\mathfrak{Y}}^{-1}$, and by definition $[\mathrm{id}]_{\mathfrak{Y}\mathfrak{X}}$ is the transpose of the coefficient array in the system of vector identities

$$\begin{aligned} \mathbf{f}_1 &= [\mathrm{id}]\ \mathbf{f}_1 = 0 + \mathbf{e}_2 + 0 \\ \mathbf{f}_2 &= [\mathrm{id}]\ \mathbf{f}_2 = -\mathbf{e}_1 + 0 + \mathbf{e}_3 \\ \mathbf{f}_3 &= [\mathrm{id}]\ \mathbf{f}_3 = \mathbf{e}_1 + 2\mathbf{e}_2 + \mathbf{e}_3 \end{aligned}$$

Thus,

$$Q^{-1} = [\mathrm{id}]_{\mathfrak{X}\mathfrak{Y}} = \begin{pmatrix} 0 & -1 & 1 \\ 1 & 0 & 2 \\ 0 & 1 & 1 \end{pmatrix}$$

and $Q = (Q^{-1})^{-1}$ can be found efficiently via row operations.

$$\left(\begin{array}{ccc|ccc} 0 & -1 & 1 & 1 & 0 & 0 \\ 1 & 0 & 2 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 \end{array}\right) \to \left(\begin{array}{ccc|ccc} 1 & 0 & 2 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & \frac{1}{2} & 0 & \frac{1}{2} \end{array}\right) \to \left(\begin{array}{ccc|ccc} 1 & 0 & 0 & -1 & 1 & -1 \\ 0 & 1 & 0 & -\frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 0 & 1 & \frac{1}{2} & 0 & \frac{1}{2} \end{array}\right)$$

Thus

$$Q = \begin{pmatrix} -1 & 1 & -1 \\ -\frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} \end{pmatrix} = \frac{1}{2}\begin{pmatrix} -2 & 2 & -2 \\ -1 & 0 & 1 \\ 1 & 0 & 1 \end{pmatrix}$$

and

$$QAQ^{-1} = \begin{pmatrix} 3 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 5 \end{pmatrix}$$

as expected. That completes the "spectral analysis" of A. $\square$

The same sort of calculations determine the eigenspaces in $\mathbb{C}^2$ when $A \in \mathrm{M}(n, \mathbb{R})$ is regarded as a matrix in $\mathrm{M}(n, \mathbb{C})$.

2.5. Example. Diagonalize the operator $L_A : \mathbb{K}^2 \to \mathbb{K}^2$ where

$$A = \begin{pmatrix} 2 & 4 \\ -1 & -2 \end{pmatrix}$$

over $\mathbb{C}$ and over $\mathbb{R}$ (insofar as this is possible).

Discussion: The characteristic polynomial of A (or L_A) is

$$p_A(\lambda) = \det \begin{pmatrix} 2-\lambda & 4 \\ -1 & -2-\lambda \end{pmatrix} = -(2-\lambda)(2+\lambda) + 4 = -4 + \lambda^2 + 4 = \lambda^2$$

The only root (real or complex) is $\lambda = 0$ so $\mathrm{sp}_{\mathbb{R}}(A) = \mathrm{sp}_{\mathbb{C}}(A) = \{0\}$. Its algebraic multiplicity is 2, but the geometric multiplicity $\dim_{\mathbb{K}} \left(E_{\lambda=0}\right)$ is equal to 1. The outcome is the same over $\mathbb{C}$ and $\mathbb{R}$.

Eigenvalue $\lambda = 0$. Here $E_{\lambda=0} = \ker(A)$. Row operations on $[A \,|\, 0]$ yield

$$\left(\begin{array}{cc|c} 2-\lambda & 4 & 0 \\ -1 & -2-\lambda & 0 \end{array}\right) \to \left(\begin{array}{cc|c} 2 & 4 & 0 \\ -1 & -2 & 0 \end{array}\right) \to \left(\begin{array}{cc|c} 2 & 4 & 0 \\ 0 & 0 & 0 \end{array}\right)$$

Solutions: In solving $(A - \lambda I)X = AX = 0$, x_2 is a free variable and $x_1 = -2x_2$ so

$$X = \begin{pmatrix} -2x_2 \\ x_2 \end{pmatrix} \quad \text{and} \quad E_{\lambda=0} = \mathbb{K} \cdot \begin{pmatrix} -2 \\ 1 \end{pmatrix} .$$

Since there are no other eigenvalues, the best we can do in trying to find a simple matrix description $[T]_{\mathfrak{Y}\mathfrak{Y}}$ is to take a basis $\mathfrak{Y} = \{\mathbf{f}_1, \mathbf{f}_2\}$ that passes first through $E_{\lambda=0}$: let $\mathbf{f}_1 = (-2, 1)$ and then include one more vector $\mathbf{f}_2 \notin \mathbb{K}\mathbf{f}_1$ to make a basis. We have

$$[T]_{\mathfrak{X}\mathfrak{X}} = \begin{pmatrix} 2 & 4 \\ -1 & -2 \end{pmatrix}$$

with respect to the standard basis $\mathfrak{X} = \{\mathbf{e}_1, \mathbf{e}_2\}$ in $\mathbb{K}^2$ (recall Exercise 4.13 of Chapter II). With respect to the basis $\mathfrak{Y} = \{\mathbf{f}_1, \mathbf{f}_2\}$ the matrix has block diagonal form,

$$[T]_{\mathfrak{Y}\mathfrak{Y}} = \begin{pmatrix} 0 & * \\ 0 & * \end{pmatrix}$$

But this operator cannot be diagonalized by any choice of basis. $\square$

2.6. Exercise. We have shown that there is a basis $\mathfrak{Y} = \{\mathbf{f}_1, \mathbf{f}_2\}$ such that

$$A = [T]_{\mathfrak{Y}\mathfrak{Y}} = \begin{pmatrix} 0 & a \\ 0 & b \end{pmatrix}$$

1. Prove that b must be 0, so
$$A = \begin{pmatrix} 0 & a \\ 0 & 0 \end{pmatrix}$$

2. Explain how to modify the basis $\mathfrak{Y}$ to get a new basis $\mathfrak{Z}$ such that
$$[T]_{\mathfrak{Z}\mathfrak{Z}} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$$

2.7. Example. The matrix

$$A = \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix} \qquad (\theta \text{ real})$$

yields an operator $L_A : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ that you will recognize as a rotation counter clockwise about the origin by θ radians. Describe its eigenspaces over $\mathbb{R}$ and over $\mathbb{C}$.

Solution: Over either field we have

$$\begin{aligned} p_A(\lambda) &= \det(A - \lambda I) \;=\; \det \begin{pmatrix} \cos(\theta) - \lambda & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) - \lambda \end{pmatrix} \\ &= \big(\cos(\theta) - \lambda\big)^2 + \sin^2(\theta) \;=\; \cos^2(\theta) + \sin^2(\theta) - 2\lambda\cos(\theta) + \lambda^2 \\ &= \lambda^2 - 2\lambda\cos(\theta) + 1 \end{aligned}$$

This is zero only when

$$\begin{aligned} \lambda &= \frac{2\cos(\theta) \pm \sqrt{4\cos^2(\theta) - 4}}{2} \;=\; \cos(\theta) \pm \sqrt{\cos^2(\theta) - 1} \\ &= \cos(\theta) \pm i\sqrt{1 - \cos^2(\theta)} \;=\; \cos(\theta) \pm i\sin(\theta) \;=\; e^{\pm i\theta} \end{aligned}$$

The roots are non-real (hence a conjugate pair as shown earlier in Figure 5.1), and they lie on the unit circle in $\mathbb{C}$ because $|e^{\pm i\theta}| = \sin^2(\theta) + \cos^2(\theta) = 1$ for all θ. When $\theta = 0$ or π we have $\lambda = \pm 1 + i0$ (real), and in this case $A = I$ or $-I$. In all other cases A has no real eigenvalues at all, but it can be diagonalized as

$$[L_A]_{\mathfrak{YY}} = \begin{pmatrix} e^{i\theta} & 0 \\ 0 & e^{-i\theta} \end{pmatrix}$$

for a suitably chosen complex basis $\mathfrak{Y} = \{\mathbf{f}_1, \mathbf{f}_2\}$ in $\mathbb{C}^2$. To find it we need to determine the eigenspaces of L_A in $\mathbb{C}^2$.

Eigenvalue: $\lambda_1 = e^{i\theta} = \cos(\theta) + i\sin(\theta)$.

$$\begin{aligned} [A - \lambda I] &= \begin{pmatrix} \cos(\theta) - e^{i\theta} & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) - e^{i\theta} \end{pmatrix} = \begin{pmatrix} -i\sin(\theta) & -\sin(\theta) \\ \sin(\theta) & -i\sin(\theta) \end{pmatrix} \\ &= \sin(\theta) \cdot \begin{pmatrix} -i & -1 \\ 1 & -i \end{pmatrix} \end{aligned}$$

Now, $(A - \lambda I)X = 0 \Leftrightarrow BX = 0$ where $B = \begin{pmatrix} -i & -1 \\ 1 & -i \end{pmatrix}$. Row operations yield:

$$\left(\begin{array}{cc|c} -i & -1 & 0 \\ 1 & -i & 0 \end{array}\right) \rightarrow \left(\begin{array}{cc|c} 1 & -i & 0 \\ 1 & -i & 0 \end{array}\right) \rightarrow \left(\begin{array}{cc|c} 1 & -i & 0 \\ 0 & 0 & 0 \end{array}\right)$$

Solutions: Here x_2 is a free variable and $x_1 = ix_2$. So,

$$X = \begin{pmatrix} ix_2 \\ x_2 \end{pmatrix} \quad \text{and} \quad E_{\lambda_1 = e^{it}} = \mathbb{C} \cdot \begin{pmatrix} i \\ 1 \end{pmatrix} \quad .$$

For λ_1, (*algebraic multiplicity*) = (*geometric multiplicity*) = 1.

The discussion for the conjugate eigenvalue $\lambda_2 = e^{-i\theta} = \cos(\theta) - i\sin(\theta)$ is almost the same, with the final result that $E_{\lambda = e^{-i\theta}} = \mathbb{C} \cdot \text{col}(-i, 1)$ Combining these observations we get

$$\mathbb{C}^2 = E_{\lambda = e^{i\theta}} \oplus E_{\lambda = e^{-i\theta}} = \mathbb{C} \cdot \begin{pmatrix} i \\ 1 \end{pmatrix} \oplus \mathbb{C} \cdot \begin{pmatrix} -i \\ 1 \end{pmatrix} = \mathbb{C}\mathbf{f}_1 \oplus \mathbb{C}\mathbf{f}_2$$

Thus, with respect to the basis

$$\mathfrak{Y} = \{ \mathbf{f}_1 = \mathrm{col}(i,1), \mathbf{f}_2 = \mathrm{col}(-i,1) \}$$

we have

$$[L_A]_{\mathfrak{Y}\mathfrak{Y}} = \begin{pmatrix} e^{i\theta} & 0 \\ 0 & e^{-i\theta} \end{pmatrix} \quad \square$$

As mentioned, the span $M = \sum_{\lambda \in \mathrm{sp}_{\mathbb{K}}(A)} E_\lambda$ is a direct sum $E_{\lambda_1} \oplus \ldots \oplus E_{\lambda_r}$ and a suitable chosen basis partially diagonalizes A, with matrix

$$[T]_{\mathfrak{Y}\mathfrak{Y}} = \left(\begin{array}{ccc|c} \boxed{\lambda_1 \cdot I_{d_1 \times d_1}} & & 0 & \\ & \ddots & & * \\ & & \boxed{\lambda_r \cdot I_{d_r \times d_r}} & \\ \hline 0 & & 0 & B \end{array}\right)$$

To proceed further and determine the structure of the lower right-hand block B we would have to develop the theory of nilpotent operators, generalized eigenspaces, and the Jordan decomposition of a linear operator over $\mathbb{C}$. We must leave all that for a subsequent course. However the following observation can be useful.

2.8. Proposition. *If $\dim_{\mathbb{K}}(V) = n$ and $T : V \rightarrow V$ has n distinct eigenvalues in $\mathbb{K}$, then T is diagonalizable and V is the direct sum $\bigoplus_{i=1}^{n} E_{\lambda_i}$. of 1-dimensional eigenspaces.*

Proof: Since $\sum_{\lambda \in \mathrm{sp}(T)} E_{\lambda_i}$ is a direct sum and each λ_i has $\dim\left(E_{\lambda_i}\right) \geq 1$, the dimension of this linear span must equal n, so $V = \bigoplus_{\lambda_i \in \mathrm{sp}(T)} E_{\lambda_i}$. $\square$

In some sense (at least for complex matrices), the "n distinct eigenvalues condition"is *generic*: If entries $a_{ij} \in \mathbb{C}$ are chosen at random, then with "probability 1" the matrix $A = [a_{ij}]$ would have distinct eigenvalues in $\mathbb{C}$, so the characteristic polynomial would split completely into distinct linear factors

$$p_A(x) = c \cdot \prod_{i=1}^{n} (x - \lambda_i) \ .$$

Unfortunately, in many important applications the matrices of interest do not have n distinct eigenvalues, which is why we need the more subtle theory of "generalized eigenvalues" developed in Linear Algebra II, as a backup when diagonalization fails.

V.3 Diagonalization and Limits of Operators.

We begin by defining limits $\lim\limits_{n\rightarrow\infty} A_n = A$ of square matrices over $\mathbb{K} = \mathbb{R}$ or $\mathbb{C}$; limits $T_n \rightarrow T$ could similarly be defined for linear operators on a finite dimensional vector space V over these fields.

3.1. Definition. *For $\mathbb{K} = \mathbb{R}$ or $\mathbb{C}$ we may define* **pointwise convergence***, or "***sup norm convergence***" of matrices in* $\mathrm{M}(N, \mathbb{K})$

$$\lim_{n\rightarrow\infty} A_n = A \quad \text{or} \quad A_n \rightarrow A \text{ as } n \rightarrow \infty$$

to mean that each entry in A_n converges in $\mathbb{C}$ to the corresponding entry in the limit matrix A:

$$\left| a_{ij}^{(n)} - a_{ij} \right| \rightarrow 0 \quad \text{in } \mathbb{C} \text{ as } n \rightarrow \infty \tag{39}$$

for each $1 \leq i, j \leq N$, where $A_n = [a_{ij}^{(n)}]$.

Later we will examine other notions of matrix (or operator) convergence. In making the present definition we are, in effect, measurng the "size" of an $N \times N$ matrix by its "**sup-norm**," the size of its largest entry:

$$\|A\|_\infty = \max\{\, |a_{ij}| : 1 \le i, j \le N\}$$

This allows us to define the distance between two matrices in $\mathrm{M}(N, \mathbb{K})$ to be $d(A, B) = \|A - B\|_\infty$, and it should be evident that the limit $A_n \to A$ defined in (39) can be recast in terms of the sup-norm:

(40) $$A_n \to A \text{ as } n \to \infty \quad \Leftrightarrow \quad \|A_n - A\|_\infty \to 0 \text{ as } n \to \infty\ .$$

The sup norm on matrix space has several important properties (easily verified):

3.2. Exercise. If A, $B \in \mathrm{M}(N, \mathbb{K})$ prove that:

1. $\|\lambda A\|_\infty = |\lambda| \cdot \|A\|_\infty$, for all $\lambda \in \mathbb{K}$,
2. Triangle Inequality: $\|A + B\|_\infty \le \|A\|_\infty + \|B\|_\infty$;
3. Multiplicative Property: $\|AB\|_\infty \le n \cdot \|A\|_\infty \cdot \|B\|_\infty$.

Hint: Use the Triangle Inequality in $\mathbb{C}$, which says $|z \pm w| \le |z| + |w|$ for any $z, w \in \mathbb{C}$

A number of theorems regarding sup-norm limits follow from these basic inequalities.

3.3. Exercise. If $A_n \to A$ and $B_n \to B$ in the sup-norm, and $\lambda_n \to \lambda$ in $\mathbb{C}$, prove that:

1. $A_n + B_n \to A + B$
2. $A_n B \to AB$ and $AB_n \to AB$;
3. $A_n B_n \to AB$. Thus matrix multiplication is a "jointly continuous" operation on its two inputs.
4. If Q is an invertible matrix then $QA_nQ^{-1} \to QAQ^{-1}$. Hence every similarity transformation $A \mapsto QAQ^{-1}$ is a continuous operation on matrix space.
5. $\lambda_n A_n \to \lambda A$.

Hint: In (3.) add and subtract $A_n B$, then apply the triangle inequality.

The triangle inequality has a "converse" that is sometimes indispensable.

3.4. Proposition (Reverse Triangle Inequality). For $A, B \in \mathrm{M}(N, \mathbb{K})$ we have

$$\Big|\, \|A\|_\infty - \|B\|_\infty \,\Big| \le \|A - B\|_\infty$$

Proof: By the Triangle Inequality

$$\|A + B\|_\infty \le \|A\|_\infty + \|B\|_\infty$$

we get

$$\|A\|_\infty = \|A - B + B\|_\infty \le \|A - B\|_\infty\ , + \|B\|_\infty$$

so that $\|A\|_\infty - \|B\|_\infty \le \|A - B\|_\infty$. Reversing roles of A, B we also get $\|B\|_\infty - \|A\|_\infty \le \|A - B\|_\infty$. Since the absolute value of a real number is either $|c| = c$ or $-c$, we conclude that

$$\Big|\, \|A\|_\infty - \|B\|_\infty \,\Big| \le \|A - B\|_\infty \quad \square$$

As an immediate consequence we have

3.5. Corollary. If $A_n \to A$ in $\mathrm{M}(n, \mathbb{C})$ then $\|A_n\|_\infty \to \|A\|_\infty$ in $\mathbb{R}$. $\quad\square$

3.6. Exercise. If A in $\mathrm{M}(n, \mathbb{C})$ is an *invertible* matrix and $A_n \to A$ in the sup-norm, prove that

1. $\det(A_n) \to \det(A)$;
2. $A_n^{-1} \to A^{-1}$ in the sup norm.

Hint: Recall Cramer's Rule for computing A^{-1} for a nonsingular matrix A.

Application #1: Computing the Exponential e^A of a Matrix. We will show that the exponential series

$$e^A = \sum_{k=0}^{\infty} \frac{1}{k!} A^k \qquad (A \in \mathrm{M}(N, \mathbb{C}))$$

converges in the sup-norm, which means that the finite partial sums of the series

$$S_n = I + A + \frac{A^2}{2!} + \ldots + \frac{A^n}{n!} \qquad n \in \mathbb{N}$$

converge to a definite limit e^A in matrix space:

$$\|S_n - e^A\|_\infty \to 0 \quad \text{as} \quad n \to \infty$$

This is not so easy to prove, but if $D = \operatorname{diag}(\lambda_1, \ldots, \lambda_N)$ is a *diagonal* matrix

$$D = \begin{pmatrix} \lambda_1 & & & 0 \\ & \lambda_2 & & \\ & & \ddots & \\ 0 & & & \lambda_N \end{pmatrix}$$

it is quite obvious that the partial sums S_n converge in the sup-norm,

$$S_n = I + D + \ldots + \frac{D^n}{n!} = \begin{pmatrix} 1 + \lambda_1 + \ldots + \frac{\lambda_1^n}{n!} & & & 0 \\ & 1 + \lambda_2 + \ldots + \frac{\lambda_2^n}{n!} & & \\ & & \ddots & \\ 0 & & & 1 + \lambda_N + \ldots + \frac{\lambda_N^n}{n!} \end{pmatrix}$$

$$\longrightarrow \begin{pmatrix} e^{\lambda_1} & & & 0 \\ & e^{\lambda_2} & & \\ & & \ddots & \\ 0 & & & e^{\lambda_N} \end{pmatrix} \qquad \text{as } n \to \infty$$

because $e^z = \sum_{k=0}^{\infty} z^k/k!$ is absolutely convergent for every complex number $z \in \mathbb{C}$. Therefore $S_n \to e^D$ in the sup-norm and

$$e^D = \sum_{k=0}^{\infty} \frac{D^k}{k!} = \lim_{n\to\infty} S_n = \begin{pmatrix} e^{\lambda_1} & & & 0 \\ & e^{\lambda_2} & & \\ & & \ddots & \\ 0 & & & e^{\lambda_N} \end{pmatrix}$$

A Digression: The Cauchy Convergence Criterion in Matrix Space. For matrices that are not diagonal it is not easy to prove that there actually *is* a matrix e^A to which the matrix-exponential series converges in sup-norm,

$$\|S_n - e^A\|_\infty \to 0 \qquad \text{as } n \to \infty \,.$$

This follows because $\mathrm{M}(N, \mathbb{K})$ equipped with the sup-norm $\|\cdot\|_\infty$ has the following **completeness** property, similar to completeness of $\mathbb{R}^n$ and $\mathbb{C}^n$ in the Euclidean norm

$$\|\mathbf{z}\|_2 = \Big(\sum_{k=1}^{N} |z_k|^2\Big)^{1/2} \qquad \text{for } \mathbf{z} = (z_1, \ldots, z_N) \text{ in } \mathbb{C}^N \ ,$$

or completeness of the number fields $\mathbb{K} = \mathbb{R}$ and $\mathbb{C}$.

THEOREM (CAUCHY CONVERGENCE CRITERION). *A sequence $\{A_n\}$ in $\mathrm{M}(N, \mathbb{K})$ converges to some limit $A_0 = lim_{n\to\infty} A_n$ in the $\|\cdot\|_\infty$-norm if and only if the sequence has the* **Cauchy property**

$$\|A_m - A_n\|_\infty \to 0 \quad \textit{eventually as} \quad m, n \to \infty$$

(41) *To be precise, this property means: Given any $r > 0$ we can find a cutoff $M > 0$ such that*

$$\|A_m - A_n\|_\infty < r \quad \textit{for all } m, n \geq M$$

Statement (41) is much stronger than saying *successive terms* in the sequence get close, with $\|A_{n+1} - A_n\| \to 0$ as $n \to \infty$; to verify the Cauchy criterion you must show that *all* the terms are eventually close together as $n \to \infty$.

When you try to prove $A_n \to A_0$ by examining the distances $\|A_n - A_0\|_\infty$ you must actually have the prospective limit A_0 in hand, and that limit might be very hard to guess. The Cauchy criterion gets around this problem. You don't need to identify the value of the limit whose existence is assured in (41), because the Cauchy criterion can be verified by inspecting the terms of the given sequence $\{A_n\}$. Similarly in $\mathbb{R}$, the Integral Test of Calculus shows that the the partial sums

$$S_n = 1 + \frac{1}{2^2} + \ldots + \frac{1}{n^2} \quad \text{of the } \textit{Harmonic Series} \ \sum_{n=1}^{\infty} 1/n^2$$

have the Cauchy property, and hence by the completeness property (41)

$$\sum_{n=1}^{\infty} \frac{1}{n^2} = \lim_{n\to\infty} \{S_n\}$$

exists. It is a lot harder to identify this limit in "closed form," and show it is exactly $\pi^2/6$. We will see one way to do this in Chapter VI.

As for the matrix exponential series $\sum_{n=0}^{\infty} A^k/k!$ we now show that its partial sums $S_n = \sum_{k=0}^{n} A^k/k!$ have the Cauchy property in $||\cdot||_\infty$-norm. By completeness of $\mathrm{M}(N, \mathbb{C})$ the partial sums actually have a limit, which we name "e^A"

$$e^A \ = \sum_{k=1}^{\infty} \frac{A^k}{k!} \ = \lim_{n\to\infty} S_n$$

Proof: To verify the Cauchy property for $\{S_n\}$ we may assume $m > n$. By the Triangle Inequality and the multiplicative property (3.) of Exercise 3.2 we have

$$\|S_m - S_n\|_\infty = \| \sum_{k=n+1}^{m} \frac{A^k}{k!} \|_\infty \leq \sum_{n+1}^{m} \frac{N^k \|A\|_\infty^k}{k!}$$

By the Ratio Test the Taylor series for $f(x) = e^x$ converges (to e^x) for all $x \in \mathbb{R}$:

$$e^x = \sum_{n=0}^{\infty} \frac{D^n f(0)}{n!} x^n = \sum_{n=0}^{\infty} \frac{x^n}{n!}$$

because $D^n\{e^x\} = e^x$ for all x. Taking $x = N \cdot \|A\|_\infty$, we get

$$\sum_{k=0}^{n} N^k \|A\|_\infty^k / k! \rightarrow \sum_{k=0}^{\infty} \frac{N^k \|A\|_\infty^k}{k!} = e^{N\|A\|_\infty} < \infty \quad \text{as } n \rightarrow \infty ,$$

hence for $m \geq n$:

$$0 \leq \|S_m - S_n\|_\infty \leq \sum_{n+1}^{m} \frac{(N \cdot \|A\|_\infty)^k}{k!} = \sum_{k=n+1}^{\infty} \frac{\big(N \cdot \|A\|_\infty\big)^k}{k!} \rightarrow 0$$

as $n \rightarrow \infty$. Thus, $\{S_n\}$ is Cauchy sequence for the $\|\cdot\|_\infty$-norm and the matrix-valued series $\sum_{k=0}^{\infty} A^k/k!$ converges in $\|\cdot\|_\infty$-norm for every matrix A. $\square$

In general, it is a difficult task to directly compute the sum of a convergent series such as $e^A = \sum_{n=0}^{\infty} A^n/n!$ For instance, consider how one might try to evaluate e^A when

$$A = \begin{pmatrix} 1 & -1 \\ -6 & 2 \end{pmatrix}$$

Computing higher and higher powers A^k is computationally prohibitive, and how many terms would be needed to compute each entry of e^A with an error of at most 1×10^{-6} (6-place accuracy)?

As mentioned earlier, computing e^A is easy if $A = D = \text{diag}(\lambda_1, \ldots, \lambda_N)$ is diagonal. Then,

$$S_n = I + D + \ldots + \frac{D^n}{n!} \rightarrow \begin{pmatrix} e^{\lambda_1} & & & 0 \\ 0 & e^{\lambda_2} & & \\ & & \ddots & \\ 0 & & & e^{\lambda_N} \end{pmatrix} = e^D$$

We now show that e^{tA} can be computed in closed form for all $t \in \mathbb{R}$, for any A that is diagonalizable over $\mathbb{R}$ or $\mathbb{C}$.

3.7. Example. Compute e^{tA} $(t \in \mathbb{R})$ for the matrix

$$A = \begin{pmatrix} 1 & -1 \\ -6 & 2 \end{pmatrix}$$

Solution: First observe that A is diagonalizable, with $QAQ^{-1} = \begin{pmatrix} 4 & 0 \\ 0 & -1 \end{pmatrix} = D$ for suitably chosen Q. The eigenvalues are the roots of the characteristic polynomial

$$\begin{aligned} p_A(x) &= \det \begin{pmatrix} 1-\lambda & -1 \\ -6 & 2-\lambda \end{pmatrix} = (\lambda - 2)(\lambda - 1) - 6 \\ &= \lambda^2 - 3\lambda + 2 - 6 = \lambda^2 - 3\lambda - 4 = (\lambda - 4)(\lambda + 1) \quad , \end{aligned}$$

so $\text{sp}(A) = \{4, -1\}$. The eigenspaces are computed by row reduction:

- **Eigenvalue** $\lambda = 4$:

$$(A - \lambda I) = \begin{pmatrix} -3 & -1 \\ -6 & -2 \end{pmatrix} \rightarrow \begin{pmatrix} 1 & \frac{1}{3} \\ 0 & 0 \end{pmatrix}.$$

Solutions of $(A - \lambda I)X = 0$ are

$$X \in \mathbb{K}\cdot \begin{pmatrix} -\frac{1}{3} \\ 1 \end{pmatrix} = \mathbb{K}\cdot \begin{pmatrix} 1 \\ -3 \end{pmatrix} = E_{\lambda=4} \ .$$

- **Eigenvalue** $\lambda = -1$:

$$(A - \lambda I) = \begin{pmatrix} 2 & -1 \\ -6 & 3 \end{pmatrix} \rightarrow \begin{pmatrix} 1 & -\frac{1}{2} \\ 0 & 0 \end{pmatrix}$$

Solutions of $(A - \lambda I)X = 0$ are $X \in \mathbb{K} \cdot \begin{pmatrix} \frac{1}{2} \\ 1 \end{pmatrix} = \mathbb{K}\cdot \begin{pmatrix} 1 \\ 2 \end{pmatrix} = E_{\lambda=-1}$.

Thus $\mathbb{K}^2 = E_{\lambda=4} \oplus E_{\lambda=-1}$ and $\mathfrak{Y} = \{ \mathbf{f}_1 = (1, -3), \mathbf{f}_2 = (1, 2) \}$ is a diagonalizing basis in $\mathbb{K}^2$. On the other hand, from our discussion of "change of basis" in Chapter II we have

$$\begin{aligned} D &= \begin{pmatrix} 4 & 0 \\ 0 & -1 \end{pmatrix} = [L_A]_{\mathfrak{Y}\mathfrak{Y}} = [\mathrm{id}]_{\mathfrak{Y}\mathfrak{X}} \cdot [L_A]_{\mathfrak{X}\mathfrak{X}} \cdot [\mathrm{id}]_{\mathfrak{X}\mathfrak{Y}} \\ &= [\mathrm{id}]_{\mathfrak{Y}\mathfrak{X}} \cdot A \cdot [\mathrm{id}]_{\mathfrak{X}\mathfrak{Y}} \end{aligned}$$

Since

$$\begin{cases} \mathbf{f}_1 = \mathbf{e}_1 - 3\mathbf{e}_2 \\ \mathbf{f}_2 = \mathbf{e}_1 + 2\mathbf{e}_2 \end{cases} \qquad \big(\text{where } \mathfrak{X} = \{\mathbf{e}_1, \mathbf{e}_2\} = \text{ standard basis in } \mathbb{K}^2\big)$$

we see that $[\mathrm{id}]_{\mathfrak{X}\mathfrak{Y}} = \begin{pmatrix} 1 & 1 \\ -3 & 2 \end{pmatrix}$. Then $QAQ^{-1} = D$ taking $Q^{-1} = [\mathrm{id}]_{\mathfrak{X}\mathfrak{Y}} = \begin{pmatrix} 1 & 1 \\ -3 & 2 \end{pmatrix}$, and since $\det(Q^{-1}) = 5$ we get $Q = (Q^{-1})^{-1} = \frac{1}{5} \cdot \begin{pmatrix} 2 & -1 \\ 3 & 1 \end{pmatrix}$. Now

$$\begin{cases} D = QAQ^{-1} \\ A = Q^{-1}DQ \end{cases} \quad \Rightarrow \quad A^k = (Q^{-1}DQ)\cdot(Q^{-1}DQ) \cdot \ldots \cdot (Q^{-1}DQ) = Q^{-1}D^kQ$$

for $k = 0, 1, 2 \ldots$, hence by (4.) of Exercise 3.3 we get

$$\begin{aligned} e^A = \sum_{k=0}^{\infty} \frac{A^k}{k!} &= \sum_{k=0}^{\infty} \frac{(Q^{-1}DQ)^k}{k!} = \sum_{k=0}^{\infty} \frac{Q^{-1}D^kQ}{k!} \\ &= Q^{-1}\Big(\sum_{k=0}^{\infty} \frac{D^k}{k!} \Big) \cdot Q = Q^{-1}e^DQ \ , \end{aligned}$$

We conclude that

$$e^A = Q^{-1} \begin{pmatrix} e^4 & 0 \\ 0 & e^{-1} \end{pmatrix} Q$$

which exhibits e^A as a product of *just three explicit matrices.*

Similarly, for $t \in \mathbb{R}$ we compute e^{tA}

$$\begin{aligned} e^{tA} &= Q^{-1} \begin{pmatrix} e^{4t} & 0 \\ 0 & e^{-t} \end{pmatrix} Q = \begin{pmatrix} 1 & 1 \\ -3 & 2 \end{pmatrix} \cdot \begin{pmatrix} e^{4t} & 0 \\ 0 & e^{-t} \end{pmatrix} \cdot \frac{1}{5} \begin{pmatrix} 2 & -1 \\ 3 & 1 \end{pmatrix} \\ &= \frac{1}{5} \begin{pmatrix} 2e^{4t} + 3e^{-t} & -e^{4t} + e^{-t} \\ -6e^{4t} + 6e^{-t} & 3e^{4t} + 2e^{-t} \end{pmatrix} = \frac{1}{5}e^{4t}\cdot \begin{pmatrix} 2 & -1 \\ -6 & 3 \end{pmatrix} + \frac{1}{5}e^{-t}\cdot \begin{pmatrix} 3 & 1 \\ 6 & 2 \end{pmatrix} \end{aligned}$$

Setting $t = 0$, we get $e^0 = I$; setting $t = 1$, we get the answer to our original question

$$e^A = \frac{1}{5}e^4 \cdot \begin{pmatrix} 2 & -1 \\ -6 & 3 \end{pmatrix} + \frac{1}{5}e^{-1} \cdot \begin{pmatrix} 3 & 1 \\ 6 & 2 \end{pmatrix} \qquad \square$$

Application #2: Solving Linear Systems of Differential Equations.

In the next application we see why one might want to compute the matrix-valued function $\phi(t) = e^{tA}$, $\phi : \mathbb{R} \to \mathrm{M}(N, \mathbb{C})$. First we must sketch some additional properties of the exponential map on matrices (mostly without proofs).

1. *If A and B commute then*

   EXPONENT LAW: $e^{A+B} = e^A \cdot e^B$

   In particular, e^A is always invertible, with $(e^A)^{-1} = e^{-A}$. Futhermore,

   ONE-PARAMETER GROUP LAW: $e^{(s+t)A} = e^{sA} \cdot e^{tA}$ *for all* $s, t \in \mathbb{R}$

   and e^{-tA} is the inverse of e^{tA} for $t \in \mathbb{R}$.

Proof (sketch): We give an informal proof involving rearrangement of a matrix-valued double series. But beware: rearrangement and regrouping of series are delicate matters even for scalar-valued series, and a proof that would pass muster with analysts would require considerably more detail – see any text on Mathematical Analysis.

The series $e^A = \sum_{k=0}^{\infty} A^k/k!$ and $e^B = \sum_{\ell=0}^{\infty} B^\ell/\ell!$ are sup-norm convergent. Expanding the product of the two series term-by-term (which in itself requires some justification!) we get

$$\begin{aligned}
e^A \cdot e^B &= \Big(\sum_{k=0}^{\infty} A^k/k!\Big) \cdot \Big(\sum_{\ell=0}^{\infty} B^\ell/\ell!\Big) = \sum_{k,\ell \geq 0} \frac{1}{k!}\frac{1}{\ell!}\, A^k B^\ell \\
&= \sum_{k,\ell \geq 0} \frac{1}{(k+\ell)!} \cdot \frac{(k+\ell)!}{k!\ell!}\, A^k B^\ell \\
&= \sum_{n=0}^{\infty} \frac{1}{n!} \cdot \Big(\sum_{k=0}^{n} \tbinom{n}{k}\, A^k B^\ell\Big) \qquad \text{where } \tbinom{n}{k} = \text{(binomial coefficient)} \\
&= \sum_{n=0}^{\infty} \frac{1}{n!}(A+B)^n \qquad \text{(Binomial Formula)} \\
&= e^{A+B} \quad \square
\end{aligned}$$

2. DIFFERENTIATION LAW. *The derivative of $\phi(t) = e^{tA}$ exists and is continuous, with*

$$\frac{d}{dt}\big(e^{tA}\big) = A \cdot e^{tA} \qquad \text{for all } t \in \mathbb{R}$$

Proof: Using the Exponent Law we get

$$\begin{aligned}
\frac{d}{dt}\big(e^{tA}\big) &= \lim_{\Delta t \to 0} \frac{e^{(t+\Delta t)A} - e^{tA}}{\Delta t} \\
&= \lim_{\Delta t \to 0} \Big(\frac{e^{(\Delta t)A} - I}{\Delta t}\Big) \cdot e^{tA} \qquad \text{(since } e^{(t+\Delta t)A} = e^{tA} \cdot e^{(\Delta t)A}\text{)} \\
&= \Big(\lim_{\Delta t \to 0} \frac{e^{(\Delta t)A} - I}{\Delta t}\Big) \cdot e^{tA}
\end{aligned}$$

Using the norm properties listed in Exercises 3.2 -3.3 it is not hard to show that

$$\begin{aligned} e^{(\Delta t)A} - I &= \Big(I + (\Delta t)A + \frac{(\Delta t)^2}{2!}A^2 + ...\Big) - I \\ &= \Delta t \cdot \Big(A + \frac{(\Delta t)^2}{2!}A^2 + ...\Big) = \Delta t\big(A + \mathcal{O}(\Delta t)\big) \end{aligned}$$

where the matrix-valued remainder $\mathcal{O}(\Delta t)$ becomes very small compared to Δt

$$\frac{\|\mathcal{O}(\Delta t)\|_\infty}{|\Delta t|} \to 0 \quad \text{as } \Delta t \to 0 \ .$$

Thus,

$$\frac{e^{(\Delta t)A} - I}{\Delta t} = \frac{\Delta t}{\Delta t}\big(A + \mathcal{O}(\Delta t)\big) \to A$$

in the $\| \cdot \|_\infty$-norm as $\Delta t \to 0$, proving the formula. $\square$

Any system of n first order linear ordinary differential equations in n unknowns can be written in matrix form as

$$\frac{d\mathbf{y}}{dt} = A \cdot \mathbf{y}(y) \qquad \text{with initial condition } \mathbf{y}(0) = \mathbf{c} \tag{42}$$

where $\mathbf{y}(t) = \big(y_1(t), \ldots, y_n(t)\big)$ is a vector-valued function of t, and the $n \times n$ matrix A provides the coefficients of the system. It is well known that once the initial value $\mathbf{c}$ is specified there is a unique infinitely differentiable vector-valued solution $\mathbf{y}(t)$ if we regard $\mathbf{y}(t)$ as an $n \times 1$ column vector. The solution can be computed explicitly as

$$\mathbf{y}(y) = e^{tA} \cdot \mathbf{y}(0) = e^{tA} \cdot \mathbf{c} \qquad \text{for } t \in \mathbb{R} \tag{43}$$

In fact,

$$\frac{d\mathbf{y}}{dt} = \frac{d}{dt}\big(e^{tA} \cdot \mathbf{c}\big) = \frac{d}{dt}\big(e^{tA}\big) \cdot \mathbf{c} = Ae^{tA} \cdot \mathbf{c} = A \cdot \mathbf{y}(t) \ ,$$

and when $t = 0$ we get $\mathbf{y}(0) = \mathbf{c}$ because $e^{0 \cdot A} = I_{n \times n}$. We must of course compute e^{tA} to arrive at $\mathbf{y}(t)$ but we have seen how to do that in the previous example, at least when the coefficient matrix can be diagonalized.

3.8. Example. If $A = \begin{pmatrix} 1 & -1 \\ -6 & 2 \end{pmatrix}$ determine the unique solution of the first order vector-valued differential equation

$$\frac{d\mathbf{y}}{dt} = A \cdot \mathbf{y}(t) \text{ such that } \mathbf{y}_0 = \mathbf{y}(0) = \begin{pmatrix} 1 \\ 0 \end{pmatrix}.$$

Likewise for the initial value $\mathbf{y}_0 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$. Then find all solutions of

$$\frac{d\mathbf{y}}{dt} = A \cdot \mathbf{y}(t) \qquad \text{for an arbitrary initial value } \mathbf{y}_0 = \begin{pmatrix} c_1 \\ c_2 \end{pmatrix}$$

Solution: Earlier we found that

$$QAQ^{-1} = \begin{pmatrix} 4 & 0 \\ 0 & -1 \end{pmatrix} \qquad \text{for} \qquad Q = \frac{1}{5}\begin{pmatrix} 2 & -1 \\ 3 & 1 \end{pmatrix}$$

and showed that

$$e^A = e^{QDQ^{-1}} = Q^{-1} \cdot e^D \cdot Q = \frac{1}{5}e^4 \cdot \begin{pmatrix} 2 & -1 \\ -6 & 3 \end{pmatrix} + e^{-1} \cdot \begin{pmatrix} 3 & 1 \\ 6 & 2 \end{pmatrix}$$

Taking e^{tA} in place of e^A, we got (with little additional effort):

$$e^{tA} = \frac{1}{5}e^{4t} \cdot \begin{pmatrix} 2 & -1 \\ -6 & 3 \end{pmatrix} + \frac{1}{5}e^{-t} \cdot \begin{pmatrix} 3 & 1 \\ 6 & 2 \end{pmatrix} \qquad \text{for all } t \in \mathbb{R}$$

Taking $\mathbf{y}_0 = \mathbf{e}_1 = (1, 0)$ we get a solution:

$$\mathbf{y}_1(t) = e^{tA}(\mathbf{e}_1) = \frac{1}{5}e^{4t} \cdot \begin{pmatrix} 2 \\ -6 \end{pmatrix} + \frac{1}{5}e^{-t} \cdot \begin{pmatrix} 3 \\ 6 \end{pmatrix}$$

If $\mathbf{y}_0 = \mathbf{e}_2 = (0, 1)$ we get another solution:

$$\mathbf{y}_2(t) = e^{tA}(\mathbf{e}_2) = \frac{1}{5}e^{4t} \begin{pmatrix} -1 \\ 3 \end{pmatrix} + \frac{1}{5}e^{-t} \cdot \begin{pmatrix} 1 \\ 2 \end{pmatrix}$$

For an arbitrary initial condition $\mathbf{y}(0) = \mathbf{c} = c_1\mathbf{e}_1 + c_2\mathbf{e}_2$, it is obvious that the solution of $d\mathbf{y}/dt = A \cdot \mathbf{y}(t)$ with this initial condition is the same linear combination of the "basic solutions" $\mathbf{y}_1(t)$ and $\mathbf{y}_2(t)$ namely:

$$\begin{aligned} \mathbf{y}(t) &= c_1\mathbf{y}_1(t) + c_2\mathbf{y}_2(t) \\ &= \frac{1}{5}e^{4t} \cdot \Big[c_1 \begin{pmatrix} 2 \\ -6 \end{pmatrix} + c_2 \begin{pmatrix} -1 \\ 3 \end{pmatrix}\Big] + \frac{1}{5}e^{-t}\Big[c_1 \begin{pmatrix} 3 \\ 6 \end{pmatrix} + c_2 \begin{pmatrix} 1 \\ 2 \end{pmatrix}\Big] \end{aligned}$$

(Check for yourself that $\mathbf{y}(0) = c_1\mathbf{e}_1 + c_2\mathbf{e}_2 = \mathbf{c}$.)

The full set of differentiable maps $f : \mathbb{R} \to \mathbb{C}^2$ such that $df/dt = A \cdot f(t)$ is a 2-dimensional subspace M in the ∞-dimensional space $\mathcal{C}^\infty(\mathbb{R}, \mathbb{C}^2)$ of infinitely differentiable vector valued maps:

$$M \ = \ \mathbb{C}\text{-span}\{\mathbf{y}_1(t), \mathbf{y}_2(t)\} \ = \ \{c_1\mathbf{y}_1 + c_2\mathbf{y}_2 : c_1, c_2 \in \mathbb{C}\}$$

and the "basic solutions" $\mathbf{y}_1,]bfy_2$ are a vector basis for M. One should check that $\mathbf{y}_1$, $\mathbf{y}_2$ are linearly independent vectors in $\mathcal{C}^\infty(\mathbb{R}, \mathbb{C}^2)$. But if there were coefficients α_1, α_2 such that $\alpha_1\mathbf{y}_1(t) + \alpha_2\mathbf{y}_2(t) \equiv 0$ in $\mathbb{C}^2$, and we take any convenient base point (say $t = 0$), we would then have the following vector identity in $\mathbb{C}^2$:

$$\begin{aligned} \begin{pmatrix} 0 \\ 0 \end{pmatrix} &= \frac{\alpha_1}{5}\Big[\begin{pmatrix} 2 \\ -6 \end{pmatrix} + \begin{pmatrix} 3 \\ 6 \end{pmatrix} \Big] + \frac{\alpha_2}{5}\Big[\begin{pmatrix} -1 \\ 3 \end{pmatrix} + \begin{pmatrix} 1 \\ 2 \end{pmatrix} \Big] \\ &\Rightarrow \ \frac{\alpha_1}{5}\begin{pmatrix} 5 \\ 0 \end{pmatrix} + \frac{\alpha_2}{5}\begin{pmatrix} 0 \\ 5 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \\ &\Rightarrow \ \alpha\mathbf{e}_1 + \alpha_2\mathbf{e}_2 \ = \ 0 \\ &\Rightarrow \ \alpha_1 = \alpha_2 = 0 \end{aligned}$$

as required. $\square$

A similar discussion holds for equations $d\mathbf{y}/dt = A \cdot \mathbf{y}(t)$ when A is $n \times n$ (and diagonalizable). If $\{\mathbf{y}_1(t), \ldots, \mathbf{y}_n(t)\} \subseteq \mathcal{C}^\infty(\mathbb{R}, \mathbb{C}^n)$ are the "basic solutions," whose initial values are $\mathbf{y}_k(0) = \mathbf{e}_k$ (the standard basis vectors in $\mathbb{C}^n$), then a solution with arbitrary initial value $\mathbf{y}(0) = \sum_{k=1}^n c_k\mathbf{e}_k \in \mathbb{C}^n$ is obtained by taking the same linear combination

$$\mathbf{y}(t) = c_1\mathbf{y}_1(t) + \ldots + c_n\mathbf{y}_n(t) \quad .$$

of basic solution $\mathbf{y}_k(t)$. As above, the $\mathbf{y}_k$ are linearly independent vectors in $\mathcal{C}^\infty(\mathbb{R}, \mathbb{C}^n)$: if $\mathbf{0} = \sum c_k \mathbf{y}_k(t)$ in $\mathcal{C}^\infty(\mathbb{R}, \mathbb{C}^n)$ for all t, then (taking $t = 0$) $\sum c_k \mathbf{e}_k = \mathbf{0}$ in $\mathbb{C}^n$; thus, $c_1 = c_2 = \ldots = c_n = 0$ because $\mathbf{y}_k(0) = \mathbf{e}_k$, by definition. We conclude that the $\{\mathbf{y}_k(t)\}$ are a basis for the full set of solutions (with arbitrary initial value) of the equation $d\mathbf{y}/dt = A \cdot \mathbf{y}(t)$.

$$\begin{aligned} M &= \left\{ f \in \mathcal{C}^\infty : \frac{df}{dt} = A \cdot f(t) \text{ for all } t \in \mathbb{R} \right\} \qquad (f : \mathbb{R} \to \mathbb{C}^n)\,\} \\ &= \mathbb{C}\text{-span}\{\mathbf{y}_1(t), \ldots, \mathbf{y}_n(t)\} \end{aligned}$$

which has dimension $\dim_{\mathbb{C}}(M) = n$. $\square$