AA/EE547: Wi 22

Lecture 0: Review of Matrix Exponential [510]

Lecturer: L.J. Ratliff

Disclaimer: *These notes have not been subjected to the usual scrutiny reserved for formal publications, meaning you should take your own notes in class and review the provided references as opposed to taking these notes as your sole resource. I provide the lecture notes to you as a courtesy; it is not required that I do this. They may be distributed outside this class only with the permission of the Instructor.*

Throughout the quarter we will use the following keys for references to books:

- **[Ax]**: Axler, Linear Algebra Done Right
- **[C&D]**: Callier and Desoer, Linear Systems Theory
- **[He]**: Hespanha, Linear Systems Theory

References: **Solutions to ODEs**: Chapter 3 **[C&D]**; **Jordan Form**: Chapter 4 **[C&D]**; Chapter 8.D **[Ax]**

1 The Matrix Exponential

First, we note that the matrix exponential has several important properties.

- $e^0 = I$
- $e^{A(t+s)} = e^{At}e^{As}$
- $e^{(A+B)t} = e^{At}e^{Bt} \iff AB = BA$
- $(e^{At})^{-1} = e^{-At}$
- $\frac{d}{dt}e^{At} = Ae^{At} = e^{At} \cdot A$
- Let $z(t) \in \mathbb{R}^{n\times n}$. Then the solution to
$$\dot{z}(t) = Az(t)$$
with $z(0) = I$ is
$$z(t) = e^{At}$$

Recall that
$$\exp(x) = \sum_{k=0}^{\infty} \frac{x^k}{k!}$$
This is also true for the matrix exponential—i.e.
$$\exp(At) = \sum_{k=0}^{\infty} \frac{(At)^k}{k!}$$

Fact. Note also that Cayley-Hamilton implies that the matrix exponential is expressible as a polynomial of order $n-1$!

Using the series representation of e^{At} to compute e^{At} is difficult unless, e.g., the matrix A is nilpotent in which case the series yields a closed form solution.

Definition. (Nilpotent) A nilpotent matrix is such that $A^k = 0$ for some k.

Example. Consider

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

Then

$$A^2 = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$$

so that

$$e^{At} = I + At = \begin{bmatrix} 1 & t \\ 0 & 1 \end{bmatrix}$$

Hence we need an alternative method to compute it.

1.1 Review of Laplace

Definition. (Laplace Transform)

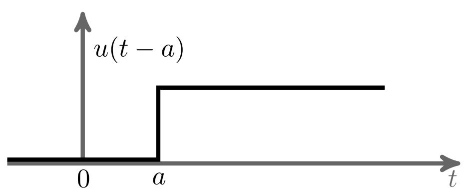
$$\mathcal{L}f(t) = \int_0^\infty f(t)e^{-st}\, dt$$

The Laplace transform has the following properties:

- Linearity:

$$\mathcal{L}(af(t) + bg(t)) = a\underbrace{F(s)}_{\mathcal{L}f(t)} + b\underbrace{G(s)}_{\mathcal{L}g(t)}$$

- Time Delay: Let u be a step function.



$$f(t-a)u(t-a) \xrightarrow{\mathcal{L}} e^{-as}F(s)$$

- First derivative (technically should be $t = 0^-$):

$$\mathcal{L}\dot{f}(t) = sF(s) - f(0)$$

- Integration:

$$\mathcal{L}\left(\int_{0^-}^\infty f(\tau)\, d\tau\right) = \frac{F(s)}{s}$$

1.2 Computation of e^{At} via Laplace

Use the Laplace transform of $\dot{X} = AX$, $X \in \mathbb{R}^{n\times n}$, $X(0) = I$:

$$s\hat{X}(s) - X(0) = A\hat{X}(s)$$

so that

$$s\hat{X}(s) - A\hat{X}(s) = I \implies \hat{X}(s) = (sI - A)^{-1}$$

We know (from property 6) that $X(t) = e^{At}$ so that

$$e^{At} = X(t) = \mathcal{L}^{-1}(\hat{X}(s)) = \mathcal{L}^{-1}\left((sI - A)^{-1}\right)$$

Example. Consider the same example above with

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

so that

$$(sI - A) = \begin{bmatrix} s & -1 \\ 0 & s \end{bmatrix}$$

Recall that the inverse of a 2×2 matrix

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

is

$$A^{-1} = \frac{1}{ad - bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}.$$

Now,

$$(sI - A)^{-1} = \frac{1}{s^2}\begin{bmatrix} s & 1 \\ 0 & s \end{bmatrix} = \begin{bmatrix} \frac{1}{s} & \frac{1}{s^2} \\ 0 & \frac{1}{s} \end{bmatrix}$$

where we recall that $\mathcal{L}(f(t)) = \frac{F(s)}{s}$ so that $\mathcal{L}(1) = \frac{1}{s}$; it is also easy to show that the ramp function transforms to $\frac{1}{s^2}$. Hence,

$$e^{At} = \mathcal{L}^{-1}\begin{bmatrix} \frac{1}{s} & \frac{1}{s^2} \\ 0 & \frac{1}{s} \end{bmatrix} = \begin{bmatrix} 1 & t \\ 0 & 1 \end{bmatrix}$$

Example. Consider

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

Then

$$(sI - A)^{-1} = \begin{bmatrix} s-1 & -1 \\ 0 & s-1 \end{bmatrix}^{-1} = \frac{1}{(s-1)^2}\begin{bmatrix} s-1 & 1 \\ 0 & s-1 \end{bmatrix}$$

where we recall that $\mathcal{L}e^{at} = \frac{1}{s-a}$, $s > a$ which can be verified by direct integration. Hence,

$$e^{At} = \mathcal{L}^{-1}\begin{bmatrix} \frac{1}{s-1} & \frac{1}{(s-1)^2} \\ 0 & \frac{1}{s-1} \end{bmatrix} = \begin{bmatrix} e^t & te^t \\ 0 & e^t \end{bmatrix}$$

2 Computing the Matrix Exponential

Computation of the matrix exponential is important for expressing the solution of a autonomous or controlled linear time invariant dynamical system. So we need ways to compute it that are tractable.

3 Distinct Eigenvalues

If matrix $A \in \mathbb{R}^{n\times n}$ (or $\in \mathbb{C}^{n\times n}$) has m distinct eigenvalues ($\lambda_i \neq \lambda_j$, $i \neq j$) then it has (at least) m linearly independent eigenvectors.

If all eigenvalues of A are distinct then A is diagonalizable.

Q: do you remember what diagonalizable means?

Diagonalizable. An $n \times n$ matrix A is diagonalizable iff the sum of the dimensions of its eigenspaces is n—aka there exists a matrix P such that

$$A = P \Lambda P^{-1}$$

where $\Lambda = \mathrm{diag}(\lambda_1, \dots, \lambda_n)$,

$$P = \begin{bmatrix} v_1 & \cdots & v_n \end{bmatrix}$$

with

$$A v_i = \lambda_i v_i$$

(i.e. col vectors of P are right eigenvectors of A)

Proof. Proof of Prop. 3 (By contradiction) Assume λ_i, $i \in \{1, \dots, m\}$ are distinct and v_i, $i = 1, \dots, m$ are linearly dependent. That is, there exists α_i such that

$$\sum_{i=1}^{m} \alpha_i v_i = 0$$

where all α_i are not zero. We can assume w.l.o.g that $\alpha_1 \neq 0$. Multiplying on the left by $(\lambda_m I - A)$,

$$0 = (\lambda_m I - A) \sum_{i=1}^{m} \alpha_i v_i = (\lambda_m I - A) \sum_{i=1}^{m-1} \alpha_i v_i + \alpha_m (\lambda_m I - A) v_m = \sum_{i=1}^{m-1} \alpha_i (\lambda_m - \lambda_i) v_i$$

since $A v_i = \lambda_i v_i$. Then multiply by $(\lambda_{m-1} I - A)$ to get that

$$0 = (\lambda_{m-1} I - A) \sum_{i=1}^{m-1} \alpha_i (\lambda_m - \lambda_i) v_i = \sum_{i=1}^{m-2} \alpha_i (\lambda_{m-1} - \lambda_i)(\lambda_m - \lambda_i) v_i = 0$$

Repeatedly multiply by $(\lambda_{m-k} I - A)$, $k = 2, \dots, m-2$ to obtain

$$\alpha \prod_{i=2}^{m} (\lambda_i - \lambda_1) v_i = 0$$

As $\lambda_1 \neq \lambda_i$, $i = 2, \dots, m$, the above implies that $\alpha_1 = 0$ which is a contradiction. □

For each $n \times n$ complex matrix A, define the exponential of A to be the matrix

$$\exp(A) = \sum_{k=0}^{\infty} \frac{A^k}{k!}$$

It is not difficult to show that this sum converges for all complex matrices A of any finite dimension. But we will not prove this here.

If A is a 1×1 matrix $[t]$, then $e^A = [e^t]$, by the Maclaurin series formula for the function $y = e^t$. More generally, if D is a diagonal matrix having diagonal entries $d_1, d_2, \dots, d_n$, then we have

$$e^D = I + D + \frac{1}{2!} D^2 + \cdots = \begin{bmatrix} 1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1 \end{bmatrix} + \mathrm{diag}(d_1, \dots, d_n) + \mathrm{diag}\left(\frac{d_1^2}{2!}, \frac{d_1^2}{2!}, \dots, \frac{d_1^2}{2!}\right) = \mathrm{diag}\left(e^{d_1}, \dots, e^{d_n}\right)$$

The situation is more complicated for matrices that are not diagonal. However, if a matrix A happens to be diagonalizable, there is a simple algorithm for computing e^A, a consequence of the following lemma.

Let A and P be complex $n \times n$ matrices, and suppose that P is invertible. Then

$$\exp(P^{-1}AP) = P^{-1}\exp(A)P$$

Proof. Recall that, for all integers $m \geq 0$, we have $(P^{-1}AP)^m = P^{-1}A^mP$. The definition for exponential then yields

$$\begin{aligned}
\exp(P^{-1}AP) &= I + P^{-1}AP + \frac{1}{2!}(P^{-1}AP)^2 + \cdots \\
&= I + P^{-1}AP + \frac{1}{2!}P^{-1}A^2P + \cdots \\
&= P^{-1}\left(I + A + \frac{A^2}{2!} + \cdots\right)P \\
&= P^{-1}\exp(A)P
\end{aligned}$$

If a matrix A is diagonalizable, then there exists an invertible P so that $A = PDP^{-1}$, where D is a diagonal matrix of eigenvalues of A, and P is a matrix having eigenvectors of A as its columns. In this case, $e^A = Pe^DP^{-1}$. □

Let A denote the matrix

$$A = \begin{bmatrix} 5 & 1 \\ -2 & 2 \end{bmatrix}$$

You can asily verify that 4 and 3 are eigenvalues of A, with corresponding eigenvectors

$$w_1 = \begin{bmatrix} 1 \\ -1 \end{bmatrix} \quad \text{and} \quad w_2 = \begin{bmatrix} 1 \\ -2 \end{bmatrix}$$

It follows that

$$A = PDP^{-1} = \begin{bmatrix} 1 & 1 \\ -1 & 2 \end{bmatrix}\begin{bmatrix} 4 & 0 \\ 0 & 3 \end{bmatrix}\begin{bmatrix} 2 & 1 \\ -1 & -1 \end{bmatrix}$$

so that

$$\exp(A) = \begin{bmatrix} 1 & 1 \\ -1 & 2 \end{bmatrix}\begin{bmatrix} e^4 & 0 \\ 0 & e^3 \end{bmatrix}\begin{bmatrix} 2 & 1 \\ -1 & -1 \end{bmatrix} = \begin{bmatrix} 2e^4 - e^3 & e^4 - e^3 \\ 2e^3 - 2e^4 & 2e^3 - e^4 \end{bmatrix}$$

The definition of the exponential as a sum immediately reveals many other familiar properties. The following proposition is easy to prove:

Let $A \in \mathbb{C}^{n \times n}$.

1. If 0 denotes the zero matrix, then $e^0 = I$.
2. $A^m e^A = e^A A^m$ for all integers m
3. $(e^A)^T = e^{(A^T)}$
4. If $AB = BA$ then $Ae^B = e^B A$ and $e^A e^B = e^B e^A$.

Unfortunately not all familiar properties of the scalar exponential function $y = e^t$ carry over to the matrix exponential. For example, we know from calculus that $e^{s+t} = e^s e^t$ when s and t are numbers. However this is often not true for exponentials of matrices. In other words, it is possible to have $n \times n$ matrices A and B such that $e^{A+B} \neq e^A e^B$. Exactly when we have equality, $e^{A+B} = e^A e^B$, depends on specific properties of the matrices A and B. What do you think they are?

Let A and B be complex $n \times n$ matrices. If $AB = BA$ then $e^{A+B} = e^A e^B$.

Proof. **DIY exercise** □

4 Generalized Eigenvectors

Last time we talked about the case when A had distinct eigenvalues and we said you could simply diagonalize as $A = P\Lambda P^{-1}$ and then write

$$\exp(At) = P\text{diag}(\exp(\lambda_1 t), \dots, \exp(\lambda_n t))P^{-1}$$

Question: What about when A is not diagonalizable?

First, some preliminaries. Consider a vector space (V, F) and a linear map $\mathcal{A} : V \to V$.

Definition. (Invariant Subspaces.) A subspace $M \subset V$ is said to be A–invariant or invariant under A if given $x \in M$, $Ax \in M$. This is often written as $A[M] \subset M$ or even $AM \subset M$.

Example.

1. $\mathcal{N}(A)$ is A–invariant.
2. $\mathcal{R}(A)$ is A–invariant.
3. $\mathcal{N}(A - \lambda_i I)$ where $\lambda_i \in \sigma(A)$ is A–invariant.
4. If
$$p(A) = A^k + \alpha_1 A^{k-1} + \cdots + \alpha_{k-1}A + \alpha_k I$$
then, $\mathcal{N}(p(A))$ is A–invariant.
5. Let the subspaces M_1 and M_2 be A–invariant. Let
$$M_1 + M_2 = \{x \in V : \ x = x_1 + x_2, \ x_i \in M_i \text{ for } i = 1, 2\}$$
Then, $M_1 \cap M_2$ and $M_1 + M_2$ are A–invariant.

Definition. (Generalized Eigenvectors) Suppose λ is an eigenvalue of the square matrix A. We say that v is a generalised eigenvector of A with eigenvalue λ, if v is a nonzero element of the null space of $(A - \lambda I)^j$—i.e. $\mathcal{N}(A - \lambda I)^j$—for some positive integer j.

Fact. Null spaces eventually stabilize—that is, the null spaces $\mathcal{N}(A-\lambda I)^j$ are increasing with j and there is a unique positive integer k such that $\mathcal{N}(A - \lambda I)^j = \mathcal{N}(A - \lambda I)^k$ for all $j \geq k$.

Definition. (Generalized Eigenspace) Consider $A \in \mathbb{F}^{n\times n}$ with spectrum $\sigma(A) = \{\lambda_1, \dots, \lambda_k\}$. Define the generalized eigenspace pertaining to λ_i by

$$E_{\lambda_i} = \{x \in \mathbb{C}^n | \ (A - \lambda_i I)^n x = 0\}$$

Intuition: Observe that all the eigenvectors pertaining to λ_i are in E_{λ_i}. For a given E_{λ_i}, we can interpret the spaces in a hierarchical viewpoint. We know that E_{λ_i} contains all the eigenvectors pertaining to λ_i. Call these eigenvectors the **first order generalized eigenvectors**. If the span of these is not equal to E_{λ_i}, then there must be a vector $x \in E_{\lambda_i}$ for which $y = (A - \lambda_i I)^2 x = 0$ but $(A - \lambda_i I)x \neq 0$. That is to say y is an eigenvector of A pertaining to λ_i. Call such vectors **second order generalized eigenvectors**. In general, we call an $x \in E_{\lambda_i}$ a **generalized eigenvector of order p** if $y = (A - \lambda_i I)^p x = 0$ but $(A - \lambda_i I)^{p-1}x \neq 0$. For this reason we will call E_{λ_i} the space of generalized eigenvectors.

Fact. Let $A \in \mathbb{C}^{n\times n}$ with spectrum $\sigma(A) = \{\lambda_1, \dots, \lambda_k\}$ and invariant subspaces E_{λ_i}, $i \in \{1, \dots, k\}$.

1. Let $x \in E_{\lambda_i}$ be a generalized eigenvector of order p. Then the vectors
$$x, (A - \lambda_i I)x, (A - \lambda_i I)^2 x, \dots, (A - \lambda_i I)^{p-1}x \tag{1}$$
are linearly independent.
2. The subspace of $\mathbb{C}^n$ generated by the above vectors is an invariant subspace of A.

Example. Consider

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

- $$\chi_A(\lambda) = (\lambda - 3)(\lambda - 1)^2$$
- eigenvalues: $\lambda = 1, 3$
- eigenvectors:
$$\begin{aligned} \lambda_1 = 3: &\quad v_1 = (1,2,2) \\ \lambda_2 = 1: &\quad v_2 = (1,0,0) \end{aligned}$$
- The last generalized eigenvector will be a vector $v_3 \neq 0$ such that
$$(A - \lambda_2 I)^2 v_3 = 0$$
but
$$(A - \lambda_2 I) v_3 \neq 0$$
Pick $v_3 = (0,1,0)$. Note that $(A - \lambda_2 I) v_3 = v_2$.

Tip. How many powers of $(A - \lambda I)$ do we need to compute in order to find all of the generalized eigenvectors for λ?

If A is an $n \times n$ matrix and λ is an eigenvalue with algebraic multiplicity k, then the set of generalized eigenvectors for λ consists of the nonzero elements of $\mathcal{N}(A - \lambda I)^k$. In other words, we need to take at most k powers of $A - \lambda I$ to find all of the generalized eigenvectors for λ.

Yet another example. Determine generalized eigenvectors for the matrix

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

- single eigenvalue of $\lambda = 1$
- single eigenvector $v_1 = (-2, 0, 1)$
- now we look at
$$(A - I)^2 = \begin{bmatrix} 2 & 0 & 4 \\ 0 & 0 & 0 \\ -1 & 0 & -2 \end{bmatrix}$$
to get generalized eigenvector $v_2 = (0,1,0)$.
- Finally, $(A - I)^3 = 0$ so that $v_3 = (1,0,0)$.

5 Jordan Normal Form

To get some intuition for why we can find a form that looks like the Jordan form (i.e. a block diagonal decomposition) let us recall the following result.

First, recall the definition of the direct sum of subspaces:

Definition. (Direct Sum.) V is the direct sum of $M_1, M_2, \ldots, M_k$, denoted as

$$V = M_1 \oplus M_2 \oplus \cdots \oplus M_k$$

if for all $x \in V$, $\exists!\ x_i \in M_i$, $i = 1, \ldots, k$ such that

$$x = x_1 + x_2 + \cdots + x_k$$

Fact. The direct sum is the generalization of linear independence; e.g., check that if $V = M_1 \oplus \cdots \oplus M_k$, then $M_i \cap M_j = \{0\}$.

Theorem. (Second Representation Theorem.) Let $A : V \to V$ be a linear map. Let $V = M_1 \oplus M_2$ where $\dim V = n$, $\dim M_1 = k$, and $\dim M_2 = n - k$ be a finite dimensional vector space. If M_1 is A–invariant, then A has a representation of the form

$$A = \begin{bmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{bmatrix}$$

where $A_{11} \in F^{k\times k}$, $A_{12} \in F^{k\times(n-k)}$, $A_{22} \in F^{(n-k)\times(n-k)}$. Moreover, if both M_1 and M_2 are A–invariant then

$$A = \begin{bmatrix} A_{11} & 0 \\ 0 & A_{22} \end{bmatrix}$$

Proof. Let $\{b_1, b_2, \ldots, b_k\}$ be a basis for M_1 and let $\{b_{k+1}, \ldots, b_n\}$ be a basis for M_2. By assumption $V = M_1 \oplus M_2$ so that $\{b_i\}_{i=1}^n$ is a basis for V and any $x \in V$ has a unique representation as

$$x = \sum_{i=1}^{n} \xi_i b_i$$

Moreover, A has a matrix representation $A = (a_{ij})$ dictated by

$$Ab_j = \sum_{i=1}^{n} a_{ij} b_i \quad \forall\ j \tag{2}$$

Now for all $j = 1, \ldots, k$, $b_j \in M_1$ which is A–invariant so that $Ab_j \in M_1$ with basis $\{b_i\}_{i=1}^k$. Thus by (2), for all $j \in \{1, \ldots, k\}$

$$Ab_j = \sum_{i=1}^{k} a_{ij} b_i$$

i.e. $A_{ij} = 0$ for all $i = k + 1, \ldots, n$, for all $j = 1, \ldots, k$.

$\square$

Essentially what this is saying is that since M_1 is A–invariant, if I apply A to a basis vector in M_1 it has to stay in M_1 so any vector $x \in M_1$ written as $x = \sum_{i=1}^k \xi_i b_i$ is such that $Ax \in M_1$ with $Ax = \sum_{i=1}^k \eta_i b_i$ and no non-zero basis vectors are coming from the basis of M_2.

Why useful? We can use the second representation theorem applied to

$$\mathbb{C}^n = \mathcal{N}(A - \lambda_1 I)^{m_1} \oplus \mathcal{N}(A - \lambda_2 I)^{m_2} \oplus \cdots \oplus \mathcal{N}(A - \lambda_p I)^{m_p}$$

to write A via similarity transform into a matrix that has 'nice structure' (Jordan blocks) so that with respect to this structure e^{At} is easy to compute.

We are also going to use this result quite a bit in terms of decomposition of controllable and observable subspaces. So keep it in your pocket.

5.1 Minimal Polynomial

In order to show this decomposition, we need to revisit the characteristic polynomial and its cousin the minimal polynomial.

We know that

$$\det(sI - A) = \chi_A(s) \quad \text{(characteristic polynomial)}$$

We can write

$$\chi_A(s) = (s - \lambda_1)^{d_1}(s - \lambda_2)^{d_2} \cdots (s - \lambda_p)^{d_p}$$

where $d_1, \ldots, d_p$ are the *multiplicities* of $\lambda_1, \ldots, \lambda_p \in \mathbb{C}$, respectively and

$$\sum_{i=1}^{p} d_i = n$$

By Cayley-Hamilton, we know that

$$\chi_A(A) = 0_{n\times n}$$

Let $\psi_A(s)$ be the polynomial of least degree such that

$$\psi_A(A) = 0_{n\times n}$$

Definition. (Minimal Polynomial.) Given a matrix $A \in \mathcal{C}^{n\times n}$, we call *minimal polynomial* of A the annihilating polynomial ψ of least degree. The minimal polynomial is of the form

$$\psi_A(s) = (s - \lambda_1)^{m_1} \cdots (s - \lambda_p)^{m_p}$$

for some integers $m_i \leq d_i$.

Proposition. $\psi_A(s)$ divides $\chi_A(s)$

That is,

$$\frac{\chi_A(s)}{\psi_A(s)} = q(s)$$

for some polynomial $q(s)$.

Example.

1. Consider

$$A_1 = \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_1 & 0 \\ 0 & 0 & \lambda_2 \end{bmatrix}$$

   Then,

$$\chi_A(s) = (s - \lambda_1)^2(s - \lambda_2) \quad \text{and} \quad \psi_A(s) = (s - \lambda_1)(s - \lambda_2)$$

2. Consider

$$A_2 = \begin{bmatrix} \lambda_1 & 1 & 0 \\ 0 & \lambda_1 & 0 \\ 0 & 0 & \lambda_1 \end{bmatrix}$$

   Then,

$$\chi_A(s) = (s - \lambda_1)^3 \quad \text{and} \quad \psi_A(s) = (s - \lambda_1)^2$$
$$\psi_A(A) = (A - \lambda_1 I)(A - \lambda_1 I) = A^2 - 2\lambda_1 A + \lambda_1^2 I = 0$$

   where

$$A^2 = \begin{bmatrix} \lambda_1^2 & 2\lambda_1 & 0 \\ 0 & \lambda_1^2 & 0 \\ 0 & 0 & \lambda_1^2 \end{bmatrix}$$

3. Consider
$$A = \begin{bmatrix} 3 & 0 & 0 \\ 1 & 3 & 0 \\ 0 & 0 & 4 \end{bmatrix}$$
Then
$$\chi_A(s) = (s-3)^2(s-4)$$
The minimal polynomial is either
$$(s-3)(s-4)$$
or
$$(s-3)^2(s-4)$$
It cannot be the former since $(A-3I)(A-4I) \neq 0$.

That is, $\psi_A(s)$ is the least degree polynomial such that $\psi_A(A) = 0$.

Theorem (Decomposition)
$$\mathbb{C}^n = \mathcal{N}(A-\lambda_1 I)^{m_1} \oplus \mathcal{N}(A-\lambda_2 I)^{m_2} \oplus \cdots \oplus \mathcal{N}(A-\lambda_p I)^{m_p}$$

Proof.
$$\frac{1}{\psi_A(s)} = \frac{1}{(s-\lambda_1)_1^m \cdots (s-\lambda_p)^{m_p}} = \frac{n_1(s)}{(s-\lambda_1)^{m_1}} + \cdots + \frac{n_p(s)}{(s-\lambda_p)^{m_p}}$$
so that
$$1 = n_1(s)q_1(s) + \cdots n_p(s)q_p(s)$$
where
$$q_i(s) = \frac{\psi_A(s)}{(s-\lambda_i)^{m_i}}$$
Thus,
$$I = n_1(A)q_1(A) + \cdots + n_p(A)q_p(A)$$
so that
$$x = \underbrace{n_1(A)q_1(A)}_{x_1} x + \cdots + \underbrace{n_p(A)q_p(A)}_{x_1} x$$
which in turn implies that
$$x_i = n_i(A)q_i(A) = n_i(A)\frac{\psi_A(A)}{(A-\lambda_i I)^{m_i}}$$
so that
$$(A-\lambda_i I)^{m_i} x_i = 0_n \quad \implies \quad x_i \in \mathcal{N}(A-\lambda_i I)_i^m$$
To show the decomposition is unique, we argue by contradiction. Let
$$x_i \in \mathcal{N}(A-\lambda_i I)^{m_i}$$
so that
$$x_1 + \cdots + x_p = 0_n$$
and wlog assume $x_1 \neq 0$. Then
$$x_1 = -x_2 - x_3 - \cdots - x_p$$
so that
$$(A-\lambda_2 I)^{m_2} \cdots (A-\lambda_p I)^{m_p} x_1 = 0_n$$
But $q_1(s)$ and $(s-\lambda_1)^{m_1}$ are *co-prime* meaning that
$$h_1(s)q_1(s) + h_2(s)(s-\lambda_1)^{m_1} = 1$$

This implies that

$$h_1(A)\underbrace{q_1(A)x_1}_{0} + h_2(A)\underbrace{(A-\lambda_1 I)^{m_1}x_1}_{0} = x_1 \implies x_1 = 0 \ \rightarrow\leftarrow$$

□

Definition. (Multiplicities)

1. The geometric multiplicity of an eigenvalue λ is the dimension of E_λ.
2. The algebraic multiplicity of an eigenvalue λ is the number of times λ appears as a root of $\chi_A(\lambda)$.

Note. In general, the algebraic multiplicity and geometric multiplicity of an eigenvalue can differ. However, the geometric multiplicity can never exceed the algebraic multiplicity.

Fact. If for every eigenvalue of A, the geometric multiplicity equals the algebraic multiplicity, then A is said to be diagonalizable.

If the minimal polynomial is

$$\psi_A(\lambda) = \prod_{i=1}^{k}(\lambda-\lambda_i)^{m_i}$$

with $1 \leq m_i \leq d_i$ and d_i the algebraic multiplicity, then

$$N_i = \mathcal{N}((A-\lambda_i I)^{m_i})$$

is the algebraic eigenspace and $\mathcal{N}(A-\lambda_i I)$ is the geometric eigenspace with $\dim(N_i)$ the algebraic multiplicity and $\dim(\mathcal{N}(A-\lambda_i I))$ the geometric multiplicity.

Proposition. $\dim \mathcal{N}(A-\lambda_i I)^{m_i} = d_i$

Proof. see C& D, p.115 □

5.2 Jordan Form Details

Definition. (Jordan Block.) Let $\lambda \in \mathbb{C}$. A Jordan block $J_k(\lambda)$ is a $k \times k$ upper triangular matrix of the form

$$J_k(\lambda) = \begin{bmatrix} \lambda & 1 & 0 & \cdots & 0 \\ 0 & \lambda & 1 & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & \lambda & 1 \\ 0 & \cdots & \cdots & 0 & \lambda \end{bmatrix}$$

A Jordan matrix is any matrix of the form

$$J = \operatorname{diag}(J_{n_1}(\lambda_1), \ldots, J_{n_k}(\lambda_k))$$

where the matrices J_{n_1} are Jordan blocks. If $J \in \mathbb{C}^{n\times n}$, then $n_1 + n_2 + \cdots + n_k = n$.

Recall that

$$\chi_A(s) = \det(sI - A) = (s-\lambda_1)^{n_1}\cdots(s-\lambda_k)^{n_k}$$

When eigenvalues are distinct, $n_i = 1$ so that A is diagonalizable.

Theorem. (semisimple system) A square complex $n \times n$ matrix is *semisimple* if and only if there exists a nonsingular complex $n \times n$ matrix T^{-1} and diagonal complex $n \times n$ matrix Λ for which

$$A = T^{-1}\Lambda T$$

or equivalently

$$\Lambda = TAT^{-1}$$

The columns $e_i \in \mathbb{C}^n$ of T^{-1} organized as

$$T^{-1} = \begin{bmatrix} | & | & & | \\ e_1 & e_2 & \cdots & e_n \\ | & | & & | \end{bmatrix} \in \mathbb{C}^{n\times n}$$

and the diagonal entries $\lambda_i \in \mathbb{C}$ of Λ organized as

$$\Lambda = \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n) \in \mathbb{C}^{n\times n}$$

may be taken as n eigenvectors associated according to (1) with n eigenvalues λ_i of A that form a *spectral list*.

Note: We call *spectral list* of A any n–tuple $(\lambda_i)_{i=1}^n$ of eigenvalues that is complete as a list of roots of the characteristic polynomial χ_A

In other words, $A \in \mathbb{C}^{n\times n}$ is semisimple iff A is *diagonalizable by a similarity transformation.*

Example 1. Modal Decomposition. Consider

$$\begin{aligned} \dot{x} &= Ax + Bu \\ y &= Cx \end{aligned}$$

Define $z = Tx$. Then

$$\begin{aligned} \dot{z} &= TAT^{-1}z + TBu \\ y &= CT^{-1}z \end{aligned}$$

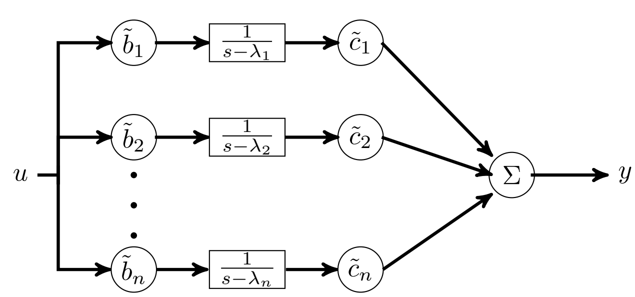
Now consider the case where we have a single input/single output (SISO, e.g., $m = 1 = p$). Define

$$TB = \begin{bmatrix} \tilde{b}_1 \\ \tilde{b}_2 \\ \vdots \\ \tilde{b}_n \end{bmatrix}, \ CT^{-1} = \begin{bmatrix} \tilde{c}_1 & \tilde{c}_2 & \cdots & \tilde{c}_n \end{bmatrix}$$

Then

$$C(sI - A)^{-1}b = \frac{\tilde{c}_1\tilde{b}_1}{s-\lambda_1} + \cdots + \frac{\tilde{c}_n\tilde{b}_n}{s-\lambda_n} = \sum_{i=1}^{n} \frac{\tilde{c}_i\tilde{b}_i}{s-\lambda_i}$$

which is called the *modal decomposition.*



Note: $c(sI-A)^{-1}b$ is called the transfer function. If $\tilde{c}_i$ or $\tilde{b}_i$ is zero, then the transfer function does not contain the term $1/(s-\lambda_i)$.

Theorem. (Jordan Normal Form.) Let $A \in \mathbb{C}^{n\times n}$. Then there is a non-singular matrix P such that

$$A = P\text{diag}(J_{n_1}(\lambda_1), \ldots, J_{n_k}(\lambda_k))P^{-1} = PJP^{-1}$$

where $J_{n_i}(\lambda_i)$ is a Jordan block and $\sum_{i=1}^k n_i = n$. The Jordan form J is unique up to permutations of the blocks. The eigenvalues $\lambda_1, \ldots, \lambda_k$ are not necessarily distinct. If A is real with real eigenvalues, then P can be taken as real.

Suppose that

$$P^{-1}AP = J = \text{diag}(J_1, \ldots, J_k)$$

where $J_i = J_{n_i}(\lambda_i)$. Express P as

$$P = \begin{bmatrix} P_1 & \cdots & P_k \end{bmatrix}$$

where $P_i \in \mathbb{C}^{n\times n_i}$ are the columns of P associated with i–th Jordan block J_i. We have that

$$AP_i = P_iJ_i$$

Let $P_i = [v_{i1}\ v_{i2}\ \cdots\ v_{in_i}]$ so that

$$Av_{i1} = \lambda_i v_{i1}$$

that is, the first column of each P_i is an eigenvector associated with eigenvalue λ_i. For $j = 2, \ldots, n_i$,

$$Av_{ij} = v_{ij-1} + \lambda_i v_{ij}$$

These $v_{i1}, \ldots, v_{in_i}$ are the generalized eigenvectors.

Example. Let A be an n by n square matrix. If its characteristic equation $\chi_A(t) = 0$ has a repeated root then A may not be diagonalizable, so we need the Jordan Canonical Form. Suppose λ is an eigenvalue of A, with multiplicity r as a root of $\chi_A(t) = 0$. The vector v is an eigenvector with eigenvalue λ if $Av = \lambda v$ or equivalently

$$(A - \lambda I)v = 0$$

The trouble is that this equation may have fewer than r linearly independent solutions for v. So we generalize and say that v is a generalized eigenvector with eigenvalue λ if

$$(A - \lambda I)^k v = 0$$

for some positive k. Now one can prove that there are exactly r linearly independent generalized eigenvectors. Finding the Jordan form is now a matter of sorting these generalized eigenvectors into an appropriate order.

To find the Jordan form carry out the following procedure for each eigenvalue λ of A.

1. First solve $(A - \lambda I)v = 0$, counting the number r_1 of linearly independent solutions.
2. If $r_1 = r$ good, otherwise $r_1 < r$ and we must now solve

   $$(A - \lambda I)^2 v = 0.$$

   There will be r_2 linearly independent solutions where $r_2 > r_1$.
3. If $r_2 = r$ good, otherwise solve

   $$(A - \lambda I)^3 v = 0$$

   to get $r_3 > r_2$ linearly independent solutions, and so on.
4. Eventually one gets $r_1 < r_2 < \cdots < r_{N-1} < r_N = r$.

Fact. The number N is the size of the largest Jordan block associated with λ, and r_1 is the total number of Jordan blocks associated to λ. If we define $s_1 = r_1$, $s_2 = r_2 - r_1$, $s_3 = r_3 - r_2$, …, $s_N = r_N - r_{N-1}$ then s_k is the number of Jordan blocks of size at least k by k associated to λ.

5. Finally put $m_1 = s_1 - s_2$, $m_2 = s_2 - s_3$, $\dots$, $m_{N-1} = s_{N-1} - s_N$ and $m_N = s_N$. Then m_k is the number of k by k Jordan blocks associated to λ. Once we've done this for all eigenvalues then we've got the Jordan form!

To find P such that $J = P^{-1}AP$ we need to do more work. We do the following for each eigenvalue λ:

1. First find the Jordan block sizes associated to λ by the above process. Put them in decreasing order
$$N_1 \geq N_2 \geq \cdots \geq N_k.$$
2. Now find a vector $v_{1,1}$ such that $(A - \lambda I)^{N_1} v_{1,1} = 0$ but $(A - \lambda I)^{N_1 - 1} v_{1,1} \neq 0$.
3. Define $v_{1,2} = (A - \lambda I)v_{1,1}$, $v_{1,3} = (A - \lambda I)v_{1,2}$ and so on until we get v_{1,N_1}. We can't go further because $(A - \lambda I)v_{1,N_1} = 0$.
4. If we only have one block we're OK, otherwise we can find $v_{2,1}$ such that $(A - \lambda I)^{N_2} v_{2,1} = 0$, and $(A - \lambda I)^{N_2 - 1} v_{2,1} \neq 0$ and AND $v_{2,1}$ is not linearly dependent on $v_{1,1}, \dots, v_{1,N_1}$.
5. Define $v_{2,2} = (A - \lambda I)v_{2,1}$ etc.
6. keep going if you have more blocks, otherwise you will have r linearly independent vectors $v_{1,1}, \dots, v_{k,N_k}$. Let
$$P_\lambda = \begin{bmatrix} v_{k,N_k} & \cdots & v_{1,1} \end{bmatrix}$$
be the n by r matrix.
7. do this for all λ's then stack the P_λ's up horizontally to get P

5.3 Functions of a matrix

Definition. (Functions of a matrix.) Let $\hat{f}(s)$ be any function of s analytic on the spectrum of A and $\hat{p}(s)$ be a polynomial such that
$$\hat{f}^k(\lambda_\ell) = \hat{p}^k(\lambda_\ell)$$
for $0 \leq k \leq m_\ell - 1$ and $1 \leq \ell \leq \sigma$. Then
$$\hat{f}(A) = \hat{p}(A)$$

In fact, if $m = \sum_{i=1}^{\sigma} m_i$ then
$$\hat{p}(s) = a_1 s^{m-1} + a_2 s^{m-2} + \cdots + a_m s^\sigma$$
where $a_1, a_2, \dots, a_n$ are functions of
$$(\hat{f}(\lambda_1), \hat{f}^1(\lambda_1), \hat{f}^2(\lambda_1), \dots, \hat{f}^{m_1}(\lambda_1), \hat{f}(\lambda_2), \dots)$$
and hence
$$\hat{f}(A) = a_1 A^{m-1} + \cdots + a_m A^0 = \sum_{\ell=1}^{\sigma} \sum_{k=0}^{m_\ell - 1} p_{k\ell}(A) f^k(\lambda_\ell)$$
where $p_{k\ell}$'s are polynomials independent of f.

Example. Define
$$J_2(\lambda, \varepsilon) = \begin{bmatrix} \lambda & 1 \\ 0 & \lambda + \varepsilon \end{bmatrix}$$
with eigenvalues $\lambda_1 = \lambda$ and $\lambda_2 = \lambda + \varepsilon$. For any $\varepsilon \neq 0$, $J_2(\lambda, \varepsilon)$ is diagonalizable. Computing eigenvector
$$\begin{aligned} [\lambda_1 I - J_2(\lambda, \varepsilon)]v_1 &= \begin{bmatrix} 0 & -1 \\ 0 & -\varepsilon \end{bmatrix} v_1 = 0 \implies v_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \\ \left[\lambda_2 I - J_2(\lambda, \varepsilon)\right] v_2 &= \begin{bmatrix} \varepsilon & -1 \\ 0 & 0 \end{bmatrix} v_2 = 0 \implies v_1 = \begin{bmatrix} 1 \\ \varepsilon \end{bmatrix} \end{aligned}$$
and
$$T = \begin{bmatrix} v_1 & v_2 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 0 & \varepsilon \end{bmatrix}, \; T^{-1} = \begin{bmatrix} 1 & -1/\varepsilon \\ 0 & 1/\varepsilon \end{bmatrix}$$

we can evaluate

$$f(J_2(\lambda,\varepsilon)) = Tf(\Lambda)T^{-1} = \begin{bmatrix} 1 & 1 \\ 0 & \varepsilon \end{bmatrix} \begin{bmatrix} f(\lambda) & 0 \\ 0 & f(\lambda+\varepsilon) \end{bmatrix} \begin{bmatrix} 1 & -1/\varepsilon \\ 0 & 1/\varepsilon \end{bmatrix} = \begin{bmatrix} f(\lambda) & (f(\lambda+\varepsilon)-f(\lambda))/\varepsilon \\ 0 & f(\lambda+\varepsilon) \end{bmatrix}$$

As $J_2(\lambda,\varepsilon) \to J_2(\lambda)$ as $\varepsilon \to 0$ and f is continuous, if f is also differentiable at λ

$$f(J_2(\lambda,\varepsilon)) = \lim_{\varepsilon\to 0} f(J_2(\lambda,\varepsilon)) = \begin{bmatrix} f(\lambda) & f'(\lambda) \\ 0 & f(\lambda) \end{bmatrix}$$

5.4 Functions of a matrix (repeated eigenvalues)

Theorem. (General Form of $f(A)$) Let $A \in \mathbb{C}^{n\times n}$ have a minimal polynomial ψ_A given by

$$\psi_A(s) = \prod_{k=1}^{\sigma} (s-\lambda_k)^{m_k}$$

Let the domain Δ contain $\sigma(A)$, then for any analytic function $f : \Delta \to \mathbb{C}$. we have

$$f(A) = \sum_{k=1}^{\sigma} \sum_{\ell=0}^{m_k-1} f^{(\ell)}(\lambda_k) p_{k\ell}(A)$$

where $p_{k\ell}$'s are polynomials independent of f.

Consider

$$J = \begin{bmatrix} \lambda & 1 & 0 & \cdots & 0 \\ 0 & \lambda & \ddots & \cdots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 1 \\ 0 & \cdots & \cdots & 0 & \lambda \end{bmatrix} \in F^{n\times n}$$

Claim:

$$f(J) = \begin{bmatrix} f(\lambda) & f'(\lambda) & \cdots & \frac{f^{(n-1)}(\lambda)}{(n-1)!} \\ 0 & \ddots & \ddots & \cdots \\ \vdots & \ddots & \ddots & f^1(\lambda) \\ 0 & \cdots & \cdots & f(\lambda) \end{bmatrix}$$

Proof. the minimum polynomial is $(s-\lambda)^n$. Thus,

$$f(J) = \sum_{\ell=0}^{n-1} f^{(\ell)}(\lambda) p_\ell(J)$$

Choose

$$f_1(s) = 1 \implies f_1(J) = I = f_1^{(0)} p_0(J) \implies p_0(J) = I$$

$$f_2(s) = s-\lambda \implies f_2(J) = J-\lambda I = f_2^{(1)}(\lambda)p_1(J) \implies p_1(J) = J-\lambda I$$

$$f_3(s) = (s-\lambda)^2 \implies f_3(J) = (J-\lambda I)^2 = f_3^{(2)}(\lambda)p_2(J) \implies 2p_2(J) = (J-\lambda I)^2$$

Hence

$$\begin{aligned} p_0(J) &= I \\ p_1(J) &= J - \lambda I \\ p_2(J) &= \frac{1}{2}(J - \lambda I)^2 \end{aligned}$$

Thus,

$$f(J) = \begin{bmatrix} f(\lambda) & f'(\lambda) & \frac{f''(\lambda)}{2} & \cdots & \frac{f^{(n-1)}(\lambda)}{(n-1)!} \\ 0 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \cdots & \ddots & \ddots & \frac{f''(\lambda)}{2} \\ \vdots & \cdots & \ddots & \ddots & f'(\lambda) \\ 0 & \cdots & \cdots & \cdots & f(\lambda) \end{bmatrix}$$

□

Hence we have

Theorem. (Spectral Mapping Theorem.)

$$\sigma(f(J)) = f(\sigma(J)) = \{f(\lambda), f(\lambda), \dots, f(\lambda)\}$$

and more generally that

$$\sigma(f(A)) = f(\sigma(A))$$

Example.

$$e^{J_i(\lambda t)} = \begin{bmatrix} e^{\lambda t} & te^{\lambda t} & \frac{t^2}{2!}e^{\lambda t} & \cdots & \frac{t^{i-1}}{(i-1)!}e^{\lambda t} \\ & \ddots & \ddots & \ddots & \vdots \\ & & \ddots & \ddots & \frac{t^2}{2!}e^{\lambda t} \\ & & & \ddots & te^{\lambda t} \\ 0 & & & e^{\lambda t} & \end{bmatrix}$$

What does this mean? Well more generally if we had

$$J = \text{diag}\left\{ \begin{bmatrix} \lambda_1 & 1 & 0 \\ 0 & \lambda_1 & 1 \\ 0 & 0 & \lambda_1 \end{bmatrix}, \begin{bmatrix} \lambda_1 & 1 \\ 0 & \lambda_1 \end{bmatrix}, \begin{bmatrix} \lambda_2 & 1 \\ 0 & \lambda_2 \end{bmatrix}, \lambda_2 \right\}$$

Recall that this Jordan form my be obtained from A by the similarity transform

$$J = TAT^{-1}$$

where

$$T^{-1} = \begin{bmatrix} e_1 & v_1 & w_1 & e_2 & v_2 & e_3 & v_3 & e_4 \end{bmatrix}$$

where $e_1, \dots, e_4$ are eigenvectors and the rest are generalized eigenvectors. Then

$$f(A) = f(T^{-1}JT) = T^{-1}f(J)T$$

where

$$f(J) = \text{diag}\left\{ \begin{bmatrix} f(\lambda_1) & f'(\lambda_1) & \frac{f''(\lambda_1)}{2} \\ 0 & f(\lambda_1) & f'(\lambda_1) \\ 0 & 0 & f(\lambda_1) \end{bmatrix}, \begin{bmatrix} f(\lambda_1) & f'(\lambda_1) \\ 0 & f(\lambda_1) \end{bmatrix}, \begin{bmatrix} f(\lambda_2) & f'(\lambda_2) \\ 0 & f(\lambda_2) \end{bmatrix}, f(\lambda_2) \right\}$$

Example. Compute e^{At} for

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

Eigenvalues:

$$\det(\lambda I - A) = \left| \begin{bmatrix} \lambda & -1 \\ 1 & \lambda_2 \end{bmatrix} \right| = \lambda^2 + 2\lambda + 1 = (\lambda+1)^2 = 0 \implies \lambda_1 = \lambda_2 = -1$$

Eigenvector:

$$(1I - A)v_1 = \begin{bmatrix} -1 & -1 \\ 1 & 1 \end{bmatrix} v_1 = 0, \implies v_1 = \begin{bmatrix} 1 \\ 01 \end{bmatrix}$$

Generalized eigenvector:

$$v_{11} = v_1,\ (A - 1I)v_{12} = \begin{bmatrix} 1 & 1 \\ -1 & -1 \end{bmatrix} v_{12} = \begin{bmatrix} 1 \\ -1 \end{bmatrix} v_{11}, \implies v_{12} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

Jordan form:

$$T = \begin{bmatrix} v_{11} & v_{12} \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ -1 & 0 \end{bmatrix},\ T^{-1} = \begin{bmatrix} 0 & -1 \\ 1 & 1 \end{bmatrix},\ J = \begin{bmatrix} -1 & 1 \\ 0 & -1 \end{bmatrix}$$

Matrix exponential:

$$e^{At} = Te^{J_2(-t)}T^{-1} = \begin{bmatrix} 1 & 1 \\ -1 & 0 \end{bmatrix} \begin{bmatrix} e^{-t} & te^{-t} \\ 0 & e^{-t} \end{bmatrix} \begin{bmatrix} 0 & -1 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} (t+1)e^{-t} & te^{-t} \\ -te^{-t} & (1-t)e^{-t} \end{bmatrix}$$