Linear Systems (034032)
lecture no. 10

Leonid Mirkin

Faculty of Mechanical Engineering
Technion—IIT

Outline

Frequency vs. step responses

State-space representation

Math background: linear algebra

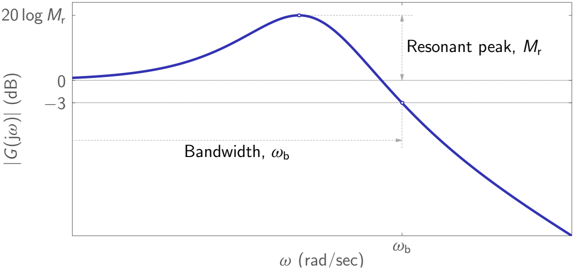
Outline

Frequency vs. step responses

State-space representation

Math background: linear algebra

Magnitude frequency response of low-pass filters



where

- bandwidth is the largest ω_b such that $|G(j\omega)| \geq 1/\sqrt{2}$ for all $\omega \leq \omega_b$
- resonance peak $M_r := \max_\omega |G(j\omega)| > 1$

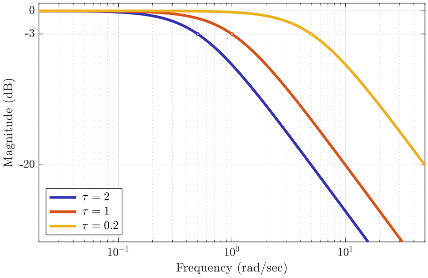
and we assume that $|G(0)| = 1$.

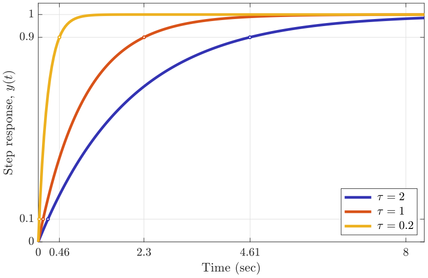
1-order systems: bandwidth vs. raise time

If

$$G(s) = \frac{1}{\tau s + 1}$$

then with $\tau \in \{0.2, 1, 2\}$,





showing that

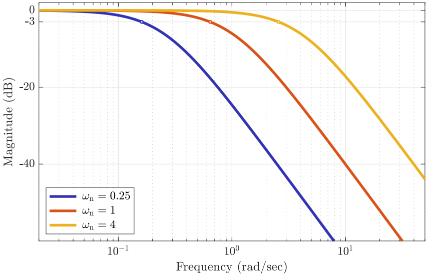
- wider $\omega_\text{b} \implies$ shorter t_r (faster transients)

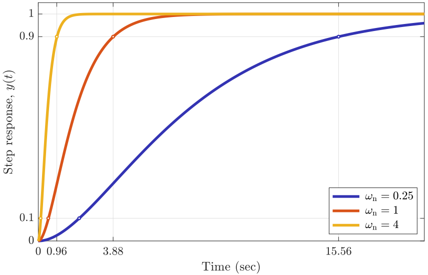
2-order systems: bandwidth vs. raise time

If

$$G(s) = \frac{\omega_n^2}{s^2 + 2\zeta\omega_n s + \omega_n^2}$$

then with $\zeta = 1$ and $\omega_n \in \{0.25, 1, 4\}$,





showing that

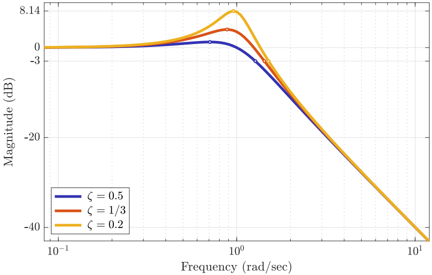
- wider $\omega_b \implies$ shorter t_r (faster transients)

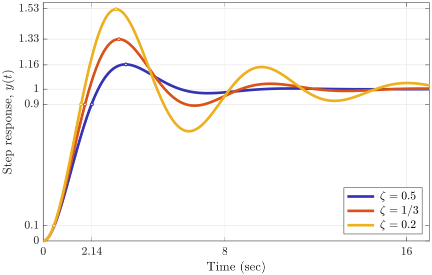
2-order systems: resonance vs. overshoot

If

$$G(s) = \frac{\omega_n^2}{s^2 + 2\zeta\omega_n s + \omega_n^2}$$

then with $\zeta \in \{0.5, 1/3, 0.2\}$ and $\omega_n = 1$,





showing that

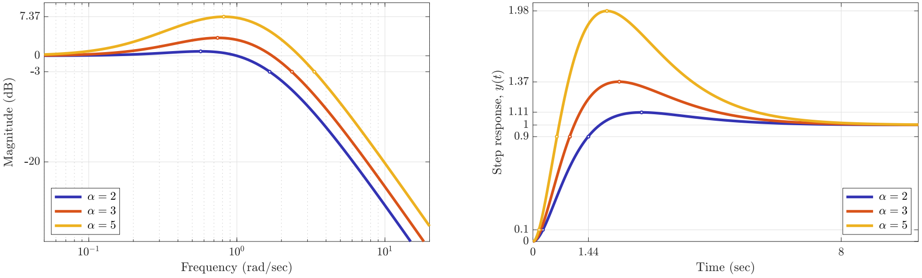
- larger $M_r \implies$ larger OS
- wider $\omega_b \implies$ shorter t_r (faster transients)

3-order systems with zeros

If

$$G(s) = \frac{\alpha \omega_n s + \omega_n^2}{(s/2+1)(s^2 + 2\zeta\omega_n s + \omega_n^2)}$$

then with $\zeta = 1$, $\omega_n = 1$, and $\alpha \in \{2, 3, 5\}$,



showing that

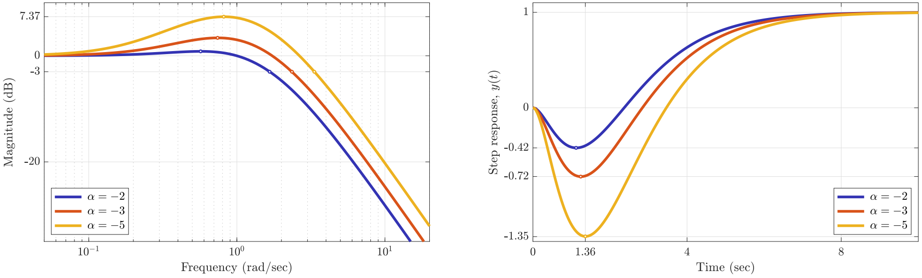
- larger $M_r \implies$ larger OS
- wider $\omega_b \implies$ shorter t_r (faster transients)

3-order systems with zeros (contd)

If

$$G(s) = \frac{\alpha\omega_n s + \omega_n^2}{(s/2+1)(s^2 + 2\zeta\omega_n s + \omega_n^2)}$$

then with $\zeta = 1$, $\omega_n = 1$, and $\alpha \in \{-2, -3, -5\}$,



showing that

- larger $M_r \implies$ larger US
- wider $\omega_b \implies$ faster leap (transients)

Rules of thumb

In general, we may *expect* that

- The higher M_r is, the larger the OS / US might be

  typically,

  - narrow peaks indicate oscillatory responses, with oscillation frequencies close to frequencies of peak on the Bode magnitude plot
  - wide peaks indicate overshoot / undershoot without oscillations

- The larger ω_b is, the faster time response is

  think of the Fourier transform frequency scaling property, $\mathfrak{F}\{\mathbb{P}_\varsigma y\} = \frac{1}{\varsigma}\,\mathbb{P}_{1/\varsigma}(\mathfrak{F}\{y\})$

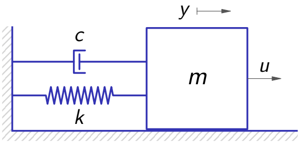
Rules of thumb

In general, we may *expect* that

- The higher M_{r} is, the larger the OS / US might be
  typically,
  - narrow peaks indicate oscillatory responses, with oscillation frequencies close to frequencies of peak on the Bode magnitude plot
  - wide peaks indicate overshoot / undershoot without oscillations
- The larger ω_{b} is, the faster time response is
  think of the Fourier transform frequency scaling property, $\mathfrak{F}\{\mathbb{P}_{\varsigma}y\} = \frac{1}{\varsigma}\,\mathbb{P}_{1/\varsigma}(\mathfrak{F}\{y\})$

Outline

Example 1: mass-spring-damper 1



Can be described by the second-order ODE $m\ddot{y}(t) + c\dot{y}(t) + ky(t) = u(t)$.
If we introduce the vector

$$x(t) := \begin{bmatrix} y(t) \\ \dot{y}(t) \end{bmatrix},$$

the system can be described by the *first-order matrix* ODE

$$\begin{cases} \dot{x}(t) = \begin{bmatrix} 0 & 1 \\ -k/m & -c/m \end{bmatrix} x(t) + \begin{bmatrix} 0 \\ 1/m \end{bmatrix} u(t) \\ y(t) = \begin{bmatrix} 1 & 0 \end{bmatrix} x(t) \end{cases}$$

Example 2: Fibonacci series

The Fibonacci series can be described as the impulse response of the system $G : u \mapsto y$ described by the second-order difference equation

$$y[t+2] - y[t+1] - y[t] = u[t+1]$$

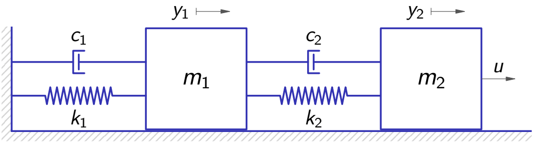
(see Tutorial 6). If we introduce the vector

$$x[t] := \begin{bmatrix} y[t] \\ y[t+1] - u[t] \end{bmatrix},$$

the system can be described by the *first-order matrix* difference equation

$$\begin{cases} x[t+1] = \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix} x[t] + \begin{bmatrix} 1 \\ 1 \end{bmatrix} u[t] \\ y[t] = \begin{bmatrix} 1 & 0 \end{bmatrix} x[t] \end{cases}$$

Example 3: mass-spring-damper 3



Assuming zero spring and dashpot forces at $y_1 = 0$ and $y_2 = \xi > 0$,

$$\begin{bmatrix} m_1 & 0 \\ 0 & m_2 \end{bmatrix} \begin{bmatrix} \ddot{y}_1(t) \\ \ddot{y}_2(t) \end{bmatrix} + \begin{bmatrix} c_1 + c_2 & -c_2 \\ -c_2 & c_2 \end{bmatrix} \begin{bmatrix} \dot{y}_1(t) \\ \dot{y}_2(t) \end{bmatrix} + \begin{bmatrix} k_1 + k_2 & -k_2 \\ -k_2 & k_2 \end{bmatrix} \begin{bmatrix} y_1(t) \\ y_2(t) \end{bmatrix} = \begin{bmatrix} -k_2\xi \\ u(t) + k_2\xi \end{bmatrix}.$$

Example 3: mass-spring-damper 3 (contd)

Define

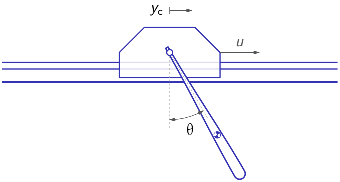
$$x(t) = \begin{bmatrix} x_1(t) \\ x_2(t) \\ x_3(t) \\ x_4(t) \end{bmatrix} := \begin{bmatrix} y_1(t) \\ y_2(t) \\ \dot{y}_1(t) \\ \dot{y}_2(t) \end{bmatrix}$$

The dynamics of the system can be written as the *first-order matrix* ODE[1]

$$\begin{cases} \dot{x}(t) = \underbrace{\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & m_1 & 0 \\ 0 & 0 & 0 & m_2 \end{bmatrix}^{-1} \begin{bmatrix} x_3 \\ x_4 \\ c_2(x_4 - x_3) - c_1 x_3 + k_2(x_2 - x_1 - \xi) - k_1 x_1 \\ c_2(x_3 - x_4) + k_2(x_1 - x_2 + \xi) + u(t) \end{bmatrix}}_{f_{\mathrm{msd3}}(x,u)} \\ y(t) = \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} \end{cases}$$

[1] The time argument of $x_i(t)$ in the right-hand side is dropped due to space limitations.

Example 4: pendulum on cart



Can be described by the ODE (see Tutorial 1)

$$\begin{cases} (M+m)\ddot{y}_c(t) + (ml\cos\theta(t))\ddot{\theta}(t) - ml\dot{\theta}^2(t)\sin\theta(t) = u(t) \\ (ml\cos\theta(t))\ddot{y}_c(t) + (J+ml^2)\ddot{\theta}(t) + mgl\sin\theta(t) = 0 \end{cases}$$

where

- M and m are cart and pendulum masses, respectively
- J is the moment of inertial of the pendulum about its center of mass
- l is the distance from the pendulum center of mass to its pivot point
- g is the standard gravity

Example 4: pendulum on cart (contd)

Define

$$x(t) = \begin{bmatrix} x_1(t) \\ x_2(t) \\ x_3(t) \\ x_4(t) \end{bmatrix} := \begin{bmatrix} y_c(t) \\ \theta(t) \\ \dot{y}_c(t) \\ \dot{\theta}(t) \end{bmatrix}$$

The dynamics of the system can be written as the *first-order matrix* ODE

$$\begin{cases} \dot{x}(t) = \underbrace{\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & M+m & ml\cos x_2(t) \\ 0 & 0 & ml\cos x_2(t) & J+ml^2 \end{bmatrix}^{-1} \begin{bmatrix} x_3(t) \\ x_4(t) \\ mlx_4^2(t)\sin x_2(t) + u(t) \\ -mlg\sin x_2(t) \end{bmatrix}}_{f_{\text{pend}}(x,u)} \\ y(t) = \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} \end{cases}$$

with the inverse well defined[2] for all $x_3 = \theta$.

[2] The determinant of the matrix to be inverted is $J(M+m) + (M + m\sin^2 x_3)ml^2 > 0$.

State-space equations

Equations of a system $G : u \mapsto y$ of the form

$$\begin{cases} \dot{x}(t) = Ax(t) + Bu(t) \\ y(t) = Cx(t) + Du(t) \end{cases} \quad \text{or} \quad \begin{cases} x[t+1] = Ax[t] + Bu[t] \\ y[t] = Cx[t] + Du[t] \end{cases}$$

are known as the state-space representations of G and the internal signal x is called its state vector.

State-space representations are widely used because they

- facilitate the use of powerful numerical tools in the analysis
- are readily extendible to MIMO systems
- are extendible to time-varying / nonlinear systems
  in the form $\dot{x}(t) = f(x, u, t)$ and $y(t) = h(x, u, t)$ for some functions g and h

Yet to understand basic properties of systems in state space

- we need linear algebra background.

State-space equations

Equations of a system $G : u \mapsto y$ of the form

$$\begin{cases} \dot{x}(t) = Ax(t) + Bu(t) \\ y(t) = Cx(t) + Du(t) \end{cases} \qquad \text{or} \qquad \begin{cases} x[t+1] = Ax[t] + Bu[t] \\ y[t] = Cx[t] + Du[t] \end{cases}$$

are known as the state-space representations of G and the internal signal x is called its state vector.

State-space representations are widely used because they

- facilitate the use of powerful numerical tools in the analysis
- are readily extendible to MIMO systems
- are extendible to time-varying / nonlinear systems
  in the form $\dot{x}(t) = f(x, u, t)$ and $y(t) = h(x, u, t)$ for some functions g and h

Yet to understand basic properties of systems in state space

- we need linear algebra background.

State-space equations

Equations of a system $G : u \mapsto y$ of the form

$$\begin{cases} \dot{x}(t) = Ax(t) + Bu(t) \\ y(t) = Cx(t) + Du(t) \end{cases} \quad \text{or} \quad \begin{cases} x[t+1] = Ax[t] + Bu[t] \\ y[t] = Cx[t] + Du[t] \end{cases}$$

are known as the state-space representations of G and the internal signal x is called its state vector.

State-space representations are widely used because they

- facilitate the use of powerful numerical tools in the analysis
- are readily extendible to MIMO systems
- are extendible to time-varying / nonlinear systems
  in the form $\dot{x}(t) = f(x, u, t)$ and $y(t) = h(x, u, t)$ for some functions g and h

Yet to understand basic properties of systems in state space

- we need linear algebra background.

Outline

Linear algebra, notation and terminology

A matrix $A \in \mathbb{R}^{n\times m}$ is an $n \times m$ table of real numbers,

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{1m} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nm} \end{bmatrix} := [a_{ij}],$$

accompanied by their manipulation rules (addition, multiplication, etc). The number of linearly independent rows (or columns) in A is called its rank and denoted $\operatorname{rank} A$. A matrix $A \in \mathbb{R}^{n\times m}$ is said to

- have full row (column) rank if $\operatorname{rank} A = n$ ($\operatorname{rank} A = m$)
- be square if $n = m$, tall if $n > m$, and fat if $n < m$
- be diagonal ($A = \operatorname{diag}\{a_i\}$) if it is square and $a_{ij} = 0$ whenever $i \neq j$
- be lower (upper) triangular if $a_{ij} = 0$ whenever $i > j$ ($i < j$)
- be symmetric if $A = A'$, where the transpose $A' := [a_{ji}]$

The identity matrix $I_n := \operatorname{diag}\{1\} \in \mathbb{R}^{n\times n}$ (if the dimension is clear, just I).

Linear algebra, notation and terminology (contd)

Let $A \in \mathbb{R}^{n\times n}$ (square). Its

- determinant, $\det(A)$ definition is long, but you shall know it already
- trace, $\mathrm{tr}(A) := \sum_{i=1}^{n} a_{ii}$
- inverse, A^{-1} such that $A^{-1}A = I$; exists iff $\det(A) \neq 0$; $(A^{-1})^{-1} = A$
- power, $A^k := \underbrace{A \cdot A \cdot \ldots \cdot A}_{k \text{ times}}$ for every $k \in \mathbb{N}$ and $A^0 = I$ whenever $A \neq 0$

Power properties:

- $A^k A^r = A^{k+r}$
- if A is diagonal, then $A^k = [a_i^k]$ is diagonal as well
- if A is triangular, then A^k is triangular as well, with a_{ii}^k on the diagonal

Also,

- A is said to be nilpotent if $\exists k \in \mathbb{N}$ such that $A^k = 0$
- $A_1, A_2 \in \mathbb{R}^{n\times n}$ are said to commute if $A_1A_2 = A_2A_1$
  αI_n for $\alpha \in \mathbb{R}$ commutes with every $n \times n$ matrix; A^k and A^l commute $\forall k, l \in \mathbb{Z}$
- $A_1, A_2 \in \mathbb{R}^{n\times n}$ are said to be similar if there is a nonsingular $T \in \mathbb{R}^{n\times n}$ such that $A_1 = TA_2T^{-1}$ or, equivalently, $A_1T = TA_2$

Eigenvalues

Given a square matrix $A \in \mathbb{R}^{n \times n}$, its eigenvalues are the solutions $\lambda \in \mathbb{C}$ to

$$\chi_A(\lambda) := \det(\lambda I - A) = \lambda^n + \chi_{n-1}\lambda^{n-1} + \cdots + \chi_1\lambda + \chi_0 = 0$$

(characteristic equation). The set of all eigenvalues of a matrix A is dubbed its spectrum, $\text{spec}(A)$. The spectral radius $\rho(A) := \max_{\lambda \in \text{spec}(A)} |\lambda|$.

Some facts:

- every $n \times n$ matrix A has n (not necessarily distinct) eigenvalues
- if $\lambda_i \in \text{spec}(A)$ is such that $\text{Im}\,\lambda_i \neq 0$, then $\overline{\lambda_i} \in \text{spec}(A)$ as well
- $\lambda_i \in \text{spec}(A) \implies \lambda_i t \in \text{spec}(At)$, $\forall t \in \mathbb{R}$
- $\lambda_i \in \text{spec}(A) \implies \lambda_i \in \text{spec}(TAT^{-1})$, $\forall T$ such that $\det(T) \neq 0$
- $\prod_{i=1}^{n} \lambda_i = \det(A)$
- $\sum_{i=1}^{n} \lambda_i = \text{tr}(A)$

If A is diagonal or triangular, then

- its eigenvalues equal the diagonal elements, i.e. $\lambda_i = a_{ii}$

Eigenvalues

Given a square matrix $A \in \mathbb{R}^{n\times n}$, its eigenvalues are the solutions $\lambda \in \mathbb{C}$ to

$$\chi_A(\lambda) := \det(\lambda I - A) = \lambda^n + \chi_{n-1}\lambda^{n-1} + \cdots + \chi_1\lambda + \chi_0 = 0$$

(characteristic equation). The set of all eigenvalues of a matrix A is dubbed its spectrum, $\text{spec}(A)$. The spectral radius $\rho(A) := \max_{\lambda\in\text{spec}(A)}|\lambda|$.

Some facts:

- every $n \times n$ matrix A has n (not necessarily distinct) eigenvalues
- if $\lambda_i \in \text{spec}(A)$ is such that $\text{Im}\,\lambda_i \neq 0$, then $\overline{\lambda_i} \in \text{spec}(A)$ as well
- $\lambda_i \in \text{spec}(A) \implies \lambda_i t \in \text{spec}(At)$, $\forall t \in \mathbb{R}$
- $\lambda_i \in \text{spec}(A) \implies \lambda_i \in \text{spec}(TAT^{-1})$, $\forall T$ such that $\det(T) \neq 0$
- $\prod_{i=1}^n \lambda_i = \det(A)$
- $\sum_{i=1}^n \lambda_i = \text{tr}(A)$

If A is diagonal or triangular, then

- its eigenvalues equal the diagonal elements, i.e. $\lambda_i = a_{ii}$

Eigenvalues: multiplicity

Given $A \in \mathbb{R}^{n \times n}$ and $\lambda_i \in \operatorname{spec}(A)$,

- algebraic multiplicity of λ_i is its multiplicity in $\chi_A(\lambda)$
- geometric multiplicity of λ_i is $n - \operatorname{rank}(\lambda_i I - A)$

They need not coincide. If $\exists \lambda_i \in \operatorname{spec}(A)$ such that its algebraic multiplicity is larger than geometric one, then A is said to be *defective*.

Example 1: if $A = \left[\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right]$, then $\chi_A(s) = (s-1)^2$ and

$$n - \operatorname{rank}(\lambda_i I - A) = 2 - \operatorname{rank}\left(\left[\begin{smallmatrix} 0 & 0 \\ 0 & 0 \end{smallmatrix}\right]\right) = 2$$

i.e. both algebraic and geometric multiplicity of $\lambda_i = 1$ are 2.

Example 2: if $A = \left[\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right]$, then $\chi_A(s) = (s-1)^2$ as well, but

$$n - \operatorname{rank}(\lambda_i I - A) = 2 - \operatorname{rank}\left(\left[\begin{smallmatrix} 0 & -1 \\ 0 & 0 \end{smallmatrix}\right]\right) = 1$$

i.e. the algebraic multiplicity of $\lambda_i = 1$ is 2, whereas the geometric one is 1. Hence, this matrix is defective.

Eigenvalues: multiplicity

Given $A \in \mathbb{R}^{n \times n}$ and $\lambda_i \in \text{spec}(A)$,

- algebraic multiplicity of λ_i is its multiplicity in $\chi_A(\lambda)$
- geometric multiplicity of λ_i is $n - \text{rank}(\lambda_i I - A)$

They need not coincide. If $\exists \lambda_i \in \text{spec}(A)$ such that its algebraic multiplicity is larger than geometric one, then A is said to be *defective*.

Example 1: if $A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$, then $\chi_A(s) = (s-1)^2$ and

$$n - \text{rank}(\lambda_i I - A) = 2 - \text{rank}\left(\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}\right) = 2$$

i.e. both algebraic and geometric multiplicity of $\lambda_i = 1$ are 2.

Example 2: if $A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$, then $\chi_A(s) = (s-1)^2$ as well, but

$$n - \text{rank}(\lambda_i I - A) = 2 - \text{rank}\left(\begin{bmatrix} 0 & -1 \\ 0 & 0 \end{bmatrix}\right) = 1$$

i.e. the algebraic multiplicity of $\lambda_i = 1$ is 2, whereas the geometric one is 1. Hence, this matrix is defective.

Eigenvectors

Right and left eigenvectors associated with $\lambda_i \in \mathrm{spec}(A)$ are nonzero vectors η_i and $\tilde{\eta}_i$, respectively, such that

$$(\lambda_i I - A)\eta_i = 0 \quad \text{and} \quad \tilde{\eta}_i'(\lambda_i I - A) = 0,$$

If η_i and $\tilde{\eta}_j$ are right and left eigenvectors associated with $\lambda_i \neq \lambda_j$, then

$$\tilde{\eta}_j' A \eta_i = \lambda_i \tilde{\eta}_j' \eta_i = \lambda_j \tilde{\eta}_j' \eta_i \iff (\lambda_i - \lambda_j)\tilde{\eta}_j' \eta_i = 0 \iff \tilde{\eta}_j' \eta_i = 0$$

i.e. right and left eigenvectors associated with distinct eigenvalues must be orthogonal (orthonormal, if they are normalized).

Diagonalization

If A is *not defective*, then it has n linearly independent eigenvectors and

$$A = \begin{bmatrix} \eta_1 & \cdots & \eta_n \end{bmatrix} \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{bmatrix} \begin{bmatrix} \eta_1 & \cdots & \eta_n \end{bmatrix}^{-1} = T \Lambda_A T^{-1}$$

or

$$A = \begin{bmatrix} \tilde{\eta}'_1 \\ \vdots \\ \tilde{\eta}'_n \end{bmatrix}^{-1} \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{bmatrix} \begin{bmatrix} \tilde{\eta}'_1 \\ \vdots \\ \tilde{\eta}'_n \end{bmatrix} = \tilde{T}^{-1} \Lambda_A \tilde{T}$$

Thus, non-defective matrices are *diagonalizable* by similarity transformation.
If all λ_i are distinct, then A is not defective and

$$A = \begin{bmatrix} \eta_1 & \cdots & \eta_n \end{bmatrix} \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{bmatrix} \begin{bmatrix} \tilde{\eta}'_1 \\ \vdots \\ \tilde{\eta}'_n \end{bmatrix} = T \Lambda_A \tilde{T}$$

assuming that all eigenvectors are normalized.

Matrix functions

Let $f(x) = \sum_{j=0}^{\infty} f_j x^j$ be analytic. Its matrix version $f(A)$ is defined as

$$f(A) := \sum_{j=0}^{\infty} f_j A^j.$$

The matrices A and $f(A)$ always commute.

If $\lambda_i \in \operatorname{spec}(A)$ with the right eigenvector η_i, then for all $j \in \mathbb{N}$

$$A^j \eta_i = \eta_i \lambda_i^j \quad \Longrightarrow \quad f(A)\eta_i = \sum_{j=0}^{\infty} f_j A^j \eta_i = \eta_i \sum_{j=0}^{\infty} f_j \lambda_i^j = \eta_i f(\lambda_i)$$

i.e. $f(\lambda_i) \in \operatorname{spec}(f(A))$ and η_i is the corresponding right eigenvector. In the same vein, we can see that a left eigenvector of A is that of $f(A)$. Also,

$$(TAT^{-1})^j = TA^jT^{-1} \quad \Longrightarrow \quad f(TAT^{-1}) = Tf(A)T^{-1}$$

for every analytic f.

Matrix functions

Let $f(x) = \sum_{j=0}^{\infty} f_j x^j$ be analytic. Its matrix version $f(A)$ is defined as

$$f(A) := \sum_{j=0}^{\infty} f_j A^j.$$

The matrices A and $f(A)$ always commute.

If $\lambda_i \in \mathrm{spec}(A)$ with the right eigenvector η_i, then for all $j \in \mathbb{N}$

$$A^j \eta_i = \eta_i \lambda_i^j \quad \Longrightarrow \quad f(A)\eta_i = \sum_{j=0}^{\infty} f_j A^j \eta_i = \eta_i \sum_{j=0}^{\infty} f_j \lambda_i^j = \eta_i f(\lambda_i)$$

i.e. $f(\lambda_i) \in \mathrm{spec}(f(A))$ and η_i is the corresponding right eigenvector. In the same vein, we can see that a left eigenvector of A is that of $f(A)$. Also,

$$(TAT^{-1})^j = TA^j T^{-1} \quad \Longrightarrow \quad f(TAT^{-1}) = Tf(A)T^{-1}$$

for every analytic f.

Example

Let

$$f(x) = \sin x = \sum_{j=0}^{\infty} \frac{(-1)^j}{(2j+1)!} x^{2j+1} \quad \text{and} \quad A = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}.$$

In this case

$$A^2 = \begin{bmatrix} 2 & 2 \\ 2 & 2 \end{bmatrix} = 2A, \quad A^3 = \begin{bmatrix} 4 & 4 \\ 4 & 4 \end{bmatrix} = 4A \quad \ldots \quad A^j = 2^{j-1} A$$

so that

$$f(A) = \sum_{j=0}^{\infty} \frac{(-1)^j}{(2j+1)!} A^{2j+1} = \frac{1}{2} \sum_{j=0}^{\infty} \frac{(-1)^j}{(2j+1)!} 2^{2j+1} A = \frac{\sin 2}{2} A$$

$$= \frac{\sin 2}{2} \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \neq \sin 1 \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} f(1) & f(1) \\ f(1) & f(1) \end{bmatrix}$$

Matrix functions via diagonalization

If $A = \text{diag}\{a_i\}$, then $A^j = \text{diag}\{a_i^j\}$ and

$$f(A) = \text{diag}\left\{\sum_{j=0}^{\infty} f_j a_i^j\right\} = \text{diag}\{f(a_i)\} = \begin{bmatrix} f(a_1) & & 0 \\ & \ddots & \\ 0 & & f(a_n) \end{bmatrix}$$

A special case of $A = \alpha I$ yields $f(A) = f(\alpha)I$ then.

Hence, if A is diagonalizable, i.e. there is T such that $A = T\Lambda_A T^{-1}$, then

$$f(A) = Tf(\Lambda_A)T^{-1} = T \begin{bmatrix} f(\lambda_1) & & 0 \\ & \ddots & \\ 0 & & f(\lambda_n) \end{bmatrix} T^{-1}$$

Example (contd)

Now note that

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

where the eigenvalues are $\lambda_1 = 2$ and $\lambda_2 = 0$ and

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

Hence,

$$\sin(A) = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} \sin 2 & 0 \\ 0 & \sin 0 \end{bmatrix} \frac{1}{2} \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} = \frac{\sin 2}{2} \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix},$$

exactly as we had before.

Cayley–Hamilton

In essence: every square matrix satisfies its own characteristic equation:

$$\chi_A(A) := A^n + \chi_{n-1}A^{n-1} + \cdots + \chi_1 A + \chi_0 I_n = 0.$$

Important consequences:

– A^k for all $k \geq n$ is a linear combination of A^i, $i \in \mathbb{Z}_{0..n-1}$, like

$$\begin{aligned} A^n &= -\chi_{n-1}A^{n-1} - \cdots - \chi_1 A - \chi_0 I_n \\ A^{n+1} &= -\chi_{n-1}A^n - \cdots - \chi_1 A^2 - \chi_0 A \\ &= \chi_{n-1}(\chi_{n-1}A^{n-1} + \cdots + \chi_1 A + \chi_0 I_n) - \cdots - \chi_1 A^2 - \chi_0 A \\ &= (\chi_{n-1}^2 - \chi_{n-2})A^{n-1} + \cdots + (\chi_{n-1}\chi_1 - \chi_0)A + \chi_{n-1}\chi_0 I_n \\ &\vdots \end{aligned}$$

– A^{-1}, if exists, is also a linear combination of A^i, $i \in \mathbb{Z}_{0..n-1}$:

$$A^{-1} = -\frac{1}{\chi_0}\bigl(\chi_1 I + \cdots + \chi_{n-1}A^{n-2} + A^{n-1}\bigr).$$

Matrix functions via Cayley–Hamilton

By CH,

$$f(A) = \sum_{j=0}^{\infty} f_j A^j = \sum_{j=0}^{n-1} g_j A^j$$

for some g_j, $j \in \mathbb{Z}_{0..n-1}$. To find those g_i, define

$$g(x) := \sum_{j=0}^{n-1} g_j x^j \qquad \text{so that } f(A) = g(A)$$

Although $g(x) \neq f(x)$ for every x,

$$f(A)\eta_i = g(A)\eta_i \iff f(\lambda_i)\eta_i = g(\lambda_i)\eta_i \iff f(\lambda_i) = g(\lambda_i)$$

for each eigenvalue-eigenvector pair. Hence,

$$f(\lambda_i) = \sum_{j=0}^{n-1} g_j \lambda_i^j = \begin{bmatrix} g_0 & g_1 & \cdots & g_{n-1} \end{bmatrix} \begin{bmatrix} 1 \\ \lambda_i \\ \vdots \\ \lambda_i^{n-1} \end{bmatrix}.$$

Matrix functions via Cayley–Hamilton (contd)

If all eigenvalues of A are simple, then we have exactly n equations

$$\begin{bmatrix} f(\lambda_1) & f(\lambda_2) & \cdots & f(\lambda_n) \end{bmatrix} = \begin{bmatrix} g_0 & g_1 & \cdots & g_{n-1} \end{bmatrix} \overbrace{\begin{bmatrix} 1 & 1 & \cdots & 1 \\ \lambda_1 & \lambda_2 & \cdots & \lambda_n \\ \vdots & \vdots & & \vdots \\ \lambda_1^{n-1} & \lambda_2^{n-1} & \cdots & \lambda_n^{n-1} \end{bmatrix}}^{\text{Vandermonde matrix, } V}$$

with $\det V = \prod_{1 \le i < j \le n} (\lambda_j - \lambda_i) \ne 0$. Hence,

$$\begin{bmatrix} g_0 & g_1 & \cdots & g_{n-1} \end{bmatrix} = \begin{bmatrix} f(\lambda_1) & f(\lambda_2) & \cdots & f(\lambda_n) \end{bmatrix} V^{-1},$$

which does not require to calculate eigenvectors.

Example (contd)

We already saw that

$$\text{spec}\left(\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}\right) = \{2, 0\} \implies V = \begin{bmatrix} 1 & 1 \\ 2 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0.5 \\ 1 & -0.5 \end{bmatrix}^{-1}$$

and

$$\begin{bmatrix} g_0 & g_1 \end{bmatrix} = \begin{bmatrix} \sin 2 & \sin 0 \end{bmatrix} \begin{bmatrix} 0 & 0.5 \\ 1 & -0.5 \end{bmatrix} = \begin{bmatrix} 0 & 0.5 \sin 2 \end{bmatrix}.$$

Hence,

$$\sin(A) = g_0 I_2 + g_1 A = \frac{\sin 2}{2} A = \frac{\sin 2}{2} \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix},$$

exactly as we had before, again.

What if there are eigenvalues with higher multiplicity?

Diagonalization is impossible for defective matrices. Rather, every matrix is similar to a block-diagonal form with $n_i \times n_i$ Jordan blocks

$$\begin{bmatrix} \lambda_i & 1 & 0 & \cdots & 0 \\ 0 & \lambda_i & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ 0 & 0 & 0 & \cdots & \lambda_i \end{bmatrix} = \lambda_i I_{n_i} + \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ 0 & 0 & 0 & \cdots & 0 \end{bmatrix} =: \lambda_i I_{n_i} + J_{0,n_i}$$

where $J_{0,n_i} \in \mathbb{R}^{n_i \times n_i}$ is nilpotent, $J_{0,n_i}^{n_i} = 0$.

If $\lambda_i \in \operatorname{spec}(A)$ has the multiplicity μ_i, then Cayley–Hamilton-based results based on $f(\lambda_i) = g(\lambda_i)$ shall be complemented with the conditions

$$\left.\frac{\mathrm{d}^j f(x)}{\mathrm{d}x^j}\right|_{x=\lambda_i} = \left.\frac{\mathrm{d}^j g(x)}{\mathrm{d}x^j}\right|_{x=\lambda_i}, \quad \forall j = 1, \ldots, \mu_i - 1$$

What if there are eigenvalues with higher multiplicity?

Diagonalization is impossible for defective matrices. Rather, every matrix is similar to a block-diagonal form with $n_i \times n_i$ Jordan blocks

$$\begin{bmatrix} \lambda_i & 1 & 0 & \cdots & 0 \\ 0 & \lambda_i & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ 0 & 0 & 0 & \cdots & \lambda_i \end{bmatrix} = \lambda_i I_{n_i} + \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ 0 & 0 & 0 & \cdots & 0 \end{bmatrix} =: \lambda_i I_{n_i} + J_{0,n_i}$$

where $J_{0,n_i} \in \mathbb{R}^{n_i \times n_i}$ is nilpotent, $J_{0,n_i}^{n_i} = 0$.

If $\lambda_i \in \operatorname{spec}(A)$ has the multiplicity μ_i, then Cayley–Hamilton-based results based on $f(\lambda_i) = g(\lambda_i)$ shall be complemented with the conditions

$$\left.\frac{\mathrm{d}^j f(x)}{\mathrm{d}x^j}\right|_{x=\lambda_i} = \left.\frac{\mathrm{d}^j g(x)}{\mathrm{d}x^j}\right|_{x=\lambda_i}, \quad \forall j = 1, \ldots, \mu_i - 1$$

Matrix exponential

The matrix exponential is defined (here $t \in \mathbb{R}$, we shall need it later on) as

$$\exp(At) = e^{At} := I + At + \frac{1}{2!}(At)^2 + \frac{1}{3!}(At)^3 + \cdots$$

Properties:

- $(e^{At})' = e^{A't}$
- e^{At} is nonsingular for every A, with $(e^{At})^{-1} = e^{-At}$
- $e^{A_1 t} e^{A_2 t} = e^{(A_1 + A_2)t}$ iff A_1 and A_2 commute
  in particular, $e^{\lambda t} e^{At} = e^{(\lambda I + A)t}$ and $e^{At_1} + e^{At_2} = e^{A(t_1 + t_2)}$
- if A is diagonal / triangular, then so is e^{At}, with diagonal elements $e^{a_{ii} t}$
- $e^{(\lambda I + J_{0,n})t} = e^{\lambda t} \begin{bmatrix} 1 & t & t^2/2! & \cdots & t^{n-1}/(n-1)! \\ 0 & 1 & t & \cdots & t^{n-2}/(n-2)! \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & t \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix}$
- Laplace transform $(\mathfrak{L}\{e^{At}\mathbb{1}\})(s) = (sI - A)^{-1}$, with RoC $= \mathbb{C}_{\max_i \operatorname{Re} \lambda_i}$

Example

Let

$$A = \begin{bmatrix} 0 & \omega \\ -\omega & 0 \end{bmatrix} = \begin{bmatrix} -\mathrm{j} & \mathrm{j} \\ 1 & 1 \end{bmatrix} \begin{bmatrix} \mathrm{j}\omega & 0 \\ 0 & -\mathrm{j}\omega \end{bmatrix} \begin{bmatrix} -\mathrm{j} & \mathrm{j} \\ 1 & 1 \end{bmatrix}^{-1}$$

Its exponential

$$\begin{aligned} \mathrm{e}^{At} &= \begin{bmatrix} -\mathrm{j} & \mathrm{j} \\ 1 & 1 \end{bmatrix} \begin{bmatrix} \mathrm{e}^{\mathrm{j}\omega t} & 0 \\ 0 & \mathrm{e}^{-\mathrm{j}\omega t} \end{bmatrix} \left(\frac{1}{2} \begin{bmatrix} \mathrm{j} & 1 \\ -\mathrm{j} & 1 \end{bmatrix} \right) \\ &= \frac{1}{2} \begin{bmatrix} \mathrm{e}^{\mathrm{j}\omega t} + \mathrm{e}^{-\mathrm{j}\omega t} & -\mathrm{j}(\mathrm{e}^{\mathrm{j}\omega t} - \mathrm{e}^{-\mathrm{j}\omega t}) \\ \mathrm{j}(\mathrm{e}^{\mathrm{j}\omega t} - \mathrm{e}^{-\mathrm{j}\omega t}) & \mathrm{e}^{\mathrm{j}\omega t} + \mathrm{e}^{-\mathrm{j}\omega t} \end{bmatrix} \\ &= \begin{bmatrix} \cos(\omega t) & \sin(\omega t) \\ -\sin(\omega t) & \cos(\omega t) \end{bmatrix} \end{aligned}$$

Hence,

$$\exp\left(\begin{bmatrix} \sigma & \omega \\ -\omega & \sigma \end{bmatrix} t \right) = \mathrm{e}^{\sigma t} \begin{bmatrix} \cos(\omega t) & \sin(\omega t) \\ -\sin(\omega t) & \cos(\omega t) \end{bmatrix}$$

where the equality $\mathrm{e}^{(\sigma I + A)t} = \mathrm{e}^{\sigma t}\mathrm{e}^{At}$ is used.

Example

Let

$$A = \begin{bmatrix} 0 & \omega \\ -\omega & 0 \end{bmatrix} = \begin{bmatrix} -\mathrm{j} & \mathrm{j} \\ 1 & 1 \end{bmatrix} \begin{bmatrix} \mathrm{j}\omega & 0 \\ 0 & -\mathrm{j}\omega \end{bmatrix} \begin{bmatrix} -\mathrm{j} & \mathrm{j} \\ 1 & 1 \end{bmatrix}^{-1}$$

Its exponential

$$\begin{aligned} \mathrm{e}^{At} &= \begin{bmatrix} -\mathrm{j} & \mathrm{j} \\ 1 & 1 \end{bmatrix} \begin{bmatrix} \mathrm{e}^{\mathrm{j}\omega t} & 0 \\ 0 & \mathrm{e}^{-\mathrm{j}\omega t} \end{bmatrix} \left(\frac{1}{2} \begin{bmatrix} \mathrm{j} & 1 \\ -\mathrm{j} & 1 \end{bmatrix} \right) \\ &= \frac{1}{2} \begin{bmatrix} \mathrm{e}^{\mathrm{j}\omega t} + \mathrm{e}^{-\mathrm{j}\omega t} & -\mathrm{j}(\mathrm{e}^{\mathrm{j}\omega t} - \mathrm{e}^{-\mathrm{j}\omega t}) \\ \mathrm{j}(\mathrm{e}^{\mathrm{j}\omega t} - \mathrm{e}^{-\mathrm{j}\omega t}) & \mathrm{e}^{\mathrm{j}\omega t} + \mathrm{e}^{-\mathrm{j}\omega t} \end{bmatrix} \\ &= \begin{bmatrix} \cos(\omega t) & \sin(\omega t) \\ -\sin(\omega t) & \cos(\omega t) \end{bmatrix} \end{aligned}$$

Hence,

$$\exp\left(\begin{bmatrix} \sigma & \omega \\ -\omega & \sigma \end{bmatrix} t \right) = \mathrm{e}^{\sigma t} \begin{bmatrix} \cos(\omega t) & \sin(\omega t) \\ -\sin(\omega t) & \cos(\omega t) \end{bmatrix}$$

where the equality $\mathrm{e}^{(\sigma I + A)t} = \mathrm{e}^{\sigma t}\mathrm{e}^{At}$ is used.

Matrix calculus

The derivative of a matrix $A(t)$ by a scalar $t \in \mathbb{R}$ is done component-wise,

$$A(t) = [a_{ij}(t)] \quad \Longrightarrow \quad \frac{\mathrm{d}}{\mathrm{d}t} A(t) = \left[\frac{\mathrm{d}a_{ij}(t)}{\mathrm{d}t} \right]$$

Some useful rules:

- $\frac{\mathrm{d}}{\mathrm{d}t}\big(A_1(t)A_2(t)\big) = \big(\frac{\mathrm{d}}{\mathrm{d}t}A_1(t)\big)A_2(t) + A_1(t)\big(\frac{\mathrm{d}}{\mathrm{d}t}A_2(t)\big)$
- $\frac{\mathrm{d}}{\mathrm{d}t}A^{-1}(t) = -A^{-1}(t)\big(\frac{\mathrm{d}}{\mathrm{d}t}A(t)\big)A^{-1}(t)$
- $\frac{\mathrm{d}}{\mathrm{d}t}(At)^k = A^k(kt^{k-1}) \quad \Longrightarrow \quad \frac{\mathrm{d}}{\mathrm{d}t}\mathrm{e}^{At} = A\mathrm{e}^{At} = \mathrm{e}^{At}A$

The derivative of $f : \mathbb{R}^m \to \mathbb{R}^n$ by its vector argument $x \in \mathbb{R}^m$,

$$\frac{\partial f(x)}{\partial x} = \left[\frac{\partial f_i(x)}{\partial x_j} \right] \in \mathbb{R}^{n \times m} \qquad \text{for every } x$$

i.e. it is a matrix-valued function, $\mathbb{R}^m \to \mathbb{R}^{n \times m}$, of x.

Matrix calculus

The derivative of a matrix $A(t)$ by a scalar $t \in \mathbb{R}$ is done component-wise,

$$A(t) = [a_{ij}(t)] \quad \Longrightarrow \quad \frac{\mathrm{d}}{\mathrm{d}t}A(t) = \left[\frac{\mathrm{d}a_{ij}(t)}{\mathrm{d}t}\right]$$

Some useful rules:

- $\frac{\mathrm{d}}{\mathrm{d}t}\big(A_1(t)A_2(t)\big) = \big(\frac{\mathrm{d}}{\mathrm{d}t}A_1(t)\big)A_2(t) + A_1(t)\big(\frac{\mathrm{d}}{\mathrm{d}t}A_2(t)\big)$
- $\frac{\mathrm{d}}{\mathrm{d}t}A^{-1}(t) = -A^{-1}(t)\big(\frac{\mathrm{d}}{\mathrm{d}t}A(t)\big)A^{-1}(t)$
- $\frac{\mathrm{d}}{\mathrm{d}t}(At)^k = A^k(kt^{k-1}) \quad \Longrightarrow \quad \frac{\mathrm{d}}{\mathrm{d}t}\mathrm{e}^{At} = A\mathrm{e}^{At} = \mathrm{e}^{At}A$

The derivative of $f : \mathbb{R}^m \to \mathbb{R}^n$ by its vector argument $x \in \mathbb{R}^m$,

$$\frac{\partial f(x)}{\partial x} = \left[\frac{\partial f_i(x)}{\partial x_j}\right] \in \mathbb{R}^{n\times m} \quad \text{for every } x$$

i.e. it is a matrix-valued function, $\mathbb{R}^m \to \mathbb{R}^{n\times m}$, of x.