Control Theory (035188)
lecture no. 6

Leonid Mirkin

Faculty of Mechanical Engineering
Technion — IIT

Outline

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). It is a convention that $A^0 = I$ for every nonzero square A.

Block matrices

Let $A \in \mathbb{R}^{n\times m}$ and let $\sum_{i=1}^{\nu} n_i = n$ and $\sum_{i=1}^{\mu} m_i = n$ for some $n_i, m_i \in \mathbb{N}$. Then A can be formally split as

$$A = \begin{bmatrix} A_{11} & A_{12} & \cdots & A_{1\mu} \\ A_{21} & A_{22} & \cdots & A_{1\mu} \\ \vdots & \vdots & & \vdots \\ A_{\nu 1} & A_{\nu 2} & \cdots & A_{\nu\mu} \end{bmatrix} = [A_{ij}]$$

for $A_{ij} = [a_{ij,kl}] \in \mathbb{R}^{n_i \times m_j}$ with entries $a_{ij,kl} = a_{pq}$, where $p = \sum_{r=1}^{i-1} n_r + k$ and $q = \sum_{r=1}^{j-1} m_r + l$. For example, we may present

$$A = \left[\begin{array}{cc:c:c} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \\ \hdashline 9 & 10 & 11 & 12 \end{array}\right] = \begin{bmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \end{bmatrix}$$

for $A_{11} = \left[\begin{smallmatrix} 1 & 2 \\ 5 & 6 \end{smallmatrix}\right]$, $A_{12} = \left[\begin{smallmatrix} 3 \\ 7 \end{smallmatrix}\right]$, $A_{13} = \left[\begin{smallmatrix} 4 \\ 8 \end{smallmatrix}\right]$, $A_{21} = \left[\begin{smallmatrix} 9 & 10 \end{smallmatrix}\right]$, $A_{22} = 11$, $A_{23} = 12$. This split is not unique and is normally done merely for convenience.

Block matrices (contd)

We say that a block matrix $A = [A_{ij}]$ is

- block-diagonal ($A = \operatorname{diag}\{A_i\}$) if $\nu = \mu$ and $A_{ij} = 0$ whenever $i \neq j$
- block lower (upper) triangular if $A_{ij} = 0$ whenever $i > j$ ($i < j$)

If A is square, i.e. $n = m$, a natural split is with $n_i = m_i$ for all i. Then all A_{ii} are square too.

Operations on block matrices are performed in terms of their blocks, like

$$\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} + \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix} = \begin{bmatrix} A_{11} + B_{11} & A_{12} + B_{12} \\ A_{21} + B_{21} & A_{22} + B_{22} \end{bmatrix},$$

$$\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix} = \begin{bmatrix} A_{11}B_{11} + A_{12}B_{21} & A_{11}B_{12} + A_{12}B_{22} \\ A_{21}B_{11} + A_{22}B_{21} & A_{21}B_{12} + A_{22}B_{22} \end{bmatrix}.$$

Sum, product, inverse of block-diagonal (block-triangular) matrices remain block-diagonal (block-triangular).

Eigenvalues & eigenvectors

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). There are n (not necessarily different) eigenvalues of A and $\operatorname{spec}(A)$ denotes their set. If $\lambda_i \in \operatorname{spec}(A)$ is such that $\operatorname{Im}\lambda_i \neq 0$, then $\bar{\lambda}_i \in \operatorname{spec}(A)$ too. Also, $\lambda_i \in \operatorname{spec}(A) \implies \lambda_i t \in \operatorname{spec}(At)$, $\forall t \in \mathbb{R}$.

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

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

respectively. Note that

$$A^k\eta_i = \eta_i\lambda_i^k \quad \text{and} \quad \tilde{\eta}_i' A^k = \lambda_i^k\tilde{\eta}_i', \qquad \forall k \in \mathbb{N}$$

so $\lambda_i \in \operatorname{spec}(A) \implies \lambda_i^k \in \operatorname{spec}(A^k)$, keeping left and right eigenvectors.

Eigenvalues & eigenvectors

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). There are n (not necessarily different) eigenvalues of A and $\operatorname{spec}(A)$ denotes their set. If $\lambda_i \in \operatorname{spec}(A)$ is such that $\operatorname{Im}\lambda_i \neq 0$, then $\bar{\lambda}_i \in \operatorname{spec}(A)$ too. Also, $\lambda_i \in \operatorname{spec}(A) \implies \lambda_i t \in \operatorname{spec}(At)$, $\forall t \in \mathbb{R}$.

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

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

respectively. Note that

$$A^k\eta_i = \eta_i\lambda_i^k \quad \text{and} \quad \tilde{\eta}_i' A^k = \lambda_i^k\tilde{\eta}_i', \qquad \forall k \in \mathbb{N}$$

so $\lambda_i \in \operatorname{spec}(A) \implies \lambda_i^k \in \operatorname{spec}(A^k)$, keeping left and right eigenvectors.

Eigenvalues & eigenvectors of block-triangular matrices

Let $A = \left[\begin{smallmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{smallmatrix}\right]$. Because

$$\begin{bmatrix} \lambda I - A_{11} & -A_{12} \\ 0 & \lambda I - A_{22} \end{bmatrix} \begin{bmatrix} \eta \\ 0 \end{bmatrix} = \begin{bmatrix} (\lambda I - A_{11})\eta \\ 0 \end{bmatrix},$$

every eigenvalue of A_{11} is that of A (its right eigenvector is $\left[\begin{smallmatrix} \eta_i \\ 0 \end{smallmatrix}\right]$). Because

$$\begin{bmatrix} 0 & \tilde{\eta}' \end{bmatrix} \begin{bmatrix} \lambda I - A_{11} & -A_{12} \\ 0 & \lambda I - A_{22} \end{bmatrix} = \begin{bmatrix} 0 & \tilde{\eta}'(\lambda I - A_{22}) \end{bmatrix},$$

every eigenvalue of A_{22} is that of A (its left eigenvector is $\begin{bmatrix} 0 & \tilde{\eta}_i' \end{bmatrix}$). Thus

- λ_i is an eigenvalue of A iff it is an eigenvalue of A_{11} or A_{22}.

The same arguments

- work for block-lower-triangular and block-diagonal matrices too.

Cayley–Hamilton

In essence: each 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, $i = 0, \ldots, n-1$, 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, $i = 0, \ldots, n-1$:

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

Matrix functions

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

$$f(A) := \sum_{i=0}^{\infty} f_i A^i.$$

It is readily seen that $f(A)\eta_i = \eta_i f(\lambda_i)$ for all eigenvalue-eigenvector pairs. By Cayley–Hamilton, we know that $\exists g_i$, $i = 0, \ldots, n-1$ such that

$$f(A) = \sum_{i=0}^{n-1} g_i A^i.$$

This does not imply that $f(x) = g(x) := \sum_{i=0}^{n-1} g_i x^i$ for all x, but still

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

for each eigenvalue λ_i of A having multiplicity μ_i. These equations can be used to calculate all n coefficients g_i and, hence, $f(A)$.

Matrix functions (contd)

If all n eigenvalues of A are different,

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

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

$$V := \begin{bmatrix} 1 & 1 & \cdots & 1 \\ \lambda_1 & \lambda_2 & \cdots & \lambda_n \\ \vdots & \vdots & & \vdots \\ \lambda_1^{n-1} & \lambda_2^{n-1} & \cdots & \lambda_n^{n-1} \end{bmatrix}$$

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

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

Matrix exponential

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

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

Example

Let $A = \left[\begin{smallmatrix} -1 & 1 \\ 0 & 2 \end{smallmatrix}\right]$ with eigenvalues $\lambda_1 = -1$ and $\lambda_2 = 2$. Then

$$\begin{bmatrix} g_0(t) & g_1(t) \end{bmatrix} = \begin{bmatrix} \mathrm{e}^{-t} & \mathrm{e}^{2t} \end{bmatrix} \begin{bmatrix} 1 & 1 \\ -1 & 2 \end{bmatrix}^{-1} \iff \begin{cases} g_0(t) = \frac{2}{3}\mathrm{e}^{-t} + \frac{1}{3}\mathrm{e}^{2t} \\ g_1(t) = -\frac{1}{3}\mathrm{e}^{-t} + \frac{1}{3}\mathrm{e}^{2t} \end{cases}$$

and hence

$$\exp\left(\begin{bmatrix} -1 & 1 \\ 0 & 2 \end{bmatrix} t \right) = g_0(t) \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + g_1(t) \begin{bmatrix} -1 & 1 \\ 0 & 2 \end{bmatrix} = \begin{bmatrix} \mathrm{e}^{-t} & \frac{1}{3}(\mathrm{e}^{2t} - \mathrm{e}^{-t}) \\ 0 & \mathrm{e}^{2t} \end{bmatrix}.$$

Note, the exponential of this triangular matrix is triangular too (general).

Matrix exponential

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

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

Example

Let $A = \left[\begin{smallmatrix} -1 & 1 \\ 0 & 2 \end{smallmatrix}\right]$ with eigenvalues $\lambda_1 = -1$ and $\lambda_2 = 2$. Then

$$\begin{bmatrix} g_0(t) & g_1(t) \end{bmatrix} = \begin{bmatrix} \mathrm{e}^{-t} & \mathrm{e}^{2t} \end{bmatrix} \begin{bmatrix} 1 & 1 \\ -1 & 2 \end{bmatrix}^{-1} \iff \begin{cases} g_0(t) = \frac{2}{3}\mathrm{e}^{-t} + \frac{1}{3}\mathrm{e}^{2t} \\ g_1(t) = -\frac{1}{3}\mathrm{e}^{-t} + \frac{1}{3}\mathrm{e}^{2t} \end{cases}$$

and hence

$$\exp\left(\begin{bmatrix} -1 & 1 \\ 0 & 2 \end{bmatrix} t\right) = g_0(t)\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + g_1(t)\begin{bmatrix} -1 & 1 \\ 0 & 2 \end{bmatrix} = \begin{bmatrix} \mathrm{e}^{-t} & \frac{1}{3}(\mathrm{e}^{2t} - \mathrm{e}^{-t}) \\ 0 & \mathrm{e}^{2t} \end{bmatrix}.$$

Note, the exponential of this triangular matrix is triangular too (general).

Similar matrices

Let $A \in \mathbb{R}^{n\times n}$ and $T \in \mathbb{R}^{n\times n}$ with $\det T \neq 0$. The matrices

$$A \quad \text{and} \quad TAT^{-1}$$

are said to be similar and $A \mapsto TAT^{-1}$ is called a similarity transformation. Similarity transformations do not affect characteristic equations:

$$\begin{aligned}\chi_{TAT^{-1}}(s) &= \det(\lambda I - TAT^{-1}) = \det\big(T(\lambda I - A)T^{-1}\big) \\ &= \det(T)\det(\lambda I - A)\det(T^{-1}) = \det(\lambda I - A) \\ &= \chi_A(s).\end{aligned}$$

Some definitions / facts:

- a matrix $A \in \mathbb{R}^{n\times n}$ is said to be diagonalizable if there is a $T \in \mathbb{R}^{n\times n}$ such that $TAT^{-1} = \operatorname{diag}\{\lambda_i\}$;
- symmetric matrices are always diagonalizable;
- $f(TAT^{-1}) = Tf(A)T^{-1}$ for any analytic function $f(x)$
  follows by the very fact that $(TAT^{-1})^k = TA^kT^{-1}$ for all $k \in \mathbb{N}$

Sign-definite matrices

A symmetric matrix $A = A' \in \mathbb{R}^{n\times n}$ is said to be

- positive definite ($A > 0$) if $x'Ax > 0$ for all $x \neq 0$,
- positive semidefinite ($A \geq 0$) if $x'Ax \geq 0$ for all x,
- negative definite ($A < 0$) if $x'Ax < 0$ for all $x \neq 0$ (or if $-A > 0$),
- negative semidefinite ($A \leq 0$) if $x'Ax \leq 0$ for all x (or $-A \geq 0$).

Clearly, $A > 0 \implies \det A \neq 0$. In fact,

- $A > 0$ (≥ 0) iff all its eigenvalues are positive (nonnegative)

Given two matrices $A = A' \in \mathbb{R}^{n\times n}$ and $B = B' \in \mathbb{R}^{n\times n}$, we say

- $A > B$ if $A - B > 0$,
- $A \geq B$ if $A - B \geq 0$,
- $A < B$ if $B - A > 0$ (equivalently, $A - B < 0$),
- $A \leq B$ if $B - A \geq 0$ (equivalently, $A - B \leq 0$).

Sign-definite matrices

A symmetric matrix $A = A' \in \mathbb{R}^{n \times n}$ is said to be

- positive definite ($A > 0$) if $x'Ax > 0$ for all $x \neq 0$,
- positive semidefinite ($A \geq 0$) if $x'Ax \geq 0$ for all x,
- negative definite ($A < 0$) if $x'Ax < 0$ for all $x \neq 0$ (or if $-A > 0$),
- negative semidefinite ($A \leq 0$) if $x'Ax \leq 0$ for all x (or $-A \geq 0$).

Clearly, $A > 0 \implies \det A \neq 0$. In fact,

- $A > 0$ (≥ 0) iff all its eigenvalues are positive (nonnegative)

Given two matrices $A = A' \in \mathbb{R}^{n \times n}$ and $B = B' \in \mathbb{R}^{n \times n}$, we say

- $A > B$ if $A - B > 0$,
- $A \geq B$ if $A - B \geq 0$,
- $A < B$ if $B - A > 0$ (equivalently, $A - B < 0$),
- $A \leq B$ if $B - A \geq 0$ (equivalently, $A - B \leq 0$).

Sign-definite matrices

A symmetric matrix $A = A' \in \mathbb{R}^{n \times n}$ is said to be

- positive definite ($A > 0$) if $x'Ax > 0$ for all $x \neq 0$,
- positive semidefinite ($A \geq 0$) if $x'Ax \geq 0$ for all x,
- negative definite ($A < 0$) if $x'Ax < 0$ for all $x \neq 0$ (or if $-A > 0$),
- negative semidefinite ($A \leq 0$) if $x'Ax \leq 0$ for all x (or $-A \geq 0$).

Clearly, $A > 0 \implies \det A \neq 0$. In fact,

- $A > 0$ (≥ 0) iff all its eigenvalues are positive (nonnegative)

Given two matrices $A = A' \in \mathbb{R}^{n \times n}$ and $B = B' \in \mathbb{R}^{n \times n}$, we say

- $A > B$ if $A - B > 0$,
- $A \geq B$ if $A - B \geq 0$,
- $A < B$ if $B - A > 0$ (equivalently, $A - B < 0$),
- $A \leq B$ if $B - A \geq 0$ (equivalently, $A - B \leq 0$).

Sign-definite matrices (contd)

Example

If $A = \left[\begin{smallmatrix} \alpha & -1 \\ -1 & 1 \end{smallmatrix}\right]$ for some $\alpha \in \mathbb{R}$, then

$$
\begin{aligned}
x'Ax &= \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} \alpha & -1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \alpha x_1^2 - 2x_1x_2 + x_2^2 \\
&= (\alpha - 1)x_1^2 + (x_1 - x_2)^2
\end{aligned}
$$

and if

- $\alpha > 1$ then $x'Ax > 0$ unless $x_1 = x_2 = 0$, for which $x'Ax = 0$ $\implies A > 0$
- $\alpha = 1$ then $x'Ax > 0$ unless $x_1 = x_2$, for which $x'Ax = 0$ $\implies A \geq 0$
- $\alpha < 1$ then $x'Ax$ might be both positive (e.g. if $x_1 = 0 \neq x_2$) and negative (e.g. if $x_1 = x_2 \neq 0$) $\implies A \not\geq 0$

Sign-definite block matrices

Let

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \quad \text{for square } A_{11} \text{ and } A_{22}$$

be symmetric, i.e. A_{11} and A_{22} are symmetric and $A_{12} = A_{21}'$. In this case

$$A > 0 \implies \begin{cases} A_{11} > 0 & \text{as } \begin{bmatrix} x_1' & 0 \end{bmatrix} A \begin{bmatrix} x_1 \\ 0 \end{bmatrix} = x_1' A_{11} x_1 > 0 \text{ for all } x_1 \neq 0 \\ A_{22} > 0 & \text{as } \begin{bmatrix} 0 & x_2' \end{bmatrix} A \begin{bmatrix} 0 \\ x_2 \end{bmatrix} = x_2' A_{22} x_2 > 0 \text{ for all } x_2 \neq 0 \end{cases}$$

Then (this technique is known as completing the square)

$$\begin{aligned} x'Ax &= x_1'A_{11}x_1 + x_1'A_{12}x_2 + x_2'A_{21}x_1 + x_2'A_{22}x_2 \\ &= x_1'(A_{11} - A_{12}A_{22}^{-1}A_{21})x_1 + x_1'A_{12}A_{22}^{-1}A_{21}x_1 + 2x_2'A_{21}x_1 + x_2'A_{22}x_2 \\ &= x_1'(A_{11} - A_{12}A_{22}^{-1}A_{21})x_1 + (x_1'A_{12}A_{22}^{-1} + x_2')A_{22}(A_{22}^{-1}A_{21}x_1 + x_2) \end{aligned}$$

Thus, $A > 0 \iff (A_{22} > 0) \wedge (A_{11} - A_{12}A_{22}^{-1}A_{21} > 0)$. We can see, by similar arguments, that $A > 0 \iff (A_{11} > 0) \wedge (A_{22} - A_{21}A_{11}^{-1}A_{12} > 0)$.

Sign-definite block matrices

Let

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \quad \text{for square } A_{11} \text{ and } A_{22}$$

be symmetric, i.e. A_{11} and A_{22} are symmetric and $A_{12} = A'_{21}$. In this case

$$A > 0 \implies \begin{cases} A_{11} > 0 & \text{as } \begin{bmatrix} x'_1 & 0 \end{bmatrix} A \begin{bmatrix} x_1 \\ 0 \end{bmatrix} = x'_1 A_{11} x_1 > 0 \text{ for all } x_1 \neq 0 \\ A_{22} > 0 & \text{as } \begin{bmatrix} 0 & x'_2 \end{bmatrix} A \begin{bmatrix} 0 \\ x_2 \end{bmatrix} = x'_2 A_{22} x_2 > 0 \text{ for all } x_2 \neq 0 \end{cases}$$

Then (this technique is known as completing the square)

$$\begin{aligned} x'Ax &= x'_1 A_{11} x_1 + x'_1 A_{12} x_2 + x'_2 A_{21} x_1 + x'_2 A_{22} x_2 \\ &= x'_1 (A_{11} - A_{12} A_{22}^{-1} A_{21}) x_1 + x'_1 A_{12} A_{22}^{-1} A_{21} x_1 + 2x'_2 A_{21} x_1 + x'_2 A_{22} x_2 \\ &= x'_1 (A_{11} - A_{12} A_{22}^{-1} A_{21}) x_1 + (x'_1 A_{12} A_{22}^{-1} + x'_2) A_{22} (A_{22}^{-1} A_{21} x_1 + x_2) \end{aligned}$$

Thus, $A > 0 \iff (A_{22} > 0) \wedge (A_{11} - A_{12} A_{22}^{-1} A_{21} > 0)$. We can see, by similar arguments, that $A > 0 \iff (A_{11} > 0) \wedge (A_{22} - A_{21} A_{11}^{-1} A_{12} > 0)$.

Sign-definite block matrices

Let

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \quad \text{for square } A_{11} \text{ and } A_{22}$$

be symmetric, i.e. A_{11} and A_{22} are symmetric and $A_{12} = A'_{21}$. In this case

$$A > 0 \implies \begin{cases} A_{11} > 0 & \text{as } \begin{bmatrix} x'_1 & 0 \end{bmatrix} A \begin{bmatrix} x_1 \\ 0 \end{bmatrix} = x'_1 A_{11} x_1 > 0 \text{ for all } x_1 \neq 0 \\ A_{22} > 0 & \text{as } \begin{bmatrix} 0 & x'_2 \end{bmatrix} A \begin{bmatrix} 0 \\ x_2 \end{bmatrix} = x'_2 A_{22} x_2 > 0 \text{ for all } x_2 \neq 0 \end{cases}$$

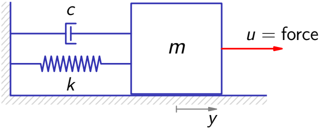
Then (this technique is known as completing the square)

$$\begin{aligned} x' A x &= x'_1 A_{11} x_1 + x'_1 A_{12} x_2 + x'_2 A_{21} x_1 + x'_2 A_{22} x_2 \\ &= x'_1 (A_{11} - A_{12} A_{22}^{-1} A_{21}) x_1 + x'_1 A_{12} A_{22}^{-1} A_{21} x_1 + 2 x'_2 A_{21} x_1 + x'_2 A_{22} x_2 \\ &= x'_1 (A_{11} - A_{12} A_{22}^{-1} A_{21}) x_1 + (x'_1 A_{12} A_{22}^{-1} + x'_2) A_{22} (A_{22}^{-1} A_{21} x_1 + x_2) \end{aligned}$$

Thus, $A > 0 \iff (A_{22} > 0) \wedge (A_{11} - A_{12} A_{22}^{-1} A_{21} > 0)$. We can see, by similar arguments, that $A > 0 \iff (A_{11} > 0) \wedge (A_{22} - A_{21} A_{11}^{-1} A_{12} > 0)$.

Outline

Spring-mass-damper system 1



Dynamics,

$$m\ddot{y}(t) + c\dot{y}(t) + ky(t) = u(t),$$

can be rewritten as

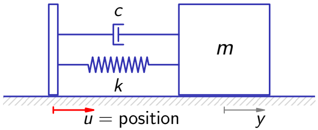
$$\begin{cases} \dot{x}(t) = \begin{bmatrix} 0 & 1 \\ -k/m & -c/m \end{bmatrix} x(t) + \begin{bmatrix} 0 \\ 1/m \end{bmatrix} u(t) \\ y(t) = \begin{bmatrix} 1 & 0 \end{bmatrix} x(t) \end{cases}$$

where

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

is the state vector, having clear physical meaning.

Spring-mass-damper system 2



Dynamics,

$$m\ddot{y}(t) + c\dot{y}(t) + ky(t) = c\dot{u}(t) + ku(t),$$

can be rewritten as

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

where

$$x(t) := \begin{bmatrix} v(t) \\ \dot{v}(t) \end{bmatrix} \quad \text{for } \ddot{v}(t) = u(t) - y(t)$$

is the state vector, having no clear physical meaning.

State-space realizations

A state-space description (realization) of a linear SISO system $G : u \mapsto y$ is

$$G : \begin{cases} \dot{x}(t) = Ax(t) + Bu(t), & x(0) = x_0 \\ y(t) = Cx(t) + Du(t), \end{cases}$$

where

- $u(t) \in \mathbb{R}$ is an input signal,
- $y(t) \in \mathbb{R}$ is an output signal,
- $x(t) \in \mathbb{R}^n$ is a state vector (internal variable) with initial condition x_0.

State and realization are not unique. For example, let $\tilde{x} := Tx$, then

$$G : \begin{cases} \dot{\tilde{x}}(t) = TAT^{-1}\tilde{x}(t) + TBu(t), & \tilde{x}(0) = Tx_0, \\ y(t) = CT^{-1}\tilde{x}(t) + Du(t) \end{cases}$$

describes the very same mapping $u \mapsto y$ (similarity transformation).

State-space realizations

A state-space description (realization) of a linear SISO system $G : u \mapsto y$ is

$$G : \begin{cases} \dot{x}(t) = Ax(t) + Bu(t), \quad x(0) = x_0 \\ y(t) = Cx(t) + Du(t), \end{cases}$$

where

- $u(t) \in \mathbb{R}$ is an input signal,
- $y(t) \in \mathbb{R}$ is an output signal,
- $x(t) \in \mathbb{R}^n$ is a state vector (internal variable) with initial condition x_0.

State and realization are not unique. For example, let $\tilde{x} := Tx$, then

$$G : \begin{cases} \dot{\tilde{x}}(t) = TAT^{-1}\tilde{x}(t) + TBu(t), \quad \tilde{x}(0) = Tx_0, \\ y(t) = CT^{-1}\tilde{x}(t) + Du(t) \end{cases}$$

describes the very same mapping $u \mapsto y$ (similarity transformation).

Solution of the state equation

The evolution of the state vector $x(t)$ at $t > 0$ is

$$x(t) = \mathrm{e}^{At}x_0 + \int_0^t \mathrm{e}^{A(t-s)}Bu(s)\mathrm{d}s.$$

Substituting this expression into the output equation, we end up with

$$y(t) = C\mathrm{e}^{At}x_0 + Du(t) + C\int_0^t \mathrm{e}^{A(t-s)}Bu(s)\mathrm{d}s.$$

This solution prompts the following interpretation of the state vector:

– the state vector $x(t)$ is a history accumulator: there is no need to know input *history* to calculate future outputs, the knowledge of the current state and *future* inputs is sufficient.

Indeed, given any t_0, then for all $t > t_0$

$$y(t) = \underbrace{C\mathrm{e}^{A(t-t_0)}x(t_0)}_{\text{history (up to } t_0)} + \underbrace{Du(t) + C\int_{t_0}^t \mathrm{e}^{A(t-s)}Bu(s)\mathrm{d}s}_{\text{future (from } t_0 \text{ and up to } t)}.$$

Solution of the state equation

The evolution of the state vector $x(t)$ at $t > 0$ is

$$x(t) = e^{At}x_0 + \int_0^t e^{A(t-s)}Bu(s)ds.$$

Substituting this expression into the output equation, we end up with

$$y(t) = Ce^{At}x_0 + Du(t) + C\int_0^t e^{A(t-s)}Bu(s)ds.$$

This solution prompts the following interpretation of the state vector:

- the state vector $x(t)$ is a history accumulator: there is no need to know input *history* to calculate future outputs, the knowledge of the current state and *future* inputs is sufficient.

Indeed, given any t_0, then for all $t > t_0$

$$y(t) = \underbrace{Ce^{A(t-t_0)}x(t_0)}_{\text{history (up to } t_0)} + \underbrace{Du(t) + C\int_{t_0}^t e^{A(t-s)}Bu(s)ds}_{\text{future (from } t_0 \text{ and up to } t)}.$$

Impulse response via state-space realizations

The response to $u(t) = \delta(t)$ and zero initial conditions $x_0 = 0$ is

$$y(t) = D\delta(t) + C\int_0^t \mathrm{e}^{A(t-s)}B\delta(s)\mathrm{d}s = D\delta(t) + C\mathrm{e}^{At}B.$$

If $D = 0$, this response is reminiscent of that to initial condition. Namely,

$$y(t) = C\mathrm{e}^{At}B = C\mathrm{e}^{At}x_0 \quad \text{whenever } x_0 = B.$$

This implies that

- the effect of initial conditions can also be expresses as the effect of external signal

(Dirac delta in this case).

From state space to transfer functions

The transfer function is the Laplace transform of the impulse response (with zero initial conditions). Hence,

$$G(s) = \mathfrak{L}\{D\delta(t) + C\mathrm{e}^{At}B\} = D + C(sI - A)^{-1}B.$$

This can also be seen via rewriting the state model in the s-domain:

$$\left.\begin{aligned} sX(s) &= AX(s) + BU(s) \\ Y(s) &= CX(s) + DU(s) \end{aligned}\right\} \quad \Longrightarrow \quad \left\{\begin{aligned} X(s) &= (sI - A)^{-1}BU(s) \\ Y(s) &= CX(s) + DU(s) \end{aligned}\right.$$

Because

$$D + C(sI - A)^{-1}B = D + \frac{C\,\mathrm{adj}(sI - A)B}{\det(sI - A)},$$

where $\mathrm{adj}(sI - A)$ is the adjugate matrix having polynomial entries,

$\Longrightarrow$ poles of $D + C(sI - A)^{-1}B$ are the eigenvalues of A, i.e. in $\mathrm{spec}(A)$ (it's a bit more complicated, we'll study that later on in the course).

From state space to transfer functions

The transfer function is the Laplace transform of the impulse response (with zero initial conditions). Hence,

$$G(s) = \mathfrak{L}\{D\delta(t) + C\mathrm{e}^{At}B\} = D + C(sI - A)^{-1}B.$$

This can also be seen via rewriting the state model in the s-domain:

$$\left.\begin{aligned} sX(s) &= AX(s) + BU(s) \\ Y(s) &= CX(s) + DU(s) \end{aligned}\right\} \quad \Longrightarrow \quad \left\{\begin{aligned} X(s) &= (sI - A)^{-1}BU(s) \\ Y(s) &= CX(s) + DU(s) \end{aligned}\right.$$

Because

$$D + C(sI - A)^{-1}B = D + \frac{C\,\mathrm{adj}(sI - A)B}{\det(sI - A)},$$

where $\mathrm{adj}(sI - A)$ is the adjugate matrix having polynomial entries,

– poles of $D + C(sI - A)^{-1}B$ are the eigenvalues of A, i.e. in $\mathrm{spec}(A)$

(it's a bit more complicated, we'll study that later on in the course).

Properness

Because

$$\lim_{|s|\to\infty} C(sI-A)^{-1}B = \lim_{|s|\to\infty} \tfrac{1}{s}C\left(I-\tfrac{1}{s}A\right)^{-1}B = \lim_{|z|\to 0} zC(I-zA)^{-1}B = 0$$

we have that

$$\lim_{|s|\to\infty} D + C(sI-A)^{-1}B = D.$$

Compare

$$\lim_{|s|\to\infty} \frac{b_n s^n + b_{n-1}s^{n-1} + \cdots + b_1 s + b_0}{s^n + a_{n-1}s^{n-1} + \cdots + a_1 s + a_0} = b_n.$$

Thus, the transfer function

- $D + C(sI-A)^{-1}B$ is strictly proper iff $D = 0$ (or bi-proper iff $D \neq 0$).

The parameter D is called the feed-through term of the realization.

Pole excess and high-frequency gain

Let

$$G(s) = \frac{b_m s^m + b_{m-1}s^{m-1} + \cdots + b_1 s + b_0}{s^n + a_{n-1}s^{n-1} + \cdots + a_1 s + a_0}$$

be strictly proper (i.e. $n > m$). Its pole excess, $n - m$, can be interpreted as

- the minimum $k \in \mathbb{N}$ such that $s^k G(s)$ is bi-proper.

Now, because

$$s(sI - A)^{-1} = I + A(sI - A)^{-1},$$

we have that

$$\begin{aligned}
sC(sI - A)^{-1}B &= CB + CA(sI - A)^{-1}B \\
s^2 C(sI - A)^{-1}B &= sCB + CAB + CA^2(sI - A)^{-1}B \\
&\vdots \\
s^k C(sI - A)^{-1}B &= \sum_{i=1}^{k-1} s^i CA^{k-1-i}B + CA^{k-1}B + CA^k(sI - A)^{-1}B
\end{aligned}$$

Pole excess and high-frequency gain (contd)

Thus, given $G(s) = D + C(sI - A)^{-1}B$ for[1] $D = 0$, its

- pole excess is the minimum $k \in \mathbb{N}$ such that $CA^{k-1}B \neq 0$.

The quantities $CA^{k-1}B$, $k \in \mathbb{N}$, are known as Markov parameters of $G(s)$.

The high-frequency gain of $G(s)$,

$$b_m = \lim_{|s| \to \infty} s^{n-m} G(s).$$

In state space,

$$s^{n-m} C(sI - A)^{-1} B = CA^{n-m-1}B + CA^{n-m}(sI - A)^{-1}B.$$

Hence, the high-frequency gain of $G(s) = C(sI - A)^{-1}B$ is

$$b_m = CA^{n-m-1}B,$$

where n is the dimension of A and $n - m$ is the pole excess of $G(s)$.

[1] If $D \neq 0$, then the pole excess is always zero.

Pole excess and high-frequency gain (contd)

Thus, given $G(s) = D + C(sI - A)^{-1}B$ for[1] $D = 0$, its

- pole excess is the minimum $k \in \mathbb{N}$ such that $CA^{k-1}B \neq 0$.

The quantities $CA^{k-1}B$, $k \in \mathbb{N}$, are known as Markov parameters of $G(s)$.

The high-frequency gain of $G(s)$,

$$b_m = \lim_{|s|\to\infty} s^{n-m}G(s).$$

In state space,

$$s^{n-m}C(sI - A)^{-1}B = CA^{n-m-1}B + CA^{n-m}(sI - A)^{-1}B.$$

Hence, the high-frequency gain of $G(s) = C(sI - A)^{-1}B$ is

$$b_m = CA^{n-m-1}B,$$

where n is the dimension of A and $n - m$ is the pole excess of $G(s)$.

[1] If $D \neq 0$, then the pole excess is always zero.

From transfer functions to state space

Let $G(s) = \dfrac{b_{n-1}s^{n-1} + \cdots + b_1 s + b_0}{s^n + a_{n-1}s^{n-1} + \cdots + a_1 s + a_0}$. Its state-space realization in

– companion form

$$\left[\begin{array}{c:c} A & B \\ \hdashline C & D \end{array}\right] = \left[\begin{array}{cccc:c} 0 & 1 & \ldots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \ldots & 1 & 0 \\ -a_0 & -a_1 & \ldots & -a_{n-1} & 1 \\ \hdashline b_0 & b_1 & \ldots & b_{n-1} & 0 \end{array}\right],$$

– observer form

$$\left[\begin{array}{c:c} A & B \\ \hdashline C & D \end{array}\right] = \left[\begin{array}{cccc:c} -a_{n-1} & 1 & \ldots & 0 & b_{n-1} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ -a_1 & 0 & \ldots & 1 & b_1 \\ -a_0 & 0 & \ldots & 0 & b_0 \\ \hdashline 1 & 0 & \ldots & 0 & 0 \end{array}\right].$$

If $b_n \neq 0$, $\frac{b_n s^n + b_{n-1}s^{n-1} + \cdots + b_0}{s^n + a_{n-1}s^{n-1} + \cdots + a_0} = b_n + \frac{(b_{n-1} - b_n a_{n-1})s^{n-1} + \cdots + b_0 - b_n a_0}{s^n + a_{n-1}s^{n-1} + \cdots + a_0}$ & $D = b_n$.

Outline

Problem

We know how to carry out parallel, cascade, and feedback interconnections in terms of transfer functions $G_1(s)$ and $G_2(s)$. Indeed:

- parallel: $G_1(s) + G_2(s)$
- series: $G_2(s)G_1(s)$
- feedback: $\dfrac{G_1(s)}{1 \mp G_1(s)G_2(s)}$

The question is: how this can be done in state space?

We assume that

$$G_1 : \begin{cases} \dot{x}_1(t) = A_1 x_1(t) + B_1 u_1(t) \\ y_1(t) = C_1 x_1(t) + D_1 u_1(t) \end{cases} \quad \text{and} \quad G_2 : \begin{cases} \dot{x}_2(t) = A_2 x_2(t) + B_2 u_2(t) \\ y_2(t) = C_2 x_2(t) + D_2 u_2(t) \end{cases}$$

A possible way of interconnecting systems is to

- determine what inputs / outputs to equate
- unite state vectors

Problem

We know how to carry out parallel, cascade, and feedback interconnections in terms of transfer functions $G_1(s)$ and $G_2(s)$. Indeed:

- parallel: $G_1(s) + G_2(s)$
- series: $G_2(s)G_1(s)$
- feedback: $\dfrac{G_1(s)}{1 \mp G_1(s)G_2(s)}$

The question is: how this can be done in state space?

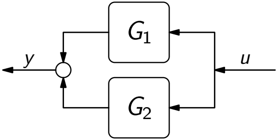
We assume that

$$G_1 : \begin{cases} \dot{x}_1(t) = A_1 x_1(t) + B_1 u_1(t) \\ y_1(t) = C_1 x_1(t) + D_1 u_1(t) \end{cases} \quad \text{and} \quad G_2 : \begin{cases} \dot{x}_2(t) = A_2 x_2(t) + B_2 u_2(t) \\ y_2(t) = C_2 x_2(t) + D_2 u_2(t) \end{cases}$$

A possible way of interconnecting systems is to

- determine what inputs / outputs to equate
- unite state vectors

Parallel interconnection



This corresponds to

- $u_1 = u_2 = u$
- $y = y_1 + y_2$

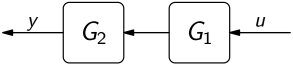
Then

$$G : \begin{cases} \begin{bmatrix} \dot{x}_1(t) \\ \dot{x}_2(t) \end{bmatrix} = \begin{bmatrix} A_1 & 0 \\ 0 & A_2 \end{bmatrix} \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} + \begin{bmatrix} B_1 \\ B_2 \end{bmatrix} u(t) \\ y(t) = \begin{bmatrix} C_1 & C_2 \end{bmatrix} \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} + (D_1 + D_2) u(t) \end{cases}$$

We already know (spectrum of block-diagonal matrices) that

- poles of $G(s)$ are $\mathrm{spec}(A_1) \cup \mathrm{spec}(A_2)$

Series / cascade interconnection



This corresponds to

- $u_1 = u$
- $u_2 = y_1$
- $y = y_2$

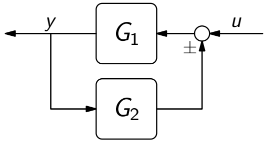
Then

$$G : \begin{cases} \begin{bmatrix} \dot{x}_1(t) \\ \dot{x}_2(t) \end{bmatrix} = \begin{bmatrix} A_1 & 0 \\ B_2 C_1 & A_2 \end{bmatrix} \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} + \begin{bmatrix} B_1 \\ B_2 D_1 \end{bmatrix} u(t) \\ y(t) = \begin{bmatrix} D_2 C_1 & C_2 \end{bmatrix} \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} + D_2 D_1 u(t) \end{cases}$$

We already know (spectrum of block-triangular matrices) that

- poles of $G(s)$ are $\mathrm{spec}(A_1) \cup \mathrm{spec}(A_2)$

Feedback interconnection



This corresponds to

- $u_1 = u \pm y_2$
- $u_2 = y_1$
- $y = y_1$

Then, assuming that $D_1 = 0$ for simplicity (enough for our purposes),

$$G : \begin{cases} \begin{bmatrix} \dot{x}_1(t) \\ \dot{x}_2(t) \end{bmatrix} = \begin{bmatrix} A_1 \pm B_1 D_2 C_1 & \pm B_1 C_2 \\ B_2 C_1 & A_2 \end{bmatrix} \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} + \begin{bmatrix} B_1 \\ 0 \end{bmatrix} u(t) \\ y(t) = \begin{bmatrix} C_1 & 0 \end{bmatrix} \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} \end{cases}$$

Inversion

Let

$$G : \begin{cases} \dot{x}(t) = Ax(t) + Bu(t) \\ y(t) = Cx(t) + Du(t) \end{cases}$$

be bi-proper (i.e. $D \neq 0$). Its inverse is the system mapping $y \mapsto u$. Then

$$u(t) = -D^{-1}Cx(t) + D^{-1}y(t)$$

and

$$\dot{x}(t) = Ax(t) + B\left(-D^{-1}Cx(t) + D^{-1}y(t)\right)$$

Therefore,

$$G^{-1} : \begin{cases} \dot{x}(t) = (A - BD^{-1}C)x(t) + BD^{-1}y(t) \\ u(t) = -D^{-1}Cx(t) + D^{-1}y(t) \end{cases}$$

and zeros of $G(s)$ (= poles of $G^{-1}(s)$) are in $\operatorname{spec}(A - BD^{-1}C)$.