Lecture Notes 06

Linear Systems: General Theory

Preliminaries: Matrix Exponentials

We have seen that

$$\dot{x} = ax, \quad x(0) = x_0$$

has solution $e^{at}x_0$. This week we will look at the vector (ordinary differential equation) ODE

$$\dot{\mathbf{x}} = A\mathbf{x}, \quad \mathbf{x}(0) = \mathbf{x}_0$$

where A is an $n \times n$ matrix and

$$\mathbf{x}(t) = \begin{bmatrix} x_1(t) \\ \vdots \\ x_n(t) \end{bmatrix}.$$

Could the solution to $\dot{\mathbf{x}} = A\mathbf{x}$ be $e^{At}\mathbf{x}_0$? And what does it even mean to write e^{At}?

Main Idea

We have learned that the Taylor series

$$\sum_{k=0}^{\infty} \frac{a^k}{k!} = 1 + a + \frac{a^2}{2} + \frac{a^3}{6} + \cdots$$

converges to a finite value for every complex value $a \in \mathbb{C}$. Furthermore, this limit value is exactly e^a. This implies the equality

$$e^a = \sum_{k=0}^{\infty} \frac{a^k}{k!}, \quad \text{for all} \quad a \in \mathbb{C}.$$

Now let A be an $n \times n$ matrix. Here A can be real or also have complex entries. We realize that as a square matrix, A^2, $A^3, \ldots$ are well defined. Based on this, we introduce the following matrix power series

$$\sum_{k=0}^{\infty} \frac{A^k}{k!} = I + A + \frac{A^2}{2} + \frac{A^3}{6} + \ldots.$$

Here I is the n-dimensional identity matrix. First of all, each term of the power series is well defined as long as A is a square matrix. Next, we can show that for any real or complex matrix A, this infinite power series of matrices converges to a finite matrix. This motivates us to formally define the exponential of a matrix A as per:

$$e^A = \sum_{k=0}^{\infty} \frac{A^k}{k!}$$

Example 1. Let A be a diagonal matrix of the form

$$A = \begin{bmatrix} b & 0 \\ 0 & c \end{bmatrix}.$$

Diect calculations give that:

$$A^2 = \begin{bmatrix} b^2 & 0 \\ 0 & c^2 \end{bmatrix}, \quad A^3 = \begin{bmatrix} b^3 & 0 \\ 0 & c^3 \end{bmatrix}, \quad \ldots$$

As a result,

$$e^A = \sum_{k=0}^{\infty} \frac{A^k}{k!} = I + A + \frac{A^2}{2} + \frac{A^3}{6} + \cdots = \begin{bmatrix} \sum_{k=0}^{\infty} b^k/k! & 0 \\ 0 & \sum_{k=0}^{\infty} c^k/k! \end{bmatrix} = \begin{bmatrix} e^b & 0 \\ 0 & e^c \end{bmatrix}.$$

Example 2. Let

$$A = \begin{bmatrix} 0 & a \\ 0 & 0 \end{bmatrix}.$$

Then $A^2 = 0$ (check yourself!), and we obtain

$$e^A = I + A + 0 + 0 + \cdots = \begin{bmatrix} 1 & a \\ 0 & 1 \end{bmatrix}.$$

Matrix exponentials do not inherit all properties of ordinary exponential!

- $e^{A+B} = e^A \cdot e^B$ if and only if $AB = BA$.
- If $AB = BA$, then we say that "A and B *commute*". Clearly, when both A and B are diagonal matrices, then they commute. Another common situation with A and B commuting is when $B = \lambda A$ for some real or complex number λ.

Matrix Exponential Functions

Let A be an $n \times n$ matrix. The definition of matrix exponential allows us to further define

$$\Phi(t) = e^{At} = \sum_{k=0}^{\infty} \frac{(At)^k}{k!}$$

which maps a real number $t \in \mathbb{R}$ to a matrix $e^{At} \in \mathbb{C}^{n \times n}$. We should be aware that each entry of $\Phi(t)$ is thus a function depending on t. The derivative of $\Phi(t)$ is then defined as the matrix with each entry being the corresponding derivative of the entry of $\Phi(t)$.

Now the key result is that $\dot{\mathbf{x}} = A\mathbf{x}$, with initial condition $\mathbf{x}(0) = \mathbf{x}_0$ has solution $\mathbf{x}(t) = \Phi(t)\mathbf{x}_0 = e^{At}\mathbf{x}_0$. This can be verified as follows:

We first note that $\mathbf{x}(0) = e^{A\cdot 0}\mathbf{x}_0 = \mathbf{x}_0$. Direct calculations give that

$$\frac{d}{dt}\frac{(At)^k}{k!} = \frac{A^k}{k!}\frac{d}{dt}t^k = \frac{A^k t^{k-1}}{(k-1)!} = A\cdot\frac{(At)^{k-1}}{(k-1)!}.$$

Thus, differentiating each term in $\sum_{k=0}^{\infty}\frac{(At)^k}{k!}$ and adding them up leads to

$$\frac{d}{dt}e^{At} = Ae^{At}. \tag{1}$$

It follows that

$$\dot{\mathbf{x}} = \frac{d}{dt}e^{At}\mathbf{x}_0 = A\big(e^{At}\mathbf{x}_0\big) = A\mathbf{x}(t),$$

which establishes that $\mathbf{x}(t) = e^{At}\mathbf{x}_0$ is the solution.

Calculating a Matrix Exponential via Similarity Transformations

For some matrices A, the matrix exponential e^A can be easily computed by directly evaluating the corresponding series. We did this in the preceding examples. For more general situations, it is convenent to recall the concept of "similarity transformations". [1]

Consider two matrices A and B. Suppose that there exists an invertible matrix T such that

$$A = TBT^{-1},$$

which means A and B are "similar" to each other. Diect calculations then give that

$$A^k = \big(TBT^{-1}\big)\big(TBT^{-1}\big)\cdots\big(TBT^{-1}\big) = TB^kT^{-1}, \text{ for all } k = 0, 1, 2, \ldots.$$

This further implies that

$$e^A = Te^BT^{-1}.$$

This gives the idea that if we can find a matrix B that is similar to A and e^B is easy to compute, then we can use the above equation to calculate e^A.

If A is diagonalisable, then we we can form

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

where T contains the eigenvectors of A.

This leads to

$$e^A = Te^{\Lambda}T^{-1} = T\begin{bmatrix} e^{\lambda_1} & & & \\ & e^{\lambda_2} & & \\ & & \ddots & \\ & & & e^{\lambda_n}\end{bmatrix}T^{-1}.$$

[1] Other interesting methods exist. If you like, have a look at "Nineteen Dubious Ways to Compute the Exponential of a Matrix", https://epubs.siam.org/doi/10.1137/1020098

More generally, **any** matrix $A \in \mathbb{R}^{n\times n}$ can be transformed into block diagonal Jordan form $J \in \mathbb{C}^{n\times n}$ with $A = TJT^{-1}$ for some invertible matrix $T \in \mathbb{C}^{n\times n}$:

$$J = \begin{bmatrix} \Lambda & & & \\ & J_1 & & \\ & & \ddots & \\ & & & J_s \end{bmatrix}$$

In the above expression, Λ contains the eigenvalues of A associated to the linearly independent eigenvectors.

The Jordan blocks contain the remaining eigenvalues λ_i:

$$J_i = \begin{bmatrix} \lambda_i & 1 & & \\ & \lambda_i & 1 & \\ & & \ddots & 1 \\ & & & \lambda_i \end{bmatrix}, \quad i = 1, 2, \ldots, s.$$

The matrix T can be determined from $AT = TJ$. However, determining Jordan forms lies outside the scope of this UoS.

Jordan forms can be used to calculate matrix exponentials using the similarity idea:

$$e^{At} = e^{TJT^{-1}t} = Te^{Jt}T^{-1}$$

In fact, due to their block diagonal structure, we have

$$e^{Jt} = \begin{bmatrix} e^{\Lambda t} & & & \\ & e^{J_1 t} & & \\ & & \ddots & \\ & & & e^{J_s t} \end{bmatrix}, \qquad \text{where}$$

$$e^{\Lambda t} = \begin{bmatrix} e^{\lambda_1 t} & & & \\ & e^{\lambda_2 t} & & \\ & & \ddots & \\ & & & e^{\lambda_m t} \end{bmatrix}, \quad e^{J_i t} = e^{\lambda_i t} \begin{bmatrix} 1 & t & \cdots & \frac{t^{m-1}}{(m-1)!} \\ 0 & 1 & t & \vdots \\ \vdots & \ddots & \ddots & t \\ 0 & \cdots & 0 & 1 \end{bmatrix} \tag{2}$$

Interestingly, in many situations (for example, when examining e^{At} when $t \to \infty$), the matrix e^{Jt} itself already states a lot about e^{At}, since T is a constant and invertible matrix. We will return to this in Section 2.

1 Higher-Order Linear Differential Equations

In past lectures we have focused on first and second-order differential equations. Here we consider differential equations of arbitrary order $n \geq 1$. The general form of such equations is

$$\frac{d^n y}{dt^n} + a_1 \frac{d^{n-1} y}{dt^{n-1}} + \cdots + a_n y = b_1 \frac{d^m u}{dt^m} + b_2 \frac{d^{m-1} u}{dt^{m-1}} + \cdots + b_{m+1} u \tag{3}$$

where y is the system output and u is the system input. Here $m \leq n$ and n is called the order of the system. We note that (linearity)

- *If $u_1(t), y_1(t)$ and $u_2(t), y_2(t)$ both satisfy the equation (3) so does $u(t) = \alpha u_1(t) + \beta u_2(t), y(t) = \alpha y_1(t) + \beta y_2(t)$.*

We can study the behaviour of this system following a similar procedure as we have treated second order systems.

Higher order systems are frequently encountered. Even if the underlying system is second-order, we may need to consider higher order systems because of dynamics in the controller.

Example 3. We have already seen PI control and PD control. The natural combination is PID control:

$$u(t) = K_P(r(t) - y(t)) + K_D(\dot{r}(t) - \dot{y}(t)) + K_I \int_0^t (r(\tau) - y(\tau))d\tau.$$

When applied to a second order system of the form

$$\ddot{y}(t) + a_1 \dot{y}(t) + a_0 y(t) = u(t)$$

we can analyse the resultig dynamics as follows:

Differentiating both of the above equations provides:

$$\dddot{y} + a_1 \ddot{y} + a_0 \dot{y} = \dot{u} = K_P(\dot{r} - \dot{y}) + K_D(\ddot{r} - \ddot{y}) + K_I(r - y)$$

and therefore the closed-loop system can be described via the third order linear ODE:

$$\dddot{y} + (a_1 + K_D)\ddot{y} + (a_0 + K_P)\dot{y} + K_I y = K_D \ddot{r} + K_P \dot{r} + K_I r.$$

We know that if this system is stable and $K_I \neq 0$, then it reaches an equilibrium with $K_I y = K_I r$, i.e. $y = r$, thus, there is no steady-state error.

How can we guarantee that it is stable? What does the transient response look like?

1.1 Homogeneous Solutions

Let $u = 0$ and then do a trial solution $y(t) = e^{st}$ to the equation (3). Noting $\frac{d^k}{dt^k} e^{st} = s^k e^{st}$, we obtain

$$\left(s^n + a_1 s^{m-1} + \cdots + a_n\right) e^{st} = 0.$$

This gives us the following polynomial equation

$$s^n + a_1 s^{m-1} + \cdots + a_n = 0 \tag{4}$$

as the characteristic equation of (3). In general, degree n polynomials have n solutions (roots) over the complex numbers[2], i.e. we have (solutions) $\lambda_1, \dots, \lambda_n \in \mathbb{C}$ to the equation (4).

$$s^n + a_1 s^{m-1} + \cdots + a_n = (s-\lambda_1)\cdot(s-\lambda_2)\dots(s-\lambda_n).$$

We have the following important points:

(i) A complex number λ can appear more than once in the $\lambda_1, \dots, \lambda_n$, and in that case it is termed a repeated root and its multiplicity is understood as how many times it has appeared within $\lambda_1, \dots, \lambda_n$.

(ii) If λ is a root, so is the complex conjugate $\bar{\lambda}$. Therefore, the complex roots, i.e., roots with nonzero imaginary parts, always appear in conjugate pairs. This follows from the fact that coefficients a_i are real numbers just by checking the expansion.

Similar to the second-order case, when the $\lambda_1, \dots, \lambda_n \in \mathbb{C}$ are all distinct, any homogeneous solution $y(t)$ of the differential equation (3) can be written as

$$y(t) = \sum_{k=1}^{n} C_k e^{\lambda_k t}$$

where $C_1, \dots, C_n$ are real numbers. With repeated roots, things become more complicated but it is possible to show that the homogeneous solutions can be represented as

$$y(t) = \sum_{k=1}^{n} C_k(t) e^{\lambda_k t}$$

where $C_k(t)$ are polynomials with degree less than the multiplicity of λ_k.

1.2 Solutions with Exponential Inputs

We next consider an exponential input function $u(t) = Ue^{st}$. We can verify directly

$$y_*(t) = \frac{b_1 s^m + b_2 s^{m-1} + \cdots + b_{m+1}}{s^n + a_1 s^{n-1} + \cdots + a_n} Ue^{st} = G(s)Ue^{st}$$

is a solution of (3). Here

$$G(s) = \frac{b_1 s^m + b_2 s^{m-1} + \cdots + b_{m+1}}{s^n + a_1 s^{n-1} + \cdots + a_n}$$

is the transfer function of the system (3). The $\lambda_1, \dots, \lambda_n$ satisfy $G(\lambda_k) = \infty$ for $k = 1, \dots, n$ and they are called *poles*. Solving the equation

$$b_1 s^m + b_2 s^{m-1} + \cdots + b_{m+1} = 0$$

yields m roots $z_1, \dots, z_m \in \mathbb{C}$. Each z_k satisfies $G(z_k) = 0$ and they are called *zeros*.

Now, the key insight is that the function

$$y(t) = \sum_{k=1}^{n} C_k(t) e^{\lambda_k t} + G(s)Ue^{st} \tag{5}$$

[2]This is the "fundamental theorem of algebra".

which is a combination of a homogeneous solution $\sum_{k=1}^{n} C_k(t)e^{\lambda_k t}$ and the particular solution $y_*(t) = G(s)Ue^{st}$ under input $u(t) = Ue^{st}$, remains a solution to (3) with input signal $u(t) = Ue^{st}$! We know from the theory of ordinary differential equations that any solution with input signal $u(t) = Ue^{st}$ for the equation (3) can be written in the form of (5).

1.3 Stability

As usual, the system (3) is said to be stable if all its homogeneous solutions converge to zero as time tends to infinity. We have learned that any homogeneous solution must be like

$$y(t) = \sum_{k=1}^{n} C_k(t)e^{\lambda_k t}$$

where $C_k(t)$ are constants or polynomials with finite degrees. From the Taylor series representation of e^{-at} it follows that exponentials will ultimately shrink (or grow) faster than any finite power of t:

$$\lim_{t\to+\infty} t^k e^{-at} = 0$$

for all integer k when $a > 0$. As a result, the following holds.

Theorem 1 *The system (3) is stable if and only if all its poles have negative real parts.*

1.4 Steady-State Response

Let the system (3) be stable. Then the first part of the solution (5) will vanish as time grows, leaving us

$$y_{ss}(t) = G(s)Ue^{st} \tag{6}$$

be the steady-state response for the system (3) under exponential input $u(t) = Ue^{st}$.

1.5 Frequency Response

Let the input $u(t)$ be a sinusoid $u(t) = A\cos(\omega t)$. Denote $\varphi = \angle G(j\omega)$. The using the linearity of the system and from (6) we can similarly derive

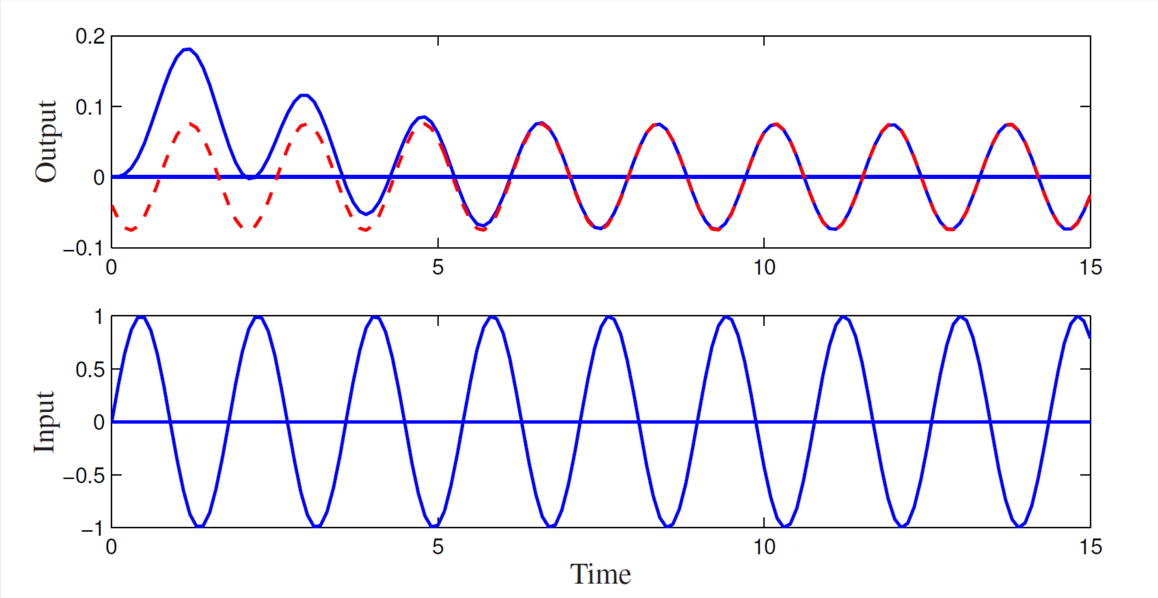
$$y_{ss}(t) = |G(j\omega)|A\cos(\omega t + \varphi).$$

1.6 Vector Representation of an ODE

Consider the differential equation

$$\frac{d^n y}{dt^n} + a_1\frac{d^{n-1}y}{dt^{n-1}} + \cdots + a_n y = u \tag{7}$$

Figura 1: An example of frequency response.



If we introduce the following notation

$$x_1(t) = y(t), \quad x_2(t) = \dot{y}(t), \quad \dots, \quad x_n(t) = y^{(n-1)}(t)$$

and then the vector

$$\mathbf{x}(t) = \begin{bmatrix} x_1(t) \\ \vdots \\ x_n(t) \end{bmatrix}.$$

Then we know

$$\begin{aligned}
\dot{x}_1(t) &= x_2(t) \\
\dot{x}_2(t) &= x_3(t) \\
&\dots \\
\dot{x}_{n-1}(t) &= x_n(t) \\
\dot{x}_n(t) &= -a_n x_1(t) - a_{n-1} x_2(t) - \cdots - a_1 x_n(t) + u(t)
\end{aligned}$$

and $y(t) = x_1(t)$.

Therefore, using vector derivatives[3] the equation (7) becomes

$$\begin{aligned}
\dot{\mathbf{x}}(t) &= A\mathbf{x}(t) + Bu(t) \\
y(t) &= C\mathbf{x}(t)
\end{aligned}$$

[3]The time-derivative of a vector $\mathbf{x}(t)$ is just the vector which has as elements the derivatives of the elements: $\dot{x}_i(t) = \frac{d}{dt} x_i(t)$.

where

$$A = \begin{pmatrix} 0 & 1 & & & \\ & 0 & 1 & & \\ & & \ddots & \ddots & \\ & & & 0 & 1 \\ -a_n & -a_{n-1} & -a_{n-1} & \dots & -a_1 \end{pmatrix}, \quad B = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}, C = \begin{bmatrix} 1 & 0 & \dots & 0 \end{bmatrix}.$$

2 State-space Model

The state-space model for linear dynamical system is

$$\dot{x} = Ax + Bu \tag{8}$$

$$y = Cx + Du \tag{9}$$

where $x \in \mathbb{R}^n$ is the *state*, $u \in \mathbb{R}^p$ is the *input*, $y \in \mathbb{R}^q$ is the *output*. Then $A \in \mathbb{R}^{n \times n}$, $B \in \mathbb{R}^{n \times p}$, $C \in \mathbb{R}^{q \times n}$, and $D \in \mathbb{R}^{q \times p}$. The system is called *single-input single-output* (SISO) if both u and y are scalars. For SISO systems B is just a column vector while C is a row vector both with n dimensions. Note that in this description the equation (8) is a differential equation describing the dynamics of the internal state of the system x, and (9) is a static equation describing how the output relies on the system state.

Example 4. A unit point mass is moving along a straight line with its position denoted by q. An external force $u \in \mathbb{R}$ is applied and there is no friction. Then the equation of motion is

$$\ddot{q} = u.$$

The output of interest is the position $y(t)$ of this point. We write $x = (x_1 \ x_2)^\top$ with $x_1 = q$ and $x_2 = \dot{q}$. The state-space description becomes

$$\frac{d}{dt}\begin{pmatrix} x_1(t) \\ x_2(t) \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} x_1(t) \\ x_2(t) \end{pmatrix} + \begin{pmatrix} 0 \\ 1 \end{pmatrix} u(t)$$

$$y(t) = \begin{pmatrix} 1 & 0 \end{pmatrix}\begin{pmatrix} x_1(t) \\ x_2(t) \end{pmatrix}.$$

Example 5. State space equations can come from other sources than just single-input single-output ODEs. Consider a simple example of heating/cooling of a 3-room building, where we simplify modelling by assuming that the temperature inside each room the same everywhere, but differs from room to room.

$$\dot{T}_1(t) = r_{12}(T_2(t) - T_1(t)) + r_{13}(T_3(t) - T_1(t)) \tag{10}$$

$$\dot{T}_2(t) = r_{21}(T_1(t) - T_2(t)) + r_{23}(T_3(t) - T_2(t)) + u_a(t) \tag{11}$$

$$\dot{T}_3(t) = r_{31}(T_3(t) - T_1(t)) + r_{32}(T_2(t) - T_3(t)) \tag{12}$$

where r_{ij} is the rate of heat flow from room j to room i, which can be derived from Newton's law of cooling, and $u_a(t)$ is a control input from a reverse-cycle air-conditioner, that can heat or cool, in room 2. We can choose the state vector to be

$$\mathbf{x}(t) = \begin{bmatrix} T_1(t) \\ T_2(t) \\ T_3(t) \end{bmatrix}$$

and write the system in the form (8) with

$$A = \begin{bmatrix} -r_{12} - r_{13} & r_{12} & r_{13} \\ r_{21} & -r_{21} - r_{23} & r_{23} \\ r_{31} & r_{32} & -r_{31} - r_{32} \end{bmatrix}, B = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix},$$

and if there were temperature sensors in rooms 1 and 2, then we could have a vector measurement of the form (9) with

$$C = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}, D = 0.$$

This model could also be expanded to include *disturbance* inputs due to, e.g., incident sunlight heating the building, or heat diffusion to ambient external temperature.

2.1 Homogeneous Solutions

The homogeneous solution of the equation (8) is found via the corresponding homogeneous equation

$$\dot{x} = Ax. \tag{13}$$

Let $x(0) = x_0 \in \mathbb{R}^n$ be the initial value of the equation (13). Consider

$$x(t) = e^{At}x_0.$$

Then (i) $x(0) = e^0 x_0 = x_0$; (ii) $\dot{x}(t) = \frac{d}{dt}e^{At}x_0 = Ae^{At}x_0 = Ax(t)$ from the equation (1). This is to say, $x(t) = e^{At}x_0$ is the unique solution of the equation (13) with $x(0) = x_0 \in \mathbb{R}^n$.

2.2 Stability

The system (8) is called to be stable if all its homogeneous solutions converge to zero. Particularly, if the system (8) is stable then

$$\lim_{t\to\infty} e^{At}v_m = 0$$

where v_m is the vector with the m'th entry being 1 and all other entries being 0. Thus, stability of the system (8) is equivalent to

$$\lim_{t\to\infty} e^{At} = 0$$

Let J be the Jordan form of the matrix A with $A = TJT^{-1}$. Then

$$e^{At} = Te^{Jt}T^{-1}.$$

Therefore, $\lim_{t\to\infty} e^{At} = 0$ is equivalent to $\lim_{t\to\infty} e^{Jt} = 0$. The expression of e^{Jt} as shown in (2) implies the following theorem.

Theorem 2 *The system (2) is stable if and only if all eigenvalues of A have negative real parts.*

Example 4. Let us consider the example in Example 3 again. Given initial value $x(0) = (x_1(0) \ \ x_2(0))^\top$, the homogeneous solution of the system is

$$\begin{pmatrix} x_1(t) \\ x_2(t) \end{pmatrix} = \begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x_1(0) \\ x_2(0) \end{pmatrix} = \begin{pmatrix} x_1(0) + tx_2(0) \\ x_2(0) \end{pmatrix}$$

The corresponding output is $y(t) = x_1(0) + tx_2(0)$. Moreover, the system is not stable since the eigenvalues of the matrix A are both zero.

2.3 Response to Exponential Input

Suppose $u = Ue^{st}$ for some (possibly complex) scalar U, and once again, guess that $y(t) = G(s)u(t) = G(s)Ue^{st}$ for some complex number $G(s)$. Furthermore, we will guess that the vector function $x(t)$ can be written as $X(s)Ue^{st}$ for some vector function $X : \mathbb{C} \to \mathbb{R}^n$.

Then

$$\dot{x}(t) = \frac{d}{dt}(X(s)Ue^{st}) = X(s)U\frac{d}{dt}(e^{st}) = sX(s)Ue^{st}$$

so from $\dot{x} = Ax + Bu$ we have

$$\begin{aligned} sX(s)Ue^{st} &= AX(s)Ue^{st} + BUe^{st} \\ sX(s) &= AX(s) + B \end{aligned} \tag{14}$$

where the second line follows from the fact that $e^{st} \neq 0$ for any s or t, and the fact that the first line must hold for all scalars U, so Ue^{st} factors can be eliminated. So we have

$$\begin{aligned} (sI - A)X(s) &= B \\ X(s) &= (sI - A)^{-1}B \end{aligned}$$

where I is the identity matrix.

Now, for the output $y(t)$ we have $y = Cx + Du$ and by assumption $y(t) = G(s)Ue^{st}$. So making substitutions for x and u:

$$\begin{aligned} G(s)Ue^{st} &= CX(s)Ue^{st} + DUe^{st} \\ G(s) &= CX(s) + D \end{aligned} \tag{15}$$

where once again, we can eliminate Ue^{st} factors.

Hence, combining (14) and (15) we have

$$G(s) = C(sI - A)^{-1}B + D \tag{16}$$

is the formula for the transfer function of the system (8), (9). For SISO systems, this is just a complex scalar function of a complex scalar s. If there are $p > 1$ inputs or $q > 1$ outputs, $G(s)$ is a $p \times q$ matrix of functions of s, sometimes called a *transfer matrix*

Recall that the poles of a transfer function $G(s)$ are the values of s for which $G(s) = \infty$, i.e. when the denominator polynomial goes to zero. Recall also that the *eigenvalues* of A are the values of s for which $\det(sI - A) = 0$. From the formula for matrix inverse:

$$(sI - A)^{-1} = \frac{\mathrm{Adj}(sI - A)}{\det(sI - A)}$$

where Adj is the adjugate matrix, which will be a matrix of polynomials in s. Now it can be seen that the poles of $G(s)$ are exactly the eigenvalues of A.

3 Coordinate Invariance

The descriptions of system state and output are certainly not unique – just see how we record temperatures in Australia or in US or by physicists. Now suppose a dynamical system has been described in state space by (8)-(9) under the state variables x. Let $T \in \mathbb{R}^{n\times n}$ be an invertible matrix and introduce $z = Tx$, and since T is constant $\dot{z} = T\dot{x}$. In this way, z is an *alternative* representation of the system state, and we have

$$\frac{dz}{dt} = T(Ax + Bu) = TAT^{-1}z + TBu \tag{17}$$

$$y = Cx + Du = CT^{-1}z + Du. \tag{18}$$

Denote $\tilde{A} = TAT^{-1}$, $\tilde{B} = TB$, $\tilde{C} = CT^{-1}$. The state-space model of the system can be re-written as

$$\frac{dz}{dt} = \tilde{A}z + \tilde{B}u \tag{19}$$

$$y = \tilde{C}z + Du. \tag{20}$$

Note that (19)-(20) describe the *same* input and output as the equations (8)-(9), where the change of representation of state variable has altered the state transition matrices in the description. We also note that A and $\tilde{A}$ are similar to each other, and therefore they have the same eigenvalues. This means a stable system is always stable no matter how we describe the state variables.

Furthermore, we also have exactly the same transfer function describing input-output relationships:

$$\begin{aligned}\tilde{C}(sI - \tilde{A})^{-1}\tilde{B} + D &= CT^{-1}(sI - TAT^{-1})^{-1}TB + D \\ &= C(T^{-1}Ts - T^{-1}TAT^{-1}T)^{-1}B + D \\ &= C(sI - A)^{-1}B + D = G(s)\end{aligned}$$

where going from the first line to the second, we have used the fact that for a sequence of inverses $P^{-1}Q^{-1}R^{-1} = (RQP)^{-1}$ and for the last line have used the fact that $T^{-1}T = I$.

4 General Solutions (advanced)

Given initial value $x(0) = x_0$ and input signal $u(t)$, the general solution of the equation (8)-(9) is

$$x(t) = e^{At}x_0 + \int_{\tau=0}^{t} e^{A(t-\tau)}Bu(\tau)d\tau, \tag{21}$$

$$y(t) = Ce^{At}x_0 + C\int_{\tau=0}^{t} e^{A(t-\tau)}Bu(\tau)d\tau + Du(t). \tag{22}$$

Perhaps the easiest way of understanding these formulas is to differentiate the given $x(t)$ in (21) directly and verify it is indeed a solution of (8).

First of all, it is straightforward to verify that

$$x(0) = e^0 x_0 + \int_{\tau=0}^{0} e^{A(t-\tau)} Bu(\tau) d\tau = x_0.$$

Next, we would like to compute the derivative of the $x(t)$. To this end we write

$$f_1(t) = e^{At}, \quad f_2(t) = \int_{\tau=0}^{t} e^{-A\tau} Bu(\tau) d\tau.$$

Thus (21) becomes

$$x(t) = e^{At} x_0 + \int_{\tau=0}^{t} e^{A(t-\tau)} Bu(\tau) d\tau = e^{At} x_0 + f_1(t) f_2(t).$$

A Calculus Trick

It looks scary but it is actually easy to calculate the functions in the form of

$$q(t) = \int_{t_0}^{t} p(\tau) d\tau$$

where the argument appears in the range of the integral. We simply have

$$\frac{d}{dt} q(t) = p(t).$$

As a result, for $f_2(t) = \int_{\tau=0}^{t} e^{-A\tau} Bu(\tau) d\tau$ there holds

$$\frac{d}{dt} f_2(t) = e^{-At} Bu(t)$$

noting the structure of $f_2(t)$.

We are now ready to compute

$$\begin{aligned}
&\frac{d}{dt}\Big(e^{At} x_0 + \int_{\tau=0}^{t} e^{A(t-\tau)} Bu(\tau) d\tau \Big) \\
&= Ae^{At} x_0 + f_1(t) \dot{f}_2(t) + \dot{f}_1(t) f_2(t) \\
&= Ae^{At} x_0 + e^{At}\Big(e^{-At} Bu(t) \Big) + Ae^{At} \int_{\tau=0}^{t} e^{-A\tau} Bu(\tau) d\tau \\
&= Ae^{At} x_0 + A \int_{\tau=0}^{t} e^{A(t-\tau)} Bu(\tau) d\tau + Bu(t) \\
&= A\Big(e^{At} x_0 + \int_{\tau=0}^{t} e^{A(t-\tau)} Bu(\tau) d\tau \Big) + Bu(t).
\end{aligned} \tag{23}$$

Therefore, (21)-(22) are the solution of equation (8)-(9) with the initial value $x(0) = x_0$ and the input signal $u(t)$.

Appendix. Linearization

Let $u \in \mathbb{R}^p$, $x \in \mathbb{R}^n$, and $y \in \mathbb{R}^q$. A nonlinear state-space model is described by

$$\frac{dx}{dt} = f(x,u) \tag{24}$$
$$y = h(x,u) \tag{25}$$

where $f : \mathbb{R}^{n+p} \mapsto \mathbb{R}^n$ and $g : \mathbb{R}^{n+p} \mapsto \mathbb{R}^q$ are two nonlinear functions. Recall that a point $(x_e, u_e) \in \mathbb{R}^n \times \mathbb{R}^p$ is an equilibrium of (24) if

$$f(x_e, u_e) = 0.$$

Multi-variable Functions and Jacobian

Let $g(x)$ be a function mapping $x = (x_1 \ \dots \ x_n)^\top \in \mathbb{R}^n$ to $\mathbb{R}^n$. This is to say, $g(x)$ has the form

$$g(x) = \begin{pmatrix} g_1(x_1, \dots, x_n) \\ g_2(x_1, \dots, x_n) \\ \vdots \\ g_n(x_1, \dots, x_n) \end{pmatrix}$$

where each $g_k(x_1, \dots, x_n)$ maps $\mathbb{R}^n$ to $\mathbb{R}$. The Jacobian of g at the point $x_* \in \mathbb{R}^n$, denoted $J(x_*) = \frac{\partial g}{\partial x}\Big|_{x=x_*}$, is given by

$$J(x_*) = \begin{pmatrix} \frac{\partial g_1(x)}{\partial x_1} & \frac{\partial g_1(x)}{\partial x_2} & \dots & \frac{\partial g_1(x)}{\partial x_n} \\ \frac{\partial g_2(x)}{\partial x_1} & \frac{\partial g_2(x)}{\partial x_2} & \dots & \frac{\partial g_2(x)}{\partial x_n} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial g_n(x)}{\partial x_1} & \frac{\partial g_n(x)}{\partial x_2} & \dots & \frac{\partial g_n(x)}{\partial x_n} \end{pmatrix}_{x=x_*}$$

Introduce a set of new variables:

$$\delta_x = x - x_e, \quad \delta_u = u - u_e, \quad \delta_y = y - h(x_e, u_e).$$

Then around the equilibrium (x_e, u_e) the system (24)-(25) can be approximated by

$$\frac{d}{dt}\delta_x = A\delta_x + B\delta_u \tag{26}$$
$$\delta_y = C\delta_x + D\delta_u \tag{27}$$

where

$$A = \frac{\partial f}{\partial x}\Big|_{(x_e,u_e)}, \ B = \frac{\partial f}{\partial u}\Big|_{(x_e,u_e)}, \ C = \frac{\partial h}{\partial x}\Big|_{(x_e,u_e)}, \ D = \frac{\partial h}{\partial u}\Big|_{(x_e,u_e)}.$$

Note that here in writing $\frac{\partial f}{\partial x}$, we have considered the variable u in $f(x,u)$ as fixed and then treat f as a function of x. The same methodology applies to $\frac{\partial f}{\partial u}$, $\frac{\partial h}{\partial x}$, and $\frac{\partial h}{\partial u}$.

Linear Approximation of Nonlinear Functions (advanced)

Let $g(x)$ be a continuously differentiable function mapping $\mathbb{R}$ to $\mathbb{R}$. Then we know

$$g(x) \approx g(x_*) + \frac{\partial g}{\partial x}(x_*)(x - x_*)$$

for x in a neighborhood of x_*. Here the approximation is understood in the sense that the difference between the two functions on the left and right sides

$$e(x) = g(x) - g(x_*) - \frac{\partial g}{\partial x}(x_*)(x - x_*)$$

satisfies that

$$\lim_{x \to x_*} \frac{|e(x)|}{|x - x_*|} = 0.$$

Now let $g(x)$ be a multi-variable function mapping $x = (x_1 \ \ldots \ x_n)^\top \in \mathbb{R}^n$ to $\mathbb{R}^n$. For $x \in \mathbb{R}^n$ we use $\|x\| = \sqrt{x_1^2 + \cdots + x_n^2}$ to quantify the *size* of the vector x, known as the Euclidean norm. Then for the function g there holds

$$g(x) \approx g(x_*) + J(x_*)(x - x_*)$$

for $x \in \mathbb{R}^n$ in a neighborhood of x_* in the sense that

$$\lim_{x \to x_*} \frac{\|g(x) - g(x_*) + J(x_*)(x - x_*)\|}{\|x - x_*\|} = 0.$$

This can be viewed as the reasoning of Jacobian approximation for nonlinear dynamical systems.

Examples. Example 6.12 (Chapter 6) and Example 7.5 (Chapter 7), Åström and Murray, *Feedback Systems*, Second Edition.