Lecture 3: Continuous-Time State-Space Models of Dynamic Systems

In this course we focus on dynamic systems described by systems of ordinary differential equations (ODEs). The main categories of dynamic systems are linear time-varying (LTV), linear time-invariant (LTI), and nonlinear. Moreover, these systems can be described in both continuous time (using ODEs) or in discrete time (using *difference equation* which are analogous to ODEs). If the system includes uncertainty (e.g., random noise) then it is considered *stochastic*, otherwise it is called *deterministic*. In this lecture we review continuous-time, deterministic systems.

Linear Time-Varying (LTV) Continuous-Time Systems

The basic representation for linear systems is the *linear state equation* (1) and the *output or measurement equation* (2) often written as:

$$\dot{\boldsymbol{x}}(t) = \boldsymbol{A}(t)\boldsymbol{x}(t) + \boldsymbol{B}(t)\boldsymbol{u}(t) \tag{1}$$

$$\boldsymbol{y}(t) = \boldsymbol{C}(t)\boldsymbol{x}(t) + \boldsymbol{D}(t)\boldsymbol{u}(t) \tag{2}$$

$$\boldsymbol{x}(t_0) = \boldsymbol{x}_0 \tag{3}$$

where

- $\boldsymbol{x}(t) \in \mathbb{R}^n$ is a time-varying *state vector* describing the states of the system that are of interest (e.g., positions and velocities of a mechanical system). Its derivative $\dot{\boldsymbol{x}}(t) \in \mathbb{R}^n$ is called the *state-rate*.
- $\boldsymbol{x}_0 \in \mathbb{R}^n$ is the initial state at the initial time $t_0 \in \mathbb{R}$ (often t_0 is taken as zero) in (3)
- $\boldsymbol{u}(t) \in \mathbb{R}^m$ is a time-varying *control input vector* that aims to achieve some control objective (e.g., to stabilize the system or drive it to a desired terminal state)
- $\boldsymbol{y}(t) \in \mathbb{R}^p$ is a time-varying *output vector* that describes a set of measurements or outputs that can be observed regarding the state
- $\boldsymbol{A}(t) \in \mathbb{R}^{n\times n}$ is the *system matrix* or *plant* that describes the natural dynamics of the system in the absence of any control inputs/disturbances
- $\boldsymbol{B}(t) \in \mathbb{R}^{n\times m}$ is the *control input matrix* or *control effect matrix* and describes how the control vector influences the state-rate
- $\boldsymbol{C}(t) \in \mathbb{R}^{p\times n}$ is the *observation matrix* that maps the state vector to an observation vector describing what can be measured about the system (e.g., by sensors)
- $\boldsymbol{D}(t) \in \mathbb{R}^{p\times m}$ is the *control-measurement influence matrix* that maps the control vector to an observation vector that modifies the system output. Very frequently this term is ignored when the sensors are not affected by control inputs, $\boldsymbol{D}(t) = 0$.

Example. The following system

$$\dot{\boldsymbol{x}} = \boldsymbol{A}(t)\boldsymbol{x} + \boldsymbol{B}(t)u(t) \tag{4}$$

$$\begin{bmatrix} \dot{x}_1(t) \\ \dot{x}_2(t) \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ t & 0 \end{bmatrix} \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \end{bmatrix} u(t) \tag{5}$$

with output equation

$$y(t) = \boldsymbol{C}(t)\boldsymbol{x}(t) \tag{6}$$

$$= \begin{bmatrix} 1 & 1 \end{bmatrix} \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} \tag{7}$$

$$= x_1(t) + x_2(t) \tag{8}$$

is a linear time-varying system. The state of the system is $\boldsymbol{x}(t) = [x_1, x_2]^{\mathrm{T}}$ and the control input is a scalar $u(t)$. Notice $\boldsymbol{D}(t)$ is zero. The output is also a scalar equal to the sum of the two states. No sources of uncertainty/noise are present in this time of system model and it is deterministic (rather than stochastic/noisy).

Peano-Baker Series. For the moment, suppose there is no control input in (1), i.e., $\boldsymbol{u}(t) = 0$ and

$$\dot{\boldsymbol{x}}(t) = \frac{d\boldsymbol{x}(t)}{dt} = \boldsymbol{A}(t)\boldsymbol{x}(t) \tag{9}$$

Consider integrating this simplified system with respect to time, starting from the initial condition $\boldsymbol{x}(t_0) = \boldsymbol{x}_0$. A state time-history or *trajectory* $\boldsymbol{x}^*(t)$ with $\boldsymbol{x}^*(t_0) = \boldsymbol{x}_0$ is a valid solution of (1) if it satisfies:

$$d\boldsymbol{x}^*(t) = \boldsymbol{A}(t)\boldsymbol{x}^*(t)dt \tag{10}$$

$$\int_{\boldsymbol{x}_0}^{\boldsymbol{x}^*(t)} d\boldsymbol{x}^* = \int_{t_0}^{t} \boldsymbol{A}(\sigma)\boldsymbol{x}^*(\sigma)d\sigma \tag{11}$$

$$\boldsymbol{x}^*(t) = \boldsymbol{x}_0 + \int_{t_0}^{t} \boldsymbol{A}(\sigma)\boldsymbol{x}^*(\sigma)d\sigma \tag{12}$$

Note the temporary switch from time t to σ as a dummy variable in the integration. Evaluating the integral on the RHS requires prior knowledge of $\boldsymbol{x}^*(t)$, hence it can only be used to check a solution (but not to construct one). To construct a solution that satisfies the initial conditions consider a series of approximations indexed by $k = 0, 1, \ldots, \infty$. The first approximation is a constant equal to the initial condition

$$\boldsymbol{x}_0(t) = \boldsymbol{x}_0 \tag{13}$$

Now, in the next approximation for $\boldsymbol{x}_1$ we use the previous solution to evaluate the integral (12) and give an even better approximation:

$$\boldsymbol{x}_1(t) = \boldsymbol{x}_0 + \int_{t_0}^{t} \boldsymbol{A}(\sigma)\boldsymbol{x}_0 d\sigma \tag{14}$$

The process repeats and in the next step $k = 2$ the previous approximation, $\boldsymbol{x}_1(t)$, is inserted into the integral on the RHS

$$\boldsymbol{x}_2(t) = \boldsymbol{x}_0 + \int_{t_0}^{t} \boldsymbol{A}(\sigma)\boldsymbol{x}_1(\sigma)d\sigma \tag{15}$$

to give a new approximation $\boldsymbol{x}_2(t)$. By substitution, we could also write this as

$$\boldsymbol{x}_2(t) = \boldsymbol{x}_0 + \int_{t_0}^{t} \boldsymbol{A}(\sigma_1)\left(\boldsymbol{x}_0 + \int_{t_0}^{\sigma_1} \boldsymbol{A}(\sigma_2)\boldsymbol{x}_0 d\sigma_2\right)d\sigma_1 \tag{16}$$

$$= \boldsymbol{x}_0 + \int_{t_0}^{t} \boldsymbol{A}(\sigma_1)\boldsymbol{x}_0 d\sigma_1 + \int_{t_0}^{t} \boldsymbol{A}(\sigma_1)\int_{t_0}^{\sigma_1} \boldsymbol{A}(\sigma_2)\boldsymbol{x}_0 d\sigma_2 d\sigma_1 \tag{17}$$

$$= \left[\boldsymbol{I} + \int_{t_0}^{t} \boldsymbol{A}(\sigma_1)d\sigma_1 + \int_{t_0}^{t} \boldsymbol{A}(\sigma_1)\int_{t_0}^{\sigma_1} \boldsymbol{A}(\sigma_2)d\sigma_2 d\sigma_1\right]\boldsymbol{x}_0 \tag{18}$$

Continuing this process of approximation and substitution onward for arbitrary k leads to

$$\boldsymbol{x}_k(t) = \left[\boldsymbol{I} + \int_{t_0}^{t} \boldsymbol{A}(\sigma_1)d\sigma_1 + \int_{t_0}^{t} \boldsymbol{A}(\sigma_1)\int_{t_0}^{\sigma_1} \boldsymbol{A}(\sigma_2)d\sigma_2 d\sigma_1\right. \tag{19}$$

$$\left. + \cdots + \int_{t_0}^{t} \boldsymbol{A}(\sigma_1)\int_{t_0}^{\sigma_1} \cdots \int_{t_0}^{\sigma_{k-1}} \boldsymbol{A}(\sigma_{k-1})d\sigma_{k-1}\cdots d\sigma_2 d\sigma_1\right]\boldsymbol{x}_0 \tag{20}$$

from which the matrix in square-brackets defines the *Peano-Baker* series

$$\boldsymbol{\Phi}(t, t_0) = \left[\boldsymbol{I} + \int_{t_0}^{t} \boldsymbol{A}(\sigma_1)d\sigma_1 + \int_{t_0}^{t} \boldsymbol{A}(\sigma_1)\int_{t_0}^{\sigma_1} \boldsymbol{A}(\sigma_2)d\sigma_2 d\sigma_1\right. \tag{21}$$

$$\left. + \cdots + \int_{t_0}^{t} \boldsymbol{A}(\sigma_1)\int_{t_0}^{\sigma_1} \cdots \int_{t_0}^{\sigma_{k-1}} \boldsymbol{A}(\sigma_{k-1})d\sigma_{k-1}\cdots d\sigma_2 d\sigma_1\right] \tag{22}$$

It can be shown (outside of the scope of this course) that the Peano-Baker series converges to a unique solution. The matrix $\boldsymbol{\Phi}(t_0, t)$ is called the *state transition matrix* or sometimes the *fundamental matrix*. According to (20) it can be used to propagate a linear system from the initial state $\boldsymbol{x}_0$ to any future time

$$\boldsymbol{x}(t) = \boldsymbol{\Phi}(t, t_0)\boldsymbol{x}_0 \ .$$

In this case (remember we assumed no control input $\boldsymbol{u}(t) = 0$) the output equation is simply

$$\boldsymbol{y}(t) = \boldsymbol{C}(t)\boldsymbol{x}(t) \tag{23}$$

$$= \boldsymbol{C}(t)\boldsymbol{\Phi}(t, t_0)\boldsymbol{x}_0 \ . \tag{24}$$

Now if repeat the procedure allowing some non-zero control input $\boldsymbol{u}(t) \neq 0$ (this process is not shown) we obtain the state and output solutions:

$$\boldsymbol{x}(t) = \boldsymbol{\Phi}(t, t_0)\boldsymbol{x}_0 + \int_{t_0}^{t} \boldsymbol{\Phi}(t, \sigma)\boldsymbol{B}(\sigma)u(\sigma)d\sigma \tag{25}$$

$$\boldsymbol{y}(t) = \boldsymbol{C}(t)\boldsymbol{x}(t) \tag{26}$$

$$= \boldsymbol{C}(t)[\boldsymbol{\Phi}(t, t_0)\boldsymbol{x}_0 + \int_{t_0}^{t} \boldsymbol{\Phi}(t, \sigma)\boldsymbol{B}(\sigma)u(\sigma)d\sigma] \tag{27}$$

The first equation is known as the *variation of constants formula*. In general, these integrals are difficult to solve for arbitrary matrices $\boldsymbol{A}(t)$ and $\boldsymbol{B}(t)$. However, the situation greatly simplifies when the system is time invariant.

Example. Here we wish to compute the Peano-Baker series for the previous example and determine the solution of the system at $t = 2$ assuming the initial condition $\boldsymbol{x}(t_0) = [1, 3]^{\mathrm{T}}$ at $t_0 = 0$ and a constant control input $u(t) = u_c = -2$. The first term in the Peano-Baker series (22) is the identity matrix, the second term is

$$\int_{t_0}^{t} \boldsymbol{A}(\sigma_1) d\sigma_1 = \int_{t_0}^{t} \begin{bmatrix} 0 & 0 \\ \sigma_1 & 0 \end{bmatrix} d\sigma_1 \tag{28}$$

$$= \frac{1}{2} \begin{bmatrix} 0 & 0 \\ (t^2 - t_0^2) & 0 \end{bmatrix}, \tag{29}$$

and the third term is

$$\int_{t_0}^{t} \boldsymbol{A}(\sigma_1) \int_{t_0}^{\sigma_1} \boldsymbol{A}(\sigma_2) d\sigma_2 d\sigma_1 = \int_{t_0}^{t} \begin{bmatrix} 0 & 0 \\ \sigma_1 & 0 \end{bmatrix} \int_{t_0}^{\sigma_1} \begin{bmatrix} 0 & 0 \\ \sigma_2 & 0 \end{bmatrix} d\sigma_2 d\sigma_1 \tag{30}$$

$$= \frac{1}{2} \int_{t_0}^{t} \begin{bmatrix} 0 & 0 \\ \sigma_1 & 0 \end{bmatrix} \begin{bmatrix} 0 & 0 \\ \sigma_1^2 - t_0^2 & 0 \end{bmatrix} d\sigma_1 \tag{31}$$

$$= 0\,. \tag{32}$$

Since all later terms (the fourth term, fifth term, etc.) contain the above expression they are all zero and the series is exactly equal to the identity matrix plus (29)

$$\boldsymbol{\Phi}(t, t_0) = \boldsymbol{I} + \int_{t_0}^{t} \boldsymbol{A}(\sigma_1) d\sigma_1 \tag{33}$$

$$= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \frac{1}{2} \begin{bmatrix} 0 & 0 \\ t^2 - t_0^2 & 0 \end{bmatrix} \tag{34}$$

$$= \begin{bmatrix} 1 & 0 \\ (1/2)(t^2 - t_0^2) & 1 \end{bmatrix} \tag{35}$$

If we had no control input $u(t) = 0$ then from (25) with $t_0 = 0$ the system would evolve according to

$$\boldsymbol{x}(t) = \boldsymbol{\Phi}(t, t_0)\boldsymbol{x}_0 \tag{36}$$

$$\begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ (1/2)t^2 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 3 \end{bmatrix} = \begin{bmatrix} 1 \\ (1/2)t^2 + 3 \end{bmatrix} \tag{37}$$

However, in our case we have $u(t) = u_c \neq 0$ and the second term of (25) is required:

$$\int_{t_0}^{t} \boldsymbol{\Phi}(t, \sigma)\boldsymbol{B}(\sigma)u(\sigma) d\sigma = \int_{t_0}^{t} \begin{bmatrix} 1 & 0 \\ (1/2)(t^2 - \sigma^2) & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} u_c d\sigma \tag{38}$$

$$= u_c \int_{t_0}^{t} \begin{bmatrix} 1 \\ (1/2)(t^2 - \sigma^2) \end{bmatrix} d\sigma \tag{39}$$

$$= \begin{bmatrix} u_c(t - t_0) \\ (1/2)t^2(t - t_0) - (1/6)(t^3 - t_0^3) \end{bmatrix} \tag{40}$$

The solution to the system is a superposition of the initial condition response (with no input) and the zero-initial condition response (with input). With $t_0 = 0$ the solution is

$$\boldsymbol{x}(t) = \boldsymbol{\Phi}(t, t_0)\boldsymbol{x}_0 + \int_{t_0}^{t} \boldsymbol{\Phi}(t, \sigma)\boldsymbol{B}(\sigma)u(\sigma)d\sigma \tag{41}$$

$$\begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ (1/2)t^2 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 3 \end{bmatrix} + \begin{bmatrix} u_c t \\ u_c t^3/6 \end{bmatrix} \tag{42}$$

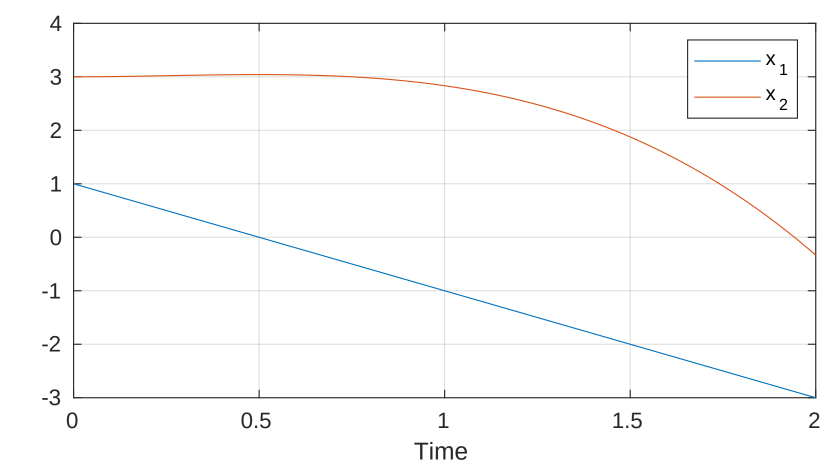
$$= \begin{bmatrix} 1 + u_c t \\ (1/2)t^2 + 3 + u_c[(1/2)t^3 - (1/6)6t^3] \end{bmatrix} \tag{43}$$

$$= \begin{bmatrix} 1 + u_c t \\ (1/2)t^2 + 3 + (4/3)u_c t^3 \end{bmatrix} \tag{44}$$

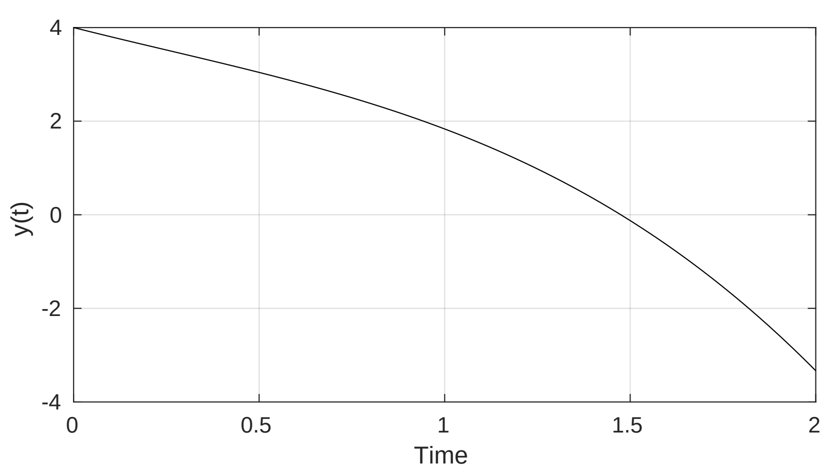
If $u_c = -2$, then after two seconds the state of the system is

$$\boldsymbol{x}(2) = \begin{bmatrix} -3 \\ -1/3 \end{bmatrix} \tag{45}$$

The solutions $x_1(t)$ and $x_2(t)$ are plotted below



and the corresponding output of the system is the sum of the two responses, $y(t)$:



Linear Time Invariant (LTI) Continuous-Time Systems

If the system is time-invariant then the matrices are all constants, i.e., $\boldsymbol{A}(t) = \boldsymbol{A}$, $\boldsymbol{B}(t) = \boldsymbol{B}$, and $\boldsymbol{C}(t) = \boldsymbol{C}$ and the LTI system is

$$\dot{\boldsymbol{x}}(t) = \boldsymbol{A}\boldsymbol{x}(t) + \boldsymbol{B}\boldsymbol{u}(t) \tag{46}$$
$$\boldsymbol{y}(t) = \boldsymbol{C}\boldsymbol{x}(t) \tag{47}$$
$$\boldsymbol{x}(t_0) = \boldsymbol{x}_0 \tag{48}$$

(again, we've ignored $\boldsymbol{D}(t)$ by setting it to zero). It is convenient for LTI systems to drop the time-dependence altogether and simply write:

$$\dot{\boldsymbol{x}} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{u} \tag{49}$$
$$\boldsymbol{y} = \boldsymbol{C}\boldsymbol{x} \tag{50}$$
$$\boldsymbol{x}(t_0) = \boldsymbol{x}_0 \tag{51}$$

For this special case the state transition matrix is given by the matrix exponential

$$\boldsymbol{\Phi}(t, t_0) = e^{\boldsymbol{A}(t-t_0)} = \sum_{k=0}^{\infty} \frac{1}{k!} \boldsymbol{A}^k (t-t_0)^k = \boldsymbol{I} + \boldsymbol{A}(t-t_0) + \frac{\boldsymbol{A}^2(t-t_0)^2}{2} + \frac{\boldsymbol{A}^3(t-t_0)^3}{3!} + \cdots \tag{52}$$

(As a reminder, the matrices are time-invariant and the term $\boldsymbol{A}(t-t_0)$ is a product of $\boldsymbol{A} \cdot (t-t_0)$ not a function of time.) Computing the matrix exponential is further simplified in certain cases:

- If the $\boldsymbol{A}$ matrix is diagonal, as in
$$\boldsymbol{A} = \begin{bmatrix} d_1 & 0 \\ 0 & d_2 \end{bmatrix},$$
then the matrix exponential is
$$\boldsymbol{\Phi}(t, t_0) = e^{\boldsymbol{A}(t-t_0)} = \begin{bmatrix} e^{d_1(t-t_0)} & 0 \\ 0 & e^{d_2(t-t_0)} \end{bmatrix}.$$
- If the matrix power $\boldsymbol{A}^k = 0$ for some integer k in the series (52) becomes zero then all remaining terms in (52) are also zero. This is true since any matrix power $\boldsymbol{A}^q$ with $q > k$ can always be factored as $\boldsymbol{A}^q = \boldsymbol{A}^k \boldsymbol{A}^{q-k} = \boldsymbol{0}\boldsymbol{A}^{q-k} = \boldsymbol{0}$.
- The infinite summations are recognized as familiar functions (e.g., trigonometric series).
- Other ways to compute the matrix exponential involves Laplace transforms, similarity transforms, and the uses the eigenvalues of $\boldsymbol{A}$ with Cayley-Hamilton theorem. We will discuss two of these methods shortly.

In the case that $\boldsymbol{u} = 0$ (i.e., there is no control input), the solution (25) to an initial condition $\boldsymbol{x}(t_0) = \boldsymbol{x}_0$ is

$$\boldsymbol{x}(t) = e^{\boldsymbol{A}(t-t_0)}\boldsymbol{x}_0 \tag{53}$$
$$\boldsymbol{y}(t) = \boldsymbol{C}\boldsymbol{x}(t) \tag{54}$$
$$= \boldsymbol{C}e^{\boldsymbol{A}(t-t_0)}\boldsymbol{x}_0 \tag{55}$$

In the case that $\boldsymbol{u} \neq 0$ (i.e., there is a control input), the solution (25) to an initial condition $\boldsymbol{x}(t_0) = \boldsymbol{x}_0$ is

$$\boldsymbol{x}(t) = e^{\boldsymbol{A}(t-t_0)}\boldsymbol{x}_0 + \int_{t_0}^{t} e^{\boldsymbol{A}(t-\sigma)}\boldsymbol{B}\boldsymbol{u}(\sigma)d\sigma \tag{56}$$

$$\boldsymbol{y}(t) = \boldsymbol{C}\boldsymbol{x}(t) \tag{57}$$

$$= \boldsymbol{C}[e^{\boldsymbol{A}(t-t_0)}\boldsymbol{x}_0 + \int_{t_0}^{t} e^{\boldsymbol{A}(t-\sigma)}\boldsymbol{B}\boldsymbol{u}(\sigma)d\sigma] \tag{58}$$

Example (Rugh). Consider the equation for a harmonic oscillator:

$$\ddot{x} = -x \tag{59}$$

If we introduce the variables $x_1 = x$ and $x_2 = \dot{x}$ then $\dot{x}_1 = \dot{x} = x_2$ and $\dot{x}_2 = \ddot{x} = -x_1$ the system becomes

$$\dot{\boldsymbol{x}} = \boldsymbol{A}\boldsymbol{x} \tag{60}$$

$$\begin{bmatrix} \dot{x}_1(t) \\ \dot{x}_2(t) \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} \tag{61}$$

The matrix exponential with $t_0 = 0$ is:

$$e^{\boldsymbol{A}(t-t_0)} = \boldsymbol{I} + \boldsymbol{A}t + \frac{\boldsymbol{A}^2 t^2}{2} + \frac{\boldsymbol{A}^3 t^3}{3!} + \cdots \tag{62}$$

where the first few terms are

$$\boldsymbol{A}^2 = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} = \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix} \tag{63}$$

and

$$\boldsymbol{A}^3 = \boldsymbol{A}^2\boldsymbol{A} = \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \tag{64}$$

and

$$\boldsymbol{A}^4 = \boldsymbol{A}^2\boldsymbol{A}^2 = \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \tag{65}$$

and

$$\boldsymbol{A}^5 = \boldsymbol{A}^4\boldsymbol{A} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \tag{66}$$

The term $\boldsymbol{A}^5$ matches $\boldsymbol{A}$ and it is clear the pattern will continue to repeat. Making use of the Taylor series expansion of sin and cos we obtain:

$$e^{\boldsymbol{A}(t-t_0)} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} t + \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix} \frac{t^2}{2} + \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \frac{t^3}{3!} + \tag{67}$$

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \frac{t^4}{4!} + \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \frac{t^5}{5!} + \cdots \tag{68}$$

$$= \begin{bmatrix} 1 - t^2/2! + t^4/4! + \cdots & t - t^3/3! + t^5/5! + \cdots \\ -(t - t^3/3! + t^5/5!) + \cdots & 1 - t^2/2! + t^4/4! + \cdots \end{bmatrix} \tag{69}$$

$$= \begin{bmatrix} \cos t & \sin t \\ -\sin t & \cos t \end{bmatrix} \tag{70}$$

The solution of the system from initial condition $\boldsymbol{x}(t_0) = [s_0, v_0]^{\mathrm{T}}$ is therefore

$$\begin{aligned} \boldsymbol{x}(t) &= e^{\boldsymbol{A}t}\boldsymbol{x}_0 && (71) \\ &= \begin{bmatrix} \cos t & \sin t \\ -\sin t & \cos t \end{bmatrix} \begin{bmatrix} s_0 \\ v_0 \end{bmatrix} && (72) \\ &= \begin{bmatrix} s_0 \cos t + v_0 \sin t \\ -s_0 \sin t + v_0 \cos t \end{bmatrix} && (73) \end{aligned}$$

State transition matrix via Laplace Transform. The state transition matrix for the system (46) satisfies the $n \times n$ matrix differential equation

$$\dot{\boldsymbol{X}}(t) = \boldsymbol{A}\boldsymbol{X}(t) \tag{74}$$

with $\boldsymbol{X}(t_0) = \boldsymbol{I}$ and has the unique solution

$$\boldsymbol{X}(t) = e^{\boldsymbol{A}t} \tag{75}$$

Taking the Laplace transform of both sides

$$\begin{aligned} s\boldsymbol{X}(s) - \boldsymbol{X}(t_0) &= \boldsymbol{A}\boldsymbol{X}(s) && (76) \\ s\boldsymbol{X}(s) - \boldsymbol{I} &= \boldsymbol{A}\boldsymbol{X}(s) && (77) \\ s\boldsymbol{X}(s) - \boldsymbol{A}\boldsymbol{X}(s) &= \boldsymbol{I} && (78) \\ (s\boldsymbol{I} - \boldsymbol{A})\boldsymbol{X}(s) &= \boldsymbol{I} && (79) \\ (s\boldsymbol{I} - \boldsymbol{A})^{-1}(s\boldsymbol{I} - \boldsymbol{A})\boldsymbol{X}(s) &= (s\boldsymbol{I} - \boldsymbol{A})^{-1}\boldsymbol{I} && (80) \\ \implies \boldsymbol{X}(s) &= (s\boldsymbol{I} - \boldsymbol{A})^{-1} && (81) \end{aligned}$$

From the definition of a matrix inverse

$$(s\boldsymbol{I} - \boldsymbol{A})^{-1} = \frac{\mathrm{adj}(s\boldsymbol{I} - \boldsymbol{A})}{\det(s\boldsymbol{I} - \boldsymbol{A})} \tag{82}$$

then

$$\boldsymbol{X}(t) = e^{\boldsymbol{A}t} = \mathcal{L}^{-1}\left[(s\boldsymbol{I} - \boldsymbol{A})^{-1}\right] \tag{83}$$

where $\mathcal{L}^{-1}$ is the familiar inverse Laplace transform. The inverse Laplace transform can be computed independently for each term in the matrix $(s\boldsymbol{I} - \boldsymbol{A})^{-1}$. The approach presented above provides the state-transition matrix $\Phi(t, 0) = e^{\boldsymbol{A}t}$ which is also the solution $\boldsymbol{X}(t)$ to (74). It is easy to confirm that at $t_0 = 0$ the matrix $e^{\boldsymbol{A}0} = \boldsymbol{I}$ satisfies the required initial condition.

Once $e^{\boldsymbol{A}t}$ is determined it can be used to solve the initial value problem of (46) for an arbitrary initial condition $\boldsymbol{x}_0$ using (56). For example, with initial time $t_0 = 0$ the free response to an initial condition is

$$\boldsymbol{x}(t) = e^{\boldsymbol{A}t}\boldsymbol{x}_0 \tag{84}$$

Note: there are two different ODEs discussed above (46) and (74), the first has an initial condition $\boldsymbol{X}(t_0) = \boldsymbol{I}$ while the second has an initial condition $\boldsymbol{x}(t_0) = \boldsymbol{x}_0$.

Example. Consider the following system (Adapted from Z.Gajic [Link])

$$\dot{\boldsymbol{x}} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{u} \tag{85}$$

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -2 & -3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} u(t) \tag{86}$$

To find the matrix exponential, we first compute the quantity

$$(s\boldsymbol{I} - \boldsymbol{A}) = s \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - \begin{bmatrix} 0 & 1 \\ -2 & -3 \end{bmatrix} \tag{87}$$

$$= \begin{bmatrix} s & -1 \\ 2 & s+3 \end{bmatrix} \tag{88}$$

and its determinant

$$\det\left(\begin{bmatrix} s & -1 \\ 2 & s+3 \end{bmatrix}\right) = s(s+3) - (-1)2 \tag{89}$$

$$= s^2 + 3s + 2 \tag{90}$$

$$= (s+1)(s+2)\ . \tag{91}$$

The adjugate is

$$\operatorname{adj}\left(\begin{bmatrix} s & -1 \\ 2 & s+3 \end{bmatrix}\right) = \begin{bmatrix} (s+3) & 1 \\ -2 & s \end{bmatrix} \tag{92}$$

so that

$$\boldsymbol{F} = e^{\boldsymbol{A}t} = \mathcal{L}^{-1}\left[\frac{\operatorname{adj}(s\boldsymbol{I} - \boldsymbol{A})}{\det(s\boldsymbol{I} - \boldsymbol{A})}\right] \tag{93}$$

$$= \mathcal{L}^{-1}\left\{\frac{1}{(s+1)(s+2)} \begin{bmatrix} (s+3) & 1 \\ -2 & s \end{bmatrix}\right\} \tag{94}$$

We may compute the inverse Laplace transform of each term individually in the matrix (e.g., using Laplace transform tables and partial fraction expansion). Begin by computing the partial fraction expansion of the first term:

$$\frac{s+3}{(s+1)(s+2)} = \frac{A}{(s+1)} + \frac{B}{(s+2)} \tag{95}$$

To evaluate for A multiply both sides by $(s+1)$ and evaluate at $s = -1$:

$$\left.\frac{s+3}{(s+2)}\right|_{s=-1} = A + \left.\frac{B(s+1)}{(s+2)}\right|_{s=-1} \tag{96}$$

$$\implies A = 2 \tag{97}$$

and similarly

$$\left.\frac{s+3}{(s+1)}\right|_{s=-2} = \left.\frac{A(s+2)}{(s+1)}\right|_{s=-2} + B \tag{98}$$

$$\implies B = -1 \tag{99}$$

and using a Laplace transform table:

$$\mathcal{L}^{-1}\left[\frac{2}{(s+1)} - \frac{1}{(s+2)}\right] = 2\mathcal{L}^{-1}\left[\frac{1}{(s+1)}\right] - \mathcal{L}^{-1}\left[\frac{1}{(s+2)}\right] \tag{100}$$

$$= 2e^{-t} - e^{-2t} \tag{101}$$

Repeating the process for the remaining three matrix entries gives

$$e^{\boldsymbol{A}t} = \begin{bmatrix} 2e^{-t} - e^{-2t} & e^{-t} - e^{-2t} \\ 2e^{-2t} - 2e^{-t} & 2e^{-2t} - e^{-t} \end{bmatrix} \tag{102}$$

Note that this procedure can also be carried out using MATLAB's symbolic toolbox:

```
syms s;
A = [0 1; -2 -3];
Q = (s*eye(2,2)-A);
F = ilaplace(adjoint(Q)./det(Q))
latex(F)
```

Once the matrix exponential is found, we can determine the free response to an initial condition. For example, if $\boldsymbol{x} = [2, 3]^{\mathrm{T}}$ then:

$$\boldsymbol{x}(t) = e^{\boldsymbol{A}t}\boldsymbol{x}_0 \tag{103}$$

$$= \begin{bmatrix} 2e^{-t} - e^{-2t} & e^{-t} - e^{-2t} \\ 2e^{-2t} - 2e^{-t} & 2e^{-2t} - e^{-t} \end{bmatrix} \begin{bmatrix} 2 \\ 3 \end{bmatrix} \tag{104}$$

$$= \begin{bmatrix} 4e^{-t} - 2e^{-2t} + 3e^{-t} - 3e^{-2t} \\ 4e^{-2t} - 3e^{-t} + 6e^{-2t} - 3e^{-t} \end{bmatrix} \tag{105}$$

State transition matrix via Cayley-Hamilton Theorem. Another method for computing the matrix exponential relies on the *Cayley-Hamilton theorem*. Consider a square matrix $\boldsymbol{A} \in \mathbb{R}^{n \times n}$. The characteristic polynomial for $\boldsymbol{A}$ is a function of a variable s found by taking the determinant of $s\boldsymbol{I} - \boldsymbol{A}$. That is,

$$\Delta(s) = \det(s\boldsymbol{I} - \boldsymbol{A}) = s^n + c_{n-1}s^{n-1} + \cdots + c_0 = 0 \tag{106}$$

and solutions of $\Delta(s) = 0$ are eigenvalues of $\boldsymbol{A}$. The corresponding matrix polynomial is obtained by substituting $\boldsymbol{A}$ for s above:

$$\Delta(\boldsymbol{A}) = \boldsymbol{A}^n + c_{n-1}\boldsymbol{A}^{n-1} + \cdots + c_0\boldsymbol{I} \tag{107}$$

The Cayley-Hamilton theorem states that every square matrix satisfies

$$\Delta(\boldsymbol{A}) = \boldsymbol{0} \tag{108}$$

where $\boldsymbol{0}$ is a matrix of all zeros. That is, the matrix $\boldsymbol{A}$ satisfies its own characteristic polynomial. This is useful because it provides a method or reducing the order of any polynomial in $\boldsymbol{A}$, including (52). The basic idea is to use polynomial long division to decompose the function into two parts: one involves an infinite sum that multiplies a quotient times $\Delta(\boldsymbol{A})$ and the other is a remainder that only contains n terms. Since the $\Delta(\boldsymbol{A}) = 0$ the first term vanishes — see here for more details. Therefore, it can be shown that the matrix exponential

$$e^{\boldsymbol{A}t} = \sum_{k=0}^{n-1} \alpha_k \boldsymbol{A}^k \tag{109}$$

where coefficients α_k are determined by solving the following system of equations that depends on the eigenvalues of $\boldsymbol{A}$

$$e^{\lambda_i t} = \sum_{k=0}^{n-1} \alpha_k \lambda_i^k \tag{110}$$

Comparing (109) to (52) we see that the summation only contains n terms rather than infinitely many!

Example. Consider the same system as in the previous example (with no control input):

$$\dot{\boldsymbol{x}} = \boldsymbol{A}\boldsymbol{x} \tag{111}$$

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -2 & -3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \tag{112}$$

The characteristic equation is

$$\det(s\boldsymbol{I} - \boldsymbol{A}) = \det\left(\begin{bmatrix} s & -1 \\ 2 & s+3 \end{bmatrix}\right) = s(s+3) + 2 = s^2 + 3s + 2 = 0 \tag{113}$$

and the eigenvalues (that satisfy the above equation) are $\lambda_1 = -1$ and $\lambda_2 = -2$. From (109) with $n = 2$

$$e^{\boldsymbol{A}t} = \alpha_0 \boldsymbol{I} + \alpha_1 \boldsymbol{A} \tag{114}$$

since $\boldsymbol{A}^0 = \boldsymbol{I}$ is the identity matrix. The coefficients α_0 and α_1 are found from (110) which defines a system of equations (one for each eigenvalue:)

$$e^{\lambda_1 t} = \alpha_0 + \alpha_1 \lambda_1 \tag{115}$$

$$e^{\lambda_2 t} = \alpha_0 + \alpha_1 \lambda_2 \tag{116}$$

where we've used $\lambda_1^0 = \lambda_2^0 = 1$. With $\lambda_1 = -1$ and $\lambda_2 = -2$ this system is

$$e^{-t} = \alpha_0 + -\alpha_1 \tag{117}$$

$$e^{-2t} = \alpha_0 - 2\alpha_1 \tag{118}$$

and its solution is $\alpha_0 = 2e^{-t} - e^{-2t}$ and $\alpha_1 = e^{-t} - e^{-2t}$. Thus, the matrix exponential is

$$e^{\boldsymbol{A}t} = (2e^{-t} - e^{-2t}) \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + (e^{-t} - e^{-2t}) \begin{bmatrix} 0 & 1 \\ -2 & -3 \end{bmatrix} \tag{119}$$

$$= \begin{bmatrix} 2e^{-t} - e^{-2t} & (e^{-t} - e^{-2t}) \\ -2(e^{-t} - e^{-2t}) & 2e^{-t} - e^{-2t} - 3(e^{-t} - e^{-2t}) \end{bmatrix} \tag{120}$$

$$\implies e^{\boldsymbol{A}t} = \begin{bmatrix} 2e^{-t} - e^{-2t} & e^{-t} - e^{-2t} \\ -2e^{-t} + 2e^{-2t} & -e^{-t} + 2e^{2-t} \end{bmatrix} \tag{121}$$

Nonlinear Continuous-Time Models

Systems that are not linear are called nonlinear and are written as:

$$\dot{\boldsymbol{x}}(t) = \boldsymbol{f}(\boldsymbol{x}(t), \boldsymbol{u}(t)) \tag{122}$$
$$\boldsymbol{y}(t) = \boldsymbol{h}(\boldsymbol{x}(t)) \tag{123}$$
$$\boldsymbol{x}(t_0) = \boldsymbol{x}_0 \tag{124}$$

For nonlinear systems there is no convenient expression for the solution $\boldsymbol{x}(t)$. In fact, nonlinear solutions may have no solutions. In some cases, as in the example below, we can integrate the system to obtain a solution. However, this is not always the case. Many state estimation and system identification techniques are applicable to nonlinear systems. One way to handle a nonlinear system is to approximate it as linear systems through the process of *linearization*.

Example. Consider the scalar system

$$\dot{x} = x^2 \tag{125}$$

with initial condition $x(t_0) = x_0$. Clearly this system is nonlinear but it is separable. Rewrite it as

$$\dot{x} = \frac{dx}{dt} = x^2 \tag{126}$$
$$\frac{1}{x^2}\, dx = dt \tag{127}$$

and integrate

$$\int_{x_0}^{x(t)} \frac{1}{x^2}\, dx = \int_{x_0}^{x(t)} dt \tag{128}$$
$$-\frac{1}{x}\bigg|_{x_0}^{x(t)} = (t - t_0) \tag{129}$$
$$-\frac{1}{x(t)} + \frac{1}{x_0} = (t - t_0) \tag{130}$$
$$-x_0 + x(t) = x(t)x_0(t - t_0) \tag{131}$$
$$x(t) - x(t)x_0(t - t_0) = x_0 \tag{132}$$
$$\implies x(t) = \frac{x_0}{1 - x_0(t - t_0)} \tag{133}$$

Rewriting Higher Order ODEs in First-Order Form

A higher-order differential equation can always be converted into a system of first-order equations form by defining additional state variables representing derivatives of the state. Consider the following example.

Example. Consider the system below

$$\dddot{x} = \dot{x} + 3x \ . \tag{134}$$

Since it is third-order define three variables: $x_1 = x$, $x_2 = \dot{x}$, $x_3 = \ddot{x}$, so that

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \dot{x}_3 \end{bmatrix} = \begin{bmatrix} \dot{x} \\ \ddot{x} \\ \dddot{x} \end{bmatrix} = \begin{bmatrix} x_2 \\ x_3 \\ x_2 + 3x_1 \end{bmatrix} \tag{135}$$

The third-order equation (134) has been converted into three first-order equations (135).

Linearizing Nonlinear Systems

Consider the nonlinear system

$$\dot{\boldsymbol{x}} = \boldsymbol{f}(\boldsymbol{x}, \boldsymbol{u}, t) \tag{136}$$

where we have dropped the time-dependency of the variables, the state vector is $\boldsymbol{x} = [x_1, x_2, \ldots, x_n]^{\mathrm{T}}$, and the control vector is $\boldsymbol{u} = [u_1, u_2, \ldots, u_m]^{\mathrm{T}}$. To show the dependence on the individual components we can re-write the system as

$$\dot{\boldsymbol{x}} = \boldsymbol{f}(x_1, x_2, \ldots, x_n, u_1, u_2, \ldots, u_m, t) \tag{137}$$

Moreover, $\boldsymbol{f} = [f_1(\cdot), f_2(\cdot), \ldots, f_n(\cdot)]^{\mathrm{T}}$ is a vector and an equivalent expression to (136) and (137) is

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \vdots \\ \dot{x}_n \end{bmatrix} = \begin{bmatrix} f_1(x_1, x_2, \ldots, x_n, u_1, u_2, \ldots, u_m, t) \\ f_2(x_1, x_2, \ldots, x_n, u_1, u_2, \ldots, u_m, t) \\ \vdots \\ f_n(x_1, x_2, \ldots, x_n, u_1, u_2, \ldots, u_m, t) \end{bmatrix} \tag{138}$$

Suppose a nominal/reference trajectory $\boldsymbol{x}^{\mathrm{ref}}(t)$ and control input pair $\boldsymbol{u}^{\mathrm{ref}}(t)$ satisfy

$$\dot{\boldsymbol{x}}^{\mathrm{ref}} = \boldsymbol{f}(\boldsymbol{x}^{\mathrm{ref}}, \boldsymbol{u}^{\mathrm{ref}}, t) \tag{139}$$

We wish to approximate the dynamics of the system when it remains "close" to this nominal trajectory. Often $\boldsymbol{x}^{\mathrm{ref}}, \boldsymbol{u}^{\mathrm{ref}}$ are chosen to be an equilibrium point such that

$$\dot{\boldsymbol{x}}^{\mathrm{ref}} = \boldsymbol{f}(\boldsymbol{x}^{\mathrm{ref}}, \boldsymbol{u}^{\mathrm{ref}}, t) = 0 \tag{140}$$

but this is not required. Define the deviation of the system from this nominal trajectory as

$$\Delta\boldsymbol{x} = \boldsymbol{x}(t) - \boldsymbol{x}^{\mathrm{ref}}(t) = \begin{bmatrix} x_1(t) - x_1^{\mathrm{ref}}(t) \\ x_2(t) - x_2^{\mathrm{ref}}(t) \\ \vdots \\ x_n(t) - x_n^{\mathrm{ref}}(t) \end{bmatrix} \quad \text{and} \quad \Delta\boldsymbol{u} = \boldsymbol{u}(t) - \boldsymbol{u}^{\mathrm{ref}}(t) = \begin{bmatrix} u_1(t) - u_1^{\mathrm{ref}}(t) \\ u_2(t) - u_2^{\mathrm{ref}}(t) \\ \vdots \\ u_n(t) - u_n^{\mathrm{ref}}(t) \end{bmatrix} \tag{141}$$

Our goal is to construct a linear system of the form

$$\Delta \dot{\boldsymbol{x}} = \boldsymbol{A}\Delta \boldsymbol{x} + \boldsymbol{B}\Delta \boldsymbol{u} \tag{142}$$

which models the deviations from the nominal trajectory. The approach is to re-write the original (nonlinear) system in terms of the nominal and deviation variables and then approximate it with a Taylor series around the nominal trajectory. Let's begin by considering just the dynamics of the first state:

$$\dot{x}_1 = f_1(x_1, \ldots, x_n, u_1, \ldots, u_m, t) \tag{143}$$

Note that $\boldsymbol{x} = \boldsymbol{x}^{\text{ref}} + \Delta \boldsymbol{x}$ and $\boldsymbol{u} = \boldsymbol{u}^{\text{ref}} + \Delta \boldsymbol{u}$ or, for just the first state, $x_1 = x_1^{\text{ref}} + \Delta x_1$ and $u_1 = u_1^{\text{ref}} + \Delta u_1$. Then inserting the reference and deviation variables into (143) gives:

$$\dot{x}_1^{\text{ref}} + \Delta \dot{x}_1 = f_1((x_1^{\text{ref}} + \Delta x_1), \ldots, (x_n^{\text{ref}} + \Delta x_n), (u_1^{\text{ref}} + \Delta u_1), \ldots, (u_m^{\text{ref}} + \Delta u_m), t) \tag{144}$$

Compute a Taylor series expansion around the nominal the trajectory:

$$\begin{aligned}
\dot{x}_1^{\text{ref}} + \Delta \dot{x}_1 \approx\ & f_1(x_1^{\text{ref}}, \ldots, x_n^{\text{ref}}, u_1^{\text{ref}}, \ldots, u_m^{\text{ref}}, t) \\
& + \left.\frac{\partial f_1}{d x_1}\right|_{\boldsymbol{x}^{\text{ref}}(t), \boldsymbol{u}^{\text{ref}}(t)} (x_1 - x_1^{\text{ref}}) + \cdots + \left.\frac{\partial f_n}{d x_n}\right|_{\boldsymbol{x}^{\text{ref}}(t), \boldsymbol{u}^{\text{ref}}(t)} (x_n - x_n^{\text{ref}}) + \text{h.o.t.} \\
& + \left.\frac{\partial f_1}{d u_1}\right|_{\boldsymbol{x}^{\text{ref}}(t), \boldsymbol{u}^{\text{ref}}(t)} (u_1 - u_1^{\text{ref}}) + \cdots + \left.\frac{\partial f_n}{d u_m}\right|_{\boldsymbol{x}^{\text{ref}}(t), \boldsymbol{u}^{\text{ref}}(t)} (u_m - u_m^{\text{ref}}) + \text{h.o.t.}
\end{aligned} \tag{145}$$

By assumption, the nominal trajectory satisfies

$$\dot{x}_1^{\text{ref}} = f_1(x_1^{\text{ref}}, \ldots, x_n^{\text{ref}}, u_1^{\text{ref}}, \ldots, u_m^{\text{ref}}, t) \tag{146}$$

hence these terms are equal on both the LHS and RHS of (145) and can be removed. Now, rewrite (145) more compactly in matrix notation

$$\Delta \dot{x}_1 \approx \left[\frac{\partial f_1}{\partial x_1}, \cdots, \frac{\partial f_1}{\partial x_n}\right]\Bigg|_{\boldsymbol{x}^{\text{ref}}(t), \boldsymbol{u}^{\text{ref}}(t)} \Delta \boldsymbol{x} + \left[\frac{\partial f_1}{\partial u_1}, \cdots, \frac{\partial f_1}{\partial u_m}\right]\Bigg|_{\boldsymbol{x}^{\text{ref}}(t), \boldsymbol{u}^{\text{ref}}(t)} \Delta \boldsymbol{u} \tag{147}$$

where each term is a product of a $1 \times n$ vector with a $n \times 1$ vector (resulting in a scalar). If we repeat all of the above steps for all n states we obtain:

$$\Delta \dot{x}_2 \approx \left[\frac{\partial f_2}{\partial x_1}, \cdots, \frac{\partial f_2}{\partial x_n}\right]\Bigg|_{\boldsymbol{x}^{\text{ref}}(t), \boldsymbol{u}^{\text{ref}}(t)} \Delta \boldsymbol{x} + \left[\frac{\partial f_2}{\partial u_1}, \cdots, \frac{\partial f_2}{\partial u_m}\right]\Bigg|_{\boldsymbol{x}^{\text{ref}}(t), \boldsymbol{u}^{\text{ref}}(t)} \Delta \boldsymbol{u} \tag{148}$$

$$\vdots \tag{149}$$

$$\Delta \dot{x}_n \approx \left[\frac{\partial f_n}{\partial x_1}, \cdots, \frac{\partial f_n}{\partial x_n}\right]\Bigg|_{\boldsymbol{x}^{\text{ref}}(t), \boldsymbol{u}^{\text{ref}}(t)} \Delta \boldsymbol{x} + \left[\frac{\partial f_n}{\partial u_1}, \cdots, \frac{\partial f_n}{\partial u_m}\right]\Bigg|_{\boldsymbol{x}^{\text{ref}}(t), \boldsymbol{u}^{\text{ref}}(t)} \Delta \boldsymbol{u} \tag{150}$$

which can be further simplified by arraying all of the partial derivatives as the *Jacobian matrices*

$$\boldsymbol{J_x f} = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_n}{\partial x_1} & \cdots & \frac{\partial f_n}{\partial x_n} \end{bmatrix} \quad \text{and} \quad \boldsymbol{J_u f} = \begin{bmatrix} \frac{\partial f_1}{\partial u_1} & \cdots & \frac{\partial f_1}{\partial u_m} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_n}{\partial u_1} & \cdots & \frac{\partial f_n}{\partial u_m} \end{bmatrix} \tag{151}$$

Next, evaluate the Jacobians along the nominal trajectory/control to obtain two (possibly time-varying) matrices:

$$\boldsymbol{A}(t) = \boldsymbol{J}_{\boldsymbol{x}}\boldsymbol{f}\Big|_{\boldsymbol{x}^{\text{ref}}(t),\boldsymbol{u}^{\text{ref}}(t)} \qquad \text{and} \qquad \boldsymbol{B}(t) = \boldsymbol{J}_{\boldsymbol{u}}\boldsymbol{f}\Big|_{\boldsymbol{x}^{\text{ref}}(t),\boldsymbol{u}^{\text{ref}}(t)} \tag{152}$$

Then the set of equations (147)–(150) is equivalent to the linear system

$$\Delta\dot{\boldsymbol{x}} = \boldsymbol{A}(t)\Delta\boldsymbol{x} + \boldsymbol{B}(t)\Delta\boldsymbol{u} \tag{153}$$

and we say that (153) is a linearization of the nonlinear system (136) around the nominal trajectory $(\boldsymbol{x}^{\text{ref}}(t), \boldsymbol{u}^{\text{ref}}(t))$. Note that if the Jacobians are independent of time and the nominal trajectory $(\boldsymbol{x}^{\text{ref}}(t), \boldsymbol{u}^{\text{ref}}(t))$ is instead a nominal point (i.e., not time-varying) then the resulting linearized system is LTI. Also, you may often encounter the notation in which the Δ notation is removed and (153) is expressed as

$$\dot{\boldsymbol{x}} = \boldsymbol{A}(t)\boldsymbol{x} + \boldsymbol{B}(t)\boldsymbol{u} \tag{154}$$

In such cases it is understood that the linearized states and controls are in fact deviation states from the reference. A similar process can be used to linearize the nonlinear output equation

$$\boldsymbol{y} = \boldsymbol{h}(\boldsymbol{x}, \boldsymbol{u}) \tag{155}$$

Define the nominal observations corresponding to the nominal trajectory $\boldsymbol{y}^{\text{ref}} = \boldsymbol{h}(\boldsymbol{x}^{\text{ref}}, \boldsymbol{u}^{\text{ref}})$ and the deviations are then

$$\Delta\boldsymbol{y} = \boldsymbol{y} - \boldsymbol{y}^{\text{ref}} \tag{156}$$

Following a similar procedure we obtain a linearized system

$$\Delta\boldsymbol{y} = \boldsymbol{C}(t)\Delta\boldsymbol{x} + \boldsymbol{D}(t)\Delta\boldsymbol{u} \tag{157}$$

where

$$\boldsymbol{C} = \boldsymbol{J}_{\boldsymbol{x}}\boldsymbol{h}\Big|_{\boldsymbol{x}^{\text{ref}}(t),\boldsymbol{u}^{\text{ref}}(t)} \qquad \text{and} \qquad \boldsymbol{D} = \boldsymbol{J}_{\boldsymbol{u}}\boldsymbol{h}\Big|_{\boldsymbol{x}^{\text{ref}}(t),\boldsymbol{u}^{\text{ref}}(t)} \tag{158}$$

Example. The rotational motion of a damped pendulum with torque input is described by the equation

$$\ddot{\theta} = -\frac{d}{m}\dot{\theta} - \frac{g}{L}\sin\theta + \tau(t) \tag{159}$$

where θ is the pendulum angle, $\dot{\theta}$ is the angular rate, L is the length, m is the mass, d is a damping coefficient, and $\tau(t)$ is a torque control input. We can write this as a set of nonlinear first-order equations by defining $x_1 = \theta$, $x_2 = \dot{\theta}$, and $u = \tau$ so that

$$\dot{\boldsymbol{x}} = \boldsymbol{f}(\boldsymbol{x}, u) \tag{160}$$

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \begin{bmatrix} x_2 \\ -ax_2 - b\sin x_1 + u(t) \end{bmatrix} \tag{161}$$

where we've introduced the new constants $a = d/m$ and $b = g/L$ for convenience. We wish to linearize this system around a nominal trajectory which we choose to be the equilibrium point

$$x_1^{\text{ref}}(t) = 0 \quad \text{and} \quad x_2^{\text{ref}}(t) = 0 \quad \text{and} \quad u^{\text{ref}}(t) = 0 \tag{162}$$

This point corresponds to the pendulum hanging down with no torque applied. Begin by computing the Jacobians with $f_1(x_1, x_2, u) = x_2$ and $f_2(x_1, x_2, u) = -ax_2 - b\sin x_1 + u(t)$:

$$\boldsymbol{J_x f} = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -b\cos x_1 & -a \end{bmatrix} \tag{163}$$

and

$$\boldsymbol{J_u f} = \begin{bmatrix} \frac{\partial f_1}{\partial u} \\ \frac{\partial f_2}{\partial u} \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \tag{164}$$

Then evaluating the Jacobians at the equilibrium

$$\boldsymbol{A} = \boldsymbol{J_x f}\Big|_{x_1^{\text{ref}}=0,\ x_2^{\text{ref}}=0,\ u^{\text{ref}}=0} = \begin{bmatrix} 0 & 1 \\ -b & -a \end{bmatrix} \tag{165}$$

and

$$\boldsymbol{B} = \boldsymbol{J_u f}\Big|_{x_1^{\text{ref}}=0,\ x_2^{\text{ref}}=0,\ u^{\text{ref}}=0} = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \tag{166}$$

Now introduce the deviation states $\Delta\boldsymbol{x} = \boldsymbol{x} - \boldsymbol{x}^{\text{ref}}$ and deviation control $\Delta u = u - u^{\text{ref}}$. (In this particular case all of the references are zero, hence $\Delta\boldsymbol{x} = \boldsymbol{x}$ and $\Delta u = u$.) The linearized system is

$$\Delta\dot{\boldsymbol{x}} = \boldsymbol{A}\Delta\boldsymbol{x} + \boldsymbol{B}\Delta u \tag{167}$$

$$\begin{bmatrix} \Delta\dot{x}_1 \\ \Delta\dot{x}_2 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -b & -a \end{bmatrix} \begin{bmatrix} \Delta x_1 \\ \Delta x_2 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} \Delta u \tag{168}$$

References

[Hespanha, 2018] Hespanha, J. P. (2018). *Linear Systems Theory*. Princeton university press.

[Rugh, 1996] Rugh, W. J. (1996). *Linear System Theory*. Prentice-Hall, Inc.

[Stilwell, 2012] Stilwell, D. (2012). Lecture notes in ECE 5744: Linear Systems Theory.