EML 4312: Control of Mechanical Engineering Systems

Mrinal Kumar©

Sep 14, 2015

Dynamic Systems Analysis

Modeling of Dynamical Systems

Consider again the generic feedback loop we have considered so far (Fig.1). This set of notes will look closer at the block named "plant", which is nothing but the "process" we are striving to control. We will study elements of state-space modeling, linearization and various analysis tools for linear systems.

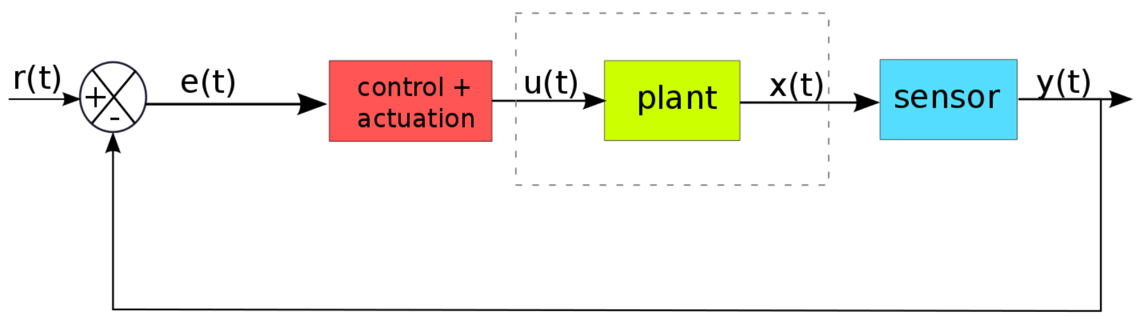


Figure 1: The Feedback Schematic

- In the broadest sense, a *model* is nothing but the mathematical representation of physical (or chemical, biological etc.) reality. It constitutes the starting point for analysis and design. In this course, we are interested in *dynamical systems*, i.e. systems that evolve over time, in which the effect of actions do not manifest instantly. E.g. it takes a while for a vehicle to come to a stop after brakes have been applied, or the temperature of a cold room to reach a comfortable level after the heater has been switched on. The most common model for time-evolving systems is ordinary differential equations (ODEs)[1].
- The first step in constructing a model for a dynamical system (DS) is the *identification of its **states***. A "state" is a variable that describes the condition, or configuration of the DS. The *minimum* number of states needed to fully describe a system is a common yardstick for measuring its complexity.
- The next step involves applying the laws of physics to obtain the dynamical equations of each state. For a mechanical system, you may employ Newton's laws of motion, for an electrical system you may need Kirchhoff's circuit laws, etc. This step, i.e. derivation of the evolution equations, typically in the form of ODE's, is skipped in this course. The governing equations of system dynamics will always be given.

[1]another common model is *partial differential equations* (PDEs), which are significantly more complex

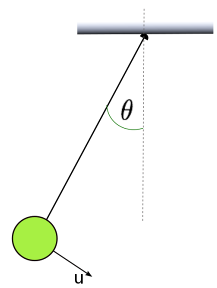


Figure 2: The Simple Pendulum

❣ **Example** A simple dynamical system that evolves over time is the so-called *simple pendulum*, shown in Fig.2. This system has two states, namely the angular position of the point mass: θ, and its angular speed: $\dot{\theta}$. Its physics is known to be described by the following second order ODE:

$$\ddot{\theta} = -\frac{g}{l}\sin\theta + u \tag{1}$$

where g is the constant acceleration due to gravity and l is the length of the string. Note that from the above governing equation, it is difficult to determine what the states of the system are. Since we already identified the states as θ and $\dot{\theta}$, so it makes sense to break down Eq.(1) to write a separate governing equation *for each state*. First, redefine the states as

$$x_1 \triangleq \theta \tag{2a}$$

$$x_2 \triangleq \dot{\theta} \tag{2b}$$

Now it is easy to see that

$$\dot{x}_1 \underset{\text{Eq.(2a)}}{=} \dot{\theta} \underset{\text{Eq.(2b)}}{=} x_2 \tag{3a}$$

$$\dot{x}_2 \underset{\text{Eq.(2b)}}{=} \ddot{\theta} \underset{\text{Eq.(1)}}{=} -\frac{g}{l}\sin\theta + u \underset{\text{Eq.(2a)}}{=} -\frac{g}{l}\sin x_1 + u \tag{3b}$$

Cleaning up, we get

$$\boxed{\begin{aligned}\dot{x}_1 &= x_2 \\ \dot{x}_2 &= -\frac{g}{l}\sin x_1 + u\end{aligned}} \qquad (4a),\ (4b)$$

Clearly, the above system of equations is more descriptive because it allows us to easily identify individual states of the dynamical system (on the LHS), and their corresponding dynamics (on the RHS).

- The above structure is indeed special, and is referred to as the **state-space form** of the dynamical system. Essentially, the state-space (ss) form has *one first-order ODE for each state of the DS*. It is very useful because as mentioned above, it enumerates all the states of the DS right off the bat. Recall that a state-variable helps describe the condition of the process and if "x" is a state, you *must* have a first-order ODE for it!

- What happened in the above developments? We took a particular example (a simple pendulum) and starting with the governing physics (Eq.(1)), wrote down its evolution in terms of generic state variables, x_1 and x_2: Eqs.(4). In other words, we added some mathematical abstraction to the otherwise highly intuitive pendulum model. Indeed, if nothing but Eqs.(4) were given, it would not be trivial to immediately identify the system of equations as a mathematical model for a simple pendulum!

The obvious giveaway would be the right-hand side, which "looks like the dynamics of a simple pendulum..." So, there is room to generalize more!

- **General state-space representation of dynamic systems.** The RHS of Eqs.(3)-(4) are nothing but functions of the state and control variables (x_1, x_2, u). So, the ss form of the simple pendulum can be written in very general fashion as:

$$\dot{x}_1 = x_2 = f_1(x_1, x_2, u), \tag{5a}$$

$$\dot{x}_2 = -\frac{g}{l}\sin x_1 + u = f_2(x_1, x_2, u) \tag{5b}$$

where, the functions f_1 and f_2 describe the dynamics of states x_1 and x_2 respectively. Or, in **vector** form:

$$\begin{Bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{Bmatrix} = \begin{Bmatrix} f_1 \\ f_2 \end{Bmatrix} = \begin{Bmatrix} f_1(x_1, x_2, u) \\ f_2(x_1, x_2, u) \end{Bmatrix} \tag{6}$$

More compactly,

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}, \mathbf{u}) \tag{7}$$

where, **bold** represents a vector. Generalize one step further to allow the "dynamics" ($\mathbf{f}$) be an explicit function of time:

$$\boxed{\dot{\mathbf{x}} = \mathbf{f}(t, \mathbf{x}, \mathbf{u})} \tag{8}$$

Eq.(8) is a general state-space model of a dynamic system.

Observations:

- A general dynamic system has n states, i.e. $\mathbf{x} = [x_1, x_2, \ldots, x_n]^{\mathrm{T}}$, where "T" stands for transpose; and n control inputs: $\mathbf{u} = [u_1, u_2, \ldots u_m]^{\mathrm{T}}$. For the special case of the simple pendulum, $n = 2$ with $x_1 = \theta$ and $x_2 = \dot{\theta}$; and $m = 1$, i.e. the control is simply a scalar.
- In Eq.(8), the LHS contains the time derivative of each state:

$$\dot{\mathbf{x}} = \begin{Bmatrix} \frac{dx_1}{dt} \\ \frac{dx_2}{dt} \\ \vdots \\ \frac{dx_n}{dt} \end{Bmatrix} \tag{9}$$

- The RHS of Eq.(8) contains the evolution law for each state and in general, can be referred to as the system **dynamics**:

$$\mathbf{f}(t, \mathbf{x}, \mathbf{u}) = \begin{Bmatrix} f_1(t, x_1, x_2, \ldots, x_n, u_1, u_2, \ldots u_m) \\ f_2(t, x_1, x_2, \ldots, x_n, u_1, u_2, \ldots u_m) \\ \vdots \\ f_n(t, x_1, x_2, \ldots, x_n, u_1, u_2, \ldots u_m) \end{Bmatrix} \tag{10}$$

- So, here's what we did: we started with an actual physical system, namely the simple pendulum (Eq.(1)) and cast it in state-space form (Eq.(4)). Then, realizing that other dynamical systems can have more than two states and different form of dynamics, formulated a general state-space form of a dynamic system, given by Eq.(8).
- *It is very important to note that the state-space form was a crucial development in control theory because it is the foundation of **Modern Control**, a paradigm that was developed in the early 1960's and quickly became the accepted framework for analysis of control systems.*

- Let us consider another example, somewhat more complex than the simple pendulum.

❣ **Example** Fig.(3) shows a two-mass spring damper system in which the second spring (k_2) ensures that masses m_1 and m_2 both have their independent degrees of freedom.

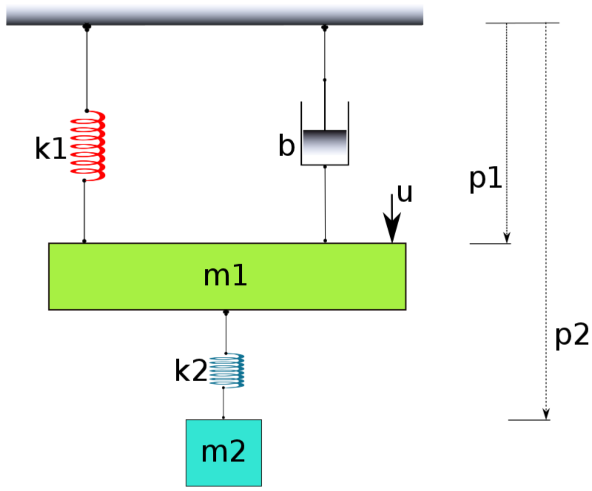


Figure 3: A two-mass spring damper system

The physics is given by:

$$\begin{aligned} m_1\ddot{p}_1 + k_2(p_1 - p_2) + k_1 p_1 + b\dot{p}_1 &= u(t) \\ m_2\ddot{p}_2 + k_2(p_2 - p_1) &= 0 \end{aligned}$$

or, written differently,

$$\ddot{p}_1 = -\left[\left(\frac{k_1}{m_1} + \frac{k_2}{m_1}\right)\right]p_1 + \frac{k_2}{m_1}p_2 - \frac{b}{m_1}\dot{p}_1 + u(t) \tag{12a}$$

$$\ddot{p}_2 = -\frac{k_2}{m_2}p_2 + \frac{k_2}{m_2}p_1 \tag{12b}$$

where, p_1 and p_2 are the positions of masses m_1 and m_2; k_1 and k_2 are the spring constants and b is the damping coefficient.

- The first observation is that the above equation is **not** in state space form (note the second derivatives!).
- Let's identify the states of this system: we have two degrees of freedom, one for each block, which is in turn governed by second-order dynamics. So, the states are:
  * $x_1 \triangleq p_1$ (position of m_1),
  * $x_2 \triangleq p_2$ (position of m_2),
  * $x_3 \triangleq \dot{p}_1$ (speed of m_1) and,
  * $x_4 \triangleq \dot{p}_2$ (speed of m_2)
- The order of the above definitions is immaterial. E.g. you may very well want to label x_2 as $\dot{p}_1$ and x_3 as p_2, etc. You just need to define the labels once and stick to your definitions.

– Using the definitions given above, the state-space form of the system becomes:

$$\begin{Bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \dot{x}_3 \\ \dot{x}_4 \end{Bmatrix} = \begin{Bmatrix} x_3 \\ x_4 \\ -\frac{k_1+k_2}{m_1}x_1 + \frac{k_2}{m_1}x_2 - \frac{b}{m_1}x_3 \\ \frac{k_2}{m_2}x_1 - \frac{k_2}{m_2}x_2 \end{Bmatrix} + \begin{Bmatrix} 0 \\ 0 \\ \frac{u}{m_1} \\ 0 \end{Bmatrix} \tag{13}$$

– It is easy to see that the "dimensionality of the state-space" is $n = 4$ and "dimensionality of the control" is $m = 1$.

- **Linear state-space form.** It is interesting to note that the dynamics of the double-mass spring damper system can be "separated" into the following very special form:

$$\dot{\mathbf{x}} = \begin{Bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \dot{x}_3 \\ \dot{x}_4 \end{Bmatrix} = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ -\frac{k_1+k_2}{m_1} & \frac{k_2}{m_1} & -\frac{b}{m_1} & 0 \\ \frac{k_2}{m_2} & -\frac{k_2}{m_2} & 0 & 0 \end{bmatrix} \begin{Bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{Bmatrix} + \begin{bmatrix} 0 \\ 0 \\ \frac{1}{m_1} \\ 0 \end{bmatrix} u \tag{14}$$

Or,

$$\boxed{\underset{(n\times 1)}{\dot{\mathbf{x}}} = \underset{(n\times n)}{\mathbf{A}}\underset{(n\times 1)}{\mathbf{x}} + \underset{(n\times m)}{\mathbf{B}}\underset{(m\times 1)}{\mathbf{u}}} \tag{15}$$

where, $\mathbf{A}$ is called the **system matrix** and $\mathbf{B}$ is called the **control influence matrix**.

So, what is the relationship between a general dynamic system in state-space form given in Eq.(8) and the form given in Eq.(15)? Essentially, the latter is a special case of the former. In particular, *if* the system dynamics "*bdyn*" can be represented in form given in Eq.(15), i.e. "$\mathbf{Ax}+\mathbf{Bu}$", *then* the dynamic system is called **linear**.

$$\boxed{\begin{array}{c}\text{Dynamic system } \dot{\mathbf{x}} = \mathbf{f}(t,\mathbf{x},\mathbf{u}) \text{ is } \textit{linear} \text{ if} \\ \mathbf{f}(t,\mathbf{x},\mathbf{u}) \equiv \mathbf{Ax} + \mathbf{Bu}\end{array}}$$

The contents of matrices $\mathbf{A}$ and $\mathbf{B}$ depends on the details of the system under consideration.

❢ **Example** Let us revisit our beloved simple pendulum (Eqs.(4)) and see if it can be cast in the form of a linear system. See below that this is not possible..

$$\begin{Bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{Bmatrix} \underset{\text{Eq.(4)}}{=} \begin{Bmatrix} x_2 \\ -\frac{g}{l}\sin x_1 + u \end{Bmatrix} \not\equiv \begin{bmatrix} 0 & 1 \\ \otimes & 0 \end{bmatrix} \begin{Bmatrix} x_1 \\ x_2 \end{Bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} u$$

however... if you made the so-called "small angle approximation".. i.e. $\sin x_1 \approx x_1$ then something amazing happens:

$$\begin{Bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{Bmatrix} \underset{\text{Eq.(4)}}{=} \begin{Bmatrix} x_2 \\ -\frac{g}{l}\sin x_1 + u \end{Bmatrix} \underset{\text{small angle..}}{\approx} \begin{Bmatrix} x_2 \\ -\frac{g}{l}x_1 + u \end{Bmatrix} \equiv \begin{bmatrix} 0 & 1 \\ -\frac{g}{l} & 0 \end{bmatrix} \begin{Bmatrix} x_1 \\ x_2 \end{Bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} u$$

what just happened? By making the "small angle approximation", you inadvertently ***linearized*** a dynamic system that was originally nonlinear! In other words, you forced the dynamics ($\mathbf{f}$) into a structure that was expressible in matrix format.. Of-course, keep in mind that the approximation is valid only for small angles. The extent to which you can stretch your definition of "small" changes from one application to another.

- Linear systems are so important because "everything that is to be known" about them is known! They are relatively easy to analyze and closed-form analytical solutions invariably exist. So, they are always desirable.

- A natural question is - *what kind of systems are called linear*? We sort of answered this question through Eq.(15), saying that any dynamic system expressible in the shown matrix form is linear..

- Rigorously speaking, a system is linear if it obeys the *principle of superposition*.

  Principle of Superposition. A system S satisfies the principle of superposition if the output of S to a signal $(ax_1(t) + bx_2(t))$ is $(ay_1(t) + by_2(t))$, where, $y_1(t)$ and $y_2(t)$ are the outputs of S to signals $x_1(t)$ and $x_2(t)$ respectively. Here, a and b are arbitrary constants such that both are not simultaneously equal to zero.

  In the context of problems and equations... A problem P (or equation E) follows the principle of superposition if, given that y_1 and y_2 are solutions, $(ay_1 + by_2)$ is also a solution, where a and b are arbitrary constants, both not simultaneously equal to zero.

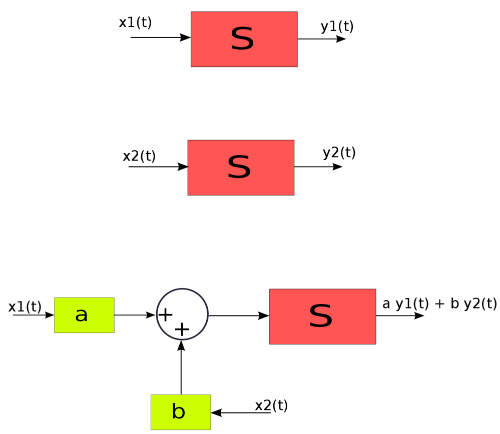


Figure 4: An illustration of the principle of superposition

Fig.(4) illustrates this idea. At the expense of devolving into a circular argument, one could say that a system is linear if the output of the system to a linear combination of two individual inputs is the linear combination of the individual outputs of the two signals.

❢ **Example** Consider the linear system again without control inputs: $\dot{\mathbf{x}} = \mathbf{A}\mathbf{x}$. We would like to confirm that this system obeys the principle of superposition. Let us assume that two signals, $\mathbf{x}_1(t)$ and $\mathbf{x}_2(t)$ solve this system. In other words,

$$\dot{\mathbf{x}}_1 = \mathbf{A}\mathbf{x}_1 \text{ and,} \tag{16a}$$

$$\dot{\mathbf{x}}_2 = \mathbf{A}\mathbf{x}_2 \tag{16b}$$

We must simply verify that $(abx_1(t) + b\mathbf{x}_2(t))$ also solves the system.. i.e.

$$\dot{\overline{a\mathbf{x}_1(t) + b\mathbf{x}_2(t)}} = \mathbf{A}(a\mathbf{x}_1(t) + b\mathbf{x}_2(t)) \tag{17}$$

Consider the LHS:

$$\dot{\overline{a\mathbf{x}_1(t) + b\mathbf{x}_2(t)}} = a\dot{\mathbf{x}}_1(t) + b\dot{\mathbf{x}}_2(t) \tag{18a}$$

$$\underset{\text{Eq.(16)}}{=} a\mathbf{A}\mathbf{x}_1(t) + b\mathbf{A}\mathbf{x}_2(t) \tag{18b}$$

$$= \mathbf{A}(a\mathbf{x}_1(t) + b\mathbf{x}_2(t)) = \text{RHS of Eq.(20)!} \tag{18c}$$

Note that Eq.(18a) exploits the fact that "differentiation" is a linear operation, i.e. it satisfies the principle of superposition, allowing us to "open the bracket".. The rest is just a rearrangement of symbols!

❢ **Example** Consider the following dynamic system

$$\dot{x} = k\sin x \tag{19}$$

We use the principle of superposition to show that this system is nonlinear. Let $x_1(t)$ and $x_2(t)$ solve this equation, i.e.

$$\dot{x}_1 = k\sin x_1 \text{ and,} \tag{20a}$$
$$\dot{x}_2 = k\sin x_2 \tag{20b}$$

We would like to check the following: For arbitrary constants a, b,

$$\text{Is } \dot{\overline{ax_1 + bx_2}} = k\sin(ax_1 + bx_2) \text{ ?} \tag{21}$$

Follow the same steps as the previous example:

$$\text{LHS: } \dot{\overline{ax_1 + bx_2}} = a\dot{x}_1 + b\dot{x}_2 \tag{22a}$$
$$\underset{\text{Eq.(19)}}{=} ak\sin x_1 + bk\sin x_2 = k(a\sin x_1 + b\sin x_2) \tag{22b}$$
$$\neq k\sin(ax_1 + bx_2) \tag{22c}$$

unless $a = b = 0$!

- The system dynamics $\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{u}$ is linear in the state ($\mathbf{x}$) **and** the control ($\mathbf{u}$)..
- In general, nonlinear systems are too difficult to deal with and it is impossible to given any guarantees. Therefore, we like to linearize them and exploit the theory of linear systems. However, this process involves simplifying assumptions (discussed below) which limits the domain of validity of linearized analysis. In the next section, we consider this supremely important engineering tool: linearization.

Linearization

- **Classification of dynamic models.** The most general ODE model for a physical process is given in state-space form in Eq.(8), in which the dynamics (RHS) is in general a function of time and state (n) and control (m) variables, and nonlinear in nature. A dynamic system is called
  - *Autonomous* if the dynamics is not an explicit function of time. In other words, $\mathbf{f}(t, \mathbf{x}, \mathbf{u}) = \mathbf{f}(\mathbf{x}, \mathbf{u})$. These are also called *time invariant* or TI.
  - *Non autonomous* or *time varying* (TV) if the dynamics is an explicit function of time; e.g. $\dot{x} = k\sin x + a\sin t$.

  Clearly, a dynamic system is **L**inear **T**ime **I**nvariant (LTI) if it is described by the following equation: $\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{u}$, where $\mathbf{A}$ and $\mathbf{B}$ are **constant matrices**.

  Similarly, a dynamic system is **L**inear **T**ime **V**arying (LTV) if it is given by: $\dot{\mathbf{x}} = \mathbf{A}(t)\mathbf{x} + \mathbf{B}(t)\mathbf{u}$, where $\mathbf{A}$ and $\mathbf{B}$ are clearly functions of time.
  For the most part in this course, we will be concerned with LTI systems.
- As previously mentioned, most physical systems (natural or man made) are nonlinear. Therefore, to exploit the many known useful results of linear systems theory, we must first linearize dynamic systems models.

- To begin talking about linearization, let us start with a scalar ($n = 1$) autonomous nonlinear system with no control ($m = 0$):

$$\dot{x} = f(x) \tag{23}$$

It is interesting to note that the LHS is simply a differentiation term, which is already a linear operation! ($\frac{ax_1+bx_2}{dt} = a\frac{dx_1}{dt} + b\frac{dx_2}{dt}$). So, we must only linearize the RHS, which is simply a function of x! It is therefore natural that in order to begin the discussion on linearization of dynamic systems, we first talk about linearization of nonlinear functions!

Linearization of functions

- Following the discussion above, consider a scalar function:

$$y = f(x) \tag{24}$$

Consider any two arbitrary points x_1 and x_2 and two arbitrary constants a and b. If $y_1 = f(x_1)$ and $y_2 = f(x_2)$, then the function obeys the principle of superposition (and is therefore linear) **if**

$$f(ax_1 + bx_2) = af(x_1) + bf(x_2) = ay_1 + by_2 \quad \text{(principle of superposition)} \tag{25}$$

❣ **Example** Is $y = 5x$ linear?
We simply need to check the principle of superposition.. let $y_1 = f(x_1) = 5x_1$ and $y_2 = f(x_2) = 5x_2$. For arbitrary constants a and b, we need to verify whether $f(ax_1 + bx_2) = af(x_1) + bf(x_2)$.

We have,

$$\begin{aligned} f(ax_1 + bx_2) &= 5(ax_1 + bx_2) \\ &= 5ax_1 + 5bx_2 \\ &= a(5x_1) + b(5x_2) \\ &= af(x_1) + bf(x_2) \quad \checkmark \end{aligned}$$

❣ **Example** Is $y = x^2$ linear?
Again, check the principle of superposition.. let $y_1 = f(x_1) = x_1^2$ and $y_2 = f(x_2) = x_2^2$. For arbitrary constants a and b, we need to verify whether $f(ax_1 + bx_2) = af(x_1) + bf(x_2)$:

$$\begin{aligned} f(ax_1 + bx_2) &= (ax_1 + bx_2)^2 \\ &= a^2x_1^2 + b^2x_2^2 + 2abx_1x_2 \\ &\neq ax_1^2 + bx_2^2 \quad \text{unless, a = b = 0} \end{aligned}$$

so, $y = x^2$ is not linear (something you already knew ☺).

- So, our objective here is to develop linear approximations for nonlinear functions..:

$$\text{Original function: } y = f(x)$$
$$\text{Linear approximation: } \hat{y}_L = \hat{f}_L(x)$$

where, the "hat" ($\hat{\cdot}$) denotes "approximation" and the subscript L represents the fact that the approximation is "linear".

⊛ Steps for linearization.

1. **Choose a reference, x^*.** The reference point is in some sense, the "center of linearization". Let us consider an example:
$$y = \frac{\mu}{x - c} \tag{26}$$
This function is shown in Fig.(5), with $\mu = 1/4$ and $c = -2$. Note that the domain of interest in this figure is $x \in [-1, 1]$. So, in the absence of any other information, it makes sense to pick $x^* = 0$ as the reference.

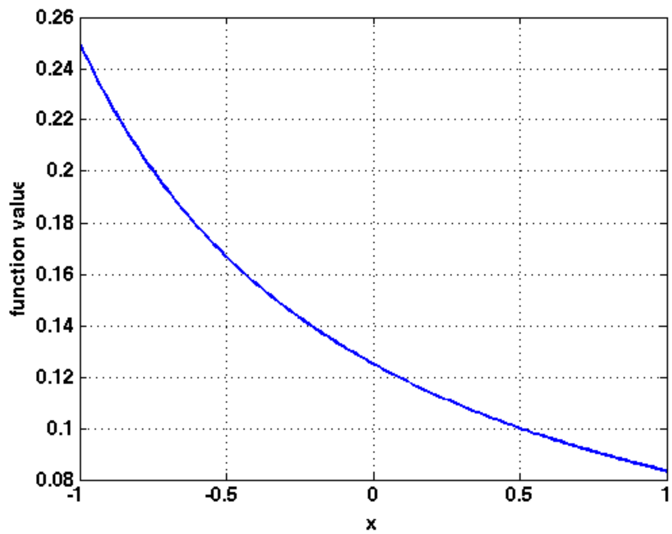


Figure 5: Nonlinear function: $y = \frac{\mu}{x-c}$

Then, $y^* = f(x^*) = -\mu/c = 1/8$.

2. **Define perturbation, δ.** Given the reference x^*, we will express a general point on the x-axis, x as a perturbation over the reference, i.e.
$$x = x^* + \delta \tag{27}$$
such that, $\delta = x - x^*$ and indeed
$$f(x) = f(x^* + \delta) \tag{28}$$
Since in the above example $x^* = 0$, we get $x = x^* + \delta = \delta$.

3. **Apply Taylor's series.** The final step is to employ the Taylor series, and expand the nonlinear function f about the reference, x^*, with respect to the perturbation, δ:
$$f(x) = f(x^* + \delta) = f(x^*) + \delta \left.\frac{\partial f}{\partial x}\right|_{x^*} + \frac{\delta^2}{2!} \left.\frac{\partial^2 f}{\partial x^2}\right|_{x^*} + \frac{\delta^3}{3!} \left.\frac{\partial^3 f}{\partial x^3}\right|_{x^*} + \ldots \tag{29}$$
The above equation is exact (i.e. the RHS exactly equals the LHS) only if we retain the expansion to an infinite number of terms. If however you decide to truncate to a finite number of terms, the "equality" (=) must be replaced with an "approximation" (≈) on account of the lost resolution due to the dropped terms. For instance, you may choose to truncate after the linear term in δ, leading to the linear perturbation approximation of f:
$$\boxed{\hat{f}_L(x) = f(x^*) + \delta \left.\frac{\partial f}{\partial x}\right|_{x^*}} \tag{30}$$

For the example at hand, we have $f'(x) = \frac{\partial f}{\partial x} = -\frac{\mu}{(x-c)^2}$, such that $f'(x^*) = f'(0) = -\mu/c^2 = -1/16$. So, the linear approximation is:

$$f(x) \approx \hat{f}_L(x) = f(0) + \delta \left.\frac{\partial f}{\partial x}\right|_0 \tag{31a}$$

$$= -\frac{\mu}{c} - \frac{\mu\delta}{c^2}$$

$$= 0.125 - 0.0625\delta \tag{31b}$$

$$= 0.125 - 0.0625(x - x^*)$$

$$= 0.125 - 0.0625x \tag{31c}$$

- An obvious question is, *when*, or *under what circumstances* is the linear approximation "good". As you can see from Eq.(29), the dropped higher order terms will be insignificant when the perturbation (δ) is small, i.e. the point x is not far from the reference, x^*!, i.e..

$$\delta \gg \delta^2; \delta^3 \ldots \tag{32}$$

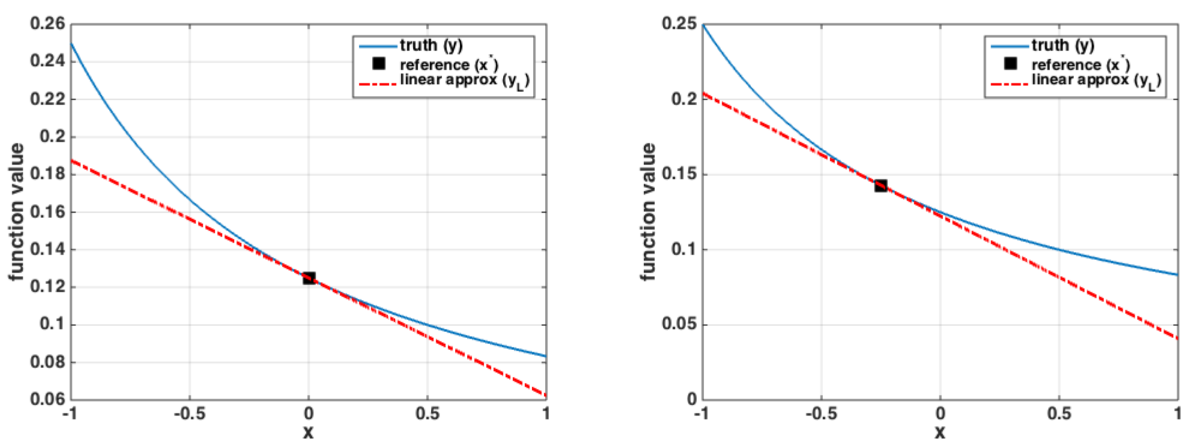
Clearly, as $\delta \uparrow$, the linear approximation generally loses accuracy because the dropped higher order terms start becoming more and more important. Indeed, we can look at the error of approximation, simply defined as

$$e_L(x) = \underbrace{f(x)}_{\text{truth}} - \underbrace{\hat{f}_L(x)}_{\text{lin. approx.}} \tag{33}$$

$$\underset{\text{Eq.30}}{=} f(x) - f(x^*) - \delta \left.\frac{\partial f}{\partial x}\right|_{x^*} \tag{34}$$

Clearly, at $x = x^*$, $\delta = 0$, and we $e_L(x^*) = 0$! I.E., the linear approximation is perfect (has zero error) at the reference point, and then (generally), the error grows as $\delta \uparrow$.

These ideas are illustrated in Fig.(6) for the example studied above. On the left, the reference point chosen is $x^* = 0$, as in the analysis above. The linear approximation is shown to be exact at the reference value. However, as x deviates from the reference, the approximation becomes increasingly bad.



(a) Reference: $x^* = 0$

(b) Reference: $x^* = -0.25$

Figure 6: Linearization of functions. Truth: $f(x) = \frac{\mu}{x-c}$

The figure on the left (Fig.(6(a)), corresponding to $x^* = 0$) is very interesting for another reason: note that the reference point is symmetric with respect to the domain of interest, i.e. for lies in the center of the domain $x \in [-1, 1]$. This is a natural choice for the reference because we do not have any other information about the nonlinear function. Also note that per intuition, error grows on either side of the reference as we move away from it. I.E., error grows in sync with the magnitude of perturbation. **However**, the growth of error is not symmetric on both sides of x^*. As you can clearly see, the error on the left end of the domain is much greater than that on the right. This implies that the dropped higher order terms have greater impact to the left of $x^* = 0$ than on the right. In other words, the "*degree of nonlinearity*" of f is greater on the left of $x^* = 0$ than on the right due to which the linearized approximation in that domain falters at a faster rate!

With this new information, it makes more sense to use a reference skewed to the left of $x = 0$! This may be asymmetric in terms of the placement of the reference point in the domain of interest, but... will hopefully *better balance the error on both sides of the reference over the entire domain.*. This is confirmed when we pick $x^* = -0.25$, shown in Fig.(6(b)). The maximum error is now about the same on both sides of the new reference point!

- Of course, as far as approximations go, there is no reason (besides the fact that linearity is always desirable) to stop at the linear term in the expansion given in Eq.(29). For example, the quadratic approximation of $f(x)$ can be written as

$$\hat{f}_Q(x) = f(x^*) + \delta \left.\frac{\partial f}{\partial x}\right|_{x^*} + \frac{\delta^2}{2} \left.\frac{\partial f}{\partial x}\right|_{x^*} \tag{35}$$

Or,

$$\hat{f}_Q(x) \underset{\text{Eq.(30)}}{=} \hat{f}_L(x) + \underbrace{\frac{\delta^2}{2} \left.\frac{\partial f}{\partial x}\right|_{x^*}}_{\text{quadractic correction over lin. approx.}} \tag{36}$$

In similar fashion, the cubic, quartic, ... n–ic approximations of the function $f(x)$ can be written! The quadratic approximation for the example considered above is shown for $x^* = 0$ in Fig.(7), which is decidedly better than the linear approximation in the sense that it has lower error over the entire domain of interest.

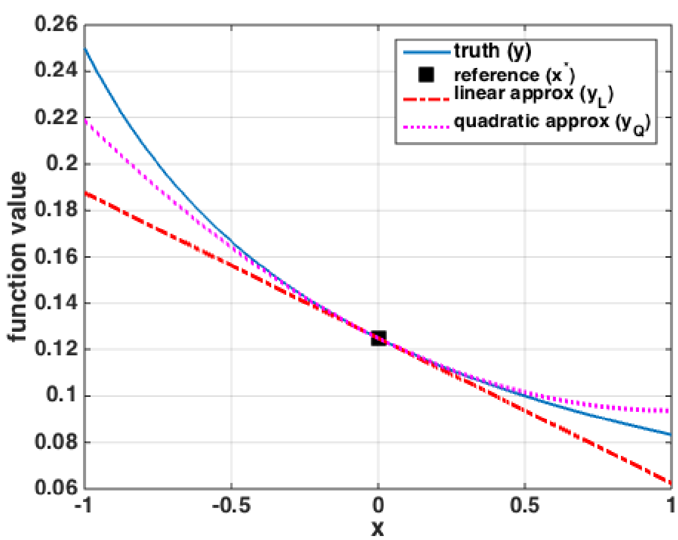


Figure 7: Comparison between the true nonlinear function: $y = \frac{\mu}{x-c}$, and its linear and quadratic approximations ($\hat{f}_L(x)$ and $\hat{f}_Q(x)$ respectively); with reference at $x^* = 0$

- **Linearization of multivariable functions.** In the above, we considered functions of a single variable (x). Since we are building up towards linearization of dynamic systems, in which the "dynamics" is invariable a function of many variables (states, controls..), we must look at the multivariable case:

$$y = f(\mathbf{x}) = f(x_1, x_2, \ldots, x_n) \tag{37}$$

where, f is a scalar nonlinear function of the vector variable $\mathbf{x} = [x_1, x_2, \ldots, x_n]^{\mathrm{T}}$. The linearization procedure is exactly the same, and begins by choosing a reference, which is a vector variable in this case: $\mathbf{x}^* = [x_1^*, x_2^*, \ldots, x_n^*]^{\mathrm{T}}$. Then, we define a vector valued perturbation as:

$$\boldsymbol{\delta} = \mathbf{x} - \mathbf{x}^* \tag{38}$$

or, written in expanded form:

$$\boldsymbol{\delta} = \begin{Bmatrix} \delta_1 \\ \delta_2 \\ \vdots \\ \delta_n \end{Bmatrix} = \mathbf{x} - \mathbf{x}^* = \begin{Bmatrix} x_1 - x_1^* \\ x_2 - x_2^* \\ \vdots \\ x_n - x_n^* \end{Bmatrix} \tag{39}$$

which essentially comprises of a perturbation term for each input variable of the function, $\delta_k = x_k - x_k^*$, $k = 1, \ldots n$. The final step is to perform the multivariable version of the Taylor series, followed by truncation to retain only the linear term, resulting in the following linear approximation:

$$\boxed{\hat{f}_L(\mathbf{x}) = f(\mathbf{x}^*) + \left[\delta_1 \left.\frac{\partial f}{\partial x_1}\right|_{\mathbf{x}^*} + \delta_2 \left.\frac{\partial f}{\partial x_2}\right|_{\mathbf{x}^*} + \ldots + \delta_n \left.\frac{\partial f}{\partial x_n}\right|_{\mathbf{x}^*}\right]} \tag{40}$$

Essentially, the second term on the RHS contains the linear corrections corresponding to each input variable, x_1, x_2, …, x_n. This can be written in compact form as:

$$\hat{f}_L(\mathbf{x}) = f(\mathbf{x}^*) + \boldsymbol{\delta}^{\mathrm{T}}\mathbf{G}(\mathbf{x}^*) \tag{41}$$

where, $\boldsymbol{\delta}^{\mathrm{T}}$ is the transpose of the perturbation vector: $\boldsymbol{\delta}^{\mathrm{T}} = [\delta_1 \;\; \delta_2 \;\; \ldots \;\; \delta_n]$, and define $\mathbf{G} = \frac{\partial f}{\partial \mathbf{x}}$ as the gradient of the function f:

$$\mathbf{G}(\mathbf{x}) = \frac{\partial f}{\partial \mathbf{x}} \triangleq \begin{Bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{Bmatrix} \tag{42}$$

such that

$$\begin{aligned} \hat{f}_L(\mathbf{x}) &= f(\mathbf{x}^*) + \boldsymbol{\delta}^{\mathrm{T}}\mathbf{G}(\mathbf{x}^*) \\ &= f(\mathbf{x}^*) + \begin{bmatrix} \delta_1 & \delta_2 & \cdots & \delta_n \end{bmatrix} \begin{bmatrix} \frac{\partial f}{\partial x_1}(\mathbf{x}^*) \\ \frac{\partial f}{\partial x_2}(\mathbf{x}^*) \\ \vdots \\ \frac{\partial f}{\partial x_n}(\mathbf{x}^*) \end{bmatrix} \\ &= f(\mathbf{x}^*) + \left[\delta_1 \frac{\partial f}{\partial x_1}(\mathbf{x}^*) + \delta_2 \frac{\partial f}{\partial x_2}(\mathbf{x}^*) + \ldots + \delta_n \frac{\partial f}{\partial x_n}(\mathbf{x}^*)\right] \end{aligned}$$

❢ **Example** Let us linearize the bivariate function $f(\mathbf{x}) = x_1^2 + x_2^2$ about the reference $\mathbf{x}^* = [1, 1]^{\mathrm{T}}$. We have, $f(\mathbf{x}^*) = 2$. Also, the gradient of f evaluated at the reference is:

$$
\begin{aligned}
\mathbf{G}(\mathbf{x}^*) &= \left.\frac{\partial f}{\partial \mathbf{x}}\right|_{\mathbf{x}^*} \triangleq \left.\begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \\ \frac{\partial f}{\partial x_2} \end{bmatrix}\right/_{\mathbf{x}^*} \\
&= \left.\begin{bmatrix} 2x_1 \\ \\ 2x_2 \end{bmatrix}\right/_{[1,1]} \\
&= \begin{bmatrix} 2 \\ \\ 2 \end{bmatrix}
\end{aligned}
$$

So, the linear approximation is:

$$
\begin{aligned}
\hat{f}_L(\mathbf{x}) &= f(\mathbf{x}^*) + \boldsymbol{\delta}^{\mathrm{T}} \mathbf{G}(\mathbf{x}^*) \\
&= 2 + [\delta_1 \ \ \delta_2] \begin{bmatrix} 2 \\ 2 \end{bmatrix} = 2 + 2\delta_1 + 2\delta_2 \qquad (43)
\end{aligned}
$$

Linearization of Dynamic Systems

The linearization of dynamic systems follows a similar route as the one described above, with the three steps of choosing a reference, defining perturbation and employing Taylor series exapnsion.

- Consider first a scalar (nonlinear) dynamic system with no control:

$$\dot{x} = f(x) \qquad (44)$$

Step 1. **Reference.** We need a reference, which is not a 'point' anymore, but an entire trajectory, written as $x^*(t)$. Now unlike linearization of functions, in which *any* point within the domain of interest is a valid reference point, we need to be more careful here. Note that we have a constraining system equation, given by Eq.(44), which must always be satisfied. So, *a reference trajectory of a dynamic system is any trajectory that satisfies its governing equation*; i.e. a trajectory $x^*(t)$ is a valid reference if

$$\dot{x}^*(t) = f(x^*(t)) \qquad (45)$$

An example is shown in Fig.(8). As you can clearly see, this trajectory is time varying in general. We will study perturbations around this reference path!

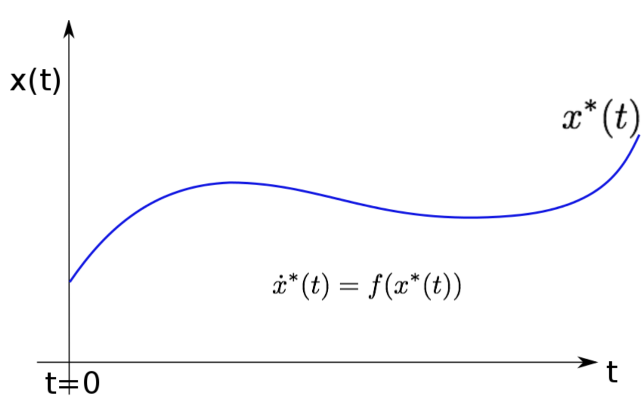


Figure 8: A reference trajectory of a dynamic system

A special reference trajectory: *steady-state reference.* A special case arises when we decide to choose a reference that *does not change with time.* In the language of dynamic systems, such references are called steady states (ss), denoted by $x^*_{\text{ss}}(t)$, and as you would suspect, they are such that

$$\dot{x}^*_{\text{ss}}(t) = 0 \tag{46}$$

This is pictorially depicted in Fig.(9). Recall that no matter what the reference, it must always satisfy the system equation. So we must have, combining Eqs.(46) and (44),

$$\dot{x}^*_{\text{ss}}(t) = \underbrace{f(x^*_{\text{ss}}(t)) = 0}_{\text{solve this!!}} \tag{47}$$

So, as you can imagine, a steady-state reference may, <u>or may not</u> exist, given the nature of the dynamic system. The above equation clearly tells us that <u>a steady-state reference is a valid reference</u> if there exists $x^*_{\text{ss}}(t)$ such that $f(x^*_{\text{ss}}(t)) = 0$. Since $x^*_{\text{ss}}(t)$ does not change with time, we can simply write $f(x^*_{\text{ss}}) = 0$.

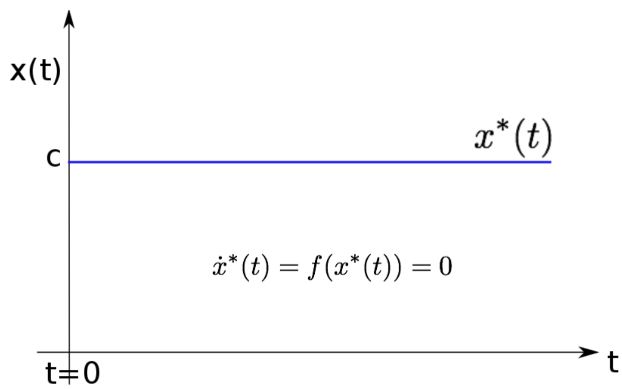


Figure 9: A *steady state* reference trajectory of a dynamic system: $x^*(t) = c =$ constant

At a steady-state (sometime also called a fixed point), the system comes to a halt because it stops changing with time.

❣ **Example** Consider a two-state dynamic system (the Duffing oscillator):

$$\dot{x}_1 = f_1(x_1, x_2) = x_2 \tag{48a}$$
$$\dot{x}_2 = f_2(x_1, x_2) = -ax_2 - bx_1 - \epsilon x_1^3 \tag{48b}$$

We are interested in finding out if this system has a steady state, and if yes, what is it? In steady state, all states must stop evolving, giving us that $f_1(x_1^*, x_2^*) = 0 = f_2(x_1^*, x_2^*)$. so we have, from Eq.(48),

$$x_2^* = 0 \tag{49a}$$
$$-ax_2^* - bx_1^* - \epsilon {x_1^*}^3 = 0 \tag{49b}$$

which solves out to give: $x_2^* = 0$ and $x_1^* = 0, \pm\sqrt{-b/\epsilon}$. So, the Duffing oscillator admits three fixed points:

$$(x_1^*, x_2^*) = (0,\ 0);\ \underbrace{\left(\sqrt{-b/\epsilon},\ 0\right);\ \left(-\sqrt{-b/\epsilon},\ 0\right)}_{\text{exist if sign}(b) = -\text{sign}(\epsilon)} \tag{49c}$$

Note the condition for existence of three steady states: $\text{sign}(b) = -\text{sign}(\epsilon)$. If not, only one steady state exists, namely $(0, 0)$, and the other two are complex (imaginary).

Step 2. **Perturbation.** Now that we have a reference, the next step is to express the current state ($x(t)$) as a perturbation over the reference, such that

$$x(t) = x^*(t) + \delta(t) \tag{50}$$

or,

$$\delta(t) = x(t) - x^*(t) \tag{51}$$

Note that the perturbation is a function of time, and measures the *current deviation of the state from the reference trajectory.* This is illustrated in Fig.(10).

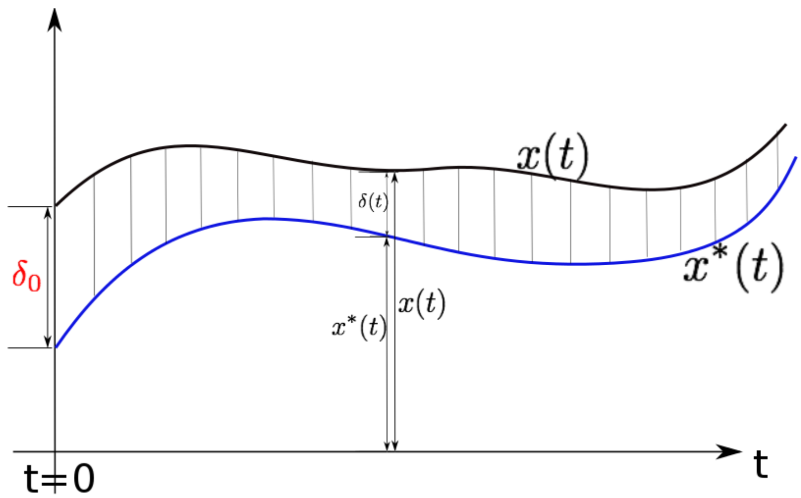


Figure 10: Perturbation over a reference trajectory: $\delta(t) = x(t) - x^*(t)$

Clearly, the analysis begins at $t = 0$, when the system receives an initial "kick" that throws it off its reference path. The initial perturbation is $\delta(t = 0) = \delta_0$, shown is red, such that $x(0) = x^*(0) + \delta_0$. From that point onward, the perturbation evolves, which is considered next.

Step 3. **(linearized) Evolution of perturbation.** We have:

$$\dot{\delta}(t) \underset{\text{Eq.51}}{=} \dot{x}(t) - \dot{x}^*(t) \tag{52a}$$

$$\underset{\text{Eq.44,45}}{=} f(x(t)) - f(x^*(t)) \tag{52b}$$

$$\underset{\text{Eq.50}}{=} f(x^*(t) + \delta(t)) - f(x^*(t)) \tag{52c}$$

As you can see, now we are in familiar territory as the first term in the RHS of Eq.(52c) looks eerily similar to Eq.(29). So, simply expand it using Taylor series about the current reference, $x^*(t)$ with respect to the current perturbation, $\delta(t)$, giving us:

$$\dot{\delta}(t) = \underbrace{\left[\cancel{f(x^*(t))} + \delta(t) \left.\frac{\partial f}{\partial x}\right|_{x^*(t)} + \underbrace{\frac{\delta(t)^2}{2} \left.\frac{\partial^2 f}{\partial x^2}\right|_{x^*(t)} + \dots}_{\text{higher order terms: ignore..}}\right]}_{=f(x^*(t)+\delta(t))\text{expanded in Taylor series!}} - \cancel{f(x^*(t))} \tag{53}$$

Giving us the following *linearized evolution of the perturbation* (after dropping the higher order terms in $\delta(t)$ as shown above:)

$$\boxed{\dot{\delta}(t) = a\delta(t)} \tag{54}$$

where, a is the **local** (i.e. at the current time) gradient of $f(x)$ evaluated at $x^*(t)$:

$$a = \left.\frac{\partial f}{\partial x}\right|_{x^*(t)} \tag{55}$$

It is very important to note that the variable a is not a constant - its instantaneous value depends on the location on the reference trajectory. We can re-write Eq.(54) explicitly as

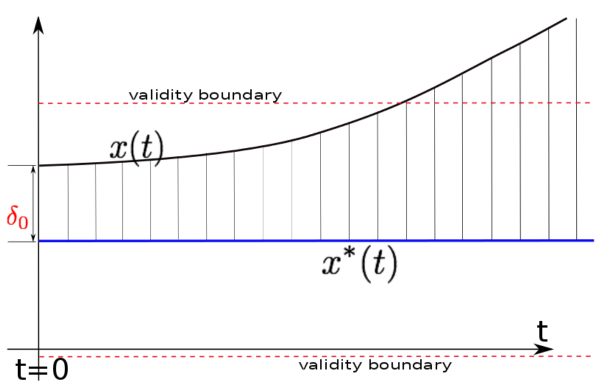
$$\dot{\delta}(t) = a(x^*(t))\,\delta(t) \tag{56}$$

In the special case of a steady state reference, things are significantly easier - this is because the reference is time-invariant, whereby, the gradient is a constant! : $a = a(x^*_{ss}(t)) = a(c)$, where $x^*_{ss}(t) = c$. Therefore, for the special case of a steady state reference, the evolution of the perturbation is LTI (see page 7)! Infact, $\dot{\delta} = a\delta$ easily integrates out to give:

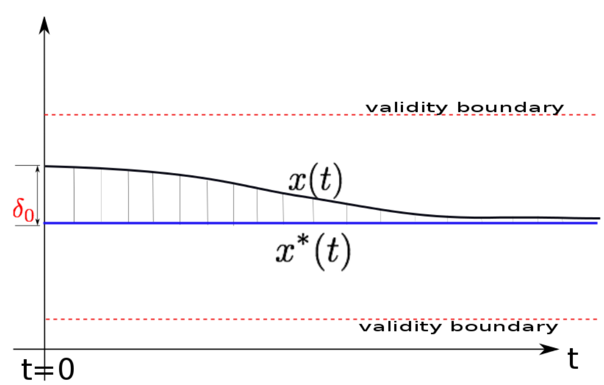
$$\delta(t) = \delta_0 e^{at} \tag{57}$$

The above equation holds clues about the stability of the reference trajectory! It is clear that if $a > 0$, i.e. the gradient of the dynamics evaluated at the fixed-point is positive, the perturbation explodes! I.E., after its initial disturbance from the steady-state, the system continues to move away and never returns to the reference... thus, the reference is **unstable**. (Fig.(11(a)))
On the other hand, if $a < 0$, the initial perturbation dies down to zero, such that the system returns to its steady-state position, whereby we conclude that the reference is **stable**. The case of $a = 0$ is called neutral stability and linear analysis is inconclusive in this situation. (Fig.(11(b)))

$$\delta(t) = e^{at} \;::\; \begin{cases} a > 0 \;\Rightarrow\; \lim_{t\to\infty}\delta(t) = \infty \;\Rightarrow\; \lim_{t\to\infty} x(t) \underset{\text{Eq.50}}{=} \infty \\ a < 0 \;\Rightarrow\; \lim_{t\to\infty}\delta(t) = 0 \;\Rightarrow\; \lim_{t\to\infty} x(t) \underset{\text{Eq.50}}{=} x^*_{ss} \\ a = 0 \;\; : \text{inconclusive} \end{cases} \tag{58}$$



(a) Unstable steady-state reference: $a > 0$



(b) Stable steady-state reference: $a < 0$

Figure 11: Perturbation evolution for stable and unstable steady-state reference trajectories.

⊛ **Some notes:**

- There is no control at this point! The behaviors described above are the system's natural responses to perturbations in the vicinity of its steady-state reference trajectories.
- "Stability" has many notions - the one developed above is the most natural in the sense that it describes whether a system returns to (stable), or continues to move away from (unstable) a steady-state reference following an initial perturbation. Therefore, stability is a property of reference trajectories, not the "system". The same system can be stable around reference trajectory $x^*_A(t)$, while unstable around reference trajectory $x^*_B(t)$.

- Recall that linearization is only valid in a relatively small domain! We are only talking about small perturbations. Both figures above show the boundary beyond which linearization is not a valid approximation, i.e. higher order terms have a dominant role to play. This is not a concern for the stable reference, in which an initial small perturbation stays small, in fact, gets smaller. However, in the unstable case, the as the perturbation starts growing in magnitude, it breaks the domain of validity after the point. Beyond this time, linear analysis cannot be trusted! Indeed, there can be "stabilizing" nonlinear (higher order) terms that may cause the system to return towards the steady-state..
- Following up on the previous point, linear analysis only tells us the **initial tendency** of the system in response to a perturbation around a reference trajectory. This is called *static stability* in the flight mechanics community, and is based solely on the sign of the gradient (term a above). We try to answer the following question:: following a perturbation around a given reference trajectory, is the system's natural first response to return to the reference or move away from it?
  For long-term analysis (*dynamic stability*), further work is usually needed.

- Recall that one of the distinguishing features of feedback control was to fundamentally alter the behavior of dynamic systems. Suppose a system is naturally unstable near the reference $x^*_{\text{ss}}(t)$, such that in terms of the above developments, $a(x^*_{\text{ss}}) > 0$. Suppose we measure the actual system response and use it in feedback fashion such that

$$u = k\delta = k(x - x^*_{\text{ss}}) \tag{59}$$

  So, we have the "closed-loop" perturbation dynamics as:

$$\dot{\delta} = a\delta + k\delta = (a+k)\delta \tag{60}$$

  where, k is the control-gain that the user must pick. Based on our discussion above, to make sure that the system exhibits stable response at x^*_{ss}, all we need to do is ensure that $(a+k) < 0$, or, $k < -a$! You have now stabilized an otherwise unstable steady-state reference ☺

- **Extension to $n > 1$ (many states)**. Most systems have more than one state. We must therefore extend our discussion on linearization to multi-state systems. We have,

$$\dot{\mathbf{x}} = \begin{Bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \vdots \\ \dot{x}_n \end{Bmatrix} = \mathbf{f}(\mathbf{x}) = \begin{Bmatrix} f_1(\mathbf{x}) \\ f_2(\mathbf{x}) \\ \vdots \\ f_n(\mathbf{x}) \end{Bmatrix} \tag{61}$$

  Note that the above is a system of n nonlinear ODEs, such that in the RHS, each function f_k, is a multivariable function, $f_k \equiv f_k(\mathbf{x}) = f_k(x_1, x_2, \ldots, x_n)$, $k = 1, \ldots, n$. The steps for linearization are the same:

Step 1. **Reference.** Obtain a reference, $\mathbf{x}^*(t)$. A special case is the steady state reference which is time invariant, i.e.

$$\mathbf{x}^*_{\text{ss}}(t) = \mathbf{c}, \quad \text{or,} \quad \begin{bmatrix} x^*_{1,\text{ss}} \\ x^*_{2,\text{ss}} \\ \vdots \\ x^*_{n,\text{ss}} \end{bmatrix} = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix} \tag{62}$$

Step 2. **Perturbation.** Define the disturbance from the reference such that

$$\boldsymbol{\delta}(t) = \mathbf{x}(t) - \mathbf{x}^*(t); \quad \text{or,} \quad \begin{bmatrix} \delta_1 \\ \delta_2 \\ \vdots \\ \delta_n \end{bmatrix} = \begin{bmatrix} x_1 - x_1^* \\ x_2 - x_2^* \\ \vdots \\ x_n - x_n^* \end{bmatrix} \tag{63}$$

So, for the perturbation in each state, we have an evolution equation:

$$\dot{\delta}_1 = f_1(\mathbf{x}(t)) - f_1(\mathbf{x}^*(t)) = f_1(\mathbf{x}^*(t) + \boldsymbol{\delta}(t)) - f_1(\mathbf{x}^*(t)) \tag{64a}$$

$$\dot{\delta}_2 = f_2(\mathbf{x}(t)) - f_2(\mathbf{x}^*(t)) = f_2(\mathbf{x}^*(t) + \boldsymbol{\delta}(t)) - f_2(\mathbf{x}^*(t)) \tag{64b}$$

$$\vdots \tag{64c}$$

$$\dot{\delta}_n = f_n(\mathbf{x}(t)) - f_n(\mathbf{x}^*(t)) = f_n(\mathbf{x}^*(t) + \boldsymbol{\delta}(t)) - f_n(\mathbf{x}^*(t)) \tag{64d}$$

Step 3. **Taylor Series**. We are now ready to expand each multivariate function using Taylor series just as in Eq.(40):

$$\dot{\delta}_1 = \underbrace{\left[\cancel{f_1(\mathbf{x}^*(t))} + \left(\delta_1 \frac{\partial f_1}{\partial x_1}\bigg|_{\mathbf{x}^*(t)} + \delta_2 \frac{\partial f_1}{\partial x_2}\bigg|_{\mathbf{x}^*(t)} + \cdots + \delta_n \frac{\partial f_1}{\partial x_n}\bigg|_{\mathbf{x}^*(t)}\right) \underbrace{+ \cdots \mathrm{H \cdot O \cdot T} \cdots}_{\text{drop..}}\right]}_{\text{Taylor series expansion of } f_1(\mathbf{x}^*(t)+\boldsymbol{\delta}(t)) \text{ about } \mathbf{x}^*(t) \text{ w.r.t } \boldsymbol{\delta}(t)} - \cancel{f_1(\mathbf{x}^*(t))} \tag{65a}$$

$$\dot{\delta}_2 = \left[\cancel{f_1(\mathbf{x}^*(t))} + \left(\delta_1 \frac{\partial f_2}{\partial x_1}\bigg|_{\mathbf{x}^*(t)} + \delta_2 \frac{\partial f_2}{\partial x_2}\bigg|_{\mathbf{x}^*(t)} + \cdots + \delta_n \frac{\partial f_2}{\partial x_n}\bigg|_{\mathbf{x}^*(t)}\right) + \cdots \mathrm{H \cdot O \cdot T} \cdots\right] - \cancel{f_2(\mathbf{x}^*(t))} \tag{65b}$$

$$\vdots$$

$$\dot{\delta}_n = \left[\cancel{f_n(\mathbf{x}^*(t))} + \left(\delta_1 \frac{\partial f_n}{\partial x_1}\bigg|_{\mathbf{x}^*(t)} + \delta_2 \frac{\partial f_n}{\partial x_2}\bigg|_{\mathbf{x}^*(t)} + \cdots + \delta_n \frac{\partial f_n}{\partial x_n}\bigg|_{\mathbf{x}^*(t)}\right) + \cdots \mathrm{H \cdot O \cdot T} \cdots\right] - \cancel{f_n(\mathbf{x}^*(t))} \tag{65c}$$

Leaving behind,

$$\dot{\delta}_1 = \begin{bmatrix} \left.\frac{\partial f_1}{\partial x_1}\right|_{\mathbf{x}^*(t)} & \left.\frac{\partial f_1}{\partial x_2}\right|_{\mathbf{x}^*(t)} & \cdots & \left.\frac{\partial f_1}{\partial x_n}\right|_{\mathbf{x}^*(t)} \end{bmatrix} \times \begin{bmatrix} \delta_1 \\ \delta_2 \\ \cdots \\ \delta_n \end{bmatrix} \tag{66a}$$

$$\dot{\delta}_2 = \begin{bmatrix} \left.\frac{\partial f_2}{\partial x_1}\right|_{\mathbf{x}^*(t)} & \left.\frac{\partial f_2}{\partial x_2}\right|_{\mathbf{x}^*(t)} & \cdots & \left.\frac{\partial f_2}{\partial x_n}\right|_{\mathbf{x}^*(t)} \end{bmatrix} \times \begin{bmatrix} \delta_1 \\ \delta_2 \\ \cdots \\ \delta_n \end{bmatrix} \tag{66b}$$

$$\vdots$$

$$\dot{\delta}_n = \begin{bmatrix} \left.\frac{\partial f_n}{\partial x_1}\right|_{\mathbf{x}^*(t)} & \left.\frac{\partial f_n}{\partial x_2}\right|_{\mathbf{x}^*(t)} & \cdots & \left.\frac{\partial f_n}{\partial x_n}\right|_{\mathbf{x}^*(t)} \end{bmatrix} \times \begin{bmatrix} \delta_1 \\ \delta_2 \\ \cdots \\ \delta_n \end{bmatrix} \tag{66c}$$

Collecting all equations together we get

$$\dot{\boldsymbol{\delta}} = \underbrace{\begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \cdots & \frac{\partial f_1}{\partial x_n} \\ \frac{\partial f_2}{\partial x_1} & \cdots & \cdots & \cdots & \frac{\partial f_2}{\partial x_n} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \frac{\partial f_n}{\partial x_1} & \frac{\partial f_n}{\partial x_2} & \cdots & \cdots & \frac{\partial f_n}{\partial x_n} \end{bmatrix}\Bigg/_{\mathbf{x}^*(t)}}_{\mathbf{A}:\ \text{Jacobian matrix}} \begin{bmatrix} \delta_1 \\ \delta_2 \\ \cdots \\ \delta_n \end{bmatrix} \tag{67}$$

Or, in short,

$$\boxed{\dot{\boldsymbol{\delta}} = \mathbf{A}\,\boldsymbol{\delta}(t)} \tag{68}$$

As mentioned above, **A** is the Jacobian matrix, which is the gradient of the dynamics evaluated along the reference trajectory:

$$\mathbf{A} = \left.\frac{\partial \mathbf{f}}{\partial \mathbf{x}}\right|_{\mathbf{x}^*(t)} \tag{69}$$

As for the scalar case, the Jacobian is a constant only when the reference trajectory is a steady-state, $\mathbf{x}^*(t) = \mathbf{x}^*_{\text{ss}}(t)$.

❣ **Example** Consider the simple pendulum again:

$$\dot{x}_1 = x_2 \tag{70a}$$

$$\dot{x}_2 = -k \sin x_1 \tag{70b}$$

Let us use the steady-state reference, which are easy to find from the above equations: $\mathbf{x}^*_{ss} = (0,0)$, $(\pm\pi, 0)$, $(\pm 2\pi, 0), \ldots$. Mathematically, there are an infinite number of fixed points of this system! (although physically there are only two – think about it). Let us determine the system's linearized response about the steady-state reference $\mathbf{x}^*_{ss}(t) = (0,0)$.

We have the perturbations:

$$\boldsymbol{\delta}(t) = \begin{Bmatrix} \delta_1(t) \\ \delta_2(t) \end{Bmatrix} = \begin{Bmatrix} x_1(t) - x_1^*(t) \\ x_2(t) - x_2^*(t) \end{Bmatrix} = \begin{Bmatrix} x_1(t) \\ x_2(t) \end{Bmatrix} \tag{71}$$

And, the Jacobian:

$$\mathbf{A} = \frac{\partial \mathbf{f}}{\partial \mathbf{x}}\Big|_{\mathbf{x}^*(t)}$$

$$= \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} \end{bmatrix}\Bigg/_{(0,0)} \tag{72}$$

$$= \begin{bmatrix} 0 & 1 \\ -k\cos x_1 & 0 \end{bmatrix}\Bigg/_{(0,0)} = \begin{bmatrix} 0 & 1 \\ -k & 0 \end{bmatrix} \tag{73}$$

So, the linearized perturbation evolution is:

$$\begin{bmatrix} \dot{\delta}_1 \\ \dot{\delta}_2 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -k & 0 \end{bmatrix} \begin{bmatrix} \delta_1 \\ \delta_2 \end{bmatrix} \tag{74}$$

Note that the above state-space form is equivalent to $\ddot{\delta} + k\delta = 0$, which is the small angle representation of the simple pendulum.. or, equivalently, a spring-mass system! In this particular case ($\mathbf{x}^* = \mathbf{0}$),

$$\begin{aligned} \mathbf{x} &= \mathbf{x}^* + \boldsymbol{\delta} \\ &= \boldsymbol{\delta} \\ \text{So that, } \dot{\mathbf{x}} &= \dot{\boldsymbol{\delta}} \\ &= \mathbf{A}\boldsymbol{\delta} \\ &= \mathbf{A}\mathbf{x} \end{aligned}$$

Given $\mathbf{A}$ in Eq.(73), the above is equivalent to,

$$\ddot{x} + kx = 0$$

Now you know the complete detail behind the "small angle approximation" of the simple pendulum: it represents the linearized perturbation model of the nonlinear pendulum about the steady-state reference $\mathbf{x}^*(t) = (0,0)$. In the linear world, the simple pendulum is equivalent to a spring-mass system!

- So, what about $\dot{\mathbf{x}}$ in general?

Note that $\mathbf{x} = \boldsymbol{\delta} + \mathbf{x}^*$. If we have a steady-state reference, i.e. $\mathbf{x}^*(t) = \mathbf{x}^*_{\text{ss}}$, then $\dot{\mathbf{x}}^*(t) = \mathbf{0}$, such that

$$\dot{\mathbf{x}} = \dot{\boldsymbol{\delta}} = \mathbf{A}\left(\mathbf{x} - \mathbf{x}^*_{\text{ss}}\right) \tag{75}$$

Moreover, if $\mathbf{x}^*_{\text{ss}} = \mathbf{0}$, then

$$\boxed{\dot{\mathbf{x}} = \mathbf{A}\,\mathbf{x}} \tag{76}$$

- Here's the summary: we started from $\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x})$, and we have ended with $\dot{\mathbf{x}} = \mathbf{A}\,\mathbf{x}$, which is the linearized dynamics about the steady-state reference $\mathbf{x}^*_{\text{ss}} = \mathbf{0}$. Clearly, things are nice and simple if we concern ourselves only with the special case of $\mathbf{x}^*(t) = \mathbf{x}^*_{\text{ss}} = \mathbf{0}$. Fortunately, this is not very difficult. If *some* steady-state reference $\mathbf{x}^*_{\text{ss}} = \mathbf{c}$ exists, we can convert it to the reference $\mathbf{x}^*_{\text{ss}} = \mathbf{0}$ via a simple coordinate transformation:

$$\mathbf{x} \xleftarrow{\text{replace with}} \mathbf{x} - \mathbf{c} \tag{77}$$

- So, the bottom line is that for the most part, we will be looking at systems linearized about a steady-state reference, which through an appropriate coordinate transformation will be ensured to be at $\mathbf{0}$.