Linear Control Theory

Chandra Gummaluru
chandra-gummaluru.github.io

Department of Electrical and Computer Engineering, University of Toronto

2022

Contents

1 Continuous-Time Signals

To begin, we consider one definition of a continuous-time signal.

Definition 1.1. (Continuous-Time Signal) A **continuous-time signal** is a function, $x : \mathbb{R} \to \mathbb{R}^n$ where $x(t)$ is the signal's value at time, t.

It is sometimes convenient to view a continuous-time signal, x, as linear combinations of basis signals, $\phi_1, \phi_2, \dots$, i.e.,

$$x(t) = \sum_k \hat{x}_k \phi_k(t).$$

. This is not always the case.

Remark 1.1. The basis may either be finite or countably infinite. In any case, we require $x \in \mathrm{Span}(\phi_1, \phi_2, \dots)$.

While the basis is otherwise arbitrary, it is preferable to choose one that is orthogonal since it allows us to derive a relatively simple formula for the $\hat{x}_k$.

Definition 1.2. (Orthogonal Basis) A set of basis functions, $\{\phi_k\}$ is **orthogonal** if $\langle \phi_h, \phi_k \rangle = 0$ whenever $k \neq h$, where $\langle \cdot, \cdot \rangle$ denotes an inner-product.

In this case, $\hat{x}_k$ may be computed as

$$\frac{\langle x, \phi_k \rangle}{\langle \phi_k, \phi_k \rangle} = \frac{1}{\langle \phi_k, \phi_k \rangle} \left\langle \sum_h \hat{x}_h \phi_h, \phi_k \right\rangle = \frac{1}{\langle \phi_k, \phi_k \rangle} \sum_h \langle \hat{x}_j \phi_h, \phi_k \rangle = \hat{x}_k.$$

If the basis is **orthonormal**, i.e., orthogonal and moreover $\langle \phi_k, \phi_k \rangle = 1, \forall k$, the above reduces to $\hat{x}_k = \langle x, \phi_k \rangle$.

1.1 Transformations of Continuous-Time Signals

We consider a few important transformations of continuous-time signals.

Definition 1.3. (Continuous-Time Time Shifting) A signal, $x_\tau : \mathbb{R} \to \mathbb{R}$ is a **time-shift** of a signal, $x : \mathbb{R} \to \mathbb{R}$, if

$$x_\tau(t) = x(t - \tau), \forall t.$$

For $\tau > 0$, the time-shift is into the past, and for $\tau < 0$, it is into the future.

1.2 Continuous-Time Distributions

A consequence of the Def. 1.1 is that a continuous-time signal, x, can be measured at any time, $t \in \mathbb{R}$. However, this is often impossible in practice. Instead, we must must resort to computing a (possibly non-uniform) mean measurement of the signal within some finite region, R_t around t.

More precisely, we may only compute

$$\int_{R_t} \phi(\tau) x(\tau) d\tau,$$

where $\phi : \mathbb{R} \to \mathbb{C}$ denotes the weighting over R_t.

Thus, rather than representing continuous-time signals as functions (as we did in Def. 1.1), we will use a new construct called a distribution.

Definition 1.4. (Distribution) A **distribution**, $T : \mathcal{D} \to \mathbb{C}^n$ is a linear functional[1] where $\mathcal{D}$ is a set of test functions on which T acts. The notation $\langle T, \phi \rangle$ is used to denote the distribution, T, acting on the test function $\phi \in \mathcal{D}$.

Definition 1.5. (Equivalent Distributions) Two distributions, T_1 and T_2, are said to be equivalent, denoted $T_1 \equiv T_2$ if

$$\langle T_1, \phi \rangle = \langle T_2, \phi \rangle,$$

for any appropriate test function, ϕ.

Many conventional functions can be represented as distributions.

Definition 1.6. (Distribution of a Function) Any (locally integrable) function, x, may be represented as a distribution via the mapping

$$\langle T_x, \phi \rangle \mapsto \int_{\operatorname{dom} x} x(t)\phi(t)dt,$$

where T_x denotes the distributional analogue or **regular distribution** of x.

Remark 1.2. In a common abuse of notation, x is often also used to denote its regular distribution, T_x.

We now consider transformations equivalent to those defined in §**??**, but for distributions.

Definition 1.7. (Sum of Distributions) Let T_1 and T_2 be two distributions. Then, we define $T_1 + T_2$ as the distribution such that

$$\langle T_1 + T_2, \phi \rangle = \langle T_1, \phi \rangle + \langle T_2, \phi \rangle.$$

Definition 1.8. (Time-Shift of a Distribution) For any distribution, T, and $d \in \mathbb{R}$, we define the backward time-shift, T_d, as a distribution such that

$$\langle T_d, \phi \rangle = \langle T, \phi_{-d} \rangle.$$

Definition 1.9. (Product with a Smooth Function) For any distribution, T, and smooth function, s, we define the distributional product $T \cdot s$ so that

$$\langle T \cdot s, \phi \rangle = \langle T, s \cdot \phi \rangle.$$

Definition 1.10. (Distributional Derivative) The **distributional derivative** of T, denoted with T' is such that

$$\langle T', \phi \rangle = -\langle T, \phi' \rangle.$$

Theorem 1.1. (Regular Distributional Derivative) $T_x' = T_{x'}$ for any regular distribution, T.

Proof. We want a definition such that $T_x' = T_{x'}$.

Using integration by parts, we can show that

$$\langle T(x'), \varphi \rangle = \int_{\mathbb{R}} x'(t)\varphi(t)dt$$

[1] A function of a function.

$$= [\varphi(t)x(t)]_{-\infty}^{\infty} - \int_{\mathbb{R}} x(t)\varphi'(t)dt$$
$$= -\int_{\mathbb{R}} x(t)\varphi'(t)dt$$
$$= -\langle T(x), \varphi' \rangle.$$

Thus, we define a distributional derivative accordingly. □

Distributions can also be used to represent signals that cannot be defined as well-behaved functions.

Definition 1.11. The **unit-impulse distribution**, δ, is defined such that

$$\langle \delta, \varphi \rangle := \varphi(0),$$

meaning that δ evaluates a test function, φ at $t = 0$.

Definition 1.12. The **unit-step distribution**, u, is defined such that

$$\langle u, \varphi \rangle := \int_{\mathbb{R}_+} \varphi(t)dt,$$

that is, u evaluates the integral of a test function, ϕ on the non-negative real number line, $\mathbb{R}_+$.

Claim 1.1. The distributional derivative of the unit-step distribution is the unit-impulse distribution, i.e., $\langle u', \varphi \rangle = \langle \delta, \varphi \rangle$.

Proof. By definition, $\langle u', \varphi \rangle = -\langle u, \varphi' \rangle$

$$= -\int_{\mathbb{R}_+} \varphi'(x)dx$$
$$= \varphi(0) := \langle \delta, \varphi \rangle$$

□

Definition 1.13. The **ramp distribution**, r, is defined such that

$$\langle r, \varphi \rangle := \int_{\mathbb{R}_+} t\varphi(t)dt.$$

Claim 1.2. The distributional derivative of the ramp distribution is the unit-step distribution, i.e., $\langle r, \varphi \rangle' = \langle u, \varphi \rangle$.

Proof. By definition, $\langle R, \varphi \rangle' = -\langle R, \varphi' \rangle$

$$= -\int_{\mathbb{R}_+} x\varphi'(x)dx$$
$$= -[xp(x)]_0^{\infty} + \int_{\mathbb{R}_+} p(x)dx$$
$$= \int_{\mathbb{R}_+} p(x)dx$$
$$:= \langle u, \varphi \rangle$$

□

By considering such signals as distributions, we can still reasonably analyse their influence in relation to how they affect other (well-behaved) signals.

2 Integral Representations of Continuous-Time Signals

In some cases, it is possible to represent a signal, x, using an integral of the form

$$x(t) = \int X(s)\psi(s,t)ds,$$

where

$$X(s) = \int_{\operatorname{dom} x} x(t)\phi(t,s)dt,$$

for some functions, $X : \mathbb{R} \to \mathbb{C}$, $\psi : \mathbb{R} \times \mathbb{R} \to \mathbb{C}$, and $\phi : \mathbb{R} \to \mathbb{R} \to \mathbb{C}$.

Of course, ψ and ϕ should be chosen appropriately to ensure the above relationships hold.

2.1 The s-Transform

Definition 2.1. The s-transform of a continuous-time signal, $x : \mathbb{R} \to \mathbb{C}$ can be expressed as

$$X(s) := \int_{-\infty}^{\infty} x(t)e^{-st}dt. \tag{2.1.1}$$

We write $X(s) = \mathscr{L}\{x(t)\}(s)$.

Remark 2.1. The set of s for which Eq. 2.1.1 converges is called the **region of convergence**, typically denoted with Γ.

Theorem 2.1. (Structure of the Region of Convergence) If $\gamma \in \Gamma$ for some $\gamma \in \mathbb{R}$, then $\gamma + j\sigma \in \Gamma$ for any $\sigma \in \mathbb{R}$.

It follow that $\Gamma = \Re\{\Gamma\} \times j\mathbb{R}$.

Proof. Omitted. □

Theorem 2.2. (Inverse s-Transform) If X is the s-transform of x, then for any $\gamma \in \Re\{\Gamma\}$, we have

$$x(t) = \frac{1}{2\pi j}\int_{\gamma - j\infty}^{\gamma + j\infty} X(s)e^{st}ds. \tag{2.1.2}$$

We write $x(t) = \mathscr{L}^{-1}\{X(s)\}(t)$.

Before proving Thm. 2.2, we prove a relevant lemma: Let ϕ be a test function, i.e., ϕ is smooth and has compact support. Then, in the distributional sense,

$$\int_{-\infty}^{\infty} e^{j\omega x}d\omega \equiv 2\pi\delta(x).$$

where $x \in \mathbb{R}$ is an arbitrary constant.

Proof. Omitted. □

Given the lemma, we can now prove Thm. 2.2.

Proof. Assuming $s \in \Gamma$, we have,

$$x(t) \stackrel{?}{=} \frac{1}{2\pi j} \int_{\gamma - j\infty}^{\gamma + j\infty} X(s) e^{st} ds$$

$$= \frac{1}{2\pi j} \int_{\gamma - j\infty}^{\gamma + j\infty} \left(\int_{-\infty}^{\infty} x(\tau) e^{-s\tau} d\tau \right) e^{st} ds$$

$$= \frac{1}{2\pi j} \int_{-\infty}^{\infty} x(\tau) \int_{\gamma - j\infty}^{\gamma + j\infty} e^{s(t-\tau)} ds d\tau.$$

We now consider the substitution, $\sigma = -j(s - \gamma)$, or equivalently, $s = \gamma + j\sigma$. However, before proceeding, we make an important observation.

Since we require that $s \in \Gamma$, and $\Gamma = \Re\{\Gamma\} \times j\mathbb{R}$ from Thm. 2.1, we must also require $\gamma \in \Re\{\Gamma\}$.

Under this assumption, we have

$$\int_{\gamma - j\infty}^{\gamma + j\infty} e^{s(t-\tau)} ds = j \int_{-\infty}^{\infty} e^{(\gamma + j\sigma)(t-\tau)} d\sigma$$

$$= je^{\gamma(t-\tau)} \underbrace{\int_{-\infty}^{\infty} e^{j\sigma(t-\tau)} d\sigma}_{2\pi\delta(t-\tau)} = 2\pi j e^{\gamma(t-\tau)} \delta(t - \tau)$$

$$= -2\pi j \delta(t - \tau).$$

Back-substituting this result, we have

$$\frac{1}{2\pi j} \int_{-\infty}^{\infty} 2\pi j x(\tau) \delta(t - \tau) dt = \int_{-\infty}^{\infty} x(\tau) \delta(t - \tau) d\tau := x(t).$$

□

Remark 2.2. Eq. 2.1.2 does not uniquely define x unless Γ is specified.

In Table **??**, we present (without proof) a few basic properties of the s-transform which we will make use of repeatedly.

Table 1: s-Transform Properties

Property	$x(t)$	$X(s)$
Linearity	$ax_1(t) + bx_2(t)$	$aX_1(s) + bX_2(s)$
Time Shift	$x(t - \tau)$	$e^{-\tau\omega} X(j\omega)$
s-domain Shift	$e^{-s_0 t} x(t)$	$X(s - s_0)$
Convolution	$(x_1 * x_2)(t)$	$X_1(s) X_2(s)$
Multiplication	$X_1(t) X_2(t)$	$\frac{1}{2\pi j}(X_1 * X_2)(s)$

3 Continuous-Time Systems

We now consider continuous-time systems.

Definition 3.1. (Continuous-Time System) A **continuous-time system** is an operator, $\mathcal{T}$ that maps one continuous-time signal, $x : \mathbb{R} \to \mathbb{R}$ to another continuous-time signal, $y = \mathcal{T}\{x\}$.

3.1 Time-Invariant Systems

One property we would like our system to have is called "time-invariance".

Definition 3.2. (Time-Invariant Continuous-Time Signal) A continuous-time signal is **time-invariant** iff for any input x, if

$$y(t) = \mathcal{T}\{x(\tau)\}(t),$$

then

$$y(t - t_0) = \mathcal{T}\{x(\tau - t_0)\}(t), \forall t_0.$$

We also want our system to be "causal".

Definition 3.3. (Causal Continuous-Time System) A continuous-time system, $\mathcal{T}$ is **causal** if its response to any signal, $x : \mathbb{R} \to \mathbb{R}$, at some time, t, depends only on the value of x at or before t

For a causal system, we can express $y(t)$ using a recursive definition:

$$\begin{aligned} \dot{x}(t) &= f(x(t), u(t), t) \\ y(t) &= g(x(t), u(t), t), \end{aligned}$$

where $x : \mathbb{R}\_+ \to \mathbb{R}^n, f : \mathbb{R}^n \times \mathbb{R}^m$ and $g : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^p$. We refer to $x(t)$ as the state of the system at time, t.

For a time-invariant causal system, f and g are independent of t. Thus, if we knew x and u for all t, it would be relatively easy to compute $y(t), \forall t$. However, we often only know the initial state, $x_0 = x(0)$. To determine $x(t)$, we need to solve:

$$\begin{aligned} \dot{x} &= f(x, u) \\ x(0) &= x_0. \end{aligned}$$

This is difficult to do in general.

3.2 Linear Systems

We will focus on linear systems.

Definition 3.4. (Linear Continuous-Time System) A continuous-time system, $\mathcal{T}$, is **linear** iff for any input signals, x_1 and x_2, and scalars a_1 and a_2, we have

$$\mathcal{T}\{a_1 x_1(t) + a_2 x_2(t)\} = a_1 \mathcal{T}\{x_1(t)\} + a_2 \mathcal{T}\{x_2(t)\}.$$

For a causal linear system, we may write

$$\begin{aligned} \dot{x}(t) &= A(t)x(t) + B(t)u(t) \\ y(t) &= C(t)x(t) + D(t)u(t) \end{aligned}$$

where $A : \mathbb{R}_+ \to \mathbb{R}^{n,n}, B : \mathbb{R}_+ \to \mathbb{R}^{n,m}, C : \mathbb{R}_+ \to \mathbb{R}^{p,n}, D : \mathbb{R}_+ \to \mathbb{R}^{p,m}$.

3.3 Linear Time-Invariant Continuous-Time Systems

For a linear time-invariant (LTI) system, A, B, C, and D are constant w.r.t. t, and

$$\dot{x}(t) = Ax(t) + Bu(t)$$
$$y(t) = Cx(t) + Du(t)$$

We will primarily focus on LTI systems. LTI systems are particularly simple to study because their behaviour can be entirely characterized by their responses to a unit impulse, i.e., their impulse response.

Theorem 3.1. (Response of a Continuous-Time Linear System) Consider a continuous-time LTI system with an impulse response, $h(t) = \mathcal{T}\{\delta(t)\}$.

Then for any other continuous-time signal, $x : \mathbb{R} \to \mathbb{R}$,

$$\mathcal{T}\{x(\tau)\}(t) = (x * h)(t) := \int_{-\infty}^{\infty} x(\tau)h(t - \tau)d\tau.$$

Proof. Omitted. □

Remark 3.1. The convolution operator, $*$, satisfies the following properties:

- **Distributivity**: $x * (h_1 + h_2) = x * h_1 + x * h_2$,
- **Commutativity**: $x * h = h * x$,
- **Associativity**: $x * (h_1 * h_2) = (x * h_1) * h_2$.

As a result, LTI systems are also distributive, commutative, and associative.

Theorem 3.2. (Impulse Response of a Causal LTI System) The impulse response, h, of a causal continuous-time LTI system is always such that

$$h(t) = 0, \forall t < 0.$$

Proof. Consider the response $y : \mathbb{R} \to \mathbb{R}$ of the system to an arbitrary continuous-time signal, $x : \mathbb{R} \to \mathbb{R}$. We have

$$y(t) = (x * h)(t) = (h * x)(t)$$

$$= \int_{-\infty}^{\infty} h(\tau)x(t - \tau)d\tau$$
$$= \int_{-\infty}^{-1} h(\tau)x(t - \tau)d\tau + \int_{0}^{\infty} h(\tau)x(t - \tau)d\tau$$

In order to satisfy causality, the first sum must evaluate to zero for any x, which must mean that $h(t) = 0, \forall t < 0$. □

Theorem 3.3. (Impulse Response of a Memoryless LTI System) The impulse response, h of a memoryless continuous-time LTI system is always such that

$$h(t) = 0, \forall n \neq 0.$$

Proof. The proof is analogous to that for Thm. 3.2. □

It turns out that exponential signals, i.e., those of the form, e^{st}, where $s \in \mathbb{C}$, are eigensignals of LTI systems.

Theorem 3.4. (Response of a Continuous-Time LTI System to an Exponential) The response of a continuous-time LTI system $\mathcal{T}$ to a signal of the form, e^{st} is

$$\mathcal{T}\{e^{st}\}(t) = H(s)e^{st},$$

where $H(s) = \mathscr{L}\{h(t)\}(s)$ is called the **system function**.

Proof. Let $x(t) = e^{st}$. We have

$$y(t) = \mathcal{T}\{x(t)\}(t) = (h * x)(t)$$

$$= \int_{-\infty}^{\infty} h(\tau)x(t-\tau)d\tau$$
$$= \int_{-\infty}^{\infty} h(\tau)e^{s(t-\tau)}d\tau$$
$$= e^{st} \underbrace{\int_{-\infty}^{\infty} h(\tau)e^{-s\tau}d\tau}_{\mathscr{L}\{h\}(s)}.$$

□

Theorem 3.5. (System Function) Consider a continuous-time LTI system, $\mathcal{T}$, with a system function, H. Suppose y is the output of $\mathcal{T}$ to a signal, x, i.e.,

$$y(t) = \mathcal{T}\{x(t)\}(t).$$

Then,

$$Y(s) = H(s)X(s),$$

where $X(s) = \mathscr{L}\{x(t)\}(s)$ and $Y(s) = \mathscr{L}\{y(t)\}(s)$.

Proof. The proof immediately follows from Thm. 3.1 and the convolution property of the Laplace transform. □

3.4 Linearisation of Non-Linear Systems

Many real-world systems are non-linear. We would like to approximate such systems using linear systems. This can done via 1^{st}-order Taylor approximations of f and h around some equilibrium of the system, $(\bar{x}, \bar{u})$. It follows that

$$f(x,u) \approx f(\bar{x},\bar{u}) + \frac{d}{dx} f(x,u)\bigg|_{x=\bar{x}} (x-\bar{x}) + \frac{d}{du} f(x,u)\bigg|_{u=\bar{u}} (u-\bar{u}),$$

and

$$h(x,u) \approx h(\bar{x},\bar{u}) \; \frac{d}{dx} h(x,u)\bigg|_{x=\bar{x}} (x-\bar{x}) + \frac{d}{du} h(x,u)\bigg|_{u=\bar{u}} (u-\bar{u}),$$

where

$$\frac{df}{dx} = \begin{bmatrix} \frac{df_1}{dx_1} & \cdots & \frac{df_1}{dx_n} \\ \vdots & \ddots & \vdots \\ \frac{df_n}{dx_1} & \cdots & \frac{df_n}{dx_n} \end{bmatrix}, \frac{d}{du} f = \begin{bmatrix} \frac{df_1}{du_1} & \cdots & \frac{df_1}{du_m} \\ \vdots & \ddots & \vdots \\ \frac{df_n}{du_1} & \cdots & \frac{df_n}{du_m} \end{bmatrix}$$

$$\frac{d}{dx} h = \begin{bmatrix} \frac{dh_1}{dx_1} & \cdots & \frac{dh_1}{dx_n} \\ \vdots & \ddots & \vdots \\ \frac{dh_p}{dx_1} & \cdots & \frac{dh_p}{dx_n} \end{bmatrix}, \frac{d}{du} h = \begin{bmatrix} \frac{dh_1}{du_1} & \cdots & \frac{dh_1}{du_m} \\ \vdots & \ddots & \vdots \\ \frac{dh_p}{du_1} & \cdots & \frac{dh_p}{du_m} \end{bmatrix}$$

Since $(\bar{x}, \bar{u})$ is an equilibrium of the system, $f(\bar{x}, \bar{u}) = 0$. Letting $\tilde{x} = x - \bar{x}$, $\tilde{u} = u - \bar{u}$ and $\tilde{y} = h(x,u) - h(\bar{x}, \bar{u})$ denote the deviations of x, u, and y from their equilibrium values, we obtain

$$\begin{aligned} \dot{\tilde{x}} = \dot{x} &= A\tilde{x} + B\tilde{u} \\ \tilde{y} &= C\tilde{x} + D\tilde{u}, \end{aligned} \tag{3.4.1}$$

where

$$A = \frac{d}{dx} f(x,u)\bigg|_{x=\bar{x}}, B = \frac{d}{dx} h(x,u)\bigg|_{u=\bar{u}}, C = \frac{d}{dx} h(x,u)\bigg|_{x=\bar{x}}, D = \frac{d}{dx} h(x,u)\bigg|_{u=\bar{u}}.$$

The system described by (**??**) is the linearisation of the one described by (**??**) at the equilibrium $(\bar{x}, \bar{u})$.

4 The Solution to $\dot{x} = Ax$

Our goal now is to find a closed form expression for x that satisfies (**??**) given some initial state $x(0) = x_0 \in \mathbb{R}^n$ and input u. For now, we will assume that either $B = \mathbf{0}$ and/or $u \equiv 0$, leading to the following initial-value problem:

$$\begin{aligned} \dot{x} &= Ax \\ x(0) &= x_0. \end{aligned} \tag{4.0.1}$$

Suppose that we replace the matrix A in (**??**) with a scalar a, i.e., we consider the initial-value problem

$$\begin{aligned} \dot{x} &= ax \\ x(0) &= x_0. \end{aligned}$$

We can easily verify that $x(t) = x_0 e^{at}$ is a solution. Indeed we have $\dot{x}(t) = ax_0 e^{at} = ax(t)$ and $x(0) = x_0$.

We might expect a similar result to hold for (**??**), provided we define the exponential of a matrix such that it generalizes the scalar definition.

4.1 The Matrix Exponential

To define the exponential of a matrix, A, we turn to the series representation for the exponential of a scalar, a:

$$e^a := \sum_{k=0}^{\infty} \frac{a^k}{k!}.$$

The advantage of the series representation is that it can readily extend to the matrix case simply by replacing a with A.

Definition 4.1. (Matrix Exponential) The exponential of a matrix A is given by the infinite series,

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

where by definition, $A^0 = I$.

Later, we will show that $x(t) = e^{At}x_0$ is indeed a solution to (**??**). However, for now, we focus on showing that the matrix series in Def. **??** actually converges.

Theorem 4.1. (Convergence of the Matrix Exponential) The series in Def. **??** converges for any A in the sense that the scalar series $\sum_{k=0}^{\infty} (A_k)_{i,j}$ converges for every i and j.

Proof. Since absolute convergence implies convergence, it is sufficient to show that the scalar series,

$$\sum_{k=0}^{\infty} \left\| \frac{A^k}{k!} \right\| = \lim_{n\to\infty} \underbrace{\sum_{k=0}^{n} \left\| \frac{A^k}{k!} \right\|}_{S_n}$$

converges for any A. However, we shall simply verify that the scalar series above does indeed converge. We first verify that S_n is an increasing sequence. Indeed we have

$$S_{n+1} = \sum_{k=0}^{n+1} \left\| \frac{A^k}{k!} \right\| = \sum_{k=0}^{n} \left\| \frac{A^k}{k!} \right\| + \left\| \frac{A^{n+1}}{(n+1)!} \right\| = S_n + \underbrace{\left\| \frac{A^{n+1}}{(n+1)!} \right\|}_{\geq 0} \geq S_n.$$

Next, we show that S_n is bounded from above. Indeed, we have

$$S_n = \sum_{k=0}^{n} \left\| \frac{A^k}{k!} \right\| = \sum_{k=0}^{n} \frac{1}{k!} \left\| A^k \right\| \leq \sum_{k=0}^{\infty} \frac{1}{k!} \left\| A^k \right\| \leq \sum_{k=0}^{\infty} \frac{1}{k!} \left\| A \right\|^k := e^{\|A\|}.$$

Since S_n is an increasing sequence bounded from above by $e^{\|A\|}$, it must eventually converge. □

Most of the properties of the scalar exponential readily extend to the matrix exponential. We shall explicitly state and prove the ones we will make extensive use of.

Theorem 4.2. (Properties of the Matrix Exponential) The matrix exponential satisfies the following properties:

1. For any non-singular matrix, P, we have $e^{PAP^{-1}} = Pe^A P^{-1}$.
2. If $AB = BA$ then $e^{AB} = e^A e^B = e^B e^A$,
3. The inverse of the exponential of A is the exponential of $-A$, i.e., $e^{-A} = (e^A)^{-1}$.

Proof. Omitted. □

4.2 Computing the Matrix Exponential

Let us now discuss how to actually compute the exponential of a matrix, A. We begin by considering the trivial but instructive special case where A is diagonal, i.e., $A = \mathrm{diag}(a_1, \ldots, a_n)$. We compute

$$\exp(A) = \exp\left(\mathrm{diag}(a_1, \ldots, a_n)\right)$$

$$= \sum_{0=1}^{\infty} \frac{\mathrm{diag}(a_1, \ldots, a_n)^k}{k!}$$

$$= \mathrm{diag}\left(\sum_{k=0}^{\infty} \frac{a_1^k}{k!}, \ldots, \sum_{k=0}^{\infty} \frac{a_n^k}{k!} \right)$$

$$= \mathrm{diag}(e^{a_1}, \ldots, e^{a_n}).$$

We see that the matrix exponential of a diagonal matrix is the scalar exponential of each of its elements. We can use this result to derive a relatively simple formula for e^A even if A is not necessarily diagonal, so long as it is diagonalizable.

To begin, we derive the conditions under which A is diagonalizable.

Theorem 4.3. (Computing e^A when A is Diagonalizable) Suppose $A \in \mathbb{R}^{n,n}$ is diagonalizable, i.e., there exists a matrix, $P \in \mathbb{R}^{n,n}$, and diagonal matrix, $\Delta = \mathrm{diag}(\lambda_1, \ldots, \lambda_n) \in \mathbb{R}^{n,n}$ such that $A = P \Delta P^{-1}$.

Then, we have

$$e^A = P e^\Delta P^{-1} \text{ where } e^\Delta = \begin{bmatrix} e^{\lambda_1} & & \\ & \ddots & \\ & & e^{\lambda_n} \end{bmatrix}.$$

Proof. To begin, assume $A = P\Delta P^{-1}$. We have

$$e^A = e^{P\Delta P^{-1}} = \sum_{k=0}^{\infty} \frac{(P\Delta P^{-1})^k}{k!} = P \sum_{k=0}^{\infty} \frac{\Delta^k}{k!} P^{-1} = Pe^{\Delta P^{-1}}.$$

□

Thm. **??** makes use of the eigenvalues and eigenvectors of A, which can be complex-valued even if the elements of A are real. While the formulation is still valid in such cases, the representation can be somewhat non-intuitive, particularly since the definition of e^A clearly implies that $e^A \in \mathbb{R}^{n,n}$ if $A \in \mathbb{R}^{n,n}$. Ideally, we would avoid intermediate computations that involve complex values. We shall now derive a procedure to do so.

To begin, we recall that the eigen-values/vectors of a matrix come in complex conjugate pairs, i.e., if $\lambda, v)$ is an eigen-value/vector pair, then so too is (λ^*, v^*).

Generally, we have $P = \begin{bmatrix} \dots & v & v^* & \dots \end{bmatrix}$ and $\Delta = \operatorname{diag}(\dots, \lambda, \lambda^*, \dots)$. Assume $\lambda = a + bi$. Now define $\tilde{P}$ by taking P and replacing v with $\mathfrak{Re}(v)$ and v^* with $\mathfrak{Im}(v)$. Similarly, define $\tilde{\Delta}$ by taking Δ and replacing the 2×2 block containing λ and λ^* with

$$\begin{bmatrix} a & b \\ -b & a \end{bmatrix}.$$

It can be shown that $e^A = \tilde{P} e^{\tilde{\Delta}} \tilde{P}^{-1}$. We now show that

$$\exp\left\{ \begin{bmatrix} a & b \\ -b & a \end{bmatrix} \right\} = e^a \begin{bmatrix} \cos b & \sin b \\ -\sin b & \cos b \end{bmatrix}.$$

Proof. Let $(\lambda, v), (\lambda^*, v^*)$ denote the eigen-value/vector pairs. We have

$$\det\left(\lambda I - \begin{bmatrix} a & b \\ -b & a \end{bmatrix} \right) = \det\left(\begin{bmatrix} \lambda - a & -b \\ b & \lambda - a \end{bmatrix} \right) = (\lambda - a)^2 + b^2 = 0.$$

We conclude that $\lambda = a + bi$. We now compute the associated v. We have

$$\begin{bmatrix} bi & -b \\ b & bi \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = b \begin{bmatrix} i & -1 \\ 1 & i \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

We can easily see that $v_1 = 1$ and $v_2 = i$. Letting $v = \begin{bmatrix} 1 & i \end{bmatrix}^\top$, we compute $\|v\| = 2$. Thus, we may write

$$\begin{bmatrix} a & b \\ -b & a \end{bmatrix} = \frac{1}{2} \begin{bmatrix} 1 & 1 \\ i & -i \end{bmatrix} \begin{bmatrix} a + bi & \\ & a - bi \end{bmatrix} \begin{bmatrix} 1 & -i \\ 1 & i \end{bmatrix},$$

which means that

$$\exp\left\{\begin{bmatrix} a & b \\ -b & a \end{bmatrix}\right\} = \begin{bmatrix} 1 & 1 \\ i & -i \end{bmatrix} \begin{bmatrix} e^{a+bi} & \\ & e^{a-bi} \end{bmatrix} \begin{bmatrix} 1 & -i \\ 1 & i \end{bmatrix}$$

$$= \frac{1}{2}e^a \begin{bmatrix} 1 & 1 \\ i & -i \end{bmatrix} \begin{bmatrix} e^{bi} & \\ & e^{-bi} \end{bmatrix} \begin{bmatrix} 1 & -i \\ 1 & i \end{bmatrix}$$

$$= \frac{1}{2}e^a \begin{bmatrix} 1 & 1 \\ i & -i \end{bmatrix} \begin{bmatrix} e^{bi} & -ie^{bi} \\ e^{-bi} & ie^{-bi} \end{bmatrix}$$

$$= \frac{1}{2}e^a \begin{bmatrix} e^{bi} + e^{-bi} & -i\left(e^{bi} - e^{-bi}\right) \\ i\left(e^{bi} - e^{-bi}\right) & e^{bi} + e^{-bi} \end{bmatrix}$$

$$= e^a \begin{bmatrix} \cos b & -i^2 \sin b \\ i^2 \sin b & \cos b \end{bmatrix}$$

$$= e^a \begin{bmatrix} \cos b & \sin b \\ -\sin b & \cos b \end{bmatrix}$$

as desired. □

Even if A is not diagonalizable, it is still possible to compute e^A by using the fact that $x(t) = e^{At}x_0$ is the solution to the initial-value problem (**??**). The procedure to do so is outlined by the following theorem:

Theorem 4.4. (Computing e^A for any A) Let $A \in \mathbb{R}^{n,n}$ and consider the initial-value problem in (**??**). Taking the Laplace transform of both sides, we have

$$\mathcal{L}(\dot{x}) = \mathcal{L}(Ax)$$

$$\Rightarrow sX(s) - \underbrace{x(0)}_{x_0} = AX(s)$$

$$\Rightarrow (sI - A)X(s) = x_0$$

$$\Rightarrow X(s) = (sI - A)^{-1}x_0.$$

Taking the inverse Laplace transform of both sides, we see that

$$x(t) = \mathcal{L}^{-1}\left((sI - A)^{-1}\right)x_0.$$

Since (we claimed that) $x(t) = e^{At}x_0$ is the unique solution to (**??**), and so

$$e^{At}x_0 = \mathcal{L}^{-1}\left((sI - A)^{-1}\right)x_0.$$

Upon computing e^{At}, we can readily evaluate the expression at $t = 1$ to obtain e^A.

4.3 Existence and Uniqueness

We now show that $x(t) = e^{At}x_0$ is in fact the unique solution to the initial-value-problem in (**??**).

Theorem 4.5. (Unique Solution to $\dot{x} = Ax$) The unique solution to the initial-value problem in (**??**) is given by $x(t) = e^{At}x_0$.

Proof. We begin by showing that $x(t) = e^{At}x_0$ is in fact a solution. Indeed, we can easily see that the initial condition is satisfied since $x(0) = e^{A(0)}x_0 = Ix_0 = x_0$. As for the differential equation, we have

$$\dot{x} = \frac{d}{dt}x(t) = \frac{d}{dt}\left(e^{At}x_0\right) = Ae^{At}x_0 = Ax(t).$$

All that remains is to show that the aforementioned solution is unique. To that end, we assume there exists another solution, $z(t)$ that solves (**??**). Then by definition, $z(0) = x_0$ and $\dot{z}(t) = Az(t)$. If we let $w(t) := e^{-At}z(t)$, we have

$$\dot{w} = \frac{d}{dt}w(t) = \frac{d}{dt}e^{-At}z(t) = \frac{d}{dt}\left(e^{-At}\right)z(t)+e^{-At}\frac{d}{dt}z(t) = -Ae^{-At}z(t)+e^{-At}Az(t) = \mathbf{0}.$$

In other words, $w(t) = e^{-At}z(t)$ is constant in t. It follows that $w(t) = w(0) = e^{-A(0)}z(0) = Ix_0 = x_0$. Therefore, $e^{-At}z(t) = x_0$, which implies that $z(t) = e^{At}x_0$. □

Let us now consider the more general case where $B \neq \mathbf{0}$.

4.4 Geometric Interpretations

5 Stability

An import aspect of a system is whether its state exhibits some sense of stability.

5.1 Notions of Stability for Autonomous Systems

We shall first consider definitions of stability for a general (possibility non-linear) system, but where we assume $u \equiv 0$.

Definition 5.1. (Stability) The system given by (**??**) is **stable** if

$$\|x(0)\| \leq \delta \Rightarrow \|x(t)\| \leq \epsilon, \forall t.$$

In other words, if the system's state starts within the δ-norm ball, it stays within the ε-norm ball. A system is **unstable** if it is not stable.

Definition 5.2. (Attractive System) The system given by (**??**) is **attractive** if

$$\|x_0\| \leq \delta \Rightarrow \lim_{t \to \infty} \|x(t)\| = 0.$$

An attractive system is not necessarily stable since it is possible for $x(t)$ to exit the ε-norm ball even if it converges to zero in the long run.

Definition 5.3. (Asymptotic Stability) The system given by (**??**) is **asymptotically stable** if it is both stable attractive.

5.2 Stability of Autonomous Linear Systems

Our goal now is to develop necessary and sufficient conditions for stability for the system given by (**??**). To begin, recall that the solution is given by $x(t) = e^{At}x_0$, where x_0 is the initial condition.

Theorem 5.1. (Asymptotic Stability) Let $\mathbb{C}_{\mathfrak{Re}<0}$ denote the open left-half of the complex plane, i.e.,

$$\mathbb{C}_{\mathfrak{Re}<0} := \{z \in \mathbb{C} | \mathfrak{Re}(z) < 0\}.$$

The system given by (**??**) is asymptotically stable if and only if the eigenvalues of A all lie within $\mathbb{C}_{\mathfrak{Re}<0}$.

Proof. Let $\lambda_1, \ldots, \lambda_n$ denote the eigen-values of A. Let $A = P\Delta P^{-1}$. Then,

$$e^{At} = Pe^{\Delta}P^{-1}.$$

In general,

$$\Delta = \begin{bmatrix} e^{J_{\lambda_1}t} & & \\ & \ddots & \\ & & e^{J_{\lambda_k}t} \end{bmatrix},$$

where

$$J_{\lambda_i}t = e^{\lambda_i t} \begin{bmatrix} 1 & t & \ldots & t^k/k! \\ & \ddots & \ddots & \\ & & 1 & t \end{bmatrix}.$$

$\Rightarrow$: If $\lambda_1, \ldots, \lambda_n \in \mathbb{C}_{\mathfrak{Re}<0}$, then, $e^{\lambda_i t} t^k \to 0$ as $t \to 0$ for every k and $i = 1, \ldots, n$. Moreover, we have

$$e^{\lambda_i t} t^k \to 0 \Rightarrow J_{\lambda_i} t \to \mathbf{0} \Rightarrow \Delta \to \mathbf{0} \Rightarrow e^{\Delta} \to \mathbf{0} \Rightarrow e^{At} \to \mathbf{0} \Rightarrow x(t) \to 0.$$

We conclude that $x(t) \to 0$ as $t \to \infty$, and the system is asymptotically stable.

$\Leftarrow$: Assume $\lambda_i \notin \mathbb{C}_{\mathfrak{Re}<0}$. Then $e^{\lambda_i t} t^k \not\to 0$ as $t \to 0$. Moreover, we have

$$J_{\lambda_i} t \not\to \mathbf{0} \Rightarrow \Delta \not\to \mathbf{0} \Rightarrow e^{\Delta} \not\to \mathbf{0} \Rightarrow e^{At} \not\to \mathbf{0} \Rightarrow x(t) := e^{At} x_0 \not\to 0, \forall x_0 \neq \mathbf{0}.$$

We conclude that $x(t) \not\to 0$ as $t \to \infty$ for any $x_0 \neq \mathbf{0}$, and the system is not asymptotically stable. □

Theorem 5.2. (Stability) Let $\mathbb{C}_{\mathfrak{Re}\leq 0}$ denote the closed left-half of the complex plane, i.e.,

$$\mathbb{C}_{\mathfrak{Re}\leq 0} = \{z \in \mathbb{C} : \mathfrak{Re}(z) \leq 0\}.$$

The system given by (**??**) is stable if and only if the eigenvalues of A all lie within $\mathbb{C}_{\mathfrak{Re}\leq 0}$ and the geometric and algebraic multiplicity of any eigenvalue, λ of A, with $\mathfrak{Re}(\lambda) = 0$ are equal.

Proof. Omitted. □

5.3 Stability of General Linear Systems

We now extend the definitions of stability for general linear systems, where $u \not\equiv 0$.

Definition 5.4. (Input/Output Stability) The system given by (**??**) is **input/output** stable under the initial condition x_0, if every bounded input, u, the output, y, is also bounded, i.e.,

$$\|u(t)\| \leq \varepsilon, \forall t \Rightarrow \|y(t)\| \leq \delta, \forall t$$

where δ and ϵ are finite.

If the system is input/output stable under every x_0, then we say it is simply **input/output** stable.

Definition 5.5. (Bounded-Input/Bounded-Output Stability) The system given by (**??**) is **bounded-input/bounded-output stable** if it is input/output stable under $x_0 = \mathbf{0}$.

We now consider necessary and sufficient conditions for stability for the system described by (**??**).

6 The Representation Theorem

Our goal in this section is to derive a key result in linear algebra called the "Representation theorem".