Linear Systems

Course Notes

Sivakumar Balasubramanian

October 7, 2018

Contents

1

Basics of Linear Algebra and Matrix Theory

1.1 Vector space

This chapter mainly consists of a bunch of definitions and concepts about vector spaces that one must be clear about to understand linear algebra. Like most topics in mathematics, or any discipline with a formal structure, the importance of being absolutely clear about the foundational ideas cannot be overstated.

> **Definition 1.1 Vector Space**
>
> A vector space $\mathcal{V}$ over the field $\mathcal{F}$ is a non-empty set of elements that is closed under the operations of scalar multiplication and vector addition.

There are four things involved in a vector space as defined above.

1. A non-empty set of elements, where the are called **vectors**.
2. A field $\mathcal{F}$ of scalar that we will be dealing with when working vector spaces. The numbers associated with the elements of the vectors spaces will be from the field $\mathcal{F}$. We will be mostly dealing with the real $\mathbb{R}$ or complex $\mathbb{C}$ scalar fields in this course.
3. A vector addition operation denoted by $\mathbf{v}_1 + \mathbf{v}_2$, where $\mathbf{v}_1, \mathbf{v}_2 \in \mathcal{V}$.
4. A scalar multiplication operation denoted by $\alpha\mathbf{v}_1$ between an element of $\mathbf{v} \in \mathcal{V}$ and $\alpha \in \mathcal{F}$.

Note: The exact nature of the vector addition and scalar multiplication operations need to be defined depending on the type of vector space we are dealing with.

Vector spaces generalize the idea of Euclidean (flat) spaces that we are very familiar with from our everyday experience and also most our basic math from school.

The most common example of vector spaces that we will encounter in this course are the n-tuple of numbers, which could be thought of as points in an n-dimensional Euclidean space. Consider the all familiar xy-plane shown in Fig. 1.1. The set of all points in this plane forms a vector space. Each point in this plane can be represented by 2-tuple of real numbers,

$$\mathbb{R}^2 = \left\{ \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \middle| x_1, x_2 \in \mathbb{R} \right\} = \{\ldots \mathbf{p}_1, \ldots \mathbf{p}_2, \ldots\}$$

The points $\mathbf{p}_1$, $\mathbf{p}_2$ shown in Fig. 1.1 would be somewhere in the set $\mathbb{R}^2$.

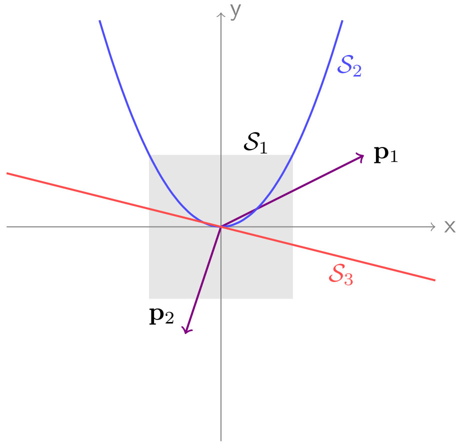


Figure 1.1: xy-plane.

Similarly, the set of all points in n-dimensional Euclidean space form a vector space,

$$\mathbb{R}^n = \left\{ \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \middle| x_1, x_2, \ldots x_n \in \mathbb{R} \right\}$$

Vector addition between two n-tuples $\mathbf{v}, \mathbf{w}$ is defined as the sum of the corresponding scalar elements,

$$\mathbf{v} + \mathbf{w} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} + \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{bmatrix} \triangleq \begin{bmatrix} v_1 + w_1 \\ v_2 + w_2 \\ \vdots \\ v_n + w_n \end{bmatrix}$$

Scalar multiplication of an n-tuple by a scalar $\alpha \in \mathbb{R}$ is defined as,

$$\alpha\mathbf{v} = \alpha \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} \triangleq \begin{bmatrix} \alpha v_1 \\ \alpha v_2 \\ \vdots \\ \alpha v_n \end{bmatrix}$$

The elements of the n-tuple results from vector addition and scalar multiplication are all real number and thus belong to $\mathbb{R}^n$.

The vector addition and scalar multiplication operations associated with a vector space $\mathcal{V}$,

1. $\mathbf{v} + \mathbf{w} = \mathbf{w} + \mathbf{v}$.
2. $\mathbf{u} + (\mathbf{v} + \mathbf{w}) = (\mathbf{u} + \mathbf{v}) + \mathbf{w}$.
3. There is a zero element in the vector space $\mathcal{V}$, $\mathbf{0} \in \mathcal{V}$, such that $\mathbf{v} + \mathbf{0} = \mathbf{v}$.

4. For every $\mathbf{v} \in \mathcal{V}$, there is an element $-\mathbf{v}$, such that $\mathbf{v} + (-\mathbf{v}) = \mathbf{0}$.
5. $(\alpha\beta)\,\mathbf{v} = \alpha\,(\beta\mathbf{v})$, $\forall \alpha, \beta \in \mathcal{F}$ and $\forall \mathbf{v} \in \mathcal{V}$.
6. $\alpha\,(\mathbf{v} + \mathbf{w}) = \alpha\mathbf{v} + \alpha\mathbf{w}$, $\forall \alpha \in \mathcal{F}$ and $\forall \mathbf{v}, \mathbf{w} \in \mathcal{V}$.
7. $(\alpha + \beta)\,\mathbf{v} = \alpha\mathbf{v} + \beta\mathbf{v}$, $\forall \alpha, \beta \in \mathcal{F}$ and $\forall \mathbf{v} \in \mathcal{V}$.
8. $1\mathbf{v} = \mathbf{v}$, $\forall \mathbf{v} \in \mathcal{V}$.

1.2 Subspace

Consider a vector space $\mathcal{V}$[1]. Let us now consider $\mathcal{S}$, a subset of $\mathcal{V}$, i.e. $\mathcal{S} \subseteq \mathcal{V}$. This means that if $x \in \mathcal{S}$ then $x \in \mathcal{V}$. A subset of $\mathcal{V}$ qualifies to a subspace of $\mathcal{V}$ only if $\mathcal{S}$ itself is a vector space.

Definition 1.2 Subspace

Let $\mathcal{S}$ be a non-empty subset of a vector space $\mathcal{V}$ over the field $\mathcal{F}$. $\mathcal{S}$ is a subspace over the same field $\mathcal{F}$ with the same vector addition and scalar multiplication operations as $\mathcal{V}$, if and only if $\mathcal{S}$ is closed under the vector addition and scalar multiplication operations.

$$\begin{aligned} \mathbf{x}, \mathbf{y} \in \mathcal{S} &\implies \mathbf{x} + \mathbf{y} \in \mathcal{S} \\ \mathbf{x} \in \mathcal{S} \text{ and } \alpha \in \mathcal{F} &\implies \alpha\mathbf{x} \in \mathcal{S} \end{aligned}$$

Let us look at some examples to get a good grasp of what subspaces are. $\mathcal{V} = \mathbb{R}^2$ is the parent vector space.

$$\mathcal{V} = \mathbb{R}^2 = \left\{ \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \middle| x_1, x_2 \in \mathbb{R} \right\}$$

The following are some of non-empty subsets of $\mathbb{R}^2$ (shown in Fig. 1.1, but not all of them are subspaces.

1. $\mathcal{S}_1 = \left\{ \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \middle| \, |x_1| \leq 1, |x_2| \leq 1 \right\}$. This is the set of all points in the gray square shown in Fig. 1.1.

   This is **not** a subspace of $\mathcal{V}$. Because, it is not closed under scalar multiplication or vector addition. For example, $\mathbf{v}_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \mathbf{v}_2 = \begin{bmatrix} -1 \\ 1 \end{bmatrix} \in \mathcal{S}_1$, but $\mathbf{v}_1 + \mathbf{v}_2 = \begin{bmatrix} 0 \\ 2 \end{bmatrix} \notin \mathcal{S}_1$.

2. $\mathcal{S}_2 = \left\{ \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \middle| x_1 \in \mathbb{R}, \, x_2 = x_1^2 \right\}$. This is the set of all points on the blue colored parabola in Fig. 1.1.

   This is also **not** a subspace of $\mathcal{V}$, as it is not closed under scalar multiplication or vector addition. Let, $\mathbf{v}_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \in \mathcal{S}_2$, but $2\mathbf{v}_1 = \begin{bmatrix} 2 \\ 2 \end{bmatrix} \notin \mathcal{S}_2$.

3. $\mathcal{S}_3 = \left\{ \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \middle| x_1 \in \mathbb{R}, \, x_2 = \beta x_1, \beta \in \mathbb{R} \right\}$. This is the set of all points on the red line passing through the origin in Fig. 1.1.

   This is a subspace of $\mathcal{V}$; it is closed under scalar multiplication or vector addition.

   Let $\mathbf{v} = \begin{bmatrix} v_1 \\ \beta v_1 \end{bmatrix}$ and $\mathbf{w} = \begin{bmatrix} w_1 \\ \beta w_1 \end{bmatrix}$. Then, $\alpha\mathbf{v} = \begin{bmatrix} \alpha v_1 \\ \alpha\beta v_1 \end{bmatrix} = \begin{bmatrix} \alpha v_1 \\ \beta\,(\alpha v_1) \end{bmatrix} \in \mathcal{S}_3$.

   And, $\mathbf{v} + \mathbf{w} = \begin{bmatrix} v_1 \\ \beta v_1 \end{bmatrix} + \begin{bmatrix} w_1 \\ \beta w_1 \end{bmatrix} = \begin{bmatrix} v_1 + w_1 \\ \beta v_1 + \beta w_1 \end{bmatrix} = \begin{bmatrix} v_1 + w_1 \\ \beta\,(v_1 + w_1) \end{bmatrix} \in \mathcal{S}_3$.

Problem Subspaces

Which of the following subsets are subspaces of $\mathbb{R}^2$?

1. $\left\{ \alpha\mathbf{p} \mid \mathbf{p} \in \mathbb{R}^2, \alpha = \mathbb{R} \right\}$.
2. $\left\{ \alpha\mathbf{p} + \mathbf{q} \mid \mathbf{p}, \mathbf{q} \in \mathbb{R}^2, \alpha = \mathbb{R} \right\}$.
3. $\left\{ \mathbf{p} = \begin{bmatrix} p_1 \\ p_2 \end{bmatrix} \middle| \, \mathbf{p} \in \mathbb{R}^2, \, p_1^2 + p_2^2 = 1 \right\}$.
4. $\left\{ \mathbf{p} = \begin{bmatrix} p_1 \\ p_2 \end{bmatrix} \middle| \, \mathbf{p} \in \mathbb{R}^2, \, p_1^2 + p_2^2 = 0 \right\}$.

1.3 Span of a set of vectors

Consider a set of vectors $\mathcal{S} = \{\mathbf{v}_1, \mathbf{v}_2, \ldots \mathbf{v}_n\}$, $\mathbf{v}_i \in \mathcal{V}$, $1 \leq i \leq n$, where $\mathcal{V}$ is a vector space over the field $\mathcal{F}$[2]. A linear combination of the elements of $\mathcal{S}$ is a vector $\mathbf{u} \in \mathcal{V}$ defined as the following,

$$\mathbf{u} = \sum_{i=1}^{n} \alpha_i \mathbf{v}_i, \, \alpha_i \in \mathcal{F}, \, \mathbf{v}_i \in \mathcal{S}$$

Definition 1.3 Span of set of vectors

The span of a set of vectors $S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots \mathbf{v}_n\}$ is defined as the set of all linear combinations of a the elements of $\mathcal{S}$, i.e.

$$span\,(\mathcal{S}) = \left\{ \sum_{i=1}^{n} \alpha_i \mathbf{v}_i \middle| \alpha \in \mathcal{F} \right\}$$

What can we say about $span\,(\mathcal{S})$? Is it a subset of $\mathcal{V}$? Is it a subspace of $\mathcal{V}$? It turns out that the span of a set of vector does form a subspace of $\mathcal{V}$. It is left as an exercise for the reader to prove this.

The subspace $span\,(\mathcal{S})$ is called the subspace spanned by $\mathcal{S}$, or $\mathcal{S}$ is said to span the subspace $span\,(\mathcal{S})$. $\mathcal{S}$ is also called the spanning set of $span\,(\mathcal{S})$.

Problem Span of a set of vectors

1. Consider a set of vectors $\mathcal{S} = \{\mathbf{v}_1, \mathbf{v}_2, \ldots \mathbf{v}_n\}$, $\mathbf{v}_i \in \mathcal{V}, 1 \leq i \leq n$, where $\mathcal{V}$ is a vector space. Prove that a linear combination of the vectors of $\mathcal{S}$ belong to $\mathcal{V}$.

[1] Whenever you hear see the term 'vector space', you must train yourself to immediately think of the the Def. 1.1

[2] From now on, we will assume that a given vector space is over the field $\mathbf{F}$, unless specified otherwise.

2. Prove that the span of a set $\mathcal{S}$ is a subspace of $\mathcal{V}$.

1.4 Linear independence

Like the span, linear independence (LI) is also a concept associated with a set of vectors $\mathcal{S} = \{\mathbf{v}_1, \mathbf{v}_2, \dots \mathbf{v}_n\}$, $\mathbf{v}_i \in \mathcal{V}$, $1 \le i \le n$ where $\mathcal{V}$ is a vector space. LI tells us if one or more vectors in $\mathcal{S}$ is in the subspace spanned by other vectors in $\mathcal{S}$ (or if one or more vectors is a linear combination of the other vectors in $\mathcal{S}$).

Definition 1.4 Linear Independene

A set of vectors $\mathcal{S} = \{\mathbf{v}_1, \mathbf{v}_2, \dots \mathbf{v}_n\}$ is a linearly independent set if and only if,

$$\sum_{i=1}^{n} \alpha_i \mathbf{v}_i = 0 \implies \alpha_i = 0, \forall i$$

The set is linearly dependent if at least one of the $\alpha_i \neq 0$.

1.5 Basis and dimension of a vector space

Vector spaces can consist of infinite number of elements. The special structure of vector spaces (closure under scalar multiplication and addition) allows us to generate all elements of a vector space using only a finite number of elements from that vector space[3]. We have seen this earlier: a spanning set $\mathcal{S}$ (consisting of a finite number of elements) can be used to generate all elements in the subspace $span\,(\mathcal{S})$.

Consider a set $\mathcal{S} = \{\mathbf{v}_1, \mathbf{v}_2, \dots \mathbf{v}_n\}$, $\mathbf{v}_i \in \mathcal{V}$, $1 \le i \le n$, a subset of the vector space $\mathcal{V}$, whose elements can be used to generate every element of $\mathcal{V}$ through a linear combination, i.e. $\forall \mathbf{x} \in \mathcal{V}, \exists \alpha_i \in \mathcal{F}, 1 \le i \le n$ such that ,

$$\mathbf{x} = \sum_{i=1}^{n} \alpha_i \mathbf{v}_i$$

A basis of a vector space $\mathcal{V}$ is a set with the least number of elements, which can be used to generate all elements of $\mathcal{V}$. Thus, a basis is a set with the sufficient of elements from $\mathcal{V}$ that are necessary to generate all elements of $\mathcal{V}$.

Definition 1.5 Basis of a vector space

A linearly independent spanning set of a vector space $\mathcal{V}$ is called a basis of $\mathcal{V}$.

There can be an infinite number of basis for a vector space $\mathcal{V}$. Apart from the all of them sharing a property of being linearly independent spanning sets of $\mathcal{V}$, there share another common property. They all have the same number of elements. This number is called the dimension of the vector space $\mathcal{V}$.

[3]This is only valid for finite dimensional vector space, and in this course we will only talk about finite dimensional vector space.

Definition 1.6 Dimension of a vector space

The dimension of a vector space dim$\mathcal{V}$ is the number of elements of a basis the basis of $\mathcal{V}$.

2

Signals and Systems

2.1 What is a signal?

A *signal* is defined as any measurable physical quantity carrying information that depends on one or more independent variables, such a time, space etc. Mathematically, a signal can be represented a function of a set of independent variable(s). For example,

$$s(t) = 0.23t^2 - 5.11t + 31.5$$

where, t is time. This is an example ofa 1-dimensional (1-D) signal, i.e. a signal that is a function of a single independent variable.

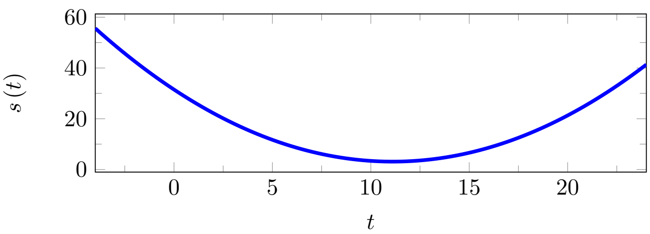


Figure 2.1: Visualization of a 1-D signal

An image is a 2-D signal that is a function of two independent spatial variables x and y. In the case of a gray-scale image, the intensity of the image at a spatial location is given by $I(x, y)$, where x is horizontal position, and y is the vertical position.

A mathematical representation a signal is not always possible for all types of signals. For example, many of the physiological signals cannot be represented mathematically, either because the exact function is not known or is too complicated.

Problem **Higher dimensional signals**

Can you think of an examples of a 3-D and 4-D signals?

2.2 Classification of signals

A particular type of signal classification was already mentioned in the previous section, based on the number of independent variables associated with the signal. Some of the other common classifications of signals are described in this section.

2.2.1 Scalar and Vector Signals

Signals can take one or more values for each value(s) of its independent variables. *Scalar* signals take on only one value, while *vector* signals take more than one value. A gray-scale image is an example of a scalar 2-D signal, while a RGB image is an example of a vector 2-D signal (with each point on the image taking three values).

$$I(x, y) \in [0, 1]$$

where, $I(x, y)$ is the intensity of the point with 0 corresponding to black and 1 corresponding to white colors. An RGB image on the other hand can be represented as the following,

$$\mathbf{I}(x, y) = \begin{bmatrix} r & g & b \end{bmatrix}^T$$

where, $r, g, b \in [0, 1]$ correspond to the amount of red, green and blue components at each point in the image. The different combinations of these three primary colors results in the different colors in the image.

2.2.2 Continuous-time and Discrete-time Signals

An important classification of signals is based on the nature of the values assumed by the independent variable of a signal. *Continuous-time* signals are ones whose independent variable can take on any value in continuous interval (a, b) on the real axis. For example, a function $e^{-0.1t^2}$ with $t \in (-\infty, \infty)$ is an example of a continuous-time signal.

On the other hand, *discrete* signals can take on values only for specific values of the independent variable. They can be represented in a tabular form that represents the mapping from the signal's domain to its range. For example consider a discrete signal $x[t]$ that takes on values only for specific time instants $t \in \{\ldots, t_n, t_{n+1}, t_{n+2}, \ldots\}$, where $n \in \mathbb{Z}$. The most common discrete signals that we will encounter are the ones where the time instants are uniformly spaced, i.e. $t \in \{\ldots, nT, (n+1)T, (n+2)T, \ldots\}$. The domain of signals of this form are simply taken to be $\mathbb{Z}$, ignoring the value of T.

We will use the square bracket notation for discrete-time signals, and the simple brackets for continuous-time signals. Consider the example of the continuous-time signal given above, $x(t) = e^{-0.1*t^2}, t \in \mathbb{R}$. The discrete-time version of this signal (sampled with $T = 1$) is $x[n] = e^{-0.1*n^2}, t \in \mathbb{Z}$. The plot of these two signals is shown below.

Note: Although these signals are called continuous-time and discrete-time, there is not restriction on the choice of the independent variable - it can be time, space, etc. The name 'time' is used in the naming convention as time signals are the most commonly encountered signals in practice.

In the above example, the discrete-time signal was obtained by selecting the values of the continuous-time signal

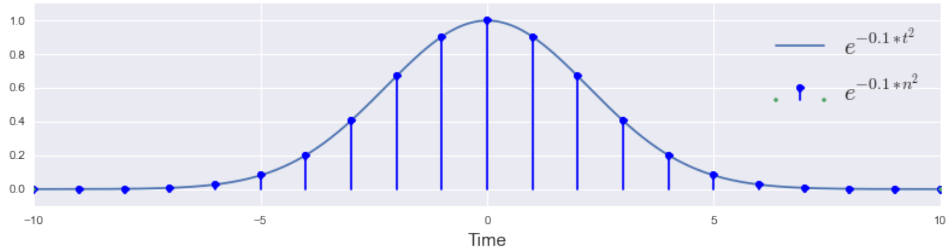


Figure 2.2: Continuous-time and discrete-time signals

at specific (equidistant) time points. This is the process of *sampling*, which we will be dealt in detail in Chapter 4.

2.2.3 Continuous-valued and Discrete-valued Signals

The previous classification was based on the discretiation of the indepndent variable, while the current classification is based on the dsicretization of the dependent variable, i.e. the values assumed by a signal for different values of its independent variable. When a signal can take any value in a continuous interval in $\mathbb{R}$, it is called a *continuous-valued* signal, while a *discrete-valued* signal can only take on specific values from a set of values. In the case of *discrete-valued* signals, the set of values the signal can take are usually equidistant. An example of a continuous-valued $(x(t))$ and discrete-valued $(\hat{x}(t))$ signals are shown in the following figure.

In the above example, the $\hat{x}(t)$ is the discrete-valued version of $x(t)$, and it can only take on values in the set $\{\ldots, -0.2, -0.1, 0, 0.1, 0.2, \ldots\}$. The conversion of $x(t)$ to $\hat{x}(t)$ is called *quantization*, which will be discussed in a later chapter.

The last two classifications can be combined to have four possible combinations of signals:

- **Continuous-time continuous-valued signals**: Both the domain and range can take on continous values.
- **Continuous-time discrete-valued signals**: Domain takes on continuous values, while the range can only take on specific discrete values.
- **Discrete-time continuous-valued signals**: Domain taken on values from a discrete set, while the range is continuous
- **Discrete-time discrete-valued signals**: Both the domain and the range take on discrete values.

These four cases are demonstrated in the following four figures.

2.2.4 Even and Odd Signals

This classification is based on the symmerty of signals about the vertical axis. A continous-time signal $x(t)$ is called even, if

$$x(t) = x(-t), \ \forall t$$

The signal $x(t)$ is called odd, if

$$x(t) = -x(-t), \ \forall t$$

Even signals are *symmertric* about the vertical axis (about the point $t = 0$), while odd signals are *antisymmetric* about the vertical axis. This classification also applied to discrete-time signals.

An interesting property of signals is that any signal $x(t)$ (even, odd or neigther) can be decomposed into an even and odd component. Lets us supposed that $x(t)$ can be decomposed into an even $x_e(t)$ and an odd $x_o(t)$ component.

$$x(t) = x_e(t) + x_o(t) \implies x(-t) = x_e(t) - x_o(t)$$

Thus,

$$x_e(t) = \frac{x(t) + x(-t)}{2} \ \& \ x_o(t) = \frac{x(t) - x(-t)}{2}$$

2.2.5 Periodic and Non-periodic Signals

A periodic signal is one that repeats itself after a fixed finite value of time (or the corresponding independent variable), i.e. if the continuous-time $x()$ is periodic, then

$$x(t) = x(t+T), \ \forall t$$

where T is a positive constant. If T is the smallest positive value for which the above relationship is satisfied, then T is called the fundamental period of the signal, and the following is also true.

$$x(t) = x(t+T) = x(t+2T) = x(t+3T)\ldots, \ \forall t$$

The inverse of T is called the fundamental frequency.

Any signal that does not satisfy the above condition is called an *aperiodic* or *non-periodic* signal.

In the case of a discrete-time signal $x[n]$, the following condition must be satisfied for the signal to be periodic,

$$x[n] = x[n+N], \ \forall n \in \mathbb{Z} \ \& \ N \in \mathbb{Z}$$

The fundamental period of a discrete-time signal must be an integer.

2.2.6 Deterministic and Stochastic Signals

All the signals we seen so far have been described through a mathematical expression. Such an expression allows one to completely characterize the signal on the domain over which it is defined. Such signals are termed *deterministic*. The representation of deterministic signals is not limited to mathematical expressions; a table of values or a well-defined rule for its calculation would also do as well. These explicit representations of the signal is known as the *signal model*. Through the signal model, the value of the signal can be be predicted for any value of its arguments in its domain.

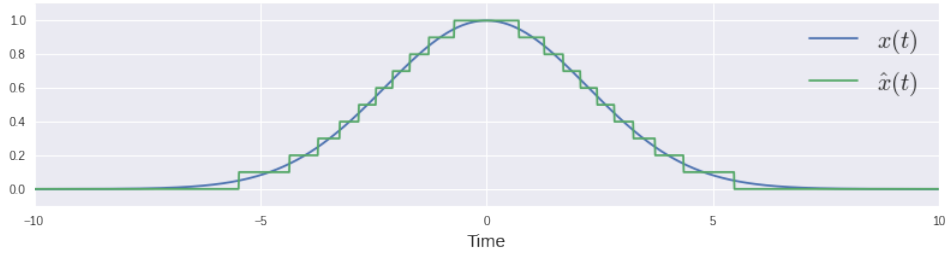


Figure 2.3: Continuous-time and discrete-time signals

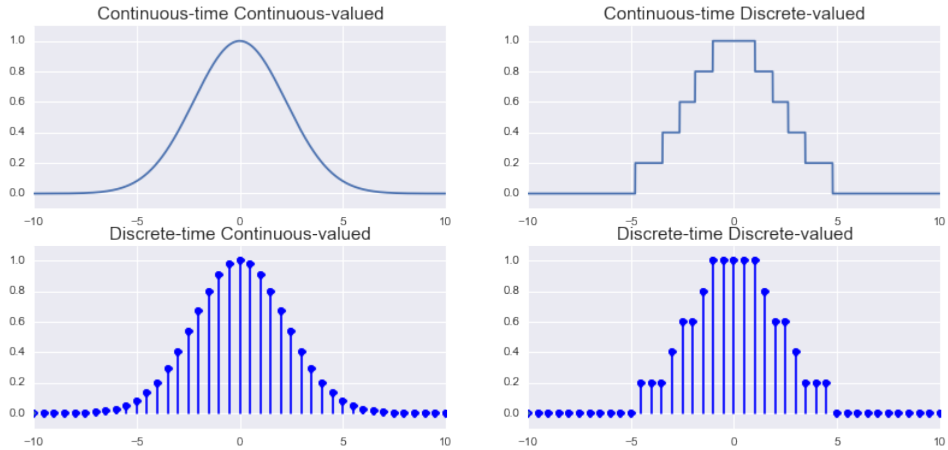


Figure 2.4: Continuous-time and discrete-time signals

A *stochastic* or *random* signal is one that cannot be represented by an explicit mathematical expression to any reasonable accuracy, or it is exceedingly complicated to do so for any practical purposes. Stochastic signals evolve in an unpredictable fashion as a functions of its arguments. For example, the EMG signal recorded using surface electrodes from a muscle is a stochastic signal. Unlike deterministic signals, which are ammenable to classical techniques of mathematical analysis, statistical techniques are the main tools for analysing stochastic signals.

2.2.7 Energy and Power Signals

Consider a continuous-time signals $v(t)$ and $i(t)$, which represent the voltage across and the current through a resistor R, respectively. Then the instantaneous power dissipated in the resistor is,

$$P(t) = \frac{v^2(t)}{R} = Ri^2(t)$$

In both cases the power is proportional to the resistance R. When the value of the resistance in 1omh, then the instantaneous power takes the same form for both $v(t)$ and $i(t)$.

In signal processing, regardless of the type of signal represented by $x(t)$, the instantaneous power is defined as,

$$p(t) = x^2(t)$$

Thus the total energy in a signal $x(t)$, $t \in (-\infty, \infty)$ is given by,

$$E = \int_{-\infty}^{\infty} x^2(t)\, dt$$

and the average power is defined as,

$$P = \lim_{T\to\infty} \frac{1}{T} \int_{-T/2}^{T/2} x^2(t)\, dt$$

Similar definitions can be given for discrete-time signals $(x[n])$ as well, by replacing the integrals by sums. The total energy in $x[n]$ is,

$$E = \sum_{n=-\infty}^{\infty} x^2[n]$$

And the average power is,

$$P = \lim_{N\to\infty} \sum_{n=-N}^{N} x^2[n]$$

Based on these definitions, an *energy* signal is defined as any signal whose total energy E is finite. i.e.

$$0 < E < \infty$$

While a signal is called a *power* signal if its average power is finite.

$$0 < P < \infty$$

Thus, an energy signal has zero time-averaged power, and a power signal have infinite energy. Periodic signals are usually viewed as power signals, while non-periodic signals are viewed as energy signals.

2.3 What is a system?

A system is any physical device or algorithm that performs some operation on a singal to transform it into another signal. Mathematically, systems can be thought of as functions or operators that map a signal to another signal. For example, an amplifier increases the amplitude of a signals, a filter removes components of a signals that are considered unwanted, etc. The following figure show a schematic of a system that takes a signal as its input, processes the signal and provides an output processed signal.

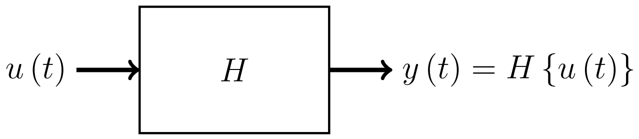


Figure 2.5: Schematic depiction of a system

2.4 Properites of systems

Systems are characterized by the nature of the operations that they perform on an input signal. The different classifications of systems, similar to that of singals, are characterized by the type of signals they deal with and the type of opertaions they perform on signals. Instead of presenting the different types of clasifications, like the one done for signals, here present some of the general characteristrics of systems which can be used as a basis for classifying systems.

2.4.1 Linearity

Linearity is an important properpty. All systems we will study and design in this book will be systems that are linear. A system is said to be linear when it satisfies the properties of scaling and superposition. Consider a system f that operates on a set of signals $x_i(t), i \in \{1, 2, \dots n\}$ and produces the outputs $y_i(t)$ respectively, i.e.

$$y_i(t) = f(x_i(t)), \, i \in \{1, 2, \dots n\}$$

Then, the system f is linear if and only if the following is true.

$$f\left(\sum_i a_i x_i(t)\right) = \sum_i a_i y_i(t)$$

where, a_is are some arbitrary constants. The scaling and superpostion properties are both contained in the previous statement.

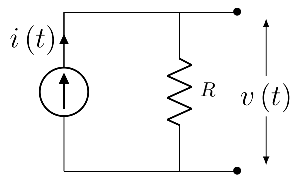
Scaling refers to the multiplication of the input signals by an arbitrary constant a results in the an output that is also equally scaled (multiplied by a), i.e.

$$f : x_1 \mapsto y_1 \implies a x_1 \mapsto a y_1$$

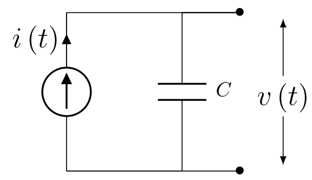
The principle of superposition says that if we know the output of a linear system to a set of inputs $x_i, i \in \{1, 2, 3, \dots n\}$, then the output of the system to the sum of these inputs is simply the sum of the outputs corresponding to these individual inputs. i.e.

$$f : x_i \mapsto y_i \implies \sum_i x_i \mapsto \sum_i y_i$$

Linearity is a very important property, and all systems that we will analyze and design in the rest of this book will be based on the linearity assumption, i.e. linear systems. Any system that does not satisfy the scaling and superposition properties are known as *non-linear* systems. Almost all real systems are non-linear, but it will be convenient to assume that they are linear (or approximately linear) in order to use the tools of signal processing and systems theory.



(a) System without memory.



(b) System with memory.

Figure 2.6: Simple examples of electrical systems with and without memory.

2.4.2 Memory

The property of memory is easy to understand in the context of a system that operates on time-domain signals. A system is said to have memory if its behavior depends on the past (and may be future) values of its input. On the contrary, it is said to be memoryless if its behavior only depends on the current values of its input.

The followings are the input-output relationships for the two systems that are shown in the above Fig. 2.6. The input to both these systems is a current source $i(t)$, and their corresponding outputs $v(t)$ are the voltages across the resistor R, and the capacitor C.

Resistor driven by a current source

The output voltage of this circuit is,

$$v(t) = R \times i(t)$$

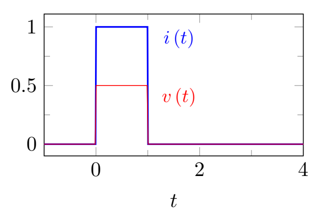
The output at any given time t only depends on the input value at that particular time. The input-output relationship of this system is shown in Fig. 2.7(a).

Capacitor driven by a current source

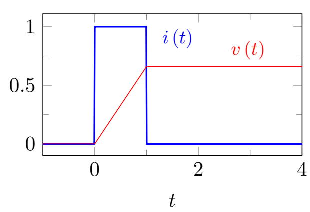
For the capacitor (Fig. 2.6), the input-output relationship is slightly more complicated. The relationship between $i(t)$ and $v(t)$ are governed by the following,

$$v(t) = \frac{1}{C} \int_{-\infty}^{t} i(\tau)\, d\tau$$

The output at any given time t depends on the entire history of the input signal $i(t)$ for all time. The input-output relationship of this system is shown in Fig. 2.7(b).



(a) Resistor circuit ($R = 2\Omega$)



(b) Capacitor circuit ($C = 1.5F$)

Figure 2.7: Input output relationship of a system (a) without and (b) with memory.

2.4.3 Causality

A system is said to be causal if its output depends only on the present and past values of its inputs, and not on the future values. While, a non-causal system's output depends also on the future values of the input signal. Causality and non-causality are purely based on what is considered the present. The choice of what is the present is fixed for systems that operate in real-time. For example, a filter t that is operating in real-time is strictly causal. However, a system that works off-line on stored data (non-real time) is free to define its present, and thus use data from the future to generate its output.

2.4.4 Time-invariance

Time-invariance property deals with whether or not a system changes over time. This change over time is characterized by the input-output relationship of the system. A system is said to time-invariant if its input-output relationship does not change with time, i.e. one get the same output from the system for a given input independent of whether the input was applied now, 10 minutes earlier or a year from now. Any system where this property does not hold is termed as time-variant. Time-invariance is another property that we will assume in the systems that we analyze and design in the upcoming chapters.

2.4.5 Stability

Stability refers to the property of a system to produce limited output when provided with finite input. Such systems are called stable system in the *bounded-input bounded-output (BIBO)* sense. Systems that do not satisfy the BIBO criteria are called unstable systems.

2.4.6 Invertibility

A system is said to invertible if an inverse of this system can be constructed, i.e. if a system f maps $x(t)$ to $y(t)$, then the inverse of the system will map $y(t)$ to $x(t)$.

3

Important Signals

We will encounter several types of signals in our study of linear systems. This chapter discusses the different signals we will be used at some point during our discussion of linear systems. In this chapter we will cover both continuous-time and discrete-time signals; in fact this will be our general approach in all chapter, where we will cover both continuous-time and discrete-time concepts.

3.1 Exponential Signals

3.1.1 Continuous-time real exponential

A continuous-time real exponential signal is represented as the following,

$$x(t) = ae^{bt}, \quad a, b \in \mathbb{R}$$

where, t is time, a is the amplitude of the signal when $t = 0$, and b the parameter indicate whether the $x(t)$ is an exponentially growing signal $(b > 0)$ or a exponentially decaying signal $(b < 0)$.

When $b = 0$, then we get a constant signal $x(t) = a, \ \forall t$. The plot of $x(t)$ for different values of b is given in Fig. 3.1.

Exponential signals are often observed in physical systems. They play an important role in the analysis of linear time-invariant system.

3.1.2 Discrete-time real exponential

A discrete-time real exponential signal is represented as,

$$x[n] = b \times a^n, \qquad a, b \in \mathbb{R} \text{ and } n \in \mathbb{Z}$$

This signal represents a growing exponential if $a > 1$, and its a decaying exponential if $0 < a < 1$; these are shown in Fig 3.2.

Problem **Discrete-time exponential signals**

1. What does a^n look like when: (a) $-1 < a < 0$; and (b) $a < -1$?

3.2 Sinusoidal Signals

3.2.1 Continuous-time sinusoids

The general form of a continuous-time sinusoidal signals is given by the following,

$$x(t) = A \sin(\omega t + \phi),$$

where A is the amplitude of the sinusoidal signal, ω is the angular frequency (radian per sec), and ϕ is the phase angle in radians. The sinusoid is an example of a periodic signal with the fundamental period $T = \frac{2\pi}{\omega}$ (Fig. 3.3).

Problem **Discrete-time exponential signals**

Please verify that T is the fundamental period of $x(t) = A\sin(\omega t + \phi)$.

The exponential signal was earlier introduced as the real exponential, as the parameters of the signal are real numbers. In fact a more general form of exponential signals is the complex exponential, where the parameters are allowed to be complex numbers. A complex exponential can be witten as,

$$x(t) = Ae^{bt}, \, t \in \mathbb{R}, \ A, b \in \mathbb{C}$$

Consider a complex exponential signal $e^{j\omega t}$. This can written as the following using the *Euler identity*.

$$e^{j\omega t} = \cos(\omega t) + j \sin(\omega t)$$

Thus a continuous-time sinusoid can be written as the real or imaginary part of the complex exponential.

$$\cos(\omega t + \phi) = \Re\left(e^{j\omega t + \phi}\right)$$

$$\sin(\omega t + \phi) = \Im\left(e^{j\omega t + \phi}\right)$$

Another representation of sinusoidal signals in terms of complex exponential is the following,

$$\cos(\omega t + \phi) = \frac{e^{(j\omega t + \phi)} + e^{-(j\omega t + \phi)}}{2}$$

$$\sin(\omega t + \phi) = \frac{e^{(j\omega t + \phi)} - e^{-(j\omega t + \phi)}}{2j}$$

3.2.2 Discrete-time sinusoids

The discrete-time equivalent of the sinusoids are represented as the following,

$$x[n] = A \sin(\Omega n + \phi)$$

where, A is the amplitude of the sinusoid, Ω is the discrete frequency (radians per sample), and ϕ is the phase. A plot of a discrete-time sinusoid is shown in Fig. 3.4.

The period of $x[n]$ is $N = k\frac{2\pi}{\Omega}$, where $k \in \mathbb{Z}$ such that $N \in \mathbb{Z}$. The discrete-time sinusoid has several features that are quite different to that of the continuous-time sinusoid. The following problems bring out some of these features.

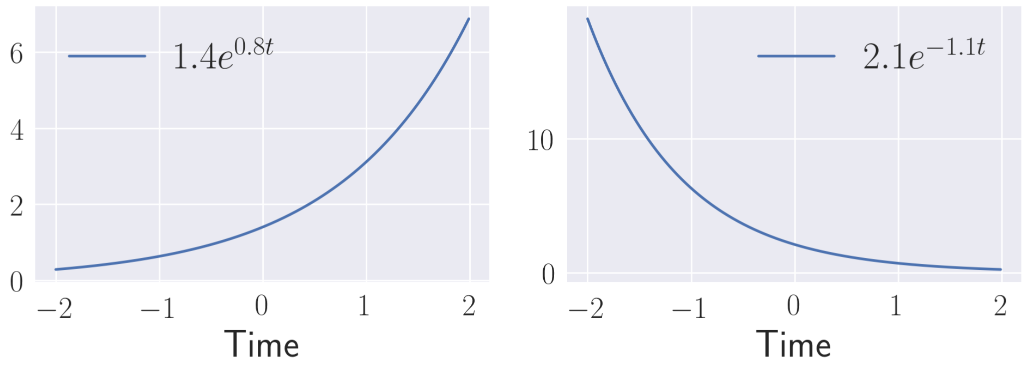


Figure 3.1: Continuous-time real exponential signals – growing exponential on the left and a decaying exponential on the right.

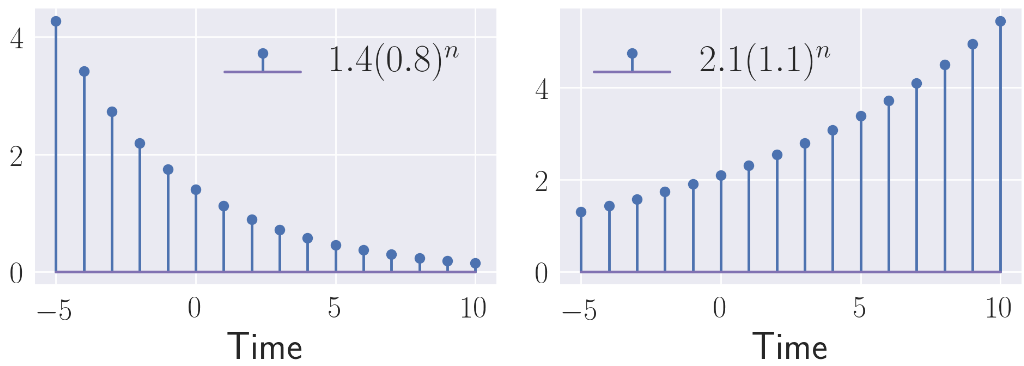


Figure 3.2: Continuous-time real exponential signals – growing exponential on the left and a decaying exponential on the right.

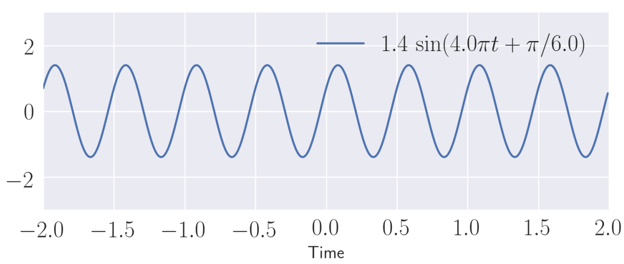


Figure 3.3: Continuous-time real exponential signals – growing exponential on the left and a decaying exponential on the right.

ch2-

Problem **Discrete-time sinusoidal signals**

1. Please verify that $N = k\frac{2\pi}{\Omega}$ ($k \in \mathbb{Z}$ such that $N \in \mathbb{Z}$) is the period of $\sin(\Omega n + \phi)$.
2. Prove that the discrete frequency Ω can only take values between $0 \leq \Omega < \pi$.
3. There are some sinusoidal signals that are not periodic. Verify that $\sin(n)$ is not periodic.

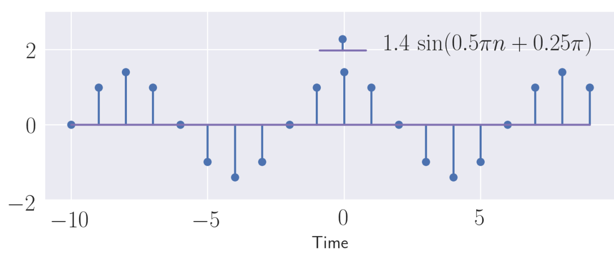


Figure 3.4: Continuous-time real exponential signals – growing exponential on the left and a decaying exponential on the right.

3.3 Exponential Sinusoidal Signals

A natural extension of the representation of sinusoids in terms of complex exponentials is to combine the real and complex exponentials, which results in the exponential sinusoidal signals. A exponentially weighted sinusoid is obtained by multiplying a sinusoidal signal ($\sin(\omega t + \phi)$) by a real exponential signal ($e^{\alpha t}$),

$$x(t) = Ae^{\alpha t} \sin(\omega t + \phi)$$

$x(t)$ can be represented using complex exponentials as,

$$x(t) = Ae^{\alpha t} \sin(\omega t + \phi) = A\frac{e^{(\alpha + j\omega)t} - e^{-(\alpha + j\omega)t}}{2j}$$

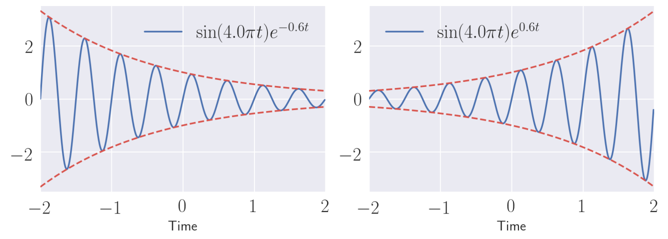


Figure 3.5: Exponential continuous-time sinusoidal signals – growing exponential on the left and a decaying exponential on the right.

$$x\left(t\right) = Ae^{\alpha t}\frac{e^{j\omega t} - e^{-j\omega t}}{2j}$$

Problem **Discrete-time exponential sinusoidal signals**

Write down the expression for a discrete-time exponential sinusoidal signal. Explain how the different parameters affect the nature of the signal.

3.4 Impulse function

3.4.1 Continuous-time impulse function

The first and most important fact to remember about the impulse function is that it is not an ordinary function. The impulse function $\delta\left(t\right)$, also know as the Dirac delta function, is not characterized by the exact values it takes for the different values of its argument, rather it is characterized by the following property,

$$\int_a^b \delta\left(t\right) dt = \begin{cases} 1, & 0 \in [a, b] \\ 0, & \text{Otherwise} \end{cases}$$

This property tells us that the impulse function is concentrated around the origin $t = 0$.

A useful property of the impulse function is that ***sifting property***, For any ordinary function $f\left(t\right)$, which is continuous at $t = t_0$,

$$\int_{-\infty}^{\infty} f\left(t - t_0\right)\delta\left(t\right) dt = f\left(t_0\right)$$

One way to think of the impulse function in terms of ordinary functions is to see it a limit of a sequence or family of functions. One sequence that is commonly encountered in books on signals and systems is the following,

$$g_n\left(t\right) = \begin{cases} n, & -\frac{1}{2n} \leq t \leq \frac{1}{2n} \\ 0, & \text{Otherwise} \end{cases}$$

This is a rectangular function that grows taller as $n \to \infty$. Now, lets take $g_n\left(t\right)$ and apply it on an ordinary function $f\left(t\right)$,

$$\int_{-\infty}^{\infty} g_n\left(t\right) f\left(t\right) dt = \int_{-\frac{1}{2n}}^{\frac{1}{2n}} nf\left(t\right) dt = f_n$$

And,

$$\lim_{n\to\infty}\int_{-\infty}^{\infty} g_n\left(t\right) f\left(t\right) dt = \lim_{n\to\infty} f_n = f\left(0\right)$$

This is demonstrated in Fig. 3.6.

3.4.2 Discrete-time impulse function

The discrete-time impulse function is theoretically a much simpler function deal with. The discrete-time impulse function is defined as,

$$\delta\left[n\right] = \begin{cases} 1, & n = 0 \\ 0, & n \neq 0 \end{cases}$$

3.4.3 Step Function

3.4.4 Continuous-time step function

The step function is defined using the impulse function as the following,

$$1\left(t\right) = \int_{-\infty}^{t} \delta\left(\tau\right) d\tau = \begin{cases} 1, & t > 0 \\ 0, & t < 0 \\ \frac{1}{2}, & t = 0 \end{cases}$$

This function has a step discontinuity at $t = 0$.

Problem **Discrete-time step signal**

What would be definition of a discrete-time step function? How would you represent the discrete-time step function in terms of the discrete-time impulse function.

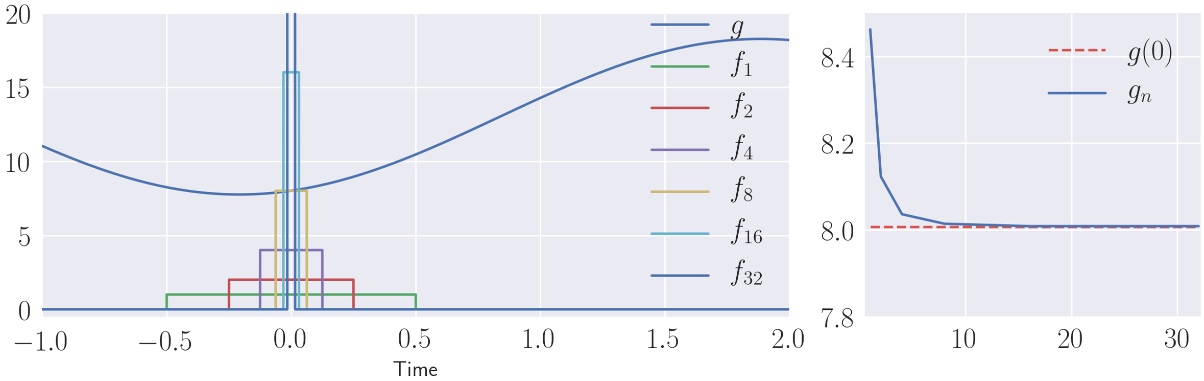


Figure 3.6: Viewing the impulse function as a limit of a sequence of rectangular pulses.

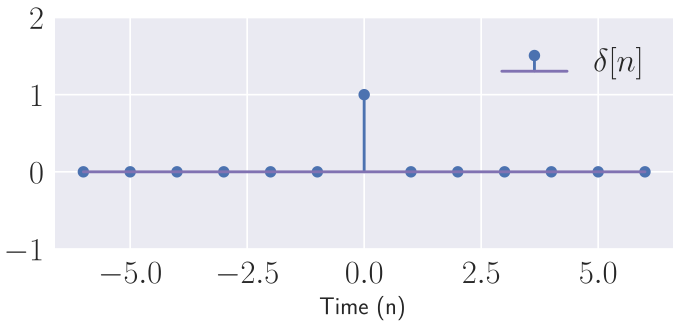


Figure 3.7: Discrete-time impulse signal.

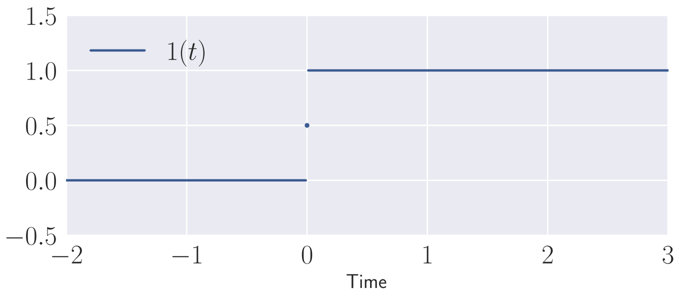


Figure 3.8: Viewing the impulse function as a limit of a sequence of rectangular pulses.

4

Continuous-time Linear Time-Invariant systems

A system is any physical device or algorithm that performs some operation on a singal to transform it into another signal. Mathematically, systems can be thought of as functions or operators that map a signal to another signal.

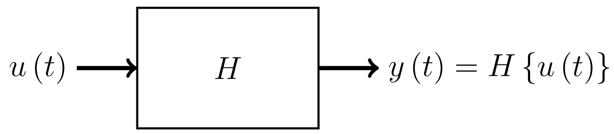


Figure 4.1: Schematic depiction of a system

Mathematically, one could think of a system as a map $H(\bullet)$ that associates a set of input signals U to a set of output signals Y, i.e.

$$H(\bullet) : U \mapsto Y \tag{4.1}$$

In this book, we will primarily talk about two different types of systems: continuous-time (CT) and discrete-time (DT). A CT system maps a set of CT input signals to a set of CT output signals, while a DT system maps DT input signals to DT output signals. In this book, we will focus primarily on linear time-invariant (LTI) systems. The current and next chapter deal with CT LTI and DT LTI systems.

4.1 Why study linear time-invariant systems?

The theroy of linear time-invariant (LTI) systems play and important role in engineering, and there are several reasons to study linear time-invariant systems.

1. Most engineering systems that we encounter in nature can be well approximated by LTI system for understanding, analyzing and controlling their behavior.

2. The mathematical theory of linear time-invariant systems is well developed and there are lots of useful tools for dealing with these systems. This is not surprising because LTI systems are simpler to deal with than a non-linear time-variant system.

4.2 Input-Output behavior of a system

Consider a general system H about which we know nothing about. We can interact with the system by applying some input and measure the corresponding output. To understand the behavior of the system, we could apply a set of inputs $U_{test} = \{u_i(t)\}_{i=1}^{N}$ and observe the corresponding output $Y_{test} = \{y_i(t)\}_{i=1}^{N}$, and tabulate the results. With the knowledge of these N input-output (IO) pairs, what can we say about the output of the system for any arbitrary input $u(t)$ that is not in U_{test}? In the case of a non-linear system, we cannot say anything; our knowledge of the system's IO characteristics are restricted to the set U_{test}, Y_{test}. But if we are told that the system is linear, then we know more about the IO characteristics of the system, beyond the known IO pairs U_{test}, Y_{test}. For a linear system, from the IO pairs, we also know the output of the system to any input of the form $u(t) = \sum_{i=1}^{N} \alpha_i u_i(t)$.

$$y(t) = H(u(t)) = H\left(\sum_{i=1}^{N} \alpha_i u_i(t)\right) = \sum_{i=1}^{N} \alpha_i H(u_i(t))$$

$$y(t) = \sum_{i=1}^{N} \alpha_i y_i(t) \tag{4.2}$$

Eq. 4.2 tells us that for a linear system, from U_{test}, Y_{test}, we also know the output of the system to any input that is a linear combination of the inputs in the set U_{test}.

5

Problems

"I hear and I forget. I see and I remember. I do and I understand."
– Confucius,

5.1 Linear Time-Invariant (LTI) Systems

1. Prove that for a memoryless LTI system the output is a scaled version of the input. Is this also true for: (a) a linear time-variant system; and (b) a non-linear system? Explain your answer.