Models to Measurements:
Observing Dynamic Systems

Brian Bingham

Version 0.1

Models to Measurements:
Observing Dynamic Systems

The original form of this book is LaTeX source code. Compiling this code has the effect of generating a device-independent representation of a textbook, which can be converted to other formats and printed.

Preface

This book is an extention of the course notes developed for the undergraduate course *ME 402: Dynamic System Laboratory* at the University of Hawaii at Manoa.

The concept of tranlating these course-specific notes into a more general textbook is the result of working with Professor Allen Downey at Olin College. Professor Downey is a prophet of free books and has authored many which he distributes at Green Tea Press `http://greenteapress.com/`.

If you have suggestions and corrections, please send them to `bsb@hawaii.edu`.

Brian Bingham
Honolulu, HI

Contents

Chapter 1

Digital Measurement

1.1 Telling Tall Tales

The first thing that comes to mind when we mention measurement is often the lowly ruler. Because of its familiarity, the ruler is a nice place to start our conversation about the basic characteristics of measurement.

Imagine you have just caught a large fish from your local pond/lake/stream/ocean. Because you practice catch-and-release (of course) you'll want to measure the length of your catch to avoid an common accusation of telling tall tales. First you'll need to find an appropriately sized ruler to make the measurement. The size of the ruler will dictate the range of the measurement. Were you fishing for Brook Trout or a Blue Marlin? Hopefully you brought along a ruler with a range that exceeds the maximum predicted length of your catch, otherwise you'll have to resort to extrapolation!

Once we lay our ruler alongside our catch we'll need to estimate the actual length based on the graduation marks on the ruler. The distance between these marks is the resolution of our measurement, which is the smallest step size available. The finest marks on a typical yardstick might be one-half inch, so our resulting measurement would have a resolution of 0.5 in. If we have a large set of calipers, the resolution of our measurement would be more like 0.001 in.

Finally, when we are back on land and telling our tale we will inevitably be challenged on the accuracy of our measurement which can attempt to quantify using the error (E)

$$E = \mathrm{measuredvalue} - \mathrm{truevalue}. \tag{1.1}$$

This is a difficult calculation to do without knowing the "true" length of our fish, but we'll return to that topic later.

1.2 Bits and Bytes

In a moment we'll discuss how we might make this measurement using a digital device, but, as an aside, we'll need to introduce the basics of binary numbers. The base-2, or binary, number system represents numbers using two symbols: 0 and 1. This number system is the foundation for all digital electronics, including computers.

Table 1.1: Four-bit binary number example

MSB			LSB	Base-10
0	0	0	0	0
0	0	0	1	1
0	0	1	0	2
0	0	1	1	3
0	1	0	0	4
0	1	0	1	5

One way to describe the binary system is to compare it to the base-10, or decimal, number system that most of us are more familiar with. Each binary digit is called a bit and can be 0 or 1. If we want to be able to count higher than 1, we need to combine multiple bits to represent an integer.

Table 1.1 illustrates how a 4-bit binary number can be used to represent an integer number. For this example we have placed the most significant bit (MSB) on the left side and the least significant bit (LSB) on the right so that the number is read similar to how we write decimal numbers (the 10's place goes to the right of the 1's place).

Exercise 1.1 *What is the largest decimal integer (base-10) that can be represented by a 4-bit binary number? (See Table 1.1) Write the integer in both binary (base-2) and decimal (base-10) forms.*

The number of bits determines the number of integers that can be represented. For a 1-bit number can represent $\{0, 1\}$, a 2-bit number can represent $\{0, 1, 2, 3\}$, a 3-bit number can represent $\{0, 1, 2, 3, 4, 5, 6, 7\}$, etc. Finally, an M-bot number can represent 2^M in integers.

1.3   The Digital Ruler

Now we can return to our fishing expedition. This time, instead of a regular (analog) ruler, we have a digital meter stick. This magical instrument is one meter long (so the full scale range (FSR) is $1.0\,\mathrm{m}$) and it reports the length of the fish as a 3-bit number. Figure 1.1 illustrates our measurement.

Remember, that our digital ruler can only report a 3-bit number. From the illustration we might anticipate that the ruler would report the binary number 100 since that is closest to the actual length of our fish. However, since most of our friends are engineers, they will expect our fishing story to have engineering units. How do we convert this binary number to an actual length?

Since we know the range of our measurement ($1.0\,\mathrm{m}$) and the total number of bits (4) we can convert to a length (L) using a ratio

$$L = FSR\left(\frac{X}{2^M - 1}\right) = (1.0m)(4/7) = 0.57m \tag{1.2}$$

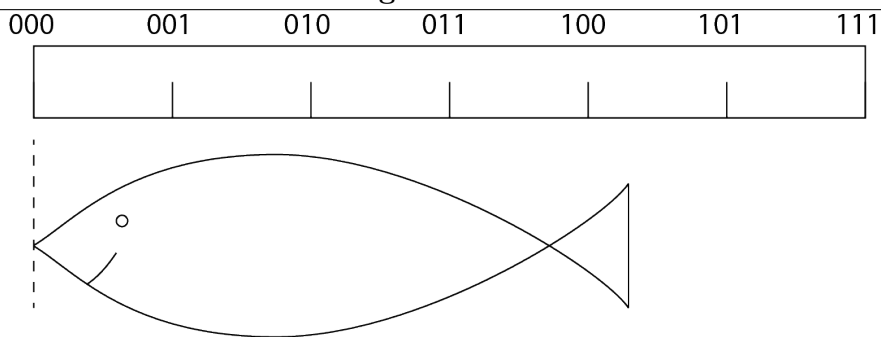where X is the decimal equivalent of our binary measurement (100).

Figure 1.1: Measuring our catch with a digital ruler.

From the illustration it should be clear that the resolution of this measurement is fairly course. In particular, the resolution (Q) can be calculated using the full sale range and the number of bits in our instrument using the relationship

$$Q = \frac{FSR}{2^M} = \frac{1.0\,\text{m}}{2^3} = 0.125\,\text{m}. \tag{1.3}$$

So, in reality we might report our measurement as $0.57\,\text{m}$ with a resolution of $0.125\,\text{m}$. In other words we are saying that the length of our fish is somewhere between $0.51\,\text{m}$ and $0.63\,\text{m}$. So any fish with a length between $0.51\,\text{m}$ and $0.63\,\text{m}$ would generate the same report from our digital ruler - 100.

Exercise 1.2 *If we had a digital ruler with twice the number of bits (6-bits instead of 3-bits) what would be the resolution of our measurement be (with engineering units)?*

1.4   Digital Measurements in an Analog World

Moving on from our tall tale we are now prepared to discuss making measurements using a digital device, for instance a computer or micro-controller. With the ubiquity of computational devices many observations we make today are done with these devices and at the foundation is the concept of an analog-to-digital converter (ADC) or (A/D). The details of how an ADC does it work is beyond the scope of this discussion. What is important is to understand the conceptual idea of converting an analog signal, which can have an infinite number of values within a range, to a digital signal, which is represented by a finite number of levels. The number of levels that this digital signal can take on is specified by the number of bit, just as our 3-bit digital meter stick from Figure 1.1 could only report 8 individual values.

To begin our discussion of digital measurements let's consider a simplified case where an analog voltage comes in from the outside world (from a sensor) and the ADC converts the voltage to a binary number. As an example let's consider an ADC with the following properties:

- FSR = 0–10 V
- 8-bit resolution

The job of the ADC is to convert the incoming voltage and convert the value to a binary number that most closely represents the value as illustrated in Figure 1.2 An important consideration of this
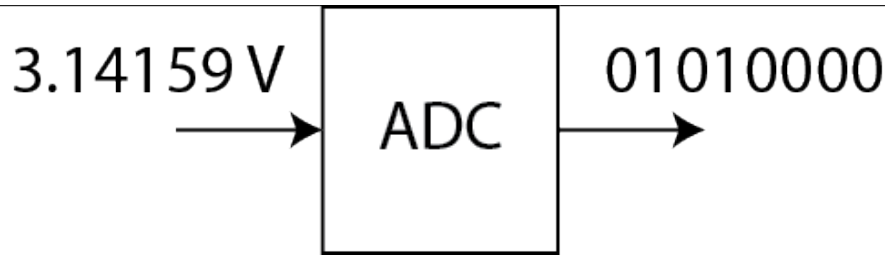
Figure 1.2: Analog-to-digital conversion: analog voltage input and binary number output.

ADC example is the resolution of the measurement, also known as the quantization error, which can be calculated using the same expression as before:

$$Q = \frac{FSR}{2^M} = \frac{10.0\,\text{V}}{2^8} = 39.1\,\text{mV} \tag{1.4}$$

Because the ADC has 8-bits of resolution, it will report a binary number that is one of 256 possible levels. Each level is separated by 39.1 mV so that the total range is 10.0 V. If the actual input voltage is 3.14159 V the ADC will report the closes binary level that correspond to this measurement which is 01010000 (80 in decimal form). So the ADC output, converted back to voltage, would be 3.125 V ± 0.0196 V, where 0.0196 V is $Q/2$.

Now if it wasn't bad enough that our 8-bit ADC has such coarse resolution, we also have to contend with the fact that the input voltage might be much smaller than the incoming range. Consider what would happen if the actual input voltage only range between 0–0.1 V, instead of the 0–10.0 V range of the ADC. In this case the ADC would only have four possible outputs corresponding to 0, 0.0391, 0.0782 and 0.1173 V. So, in effect we would only be using 2-bits of our 8-bit ADC!

Exercise 1.3 *Consider an analog-to-digital converter with the following properties:*

- *FSR = ± 1.0 V*
- *12-bit resolution*

What is the resolution (in Volts) of the ADC? If the input to the ADC was 3.14 V, what binary number would be output by the ADC?

1.5   Summary

The point of all this is the following:

- Digital representations of analog measurements have only finite resolution.
- The input signal must be well matched to the ADC range to effectively use the full resolution of the ADC.

So far we've just considered the simple idea of converting static analog measurement to a digital representation. In the next chapter we'll consider what happens when we consider another dimension—time—in our understanding of measurement systems. We'll also look at how to couple the ADC component with the rest of our system to make quality observations.

1.6   Exercises

Exercise 1.4 *Visit your favorite on-line vendor and search for a computer (PC or Mac) sound card. Report the following specifications:*

- *Make and Model*
- *Price*
- *Audio input resolution (in bits)*
- *Audio input range (in Volts)*
- *Maximum sample rate (in Hz)*

1.7   Glossary

Notation	Description	Symbol	Page List
accuracy	The agreement between the measured value and the actual (true) value.		1
analog-to-digital converter	A device for the transforms an analog signal, with infinite resolution, to a binary number, with finite resolution. symbol		3
binary numbers	Integer numbers represented using the base-2 number system.		1
bit	A single binary digit. A bit can be either 0 or 1.		2
error	A quantification of the inaccuracy in a measurement or set of measurements. Often used as a synonym for accuracy.		1
range	The span of measurements for a particular instrument. Also known as the *full-scale range* (FSR)		1
resolution	The smallest increment of the measured value that can be discerned using a given instrument.		1

Chapter 2

Data Acquisition

2.1 Data Acquisition Systems

Figure 2.1 illustrates a typical data acquisition system which consists of the following components:

- Sensor(s): Device for converting a physical quantity into an analog signal, e.g., accelerometer, thermocouple or strain gage.
- Signal Conditioning: Electronics for adapting sensor output to comply with the input of a DAC, e.g., filtering, Wheatstone bridge or amplification.
- Digital-to-Analog Converter (DAC): See Section 1.4.
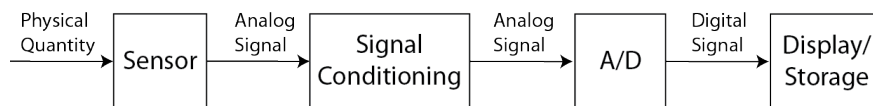- Destinations: Digital display (e.g., 2D graph) or archive (e.g., log file).

Figure 2.1: Simple data acquisition system diagram.

In this chapter we'll work through each of the subsystems in Figure 2.1 to discuss how they operate together to create a system for making digital measurements.

2.2 Sensors

A transducer is a device that converts one type of energy into another type of energy. A sensor is a transducer, but not all transducers are sensors. (The opposite of a sensor is an actuator). Here are some examples:

- A microphone is a transducer that converts acoustic energy (sound pressure) to electrical energy (voltage). A microphone is a sensor.
- A speaker is a transducer that converts electrical energy to acoustic energy. A speaker is an actuator.
- An accelerometer is a transducer that converts mechanical energy (vibration) to electrical energy (voltage). An accelerometer is a sensor.

- An electric motor is a transducer that converts electrical energy to mechanical energy (rotation). An electric motor is an actuator.
- A thermocouple is a transducer that converts thermal energy (temperature) to electrical energy (voltage). A thermocouple is a sensor.
- A heating coil is a transducer that converts electrical energy to thermal energy. A heating coil is an actuator.

Exercise 2.1 *Following the pattern in the list above, list two more transducers that are sensors and two more transducers that are actuators.*

There are many, many types of sensors. For just about everything we can imagine wanting to measure there is a device that will convert the physical parameter to an electrical signal that we can use for data aquisition. For our purposes we will focus on sensors that measure mechanical quantities (motion, strain, temperature, etc.) by producing an electrical output. The next step is converting that electrical output into a form that our D/A can accept.

2.2.1 Sensitivity

The sensitivity of a sensor is the parameter of the sensor that describes the transduction. In simple cases the sensitivity of a sensor is given as a ratio between the output (typically voltage) and the physical input. For example, the ADXL 335 accelerometer has a typical sensitivity of $300\,\mathrm{mV/g}$ which means that for an input acceleration of $1.0\,\mathrm{g}$ the sensor output will be $300\,\mathrm{mV}$.

Exercise 2.2 *Suppose we are going to use an ADXL 335 accelerometer to measure the acceleration of a radio-controlled car. We know that the car has a maximum forward acceleration of $2.5\,\mathrm{m/s^2}$. What is the maximum output voltage we could expect from the accelerometer?*

Exercise 2.3 *A Type K thermocouple is a common sensor for measuring temperature, and the typical sensitivity for such a device is $41\,\mathrm{\mu V/^\circ C}$. What output voltage could we expect if we were using this device to measure the temperature of boiling water (at sea-level)?*

A complication that can arise is when the sensitivity of a sensor is not a constant. This can happen if there is a non-linear relationship between the physical parameter and the output. In this case we may have to use a mathematical relation (i.e., a formula, hopefully supplied by the sensor datasheet) to convert the output signal (voltage) to the physical measurement. Another situation is when the output depends on the supplied voltage, which is called a ratiometric sensitivity. In fact, I oversimplified the ADXL 335 sensitivity. This sensor requires a supply voltage, sometimes referred to as the excitation voltage, to power the sensor. The typical supply voltage is $3.0\,\mathrm{V}$, but the value can range between 1.8–$3.6\,\mathrm{V}$. The sensitivity is directly proportional to this supply voltage, so with a supply of $1.8\,\mathrm{V}$ we can expect a sensitivity of $180\,\mathrm{mV/g}$.

Another common example of a ratiometric sensitivity is a load cell. As its name implies, a load cell is a sensor that reports a load (tension or compression) as a voltage. A common type of load cell is S-Beam load cell that uses a configuration of strain gages to measure the strain in a mechanical beam. The beam is calibrated so that the relationship between the strain and the applied force is known. The sensitivity of this type of sensor is specified in a confusing manner. For example, a typical device will come in a range of load capacities, say 25-10,000 lb. For this family of devices

the sensitivity might be quoted as $3\,\mathrm{mV/V}$. What on earth does this mean you ask? Well, it means that at the maximum load, the output voltage is $3\,\mathrm{mV}$ multiplied by the excitation voltage. They typical excitation voltage of this particular device is $10\,\mathrm{V}$. So, if we have a load cell rated at $100\,\mathrm{lb}$ and we apply the recommended excitation of $10\,\mathrm{V}$, the sensitivity is $0.3\,\mathrm{mV/lb}$. In other words, we can find the sensitivity ($S\,\mathrm{mV/lb}$) as

$$S\,\mathrm{mV/lb} = \frac{(O\,\mathrm{mV/V})(E\,\mathrm{V})}{L_{max}\,\mathrm{lb}}$$

where $O\,\mathrm{mV/V}$ is the ratiometric output of the load cell, $E\,\mathrm{V}$ is the excitation voltage and $L_{max}\,\mathrm{lb}$ is the maximum load capacity for the device.

Exercise 2.4 *Suppose we want to use two AA alkaline batteries to power our ADXL 335 accelerometer. What sensitivity could we expect with this arrangement?*

Exercise 2.5 *Consider that we are using a thin-beam load cell (Omega LCL Series) with a maximum load capacity of 1 lb. The rated output of the load cell is $2\,\mathrm{mV/V}$ and the recommended excitation is 5 V.*

- *What is the sensitivity of the load cell in $\mathrm{mV/lb}$?*
- *What is the maximum voltage we could expect when the load cell is at its maximum capacity load?*
- *What is the voltage we could expect if we apply a 0.25 lb load?*

2.3   Signal Conditioning

Sensors often have an output that is inconvenient to use with digital-to-analog converters. As we alluded to in Chapter 1, digital-to-analog converters typically take a voltage as input and it is important that we use the full range of the voltage that the D/A will accept (e.g., 0–10 V).

Consider the case of a strain gage, illustrated in Figure 2.2. The specifics will be discussed in more detail later, but for now it will suffice to say that a strain gage is a sensor that converts a strain (a change in length) to a change in electrical resistance. This is great, but how do we plug that into our D/A which wants a voltage between 0–10 V? It turns out that Sir Charles Wheatstone found a solution to this problem in an elegant circuit that is now called the Wheatstone bridge. This circuit converts the change in resistance to a very small voltage, say 0–10 mV. Now we have another problem, our D/A wants a much larger voltage that the 10 mV coming from the Wheatstone bridge. Perhaps we could add an amplifier that will multiply the output of the bridge by a gain of 1,000 so that we now have a voltage that is suitable for the D/A.
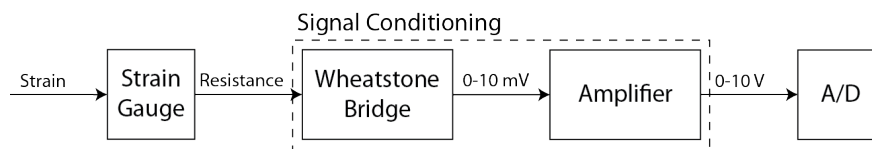

Figure 2.2: Signal conditioning example. In the case of the strain gage, a Wheatstone bridge and amplifier are used to condition the signal before it is passed to the analog-to-digital converter.

For our purposes two general types of signal conditioning will play an important roll in designing our measurement systems, amplification and anti-aliasing. For now we'll discuss amplification and we'll save the topic of anti-aliasing for Section 2.4.5.

2.3.1   Amplifiers

The roll of an amplifier in our signal conditioning system is to scale the electrical output of a sensor (typically voltage) to match the input range of the analog-to-digital converter. As we discussed in Section **??**, in order to take full advantage of the resolution of our A/D, the signal range must be approximately equivalent to the input range of our A/D. In theory (always a dangerous first two words), the function of an amplifier is to provide a multiplicative gain in the signal path, i.e.,

$$V_{out} = K\left(V_{in}\right)$$

where V_{in} is the input to the amplifier, V_{out} is the amplifier output and K is the amplifier gain.

Exercise 2.6 *Building on Exercise **??** consider that we want to interface our ADXL 335 into an A/D that has a range of 0–10 V and 12-bits of resolution.*

- *If we did not use any amplification, what would be the effective resolution (in bits) of our A/D? (Hint: See Section 1.4.)*
- *What amplifier gain would you suggest so that we could use the full resolution of the A/D?*

2.4   Analog-to-Digital Conversion

We started our discussion of A/D devices in Section 1.4. Here we'll add some detail to our description of these devices. Again, we'll focus on *what* these devices do and leave the specifics of *how* they do this (successive-approximation, sigma-delta, etc.) to another book.

2.4.1   Range and Resolution

Recall that the function of an ADC is to convert an voltage input (analog) to a binary number (digital). The full scale voltage range (E_{FSR}) is the span of voltage inputs that the A/D will accept. This range comes in two flavors: bilateral or unilateral. A bilateral input range is typically symmetric about 0 V while a unilateral input can only accept positive voltage values. For example,

- Bilateral Input Range: $\pm 10\,\mathrm{V}$
- Unilateral Input Range: 0–10 V

In either case the full scale range is simply $E_{\mathrm{FSR}} = V_{\max} - V_{\min}$ where the maximum input voltage is $V_{\max}$ and the minimum input voltage is $V_{\min}$.

Exercise 2.7 *For the bilateral and unilateral examples above what is E_{FSR} for each case?*

Any voltage within the input range is converted to a binary value. The length of this binary number is specified by the number of bits (M) inherent to the A/D, e.g., 12-bits. The length of the binary representation determines the finite number of voltage levels that can be represented (see

Section 1.4), so the ratio of the full voltage range to the number of levels that we can use in that range determines the resolution Q (in Volts) of the A/D,

$$Q = \frac{E_{\text{FSR}}}{2^M}. \tag{2.1}$$

You might notice that the resolution of an A/D is often referred to as either the number of bits used in the conversion (M) or the resolution in volts (Q). This is somewhat sloppy nomenclature, but common.

A related concept is quantization error ("ensuremath –"pm Q/2") which is another way of describing the finite resolution of converting a continuous analog value to a discrete digital value. As we discussed in Section 1.4, the A/D will associate the incoming voltage level (which can assume an infinite number of possible values) with the binary representation (which can only assume one of 2^M possible values) that is nearest that actual value. Based on this nearest-neighbor concept the error between the actual incoming analog voltage and the reported digital value will be $\pm Q/2$, the quantization error.

Exercise 2.8 *An alternative to (2.1) is to express the resolution as*

$$Q = \frac{E_{\text{FSR}}}{2^M - 1}.$$

- *Can you explain why this might be more correct? Try drawing a picture of a 1-bit or 2-bit scenario to understand why we might actually only have $2^M - 1$ levels in our A/D.*
- *Given that typical A/D hardware has more than 10-bits, and as much as 24-bits, of resolution, is the difference between these alternatives significant? Why or why not?*

2.4.2 Differential Versus Single-Ended

When connecting a voltage input to an A/D there are two important options: differential or single-ended input. Single-ended input consists of a common ground (GND) that is shared between all the inputs to the A/D. The A/D output is based on the voltage measured between the signal input (+) and ground. In contrast differential input consists of two voltage inputs, high (+) and low (-), which may or may not be referenced to a common GND potential. The A/D output is then based on the voltage difference (hence differential) between the signal high and signal low potentials. Here are some advantages of each approach:

- Single-ended inputs are simpler to implement
- Single-ended inputs use the full number of channels in the A/D device. For example, a device advertised to have "16 inputs" will actually have 16 single-ended inputs, but only 8 differential inputs.
- Single-ended inputs are lower in cost
- Differential inputs are less susceptible to noise (electromagnetic or radio-frequency interference). This is particularly important for sensors that produce very small signals, e.g., strain gages, thermocouples, load cells, etc.
- Differential inputs can be applied in cases (such as Wheatstone bridge outputs) where there is no true connection to a common ground.

2.4.3   Sampling

So far we have been concerned with the resolution of a digital measurement in terms of the finite levels, typically of voltage, that an ADC can use to represent the analog input—the y-axis in Figure 2.3. In this section we'll discuss the ramifications of the fact that the ADC operates at a finite sampling period (dt), the output of the ADC changes only at discrete times. Therefore, an ADC has finite resolution in both amplitude (i.e., voltage) and time, illustrated by the fact that the individual samples in Figure 2.3 are discrete points in the x and y axes.
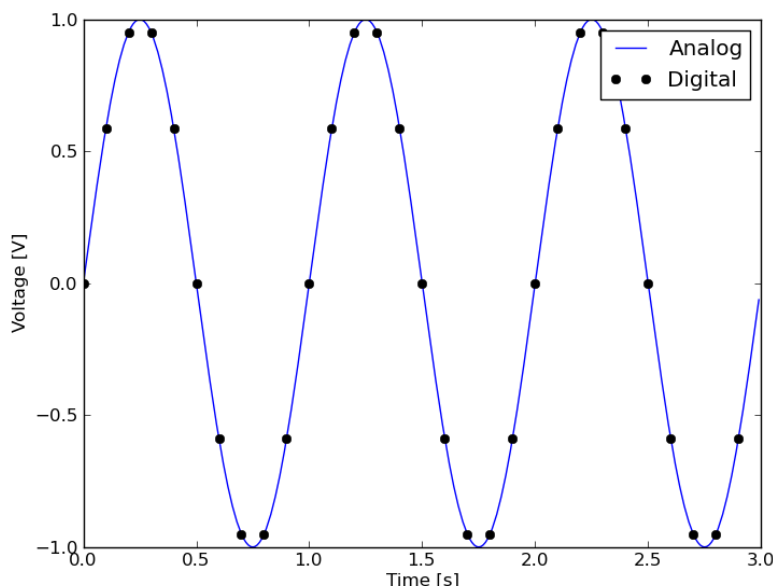

Figure 2.3: Illustration of a sampled signal. The analog signal ($y(t) = \sin \omega t$) is represented by a continuous curve with an infinite set of values in both time (x-axis) and amplitude (y-axis). The A/D converts this signal discrete samples at a finite set of time and amplitude values.

The typical function of an A/D is to regularly sample the input voltage level and record these samples as digital numbers. The speed at which this happens is specified by the sampling rate, or equivalently the sampling frequency (f_s) where $f_s = dt$ Hz. (Beware, the terms "sample rate" or "sampling rate" are often used to refer to both sampling period and sampling frequency.) The sampling is illustrated in Figure 2.3 by the black markers; the time between each sample ($dt = 0.1$ s) is the sampling period.

Exercise 2.9 *Consider the same analog input as illustrated in Figure 2.3 and Figure 2.4, a continuous sine wave input with a period of 1.0 s and an amplitude of 1.0 V. The signal is input to a A/D with the following characteristics:*

- *4-bit voltage resolution*
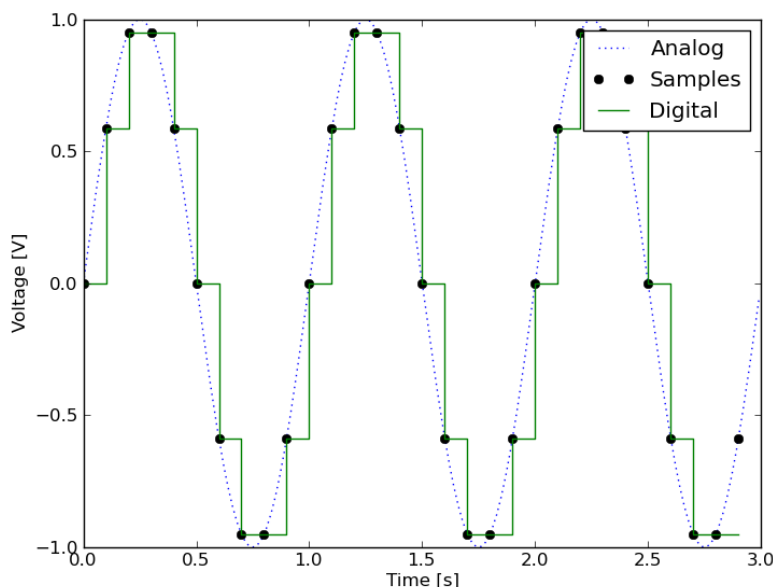- *$E_{\mathrm{FSR}} = \pm 1.0$ V*
- *$dt = 0.5$ s*

Figure 2.4: A second illustration of the sampled signal from Figure 2.3. The "Digital" graph shows how the digital output of the A/D changes over time. The reported value from the A/D only changes after each sampling event.

Sketch a 2D graph (Voltage versus Time) that illustrates the analog signal and the digital samples.

Sketch the same 2D graph if $dt = 0.33$ s.

2.4.4   How Fast to Sample?

The sampling rate (see the warning above!) for A/D devices can vary by orders of magnitudes. Consider the contrast between two processes that we might want to measure: the rise of global temperature or the incoming voice signal from an iPhone user. These processes evolve on very different time-scales. To measure global temperature rise it might be sufficient to take an average measurement once per year, but a typical sampling frequency for audio input is 44,000 samples-per-second (Hz).

Exercise 2.10 *Suppose we wanted to sample the global temperature at a sampling frequency of 44kHz using a 16-bit A/D for a year. Estimate the size of the hard drive (in bytes) that would be necessary to store this much data. (Hint: There are 8-bits per byte.)*

The sampling frequency (or period) can often be specified in software when using a ADC. So what setting should we choose? The simple answer is that we typically sample as fast as we can!

The longer (and hopefully more interesting) answer is that we can set a minimum necessary sampling frequency based on the characteristics of the input signal. Let's start with our audio (iPhone) example. The key thing you need to know is that the typical range of human hearing is from

20–20,000 Hz. So we'll need to sample fast enough that our digital representation (the output of the ADC) can represent this complete frequency range. It turns out that if we collect samples at *twice* the maximum frequency of interest, then we will be able to represent the analog signal with our digital representation, i.e.,

$$f_s \geq 2\left(f_{\max}\right)$$

where $f_{\max}$ is the Nyquist frequency, the maximum frequency of interest in the analog signal we are trying to measure. The Nyquist frequency dictates the theoretical minimum required sampling frequency, but in practice we can often sample much faster than this minimum; a typical rule-of-thumb is to set $f_s = 10(f_{\max})$

Exercise 2.11   *What would be the minimum sampling frequency for the analog signal illustrated in Figure 2.3? What would be the cooresponding Nyquist frequency?*

2.4.5   Anti-Aliasing

What happens when we violate this sampling rule-of-thumb? Let's work through an illustrative example similar to the illustration from Figure 2.3. Considering our simple sine wave input that changes amplitude with a frequency of 1.0 Hz. If we use our rule-of-thumb and set the sampling frequency to $f_s = 10\,\text{Hz}$ we will end up with a digital representation, illustrated in Figure 2.5, that looks pretty much like the original analog signal. *If we sample "fast enough" the digital output will be a close approximation to the analog input.*
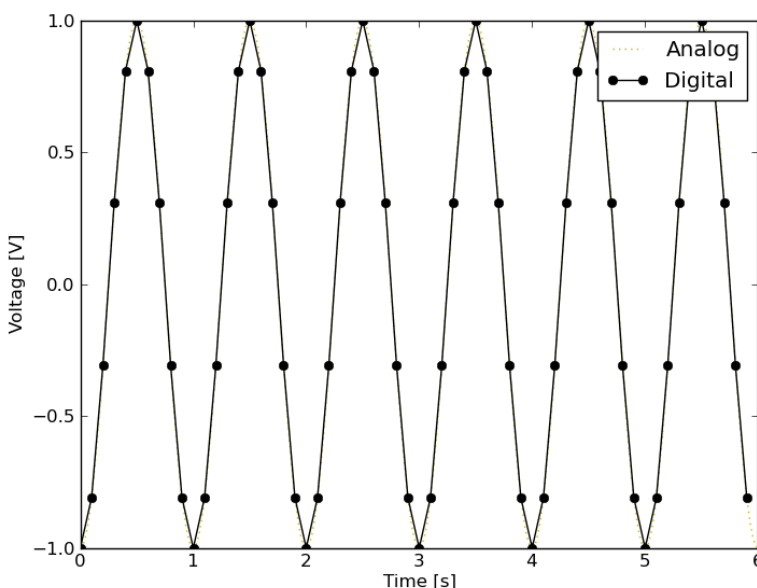

Figure 2.5: A second illustration of the sampled signal from Figure 2.3. The "Digital" graph shows how the digital output of the A/D changes over time. The reported value from the A/D only changes after each sampling event.

Now let's see how things change if we cut it close and use the minimum frequency ($f_s = 2\,\text{Hz}$), i.e., set the sampling frequency equal to twice the Nyquist frequency. Figure 2.6 illustrates that the sampled signal still roughly resembles the original sine wave, but things are starting to look a bit fishy.
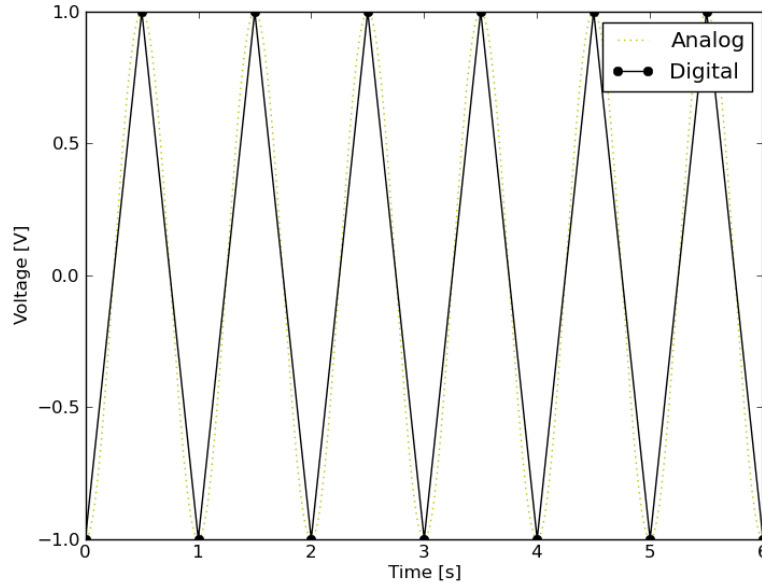

Figure 2.6: Sampling at the minimum frequency, twice the Nyquist frequency: $f_s = 2\,\text{Hz}$, $dt = 0.5\,\text{s}$.

Finally, the sampling frequency becomes less than twice the Nyquist frequency the digital output is significantly different than the analog input. Figure 2.7 illustrates that the digital output appears to be a sawtooth pattern with a much slower frequency of oscillation than the analog input, the term for this phenomenon is aliasing.

To generalize from our specific example, aliasing is when the analog input has significant frequency content above the half the sampling frequency. These portions of the signal are folded back into the digital representation as lower frequency components that can corrupt the digital representation. This can be a particular problem when there is high frequency noise. In this case aliasing can cause high frequency noise to corrupt your lower frequency data.

To address this issue an anti-aliasing filter is often included in the signal conditioning portion of the data acquisition system. This analog low-pass filter attenuates the higher frequency portions of the analog input, those portions of the input above the Nyquist frequency, to prevent these high frequency portions of the input from corrupting the measurement.

Exercise 2.12 *Anti-aliasing filters must be implemented on the analog signal. However, digital filters can be used on the digital measurement (after the A/D conversion). Why is it not possible to have a digital anti-aliasing filter?*
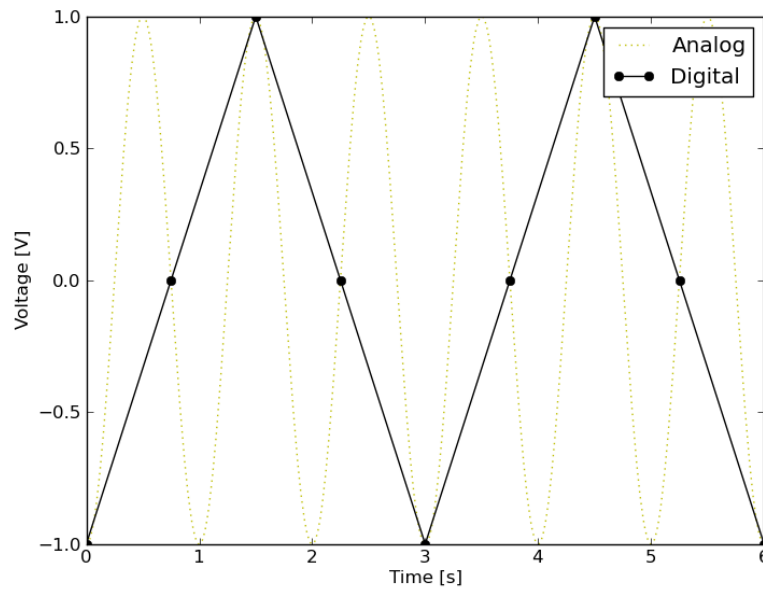
Figure 2.7: Sampling below twice the Nyquist frequency: $f_s = 1.33\,\mathrm{Hz}$, $dt = 0.75\,\mathrm{s}$.

2.5 Glossary

Notation	Description	Symbol	Page List
aliasing	The result of having energy in the input to an A/D with frequency content higher than the Nyquist frequency. These higher frequencies are aliased into the digital representation to cause errors.		15
data acquisition system	Assembly of sensors, signal conditioning, digital-to-analog converter and a display or storage device for collecting physical measurements as digital records.	DAS or DAQ	7
differential input	the input configuration of an A/D where the output measures the difference between two voltages (+ and -) that are independent of electrical ground.		11
excitation voltage	The constant voltage supplied to power a sensor, typically a strain gage or Wheatstone bridge.		8
full scale voltage range	The specified input voltage range to an ADC. The range can be bipolar (both positive and negative, e.g., $E_{\mathrm{FSR}} = \pm 1.0\,\mathrm{V}$) or unipolar (only positive, e.g., $E_{\mathrm{FSR}} = 0--1.0\,\mathrm{V}$.) symbol		10
Nyquist frequency	For the output of an A/D the Nyquist frequency is the highest frequency of the signal that can be observed in the digital version of the signal. Because the Nyquist frequency is half of the sampling frequency, it is important to sample sufficiently fast to capture all the dynamics in the input.		14
quantization error	the difference between the actual analog input to a analog-to-digital converter and the reported digital (or quantized) value.	$\pm Q/2$	11
ratiometric	Refers to the situation where the output (voltage) is a ratio of the supply or excitation voltage.		8

Notation	Description	Symbol	Page List
sampling frequency	For an A/D, the frequency of sampling events which is the reciprocal of the sampling period, typically in units of Hz.	f_s	12
sampling period	For an A/D, the duration in time between sampling events in units of time, e.g, seconds.	dt	12
single-ended input	the input configuration of an A/D where the output measures the voltage of a single voltage (+) relative to a common electrical ground (GND).		11
transducer	A device for converting on type of energy to another type of energy. Examples include both sensors and actuators.		7

Chapter 3

Dynamic Models: Part 1

3.1 All Models Are Wrong!

For our purposes a *model* is equivalent to a first-order model, an equation. We can use to understand and design a *system*, a physical system which we wish to analyze. Over the last few hundred years people with last names like Newton, Maxwell, Euler, Lagrange and Bernoulli have come with with a variety of simple equations that describe physical processes. The key point, that is often overlooked, is that this description is only approximate; the mathematical model and the physical system never share the same behavior.

So why do engineers spend so much time and energy learning mathematical equations and how to solve them if these models never completely capture the behavior of a physical system? It turns out that under specific conditions, these models allow us to *predict* the behavior of systems that we might want to build.

As an example, consider the Hoover Dam, a 726 ft tall structure constructed with over 3 million cubic yards of cement. In designing this structure we would hope that engineers spent a large amount of time on structural models to attempt to predict if their designs would last through floods, earthquakes and any situation that might arrive. Without any models they would be have to fall back on the build-and-test process of engineering, building a prototype dam and sitting back to see if it worked; obviously not a very safe or cost effect way to build a dam. On the other hand, we have to imagine that the engineers included some conservative factors of safety to build the structure somewhat larger and stronger than their models predicted. They did this because they understood that the model was only an approximation of the dam; the model was wrong, but still useful.

So, even though models are wrong (they are always simplified approximations), they are useful for the following reasons:

- under certain specific conditions mathematical models can predict (approximately) the behavior of physical systems
- analysis of the model behavior can be an efficient way to test the behavior of system designs before we go to the trouble (cost and time) to build the physical system
- we can analyze the behavior of models in situations that would be hard or impossible to

create in the physical system
- the model analysis enables us to understand the system and build intuition about how a system behaves

3.2   First-Order Model

Our first-order model, a first-order, linear, ordinary differential equation, is

$$\frac{dy(t)}{dt} + \frac{1}{\tau}y(t) = f(t). \tag{3.1}$$

where $f(t)$ is the input (forcing function), $y(t)$ is the output (response function) and τ is the time constant. That's it; there is really nothing more to say.

But of course we do have much more to say. I want to start by emphasizing that this model is just as wrong as the rest of them (but it can be very useful). There are two reasons that this model is useful:

1. The time-response of this model is sufficiently similar to that of many physical systems to allow us to use this fictitious model (equation) to predict how an actual system might behave.

2. We know (or we will know shortly) how to solve this equation to calculate the output of the model in response to a variety of inputs.

3.2.1   An Example

One example of a physical system that can be approximated by our first-order model is a automobile. Let's consider how the speed of an automobile (the output) responds to changes in the gas petal (the input) as it moves in a constant direction. The sketch if Figure 3.1 illustrates this simplification of a complex physical system. Based on this sketch we can do things like draw a free-body-diagram and
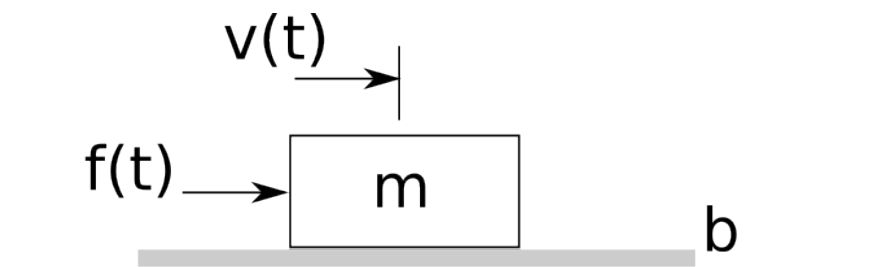

Figure 3.1: Sketch of a simplified automobile model. The point mass (m) moves to the right at a velocity ($v(t)$). The input is a force ($f(t)$). There is a drag force that resists the motion.

apply Newtonian principles to arrive at an equation of motion. We might also make an ambitious simplifying assumption that the drag force that resists the motion of the mass is linearly related to the velocity, i.e., $f_{\mathrm{drag}} = b(v(t))$. We could come up with an equation of motion, our mathematical model,

$$m\left(\dot{v}(t)\right) + b(v(t)) = f(t) \tag{3.2}$$

where m is the mass of the car, $v(t)$ is the speed, $\dot{v}(t)$ is the acceleration, b is the coefficient of linear drag and $f(t)$ is the input force. This equation has the same form (a first-order, linear, ordinary differential equation) as our first-order model (3.1).

This is meant to be an example of a model that is obviously wrong. A physical automobile is a complex system with many, many degrees of freedom. Our simplified model (3.2) is meant to do just one thing, allow us to predict how, in general, the speed of a car responds to changes in the throttle input.

3.3   Step Response

One useful aspect of our first-order model (3.1) is that the solution to this differential equation is well known for a variety of input functions. A particularly interesting input function is the step input ($\mu(t)$) defined as

$$\mu(t) = \begin{cases} 0 & \text{for } t < 0 \\ 1 & \text{for } t \geq 0 \end{cases}. \tag{3.3}$$

Figure 3.2 illustrates how the value of $\mu(t)$ is zero before $t = 0$ and then instantaneously changes to unity.
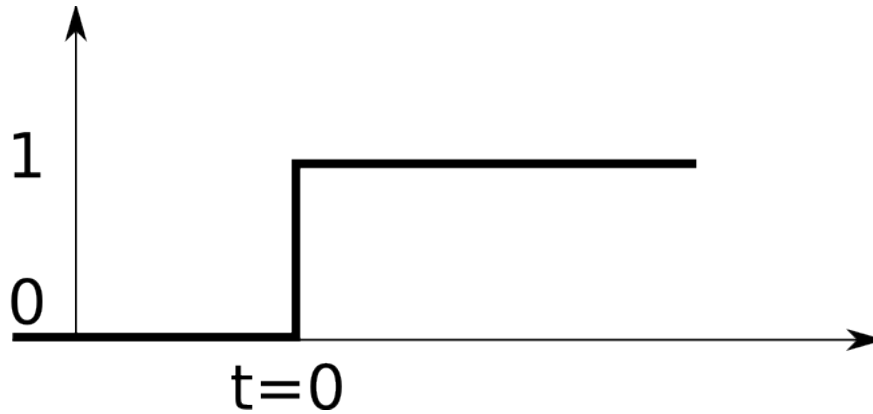

Figure 3.2: Illustration of a step function $\mu(t)$.

The step response of our first-order model is the solution to (3.1) with step function as the input ($f(t) = A\mu(t)$) and an initial condition of zero ($y(0) = 0$). (You should notice that the step function here has an amplitude of A rather than an amplitude of 1 as indicated in (3.3).) The step response is the well known solution

$$y(t) = A\tau \left(1 - e^{-t/\tau}\right) \tag{3.4}$$

for $t > 0$.

Listing 3.1: code

```
% Example of time−response of first−order system
% First Order Model: dy/dt = 1/tau*y = f(t)
tau = 2;  % time constant [s]
% A vector of time values for evaluating the response
```
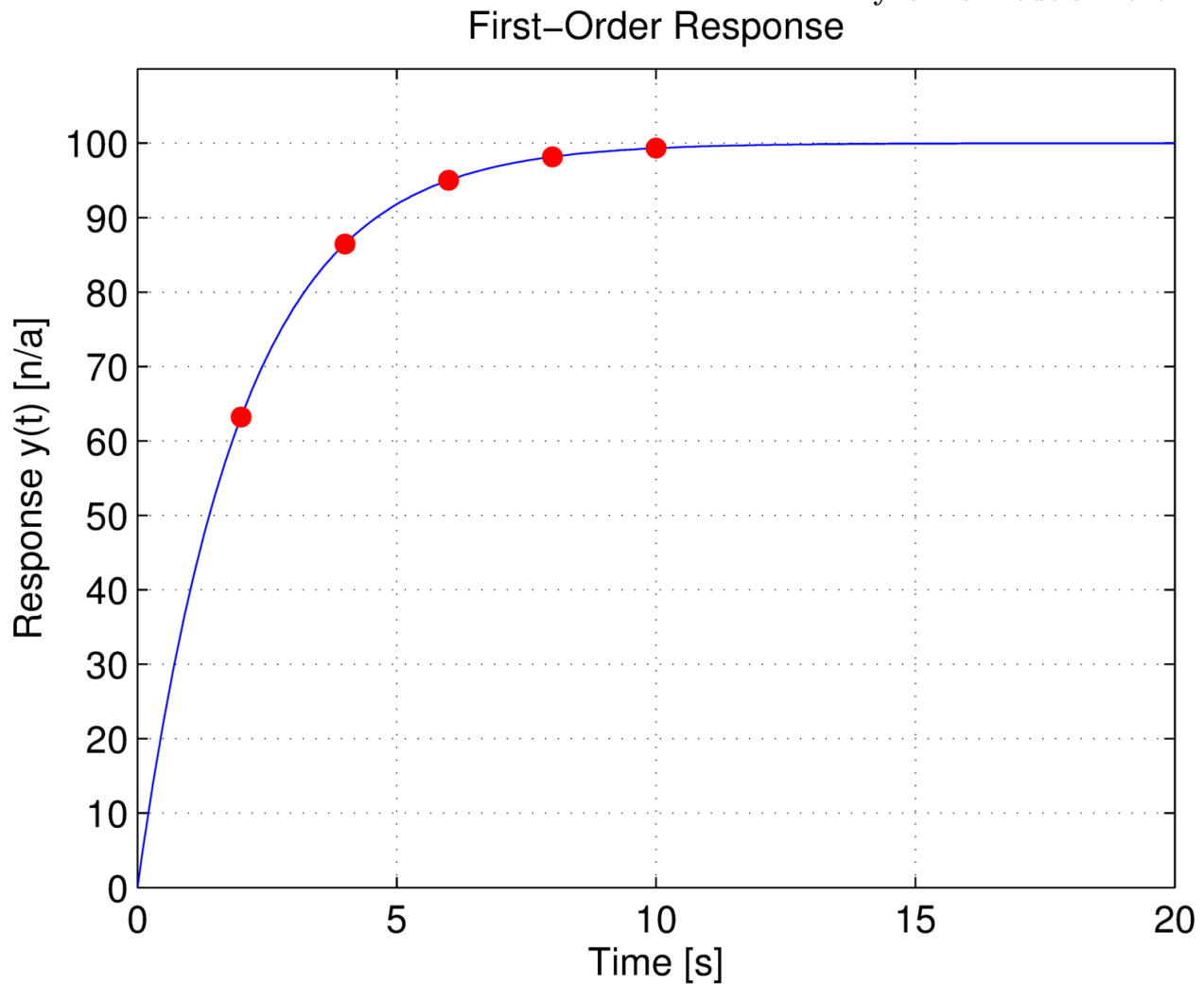
First–Order Response

Figure 3.3: Graph of the step response (3.4) with $\tau = 2\,\mathrm{s}$ and $A = 50$. The red dot markers indicate the value of the solution at $t = [\tau, 2\tau, 3\tau, 4\tau, 5\tau]$.

```
t = 0:0.1:20;
% Step Response: f(t)=A*mu(t) , y(0)=0
A = 50;
% Evaluate the solution at each value of t
y = A*tau*(1-exp(-t/tau));
% Plot
figure(1); clf();
plot(t,y,'-')
xlabel('Time [s]')
ylabel('Response y(t) [n/a]')
title('First-Order Response')
grid on
ylim([0 max(y)*1.1])
```

This solution is graphed in Figure 3.3 using the MATLAB script in Listing **??**. This time response to a step input (or step response for short) has a characteristic shape as it rises and approaches the steady-state response exponentially. One important aspect of the response is that it approaches a new constant value, the steady-state response. A simple way to find the steady state value is to set $\dot{y}(t) = 0$ in (3.1). Since we are looking for the "steady-state" value we can anticipate that all rates of change (derivatives) will be zero. Now the steady-state value (y_{ss}) is

$$y_{ss} = \tau A \tag{3.5}$$

Another important characteristic of this model is that the response ($y(t)$) moves exponentially from the initial value to the steady state value. This response is characterized by the the first-order model (τ). If we evaluate time response (3.4) at $t = \tau$ we find that

$$y(t = \tau) = A\tau(0.632) = 0.632(y_{ss})$$

which means that after one time constant has passed the response is 63.2% of the way from the initial value to the steady state value. Take a look at Figure 3.3 to make sure that we did the math correctly!

3.3.1   An Example

Now that we've discussed the time response of our first-order model let's see how this applies to our automobile model. We can rearrange the mathematical model (3.2) so that it looks more like our generic model (3.1)

$$\dot{v}(t) + \frac{1}{m/b}(v(t)) = \frac{F}{m}\mu(t)) \tag{3.6}$$

which highlights that the time constant is $\tau = m/b$ and that the steady state speed is $v_{ss} = F/b$. Take a moment to think about this. It means that the larger the mass of our car, the slower the acceleration (because the time constant is larger). You might also notice that the mass has no effect on the steady state speed; our model says that a heavy "car" will reach the same final speed as a light car, it will just get there slower. Finally we might notice that the more drag in our model, the slower our "car" will go for the same input.

Exercise 3.1   *The step function in (3.3) could be more precisely called the "unit step function" because it has an amplitude of 1. Write an equation (similar to (3.3)) and sketch a graph (similar to Figure 3.2) for the more general step function input $f(t) = A(\mu(t - t_o))$.*

Exercise 3.2   *We justified (3.5) by starting with the model (3.1). Another way to calculate the steady state response is to evaluate the solution (3.4) as $t \to \infty$. Show that using this method yields the same result.*

Exercise 3.3   *We showed that the response of a first-order model to a step function input has an exponential increase characterized by the time constant. Furthermore we showed that after a duration of one time constant beyond the rise in the step input the output will be 63.2% of the way to its steady state value. Figure 3.3 illustrates the value of the response at $t = [\tau, 2\tau, 3\tau, 4\tau, 5\tau]$. Evaluate 3.4 for these values of time and report the results in a two column table where the first column is the ratio t/τ and the second column is the ratio $y(t)/y_{ss}$.*

Exercise 3.4 *Consider the step response to our car model (3.6) with the following parameters: mass = 750 kg, drag coefficient = 150 Ns/m, step input amplitude = 10,000 N. Create a graph similar to Figure 3.3 to illustrate the step response of our "car". Annotate the graph to show the time constant and the steady state speed. Also, make sure to use appropriate units for each axis.*

Exercise 3.5 *Consider the step response to our car model (3.6) with the following parameters: mass = 750 kg, drag coefficient = 150 m/s.*

- *What is the minimum step input amplitude ($F_{\min}$ that would cause our "car" to go from 0–60 mph?*
- *Based on this minimum step input amplitude, how long does it take for the "car" to go from 0–60 mph?*
- *If we double ($F_{\min}$*
  - *How long does it take for the "car" to go from 0–60 mph?*
  - *What is the new top speed (in mph)?*
  - *How long does it take for the "car" to achieve this new top speed?*

3.4   Free Response

Another useful time response of our first-order model is the response with no input ($f(t) = 0$) and an initial condition ($y(0) = y_o$). The solution to (3.1) under these conditions is

$$y(t) = y_o \, e^{-t/\tau} \tag{3.7}$$

for $t > 0$. Figure 3.4 illustrates this free response. Just as before the time constant is the key factor that determines the characteristic of the response. Again, at $t = \tau$ the output has fallen by 63.2%.

Exercise 3.6 *Using our automobile model (3.2), write an expression for the free response of this model with appropriate physical parameters. Create a graph of the free response with the initial condition $v(0) = 60$ mph. For the x-axis of the graph use the units mph and for the y-axis of the graph use the units seconds.*

3.5   Superposition

One the characteristics that make linear system such as our first-order model useful is the principle of superposition which allows determine the solution to a linear model by adding together multiple solutions. As an example, suppose there are two input functions, $f_1(t)$ and $f_2(t)$, and that the corresponding solutions are $y_1(t)$ and $y_2(t)$. If we want to know the response of the system to both inputs, $f(t) = f_1(t) + f_2(t)$, we can simply add the two solutions, $y(t) = y_1(t) + y_2(t)$.

This principle simplifies the task of finding the response to our first-order model when there are both initial conditions and a forcing function. Armed with this we can determine the step response of our first-order model when there is a non-zero initial condition

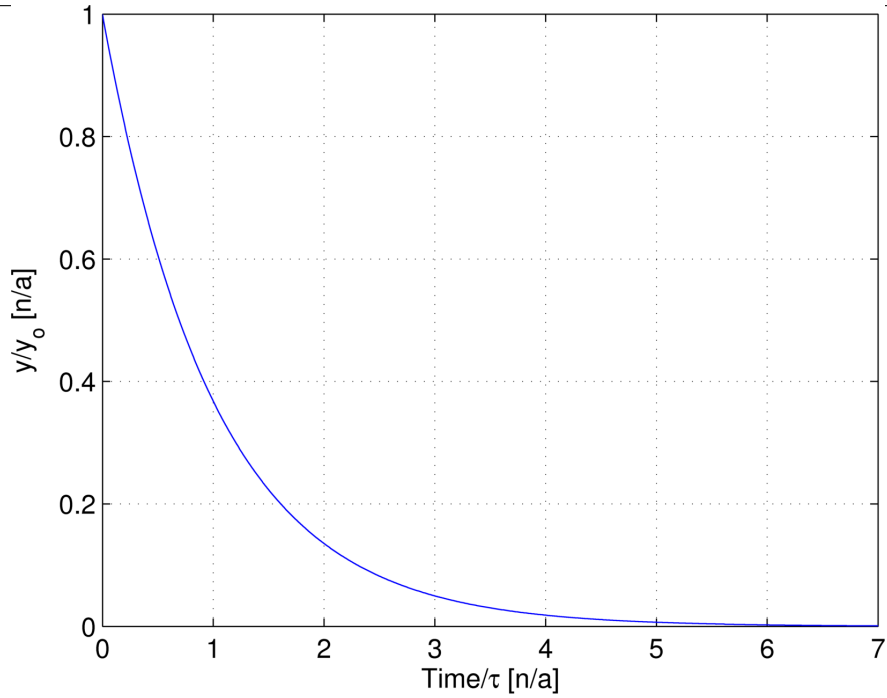$$y(t) = y_o \, e^{-t/\tau} + A\tau \left(1 - e^{-t/\tau}\right) \tag{3.8}$$

Figure 3.4: Graph of the free response (3.7). Notice that the x and y axes have been normalized so that the output is the ratio of the response and the initial condition and the time is the ratio of time and the time constant.

Another handy aspect of superposition is the scaling property. This means that if we scale (multiply) the input by a factor (a) then the output is simply the response to the original input scaled (multiplied) by the same factor. As an example, consider our first-order car model (3.2) where the input is a force ($f(t)$). If we know that a force of 1,000 N results in a steady state speed of 25 mph then superposition tells us that a force of 2,000 N will result in a final speed of 50 mph.

Exercise 3.7 *Consider our first-order car model discussed above where a 1,000 N input yields a steady state speed of 25 mph. Furthermore, we know that the time constant of this system is 2 s. Again we double the input to 2,000 N. What is the time constant with the doubled input?*

3.6 Thermometer Example

Our first-order model can be used to describe the response of a temperature sensor to changes in the ambient environment. This is analogous to our "car" model in that the temperature sensor has thermal inertia (analogous to the mass of our "car") and thermal resistance (analogous to the drag of our "car").

3.6.1   Mathematical Model

Suppose we have a bulb thermometer that is at equilibrium the instant before it is plunged into boiling water. The measured temperature will not instantly move to $100\,°C$, but instead the temperature of the device ($T(t)$) will move toward the new equilibrium over time. The speed of that response is determined by the inertia and resistance of the system.

Figure 3.5 illustrates how we might analyze this system. Newton's law of cooling says that "the rate of heat loss of a body is proportional to the difference in temperatures between the body and the surroundings", or, stated mathematically

$$\dot{Q} = hA(T_f - T(t)) \tag{3.9}$$

where $\dot{Q}$ is the rate of heat transfer, h is the heat transfer coefficient, A is the surface area of the thermometer, T_f is temperature of the surroundings and $T(t)$ is the temperature of the thermometer. Furthermore the thermal energy in the thermometer mass is $Q = C(T(t))$ where C is the total heat capacity (the product of the mass (m) and the specific heat capacity of the material (c_p)). Putting this together we can come up with a first-order model for a thermometer

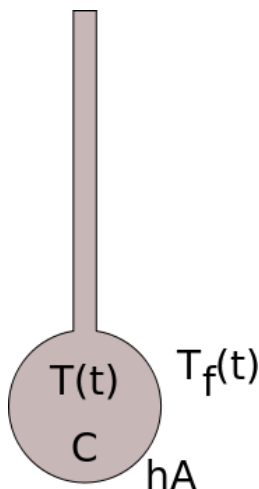$$\dot{T}(t) = \frac{hA}{C}(T_f - T(t)). \tag{3.10}$$

Figure 3.5: Illustration of a simple model of the response of a thermometer.

Exercise 3.8   *Using (3.10) rearrange the terms so that the expression has the same form as our canonical first-order model (3.1). What is the input to our model? What is the output of our model? What is the time constant of this model? What is the steady-state temperature predicted by this model?*

3.7   Glossary

Notation	Description	Symbol	Page List
first-order model	A first-order, linear, time-invariant ordinary differential equation.		20
first-order model	An equation or set of equations that behave in a similar way to a physical system under certain conditions.		19
first-order model	For a first-order model with a step input, the time duration from the initiation of the step to the time when the model output is 63.2% of the way to the steady-state value.	τ	23
steady-state response	The output of a mathematical model or physical system when in equilibrium. The steady-state response is often contrasted with the transient response.		23
step input	A mathematical function that is zero for all time less than $t = 0$ and unity for all time greater than or equal to $t = 1$.	$\mu(t)$	21
step response	The output of a mathematical function, as a function of time, when a step input is given as the input or forcing function.		21
superposition	The property of linear mathematical models that states that the total response of the system caused by two (or more) inputs is the sum of the responses to each input considered independently		24