EECS 16A Designing Information Devices and Systems I
Spring 2021 Lecture Notes Note 1A

Overview

In this note, we will introduce the fundamental objects of **vectors** and **matrices**, and discuss how to use them to represent systems of linear equations. We will further introduce the concept of **linearity**, and discuss how its formal definition relates to our intuitive understanding of "linear" functions. To illustrate all of these concepts, we will discuss the real-world technique of tomographic imaging.

1.1 What is Linear Algebra?

- Linear algebra is the study of linear functions and linear equations, typically via their representation using vectors and matrices.
- A lot of objects in EECS can be treated as vectors and studied with linear algebra.
- Linearity is a good first-order approximation to the complicated real world.
- There exist good fast algorithms to do many of these manipulations in computers.
- Linear algebra concepts are an important tool for modeling the real world.

As you will see in the homeworks and labs, these concepts can be used to do many interesting things in real-world-relevant application scenarios. In the previous note, we introduced the idea that all information devices and systems (1) take some piece of information from the real world, (2) convert it to the electrical domain for measurement, and then (3) process these electrical signals. Because so many efficient algorithms exist that perform linear algebraic manipulations with computers, linear algebra is often a crucial component of this processing step.

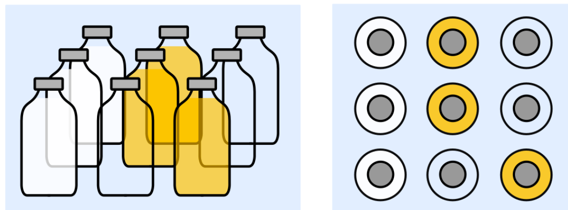
1.2 Application: Tomography

Throughout this course, we will motivate the introduction of concepts by considering a real-world application - this is the first one!

Tomography allows us to "see inside" a solid object, such as the human body or even the earth, by taking measurements with a penetrating wave, such as X-rays. CT scans in medical imaging are perhaps the most famous such example — in fact, CT stands for "computed tomography."

Let's look at a specific toy example, using tomography to help with a (fairly unlikely!) real-world scenario.

A grocery store employee just had a truck load of bottles given to him. Each bottle is either empty, contains milk, or contains juice, and the bottles are packaged in boxes, with each box containing 9 bottles in a 3×3 grid. Inside a single box, it might look something like this:



Side view Top view

If we choose symbols such that M represents milk, J represents juice, and O represents an empty bottle, we can represent the stack of bottles shown above as follows:

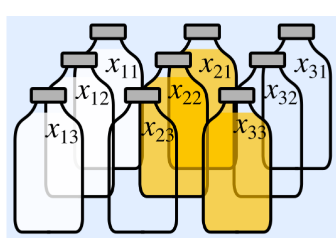
$$\begin{array}{ccc} M & J & O \\ M & J & O \\ M & O & J \end{array} \tag{1}$$

Imagine that our grocer cannot see directly into the box, but needs to determine its contents using a light source and light sensor. How can we help him do this?

Let the light source emit light with a certain known intensity. As the light passes through a bottle, its intensity diminishes by an amount that depends on the contents of the bottle - milk absorbs 3 units of light, juice absorbs 2 units of light and an empty bottle absorbs 1 unit of light. The box itself does not affect the intensity of the light. After the light emitted exits the box, we can use our sensor to measure the final intensity, and so determine the amount of light absorbed by each bottle.

Thus, if we shine light in a straight line through some bottles within the box, we can determine the total amount of light absorbed by the bottles as the sum of the light absorbed by each bottle. For instance, in our specific example, shining a light from left to right would look like this, with each row observed to absorb 6 total units of light:

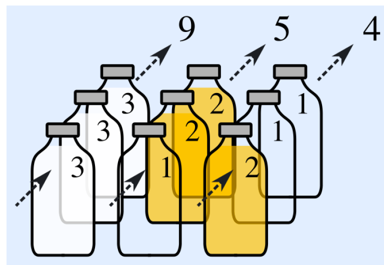
In order to deal with this more generally, let's assign variables to the amount of light absorbed by each bottle:



This means that x_{11} would be the amount of light the top left bottle absorbs, x_{21} would be the amount of light the top middle bottle absorbs, and so forth. Shining the light from left to right for our specific example gives the following equations:

$$x_{11} + x_{21} + x_{31} = 6 \tag{2}$$
$$x_{12} + x_{22} + x_{32} = 6 \tag{3}$$
$$x_{13} + x_{23} + x_{33} = 6 \tag{4}$$

Similarly, we could consider shining a light from bottom to top:



Which would give the following equations:

$$x_{13} + x_{12} + x_{11} = 9 \tag{5}$$
$$x_{23} + x_{22} + x_{21} = 5 \tag{6}$$
$$x_{33} + x_{32} + x_{31} = 4 \tag{7}$$

Thus, we now know how to determine our observations given the contents of the box. But can we do the reverse? That is to say, given the amounts of light absorbed by each row and column of bottles, can we reconstruct the box's original contents?

From our above observations, one possible solution for the values of x_{ij} (corresponding to the actual configuration of bottles) is

$$\begin{matrix} 3 & 2 & 1 \\ 3 & 2 & 1 \\ 3 & 1 & 2 \end{matrix} \tag{8}$$

However, the following solution also works:

$$\begin{matrix} 3 & 2 & 1 \\ 3 & 1 & 2 \\ 3 & 2 & 1, \end{matrix} \tag{9}$$

which corresponds to a different configuration of bottles within the box. In other words, our observations are not sufficient to **uniquely** identify the configuration of bottles within the box. This is a problem!

Usually, if we can't identify an object from a set of observations, we make more observations. To get these additional observations, we could shine light at different angles through the box.

This brings up some very natural questions: Would shining light through the diagonals of the box provide us with enough information? If not, how many different directions do we need to shine light through before we are certain of the configuration of bottles? Do some measurements provide us with more information than others do? What happens as we vary the number of bottles in our box?

In Module 1 of this course, we will develop the tools to answer all of these questions.

1.3 What is a Linear Equation?

1.3.1 Intuition

In our bottle-sorting tomography example, we represented each measurement in a row or column as an equation. The collection of equations is an example of a **system of linear equations**, which summarizes the known relationships between the variables we want to solve for (x_{11}, x_{12}, x_{13}, etc.) and our measurements.

But what makes an equation "linear"? Essentially, a linear equation is one where each variable has degree 1. For instance, for unknowns x and y, the equation

$$5x + 6y = 7 \tag{10}$$

is made up of the two coefficients 5 and 6 multiplied by the two unknowns x and y respectively, summed together, that are then set equal to the scalar constant 7.

In contrast, the equation

$$y^2 = 5 \tag{11}$$

is not a linear equation, since we multiply our unknown y with itself, rather than a constant scalar.

Observe that equations such as

$$8x = 4y,$$

or

$$8x - 4y = 0,$$

or

$$2x - y = 0$$

are all examples of linear equations. Notice that the scalar constant is allowed to be 0.

1.3.2 Linear Functions

Let's try to formalize this intuition a little. First, we will define the concept of a **linear function**. At this point, we can define a linear function as simply a function f of one scalar argument with the property that, for arbitrary scalars α and x,

$$f(\alpha x) = \alpha \cdot f(x).$$

In other words, if the input to a linear function is multiplied by some scalar α, the output of the function will be multiplied by α as well.

To gain a better understanding of this definition, we'll consider a few examples. First, consider the function

$$f_1(x) = x^2.$$

Plugging in the values $x = 1$ and $x = 2$, we see that

$$f_1(1) = 1 \text{ and } f_1(2) = 4.$$

So when x is scaled by a factor of 2 (from $x = 1$ to $x = 2$), $f_1(x)$ scales by a factor of 4, which is not the same factor. Due to this counterexample, we have shown that x^2 is not a linear function.

Now, let's look at the function

$$f_2(x) = 5x.$$

Plugging in the same values for x as before, we see that

$$f_2(1) = 5 \text{ and } f_2(2) = 10.$$

Here, when x was scaled by a factor of 2 (from 1 to 2), $f_2(x)$ scaled by a factor of 2 as well (from 5 to 10)! This suggests that $f_2(x)$ *might* be a linear function.

Let's try to quickly prove it. For arbitrary scalars α and x, we see from the definition of $f_2(x)$ that

$$f_2(\alpha x) = 5 \cdot (\alpha x) = \alpha \cdot (5x) = \alpha \cdot f_2(x),$$

so $f_2(x)$ is indeed a linear function.

More generally, a similar proof to the above can be done to show that *all* functions of the form $f(x) = kx$ are linear - in fact, we can even show that *any* linear function with a single scalar input and output can be written in the form $f(x) = kx$, for some constant k! Let's try to prove this!

Let $f(x)$ be an arbitrary linear function. Observe that, by the definition of linear functions,

$$f(x) = f(x \cdot 1) = x \cdot f(1).$$

Clearly, $f(1)$ does not depend on x, and so is a scalar constant. Let this scalar constant be $k = f(1)$. Thus, we have shown that there exists a k such that

$$f(x) = kx,$$

no matter what $f(x)$ we choose, so long as it is a linear function with scalar inputs and outputs.

1.3.3 Affine Functions

But what about functions like

$$f_3(x) = 2x + 1?$$

Plotting this function, we see that it is a line. But it doesn't seem to fit into the form $f(x) = kx$, so is it linear? A simple check, if we're ever unsure about the behavior of a function, is to plug in some simple input values and see how the output behaves. Let's do that here, for $x = 1$ and $x = 2$. We see that

$$f_3(1) = 3 \text{ and } f_3(2) = 5,$$

so doubling the input value from 1 to 2 changes the output by a factor of $5/3$. Thus, this function is not linear, *even though* it describes the equation of a line.

Nevertheless, it still seems to have some "linear" properties, compared to an expression like x^2. More precisely, observe that it can be written as a sum of a linear function of x and a scalar constant. We define **affine functions** to be the set of functions that can be written as a sum of a linear function and a scalar constant, so though $f_3(x)$ is not *linear*, it is still *affine*.

Notice that the definition of affine functions includes all linear functions as a subset (by setting the scalar constant to 0), so every linear function is also affine, though not vice-versa.

These definitions mean that while all functions describing a line can be shown to be affine, not all of them are linear. This has the unfortunate consequence that, in informal conversation, *affine* functions may be called *linear*, since both describe a line. This usage, though common, is **wrong**.

1.3.4 Linear Equations

So what's a linear equation, then? Formally, a linear equation with the unknown scalars $x_1, x_2, \ldots, x_n$ is an equation where each side is a sum of scalar-valued linear functions of each of the unknowns plus a scalar constant.

What does this mean? Expressed algebraically, we obtain the most general form of a linear equation, where the f_i and g_i are each linear functions with a single scalar input and output, and b_f and b_g are two scalar constants:

$$f_1(x_1) + f_2(x_2) + \ldots + f_n(x_n) + b_f = g_1(x_1) + g_2(x_2) + \ldots + g_n(x_n) + b_g.$$

Now, recall that linear functions with a single scalar input and output can be expressed in a very particular form - we know that we can write $f_i(x) = a_i \cdot x$, and $g_i(x) = a_i' \cdot x$, where all the a_i and a_i' are scalar constants. Substituting, we find that the general form of a linear equation can be rewritten as

$$a_1x_1 + a_2x_2 + \ldots + a_nx_n + b_f = a_1'x_1 + a_2'x_2 + \ldots + a_n'x_n + b_g.$$

Let's take a closer look at part of this equation - specifically, the expression

$$a_1x_1 + a_2x_2 + \ldots + a_nx_n.$$

Notice that this expression can be thought of as a "weighted sum" of the x_i, where the weights are the scalar constants a_i. When the weights do not depend on any of the terms (such as when the weights are constants), we call the weighted sum a **linear combination** of said terms. So the above expression is typically referred to as a *linear combination* of the x_i.

Thus, we can simplify our definition of a linear equation, to the following: A linear equation is one that equates two linear combinations of the unknowns plus a constant term.

1.4 Vectors and Matrices

We will now introduce some new notation that will help us deal with systems of linear equations in a more compact form.

1.4.1 Vectors

Definition 1.1 (Vector):

A vector is an ordered list of numbers. Suppose we have a collection of n real numbers: $x_1, x_2, \cdots, x_n$. This collection can be written as a single point in an n-dimensional space, denoted as:

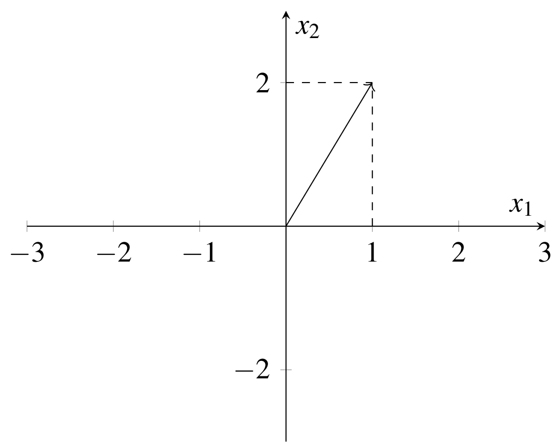
$$\vec{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}. \tag{12}$$

We call $\vec{x}$ a **vector**. Because $\vec{x}$ contains n real numbers, we can use the $\in$ ("in" — i.e., is a member of) symbol to write $\vec{x} \in \mathbb{R}^n$ ($\mathbb{R}$ represents the set of real numbers). If the elements of $\vec{x}$ were complex numbers, we would write $\vec{x} \in \mathbb{C}^n$. Each x_i (for i between 1 and n) is called a **component**, or **element**, of the vector. The **size** of a vector is the number of components it contains (so the example vector is of size n, and the example below is of size two).

Example 1.1 (Vector of size two):

$$\vec{x} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

In the above example, $\vec{x}$ is a vector with two components. Because the components are both real numbers, $\vec{x} \in \mathbb{R}^2$. We can represent the vector graphically on a 2-D plane, using the first element, x_1, to denote the horizontal position of the vector and the second element, x_2, to denote its vertical position:



Additional Resources For more on vectors, read pages 1-6 of *Strang* and try Problem Set 1.1. *Extra: Try reading the portions on linear combinations which generate a"space."*

Read more on vectors in *Schuam's* on pages 1-3 and try Problems 1.1 to 1.6.

1.4.2 Matrices

Definition 1.2 (Matrix): A **matrix** is a rectangular array of numbers, written as:

$$A = \begin{bmatrix} A_{11} & \cdots & A_{1n} \\ \vdots & \ddots & \vdots \\ A_{m1} & \cdots & A_{mn} \end{bmatrix} \tag{13}$$

Each A_{ij} (where i is the row index and j is the column index) is a **component**, or **element** of the matrix A.

Example 1.2 (4×3 **Matrix)**:

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 4 & 6 \\ 3 & 5 & 7 \\ 4 & 8 & 12 \end{bmatrix}$$

In the example above, A has $m = 4$ rows and $n = 3$ columns (a 4×3 matrix).

1.5 Representing Linear Systems using Matrices

1.5.1 Augmented Matrices

First, we will use what we call an **augmented matrix** to represent the coefficients of a linear system of equations, like the system we saw earlier when experimenting with tomography:

$$x_{11} + x_{21} + x_{31} = 6 \tag{14}$$
$$x_{12} + x_{22} + x_{32} = 6 \tag{15}$$
$$x_{13} + x_{23} + x_{33} = 6 \tag{16}$$
$$x_{13} + x_{12} + x_{11} = 9 \tag{17}$$
$$x_{23} + x_{22} + x_{21} = 5 \tag{18}$$
$$x_{33} + x_{32} + x_{31} = 4. \tag{19}$$

Recall that the first three equations of the above system came from measuring the light absorbed through each row of bottles, while the second three equations came from measurements of the columns.

First, we will rewrite our system slightly to introduce some additional structure. Specifically, notice that each equation involves only a subset of our unknowns: for instance, the first equation only includes x_{11}, x_{21}, and x_{31}, and not the other six unknowns in the problem. We can make sure all nine unknowns are included in each equation by introducing linear terms with a zero coefficient, as follows:

$$1 \times x_{11} + 0 \times x_{12} + 0 \times x_{13} + 1 \times x_{21} + 0 \times x_{22} + 0 \times x_{23} + 1 \times x_{31} + 0 \times x_{32} + 0 \times x_{33} = 6 \tag{20}$$
$$0 \times x_{11} + 1 \times x_{12} + 0 \times x_{13} + 0 \times x_{21} + 1 \times x_{22} + 0 \times x_{23} + 0 \times x_{31} + 1 \times x_{32} + 0 \times x_{33} = 6 \tag{21}$$
$$0 \times x_{11} + 0 \times x_{12} + 1 \times x_{13} + 0 \times x_{21} + 0 \times x_{22} + 1 \times x_{23} + 0 \times x_{31} + 0 \times x_{32} + 1 \times x_{33} = 6 \tag{22}$$
$$1 \times x_{11} + 1 \times x_{12} + 1 \times x_{13} + 0 \times x_{21} + 0 \times x_{22} + 0 \times x_{23} + 0 \times x_{31} + 0 \times x_{32} + 0 \times x_{33} = 9 \tag{23}$$
$$0 \times x_{11} + 0 \times x_{12} + 0 \times x_{13} + 1 \times x_{21} + 1 \times x_{22} + 1 \times x_{23} + 0 \times x_{31} + 0 \times x_{32} + 0 \times x_{33} = 5 \tag{24}$$
$$0 \times x_{11} + 0 \times x_{12} + 0 \times x_{13} + 0 \times x_{21} + 0 \times x_{22} + 0 \times x_{23} + 1 \times x_{31} + 1 \times x_{32} + 1 \times x_{33} = 4 \tag{25}$$

So far, we've made some fairly elementary transformations to our system of linear equations. But what's the point? While our resultant system is both longer and harder to read than our original system, it is structured. This will allow us to easily use a standard algorithm to process and solve the system.

A natural next step is to put the coefficients into what we call an "augmented matrix" as below. Each row of the augmented matrix represents one equation. Each column corresponds to the coefficients for a given variable across all equations - i.e. the first column corresponds to x_{11}, the second to x_{12}, and so on:

$$\left[\begin{array}{ccccccccc|c} 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 6 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 6 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 6 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 9 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 5 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 4 \end{array}\right] \tag{26}$$

Typically, when representing a system of linear equations as a single matrix, it is conventional to draw a line before the last column as done above, since that column contains the constants, not the coefficients, of our system:

This representation of a linear system is known as the **augmented matrix representation**. An interesting thing to notice about this representation is that the symbols corresponding to our unknowns have vanished entirely!

1.5.2 Matrix-Vector Form

Intuitively, we can imagine that augmented matrices convey the "underlying" linear system corresponding to our problem, without involving the actual variable names that we chose. Still, we might want to preserve the variable names, since without them we are losing some information about our problem. To do so, we can write our above system of equations in the following form:

$$\begin{bmatrix} 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} x_{11} \\ x_{12} \\ x_{13} \\ x_{21} \\ x_{22} \\ x_{23} \\ x_{31} \\ x_{32} \\ x_{33} \end{bmatrix} = \begin{bmatrix} 6 \\ 6 \\ 6 \\ 9 \\ 5 \\ 4 \end{bmatrix}.$$

What's going on here? Compared to the augmented matrix form, notice that the constant terms have not been included into the main matrix (known as the *coefficient matrix*), but instead have been placed in a vector on the right-hand-side of an equality. In addition, observe that the unknowns have been placed in a vector that is right-multiplied with the coefficient matrix on the left-hand-side of the equality. Often, the letter A is used to represent the coefficient matrix, $\vec{x}$ for the vector of unknowns, and $\vec{b}$ for the vector of coefficients. After these substitutions, the above equation becomes

$$\begin{aligned} A &= \begin{bmatrix} 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \end{bmatrix} \\ \vec{x} &= \begin{bmatrix} x_{11} \\ x_{12} \\ x_{13} \\ x_{21} \\ x_{22} \\ x_{23} \\ x_{31} \\ x_{32} \\ x_{33} \end{bmatrix} \\ \vec{b} &= \begin{bmatrix} 6 \\ 6 \\ 6 \\ 9 \\ 5 \\ 4 \end{bmatrix} \\ \implies \quad A\vec{x} &= \vec{b}. \end{aligned}$$

The above representation is fundamentally an alternative shorthand for systems of linear equations that preserves the variable names of the original system.

1.6 Practice Problems

These practice problems are also available in an interactive form on the course website, along with their solutions.

1. Is $x+2y=4z$ linear?
2. Is $\sin x - 2 = 6$ linear?
3. Is $\sum_{i=1}^{50} i \cdot x - e^{-3}y = \sin\frac{\pi}{3}$ linear?
4. Write $\begin{cases} 2x-3y=1 \\ 3x+y=-2 \end{cases}$ in matrix form.

   (a) $\begin{bmatrix} 2 & -3 & 3 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 1 \\ -2 \end{bmatrix}$

   (b) $\begin{bmatrix} 2 & 3 \\ -3 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 1 \\ -2 \end{bmatrix}$

   (c) $\begin{bmatrix} 2 & -3 \\ 3 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 1 \\ -2 \end{bmatrix}$

   (d) $\begin{bmatrix} 2 & -3 \\ 3 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ -2 \end{bmatrix} = \begin{bmatrix} x \\ y \end{bmatrix}$