EECS 16A Designing Information Devices and Systems I
Spring 2021 Lecture Notes Note 2A

This note covers the following topics:

- **Solutions to Systems of Linear Equations:** How do we know when a system of linear equations has a solution? If it does have a solution, how do we solve it?
- **Vectors:** The notion of a vector is introduced, examples are given, and various mathematical operations on vectors are discussed.
- **Matrices:** The notion of a matrix is introduced, examples are given, and various mathematical operations on matrices are discussed.
- **Matrix-vector Form for Systems of Linear Equations:** We'll see that systems of linear equations can be expressed algebraically using matrices and vectors. This is the starting point for harnessing the full power of linear algebra to solve and understand systems of linear equations.

2.1 Solutions to Systems of Linear Equations

In the previous note, we introduced systems of linear equations and discussed their representation in terms of augmented matrices. We introduced Gaussian elimination, which is a systematic method for solving systems of equations. In terms of the augmented matrix representation, Gaussian elimination proceeds by performing a sequence of elementary row operations until the augmented matrix is in reduced row echelon form. Once this point is reached, it was generally possible to determine whether the given system of equations had a solution or not by inspection of the augmented matrix. The aim of the present section is to make this last step more precise.

To start, we make two important definitions. We say that a system of linear equations possessing one or more solutions is **consistent**. Conversely, if the system of linear equations does not have a solution, we say that it is **inconsistent**. This terminology makes sense when we consider real-life examples of systems of linear equations. In particular, for the tomography example, inconsistency would imply that the set of measurements could not have occurred given our model (hence, the measurements would be "inconsistent" with the model). On the other hand, if the system of equations was consistent, it would imply that the measurements we obtained were "consistent" with the model proposed, and the solution to the system of equations would correspond to a possible configuration of beverages in the crate. While it might seem that inconsistency would not be encountered in real life, it often is. This can be a consequence of our modeling assumptions being imperfect, or it could also be the result of "noise" in our measurements. We will discuss how to deal with inconsistent systems of equations later in this course.

2.1.1 Determining Consistency of Systems of Linear Equations

We state the following useful fact as a Theorem, to highlight it as an important result to keep in mind. In particular, it provides a simple way to determine consistency of a system of linear equations by inspecting the corresponding augmented matrix representation in reduced row echelon form.

Theorem 2.1: A system of linear equations is inconsistent if and only if the corresponding augmented matrix in reduced row echelon form contains the row $[0,0,\ldots,0|1]$.

Why should this be true? Well, the interpretation of such a row is the linear equation $0 = 1$, which is absurd! Since we can never make this true for any assignment of values to the unknowns, the system of equations under consideration does not have a solution (it is inconsistent). In fact, we can make the following similar observation: If the augmented matrix corresponding to a system of linear equations contains a row of the form $[0,0,\ldots,0|b]$ for $b \neq 0$, then the system of linear equations must be inconsistent. This observation can allow us to terminate the Gaussian elimination algorithm early, since we can immediately determine inconsistency without fully reducing the augmented matrix to reduced row echelon form.

On the other hand, if the row $[0,0,\ldots,0|1]$ does not appear in the augmented matrix in reduced row echelon form, then the leading entry of every non-zero row corresponds to a variable (more specifically, a basic variable by definition). By definition of reduced row echelon form, each leading entry is the only non-zero entry in its column. These two observations can be concisely summarized in terms of variables and linear equations: The equations in the linear system (corresponding to the augmented matrix in reduced row echelon form) are in one-to-one correspondence with the basic variables. Hence, the set of solutions to the system of linear equations can be obtained as follows: set the free variables to whatever you like, and solve for the unique values of the basic variables. In particular, if a consistent system of linear equations has at least one free variable, then it has infinitely many solutions. On the other hand, if a consistent system of linear equations has no free variables, then it has a unique solution.

Let's take a few examples to illustrate the ideas discussed above. In each example, we consider a system of equations that has already been transformed to reduced row echelon form (e.g., by running Gaussian elimination). In each case, we determine consistency and solutions, provided they exist.

Example 2.1 (Inconsistent System of Linear Equations):

$$\begin{bmatrix} x & + & & & 2z & = & 2 \\ & & y & - & 3z & = & 1 \\ & & & & 0 & = & 1 \end{bmatrix} \rightarrow \left[\begin{array}{ccc|c} 1 & 0 & 2 & 2 \\ 0 & 1 & -3 & 1 \\ 0 & 0 & 0 & 1 \end{array}\right].$$

The augmented matrix contains the row $[0,0,0|1]$, and therefore the system of equations is inconsistent (i.e., it does not have any solution).

Example 2.2 (Consistent System of Linear Equations with Unique Solution):

$$\begin{bmatrix} x & & & = & 2 \\ & y & & = & 3 \\ & & z & = & 1 \\ & & & & \end{bmatrix} \rightarrow \left[\begin{array}{ccc|c} 1 & 0 & 0 & -2 \\ 0 & 1 & 0 & 3 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 \end{array}\right].$$

The system is consistent since the augmented matrix does not contain the row $[0,0,0|1]$. There are no free variables, and the system has the unique solution $x = -2$, $y = 3$ and $z = 1$.

Example 2.3 (Consistent System of Linear Equations with Infinitely Many Solutions):

$$\begin{bmatrix} v & + & 2w & & - & 2y & & = & 3 \\ & & & x & + & 3y & & = & -1 \\ & & & & & & z & = & 8 \end{bmatrix} \rightarrow \left[\begin{array}{ccccc|c} 1 & 2 & 0 & -2 & 0 & 3 \\ 0 & 0 & 1 & 3 & 0 & -1 \\ 0 & 0 & 0 & 0 & 1 & 8 \end{array}\right].$$

The system is consistent since the augmented matrix does not contain the row $[0,0,0,0,0|1]$. The variables v,x,z are basic variables, and w,y are free variables. For any choice of $w,y \in \mathbb{R}$, the following choices of v,x,z yield a solution:

$$\begin{aligned} v &= 3 - 2w + 2y \\ x &= -1 - 3y \\ z &= 8. \end{aligned}$$

Important Note: To determine consistency, it is a good idea to first transform the augmented matrix to reduced row echelon form (e.g., by Gaussian elimination). Indeed, if the augmented matrix is *not* in reduced row echelon form and does *not* contain the row $[0,\dots,0|1]$, we can *not* guarantee that the system of equations is consistent.

2.2 Vectors and Matrices

In the previous note, we mentioned that linear algebra is the study of linear functions and linear equations, usually in terms of vectors and matrices. We now formally introduce the concepts of vectors and matrices, and discuss how to perform associated computations such as addition and multiplication. As a consequence, we will find that matrices and vectors can be used to algebraically express systems of linear equations. In future notes, we will consider additional properties of vectors and matrices and see how these can help us understand and analyze real-world systems.

2.3 Introduction to Vectors

What is a vector? It is an ordered list of numbers. More specifically, if we are given an ordered collection of n real numbers such as $x_1, x_2, \cdots, x_n$, we can represent this collection as a single point in an n-dimensional real space, $\mathbb{R}^n$, denoted as a **vector** $\vec{x}$:

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

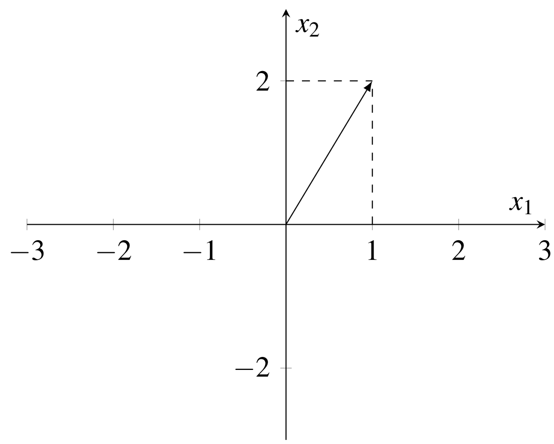
Each x_i (for i between 1 and n) is called a **component**, or **element**, of the vector. The elements of a vector do not have to be real — they could be complex-valued, meaning $\vec{x} \in \mathbb{C}^n$ — but the majority of vectors you will encounter in 16A will be real. The **dimension** (or **size**) of a vector is the number of components it contains. Two vectors $\vec{x}$ and $\vec{y}$ are said to be **equal**, denoted $\vec{x} = \vec{y}$, if they have the same dimension and $x_i = y_i$ for all i.

Remark. Depending on the text, vectors may be written simply as x, or they might be written as letters in boldface $\mathbf{x}$, or they might be denoted with a small arrow on top $\vec{x}$. In these lecture notes we will use the latter arrow notation, and the dimension of the vector will be given (for example, $\vec{x} \in \mathbb{R}^3$ means that $\vec{x}$ contains three real numbers).

Example 2.4 (Vector of size two):

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

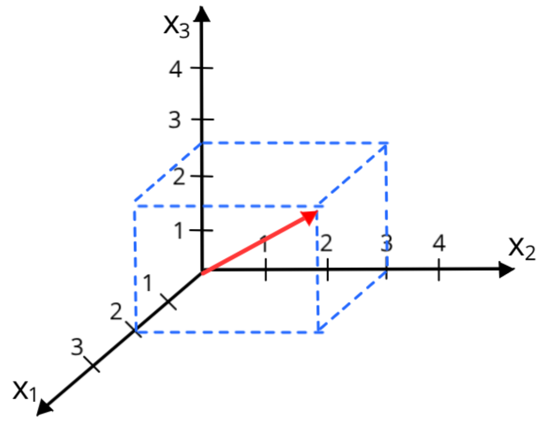
In the above example, $\vec{x}$ is a vector with two components. Because the components are both real numbers, $\vec{x} \in \mathbb{R}^2$. We can represent the vector graphically on a 2-D plane, using the first element, x_1, to denote the horizontal position of the vector and the second element, x_2,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



Additional Resources For more on vectors, read pages 1-6 of *Strang* and try Problem Set 1.1 in *Strang*. *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.

Example 2.5 (3D Vector): Assume we have $x_1 = 2$, $x_2 = 3$, and $x_3 = 2.5$, which can be represented as the vector $\vec{x} = \begin{bmatrix} 2 \\ 3 \\ 2.5 \end{bmatrix}$. This represents a point in 3-D space ($\mathbb{R}^3$), which can be drawn as an arrow from the origin to this point in space:



As we shall see shortly, vectors can be used to help write systems of equations compactly. However, they are also useful for representing a multitude of things — anything that can be represented as an ordered list of numbers can be expressed as a vector. For instance, in the tomography example, we can write a vector to represent the amount of light absorbed by each bottle in a row or column.

2.3.1 More Examples of Vectors

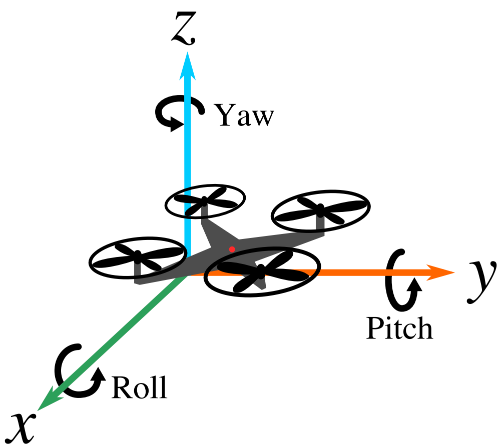
Example 2.6 (Position vs. time): $\vec{x} \in \mathbb{R}^n$ can represent samples of a quantity at n time points. Imagine a car moving along a line. Its position at time $t_1, t_2, \cdots, t_n$ can be represented with a vector:

$$\vec{x} = \begin{bmatrix} x_{t_1} \\ x_{t_2} \\ \vdots \\ x_{t_n} \end{bmatrix}$$

Here, x_{t_i} represents the position of the car at time t_i.

Example 2.7 (Quadrotor state): Vectors can be used to represent the *state* of a system, which is the minimum information you need to completely characterize a system at a given point in time, without any need for more information about the past of the system. State is a powerful concept because it lets us separate the past from the future. The state completely captures the present — and the past can only affect the future through the present. We will revisit the concept of state in EECS 16B, and you will see this in your homework.

As an example, consider modeling the dynamics of a quadrotor. A quadrotor is a type of helicopter that has four rotors. The state of a quadrotor at a particular time can be summarized by its: 3D position (x, y, z), angular position (roll, pitch, yaw), velocity ($\dot{x}$, $\dot{y}$, $\dot{z}$) and angular velocity ($\dot{roll}$, $\dot{pitch}$, $\dot{yaw}$), which can be represented as a vector $\vec{q} \in \mathbb{R}^{12}$, as illustrated in Figure 1.



$$\vec{q} = \begin{bmatrix} x \\ y \\ z \\ \text{roll} \\ \text{pitch} \\ \text{yaw} \\ \dot{x} \\ \dot{y} \\ \dot{z} \\ \dot{\text{roll}} \\ \dot{\text{pitch}} \\ \dot{\text{yaw}} \end{bmatrix}$$

Figure 1: A quadrotor drone (left) and the vector containing the information needed to determine its state (right). The 3D position of the quadrotor is captured in its x, y, and z coordinates, while its angular position is measured as its roll, pitch, and yaw. A dot above a variable implies taking a derivative with respect to time, giving the velocity of each component (for instance, $\dot{x} = \frac{dx}{dt}$ is the quadrotor's velocity in the x direction).

Example 2.8 (Electrical circuit quantities): Later in this course, we will learn about *current* and *voltage*, which are physical quantities present in electrical circuits. The current and voltage vary at different points in the electrical circuit, and we can represent these quantities as elements in a vector where the size of the vector will depend on the complexity of the circuit. We'll get into this more in Module 2!

Example 2.9 (Vectors representing functions): Consider a simple game in which points are earned or lost by rolling a 6-sided die. If you roll a 5 or lower, you earn the number of points that you roll, but if you roll a 6, you earn -6 points. We could define a function that maps the die roll outcome, d, to the number of points earned by that roll, $f(d)$:

$$f(d) = \begin{cases} d & \text{if } d \leq 5 \\ -d & \text{if } d = 6 \end{cases}$$

We could also capture this information in a vector that summarizes all possible values of $f(d)$:

$$\begin{bmatrix} f(1) \\ f(2) \\ f(3) \\ f(4) \\ f(5) \\ f(6) \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ -6 \end{bmatrix}$$

Because the input domain of the function $f(d)$ is restricted to the integers 1 through 6, it is easy to see how it can be represented as a vector. It turns out that we can build a vector representation of any function with a numerical output — even functions that operate on an input domain of infinite size, such as all real numbers or all integers. Such functions can be represented by an *infinite-dimensional* vector. But for the most part, "vectors" in 16A will be finite-dimensional.

Example 2.10 (Color): This is relevant to the lab we will do! The vector

$$\vec{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}, \quad \vec{x} \in \mathbb{R}^3$$

can represent color with its components giving red, green, and blue (RGB) intensity values. Typically, the entries of $\vec{x}$ will fall in a range that reflects the range of the color (often 0 to 255)[1]. However, the human perception of color is quite subtle and involves psychophysics as well as ideas that are related to sampling that we will talk about in EECS 16B.

2.3.2 Special Vectors

Definition 2.1 (Zero Vector): A **zero vector** is a vector with all the components equal to zero, usually just represented as $\vec{0}$. You can usually tell the size of the zero vector $\vec{0}$ from the context: if $\vec{x} \in \mathbb{R}^n$ is added to $\vec{0}$, then $\vec{0}$ must also be in $\mathbb{R}^n$.

Definition 2.2 (Standard Unit Vector): A **standard unit vector** is a vector with all components equal to 0 except for one element, which is equal to 1. A standard unit vector where the ith position is equal to 1 is written as $\vec{e}_i$. We can denote the 3 standard unit vectors in $\mathbb{R}^3$ as:

$$\vec{e}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \vec{e}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \vec{e}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

When talking about standard unit vectors in the context of states, we might also use the word “pure” to refer to such states. This is because they only have one non-zero component in them. Other states are mixtures of pure states.

2.4 Vector Operations

2.4.1 Vector Addition

Two vectors of the same size and in the same space (e.g. complex numbers, real numbers, etc.) can be added together by adding their corresponding components. For example, we can add two vectors in $\mathbb{R}^3$:

$$\begin{bmatrix} -1 \\ 3.5 \\ 0 \end{bmatrix} + \begin{bmatrix} 2 \\ -1 \\ 3 \end{bmatrix} = \begin{bmatrix} -1+2 \\ 3.5-1 \\ 0+3 \end{bmatrix} = \begin{bmatrix} 1 \\ 2.5 \\ 3 \end{bmatrix}$$

Vector addition can be performed graphically as well. In $\mathbb{R}^n$, you place the tail of the first vector at the origin, and then place the tail of the second vector at the first vector’s head. The vector from the origin to the head of the second vector is the resulting vector sum. The following illustration shows how this can be done for two vectors, $\vec{x}$ and $\vec{y}$, in $\mathbb{R}^2$:

[1] Why 0 to 255? We will cover this more in 16B, but generally speaking, the 256 integers in this range can be represented easily as an 8-bit binary number because $256 = 2^8$.

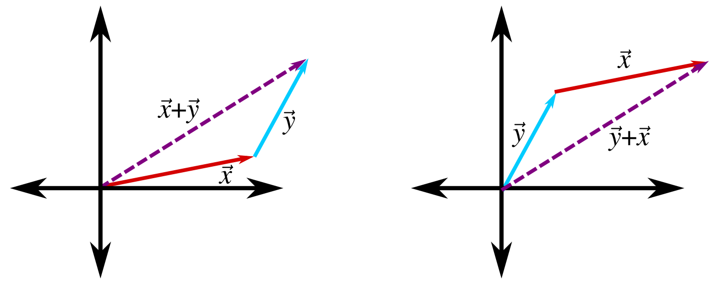


Figure 2: Adding two vectors in $\mathbb{R}^2$. We can see that $\vec{x}+\vec{y}=\vec{y}+\vec{x}$, so vector addition is commutative.

Properties of Vector Addition: Many of the properties of addition you are already familiar with when adding individual numbers hold for vector addition as well. For three vectors $\vec{x},\vec{y},\vec{z}\in\mathbb{R}^n$ (and $\vec{0}\in\mathbb{R}^n$), the following properties hold:

$$\begin{aligned}\textbf{Commutativity}\text{ (shown in Figure 2)}:&\quad \vec{x}+\vec{y}=\vec{y}+\vec{x}\\ \textbf{Associativity}:&\quad (\vec{x}+\vec{y})+\vec{z}=\vec{x}+(\vec{y}+\vec{z})\\ \textbf{Additive identity}:&\quad \vec{x}+\vec{0}=\vec{x}\\ \textbf{Additive inverse}:&\quad \vec{x}+(-\vec{x})=\vec{0}\end{aligned}$$

2.4.2 Scalar Multiplication

We can multiply a vector by a number, called a **scalar**.[2]

Definition 2.3 (Scalar): A **scalar** is a number. In mathematics and physics, scalars can be used to describe magnitude or used to scale things (e.g. cut every element of a vector in half by multiplying by 0.5, or flip the signs of each element in a vector by multiplying by -1).

To perform scalar multiplication, just multiply each of the components of the vector by the scalar:

$$-3\times\begin{bmatrix}-1\\3.5\\0\end{bmatrix}=\begin{bmatrix}3\\-10.5\\0\end{bmatrix}$$

In general, for a scalar α and vector $\vec{x}$, this looks like

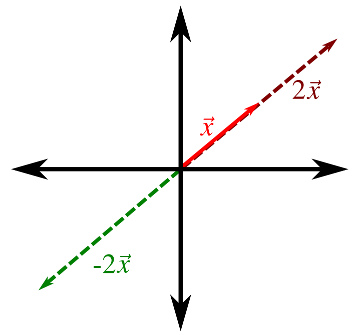
$$\alpha\begin{bmatrix}x_1\\x_2\\\vdots\\x_n\end{bmatrix}=\begin{bmatrix}\alpha x_1\\\alpha x_2\\\vdots\\\alpha x_n\end{bmatrix}$$

Example 2.11 (Negative Vector): To obtain a negative vector $-\vec{x}$, where $\vec{x}\in\mathbb{R}^n$, we can just multiply the vector $\vec{x}$ by the scalar -1. In other words, $-\vec{x}=-1\times\vec{x}$.

Example 2.12 (Zero Vector): We can obtain the zero vector by multiplying any vector by 0: $0\vec{x}=\vec{0}$

[2]Note that this scalar must be in the same space as the vector (e.g. also in the reals, $\mathbb{R}$, or in the complex numbers, $\mathbb{C}$). Since we'll only be working over the reals in this class, you don't have to worry about this detail.

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



Properties of Scalar Multiplication: Just as multiplying a scalar by a scalar is associative ($(xy)z = x(yz)$), distributive ($x(y+z) = xy+xz$), and the multiplicative identity holds (multiplying any number by 1 returns that original number), multiplying a scalar by a vector has the same properties:

$$\begin{aligned} \textbf{Associativity}: &\quad (\alpha\beta)\vec{x} = \alpha(\beta\vec{x}) \\ \textbf{Distributivity}: &\quad (\alpha+\beta)\vec{x} = \alpha\vec{x} + \beta\vec{x} \\ \textbf{Multiplicative identity}: &\quad 1\vec{x} = \vec{x} \end{aligned}$$

2.4.3 Vector Transpose

The transpose of a vector (or a matrix) is a crucial piece of notation for vector and matrix calculations. We write the transpose of a vector $\vec{x}$ as $\vec{x}^T$. If $\vec{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$, then $\vec{x}^T = \begin{bmatrix} x_1 & \cdots & x_n \end{bmatrix}$. The transpose of a **column vector** (a vertical vector) is called a **row vector** (a horizontal vector).The transpose of a row vector is a column vector. Consequently, the transpose of the transpose of a vector recovers the original vector.

It is important to recognize that, though the transpose of a vector contains the same elements as the original vector, it is still a different vector! That is to say, for any vector $\vec{x}$ (with at least two components), $\vec{x}^T \neq \vec{x}$!

2.4.4 Vector-Vector Multiplication

While the operations we have discussed thus far may be largely consistent with your intuition from working with scalars, vector-vector multiplication is not quite as straightforward. By convention, a row vector can only be multiplied by a column vector (and vice versa). Moreover, vector-vector multiplication is not commutative, so $\vec{y}^T\vec{x} \neq \vec{x}\vec{y}^T$. $\vec{x}$ and $\vec{y}$ are both column vectors. In fact, while $\vec{y}^T\vec{x}$ is only defined when both vectors have the same number of elements, $\vec{x}\vec{y}^T$ is defined for vectors of any size.

Multiplying a column vector on the right by a row vector on the left is called computing the **inner product** (also called the **dot product**) of two vectors,[3] which we will discuss further in later notes. An alternative notation for the inner product of two vectors is $\langle\vec{x},\vec{y}\rangle$ - we will see a lot more of this notation in Module 3 of the course. Below is an illustration of $\vec{y}^T\vec{x}$:

[3] These terms are not identical mathematically, but will give the same result for the real vectors we will be working with in 16A.

$$\vec{y}^T\vec{x} = \begin{bmatrix} \boxed{y_1} & \boxed{y_2} & \cdots & \boxed{y_n} \end{bmatrix} \begin{bmatrix} \boxed{x_1} \\ \boxed{x_2} \\ \vdots \\ \boxed{x_n} \end{bmatrix} = y_1x_1 + y_2x_2 + \cdots + y_nx_n$$

Fundamentally, $\vec{y}^T\vec{x}$ is a sum of products — each element in $\vec{y}$ is multiplied by the corresponding element in $\vec{x}$, and all of these products are added together, resulting in a scalar. This form of vector-vector multiplication thus maps a pair of vectors in $\mathbb{R}^n$ to a single scalar in $\mathbb{R}$. Notice that the inner product as we have defined it is *commutative*, since $\vec{y}^T\vec{x} = \vec{x}^T\vec{y}$. As it turns out, this ceases to be true when we start working with complex numbers, as will be seen in EECS 16B.

Multiplying a row vector on the right by a column vector on the left, on the other hand, generates a **matrix**, which we will define formally in the next section. In particular, the product of $\vec{x} \in \mathbb{R}^n$ and $\vec{y}^T \in \mathbb{R}^m$ is given by:

$$\vec{x}\,\vec{y}^T = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \begin{bmatrix} y_1 & y_2 & \cdots & y_m \end{bmatrix} = \begin{bmatrix} x_1y_1 & x_1y_2 & \cdots & x_1y_m \\ x_2y_1 & x_2y_2 & \cdots & x_2y_m \\ \vdots & \vdots & \vdots & \vdots \\ x_ny_1 & x_ny_2 & \cdots & x_ny_m \end{bmatrix}$$

The entry in row i and column j of $\vec{x}\,\vec{y}^T$ is the product of x_i and y_j. To distinguish it from the inner product, this type of vector-vector multiplication is often called the **outer product**. This is perhaps best understood as a particular form of matrix-matrix multiplication, which we will discuss more in the next section.

2.5 Introduction to Matrices

Definition 2.4 (Matrix): A **matrix** is a rectangular array of numbers, written as:

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

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

We will generally consider the case of *real* matrices (meaning that the elements are real numbers), in which case we write $A \in \mathbb{R}^{m\times n}$ to denote that A is an $m \times n$ matrix (i.e., it has m rows and n columns).

Remark. Matrices are often represented by capital letters (e.g. A), sometimes in boldface (e.g. $\mathbf{A}$). In these notes, the notations are used interchangeably. Subscripts are typically used to specify an element of a matrix, with the first subscript corresponding the row index and the second corresponding to the column index.

A matrix is said to be **square** if $m = n$ (that is, if the number of rows and number of columns are equal). Just as we could compute the **transpose** of a vector by transforming rows into columns, we can compute the transpose of a matrix, A^T, where the rows of A^T are the (transposed) columns of A:

$$A^T = \begin{bmatrix} A_{11} & \cdots & A_{m1} \\ \vdots & \ddots & \vdots \\ A_{1n} & \cdots & A_{nm} \end{bmatrix}$$

Mathematically, A^T is the $n \times m$ matrix given by $(A^T)_{ij} = A_{ji}$. A square matrix is said to be **symmetric** if $A = A^T$, which means that $A_{ij} = A_{ji}$ for all i and j.

2.5.1 Examples of Matrices

There are many ways matrices can be useful for representing objects; we will provide a few examples below, beginning with an array of numbers.

Example 2.13 (4×3 **Matrix**):

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

In order to describe the dimensions of the above matrix, we would say $A \in \mathbb{R}^{4\times 3}$, or that A has $m = 4$ rows and $n = 3$ columns, or that A is a 4×3 matrix.

Example 2.14 (Image): A grayscale image of $m \times n$ pixels may be regarded as a matrix with m rows and n columns, with entries corresponding to the grayscale levels at the pixel location. This matrix in $\mathbb{R}^{m\times n}$ could also be represented by a vector of length mn (and therefore, a vector in $\mathbb{R}^{mn}$) by placing all pixel readings in a single list, as illustrated explained in the figure below. The choice of representation is somewhat arbitrary, as long as we agree on what each representation signifies. This brings up an interesting philosophical point:

$$\begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mn} \end{bmatrix} \Rightarrow \begin{bmatrix} x_{11} \\ \vdots \\ x_{1n} \\ x_{21} \\ \vdots \\ x_{2n} \\ \vdots \\ x_{m1} \\ \vdots \\ x_{mn} \end{bmatrix}$$

Why not represent an image as a matrix? As we will see in the rest of this note, we can perform any linear operation on a vector using matrix-vector multiplication. So even though an image looks like a matrix, it is mathematically easier to represent it as a vector.

What about a color image? The image could be stored in a vector in $\mathbb{R}^{3mn}$, because each of the mn pixels has values for red, green, and blue.

Example 2.15 (Illumination pattern): The imaging lab, in which individual pixels are illuminated by a projector, is very similar to the tomography example, in which individual bottles are illuminated depending on the angle of light being applied to the box. The pixel illumination pattern can similarly be represented as a square matrix of 1's and 0's, with a row for each photodiode measurement and a column for each pixel in the image. If a particular pixel is illuminated during a particular measurement, the matrix contains a 1 at the appropriate row and column index; otherwise, it contains a 0. In the tomography example, we set up a system of linear equations and built a matrix of coefficient weights corresponding to which bottle was included within a particular measurement.

2.5.2 Special Matrices

Definition 2.5 (Zero Matrix): Similar to a zero vector, a zero matrix is a matrix with all components equal to zero. It is usually just represented as 0, where its size is implied from context. Here is an example of a 2×2 zero matrix:

$$0 = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$$

Definition 2.6 (Identity Matrix): An identity matrix, often written as I, is a square matrix whose diagonal elements are 1 and whose off-diagonal elements are all 0. Here is an example of the 3×3 identity matrix:

$$I = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Note that the n column vectors (and the transpose of the row vectors) that make up an $n \times n$ identity matrix are the unit vectors in $\mathbb{R}^n$. The identity matrix is useful because multiplying it with a vector $\vec{x}$ will leave the vector unchanged - in other words, $I\vec{x} = \vec{x}$. We will prove this fact later in this note.

In fact, we will see that multiplying a square matrix by an identity matrix of the same size will yield the same initial matrix: $AI = IA = A$ (we will discuss matrix-matrix multiplication further in the next note).

2.5.3 Column vs Row Perspective

The interpretation of rows and columns from the system-of-linear-equations perspective of doing experiments. This highlights that both the rows and columns of a matrix have importance. Each row of the matrix represents a particular experiment that took one particular measurement. For a given row, the coefficient entries represent how much the corresponding state variable affects the outcome of this *particular* experiment. In contrast, the columns represent the influence of a particular state variable on the *collection* of experiments taken together. These perspectives come in handy in interpreting matrix multiplication.

2.6 Matrix Operations

2.6.1 Matrix Addition

Two matrices of the same size can be added together by adding their corresponding components. For instance, we can add two matrices A and B (both in $\mathbb{R}^{m\times n}$) as follows:

$$A+B=\begin{bmatrix} A_{11} & A_{12} & \dots & A_{1n} \\ A_{21} & A_{22} & \dots & A_{2n} \\ \vdots & \vdots & & \vdots \\ A_{m1} & A_{m2} & \dots & A_{mn} \end{bmatrix}+\begin{bmatrix} B_{11} & B_{12} & \dots & B_{1n} \\ B_{21} & B_{22} & \dots & B_{2n} \\ \vdots & \vdots & & \vdots \\ B_{m1} & B_{m2} & \dots & B_{mn} \end{bmatrix}=\begin{bmatrix} A_{11}+B_{11} & A_{12}+B_{12} & \dots & A_{1n}+B_{1n} \\ A_{21}+B_{21} & A_{22}+B_{22} & \dots & A_{2n}+B_{2n} \\ \vdots & \vdots & & \vdots \\ A_{m1}+B_{m1} & A_{m2}+B_{m2} & \dots & A_{mn}+B_{mn} \end{bmatrix}$$

Here is a specific example of adding two 3×2 matrices:

$$\begin{bmatrix} -1 & 3 \\ 3.5 & 2 \\ 0 & -0.1 \end{bmatrix}+\begin{bmatrix} 2 & -1 \\ -1 & -2 \\ 3 & 0.1 \end{bmatrix}=\begin{bmatrix} 1 & 2 \\ 2.5 & 0 \\ 3 & 0 \end{bmatrix}$$

Note that if two matrices are not the same size, we cannot add them. Intuitively, this is because there is no obvious way to pair up each of their entries.

Properties of Matrix Addition: Just as many of the properties of scalar addition hold for vector addition, many of the same properties hold for matrix addition as well. For three matrices A, B, and C of the same size (and a zero matrix of the same size), the following properties are true:

$$\begin{array}{rrcl} \textbf{Commutativity}: & A+B & = & B+A \\ \textbf{Associativity}: & (A+B)+C & = & A+(B+C) \\ \textbf{Additive identity}: & A+0 & = & A \\ \textbf{Additive inverse}: & A+(-A) & = & 0 \end{array}$$

2.6.2 Scalar-Matrix Multiplication

As with scalar-vector multiplication, multiplying a matrix by a scalar requires multiplying each component of the matrix by that scalar. For instance, to multiply a scalar α by a matrix $A\in\mathbb{R}^{m\times n}$, we do the following:

$$\alpha A=\alpha\begin{bmatrix} A_{11} & A_{12} & \dots & A_{1n} \\ A_{21} & A_{22} & \dots & A_{2n} \\ \vdots & \vdots & & \vdots \\ A_{m1} & A_{m2} & \dots & A_{mn} \end{bmatrix}=\begin{bmatrix} \alpha A_{11} & \alpha A_{12} & \dots & \alpha A_{1n} \\ \alpha A_{21} & \alpha A_{22} & \dots & \alpha A_{2n} \\ \vdots & \vdots & & \vdots \\ \alpha A_{m1} & \alpha A_{m2} & \dots & \alpha A_{mn} \end{bmatrix}$$

Here is a specific example for a 3×2 matrix:

$$(-3)\begin{bmatrix} -1 & 3 \\ 3.5 & 2 \\ 0 & -0.1 \end{bmatrix}=\begin{bmatrix} 3 & -9 \\ -10.5 & -6 \\ 0 & 0.3 \end{bmatrix}$$

Example 2.16 (Negative Matrix): To obtain a negative matrix $-A$, we can just multiply the original matrix A by the scalar -1. In other words, $-A = -1 \times A$.

Example 2.17 (Zero Matrix): We can obtain the zero matrix by multiplying any matrix by 0: $0 \times A = 0$

Properties of Scalar-Matrix Multiplication: Scalar-matrix multiplication has many of the same properties as scalar-scalar multiplication and scalar-vector multiplication. Specifically, for scalars α, β and matrices $A, B \in \mathbb{R}^{m \times n}$, the following properties hold:

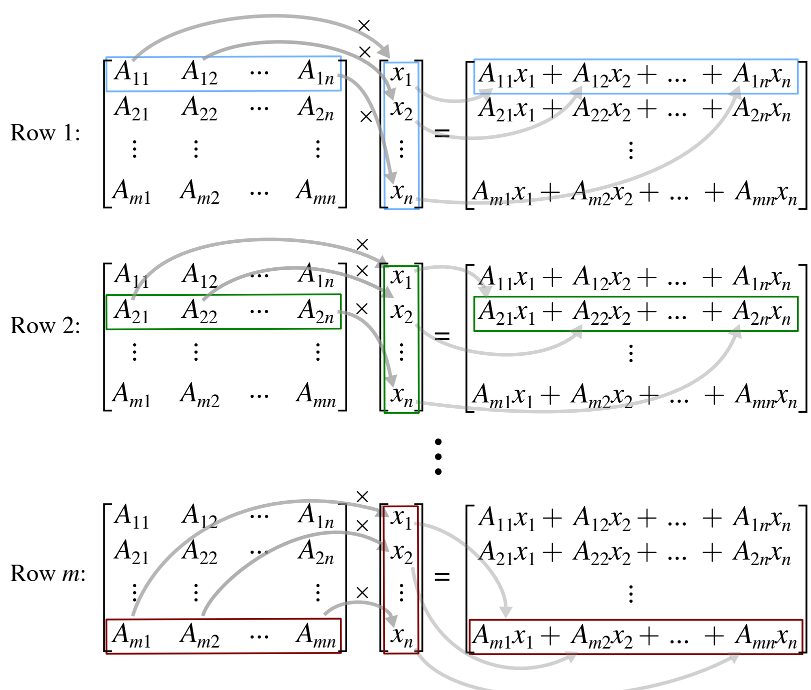
$$\begin{aligned}
\textbf{Associativity}: \quad (\alpha\beta) \times A &= (\alpha) \times (\beta A) \\
\textbf{Distributivity}: \quad (\alpha + \beta) \times A &= \alpha A + \beta A \text{ and } \alpha(A+B) = \alpha A + \alpha B \\
\textbf{Multiplicative identity}: \quad 1 \times A &= A
\end{aligned}$$

2.6.3 Matrix-Vector Multiplication

A matrix $A \in \mathbb{R}^{m \times n}$ and a vector $\vec{x} \in \mathbb{R}^n$ can be multiplied together to create a new vector of length m as follows:

$$A\vec{x} = \begin{bmatrix} A_{11} & A_{12} & \dots & A_{1n} \\ A_{21} & A_{22} & \dots & A_{2n} \\ \vdots & \vdots & & \vdots \\ A_{m1} & A_{m2} & \dots & A_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} A_{11}x_1 + A_{12}x_2 + \cdots + A_{1n}x_n \\ A_{21}x_1 + A_{22}x_2 + \cdots + A_{2n}x_n \\ \vdots \\ A_{m1}x_1 + A_{m2}x_2 + \cdots + A_{mn}x_n \end{bmatrix}$$

Each entry of the resulting vector is the inner product of the corresponding row of A with $\vec{x}$. We perform the multiplication by doing inner product calculations on each row, one at a time:



Mathematically speaking, if we say $A\vec{x} = \vec{b}$, with $A \in \mathbb{R}^{m \times n}$, $\vec{x} \in \mathbb{R}^n$, and $\vec{b} \in \mathbb{R}^m$, element i of $\vec{b}$ can be calculated as follows:

$$b_i = \sum_{j=1}^{n} A_{ij} x_j$$

We can also see that the matrix-vector product can be rewritten as a sum of vectors, which can be expressed as a linear combination of the columns of A:

$$\begin{aligned} A\vec{x} = \begin{bmatrix} A_{11}x_1 + A_{12}x_2 + \cdots + A_{1n}x_n \\ A_{21}x_1 + A_{22}x_2 + \cdots + A_{2n}x_n \\ \vdots \\ A_{m1}x_1 + A_{m2}x_2 + \cdots + A_{mn}x_n \end{bmatrix} &= \begin{bmatrix} A_{11}x_1 \\ A_{21}x_1 \\ \vdots \\ A_{m1}x_1 \end{bmatrix} + \begin{bmatrix} A_{12}x_2 \\ A_{22}x_2 \\ \vdots \\ A_{m2}x_2 \end{bmatrix} + \cdots + \begin{bmatrix} A_{1n}x_n \\ A_{2n}x_n \\ \vdots \\ A_{mn}x_n \end{bmatrix} \\ &= x_1 \begin{bmatrix} A_{11} \\ A_{21} \\ \vdots \\ A_{m1} \end{bmatrix} + x_2 \begin{bmatrix} A_{12} \\ A_{22} \\ \vdots \\ A_{m2} \end{bmatrix} + \cdots + x_n \begin{bmatrix} A_{1n} \\ A_{2n} \\ \vdots \\ A_{mn} \end{bmatrix} \end{aligned}$$

How can this be interpreted? Consider the following example:

$$\begin{bmatrix} -1 & 3 \\ 3.5 & 2 \\ 0 & -0.1 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 3 \\ 2 \\ -0.1 \end{bmatrix}$$

Multiplying a matrix by a standard unit vector effectively selects a single column of the matrix — in this case, the second column. Matrix-vector multiplication is a way to compute a linear combination of the columns in the matrix, where the weights are given by the vector that is being multiplied.

A particular example of matrix-vector multiplication occurs when the matrix we're using is the identity matrix. Earlier, we mentioned that $I\vec{x} = \vec{x}$ for all vectors $\vec{x}$. Now, let's try to prove this. Assume that $\vec{x}$ is an n-dimensional column vector, so we can write

$$\vec{x} = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix}^T.$$

Now, by the above definition of matrix-vector multiplication, we see that

$$I\vec{x} = \begin{bmatrix} \sum_{j=1}^{n} I_{1j}x_j \\ \sum_{j=1}^{n} I_{2j}x_j \\ \cdots \\ \sum_{j=1}^{n} I_{nj}x_j \end{bmatrix}$$

However, recall that the identity matrix has 0s everywhere except along its main diagonal, where it has 1s. In other words, we can express its elements as

$$\begin{aligned} I_{ij} &= 0 \quad \text{if } i \neq j \\ I_{ij} &= 1 \quad \text{if } i = j. \end{aligned}$$

Thus, we can write a single term of $I\vec{x}$ as

$$\begin{aligned}(I\vec{x})_i &= \sum_{j=1}^{n} I_{ij}x_j \\ &= I_{i1}x_1 + I_{i2}x_2 + \ldots + I_{in}x_n \\ &= 0\cdot x_1 + 0\cdot x_2 + \ldots + 0\cdot x_{i-1} + 1\cdot x_i + 0\cdot x_{i+1} + \ldots + 0\cdot x_n \\ &= \vec{x}_i.\end{aligned}$$

If $(I\vec{x})_i = x_i$ for all i, then

$$I\vec{x} = \vec{x},$$

since if all the components of two vectors are equal, then the two vectors are themselves equal. So we have now proven that multiplying the identity matrix with a vector leaves the vector unchanged, as we expected.

2.7 Matrix-Vector Form for Systems of Linear Equations

Now, let us finally return to systems of linear equations. Consider a system of m linear equations with n unknowns $x_1,\ldots,x_n$. As we've seen, such a system of equations can be written in the general form

$$\begin{aligned}a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n &= b_2 \\ &\vdots \\ a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n &= b_m,\end{aligned} \tag{2}$$

where the a_{ij} and b_i are all real-valued constants.

We previously introduced the augmented matrix notation

$$\left[\begin{array}{cccc|c} a_{11} & a_{12} & \ldots & a_{1n} & b_1 \\ a_{21} & a_{22} & \ldots & a_{2n} & b_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{m1} & a_{m2} & \ldots & a_{mn} & b_m \end{array}\right],$$

which was useful as a bookkeeping device for systematically determining solutions (e.g., using Gaussian elimination).

Now, using matrix-vector multiplication, observe that we can also write this system in **matrix-vector form**:

$$\begin{bmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \ldots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}.$$

If we denote the coefficient matrix by A, the variable vector by $\vec{x}$ and the vector of constant terms by $\vec{b}$ then the system of equations can be concisely written as

$$A\vec{x} = \vec{b}.$$

Unsurprisingly, this is called the **matrix-vector form** (or **matrix-vector representation**) of the system of linear equations Eq. (Eq. 2).

Now that we understand how to multiply matrices with vectors, we can revisit this representation of systems of equations, to demonstrate that it is algebraically equivalent to the original system.

To do so, we can evaluate the matrix-vector multiplication on the left-hand-side, to obtain

$$\begin{bmatrix} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{1n}x_n \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}.$$

Notice that the left-hand-sides of our original system of equations have suddenly appeared as components of the vector on the left-hand-side! Since equating two vectors is the same as equating their corresponding coefficients, we can rewrite the above vector equation as a set of n scalar equations:

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n &= b_2 \\ &\vdots \\ a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n &= b_m, \end{aligned}$$

which are exactly the equations in our linear system.

Thus, we have seen that matrix-vector form is not merely an alternate *notation* for a linear system of equations, but in fact is *algebraically equivalent* to the linear system, since we can use the definition of matrix-vector multiplication to go from one to another.

Additional Resources For more on matrix-vector multiplication, read *Strang* page 59. For additional practice with these ideas, try Problem Set 2.3 in *Strang*.

2.8 Practice Problems

1. Is $(\vec{x}+\vec{y})\mathbf{A}$ equivalent to $\mathbf{A}(\vec{x}+\vec{y})$ where $\vec{x}$ and $\vec{y}$ are column vectors in $\mathbb{R}^n$ and $\mathbf{A}$ is an $n \times n$ matrix?
2. True or False: If $\mathbf{A}^2 = \mathbf{0}$, where $\mathbf{0}$ is the zero matrix, then $\mathbf{A} = \mathbf{0}$.
3. Let $\mathbf{A}$ be a 4×4 transformation matrix. You know that $\mathbf{A}\begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 1 \\ 0 \end{bmatrix}$. If possible, find the second column of $\mathbf{A}$.

4. Multiply $\begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix}$ with $\begin{bmatrix} 7 \\ 8 \end{bmatrix}$ using both the element-wise and linear combination interpretations of matrix-vector multiplication, and verify that you get the same answer for this particular example.

5. The effects on the current state, $\vec{x}$, of a fighter jet performing a Barrel Roll, an Immelman, or a Nose-Dive can be described by the matrices $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$, respectively. The jet's final state after a Nose-Dive, followed by a Barrel Roll, followed by an Immelman, can be described by the expression:

   (a) $\mathbf{CAB}\vec{x}$

   (b) $\mathbf{BAC}\vec{x}$

   (c) $\mathbf{CBA}\vec{x}$

   (d) $\mathbf{ABC}\vec{x}$