STEM for VA:
Linear Algebra

Mohan Shankar
mjs7eek@virginia.edu

These notes entail what I used when giving the very first talk of STEM for Virginia on **February 28th, 2024**. They are self-contained for the scope of the talk, requiring nothing more than middle school math—I think. It's been a while since I've been in middle school though. This is not to say that linear algebra is easy or limited in scope; all I mean is that the things I show will be fairly straightforward, belying the full complexity of the subject. That being said, I hope the talk still conveys some of its beauty and utility.

Abstract: The hero of today's story will be a matrix. The "villain," a system of linear equations. This document will introduce readers to two ways of solving a linear system of equations using matrix-vector notation. We also introduce the concept of an eigenvalue (λ_i) and eigenvector ($\vec{v}_\lambda$) though we omit any explanation for a systematic way to find them. Equipped with this knowledge, we will present readers a few simplified models to display the real-world utility of these concepts.

Contents

0 Introduction

To avoid shooting myself in the foot when considering future talks, I'd like to define some things more formally here. The three main objects you will encounter in linear algebra are sets, vector spaces, and vectors.

Set: A collection of non-repeating objects. If a certain object is in a set, it is referred to as an "element" of the set. Curly brackets { } will be used to denote something is a set.

Example. The set of all pitbulls, the set of all real numbers ($\mathbb{R}$), the set of all complex numbers ($\mathbb{C}$), etc.

Vector Space: A *linear* vector space $\mathcal{V}$ is a collection of objects $|1>, |2>, ..., |v>, |w>, ...$ called vectors for which there exists

I. A definite rule for a vector sum denoted $|v> + |w>$

II. A definite rule for multiplication by scalars (i.e. numbers) $\alpha, \beta, \gamma, ...$ denoted by $\alpha|v>$

where the result of these operations results in another element of the vector space— a result known as closure:

$$|v> + |w> = |x> \in \mathcal{V}$$

$$\alpha|v> = |z> \in \mathcal{V}$$

Here, '$\in$' denotes membership of a set. Here, I use $|v>$ rather than $\vec{v}$ to introduce the abstraction that says a vector needn't be an arrow.

Example. Examples of vector spaces include $\mathbb{R}^n$ which is a vector space of a column vector with n elements which are real numbers, $P_n(\mathbb{R})$ which is the vector space containing polynomials of degree n, $\mathbb{R}^{nxm}$ which is a vector space of nxm matrices where n can equal m, etc. Similar to subsets of sets, you can have subspaces of vector spaces.

Numbers $\alpha, \beta, ...$ are the *field* over which the vector space is defined. A field is an *algebraic structure*, and is a set on which addition, subtraction, multiplication, and division are defined. Frankly, I'm not sure what the implication of this is other than to say vectors themselves are neither real nor complex, only their elements.

For the sake of this talk, we can take a "vector" to mean a column vector which I'll denote with an arrow over a lowercase letter as such: $\vec{v}$. If there is a subscript with the arrow like $\vec{v}_i$, I'm denoting the vector **has** i components. A lowercase letter without the arrow but with a subscript like v_i corresponds to the ith component of a vector.

Example.

$$\vec{v} = \begin{bmatrix} 1 \\ 5 \\ \pi \end{bmatrix}$$

is an example of a vector in $\mathbb{R}^3$ meaning our column has three components, all of which are real numbers. $\vec{v}_i$ would mean i = 3, and starting our indexing at 1, v_2 would be 5 while $v_3 = \pi$.

Matrices will be denoted by capital letters with a fancy hat such as $\hat{A}$. They will be vectors of the vector space, $\mathbb{R}^{n\times m}$, n referring to the number of rows and m the number of columns. Similarly to the vector, dropping the hat and adding subscripts means I'm referring to a particular component of that matrix.

Example.

$$\hat{A} = \begin{bmatrix} 1 & 7 \\ 2 & e \end{bmatrix}$$

$\hat{A}_{ij} = \hat{A}_{2,2}$ where i corresponds to the number of rows the matrix has and j the number of columns. Indexing the same way, $A_{1,1} = 1, A_{1,2} = 7, A_{2,1} = 2, A_{2,2} = e$.

1 Linear Systems of Equations and Gaussian Elimination

The central problem of linear algebra is solving *linear* equations. The nicest case is when the number of unknowns matches the number of equations:

$$1x + 2y = 3 \tag{1.1}$$
$$4x + 5y = 6 \tag{1.2}$$

where x and y are unknowns. There are two key procedures we can use here:

I. Gaussian Elimination: The essence of this method is to use "row operations" to simplify your system. The three row operations are multiplying both sides of an equation by a number (scalar), adding/subtracting equations from one another, and re-ordering equations (i.e. 1.1 becomes 1.2 and vice versa).

Example. Here, we can do the following: 1.2 - 4 $\times$ 1.1 which leaves

$$\begin{aligned} 1x + 2y &= 3 \\ 0x - 3y &= -6 \end{aligned}$$

We can easily see that y = 2. From here, we can substitute 2 for y in the first equation to see $1x + 2(2) = 3 \rightarrow x = -1$. You can, of course, check to see that these values of x and y satisfy 1.1 and 1.2 by substituting them back in.

II. Determinants I'll write down the formula to solve for y and x before explaining why it holds true. The key is that y and x are completely specified by the six numbers (1, 2, 3, 4, 5, 6) in the equations.

$$y = \frac{\begin{vmatrix} 1 & 3 \\ 4 & 6 \end{vmatrix}}{\begin{vmatrix} 1 & 2 \\ 4 & 5 \end{vmatrix}} = \frac{1 \cdot 6 - 3 \cdot 4}{1 \cdot 5 - 2 \cdot 4} = \frac{-6}{-3} = 2$$

$$x = \frac{\begin{vmatrix} 3 & 2 \\ 6 & 5 \end{vmatrix}}{\begin{vmatrix} 1 & 2 \\ 4 & 5 \end{vmatrix}} = \frac{3 \cdot 5 - 2 \cdot 6}{1 \cdot 5 - 2 \cdot 4} = \frac{3}{-3} = -1$$

Here, the vertical lines || denote the determinant of the entries they enclose. You can alternatively write $\det(\hat{A})$ where $\hat{A}$ is a matrix as both specify the same procedure. To motivate where it comes from, let's imagine a general system of two equations where we wish to find x and y:

$$ax + by = r_1 \tag{1.3}$$
$$cx + dy = r_2 \tag{1.4}$$

If we multiply 1.3 by d and 1.4 by b:

$$\begin{aligned} dax + dby &= dr_1 \\ bcx + bdy &= br_2 \end{aligned}$$

Subtracting the two equations produces

$$\begin{aligned} (da - bc)x &= dr_1 - br_2 \\ bcx + bdy &= br_2 \end{aligned}$$

Focusing on the top equation now, we see

$$x = \frac{dr_1 - br_2}{da - bc}$$

in general. This is the exact same procedure we leverage above! If you want to find y, you instead multiply by a and c then follow the same process.

Note. If you had a very large number of equations and unknowns, the determinant route of solving the system of equations would work, but it would be remarkably slow. Instead, we'd opt for Gaussian elimination as the algorithms developed for its implementation have a much better (i.e. faster) run time.

1.1 Matrix and Vector Notation

As mentioned in the introduction of this text, I'll denote a **matrix** using an uppercase letter with a hat, and, if I remember, subscripts that represent the number of rows and columns the matrix has.

$$\hat{A}_{ij} = \begin{bmatrix} A_{11} & A_{12} & A_{13} & \dots & A_{1j} \\ A_{21} & A_{22} & A_{23} & \dots & A_{2j} \\ A_{31} & A_{32} & A_{33} & \dots & A_{3j} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ A_{i1} & A_{i2} & A_{i3} & \dots & A_{ij} \end{bmatrix}$$

Here, we see that $\hat{A}$ has i rows and j columns as specified by its subscript. Dropping the hat but leaving the subscript refers to that specific element of $\hat{A}$.

A **vector** will be denoted using a lowercase letter with an arrow above it. A subscript will also be added should I remember. Similarly to the matrix case, a lowercase letter with subscript but without an arrow refers to that component of the vector.

$$\vec{v}_i = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_i \end{bmatrix}$$

The above mathematical object is called a "column vector" and its counterpart, a "row vector" would be the same thing just turned on its side. To get from a column to a row vector and vice versa, you take what's called the *transpose* denoted by a superscript T. Going back to $\vec{v}_i$ above, we can write:

$$\vec{v}_i = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_i \end{bmatrix} \rightarrow \vec{v}_i^T = \begin{bmatrix} v_1 & v_2 & \dots & v_i \end{bmatrix}$$

Though this idea won't come back up for a bit, I introduce you now to acquaint you with the idea that a column vector is an $i \times 1$ matrix. That is, a column vector has i rows and 1 column. On the other hand, a row vector is a $1 \times j$ matrix meaning it has 1 row and j columns. The heart of the transpose lies in *swapping* these indices, thus it can apply to $i \times j$ matrices as well.

Now, let's get back to solving equations. Suppose we now have 3 equations and 3 unknowns:

$$2u + v + w = 5 \tag{1.5}$$
$$4u - 6v + 0w = -2 \tag{1.6}$$
$$-2u + 7v + 2w = 9 \tag{1.7}$$

We can re-package this system of equations using vectors and matrices!

$$\hat{A} = \begin{bmatrix} 2 & 1 & 1 \\ 4 & -6 & 0 \\ -2 & 7 & 2 \end{bmatrix} \longrightarrow \vec{x} = \begin{bmatrix} u \\ v \\ w \end{bmatrix} \longrightarrow \vec{b} = \begin{bmatrix} 5 \\ -2 \\ 9 \end{bmatrix} \Longrightarrow \hat{A}\vec{x} = \vec{b}$$

Note. Using A, x, b are entirely arbitrary but have become convention.

For this re-packaging to work, we must **define** matrix vector multiplication (i.e. $\hat{A}\vec{x}$) to reproduce the original system. To do this, we will multiply the elements of row i of $\hat{A}$ by the elements in the column vector before summing them up. As an example, let's consider the first row of $\hat{A}$:

$$\begin{bmatrix} 2 & 1 & 1 \end{bmatrix} \begin{bmatrix} u \\ v \\ w \end{bmatrix} = \begin{bmatrix} 2 \cdot u + 1 \cdot v + 1 \cdot w \end{bmatrix} \tag{1.8}$$

Since we consider things element wise in the equation $\hat{A}\vec{x} = \vec{b}$, we then see that [2u + w + v] = [5], the first element in $\vec{b}$. The key thing to note here is that this process only works if the number of columns in $\hat{A}$ is equal to the number of rows in $\vec{x}$. As a shortcut, we can write out $\hat{A}\vec{x}$ by column:

$$\hat{A}\vec{x} = u \begin{bmatrix} 2 \\ 4 \\ -2 \end{bmatrix} + v \begin{bmatrix} 1 \\ -6 \\ 7 \end{bmatrix} + w \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix} = \begin{bmatrix} 5 \\ -2 \\ 9 \end{bmatrix}$$

Why this is useful is because it will allow us to "drop" $\vec{x}$ and focus solely on the coefficients and constants. We do this by creating an *augmented matrix*.

$$\left[\begin{array}{ccc|c} 2 & 1 & 1 & 5 \\ 4 & -6 & 0 & -2 \\ -2 & 7 & 2 & 9 \end{array}\right]$$

Here, the numbers arranged to the left of the vertical line is the exact same as $\hat{A}$ while the numbers to the right represent $\vec{b}$. We can and will use Gaussian elimination to shave down $\hat{A}$ until it has only 1's on the diagonal. Considering the initial system of equations, this would be the same as writing

$$\begin{aligned} 1u + 0v + 0w &= * \\ 0u + 1v + 0w &= * \\ 0u + 0v + 1w &= * \end{aligned}$$

thus we wouldn't have to bother with back substitution.

$$\left[\begin{array}{ccc|c} 2 & 1 & 1 & 5 \\ 4 & -6 & 0 & -2 \\ -2 & 7 & 2 & 9 \end{array}\right] \cong \left[\begin{array}{ccc|c} 1 & 1/2 & 1/2 & 5/2 \\ 4 & -6 & 0 & -2 \\ -2 & 7 & 2 & 9 \end{array}\right] \cong \left[\begin{array}{ccc|c} 1 & 1/2 & 1/2 & 5/2 \\ 0 & -8 & 2 & -12 \\ -2 & 7 & 2 & 9 \end{array}\right] \rightarrow \ldots \rightarrow\cong \left[\begin{array}{ccc|c} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 2 \end{array}\right]$$

Here, "$\cong$" denotes an equivalence that is not the same as equality. The systems remain unchanged in the sense that we haven't done anything mathematically illegal, but one form isn't *equal* to the next.

To get from the first form to the second, we divide the first row by 2 to get a 1 in the first column. From there, we perform Row 2 - 4 x Row 1 to get a zero in the second row's first column. We would then make the second column's second row element 1 and clear out everything else in the column, leaving only 1's on the diagonal. While we do this, we must make sure we're also updating the value of $\vec{b}$ when performing these row operations. Using our shortcut:

$$\hat{A}\vec{x} = u \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + v \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} + w \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix} = \vec{b}$$

thus we see u = 1, v = 1, and w = 2.

2 Matrix Operations

As shown in 1.8, matrix vector multiplication works by multiplying across rows and down columns. We can extend this definition to matrix-matrix multiplication by iterating through the columns of the left matrix and down the rows of the right matrix. Pictorially, it would look something like this:

$$\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix} = \begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix}$$

$$\begin{aligned} A_{11} \times B_{11} + A_{12} \times B_{21} &= C_{11} \\ A_{11} \times B_{12} + A_{12} \times B_{22} &= C_{12} \\ A_{12} \times B_{11} + A_{22} \times B_{21} &= C_{21} \\ A_{12} \times B_{12} + A_{22} \times B_{22} &= C_{22} \end{aligned}$$

Figure 1: Schematic for matrix multiplication.

If we want c_2, we would take the same row of A next column of B,

$$\begin{bmatrix} a_1 & a_2 & a_3 \end{bmatrix} \begin{bmatrix} b_2 \\ b_5 \\ b_8 \end{bmatrix}$$

Symbolically, we can express matrix-matrix multiplication as

$$\hat{A}_{ij}\hat{B}_{jk} = \hat{C}_{ik} \rightarrow \sum_j A_{ij}B_{jk} = C_{ik} \tag{2.1}$$

Here, the '$\sum$' symbol means "sum" while the j means sum over the jth elements. i corresponds to the rows of $\hat{A}$ while j the columns. In the case of $\hat{B}$, j corresponds to the rows of $\hat{B}$ while k the columns.

To find the element C_{11}, we let i, j, k be 1. If we refer back to Fig. 1, $A_{1,j}$ refers to the row a_1, a_2, a_3 and $B_{j,1}$ is the column b_1, b_4, b_7 since j starts at 1 and ends at 3. $a_1 \cdot b_1 + a_2 \cdot b_4 + a_3 \cdot b_7 = c_{11}$. Based on this formula, we see that for matrix multiplication to work, the left matrix must have the same number of columns as the number of rows in the right columns.

Example.

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} 5 \\ 6 \end{bmatrix} = \begin{bmatrix} 1\cdot 5 + 2\cdot 6 \\ 3\cdot 5 + 4\cdot 6 \end{bmatrix} = \begin{bmatrix} 17 \\ 35 \end{bmatrix}$$

Example.

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} 5 & 6 \\ 7 & 8 \end{bmatrix} = \begin{bmatrix} 1\cdot 5 + 2\cdot 6 & 1\cdot 6 + 2\cdot 8 \\ 3\cdot 5 + 4\cdot 6 & 3\cdot 6 + 4\cdot 8 \end{bmatrix} = \begin{bmatrix} 17 & 22 \\ 35 & 50 \end{bmatrix}$$

Out of curiosity, what happens if we switch the order of multiplication?

$$\begin{bmatrix} 5 & 6 \\ 7 & 8 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} = \begin{bmatrix} 5\cdot 1 + 6\cdot 3 & 5\cdot 2 + 6\cdot 4 \\ 7\cdot 1 + 8\cdot 3 & 7\cdot 2 + 8\cdot 4 \end{bmatrix} = \begin{bmatrix} 23 & 34 \\ 31 & 46 \end{bmatrix}$$

From this example, we see that matrix multiplication **does not always** commute. That is, $\hat{A}\hat{B} \neq \hat{B}\hat{A}$ all the time.

2.1 Eigenstuff

For our next set of examples, let

$$\hat{A} = \begin{bmatrix} 1 & 2 \\ 0 & 3 \end{bmatrix} \longrightarrow \vec{v} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \longrightarrow \vec{u} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}.$$

If we multiply $\hat{A}\vec{v}$ and $\hat{A}\vec{u}$,

$$\begin{bmatrix} 1 & 2 \\ 0 & 3 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1\cdot 1 + 2\cdot 1 \\ 0\cdot 1 + 3\cdot 1 \end{bmatrix} = \begin{bmatrix} 3 \\ 3 \end{bmatrix} = 3 \cdot \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 2 \\ 0 & 3 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 1\cdot 1 + 2\cdot 0 \\ 0\cdot 1 + 3\cdot 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} = 1 \cdot \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

Unlike our last examples, these matrix-vector multiplications returned our initial vectors and a scalar. We've found something very special! It turns out that there is a special name for this relationship between matrix, vector, and scalar.

$\vec{v}$ and $\vec{u}$ are *eigenvectors* of the matrix $\hat{A}$. Their respective *eigenvalues* are 3 and 1. It's interesting to note that *eigen-* means "own" or "inherent" in German, which is why it crops up here. The eigenvectors and their corresponding eigenvalues are inherent to the matrix itself as they don't get transformed beyond scaling (i.e. getting stretched or squished). There's a way to systematically determine all eigenvalues and eigenvectors in that order using the determinant, but I'll mention it at a later talk. If you look into it yourself, you might find the equation

$$\hat{A}\vec{v} = \lambda\vec{v} \tag{2.2}$$

Where $\lambda \in \mathbb{R}$ is the eigenvalue and $\vec{v}$ its associated eigenvector for $\hat{A}$.

2.2 General Topics

In the event that future talks require linear algebra, I'd like to cover these topics now.

Matrix Exponentiation: Repeated matrix multiplication

$$\hat{A}^n = \underbrace{\hat{A} \cdot \hat{A} \cdot ... \cdot \hat{A}}_{\text{n times}}$$

Matrix Inverse:

$$\text{If } \exists\ \hat{A}^{-1} \text{ s.t. } \hat{A}\hat{A}^{-1} = \hat{A}^{-1}\hat{A} = I$$

$\hat{A}^{-1}$ is the inverse of $\hat{A}$ as multiplying $\hat{A}$ with its inverse produces I the identity matrix.

Example.

$$\hat{A} = \begin{bmatrix} 2 & 1 \\ 1 & 3 \end{bmatrix} \text{ and } \hat{A}^{-1} = \begin{bmatrix} 3/5 & -1/5 \\ -1/5 & 2/5 \end{bmatrix}$$

Let's try $\hat{A}\hat{A}^{-1}$ first:

$$\hat{A}\hat{A}^{-1} = \begin{bmatrix} 2 & 1 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} 3/5 & -1/5 \\ -1/5 & 2/5 \end{bmatrix} = \begin{bmatrix} \{2 \cdot (3/5) + 1 \cdot (-1/5)\} & \{2 \cdot (-1/5) + 1 \cdot (2/5)\} \\ \{1 \cdot (3/5) + 3 \cdot (-1/5)\} & \{1 \cdot (-1/5) + 3 \cdot (2/5)\} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Now, $\hat{A}^{-1}\hat{A}$

$$\hat{A}^{-1}\hat{A} = \begin{bmatrix} 3/5 & -1/5 \\ -1/5 & 2/5 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ 1 & 3 \end{bmatrix} = \begin{bmatrix} \{(3/5) \cdot 2 + (-1/5) \cdot 1\} & \{(3/5) \cdot 1 + (-1/5) \cdot 3\} \\ \{-(1/5) \cdot 2 + (2/5) \cdot 1\} & \{-(1/5) \cdot 1 + (2/5) \cdot 3\} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Transpose: Swapping elements via flipping indices. $\hat{A}^T_{ij} = \hat{A}_{ji}$ which is useful when making dimensions for matrix-matrix, matrix-vector, or vector-vector multiplication work.

Example.

$$\hat{A} = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} \longrightarrow \hat{A}^T = \begin{bmatrix} 1 & 4 & 7 \\ 2 & 5 & 8 \\ 3 & 6 & 9 \end{bmatrix}$$

Note that elements along the diagonal (where i = j in A_{ij}) do not get flipped.