Linear algebra explained in four pages

Excerpt from the NO BULLSHIT GUIDE TO LINEAR ALGEBRA by Ivan Savov

***Abstract*—This document will review the fundamental ideas of linear algebra. We will learn about matrices, matrix operations, linear transformations and discuss both the theoretical and computational aspects of linear algebra. The tools of linear algebra open the gateway to the study of more advanced mathematics. A lot of *knowledge buzz* awaits you if you choose to follow the path of *understanding*, instead of trying to memorize a bunch of formulas.**

I. INTRODUCTION

Linear algebra is the math of vectors and matrices. Let n be a positive integer and let $\mathbb{R}$ denote the set of real numbers, then $\mathbb{R}^n$ is the set of all n-tuples of real numbers. A vector $\vec{v} \in \mathbb{R}^n$ is an n-tuple of real numbers. The notation "$\in S$" is read "element of S." For example, consider a vector that has three components:

$$\vec{v} = (v_1, v_2, v_3) \;\in\; (\mathbb{R}, \mathbb{R}, \mathbb{R}) \equiv \mathbb{R}^3.$$

A matrix $A \in \mathbb{R}^{m \times n}$ is a rectangular array of real numbers with m rows and n columns. For example, a 3×2 matrix looks like this:

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{bmatrix} \in \begin{bmatrix} \mathbb{R} & \mathbb{R} \\ \mathbb{R} & \mathbb{R} \\ \mathbb{R} & \mathbb{R} \end{bmatrix} \equiv \mathbb{R}^{3\times 2}.$$

The purpose of this document is to introduce you to the mathematical operations that we can perform on vectors and matrices and to give you a feel of the power of linear algebra. Many problems in science, business, and technology can be described in terms of vectors and matrices so it is important that you understand how to work with these.

Prerequisites

The only prerequisite for this tutorial is a basic understanding of high school math concepts[1] like numbers, variables, equations, and the fundamental arithmetic operations on real numbers: addition (denoted $+$), subtraction (denoted $-$), multiplication (denoted implicitly), and division (fractions).

You should also be familiar with *functions* that take real numbers as inputs and give real numbers as outputs, $f : \mathbb{R} \to \mathbb{R}$. Recall that, by definition, the *inverse function* f^{-1} *undoes* the effect of f. If you are given $f(x)$ and you want to find x, you can use the inverse function as follows: $f^{-1}(f(x)) = x$. For example, the function $f(x) = \ln(x)$ has the inverse $f^{-1}(x) = e^x$, and the inverse of $g(x) = \sqrt{x}$ is $g^{-1}(x) = x^2$.

II. DEFINITIONS

A. Vector operations

We now define the math operations for vectors. The operations we can perform on vectors $\vec{u} = (u_1, u_2, u_3)$ and $\vec{v} = (v_1, v_2, v_3)$ are: addition, subtraction, scaling, norm (length), dot product, and cross product:

$$\begin{aligned} \vec{u} + \vec{v} &= (u_1 + v_1, u_2 + v_2, u_3 + v_3) \\ \vec{u} - \vec{v} &= (u_1 - v_1, u_2 - v_2, u_3 - v_3) \\ \alpha\vec{u} &= (\alpha u_1, \alpha u_2, \alpha u_3) \\ ||\vec{u}|| &= \sqrt{u_1^2 + u_2^2 + u_3^2} \\ \vec{u} \cdot \vec{v} &= u_1 v_1 + u_2 v_2 + u_3 v_3 \\ \vec{u} \times \vec{v} &= (u_2 v_3 - u_3 v_2,\ u_3 v_1 - u_1 v_3,\ u_1 v_2 - u_2 v_1) \end{aligned}$$

The dot product and the cross product of two vectors can also be described in terms of the angle θ between the two vectors. The formula for the dot product of the vectors is $\vec{u} \cdot \vec{v} = \|\vec{u}\|\|\vec{v}\| \cos\theta$. We say two vectors $\vec{u}$ and $\vec{v}$ are *orthogonal* if the angle between them is $90°$. The dot product of orthogonal vectors is zero: $\vec{u} \cdot \vec{v} = \|\vec{u}\|\|\vec{v}\| \cos(90°) = 0$.

The *norm* of the cross product is given by $\|\vec{u} \times \vec{v}\| = \|\vec{u}\|\|\vec{v}\| \sin\theta$. The cross product is not commutative: $\vec{u} \times \vec{v} \neq \vec{v} \times \vec{u}$, in fact $\vec{u} \times \vec{v} = -\vec{v} \times \vec{u}$.

B. Matrix operations

We denote by A the matrix as a whole and refer to its entries as a_{ij}. The mathematical operations defined for matrices are the following:

- addition (denoted $+$)

$$C = A + B \qquad \Leftrightarrow \qquad c_{ij} = a_{ij} + b_{ij}.$$

- subtraction (the inverse of addition)
- matrix product. The product of matrices $A \in \mathbb{R}^{m\times n}$ and $B \in \mathbb{R}^{n\times \ell}$ is another matrix $C \in \mathbb{R}^{m\times \ell}$ given by the formula

$$C = AB \qquad \Leftrightarrow \qquad c_{ij} = \sum_{k=1}^{n} a_{ik} b_{kj},$$

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{bmatrix} \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} = \begin{bmatrix} a_{11}b_{11} + a_{12}b_{21} & a_{11}b_{12} + a_{12}b_{22} \\ a_{21}b_{11} + a_{22}b_{21} & a_{21}b_{12} + a_{22}b_{22} \\ a_{31}b_{11} + a_{32}b_{21} & a_{31}b_{12} + a_{32}b_{22} \end{bmatrix}$$

- matrix inverse (denoted A^{-1})
- matrix transpose (denoted $^{\mathsf{T}}$):

$$\begin{bmatrix} \alpha_1 & \alpha_2 & \alpha_3 \\ \beta_1 & \beta_2 & \beta_3 \end{bmatrix}^{\mathsf{T}} = \begin{bmatrix} \alpha_1 & \beta_1 \\ \alpha_2 & \beta_2 \\ \alpha_3 & \beta_3 \end{bmatrix}.$$

- matrix trace: $\text{Tr}[A] \equiv \sum_{i=1}^{n} a_{ii}$
- determinant (denoted $\det(A)$ or $|A|$)

Note that the matrix product is not a commutative operation: $AB \neq BA$.

C. Matrix-vector product

The matrix-vector product is an important special case of the matrix-matrix product. The product of a 3×2 matrix A and the 2×1 column vector $\vec{x}$ results in a 3×1 vector $\vec{y}$ given by:

$$\vec{y} = A\vec{x} \qquad \Leftrightarrow \qquad \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} a_{11}x_1 + a_{12}x_2 \\ a_{21}x_1 + a_{22}x_2 \\ a_{31}x_1 + a_{32}x_2 \end{bmatrix}$$

$$= x_1 \begin{bmatrix} a_{11} \\ a_{21} \\ a_{31} \end{bmatrix} + x_2 \begin{bmatrix} a_{12} \\ a_{22} \\ a_{32} \end{bmatrix} \tag{C}$$

$$= \begin{bmatrix} (a_{11}, a_{12}) \cdot \vec{x} \\ (a_{21}, a_{22}) \cdot \vec{x} \\ (a_{31}, a_{32}) \cdot \vec{x} \end{bmatrix}. \tag{R}$$

There are two[2] fundamentally different yet equivalent ways to interpret the matrix-vector product. In the column picture, **(C)**, the multiplication of the matrix A by the vector $\vec{x}$ produces a **linear combination of the columns of the matrix**: $\vec{y} = A\vec{x} = x_1 A_{[:,1]} + x_2 A_{[:,2]}$, where $A_{[:,1]}$ and $A_{[:,2]}$ are the first and second columns of the matrix A.

In the row picture, **(R)**, multiplication of the matrix A by the vector $\vec{x}$ produces a column vector with coefficients equal to the **dot products of rows of the matrix** with the vector $\vec{x}$.

D. Linear transformations

The matrix-vector product is used to define the notion of a *linear transformation*, which is one of the key notions in the study of linear algebra. Multiplication by a matrix $A \in \mathbb{R}^{m\times n}$ can be thought of as computing a *linear transformation* T_A that takes n-vectors as inputs and produces m-vectors as outputs:

$$T_A : \mathbb{R}^n \to \mathbb{R}^m.$$

[1] A good textbook to (re)learn high school math is `minireference.com`

[2] For more info see the video of Prof. Strang's MIT lecture: `bit.ly/1OvmKcL`

Instead of writing $\vec{y} = T_A(\vec{x})$ for the linear transformation T_A applied to the vector $\vec{x}$, we simply write $\vec{y} = A\vec{x}$. Applying the linear transformation T_A to the vector $\vec{x}$ corresponds to the product of the matrix A and the column vector $\vec{x}$. We say T_A is *represented by* the matrix A.

You can think of linear transformations as "vector functions" and describe their properties in analogy with the regular functions you are familiar with:

$$
\begin{aligned}
\text{function } f : \mathbb{R} \to \mathbb{R} \ &\Leftrightarrow\ \text{linear transformation } T_A : \mathbb{R}^n \to \mathbb{R}^m \\
\text{input } x \in \mathbb{R} \ &\Leftrightarrow\ \text{input } \vec{x} \in \mathbb{R}^n \\
\text{output } f(x) \ &\Leftrightarrow\ \text{output } T_A(\vec{x}) = A\vec{x} \in \mathbb{R}^m \\
g \circ f = g(f(x)) \ &\Leftrightarrow\ T_B(T_A(\vec{x})) = BA\vec{x} \\
\text{function inverse } f^{-1} \ &\Leftrightarrow\ \text{matrix inverse } A^{-1} \\
\text{zeros of } f \ &\Leftrightarrow\ \mathcal{N}(A) \equiv \text{null space of } A \\
\text{range of } f \ &\Leftrightarrow\ \mathcal{C}(A) \equiv \text{column space of } A = \text{range of } T_A
\end{aligned}
$$

Note that the combined effect of applying the transformation T_A followed by T_B on the input vector $\vec{x}$ is equivalent to the matrix product $BA\vec{x}$.

E. Fundamental vector spaces

A *vector space* consists of a set of vectors and all linear combinations of these vectors. For example the vector space $\mathcal{S} = \text{span}\{\vec{v}_1, \vec{v}_2\}$ consists of all vectors of the form $\vec{v} = \alpha\vec{v}_1 + \beta\vec{v}_2$, where α and β are real numbers. We now define three fundamental vector spaces associated with a matrix A.

The *column space* of a matrix A is the set of vectors that can be produced as linear combinations of the columns of the matrix A:

$$\mathcal{C}(A) \equiv \{\vec{y} \in \mathbb{R}^m \mid \vec{y} = A\vec{x} \text{ for some } \vec{x} \in \mathbb{R}^n\}.$$

The column space is the *range* of the linear transformation T_A (the set of possible outputs). You can convince yourself of this fact by reviewing the definition of the matrix-vector product in the column picture **(C)**. The vector $A\vec{x}$ contains x_1 times the 1st column of A, x_2 times the 2nd column of A, etc. Varying over all possible inputs $\vec{x}$, we obtain all possible linear combinations of the columns of A, hence the name "column space."

The *null space* $\mathcal{N}(A)$ of a matrix $A \in \mathbb{R}^{m\times n}$ consists of all the vectors that the matrix A sends to the zero vector:

$$\mathcal{N}(A) \equiv \{\vec{x} \in \mathbb{R}^n \mid A\vec{x} = \vec{0}\}.$$

The vectors in the null space are *orthogonal* to all the rows of the matrix. We can see this from the row picture **(R)**: the output vectors is $\vec{0}$ if and only if the input vector $\vec{x}$ is orthogonal to all the rows of A.

The *row space* of a matrix A, denoted $\mathcal{R}(A)$, is the set of linear combinations of the rows of A. The row space $\mathcal{R}(A)$ is the orthogonal complement of the null space $\mathcal{N}(A)$. This means that for all vectors $\vec{v} \in \mathcal{R}(A)$ and all vectors $\vec{w} \in \mathcal{N}(A)$, we have $\vec{v} \cdot \vec{w} = 0$. Together, the null space and the row space form the domain of the transformation T_A, $\mathbb{R}^n = \mathcal{N}(A) \oplus \mathcal{R}(A)$, where $\oplus$ stands for *orthogonal direct sum*.

F. Matrix inverse

By definition, the inverse matrix A^{-1} *undoes* the effects of the matrix A. The cumulative effect of applying A^{-1} after A is the identity matrix $\mathbb{1}$:

$$A^{-1}A = \mathbb{1} \equiv \begin{bmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{bmatrix}.$$

The identity matrix (ones on the diagonal and zeros everywhere else) corresponds to the identity transformation: $T_{\mathbb{1}}(\vec{x}) = \mathbb{1}\vec{x} = \vec{x}$, for all $\vec{x}$.

The matrix inverse is useful for solving matrix equations. Whenever we want to get rid of the matrix A in some matrix equation, we can "hit" A with its inverse A^{-1} to make it disappear. For example, to solve for the matrix X in the equation $XA = B$, multiply both sides of the equation by A^{-1} from the right: $X = BA^{-1}$. To solve for X in $ABCXD = E$, multiply both sides of the equation by D^{-1} on the right and by A^{-1}, B^{-1} and C^{-1} (in that order) from the left: $X = C^{-1}B^{-1}A^{-1}ED^{-1}$.

III. Computational linear algebra

Okay, I hear what you are saying "Dude, enough with the theory talk, let's see some calculations." In this section we'll look at one of the fundamental algorithms of linear algebra called Gauss–Jordan elimination.

A. Solving systems of equations

Suppose we're asked to solve the following system of equations:

$$\begin{aligned} 1x_1 + 2x_2 &= 5, \\ 3x_1 + 9x_2 &= 21. \end{aligned} \tag{1}$$

Without a knowledge of linear algebra, we could use substitution, elimination, or subtraction to find the values of the two unknowns x_1 and x_2.

Gauss–Jordan elimination is a systematic procedure for solving systems of equations based the following *row operations*:

α) Adding a multiple of one row to another row
β) Swapping two rows
γ) Multiplying a row by a constant

These row operations allow us to simplify the system of equations without changing their solution.

To illustrate the Gauss–Jordan elimination procedure, we'll now show the sequence of row operations required to solve the system of linear equations described above. We start by constructing an *augmented matrix* as follows:

$$\left[\begin{array}{cc|c} \mathbf{1} & 2 & 5 \\ 3 & 9 & 21 \end{array}\right].$$

The first column in the augmented matrix corresponds to the coefficients of the variable x_1, the second column corresponds to the coefficients of x_2, and the third column contains the constants from the right-hand side.

The Gauss-Jordan elimination procedure consists of two phases. During the first phase, we proceed left-to-right by choosing a row with a leading one in the leftmost column (called a *pivot*) and systematically subtracting that row from all rows below it to get zeros below in the entire column. In the second phase, we start with the rightmost pivot and use it to eliminate all the numbers above it in the same column. Let's see this in action.

1) The first step is to use the pivot in the first column to eliminate the variable x_1 in the second row. We do this by subtracting three times the first row from the second row, denoted $R_2 \leftarrow R_2 - 3R_1$,
$$\left[\begin{array}{cc|c} \mathbf{1} & 2 & 5 \\ 0 & 3 & 6 \end{array}\right].$$
2) Next, we create a pivot in the second row using $R_2 \leftarrow \frac{1}{3}R_2$:
$$\left[\begin{array}{cc|c} 1 & 2 & 5 \\ 0 & \mathbf{1} & 2 \end{array}\right].$$
3) We now start the backward phase and eliminate the second variable from the first row. We do this by subtracting two times the second row from the first row $R_1 \leftarrow R_1 - 2R_2$:
$$\left[\begin{array}{cc|c} 1 & 0 & 1 \\ 0 & 1 & 2 \end{array}\right].$$
The matrix is now in *reduced row echelon form* (RREF), which is its "simplest" form it could be in. The solutions are: $x_1 = 1$, $x_2 = 2$.

B. Systems of equations as matrix equations

We will now discuss another approach for solving the system of equations. Using the definition of the matrix-vector product, we can express this system of equations (1) as a matrix equation:

$$\begin{bmatrix} 1 & 2 \\ 3 & 9 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 5 \\ 21 \end{bmatrix}.$$

This matrix equation had the form $A\vec{x} = \vec{b}$, where A is a 2×2 matrix, $\vec{x}$ is the vector of unknowns, and $\vec{b}$ is a vector of constants. We can solve for $\vec{x}$ by multiplying both sides of the equation by the matrix inverse A^{-1}:

$$A^{-1}A\vec{x} = \mathbb{1}\vec{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = A^{-1}\vec{b} = \begin{bmatrix} 3 & -\frac{2}{3} \\ -1 & \frac{1}{3} \end{bmatrix} \begin{bmatrix} 5 \\ 21 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}.$$

But how did we know what the inverse matrix A^{-1} is?

IV. Computing the inverse of a matrix

In this section we'll look at several different approaches for computing the inverse of a matrix. The matrix inverse is *unique* so no matter which method we use to find the inverse, we'll always obtain the same answer.

A. Using row operations

One approach for computing the inverse is to use the Gauss–Jordan elimination procedure. Start by creating an array containing the entries of the matrix A on the left side and the identity matrix on the right side:

$$\left[\begin{array}{cc|cc} 1 & 2 & 1 & 0 \\ 3 & 9 & 0 & 1 \end{array}\right].$$

Now we perform the Gauss-Jordan elimination procedure on this array.

1) The first row operation is to subtract three times the first row from the second row: $R_2 \leftarrow R_2 - 3R_1$. We obtain:
$$\left[\begin{array}{cc|cc} 1 & 2 & 1 & 0 \\ 0 & 3 & -3 & 1 \end{array}\right].$$
2) The second row operation is divide the second row by 3: $R_2 \leftarrow \frac{1}{3}R_2$
$$\left[\begin{array}{cc|cc} 1 & 2 & 1 & 0 \\ 0 & 1 & -1 & \frac{1}{3} \end{array}\right].$$
3) The third row operation is $R_1 \leftarrow R_1 - 2R_2$
$$\left[\begin{array}{cc|cc} 1 & 0 & 3 & -\frac{2}{3} \\ 0 & 1 & -1 & \frac{1}{3} \end{array}\right].$$

The array is now in reduced row echelon form (RREF). The inverse matrix appears on the right side of the array.

Observe that the sequence of row operations we used to solve the specific system of equations in $A\vec{x} = \vec{b}$ in the previous section are the same as the row operations we used in this section to find the inverse matrix. Indeed, in both cases the combined effect of the three row operations is to "undo" the effects of A. The right side of the 2×4 array is simply a convenient way to record this sequence of operations and thus obtain A^{-1}.

B. Using elementary matrices

Every row operation we perform on a matrix is equivalent to a left-multiplication by an *elementary matrix*. There are three types of elementary matrices in correspondence with the three types of row operations:

$$\mathcal{R}_\alpha : R_1 \leftarrow R_1 + mR_2 \quad \Leftrightarrow \quad E_\alpha = \begin{bmatrix} 1 & m \\ 0 & 1 \end{bmatrix}$$

$$\mathcal{R}_\beta : R_1 \leftrightarrow R_2 \quad \Leftrightarrow \quad E_\beta = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

$$\mathcal{R}_\gamma : R_1 \leftarrow mR_1 \quad \Leftrightarrow \quad E_\gamma = \begin{bmatrix} m & 0 \\ 0 & 1 \end{bmatrix}$$

Let's revisit the row operations we used to find A^{-1} in the above section representing each row operation as an elementary matrix multiplication.

1) The first row operation $R_2 \leftarrow R_2 - 3R_1$ corresponds to a multiplication by the elementary matrix E_1:
$$E_1 A = \begin{bmatrix} 1 & 0 \\ -3 & 1 \end{bmatrix}\begin{bmatrix} 1 & 2 \\ 3 & 9 \end{bmatrix} = \begin{bmatrix} 1 & 2 \\ 0 & 3 \end{bmatrix}.$$
2) The second row operation $R_2 \leftarrow \frac{1}{3}R_2$ corresponds to a matrix E_2:
$$E_2(E_1 A) = \begin{bmatrix} 1 & 0 \\ 0 & \frac{1}{3} \end{bmatrix}\begin{bmatrix} 1 & 2 \\ 0 & 3 \end{bmatrix} = \begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix}.$$
3) The final step, $R_1 \leftarrow R_1 - 2R_2$, corresponds to the matrix E_3:
$$E_3(E_2 E_1 A) = \begin{bmatrix} 1 & -2 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

Note that $E_3 E_2 E_1 A = \mathbb{1}$, so the product $E_3 E_2 E_1$ must be equal to A^{-1}:

$$A^{-1} = E_3 E_2 E_1 = \begin{bmatrix} 1 & -2 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 0 & \frac{1}{3} \end{bmatrix}\begin{bmatrix} 1 & 0 \\ -3 & 1 \end{bmatrix} = \begin{bmatrix} 3 & -\frac{2}{3} \\ -1 & \frac{1}{3} \end{bmatrix}.$$

The elementary matrix approach teaches us that every invertible matrix can be decomposed as the product of elementary matrices. Since we know $A^{-1} = E_3 E_2 E_1$ then $A = (A^{-1})^{-1} = (E_3 E_2 E_1)^{-1} = E_1^{-1} E_2^{-1} E_3^{-1}$.

C. Using a computer

The last (and most practical) approach for finding the inverse of a matrix is to use a computer algebra system like the one at `live.sympy.org`.

```
>>> A = Matrix( [[1,2],[3,9]] )            # define A
    [1, 2]
    [3, 9]

>>> A.inv()                    # calls the inv method on A
    [ 3, -2/3]
    [-1,  1/3]
```

You can use `sympy` to "check" your answers on homework problems.

V. Other topics

We'll now discuss a number of other important topics of linear algebra.

A. Basis

Intuitively, a basis is any set of vectors that can be used as a coordinate system for a vector space. You are certainly familiar with the standard basis for the xy-plane that is made up of two orthogonal axes: the x-axis and the y-axis. A vector $\vec{v}$ can be described as a coordinate pair (v_x, v_y) with respect to these axes, or equivalently as $\vec{v} = v_x\hat{\imath} + v_y\hat{\jmath}$, where $\hat{\imath} \equiv (1, 0)$ and $\hat{\jmath} \equiv (0, 1)$ are unit vectors that point along the x-axis and y-axis respectively. However, other coordinate systems are also possible.

Definition (Basis). *A basis for a n-dimensional vector space $\mathcal{S}$ is any set of n linearly independent vectors that are part of $\mathcal{S}$.*

Any set of two linearly independent vectors $\{\hat{e}_1, \hat{e}_2\}$ can serve as a basis for $\mathbb{R}^2$. We can write any vector $\vec{v} \in \mathbb{R}^2$ as a linear combination of these basis vectors $\vec{v} = v_1\hat{e}_1 + v_2\hat{e}_2$.

Note the *same* vector $\vec{v}$ corresponds to different coordinate pairs depending on the basis used: $\vec{v} = (v_x, v_y)$ in the standard basis $B_s \equiv \{\hat{\imath}, \hat{\jmath}\}$, and $\vec{v} = (v_1, v_2)$ in the basis $B_e \equiv \{\hat{e}_1, \hat{e}_2\}$. Therefore, it is important to keep in mind the basis with respect to which the coefficients are taken, and if necessary specify the basis as a subscript, e.g., $(v_x, v_y)_{B_s}$ or $(v_1, v_2)_{B_e}$.

Converting a coordinate vector from the basis B_e to the basis B_s is performed as a multiplication by a *change of basis* matrix:

$$\begin{bmatrix} \vec{v} \end{bmatrix}_{B_s} = {}_{B_s}\begin{bmatrix} \ \mathbb{1}\ \end{bmatrix}_{B_e} \begin{bmatrix} \vec{v} \end{bmatrix}_{B_e} \quad \Leftrightarrow \quad \begin{bmatrix} v_x \\ v_y \end{bmatrix} = \begin{bmatrix} \hat{\imath}\cdot\hat{e}_1 & \hat{\imath}\cdot\hat{e}_2 \\ \hat{\jmath}\cdot\hat{e}_1 & \hat{\jmath}\cdot\hat{e}_2 \end{bmatrix}\begin{bmatrix} v_1 \\ v_2 \end{bmatrix}.$$

Note the change of basis matrix is actually an identity transformation. The vector $\vec{v}$ remains unchanged—it is simply expressed with respect to a new coordinate system. The change of basis from the B_s-basis to the B_e-basis is accomplished using the inverse matrix: ${}_{B_e}[\mathbb{1}]_{B_s} = ({}_{B_s}[\mathbb{1}]_{B_e})^{-1}$.

B. Matrix representations of linear transformations

Bases play an important role in the representation of linear transformations $T\colon \mathbb{R}^n \to \mathbb{R}^m$. To fully describe the matrix that corresponds to some linear transformation T, it is sufficient to know the effects of T to the n vectors of the standard basis for the input space. For a linear transformation $T\colon \mathbb{R}^2 \to \mathbb{R}^2$, the matrix representation corresponds to

$$M_T = \begin{bmatrix} | & | \\ T(\hat{\imath}) & T(\hat{\jmath}) \\ | & | \end{bmatrix} \in \mathbb{R}^{2\times 2}.$$

As a first example, consider the transformation Π_x which projects vectors onto the x-axis. For any vector $\vec{v} = (v_x, v_y)$, we have $\Pi_x(\vec{v}) = (v_x, 0)$. The matrix representation of Π_x is

$$M_{\Pi_x} = \left[\Pi_x\!\left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}\right) \ \Pi_x\!\left(\begin{bmatrix} 0 \\ 1 \end{bmatrix}\right) \right] = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}.$$

As a second example, let's find the matrix representation of R_θ, the counterclockwise rotation by the angle θ:

$$M_{R_\theta} = \left[R_\theta\!\left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}\right) \ R_\theta\!\left(\begin{bmatrix} 0 \\ 1 \end{bmatrix}\right) \right] = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}.$$

The first column of M_{R_θ} shows that R_θ maps the vector $\hat{\imath} \equiv 1\angle 0$ to the vector $1\angle\theta = (\cos\theta, \sin\theta)^{\mathsf{T}}$. The second column shows that R_θ maps the vector $\hat{\jmath} = 1\angle\frac{\pi}{2}$ to the vector $1\angle(\frac{\pi}{2} + \theta) = (-\sin\theta, \cos\theta)^{\mathsf{T}}$.

C. Dimension and bases for vector spaces

The *dimension* of a vector space is defined as the number of vectors in a basis for that vector space. Consider the following vector space $\mathcal{S} = \text{span}\{(1,0,0),(0,1,0),(1,1,0)\}$. Seeing that the space is described by three vectors, we might think that $\mathcal{S}$ is 3-dimensional. This is not the case, however, since the three vectors are not linearly independent so they don't form a basis for $\mathcal{S}$. Two vectors are sufficient to describe any vector in $\mathcal{S}$; we can write $\mathcal{S} = \text{span}\{(1,0,0),(0,1,0)\}$, and we see these two vectors are linearly independent so they form a basis and $\dim(\mathcal{S}) = 2$.

There is a general procedure for finding a basis for a vector space. Suppose you are given a description of a vector space in terms of m vectors $\mathcal{V} = \text{span}\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_m\}$ and you are asked to find a basis for $\mathcal{V}$ and the dimension of $\mathcal{V}$. To find a basis for $\mathcal{V}$, you must find a set of linearly independent vectors that span $\mathcal{V}$. We can use the Gauss–Jordan elimination procedure to accomplish this task. Write the vectors $\vec{v}_i$ as the rows of a matrix M. The vector space $\mathcal{V}$ corresponds to the row space of the matrix M. Next, use row operations to find the reduced row echelon form (RREF) of the matrix M. Since row operations do not change the row space of the matrix, the row space of reduced row echelon form of the matrix M is the same as the row space of the original set of vectors. The nonzero rows in the RREF of the matrix form a basis for vector space $\mathcal{V}$ and the numbers of nonzero rows is the dimension of $\mathcal{V}$.

D. Row space, columns space, and rank of a matrix

Recall the fundamental vector spaces for matrices that we defined in Section II-E: the column space $\mathcal{C}(A)$, the null space $\mathcal{N}(A)$, and the row space $\mathcal{R}(A)$. A standard linear algebra exam question is to give you a certain matrix A and ask you to find the dimension and a basis for each of its fundamental spaces.

In the previous section we described a procedure based on Gauss–Jordan elimination which can be used "distill" a set of linearly independent vectors which form a basis for the row space $\mathcal{R}(A)$. We will now illustrate this procedure with an example, and also show how to use the RREF of the matrix A to find bases for $\mathcal{C}(A)$ and $\mathcal{N}(A)$.

Consider the following matrix and its reduced row echelon form:

$$A = \begin{bmatrix} 1 & 3 & 3 & 3 \\ 2 & 6 & 7 & 6 \\ 3 & 9 & 9 & 10 \end{bmatrix} \qquad \text{rref}(A) = \begin{bmatrix} \mathbf{1} & 3 & 0 & 0 \\ 0 & 0 & \mathbf{1} & 0 \\ 0 & 0 & 0 & \mathbf{1} \end{bmatrix}.$$

The reduced row echelon form of the matrix A contains three pivots. The locations of the pivots will play an important role in the following steps.

The vectors $\{(1,3,0,0),(0,0,1,0),(0,0,0,1)\}$ form a basis for $\mathcal{R}(A)$.

To find a basis for the column space $\mathcal{C}(A)$ of the matrix A we need to find which of the columns of A are linearly independent. We can do this by identifying the columns which contain the leading ones in rref(A). The corresponding columns in the original matrix form a basis for the column space of A. Looking at rref(A) we see the first, third, and fourth columns of the matrix are linearly independent so the vectors $\{(1,2,3)^\mathsf{T},(3,7,9)^\mathsf{T},(3,6,10)^\mathsf{T}\}$ form a basis for $\mathcal{C}(A)$.

Now let's find a basis for the null space, $\mathcal{N}(A) \equiv \{\vec{x} \in \mathbb{R}^4 \mid A\vec{x} = \vec{0}\}$. The second column does not contain a pivot, therefore it corresponds to a *free variable*, which we will denote s. We are looking for a vector with three unknowns and one free variable $(x_1, s, x_3, x_4)^\mathsf{T}$ that obeys the conditions:

$$\begin{bmatrix} \mathbf{1} & 3 & 0 & 0 \\ 0 & 0 & \mathbf{1} & 0 \\ 0 & 0 & 0 & \mathbf{1} \end{bmatrix} \begin{bmatrix} x_1 \\ s \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \qquad \Rightarrow \qquad \begin{aligned} 1x_1 + 3s &= 0 \\ 1x_3 &= 0 \\ 1x_4 &= 0 \end{aligned}$$

Let's express the unknowns x_1, x_3, and x_4 in terms of the free variable s. We immediately see that $x_3 = 0$ and $x_4 = 0$, and we can write $x_1 = -3s$. Therefore, any vector of the form $(-3s, s, 0, 0)$, for any $s \in \mathbb{R}$, is in the null space of A. We write $\mathcal{N}(A) = \text{span}\{(-3,1,0,0)^\mathsf{T}\}$.

Observe that the $\dim(\mathcal{C}(A)) = \dim(\mathcal{R}(A)) = 3$, this is known as the *rank* of the matrix A. Also, $\dim(\mathcal{R}(A)) + \dim(\mathcal{N}(A)) = 3 + 1 = 4$, which is the dimension of the input space of the linear transformation T_A.

E. Invertible matrix theorem

There is an important distinction between matrices that are invertible and those that are not as formalized by the following theorem.

Theorem. *For an $n \times n$ matrix A, the following statements are equivalent:*

1) *A is invertible*
2) *The RREF of A is the $n \times n$ identity matrix*
3) *The rank of the matrix is n*
4) *The row space of A is $\mathbb{R}^n$*
5) *The column space of A is $\mathbb{R}^n$*
6) *A doesn't have a null space (only the zero vector $\mathcal{N}(A) = \{\vec{0}\}$)*
7) *The determinant of A is nonzero* $\det(A) \neq 0$

For a given matrix A, the above statements are either all true or all false.

An invertible matrix A corresponds to a linear transformation T_A which maps the n-dimensional input vector space $\mathbb{R}^n$ to the n-dimensional output vector space $\mathbb{R}^n$ such that there exists an inverse transformation T_A^{-1} that can faithfully undo the effects of T_A.

On the other hand, an $n \times n$ matrix B that is not invertible maps the input vector space $\mathbb{R}^n$ to a subspace $\mathcal{C}(B) \subsetneq \mathbb{R}^n$ and has a nonempty null space. Once T_B sends a vector $\vec{w} \in \mathcal{N}(B)$ to the zero vector, there is no T_B^{-1} that can undo this operation.

F. Determinants

The determinant of a matrix, denoted $\det(A)$ or $|A|$, is a special way to combine the entries of a matrix that serves to check if a matrix is invertible or not. The determinant formulas for 2×2 and 3×3 matrices are

$$\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}a_{22} - a_{12}a_{21}, \qquad \text{and}$$

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = a_{11}\begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - a_{12}\begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} + a_{13}\begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix}.$$

If the $|A| = 0$ then A is not invertible. If $|A| \neq 0$ then A is invertible.

G. Eigenvalues and eigenvectors

The set of eigenvectors of a matrix is a special set of input vectors for which the action of the matrix is described as a simple *scaling*. When a matrix is multiplied by one of its eigenvectors the output is the same eigenvector multiplied by a constant $A\vec{e}_\lambda = \lambda\vec{e}_\lambda$. The constant λ is called an *eigenvalue* of A.

To find the eigenvalues of a matrix we start from the eigenvalue equation $A\vec{e}_\lambda = \lambda\vec{e}_\lambda$, insert the identity $\mathbb{1}$, and rewrite it as a null-space problem:

$$A\vec{e}_\lambda = \lambda\mathbb{1}\vec{e}_\lambda \qquad \Rightarrow \qquad (A - \lambda\mathbb{1})\,\vec{e}_\lambda = \vec{0}.$$

This equation will have a solution whenever $|A - \lambda\mathbb{1}| = 0$. The eigenvalues of $A \in \mathbb{R}^{n\times n}$, denoted $\{\lambda_1, \lambda_2, \ldots, \lambda_n\}$, are the roots of the *characteristic polynomial* $p(\lambda) = |A - \lambda\mathbb{1}|$. The *eigenvectors* associated with the eigenvalue λ_i are the vectors in the null space of the matrix $(A - \lambda_i\mathbb{1})$.

Certain matrices can be written entirely in terms of their eigenvectors and their eigenvalues. Consider the matrix Λ that has the eigenvalues of the matrix A on the diagonal, and the matrix Q constructed from the eigenvectors of A as columns:

$$\Lambda = \begin{bmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & 0 \\ 0 & 0 & \lambda_n \end{bmatrix}, \quad Q = \begin{bmatrix} | & & | \\ \vec{e}_{\lambda_1} & \cdots & \vec{e}_{\lambda_n} \\ | & & | \end{bmatrix}, \quad \text{then} \quad A = Q\Lambda Q^{-1}.$$

Matrices that can be written this way are called diagonalizable.

The decomposition of a matrix into its eigenvalues and eigenvectors gives valuable insights into the properties of the matrix. Google's original `PageRank` algorithm for ranking webpages by "importance" can be formalized as an eigenvector calculation on the matrix of web hyperlinks.

VI. Textbook plug

If you're interested in learning more about linear algebra, you can check out my new book, the No bullshit guide to linear algebra.
A pre-release version of the book is available here: `gum.co/noBSLA`