Rethink Matrix

xiaohe

January 04, 2026

Contents

1 Matrix as Linear Map

The concept of matrix is closely related to linear maps between vector spaces – actually, they are two sides of the same coin. In this section, we will explore this relationship in detail.

Let's first think about a linear map $\mathcal{T}$ from a m-dimensional vector space V to an n-dimensional vector space W. We can choose a basis $\{v_1, v_2, ..., v_m\}$ for V and a basis $\{w_1, w_2, ..., w_n\}$ for W. For each basis vector v_j in V, the image under the linear map $\mathcal{T}$ can be expressed as a linear combination of the basis vectors in W:

$$\mathcal{T} v_j = \sum_{i=0}^{n} A_{i,j} w_i$$

Once all the m^2 coeffficients $A_{i,j}$ are given, we have the full information of the linear map $\mathcal{T}$ since any vector $v \in V$ can be expressed as a linear combination of the basis vectors thus its image $\mathcal{T} v$ can be computed accordingly. Now we randomly pick a vector $v \in V$:

$$v = \sum_{j=0}^{m} x_j v_j$$

where x_j are the **coordinates** of the vector v in the basis $\{v_1, v_2, ..., v_m\}$. Now we can compute the image of v under the linear map $\mathcal{T}$:

$$\mathcal{T} v = \mathcal{T}\left(\sum_{j=0}^{m} x_j v_j\right) = \sum_{j=0}^{m} x_j \mathcal{T} v_j = \sum_{j=0}^{m} x_j \left(\sum_{i=0}^{n} A_{i,j} w_i\right) = \sum_{i=0}^{n} \left(\sum_{j=0}^{m} A_{i,j} x_j\right) w_i$$

Wait... What the heck is "coordinate"? Let's take a step back. Rethink how we define coordinate of 2- and 3-dimensional vectors in high-school level math. We pick three perpendicular axes x, y and z in the 3D space. Then any vector v in this 3D space can be represented as a combination of these three axes:

$$v = v_x \hat{\imath} + v_y \hat{\jmath} + v_z \hat{k}$$

where $\hat{\imath}$, $\hat{\jmath}$ and $\hat{k}$ are unit vectors along the x, y and z axes respectively. The coefficients v_x, v_y and v_z are called the coordinates of the vector v in this coordinate system. Similarly, in the general vector space V, we pick a set of basis vectors $\{v_1, v_2, ..., v_m\}$ to define a coordinate system. Any vector $v \in V$ can be represented as a linear combination of these basis vectors, and the coefficients in this linear combination are called the coordinates of the vector v in this basis. Therefore, the coordinates x_j in the expression $v = \sum_{j=0}^{m} x_j v_j$ are the coordinates of the vector v in the basis $\{v_1, v_2, \ldots, v_m\}$.

Similarly, $\left(\sum_{j=0}^{m} A_{i,j} x_j\right)$ are the coordinates of the image vector $\mathcal{T}v$ in the basis $\{w_1, w_2, \ldots, w_n\}$ of the vector space W.

Now we can see the motivation behind the matrix-vector multiplication. Let's consider the $n \times m$ matrix $M(\mathcal{T})$ **corresponding** (we don't actually see why they are **corresponding** up to this point) to the linear map $\mathcal{T}$:

$$M(\mathcal{T}) = \begin{bmatrix} A_{1,1} & A_{1,2} & \cdots & A_{1,m} \\ A_{2,1} & A_{2,2} & \cdots & A_{2,m} \\ \vdots & \vdots & \ddots & \vdots \\ A_{n,1} & A_{n,2} & \cdots & A_{n,m} \end{bmatrix}.$$

When we multiply this matrix $M(\mathcal{T})$ with the coordinates of v in the basis $\{v_1, v_2, \ldots, v_m\}$, we get:

$$M(\mathcal{T}) \begin{bmatrix} x_1 \\ x_2 \\ \cdots \\ x_m \end{bmatrix} = \begin{bmatrix} \sum_{j=0}^{m} A_{1,j} x_j \\ \sum_{j=0}^{m} A_{2,j} x_j \\ \cdots \\ \sum_{j=0}^{m} A_{n,j} x_j \end{bmatrix}.$$

The left side is exactly the coordinates of the image vector $\mathcal{T}v$ in the basis $\{w_1, w_2, \ldots, w_n\}$ of the vector space W. Therefore, we can conclude that multiplying the matrix $M(\mathcal{T})$ with the coordinates of a vector v in the basis $\{v_1, v_2, \ldots, v_m\}$ gives us the coordinates of the image vector $\mathcal{T}v$ in the basis $\{w_1, w_2, \ldots, w_n\}$. So matrix-vector multiplication is actually a way to compute the image of a vector under a linear map when we know the matrix corresponding to that linear map and the coordinates of the vector in a chosen basis.

This insight gives us a powerful tool to think about linear equation systems, transformations in geometry, and many other applications in various fields such as physics, computer science, and economics. Take homogeneous linear equation systems as an example. A homogeneous linear equation system can be represented as:

$$Av = 0$$

where A is a $m \times n$ matrix and v is a n-dimensional vector. Once we view the matrix A as corresponding matrix to a linear map $\mathcal{A}$ from $\mathbb{R}^n$ to $\mathbb{R}^m$, under standard basis, this equation can be reinterpreted as finding the kernel (or null space) of the linear map $\mathcal{A}$. The system has non-zero solutions iff the kernel of $\mathcal{A}$ is non-trivial. And we know that linear maps from a higher-dimensional space to a lower-dimensional space (i.e., $n > m$) must have a non-trivial kernel. Thus, we can conclude

Theorem 1.1: A homogeneous linear equation system $Av = 0$ has non-zero solutions iff the number of equations is less than the number of unknowns (i.e., the number of rows of A is less than the number of columns of A).

2 Matrix as Linear Combination

In the previous section, we have seen how matrix-vector multiplication can be understood as applying a linear map to a vector when we know the matrix corresponding to that linear map and the coordinates of the vector in a chosen basis. Now, let's extend this idea to matrix-matrix multiplication.

First, let's consider the matrix-vector multiplication example Av from previous section, but in now we see the n-dimensional vector v as a $n \times 1$ matrix V:

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

Then the matrix-vector multiplication Av can be viewed as multiplying the $m \times n$ matrix A with the $n \times 1$ matrix v, resulting in a $m \times 1$ matrix. For each column of A, we may see it as a vector in $\mathbb{R}^m$. Thus, the multiplication Av can be interpreted as taking a linear combination of the columns of A, weighted by the entries of the matrix v. Specifically, if we denote the columns of A as $A_{,1}, A_{,2}, \ldots, A_{,n}$, then we have: $Av = x_1 A_{,1} + x_2 A_{,2} + \ldots + x_n A_{,n}$.

The same result holds when we multiply two general matrices. Let's consider two matrices A and B, where A is an $m \times n$ matrix and B is an $n \times p$ matrix. The matrix-matrix multiplication AB results in an $m \times p$ matrix. Similar to the previous case, we can interpret this multiplication as taking linear combinations of the columns of A, weighted by the entries of the corresponding columns of B. Specifically, if we denote the columns of A as $A_{,1}, A_{,2}, \ldots, A_{,n}$ and the columns of B as $B_{,1}, B_{,2}, \ldots, B_{,p}$, then the resulting matrix $C = AB$ can be expressed as:

$$C_{,j} = b_{1,j} A_{,1} + b_{2,j} A_{,2} + \ldots + b_{n,j} A_{,n}$$

for each column $C_{,j}$ of the resulting matrix C, where $b_{i,j}$ are the entries of the matrix B.

This gives us a more clear way to understand the **rank** of a matrix and so called **column-row factorization**. Let's define the rank of a matrix first.

Definition 2.1: The **column rank** of a matrix is defined as the maximum number of linearly independent column vectors in the matrix. In other words, it is the dimension of the column space of the matrix. Similarly, the **row rank** of a matrix is defined as the maximum number of linearly independent row vectors in the matrix. It is the dimension of the row space of the matrix.

Now we can give the theorem of column-row factorization.

Theorem 2.1: For any $m \times n$ matrix A with column rank r, there exist an $m \times r$ matrix C and a $r \times n$ matrix R such that $A = CR$.

Proof: The proof is straightforward. The dimension of the column space of A is r, so we can find a basis of r column vectors $\{c_1, c_2, ..., c_r\}$ that span the column space of A. We can form the matrix C by taking these basis vectors as its columns:

$$C = [c_1 \ \ c_2 \ \ \dots \ \ c_r].$$

Next, since each column of A can be expressed as a linear combination of the basis vectors, we can find the coefficients for these linear combinations and form the matrix R such that:

$$A = CR.$$

□

We can use the column-row factorization to prove a basic property of matrix rank.

Theorem 2.2: The column rank and row rank of any matrix are equal.

Proof: Let's consider an $m \times n$ matrix A with column rank r. According to the column-row factorization theorem, we can express A as:

$$A = CR$$

where C is an $m \times r$ matrix and R is a $r \times n$ matrix. The rows of A are linear coombinations of the rows of R. Since R has at most r linearly independent rows, the row rank of A cannot exceed r. So we have the row rank of A less than or equal to its column rank. This holds for any matrix, including A^T. And we know that

$$\begin{aligned}\text{row rank } A^T &= \text{column rank } A \\ &\leq \text{column rank } A^T \\ &= \text{row rank } A.\end{aligned}$$

Therefore, the column rank and row rank of any matrix are equal. □

So we don't need to distinguish column rank and row rank anymore. We can simply refer to them as the **rank** of a matrix. Now let's get back to the homogeneous linear equation system $Av = 0$. We have shown that this system has non-zero solutions iff the number of equations is less than the number of unknowns. We can restate this result in terms of matrix rank.

Theorem 2.3: A homogeneous linear equation system $Av = 0$ has non-zero solutions iff the rank of the matrix A is less than the number of columns of A.

Proof: The rank of matrix A is less than the number of columns of A, thus the column vectors of A are linearly dependent. Therefore, there exists $\{v_1, v_2, ..., v_n\}$, not all zero, such that $v_1 A_{,1} + v_2 A_{,2} + ... + v_n A_{,n} = 0$. Let $v = [v_1, v_2, ..., v_n]^T$. Then we have $Av = 0$.

Conversely, if there exists a non-zero vector v such that $Av = 0$, then the column vectors of A are linearly dependent, which implies that the rank of A is less than the number of columns of A. □