

Rethink Matrix

xiaohe

January 03, 2026

Contents

1 Matrix as Linear Map: Motivation behind Matrix-Vector Multiplication	1
2 Matrix as Linear Combination: Motivation behind Matrix-Matrix Multiplication	3

1 Matrix as Linear Map: Motivation behind Matrix-Vector Multiplication

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 $\{v_1, v_2, \dots, v_m\}$ for V and a basis $\{w_1, w_2, \dots, w_n\}$ 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 :

$$\mathcal{T}v_j = \sum_{i=1}^n A_{i,j} w_i$$

Once all the m^2 coefficients $A_{i,j}$ 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 $v \in V$:

$$v = \sum_{j=1}^m x_j v_j$$

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

$$\mathcal{T}v = \mathcal{T}\left(\sum_{j=1}^m x_j v_j\right) = \sum_{j=1}^m x_j \mathcal{T}v_j = \sum_{j=1}^m x_j \left(\sum_{i=1}^n A_{i,j} w_i\right) = \sum_{i=1}^n \left(\sum_{j=1}^m A_{i,j} x_j\right) w_i$$

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:

$$v = v_x \hat{i} + v_y \hat{j} + v_z \hat{k}$$

where \hat{i} , \hat{j} 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 $\{v_1, v_2, \dots, v_m\}$ 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 $\{v_1, v_2, \dots, v_m\}$.

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 $\{w_1, w_2, \dots, w_n\}$ 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} :

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

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

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

The left side is exactly the coordinates of the image vector $\mathcal{T}v$ in the basis $\{w_1, w_2, \dots, w_n\}$ 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 $\{v_1, v_2, \dots, v_m\}$ gives us the coordinates of the image vector $\mathcal{T}v$ in the basis $\{w_1, w_2, \dots, w_n\}$. 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: Motivation behind Matrix-Matrix Multiplication

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 :

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

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 .