A mystery of transpose of matrix multiplication
Why $(AB)^T = B^T A^T$?

Yamauchi, Hitoshi
Nichiyou kenkyu aikouka
Berlin, Germany

2009-5-24(Sun)

Contents

Abstract

Why does the transpose of matrix multiplication is $(AB)^T = B^T A^T$? I was told this is the rule when I was a university student. But, there must be something we could understand this. First, I would like to show you that the relationship of dot product of vectors and its transpose is $\left[\boldsymbol{u}^T\boldsymbol{v}\right]^T = \boldsymbol{v}\boldsymbol{u}^T$. Then I will point out to you that a matrix is a representation of transform rather a representation of simultaneous equations. This point of view gives us that a matrix multiplication includes dot products. Combining these two point of views and one more, a vector is a special case of a matrix, we could understand the first equation.

1 Introduction

This article talks about the transpose of matrix multiplication

$$(AB)^T = B^T A^T. \tag{1}$$

When I saw this relationship, I wonder why this happens. For me, transpose is an operator, it looks like this is a special distribution law. If multiply something (for instance, -1),

$$(-1) \cdot (a+b) \quad = \quad (-1 \cdot a) + (-1 \cdot b).$$

There is no place exchange of a and b, it does not become b a. But in the case of transpose,

$$(AB)^T = \underbrace{B^T A^T}_{\text{The order } AB \text{ becomes } BA}. \tag{2}$$

Why this happens? If I compare the each of element, I could see this should be, however, I feel something magical and could not feel to understand it. Recently, I read a book and felt a bit better. So I would like to introduce the explanation.

Before we start to talk about the matrix multiplication, I would like to start from dot product of vectors since vector is a special case of matrix. Then we will generalize back this idea to matrix multiplication. Because a simpler form is usually easier to understand, we will start a simple one and then go further.

2 Dot product of vectors

Let's think about two vectors $\boldsymbol{u}, \boldsymbol{v}$. Most of the linear algebra textbooks omit the elements of vector since it is too cumbersome, however, I will put the elements here. When we wrote elements like

$$\boldsymbol{u} = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix} \tag{3}$$

as a general vector, this is also cumbersome. So, I will start with three dimensional vector. Then we could extend this to general dimensional vectors. Here the main actor of this story is transpose ^{T}, this is an operator to exchange the row and column of a matrix or a vector.

$$
\begin{aligned}
\boldsymbol{u}^T &= \begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix}^T \\
&= \begin{pmatrix} u_1 & u_2 & u_3 \end{pmatrix} \qquad (4)
\end{aligned}
$$

I assume you know what is a dot product of vectors (dot product is also called as an inner product). The dot product $(\boldsymbol{uv})$ is

$$
\begin{aligned}
(\boldsymbol{uv}) &= \boldsymbol{u}^T \boldsymbol{v} \\
&= \begin{pmatrix} u_1 & u_2 & u_3 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} \\
&= u_1 v_1 + u_2 v_2 + u_3 v_3 \qquad (5)
\end{aligned}
$$

Let's think about one more step. Why I want to go one more step? Because this article is for explaining what is transpose. Therefore I would like to go one more step. Then some of you want to know why transpose matters for me. This is because I matter it. There is no way to prove this is interesting subject in a mathematical way. I just feel it is intriguing. There are many good mathematics book which has deep insight of mathematics. I have nothing to add to these books regarding mathematical insight. However, the emotion of my interest is mine and this is I can explain. Many of mathematical book tries to extract the beauty of mathematics in the mathematical way. I totally agree this is the way to see the beauty of mathematics. But this beauty is a queen beauty who has high natural pride. It is almost impossible to be near by for me. I would like to understand this beauty in more familiar way. Therefore, I am writing this with my feelings to the queen. Of course this method has a dangerous aspect since I sometimes can see only one view of the beauty, sometimes I lost other view points.

Now, Let's think about one more step and think about transpose of this.

$$
(u_1 v_1 + u_2 v_2 + u_3 v_3)^T = ? \qquad (6)
$$

A dot product produces a scalar value, therefore transpose does not change the result. Therefore, it should be

$$
\begin{aligned}
(\boldsymbol{uv})^T &= (u_1 v_1 + u_2 v_2 + u_3 v_3)^T \\
&= (u_1 v_1 + u_2 v_2 + u_3 v_3) \\
&= (\boldsymbol{uv}). \qquad (7)
\end{aligned}
$$

We could think about the transpose operation further using this clue – transpose does not affect the result of vector dot product.

If the distribution law is kept also for transpose,

$$
\begin{aligned}
\left[\boldsymbol{u}^T \boldsymbol{v}\right]^T &= \left[\boldsymbol{u}^T\right]^T [\boldsymbol{v}]^T \\
&= \boldsymbol{u}\boldsymbol{v}^T. \qquad (8)
\end{aligned}
$$

(Where we use transpose of transpose produces original form $(\boldsymbol{u}^{T^T} = \boldsymbol{u})$.) However,

$$
\begin{aligned}
(\boldsymbol{uv})^T &= \left[\boldsymbol{u}^T \boldsymbol{v}\right]^T \\
&= \left[\begin{pmatrix} u_1 & u_2 & u_3 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix}\right]^T \\
&= \left[\begin{pmatrix} u_1 & u_2 & u_3 \end{pmatrix}^T \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix}^T\right] \\
&= \begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix} \begin{pmatrix} v_1 & v_2 & v_3 \end{pmatrix} \\
&= \begin{pmatrix} u_1 v_1 & u_1 v_2 & u_1 v_3 \\ u_2 v_1 & u_2 v_2 & u_2 v_3 \\ u_3 v_1 & u_3 v_2 & u_3 v_3 \end{pmatrix} \\
&\neq (\boldsymbol{uv}) \qquad (9)
\end{aligned}
$$

This is even not a scalar, so something is wrong. (Some of the reader wonder why a vector multiplication becomes a matrix. Equation 9 is a Tensor product. I could not explain this in this article, but, you could look up this with keyword “Tensor product.”) From Equation 9,

$$
\left[\boldsymbol{u}^T \boldsymbol{v}\right]^T \neq \boldsymbol{u}^{T^T} \boldsymbol{v}^T. \qquad (10)
$$

To make transposed and original dot product the same,

$$
(\boldsymbol{uv})^T = \begin{pmatrix} v_1 & v_2 & v_3 \end{pmatrix} \begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix} \qquad (11)
$$

is necessary. This means we need to exchange the vector position of $\boldsymbol{u}$ and $\boldsymbol{v}$. This means

$$
\begin{aligned}
(\boldsymbol{uv})^T &= \left[\boldsymbol{u}^T \boldsymbol{v}\right]^T \\
&\rightarrow \begin{pmatrix} v_1 & v_2 & v_3 \end{pmatrix} \begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix} \\
&= \boldsymbol{v}^T \boldsymbol{u} \\
&= (\boldsymbol{v}^{T^T} \boldsymbol{u}) \\
&= (\boldsymbol{vu}) \qquad (12)
\end{aligned}
$$

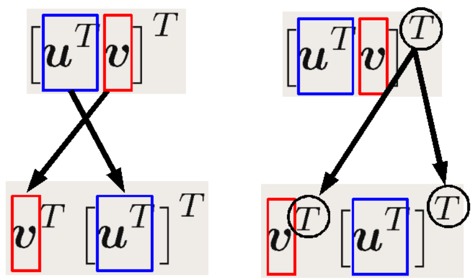

Figure 1: Anatomy of dot product vector transpose

The $\rightarrow$ in Equation 12 is required. There is no objective reason, I think this is correct. At the end,

$$\begin{aligned} \left[\boldsymbol{u}^T \boldsymbol{v}\right]^T &= \boldsymbol{v}^T \boldsymbol{u} \\ &= \boldsymbol{v}^T \left[\boldsymbol{u}^T\right]^T. \end{aligned} \tag{13}$$

Figure 1 shows that: 1. exchange the vector position, 2. distribute the transpose operator.

You might not like this since here is only a sufficient condition. But from the one transpose for each element and two vectors combination, we could find only this combination is the possibility. So, I believe this is sufficient explanation.

Now the remaining problem is what is the relationship between matrix multiplication and vector dot products. You might see the rest of the story.

This explanation is based on Farin and Hansford's book [1]. I recommend this book since this book explains these basic ideas quite well. Some knows the Farin's book is difficult. I also had such an impression and hesitated to look into this book. But, this book is easy to read. I enjoy the book a lot.

3 Matrix Multiplication

My calculum of mathematics introduced matrix as a representation of simultaneous equation.

$$\begin{aligned} y_1 &= a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \\ y_2 &= a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n \\ \vdots & \;\vdots\;\; \vdots \\ y_n &= a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n \end{aligned} \tag{14}$$

This is a natural introduction, but when I thought matrix multiplication, I needed a jump to extends the idea. I wonder, how many students realize a matrix multiplication includes dot products? I realized it quite later. Here, y_1 is a dot product of $a_{[1\ldots n]1}$ and $x_{[1\ldots n]}$.

$$\begin{aligned} y_1 &= a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \\ &= \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \end{aligned} \tag{15}$$

Another my favorite book [2] shows that matrix has an aspect of a representation of transformation. According to this idea, matrix is composed of coordinate vectors. By any ideas or aspects, these are all property of matrix. The formal operations of matrix are all the same in any way to see matrix.

One powerful mathematical idea is "when the forms are the same, they are the same." We could capture the same aspect of different things. If one aspect is the same, then we could expect these different two things have a common behavior. For example, every person has personality, there is no exact the same person. But, some of them share their hobby, they could have similar property. People who share the hobby might buy the same things. To extends this idea further, in totally different country, totally different generation people could buy the same thing because of the sharing aspects. "Found a similar aspect in the different things" is the one basic mathematical thinking. Here, we could find the same form in coordinate transformation and simultaneous equations.

When we see a matrix as a coordinate system, such matrix is composed of coordinate axis vectors. Here, we think three dimensional coordinates only.

$$\begin{aligned} \boldsymbol{a}_1 &= \begin{pmatrix} a_{11} \\ a_{21} \\ a_{31} \end{pmatrix} \\ \boldsymbol{a}_2 &= \begin{pmatrix} a_{12} \\ a_{22} \\ a_{32} \end{pmatrix} \\ \boldsymbol{a}_3 &= \begin{pmatrix} a_{13} \\ a_{23} \\ a_{33} \end{pmatrix} \end{aligned} \tag{16}$$

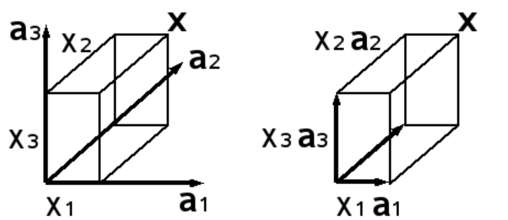

Figure 2: A three dimensional coordinate system

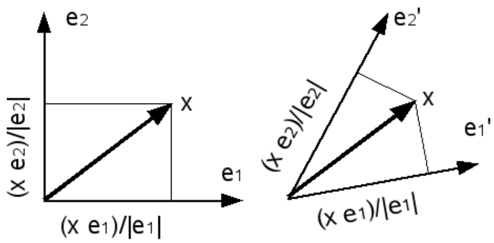

Figure 3: Coordinate system and projection. In any coordinate system, the coordinate value itself is given by projection = dot products.

We could write a matrix A as

$$\begin{aligned} A &= \begin{pmatrix} \boldsymbol{a}_1 & \boldsymbol{a}_2 & \boldsymbol{a}_3 \end{pmatrix} \\ &= \begin{pmatrix} a_{11} & a_{21} & a_{31} \\ a_{12} & a_{22} & a_{32} \\ a_{13} & a_{23} & a_{33} \end{pmatrix}. \end{aligned} \tag{17}$$

Figure 2 shows an example coordinate system in this construction. A coordinate of a point $\boldsymbol{x}$ is the following:

$$\begin{aligned} \boldsymbol{x} &= \boldsymbol{a}_1 x_1 + \boldsymbol{a}_2 x_2 + \boldsymbol{a}_3 x_3 \\ &= \begin{pmatrix} a_{11} \\ a_{12} \\ a_{13} \end{pmatrix} x_1 + \begin{pmatrix} a_{21} \\ a_{22} \\ a_{23} \end{pmatrix} x_2 + \\ & \quad \begin{pmatrix} a_{31} \\ a_{32} \\ a_{33} \end{pmatrix} x_3 \end{aligned} \tag{18}$$

Where x_1, x_2, x_3 are projected length of $\boldsymbol{x}$ onto $\boldsymbol{a}_1, \boldsymbol{a}_2, \boldsymbol{a}_3$ axis, respectively. Now we said "projected." This is dot product. Let's see in two dimensional case in Figure 3 since I think a three dimensional case is still a bit cumbersome. You can see each coordinate value is dot product to each coodinate axis. Equation 18 can be rewritten to express dot product form explicitly:

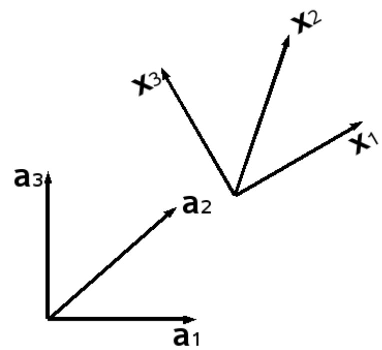

Figure 4: Two coordinate systems

$$\begin{aligned} \boldsymbol{x} &= \boldsymbol{a}_1 x_1 + \boldsymbol{a}_2 x_2 + \boldsymbol{a}_3 x_3 \\ &= \begin{pmatrix} \boldsymbol{a}_1 & \boldsymbol{a}_2 & \boldsymbol{a}_3 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \end{aligned} \tag{19}$$

$\boldsymbol{x}$ is only one element of coordinate system axis. In three dimensional case, there are three axes, $\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3$. Each coordinate axis is projected to the other coordinate axes to transform the coordinate system. Figure 4 shows two coordinate systems $\boldsymbol{a}_1, \boldsymbol{a}_2, \boldsymbol{a}_3$ and $\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3$.

$$\begin{aligned} & AX \\ &= \begin{pmatrix} \boldsymbol{a}_1 & \boldsymbol{a}_2 & \boldsymbol{a}_3 \end{pmatrix} \begin{pmatrix} \boldsymbol{x}_1 & \boldsymbol{x}_2 & \boldsymbol{x}_3 \end{pmatrix} \\ &= \begin{pmatrix} a_{11} & a_{21} & a_{31} \\ a_{12} & a_{22} & a_{32} \\ a_{13} & a_{23} & a_{33} \end{pmatrix} \begin{pmatrix} x_{11} & x_{21} & x_{31} \\ x_{12} & x_{22} & x_{32} \\ x_{13} & x_{23} & x_{33} \end{pmatrix} \end{aligned} \tag{20}$$

I hope you now see the dot product inside the matrix.

4 Transpose of Matrix Multiplication

A transpose of matrix multiplication has vector dot products, therefore, it should be $(AB)^T = B^T A^T$.

This article's discussion is based on the difference between dot product and tensor product, however, I was suggested the standard proof is still simple and understandable, so, I will show the proof.

$$
\begin{aligned}
(B^T A^T)_{ij} &= (b^T)_{ik}(a^T)_{kj} \\
&= \sum_k b_{ki} a_{jk} \\
&= \sum_k a_{jk} b_{ki} \qquad (21) \\
&= (AB)_{ji} \\
&= (AB)^T
\end{aligned}
$$

I talked about the reason of Equation 21 in this article. The proof is indeed simple and beautiful. It is just too simple for me and could not think about the behind when I was a student. I enjoyed to find the behind of Equation 21. I hope you can also enjoy the behind of this simple proof.

Acknowledgments

Thanks to C. Rössl for explaining me about coordinate transformation. Thanks to L. Gruenschloss to make a suggestion to add the last proof.

References

[1] Gerald Farin Dianne Hansford. *Practical Linear Algebra; A Geometry Toolbox*. A K Peters, Ltd., 2005.

[2] Koukichi Sugihara. *Mathematical Theory of Graphics (Gurafikusu no Suuri)*. Kyouritu Shuppan, 1995.