

矩阵乘法定义的来源

理试A2410周昭宇

在人工智能愈发强大的今天，人们越来越离不开AI。而矩阵乘法作为各大模型算法的根基，有着不可磨灭的作用。然而，矩阵乘法究竟是从何而来，是什么原因导致了矩阵乘法的产生以及矩阵乘法的运算定义的来由却少有人提及。因此，我想趁此机会，多了解一些关于矩阵乘法的来历，以及在矩阵乘法之后的故事。

在我们的教材中提到的关于矩阵乘法的定义似乎非常的理所当然，其举了一个旋转的例子。并且在其中提到“从旋转这个例子收到启发，我们给出矩阵的乘法运算的定义。”一切都似乎顺理成章，但是在学习的时候就让我非常的迷惑，这并不是对于矩阵乘法的合适解释。那么，矩阵乘法的定义究竟是从何而来呢？

根据所能查询到的资料，我了解到，矩阵的概念最初并没有被广泛应用，而是主要出现在解线性方程组的背景下。线性方程组的系数可以表示为矩阵，求解这些方程组常常需要一些计算手段。然而，在19世纪之前，尽管数学家们已经在某些情况下使用了矩阵表示，但并没有正式的定义和运算规则。

在19世纪，随着线性代数的发展，数学家们开始更加关注如何表示和组合线性变换。一个重要问题是，如何将多个线性变换（比如缩放、旋转、平移等）组合成一个更复杂的变换。在几何和物理学中，这类问题非常常见。这也许是我们的教材举了旋转的例子的来源吧。

英国数学家阿瑟·凯莱特别在矩阵理论方面做出了重要贡献，他在1858年发表了关于矩阵的重要论文(A Memoir on the Theory of Matrices)，首次系统地定义了矩阵的乘法概念。而其设计矩阵乘法规则的动机是解决如何表示和组合线性变换的问题。凯莱定义的矩阵乘法的规则使得我们可以通过“矩阵乘法”来表示线性变换的复合，即线性复合映射。这里提一个我看到的还不错的例子。有两个映射f和g：

$$f \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} ax + by \\ cx + dy \end{pmatrix}$$
$$g \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} px + qy \\ rx + sy \end{pmatrix}$$

而将f与g复合，就可以得到

$$\begin{aligned}
h \begin{pmatrix} x \\ y \end{pmatrix} &= f \left(g \begin{pmatrix} x \\ y \end{pmatrix} \right) = f \begin{pmatrix} px + qy \\ rx + sy \end{pmatrix} \\
&= \begin{pmatrix} a(px + qy) + b(rx + sy) \\ c(px + qy) + d(rx + sy) \end{pmatrix} \\
&= \begin{pmatrix} (ap + br)x + (aq + bs)y \\ (cp + dr)x + (cq + ds)y \end{pmatrix}
\end{aligned}$$

而用矩阵来表示线性代数的映射来表示线性映射的系数，将线性映射f、g和h分别表示成：

$$F = \begin{pmatrix} a & b \\ c & d \end{pmatrix}, G = \begin{pmatrix} p & q \\ r & s \end{pmatrix}, H = \begin{pmatrix} ap + br & aq + bs \\ cp + dr & cq + ds \end{pmatrix}$$

因此，矩阵乘法的运算规则就了然了。即矩阵乘法的最核心意义在于它表示了线性变换的复合，当你将两个矩阵相乘的时候，实际上是将两个线性变换合并成一个新的线性变换。

在查阅了这么多资料之后，我感觉到数学是一个一脉相承的学科。并且它其实并不简单，我其实并没有很看懂关于矩阵乘法背后的各种关联。网上所能查到的资料更多的在于强调凯莱的直觉“最后历史证明凯莱异于常人的洞察力为矩阵理论与线性代数的发展开启了一扇大门”。但是总而言之，各个理论在被发现被定义之前，都应该有他的动机。正如原始人并不会去研究什么矩阵什么代数一样，数学这门学科在不断的前进，出现了新的问题，所以需要更先进更好的数学工具来解决问题。如今人工智能蓬勃发展，矩阵乘法作为人工智能的基石，给与了人工智能“智能”，也给人工智能的发展造成了很大的制约。人类需要更优化的算法训练模型，需要更先进的理论去解释模型。我们正处于进步的时代。