

作者	生姜 DrGinger
脚本	生姜 DrGinger
视频	崔崔 CuiCui
开源学习资源	https://github.com/Visualize-ML
平台	https://www.youtube.com/@DrGinger_Jiang https://space.bilibili.com/3546865719052873 https://space.bilibili.com/513194466

2.6 矩阵乘法性质



本节你将掌握的核心技能：

- ▶ 矩阵连乘：多个几何操作按顺序施加到向量上，从右向左生效。
- ▶ 连乘转置：反序，从行向量视角理解变换。
- ▶ 结合律：改变括号位置不影响结果，可优化计算顺序减少运算量。
- ▶ 矩阵分解：将复杂矩阵拆解成基本几何变换，更直观理解其作用。
- ▶ 交换律通常不成立：即使乘积存在， AB 、 BA 也可能有不同几何含义。
- ▶ 特殊情况，交换律成立。
- ▶ 矩阵幂：方阵反复相乘表示相同线性变换多次作用。

有了矩阵乘法几何视角的铺垫，理解矩阵乘法常见性质就变得十分容易了。

矩阵连乘：连续几何变换

若干矩阵顺序相乘，相当于这些几何变换依次作用于几何体上。

以如下三个 2×2 矩阵为例

$$\mathbf{A} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \quad \mathbf{C} = \begin{bmatrix} 1/2 & 0 \\ 0 & 2 \end{bmatrix} \quad (1)$$

矩阵 \mathbf{A} 对应绕原点旋转 (rotate)， \mathbf{B} 对应剪切 (shear)， \mathbf{C} 对应缩放 (scale)。

具体来说，矩阵 \mathbf{A} 对应绕原点逆时针旋转 90 度。

矩阵 \mathbf{B} 对应沿 x_1 横轴方向剪切；简单来说，剪切将形状沿某个方向倾斜的变换，使得原本垂直或水平的线段变成倾斜状态，同时保持平行关系不变。

矩阵 \mathbf{C} 让横轴缩小为 1/2，纵轴放大至 2 倍。

如果列向量 \mathbf{x} 、 \mathbf{y} 均代表平面上一点，如下矩阵乘法

本 PDF 文件为作者草稿，发布目的为方便读者在移动终端学习，终稿内容以清华大学出版社纸质出版物为准。

版权归清华大学出版社所有，请勿商用，引用请注明出处。

代码及 PDF 文件下载：<https://github.com/Visualize-ML>

本书配套微视频均发布在 B 站——生姜 DrGinger：<https://space.bilibili.com/513194466>

欢迎大家批评指教，本书专属邮箱：jiang.visualize.ml@gmail.com

$$A @ B @ C @ x = y$$

(2)

代表对列向量 x 先后进行缩放 (C)、剪切 (B)、旋转 (A)。

整个几何变换过程如图 1 所示。

⚠ 请大家格外注意矩阵乘法 $ABCx = y$ 先后顺序，从右向左，即 $C \rightarrow B \rightarrow A$ 。

❓ 请大家手算 ABC 、 CBA 、 ACB 、 BAC 等等各种排列组合的矩阵乘法结果。

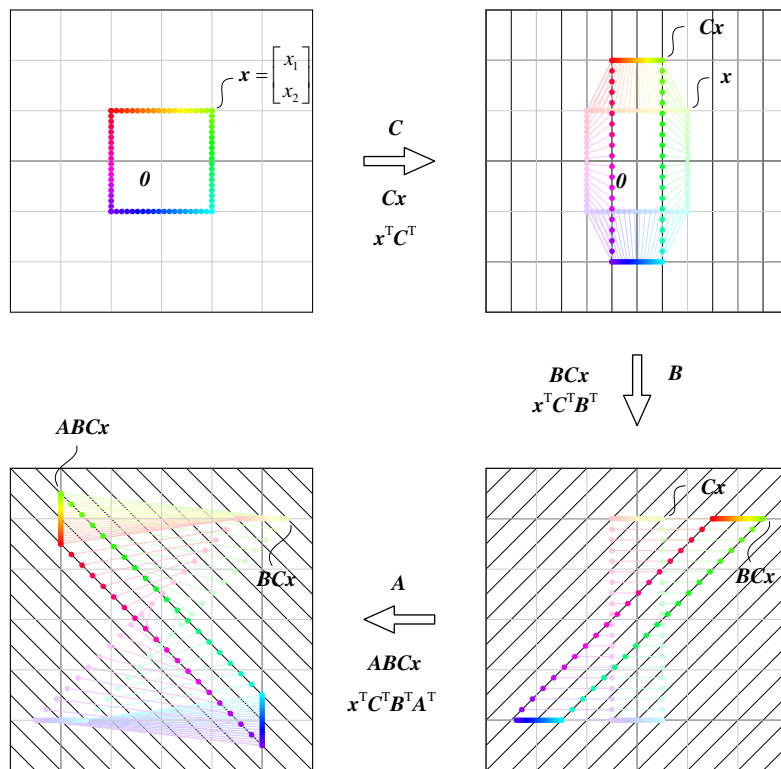


图 1. $ABCx = y$ 对应的分步几何操作

当然，我们也可以先计算 (ABC) ，然后再把 (ABC) 作为一个整体施加到 x 上。

矩阵乘法 (ABC) 相当于复合几何操作，“一步到位”完成几何变换！具体如图 2 所示。

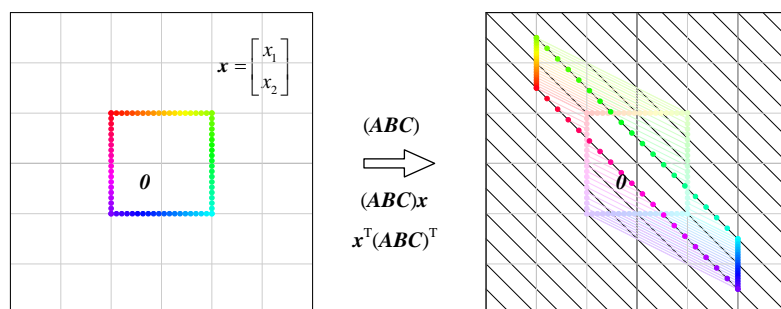


图 2. $(ABC)x = y$ 对应的“一步到位”几何操作

矩阵连乘的转置

矩阵连乘的转置如下几个重要性质值得大家重视：

$$\begin{aligned}(AB)^T &= B^T A^T \\ (ABC)^T &= C^T B^T A^T \\ (A_1 A_2 A_3 \cdots A_k)^T &= A_k^T \cdots A_3^T A_2^T A_1^T\end{aligned}\tag{3}$$

这组性质实际上并不需要“死记硬背”！

以 (2) 为例，对这个矩阵连乘转置

$$x^T @ C^T @ B^T @ A^T = y^T\tag{4}$$

列向量 $x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ 转置后得到行向量 $x^T = [x_1 \ x_2]$ ，代表水平面上一点。上式乘法告诉我们，对于行向量，依次施加缩放 (C)、剪切 (B)、旋转 (A) 几何变换，结果还是如图 1。

结合律

在矩阵运算中，如果有三个矩阵 A 、 B 、 C 参与相乘，先将 A 和 B 相乘，再与 C 相乘，或者先将 B 和 C 相乘，再左乘 A ，最终的计算结果是相同的

$$(AB)C = A(BC)\tag{5}$$

换句话说，无论我们先计算前两个矩阵的乘积，还是先计算后两个矩阵的乘积，结果都不会受到影响。这一性质可以简化矩阵运算，使我们可以灵活地选择计算顺序，以提高计算效率或便于推导数学公式。本节最后会讲到矩阵连乘中，如何通过结合律提高运算效率。

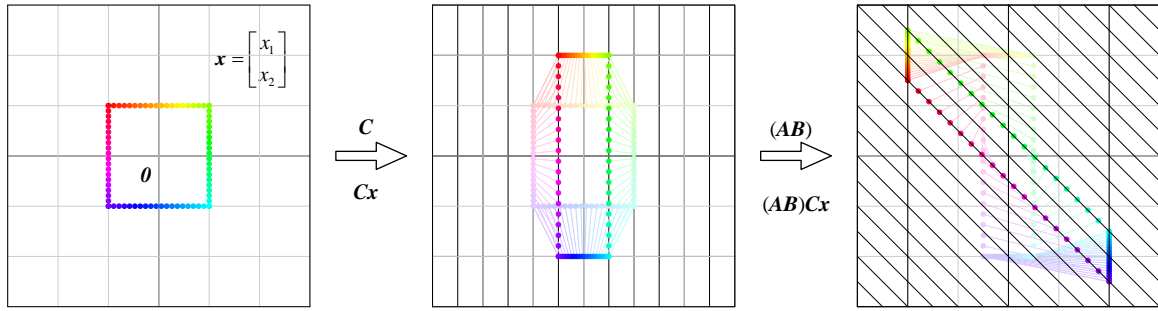
⚠ 注意，矩阵先后次序不能变。也就是说，矩阵乘法通常不满足交换律，即 $AB \neq BA$ 。

怎么理解 (5) 呢？

还是利用几何视角， (AB) 、 (BC) 相当于“局部”复合几何变换。

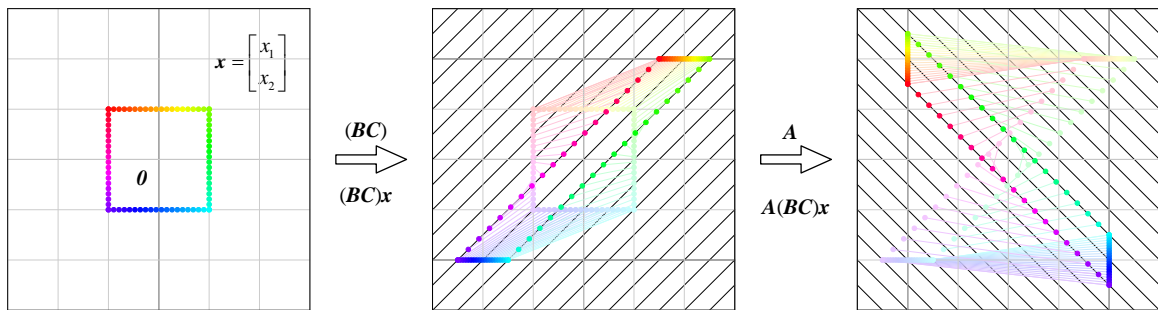
还是以 (1) 为例， (AB) 相当于复合旋转、剪切(剪切再先)； (BC) 相当于复合剪切、缩放(缩放再先)。

对于平面列向量 x ， $(AB)Cx$ 相当于对 x 先缩放 (C)，然后再 (AB) ，具体如图 3 所示。

图 3. $(AB)Cx = y$ 对应的分步几何操作

对于平面列向量 x , $A(BC)x$ 相当于对 x 先 (BC) , 然后再旋转 (A) , 具体如图 4 所示。

对比图 1、图 2、图 3、图 4, 虽然过程中有展开, 有合并, 我们发现它们的结果是完全一致的。

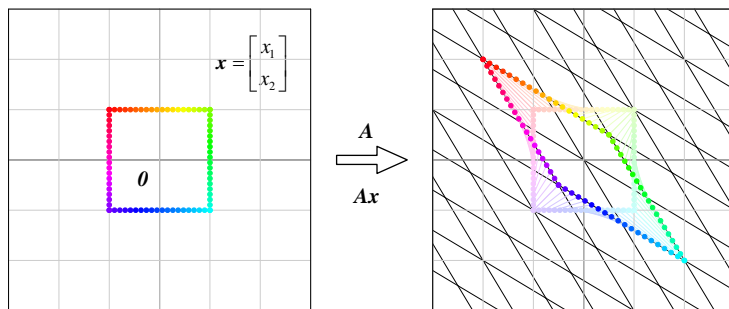
图 4. $A(BC)x = y$ 对应的分步几何操作

矩阵分解

给定矩阵 A

$$A = \begin{bmatrix} 5/4 & -3/4 \\ -3/4 & 5/4 \end{bmatrix} \quad (6)$$

如图 5 所示, $Ax = y$ 显然不是我们不熟悉的几何变换。

图 5. $Ax = y$ 对应的几何操作

要想理解 $Ax = y$ ，需要通过矩阵分解 (matrix decomposition) 把图 5 拆解成我们熟悉的几何变换。

简单来说，矩阵分解将一个矩阵分解成几个矩阵的连乘。可以这样理解，矩阵分解是矩阵连乘的逆操作。

几何角度来看，一个“复杂”几何操作可以分解为若干“我们熟悉的”几何操作。

比如，如图 6 所示，我们可以把矩阵 A 拆解成“旋转 \rightarrow 缩放 \rightarrow 旋转”，对应的矩阵乘法为

$$\begin{bmatrix} 5/4 & -3/4 \\ -3/4 & 5/4 \end{bmatrix} = \begin{bmatrix} \sqrt{2}/2 & \sqrt{2}/2 \\ -\sqrt{2}/2 & \sqrt{2}/2 \end{bmatrix} @ \begin{bmatrix} 2 & 0 \\ 0 & 1/2 \end{bmatrix} @ \begin{bmatrix} \sqrt{2}/2 & -\sqrt{2}/2 \\ \sqrt{2}/2 & \sqrt{2}/2 \end{bmatrix} \quad (7)$$

需要大家注意的是，图 6 中两个旋转方向正好相反。

(7) 对应的分解叫特征值分解 (Eigen Value Decomposition, EVD)；确切地说，由于矩阵 A 为对称矩阵，这个分解为谱分解 (spectral decomposition)。

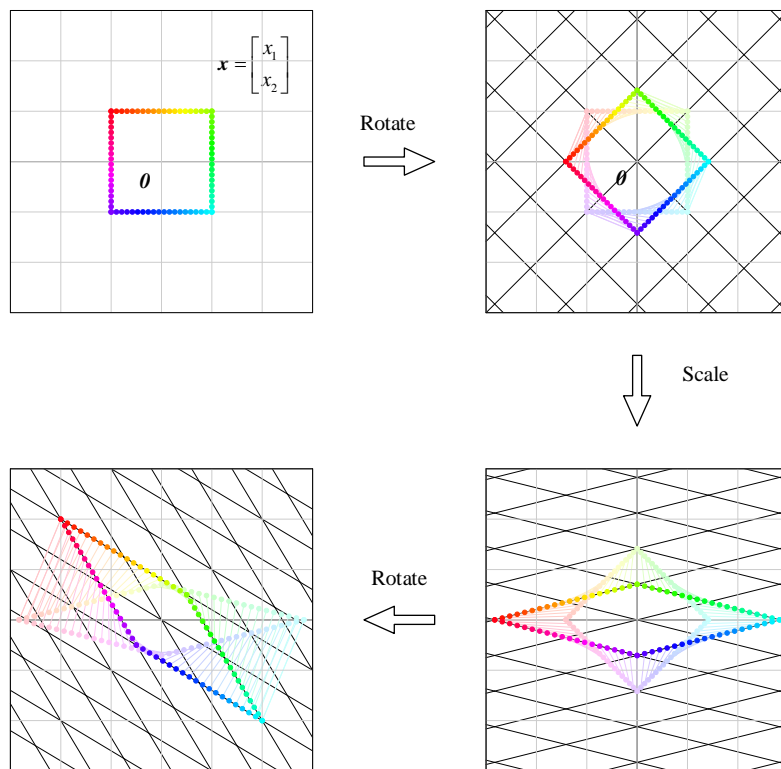


图 6. 把 A 分解成“旋转 \rightarrow 缩放 \rightarrow 旋转”

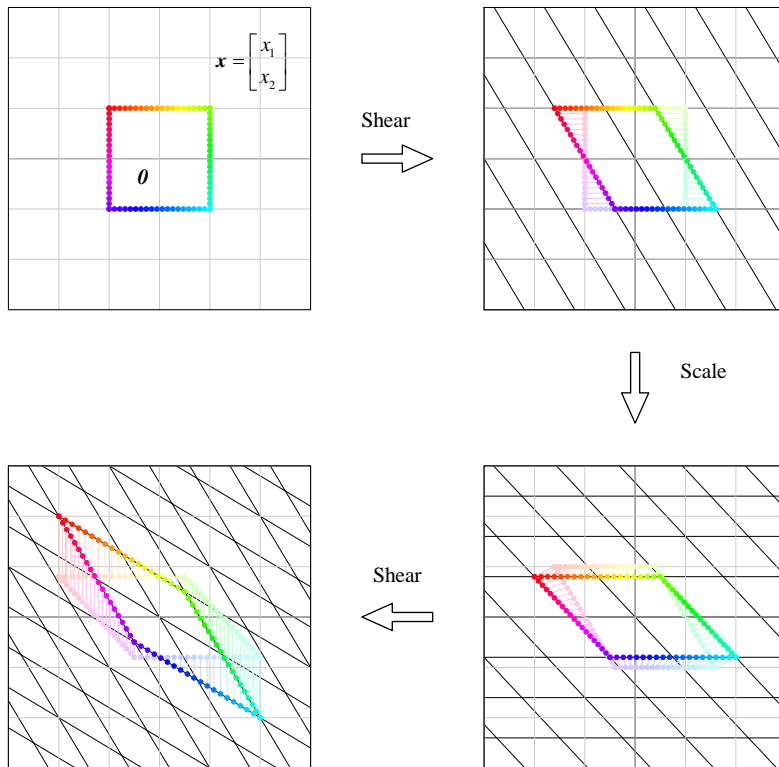
再如，如图 7 所示，我们可以把矩阵 A 拆解成“剪切 \rightarrow 缩放 \rightarrow 剪切”，对应的矩阵乘法为

$$\begin{bmatrix} 5/4 & -3/4 \\ -3/4 & 5/4 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ -0.6 & 1 \end{bmatrix} @ \begin{bmatrix} 1.25 & 0 \\ 0 & 0.8 \end{bmatrix} @ \begin{bmatrix} 1 & -0.6 \\ 0 & 1 \end{bmatrix} \quad (8)$$

需要大家注意的是，图 7 中一个剪切沿横轴，另一个沿纵轴。

(8) 这个分解叫做 LDL 分解，和它类似的分解还有 LU 分解、Cholesky 分解。

不同矩阵分解对应不同的算法，它们也都有各自的几何解读；本书后文将介绍各种常见矩阵分解。

图 7. 把 A 分解成“剪切 → 缩放 → 剪切”

一般情况， AB 不等于 BA

本书前文提过，即便矩阵乘法 AB 、 BA 都存在，一般情况

$$AB \neq BA \quad (9)$$

比如， AB 、 BA 结果的矩阵形状可能不同。

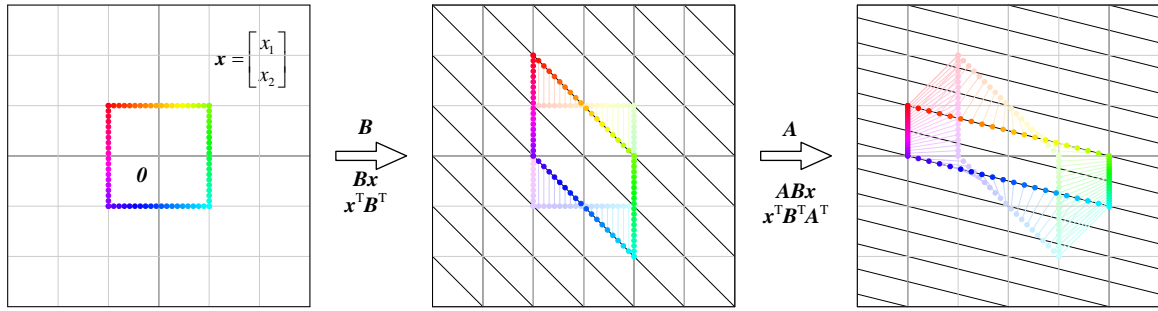
即便 AB 、 BA 形状相同，两者代表的几何变换也可能不同。

给定如下矩阵 A 、 B

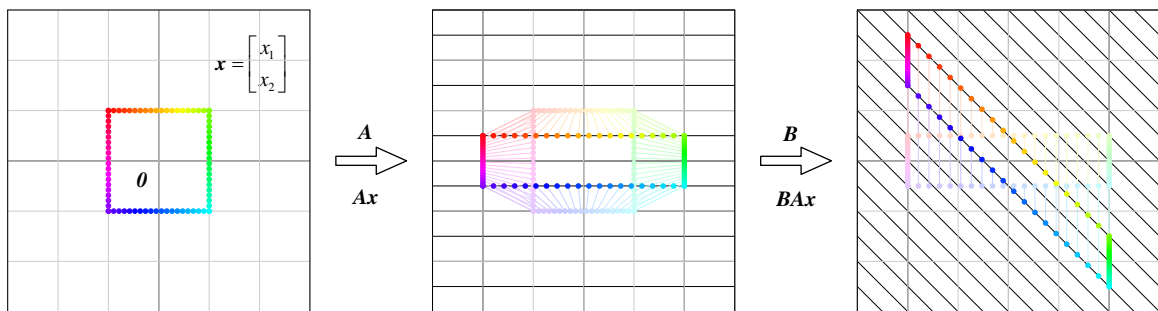
$$A = \begin{bmatrix} 2 & 0 \\ 0 & 1/2 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix} \quad (10)$$

矩阵 A 对应缩放， B 对应剪切。

矩阵乘法 ABx 代表先对 x 进行剪切 (B)，再进行缩放 (A)，具体如图 8 所示。

图 8. $ABx = y$ 对应的分步几何操作

矩阵乘法 BAx 代表先对 x 进行缩放 (A)，再进行剪切 (B)，具体如图 9 所示。

图 9. $BAx = y$ 对应的分步几何操作

特殊情况， $AB = BA$

有一些特殊情况，矩阵乘法 $AB = BA$ 。下面让我们聊聊。

首先，如果 A 、 B 都是形状相同的单位矩阵 I ，显然 $AB = BA$ 。单位矩阵 I 意味着几何体没有任何几何变化。

如果 2×2 矩阵 A 、 B 都是缩放矩阵，比如

$$A = \begin{bmatrix} 1/2 & 0 \\ 0 & 2 \end{bmatrix}, \quad B = \begin{bmatrix} 2 & 0 \\ 0 & 2/3 \end{bmatrix} \quad (11)$$

计算一下，大家会发现

$$AB = BA \quad (12)$$

如图 10 所示，哪怕调换缩放矩阵 A 、 B 的先后，最后的结果完全一致。



请大家计算 (11) 中矩阵乘法 AB 、 BA 。

这说明，这种特殊情况矩阵乘法满足交换律。

再看个例子。

给定 2×2 矩阵 A 、 B 都是沿纵轴剪切矩阵，比如

$$A = \begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & 0 \\ 3/2 & 1 \end{bmatrix} \quad (13)$$

如图 11 所示，显然 ABx 和 BAx 结果完全一致。

? 请大家计算 (13) 中矩阵乘法 AB 、 BA 。

看第三个例子。

给定 2×2 矩阵 A 、 B 都绕原点旋转矩阵，比如

$$A = \begin{bmatrix} 1/2 & -\sqrt{3}/2 \\ \sqrt{3}/2 & 1/2 \end{bmatrix}, \quad B = \begin{bmatrix} -1/2 & -\sqrt{3}/2 \\ \sqrt{3}/2 & -1/2 \end{bmatrix} \quad (14)$$

如图 12 所示，显然 ABx 和 BAx 结果完全一致。

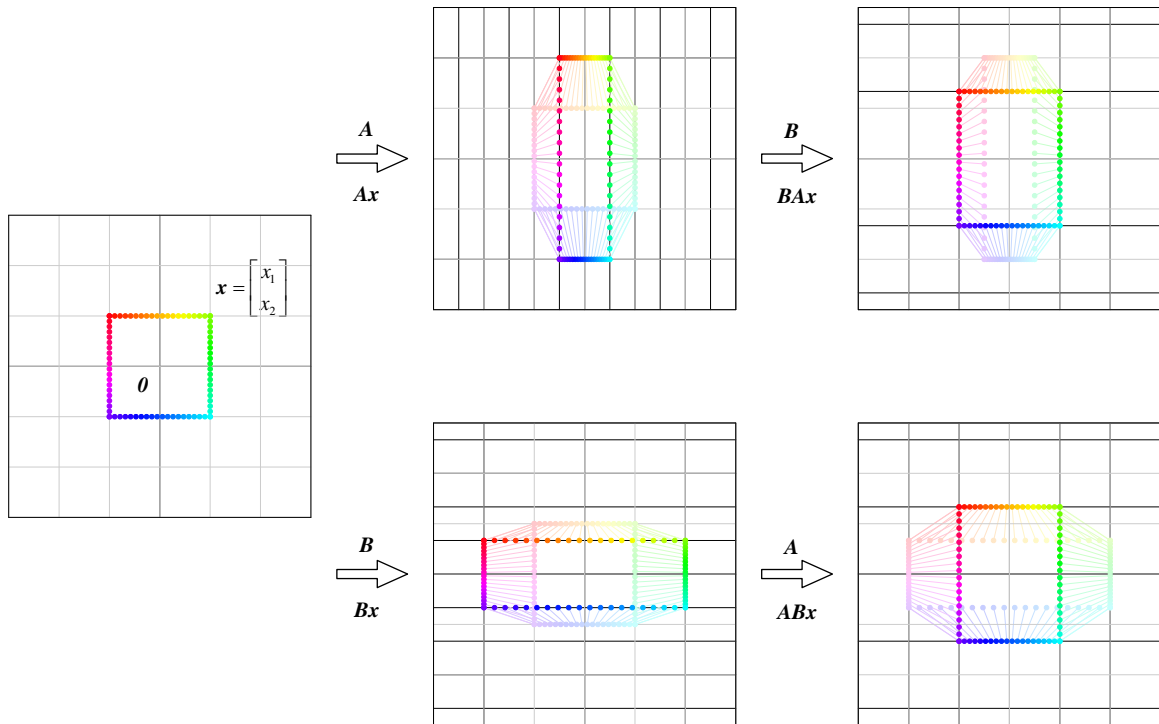
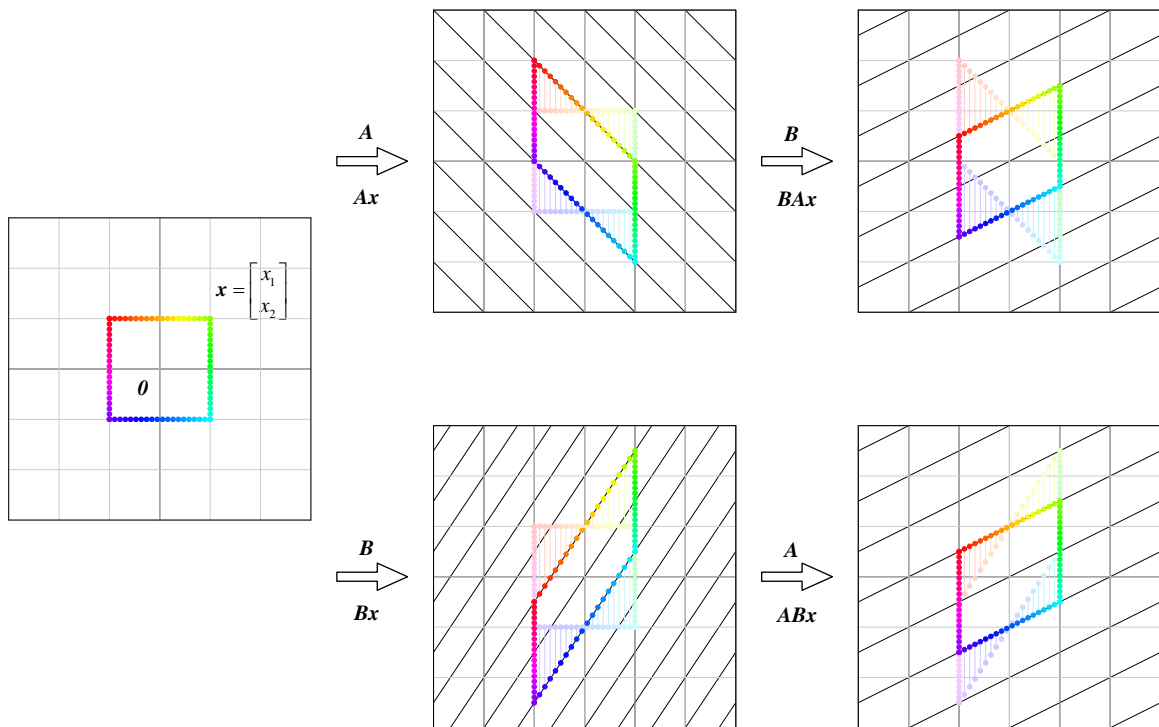
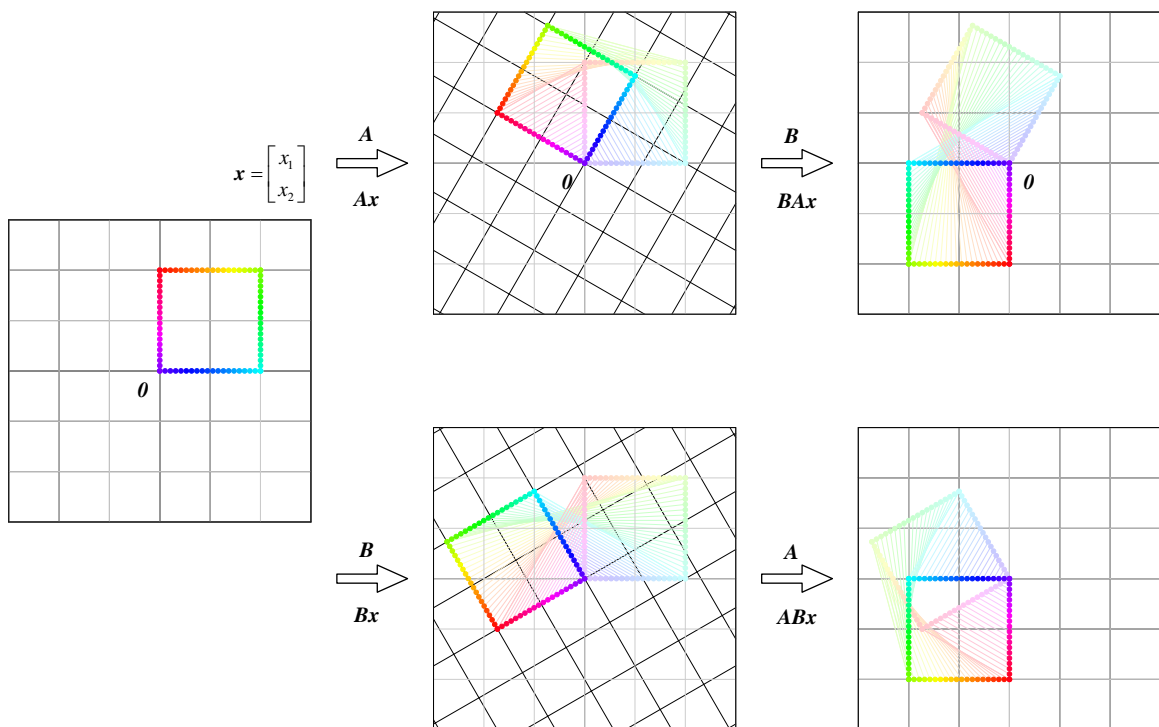


图 10. 2×2 矩阵 A 、 B 都是缩放矩阵， $AB = BA$

图 11. 2×2 矩阵 A 、 B 都是沿纵轴方向剪切矩阵, $AB = BA$ 图 12. 2×2 矩阵 A 、 B 都是绕原点旋转矩阵, $AB = BA$

本书后续将会从这些几何变化机理角度讲解上述特殊矩阵乘法。

矩阵幂

矩阵幂 (power of a matrix) 指的是方阵 A 的多次乘积,

$$\begin{aligned} A^0 &= I \\ A^1 &= A \\ A^k &= \underbrace{A @ A @ \cdots @ A}_{k \text{ times}} \end{aligned} \quad (15)$$

注意, 矩阵幂的前提是矩阵必须是方阵。如果不是方阵, 则矩阵的乘法无法进行多次迭代, 因为矩阵维度无法匹配。

此外, 上式中 k 可以为负整数, 此时要求矩阵 A 可逆。

几何角度来看, 矩阵幂可以解释为线性变换 A 的反复作用,

$$A^k x = A(A^{k-1}x) \quad (16)$$

举个例子, 给定矩阵 A 如下

$$A = \begin{bmatrix} 1 & -1/2 \\ -1/2 & 1 \end{bmatrix} \quad (17)$$

如图 13 所示, 矩阵 A 连续作用在向量 x 上 ($A^k x$) 让平面几何形状放大的同时不断旋转。

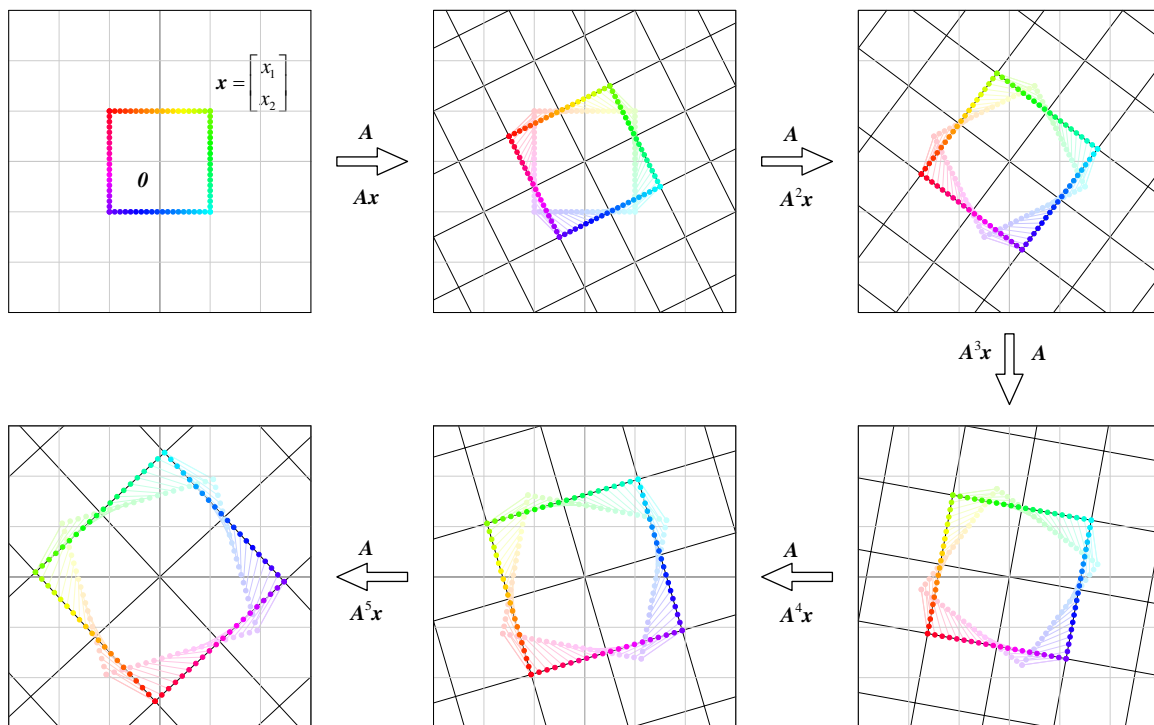


图 13. 2×2 矩阵 A 幂

$AB = O$

如果 A 、 B 矩阵乘积为零矩阵, 即 $AB = O$, 不意味着 A 、 B 为零矩阵; 也不意味着 $BA = O$ 。

举个例子，给定 2×2 矩阵 A 、 B

$$A = \begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix} \quad (18)$$

如图 14 所示， ABx 让正方形“坍塌”到原点。

但是， BA 结果并不是零矩阵；如图 15 所示，正方形发生了降维。

? 请大家计算 (18) 对应的 AB 、 BA 两个矩阵乘法。

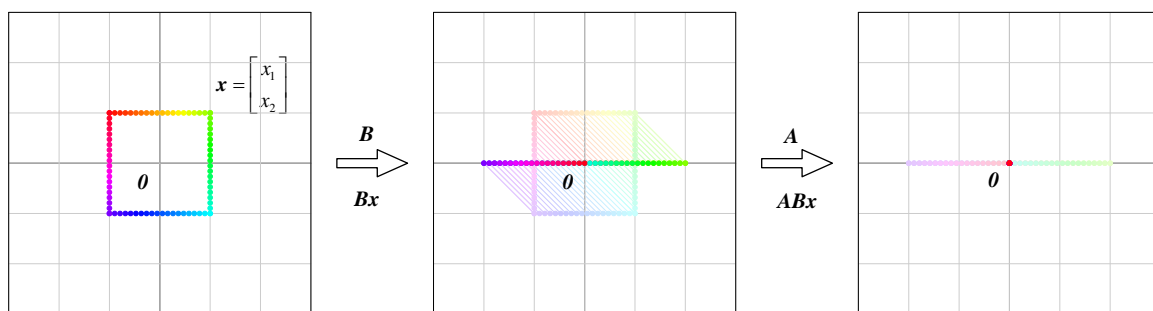


图 14. 2×2 矩阵 A 、 B 矩阵乘积为零矩阵， $AB = O$



图 15. 2×2 矩阵 A 、 B 矩阵乘积为零矩阵， BA 不为 O

此外，如果 $A = B$ ，则 $AC = BC$ 或 $CA = CB$ 。

除非 C 可逆，否则， $AC = BC$ 不能得出 $A = B$ 。

复杂度

举个例子，如下 5 个矩阵相乘

$$A_{m \times p_1} A_{p_1 \times p_2} A_{p_2 \times p_3} A_{p_3 \times p_4} A_{p_4 \times n} \quad (19)$$

如下图所示，所有中间维度都会在计算过程中消去，只保留首尾维度 m 和 n 。

第一种矩阵连乘的总复杂度为 1100 ($200 + 400 + 500$)。

第二种矩阵连乘的分步复杂度为：

- ▶ **第 1 步**， $A @ B$ 复杂度为 200 ($5 \times 10 \times 4$)。
- ▶ **第 2 步**， $C @ D$ 复杂度为 400 ($4 \times 20 \times 5$)。
- ▶ **第 3 步**， $(A @ B) @ (C @ D)$ 复杂度 100 ($5 \times 4 \times 5$)； $(A @ B)$ 形状为 5×4 ， $(C @ D)$ 形状为 4×5 。

第二种矩阵连乘的总复杂度为 700 ($200 + 400 + 100$)。

这个例子告诉我们，在矩阵连乘的运算中，选择适当的计算顺序至关重要。优先在计算过程中消去较大的中间维度，能够显著减少后续运算量，从而大幅降低整体复杂度。

优化计算顺序不仅仅是一种数学技巧，它还是高效完成大规模线性代数任务的重要基础。特别是在机器学习和数据科学中，矩阵运算广泛用于训练模型、处理高维数据以及计算特征分解。通过优化矩阵乘法顺序，可以显著提高计算性能，尤其是在处理神经网络的前向传播和反向传播时。

此外，现代优化方法还包括利用矩阵的结构特性。例如，稀疏矩阵可以大幅减少非零元素的参与，提高运算速度；块矩阵能够在分块计算中充分利用并行计算资源。这些方法为矩阵运算的进一步优化提供了多样化的工具和思路，推动了线性代数在计算领域的广泛应用。优化计算顺序、利用结构特性和先进算法，共同构成了高效矩阵运算的理论与实践框架。



请大家用 DeepSeek/ChatGPT 等工具完成本节如下习题。

Q1. 用随机正整数发生器生成四个 2×2 矩阵 A 、 B 、 C 、 D ，自己写 Python 代码计算所有全排列矩阵 (比如 $A @ B @ D @ C$) 的连乘结果。

Q2. 如下矩阵 A 沿横轴缩放， B 代表沿纵轴方向缩放，是否满足 $AB = BA$

▶ $A = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}$, $B = \begin{bmatrix} 1 & 0 \\ 0 & 3 \end{bmatrix}$

Q3. 如下矩阵 A 沿横轴剪切， B 代表沿纵轴剪切，是否满足 $AB = BA$

▶ $A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$, $B = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$

Q4. 若干沿横轴剪切矩阵，请计算矩阵乘法 $ABCD$ ，有什么规律

▶ $A = \begin{bmatrix} 1 & k_1 \\ 0 & 1 \end{bmatrix}$, $B = \begin{bmatrix} 1 & k_2 \\ 0 & 1 \end{bmatrix}$, $C = \begin{bmatrix} 1 & k_3 \\ 0 & 1 \end{bmatrix}$, $D = \begin{bmatrix} 1 & k_4 \\ 0 & 1 \end{bmatrix}$

Q5. 给定如下不同矩阵 A ，请计算 A^8 ，并从几何角度解释线性变换的作用。

▶ $\begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix}$

$$\rightarrow \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$$

$$\rightarrow \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$$

$$\rightarrow \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$$

$$\rightarrow \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}$$

$$\rightarrow \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$$

$$\rightarrow \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}$$