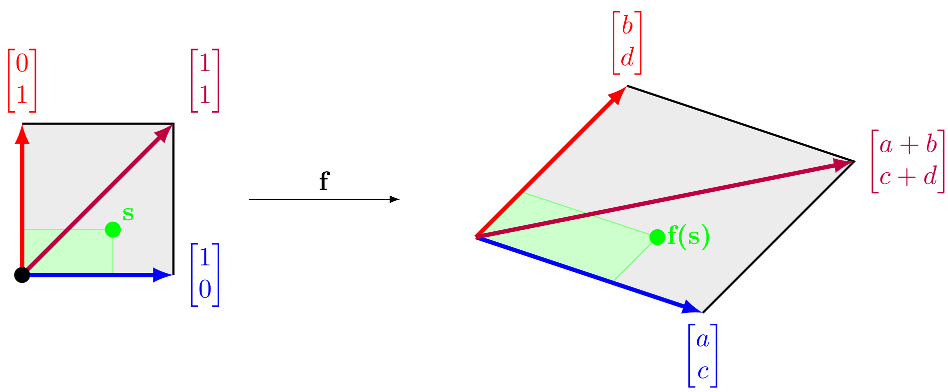
Topic(s): linear transformations

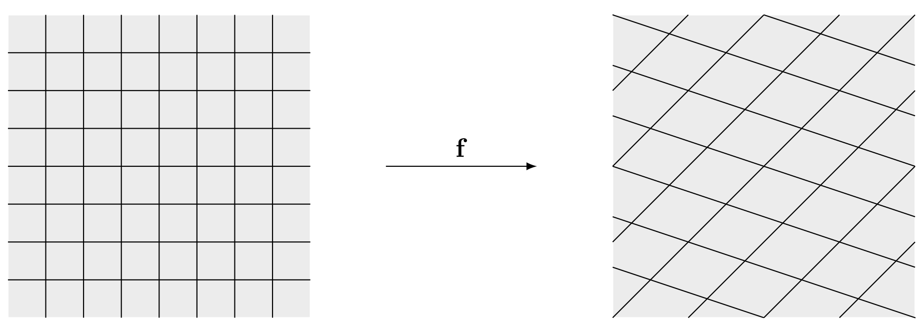
A function $\mathbf{f} : \mathbb{R}^n \to \mathbb{R}^m$ is called

- a **linear function** or a **linear transformation** if there is an $m \times n$ matrix for which $\mathbf{f}(\mathbf{x}) = A\mathbf{x}$ for all vectors $\mathbf{x} \in \mathbb{R}^n$. Recall that this implies that the jth column of A is $A\mathbf{e}_j = f(\mathbf{e}_j)$.
- an **affine function** or an **affine transformation** if there is an $m \times n$ matrix A and a vector $\mathbf{b} \in \mathbb{R}^m$ for which $\mathbf{f}(\mathbf{x}) = A\mathbf{x} + \mathbf{b}$ for all vectors $\mathbf{x} \in \mathbb{R}^n$.

Let us try to visualize the effect of the linear transformation $\mathbf{f} : \mathbb{R}^2 \to \mathbb{R}^2$ associated with $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$, where the two column vectors are not scalar multiples of each other. Let $S = \{(x, y) : 0 \le x, y \le 1\}$ be the unit square in $\mathbb{R}^2$. We can think of $\mathbb{R}^2$ is tiled by copies of S laid out in all directions. Let $\mathbf{f}(S)$ denote the output of $\mathbf{f}$ on S (the collection of points $\mathbf{f}(\mathbf{s})$ for $\mathbf{s} \in S$), called the **image** of S under $\mathbf{f}$. The image $\mathbf{f}(S)$ is a paralleologram; two principles that help us visualize the effect of $\mathbf{f}$ on $\mathbb{R}^2$ are

- **Linearity Principle.** For $c_1, c_2 \in \mathbb{R}$ and $\mathbf{v}_1, \mathbf{v}_2 \in \mathbb{R}^2$, we have $\mathbf{f}(c_1\mathbf{v}_1 + c_2\mathbf{v}_2) = c_1\mathbf{f}(\mathbf{v}_1) + c_2\mathbf{f}(\mathbf{v}_2)$.
- **Tiling Principle.** $\mathbf{f}$ transforms the tiling of $\mathbb{R}^2$ by copies of S into a tiling of $\mathbb{R}^2$ by copies of $\mathbf{f}(S)$.





Theorem 14.2.1. A function $\mathbf{g} : \mathbb{R}^n \to \mathbb{R}^m$ is linear precisely when it respects the vector operations

- $\mathbf{g}(c\mathbf{x}) = c\mathbf{g}(\mathbf{x})$, and (scalar multiplication)
- $\mathbf{g}(\mathbf{x} + \mathbf{y}) = \mathbf{g}(\mathbf{x}) + \mathbf{g}(\mathbf{y})$ (vector addition)

for all scalars $c \in \mathbb{R}$ and vector $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$. Equivalently, $\mathbf{g}$ is linear if, for all scalars a, b and vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$,

$$\mathbf{g}(a\mathbf{x} + b\mathbf{y}) = a\mathbf{g}(\mathbf{x}) + b\mathbf{g}(\mathbf{y}).$$

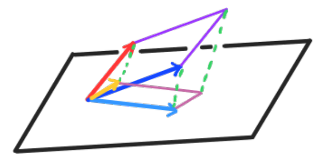
If $\mathbf{g} : \mathbb{R}^n \to \mathbb{R}^m$ and $\mathbf{h} : \mathbb{R}^p \to \mathbb{R}^n$ are linear, the composition $\mathbf{g} \circ \mathbf{h} : \mathbb{R}^p \to \mathbb{R}^m$ is linear also.

Example 1. For any non-zero linear subspace V of $\mathbb{R}^n$, the projection function $\mathbf{Proj}_V : \mathbb{R}^n \to \mathbb{R}^n$ is a linear function.

Suppose $\{\vec{v}_1, \dots, \vec{v}_k\}$ is an orthogonal basis for V.

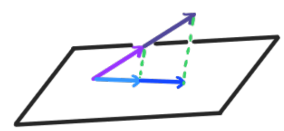
Vector addition: For $\vec{w}, \vec{u} \in \mathbb{R}^n$,

$$\begin{aligned}
\mathrm{Proj}_V(\vec{w}+\vec{u}) &= \mathrm{Proj}_{\vec{v}_1}(\vec{w}+\vec{u}) + \mathrm{Proj}_{\vec{v}_2}(\vec{w}+\vec{u}) + \cdots + \mathrm{Proj}_{\vec{v}_k}(\vec{w}+\vec{u}) \\
&= (\mathrm{Proj}_{\vec{v}_1}(\vec{w}) + \mathrm{Proj}_{\vec{v}_2}(\vec{w}) + \cdots + \mathrm{Proj}_{\vec{v}_k}(\vec{w})) \\
&\quad + (\mathrm{Proj}_{\vec{v}_1}(\vec{u}) + \mathrm{Proj}_{\vec{v}_2}(\vec{u}) + \cdots + \mathrm{Proj}_{\vec{v}_k}(\vec{u})) \\
&= \mathrm{Proj}_V(\vec{w}) + \mathrm{Proj}_V(\vec{u}).
\end{aligned}$$



Scalar multiplication: For $\vec{u} \in \mathbb{R}^n$ and $c \in \mathbb{R}$,

$$\begin{aligned}
\mathrm{Proj}_V(c\vec{u}) &= \mathrm{Proj}_{\vec{v}_1}(c\vec{u}) + \mathrm{Proj}_{\vec{v}_2}(c\vec{u}) + \cdots + \mathrm{Proj}_{\vec{v}_k}(c\vec{u}) \\
&= c(\mathrm{Proj}_{\vec{v}_1}(\vec{u}) + \cdots + \mathrm{Proj}_{\vec{v}_k}(\vec{u})) \\
&= c\,\mathrm{Proj}_V(\vec{u}).
\end{aligned}$$



Hence, Proj_V is linear.

Let A be an $m \times n$ matrix and B an $n \times p$ matrix defined by

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdot & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1p} \\ b_{21} & b_{22} & \cdots & b_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ b_{n1} & b_{n2} & \cdots & b_{np} \end{bmatrix}.$$

Let $T_A : \mathbb{R}^n \to \mathbb{R}^m$ and $T_B : \mathbb{R}^p \to \mathbb{R}^n$ be the linear transformations defined by $T_A(\mathbf{x}) = A\mathbf{x}$ for $\mathbf{x} \in \mathbb{R}^n$ and $T_B(\mathbf{y}) = B\mathbf{y}$ for $\mathbf{y} \in \mathbb{R}^p$. Then, the composition $T_A \circ T_B : \mathbb{R}^p \to \mathbb{R}^m$ is a linear transformation, and the associated matrix is AB, the **matrix product** of A and B.

Theorem 14.3.2. The entries of AB are the dot products of the *rows* of A and the *columns* of B. If we write

$$A = \begin{bmatrix} — & \mathbf{a}_1 & — \\ — & \mathbf{a}_2 & — \\ & \vdots & \\ — & \mathbf{a}_m & — \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} | & | & & | \\ \mathbf{b}_1 & \mathbf{b}_2 & \cdots & \mathbf{b}_p \\ | & | & & | \end{bmatrix},$$

then we have

$$AB = \begin{bmatrix} \mathbf{a}_1 \cdot \mathbf{b}_1 & \mathbf{a}_1 \cdot \mathbf{b}_2 & \cdots & \mathbf{a}_1 \cdot \mathbf{b}_p \\ \mathbf{a}_2 \cdot \mathbf{b}_1 & \mathbf{a}_2 \cdot \mathbf{b}_2 & \cdots & \mathbf{a}_2 \cdot \mathbf{b}_p \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{a}_m \cdot \mathbf{b}_1 & \mathbf{a}_m \cdot \mathbf{b}_2 & \cdots & \mathbf{a}_m \cdot \mathbf{b}_p \end{bmatrix} = \begin{bmatrix} | & | & & | \\ A\mathbf{b}_1 & A\mathbf{b}_2 & \cdots & A\mathbf{b}_p \\ | & | & & | \end{bmatrix}.$$

More explicitly, the ij-entry of AB is

$$\sum_{k=1}^{n} a_{ik} b_{kj},$$

the dot product of $\mathbf{a}_i$ and $\mathbf{b}_j$.

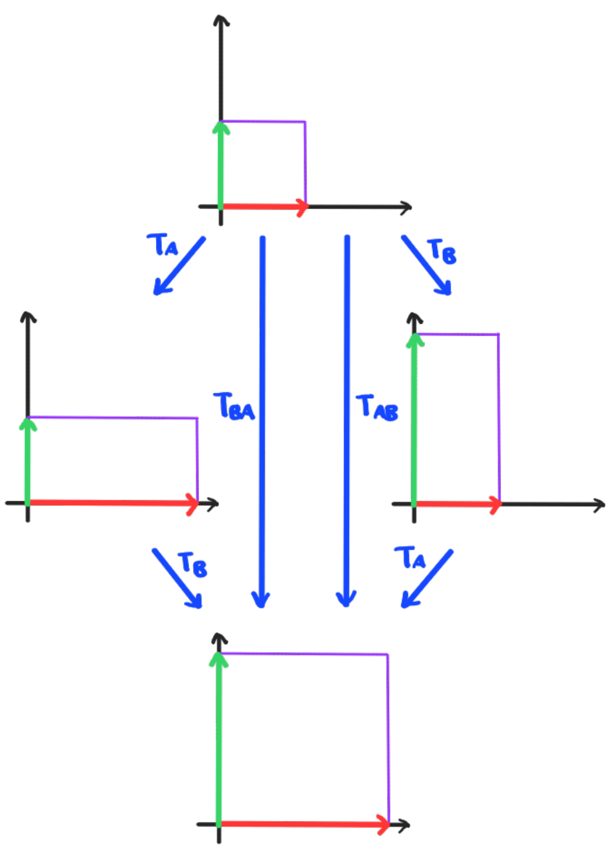
Example 2. Let $A = \begin{bmatrix} 2 & 5 \\ 1 & 1 \end{bmatrix}$ and $B = \begin{bmatrix} 3 & 1 \\ 0 & -1 \end{bmatrix}$. Compute AB.

$$AB = \begin{bmatrix} 2 & 5 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} 3 & 1 \\ 0 & -1 \end{bmatrix} = \begin{bmatrix} 6 & -3 \\ 3 & 0 \end{bmatrix}.$$

Let $C = \begin{bmatrix} 2 & 1 \\ 5 & 1 \end{bmatrix}$. Compute $C^2 = CC$.

$$C^2 = \begin{bmatrix} 2 & 1 \\ 5 & 1 \end{bmatrix}\begin{bmatrix} 2 & 1 \\ 5 & 1 \end{bmatrix} = \begin{bmatrix} 9 & 3 \\ 15 & 6 \end{bmatrix}.$$

Example 3. Let $A = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}$ and $B = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}$. What does T_A and T_B do, respectively? Compute AB and BA. What does T_{AB} and T_{BA} do respectively?



$$T_A : \begin{bmatrix} 1 \\ 0 \end{bmatrix} \longmapsto \begin{bmatrix} 2 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix} \longmapsto \begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$

T_A stretches horizontally by a factor of 2.

$$T_B : \begin{bmatrix} 1 \\ 0 \end{bmatrix} \longmapsto \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix} \longmapsto \begin{bmatrix} 0 \\ 2 \end{bmatrix}.$$

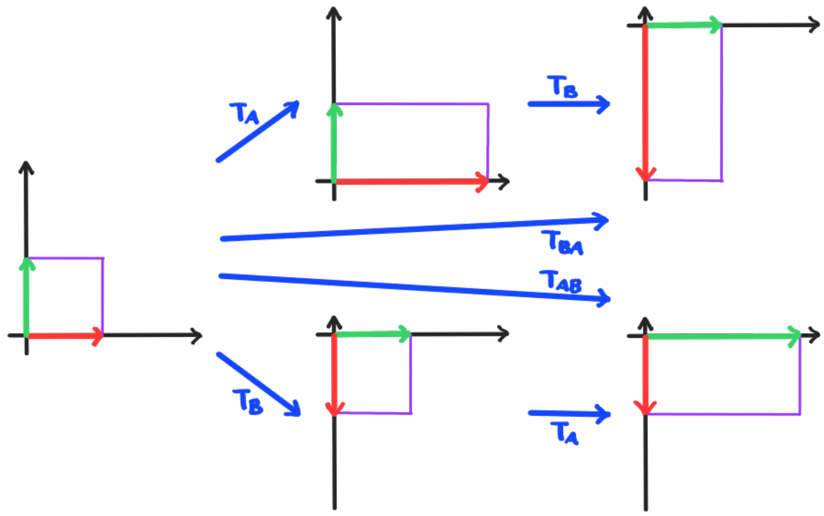
T_B stretches vertically by a factor of 2.

$$AB = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix} = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}.$$

$$BA = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}\begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}.$$

T_{AB} and T_{BA} both stretch horizontally and vertically by a factor of 2, i.e. doubles everything.

Example 4. Let $A = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}$ and $B = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$. What does T_A and T_B do, respectively? Compute AB and BA. What does T_{AB} and T_{BA} do respectively?



T_A stretches horizontally by a factor of 2.

T_B rotates 90° clockwise.

$$AB = \begin{bmatrix} 0 & 2 \\ -1 & 0 \end{bmatrix}$$

$$BA = \begin{bmatrix} 0 & 1 \\ -2 & 0 \end{bmatrix}$$

T_{AB} ($= T_A \circ T_B$) rotates 90° clockwise, then stretches horizontally by a factor of 2.

T_{BA} ($= T_B \circ T_A$) stretches horizontally by a factor of 2, then rotates 90° clockwise.

Example 5. Let $A = \begin{bmatrix} 1 & 2 & 3 \\ 0 & 4 & 5 \\ 0 & 0 & 6 \end{bmatrix}$ and $B = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 2 & 2 \\ 0 & 0 & 3 \end{bmatrix}$. Compute AB and BA.

$$AB = \begin{bmatrix} 1 & 2 & 3 \\ 0 & 4 & 5 \\ 0 & 0 & 6 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 \\ 0 & 2 & 2 \\ 0 & 0 & 3 \end{bmatrix} = \begin{bmatrix} 1 & 5 & 14 \\ 0 & 8 & 23 \\ 0 & 0 & 18 \end{bmatrix}$$

$$BA = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 2 & 2 \\ 0 & 0 & 3 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 \\ 0 & 4 & 5 \\ 0 & 0 & 6 \end{bmatrix} = \begin{bmatrix} 1 & 6 & 14 \\ 0 & 8 & 22 \\ 0 & 0 & 18 \end{bmatrix}$$

In general, $AB \neq BA$.

Example 6. Let $A = \begin{bmatrix} 1 & 2 & 3 \end{bmatrix}$ and $B = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$. Compute AB and BA.

$$AB = \begin{bmatrix} 1 & 2 & 3 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} = \begin{bmatrix} 14 \end{bmatrix}$$

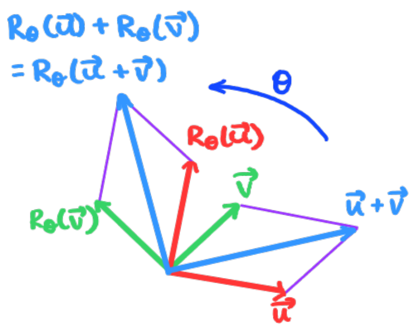
$$BA = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 \end{bmatrix} = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 4 & 6 \\ 3 & 6 & 9 \end{bmatrix}$$

In $\mathbb{R}^2$, if you rotate $\mathbf{e}_1 = \begin{bmatrix}1\\0\end{bmatrix}$ and $\mathbf{e}_2 = \begin{bmatrix}0\\1\end{bmatrix}$ counterclockwise by θ, you get $\begin{bmatrix}\cos\theta\\ \sin\theta\end{bmatrix}$ and $\begin{bmatrix}-\sin\theta\\ \cos\theta\end{bmatrix}$, respectively. Hence, if we consider the map R_θ, which rotates a 2-vector counterclockwise by θ, the associated matrix would be

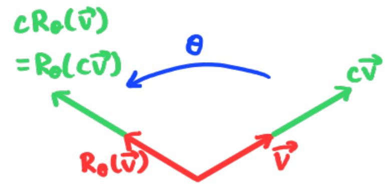
$$A_\theta = \begin{bmatrix}\cos\theta & -\sin\theta\\ \sin\theta & \cos\theta\end{bmatrix}.$$

In other words, $R_\theta(\mathbf{x}) = A_\theta\mathbf{x}$. **The rotation map R_θ is linear.**

Vector addition:



Scalar multiplication:



Example 7. Consider the matrix $A = \begin{bmatrix}0&1&0\\0&0&1\\0&0&0\end{bmatrix}$.

(a) Describe geometrically what A does to each of the coordinate axes.

$A\vec{e}_1 = \vec{0}$ The x-component vanishes.

$A\vec{e}_2 = \vec{e}_1$ The y-component rotates to the x-axis.

$A\vec{e}_3 = \vec{e}_2$ The z-component rotates to the y-axis.

(b) Calculate A^3 in two ways: the geometric viewpoint of linear transformations and algebraically by multiplying matrices.

Geometrically:

$$\begin{aligned} A\vec{e}_1 &= \vec{0} & A^2\vec{e}_1 &= A\vec{0} = \vec{0} & A^3\vec{e}_1 &= A\vec{0} = \vec{0}\\ A\vec{e}_2 &= \vec{e}_1 & A^2\vec{e}_2 &= A\vec{e}_1 = \vec{0} & A^3\vec{e}_2 &= A\vec{0} = \vec{0}\\ A\vec{e}_3 &= \vec{e}_2 & A^2\vec{e}_3 &= A\vec{e}_2 = \vec{e}_1 & A^3\vec{e}_3 &= A\vec{e}_1 = \vec{0} \end{aligned}$$

Hence, $A^3 = 0$.

Algebraically:

$$A^2 = \begin{bmatrix}0&0&1\\0&0&0\\0&0&0\end{bmatrix} \quad \text{and} \quad A^3 = \begin{bmatrix}0&0&0\\0&0&0\\0&0&0\end{bmatrix}.$$

Example 8. Recall that $\mathbf{1} = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \in \mathbb{R}^n$. Define a function $f_{\mathbf{1}} : \mathbb{R}^n \to \mathbb{R}$ by $f_{\mathbf{1}}(\mathbf{x}) = \mathbf{x} \cdot \mathbf{1}$.

(a) Show that $f_{\mathbf{1}}$ is linear by checking that it interacts well with vector addition and scalar multiplication.

Vector addition:

$$f_{\vec{1}}(\vec{u}+\vec{w}) = (\vec{u}+\vec{w})\cdot\vec{1} = \vec{u}\cdot\vec{1} + \vec{w}\cdot\vec{1} = f_{\vec{1}}(\vec{u}) + f_{\vec{1}}(\vec{w})$$

Scalar multiplication:

$$f_{\vec{1}}(c\vec{u}) = (c\vec{u})\cdot\vec{1} = c(\vec{u}\cdot\vec{1}) = c f_{\vec{1}}(\vec{u})$$

Hence, $f_{\vec{1}}$ is linear.

(b) Find the $1 \times n$ matrix representation of $f_{\mathbf{1}}$ (the matrix entries will be in terms of the v_i's).

$$\begin{aligned} A_{f_{\vec{1}}} &= [\, f_{\vec{1}}(\vec{e}_1) \;\; f_{\vec{1}}(\vec{e}_2) \;\; \cdots \;\; f_{\vec{1}}(\vec{e}_n) \,] \\ &= [\, \vec{e}_1\cdot\vec{1} \;\; \vec{e}_2\cdot\vec{1} \;\; \cdots \;\; \vec{e}_n\cdot\vec{1} \,] \\ &= [\, 1 \;\; 1 \;\; \cdots \;\; 1 \,] \end{aligned}$$

Example 9. Find a 3×3 matrix A for which A^3 acts on $\mathbb{R}^3$ as the identity function (carrying each 3-vector to itself) but A and A^2 are not the identity function. (There are many possible answers.) Hint: think geometrically, and try to solve the analogous problem on $\mathbb{R}^2$ for 2×2 matrices first.

We can think about rotating $120°$ counterclockwise on the xy-plane. Then,

$$A\vec{e}_1 = \begin{bmatrix} -\frac{1}{2} \\ \frac{\sqrt{3}}{2} \\ 0 \end{bmatrix}, \quad A\vec{e}_2 = \begin{bmatrix} -\frac{\sqrt{3}}{2} \\ -\frac{1}{2} \\ 0 \end{bmatrix}, \quad \text{and} \quad A\vec{e}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix},$$

and so, $A = \begin{bmatrix} -\frac{1}{2} & -\frac{\sqrt{3}}{2} & 0 \\ \frac{\sqrt{3}}{2} & -\frac{1}{2} & 0 \\ 0 & 0 & 1 \end{bmatrix}$.

Indeed, $A^2 = \begin{bmatrix} -\frac{1}{2} & \frac{\sqrt{3}}{2} & 0 \\ -\frac{\sqrt{3}}{2} & -\frac{1}{2} & 0 \\ 0 & 0 & 1 \end{bmatrix}$ and $A^3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$.