Perhaps matrix decomposition techniques are among the most important concepts in Machine Learning. This article will introduce some of the essences in linear algebra along with techniques that are widely popular in matrix decomposition. The content of this article is inspired by UCSD CSE 190: **Discrete and Continuous Optimization**

The Brief Essence of Linear Algebra

Lets consider this $\vec{x}$:

$$\begin{bmatrix} a \\ b \end{bmatrix}$$

It lives inside a 2D space with two basis vectors $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $\begin{bmatrix} 0 \\ 1 \end{bmatrix}$. Thus, the vector can be rewriten as:

$$\vec{x} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} a \\ b \end{bmatrix}$$

By observing the equation above, we can see that the underlying basis matrix is essential in determining the space in which $\vec{x}$ lives. By changing the basis, we can translate $\vec{x}$ into completely different representations in another space with a distinct set of basis vectors, and this change of basis in linear algebra is called **linear transformation**

Linear Transformation

The official definition of linear transformation from Wolfram Alpha is: a transformation between two vector spaces V and W is a map $T{:}V \to W$ such that the following hold:

- $T(v_1 + v_2) = T(v_1) + T(v_2)$ for any vectors v_1 and v_2 in V, and
- $T(\alpha v) = \alpha T(v)$ for any scalar α

Intuitively, a linear transformation rotate and scale each vectors inside one space and translate their representations onto another space such that all the parallel and orthogonal properties among vectors are being preserved. Mathematically, this can be represented as **a change of basis** for $\vec{x_1}$ and $\vec{x_2}$:

$$M \begin{bmatrix} \vec{x_1} & \vec{x_2} \end{bmatrix} = \begin{bmatrix} \begin{bmatrix} q_1 & p_1 \\ q_2 & p_2 \end{bmatrix} \vec{x_1} & \begin{bmatrix} q_1 & p_1 \\ q_2 & p_2 \end{bmatrix} \vec{x_2} \end{bmatrix} = \begin{bmatrix} \begin{bmatrix} q_1 & p_1 \\ q_2 & p_2 \end{bmatrix} \begin{bmatrix} a_1 \\ b_1 \end{bmatrix} & \begin{bmatrix} q_1 & p_1 \\ q_2 & p_2 \end{bmatrix} \begin{bmatrix} a_2 \\ b_2 \end{bmatrix} \end{bmatrix}$$

The tricky concepts of here is that although $\vec{x_1}$ and $\vec{x_2}$ are being transformed by a matrix M with two basis vectors $\begin{bmatrix} q_1 \\ p_1 \end{bmatrix}$ $\begin{bmatrix} q_2 \\ p_2 \end{bmatrix}$, the result of the two vectors actually lives inside the dimension space with basis vectors $\begin{bmatrix}1\\0\end{bmatrix}$ and $\begin{bmatrix}0\\1\end{bmatrix}$. This means the function of matrix M is essentially translating a foreign vector into the xy dimension that we are familiar with without changing any of the properties of the translated vectors.

A **different matrix-vector multiplication** can be represented as:

$$M\begin{bmatrix}a_1 & b_1\end{bmatrix} = \begin{bmatrix}\vec{c_1}\\\vec{c_2}\end{bmatrix}\begin{bmatrix}a_1\\b_1\end{bmatrix} = \begin{bmatrix}\vec{c_1}\end{bmatrix}a_1 + \begin{bmatrix}\vec{c_2}\end{bmatrix}$$

Here we can see that a_1 and b_1 serves as **coefficients** for the columns of matrix A.

Determinant

The determinant of a matrix A can be denotated as $\det(A)$. When we perform a linear transformation, we know that all the vectors from one space are rotated, stretched and translated evenly into another space with a different set of basis (**change of basis**). Therefore, the two important aspects of a linear transformation or matrix A are the rotation and the scaling factor. Determinant indicates the scaling factor of the matrix A.

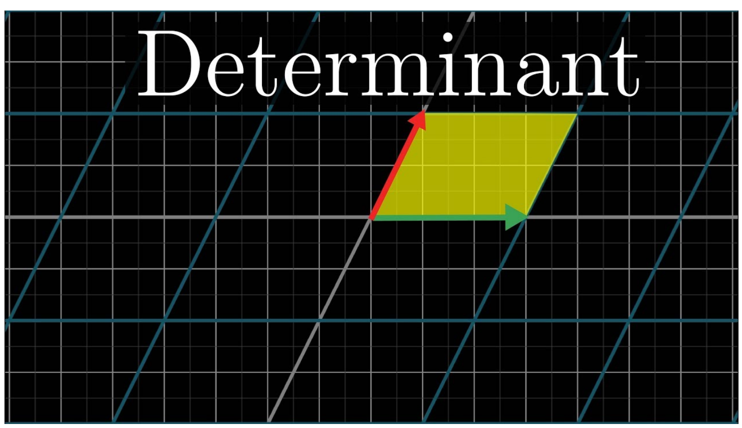

Figure 1: Determinant: Area Demonstration

If the determinant of A is c, then the post-transformation area between $\vec{x_1}$ and $\vec{x_2}$ will be scaled by c. This means that if $\det(A) = 1$, the matrix A does not change the magnitude of the vectors in their original space, and sometimes we call it a **rotation matrix**. If $\det(A) = 0$, we know that after rotation and scaling factor, the area between $\vec{x_1}$ and $\vec{x_2}$ becomes 0, which indicates that the rotation aligned the two vectors onto the same axis. This property becomes particularly important when we are trying to use determinant in finding the **eigenvalues and eigen-decomposition** of a matrix.

Dot Product

The dot product represents the concept of projection in linear algebra. It is denotated as $\dot{\vec{x}\vec{y}}$ with x and y living in the space with same set of basis (**this is important**).

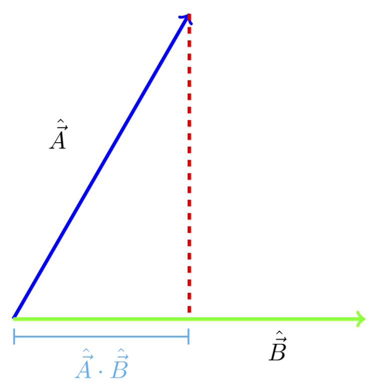

Figure 2: Dot Product: Projection Visualization

$$\mathbf{u} \cdot \mathbf{v} = ||u|| \times ||v|| \cos\theta \cos\theta = \frac{\mathbf{u} \cdot \mathbf{v}}{||u|| \times ||v||}$$

As we have seen in the equations, the dot product is associative, which means the projection is symmetric. Thus, $u\dot{v}$ can either be considered as $\vec{u}$ projected onto $\vec{v}$ or the other way around.

The implication of dot product are used in many applications such as linear regression (**least square**) and image projections. For example, the error of least square is calculated as:

$$\text{residual} = ||\vec{a} - \text{proj}_b(a) = \vec{a} - \frac{\dot{\vec{a}\vec{b}}}{\vec{b}\vec{b}^T}\vec{b}||_2^2$$

where $\frac{\dot{\vec{a}\vec{b}}}{\vec{b}\vec{b}^T}\vec{b}$ is the shadow project of $\vec{a}$ onto $\vec{b}$

Eigenvalue and Eigenvector

Matrix Decomposition

Gaussian Elimination

There are two separate processes in **Gaussian Elimination**: the forward substitution and the backward substitution. The forward substitution requires us to perform substitution from **left to right** and **top to bottom**. This is an essential property that makes **Gaussian Elimination** works, and intuitively gaussian elimination preserves the property of the matrix as we will show in later paragraphs.

$$\begin{aligned} x+y-2z &= -3 \\ y-z &= -1 \\ 3x-y+z &= 4 \end{aligned} \qquad \left(\begin{array}{ccc|c} 1 & 1 & -2 & -3 \\ 0 & 1 & -1 & -1 \\ 3 & -1 & 1 & 4 \end{array}\right)$$

We then *substitute* the first equation into the third to eliminate the $3x$ term. This is the same as scaling the relationship $x+y-2z=-3$ by -3 and adding the result to the third equation:

$$\begin{aligned} x+y-2z &= -3 \\ y-z &= -1 \\ -4y+7z &= 13 \end{aligned} \qquad \left(\begin{array}{ccc|c} 1 & 1 & -2 & -3 \\ 0 & 1 & -1 & -1 \\ 0 & -4 & 7 & 13 \end{array}\right)$$

Similarly, to eliminate y from the third equation, we scale the second equation by 4 and add the result to the third:

$$\begin{aligned} x+y-2z &= -3 \\ y-z &= -1 \\ 3z &= 9 \end{aligned} \qquad \left(\begin{array}{ccc|c} 1 & 1 & -2 & -3 \\ 0 & 1 & -1 & -1 \\ 0 & 0 & 3 & 9 \end{array}\right)$$

We have now isolated z! We scale the third row by $1/3$ to yield an expression for z:

$$\begin{aligned} x+y-2z &= -3 \\ y-z &= -1 \\ z &= 3 \end{aligned} \qquad \left(\begin{array}{ccc|c} 1 & 1 & -2 & -3 \\ 0 & 1 & -1 & -1 \\ 0 & 0 & 1 & 3 \end{array}\right)$$

Now, we substitute $z=3$ into the other two equations to remove z from all but the final row:

$$\begin{aligned} x+y &= 3 \\ y &= 2 \\ z &= 3 \end{aligned} \qquad \left(\begin{array}{ccc|c} 1 & 1 & 0 & 3 \\ 0 & 1 & 0 & 2 \\ 0 & 0 & 1 & 3 \end{array}\right)$$

Figure 3: Gaussian Elimination Example

As we have shown in the figure image, for each row operation, **forward substitution** ensures that the element before the matrix diagonal becomes 0, and **backward substitution** ensure the elements after the matrix diagonal becomes 0.

We can also consider Gaussian Elimination as a set of matrix multiplication. Consider $\vec{e_i}$ as a vector with only the i^{th} element equal to 1 and the rest equals to 0.

$$e_i = \begin{bmatrix} 0 \\ \vdots \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

With this definition, the inner product of e_i and e_j has a value of 1, and the outter product of e_i and e_j becomes a matrix with all elements 0 except element a_ij equals to 1.

$$e_1 e_1^T = \begin{bmatrix} 1 & \dots & 0 & 0 \\ 0 & \dots & 0 & 0 \\ 0 & \dots & 0 & 0 \end{bmatrix}$$

Because a forward substitution elimination matrix is a **lower triangular** matrix, we can note such matrix as L matrix. And turns out we can construct such a L matrix as the sum of multiple outter product of e_i and e_j

$$L = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 3 & 1 & 0 & 0 \\ 2 & 2 & 1 & 0 \\ 5 & -1 & -2 & 2 \end{bmatrix} = 3 \begin{bmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} + \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 2 & 2 & 1 & 0 \\ 5 & -1 & -2 & 1 \end{bmatrix}$$

With this elimination matrix decomposition we can decompose L matrix as:

$$L = E_1 + E_2 + \ldots + E_n$$

where $E = e_i e_j^T$