Understanding SVD

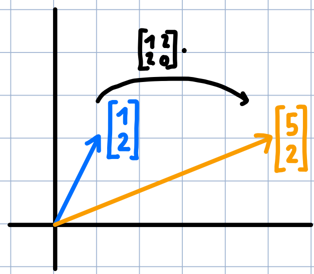
Álvaro Fernández García

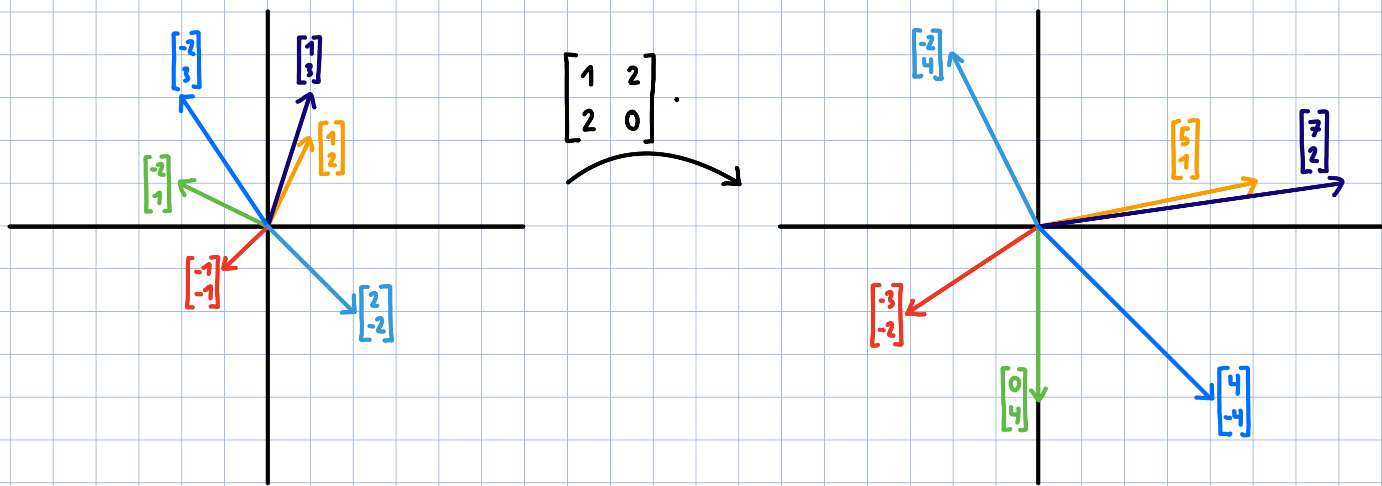
Index:

1.- Different Matrices:

Multiplying a matrix with a vector:

$$\begin{bmatrix} 1 & 2 \\ 2 & 0 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 5 \\ 2 \end{bmatrix}$$





A matrix applied to a vector convert the vector into another vector.

A vector can be represented as a point in the space.

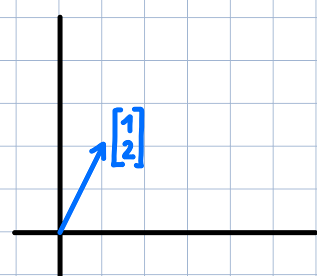
Identity Matrix:

Needs to be square (n by n).

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

It doesn't apply any transformations:

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



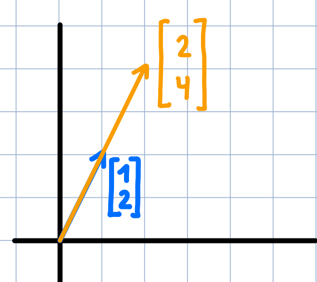
Scalar Matrix:

Needs to be square (n by n).

Constant values on the diagonal.

No distortion. Every single dot moves uniformly.

$$\begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 2 \\ 4 \end{bmatrix}$$



We can apply it to matrices, so with images, so with 3D shapes.

Off-one matrix:

Like the Identity matrix, but one number on diagonal is not 1 or 0.

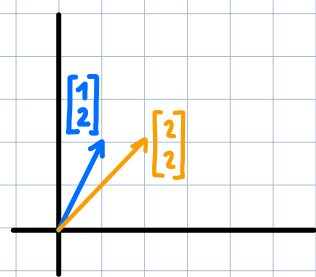
scale the x-axis by K

$$\begin{bmatrix} K & 0 \\ 0 & 1 \end{bmatrix} \qquad \begin{bmatrix} 1 & 0 \\ 0 & K \end{bmatrix}$$

scale the y-axis by K

It's the same in $\mathbb{R}^3$ but with 1 dimension more.

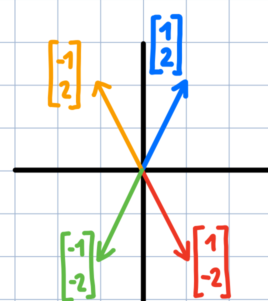
$$\begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 2 \\ 2 \end{bmatrix}$$



Reflexion Matrix:

Like the Identity matrix, but some numbers on the diagonal are -1.

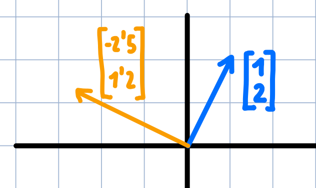
$$\begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} -1 \\ 2 \end{bmatrix} \qquad \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 1 \\ -2 \end{bmatrix} \qquad \begin{bmatrix} -1 & 0 \\ 0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



Diagonal Matrix:

Square matrix where the diagonal entries can be anything.

$$\begin{bmatrix} \text{-2'5} & 0 \\ 0 & \text{0'6} \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} \text{-2'5} \\ \text{1'2} \end{bmatrix}$$



Shear Matrix:

Square matrix applying shear transformation.

It changes the shape but preserves the area.

$\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$ Move all the dots where $y>0$ to the right 1 position & all the dots where $y<0$ to the left 1 position.

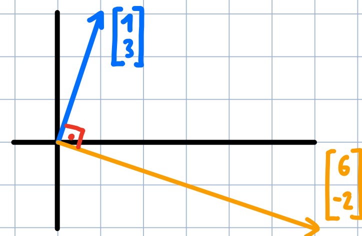
$\begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$ Move all the dots where $x>0$ up 1 position & all the dots where $x<0$ down.

Orthogonal Matrix:

It's a square matrix where all column vectors are unit vectors and all column vectors are orthogonal.

For a column vector to be a unit vector it means that the magnitude is 1.

Two vectors are orthogonal when their dot product equals to 0: $\begin{bmatrix}1 & 3\end{bmatrix} \cdot \begin{bmatrix}6\\-2\end{bmatrix} = \begin{bmatrix}0\end{bmatrix}$



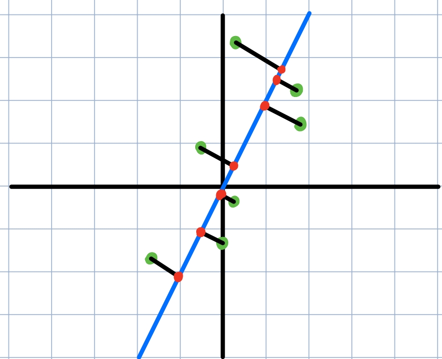
It applies a perfect rotation to a plane or cube.

The sequence of applying the transformation is important!!

$$A\ B \text{ not always} = A\ B$$

Projection Matrix:

It projects all vectors onto a subspace. Every vector moves to it's closest point on the subspace.



Multiplying a matrix with another matrix:

We are combining transformations.

some other transformation

$$B\ A = C$$

combined transformation

some transformation

Inverse of a Matrix:

Untransformation of the effect produces by the original matrix.

There isn't one matrix that does the <u>universal</u> untransformation.

$$A^{-1} A = I$$

Zero matrix and projection matrices can't be untransformed. There is a loss of information.

What exactly is a Matrix?:

We have seen here only a representation of what is a matrix, but it can change depending on field and the context (ML, circuits, probability, etc).

There isn't <u>the definition</u> of matrix, but only interpretations of the matrices.

2.- Spectral Decomposition:

The goal in this section is to understand what a symmetric matrix with a complicated transformation is doing by dividing it into 3 matrices with simpler transforms.

$$\begin{bmatrix} S \end{bmatrix} = \begin{bmatrix} ? \end{bmatrix} \begin{bmatrix} ? \end{bmatrix} \begin{bmatrix} ? \end{bmatrix}$$

Symmetric Matrix:

$$\begin{bmatrix} 2 & 4 & 5 \\ 4 & 7 & 8 \\ 5 & 8 & 0 \end{bmatrix}$$

Rectangular matrix is <u>never</u> symmetrical.

$$\begin{bmatrix} 1 & 2 & 4 & 6 \\ 0 & 5 & 0 & 1 \end{bmatrix}$$

Transpose and its properties:

$$\underset{A}{\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}} \xrightarrow{\text{Transpose}} \underset{A^T}{\begin{bmatrix} 1 & 4 & 7 \\ 2 & 5 & 8 \\ 3 & 6 & 9 \end{bmatrix}}$$

$$\underset{R}{\begin{bmatrix} 2 & 3 \\ 5 & 7 \\ 11 & 13 \end{bmatrix}_{3x2}} \xrightarrow{\text{Transpose}} \underset{R^T}{\begin{bmatrix} 2 & 5 & 11 \\ 3 & 7 & 13 \end{bmatrix}_{2x3}}$$

Transposing a <u>symmetric matrix</u> you get the <u>same matrix</u> back:

$$\underset{S}{\begin{bmatrix} 1 & 2 & 3 \\ 2 & 5 & 6 \\ 3 & 6 & 9 \end{bmatrix}} \xrightarrow{\text{Transpose}} \underset{S^T}{\begin{bmatrix} 1 & 2 & 3 \\ 2 & 5 & 6 \\ 3 & 6 & 9 \end{bmatrix}} , \quad S = S^T$$

Transposing an <u>orthogonal matrix</u> you get its inverse:

$$\underset{Q}{\begin{bmatrix} \sqrt{2}/2 & -\sqrt{2}/4 & \sqrt{6}/4 \\ \sqrt{2}/2 & \sqrt{2}/4 & -\sqrt{6}/4 \\ 0 & \sqrt{3}/2 & 1/2 \end{bmatrix}} \xrightarrow{\text{Transpose}} \underset{Q^T}{\begin{bmatrix} \sqrt{2}/2 & \sqrt{2}/2 & 0 \\ -\sqrt{2}/4 & \sqrt{2}/4 & \sqrt{3}/2 \\ \sqrt{6}/4 & -\sqrt{6}/4 & 1/2 \end{bmatrix}} , \quad Q^T = Q^{-1}$$

Remember:

- 1. Orthogonal matrix produces rotation transformation.
- 2. The inverse of a matrix means untransformation.
- 3. The untransformation of rotation means rotation in the reverse direction.
- 4. The transpose of orthogonal matrix is also the inverse matrix.
- 5. The transpose of orthogonal matrix means rotation in the reverse direction.

Matrix Decomposition:

Matrix Composition: whenever we multiply matrices we are composing their distinct transformations together into one transformation.

$$\begin{bmatrix} -2'5 & 0 \\ 0 & 0'6 \end{bmatrix} = \underbrace{\begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}}_{\text{Reflect around y-axis}} \underbrace{\begin{bmatrix} 1 & 0 \\ 0 & 0'6 \end{bmatrix}}_{\text{Scale y-axis by 0'6}} \underbrace{\begin{bmatrix} 2'5 & 0 \\ 0 & 1 \end{bmatrix}}_{\text{Scale x-axis by 2'5}}$$

Matrix Decomposition is essentially going in the reverse direction, re-expressing the complex transformation as a sequence of much simpler transformations.

It's a difficult task, but hopefully Spectral Decomposition can be the beacon of light.

One interpretation of Eigen Vectors:

They are vectors in matrices and images that do not modify their direction when a deformation is applied. Unlike most of the rest of the vectors, they only change their magnitude.

Different matrices and images <u>can</u> have different Eigen Vectors.

A 2x2 matrix can have less than 2 Eigen Vectors.

Eigen Values & Eigen Vectors always come in pairs.

Eigen Values are numbers. They tell how much its corresponding Eigen Vector is scaled during the matrix transformation.

Formal Definition:

$$\overset{\text{linear transformation}}{A}\,\underset{\text{eigen vector}}{\vec{v}} = \overset{\text{eigen value}}{\lambda}\,\underset{\text{eigen vector}}{\vec{v}}$$

Appreciate orthogonal eigen vectors:

S is symmetric $\longrightarrow$ eigen vectors are orthogonal (perpendicular).

This is really probabilistically unlikely.

Then, this means that there always exists an orthogonal matrix which rotates the bases to align with the eigen vector. An its inverse would rotate the eigen vector to align with the basis.

Spectral Decomposition:

Theorem: Whenever you have a <u>symmetric matrix</u> you can always unconditionally decompose it into a sequence of three simple matrices

$$S = Q \Lambda Q^T$$

orthogonal rotation

diagonal = stretching

$$\begin{bmatrix} S \end{bmatrix}_{2x2} = \begin{bmatrix} | & | \\ \vec{e_1} & \vec{e_2} \\ | & | \end{bmatrix} \begin{bmatrix} \lambda_1 & \\ & \lambda_2 \end{bmatrix} \begin{bmatrix} | & | \\ \vec{e_1} & \vec{e_2} \\ | & | \end{bmatrix}^T$$

Eigen Vectors of S

Scaling x-axis by λ_1

Scaling y-axis by λ_2

Q: rotates the standard basis $(x, y, z, \ldots)$ to align with the Eigen Vectors.

$Q^T = Q^{-1}$: rotates in the reverse direction to the original position of the standard basis.

$$\begin{bmatrix} 4 & 2 \\ 2 & 4 \end{bmatrix} = \begin{bmatrix} \sqrt{2}/2 & -\sqrt{2}/2 \\ \sqrt{2}/2 & \sqrt{2}/2 \end{bmatrix} \begin{bmatrix} 6 & 0 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} \sqrt{2}/2 & \sqrt{2}/2 \\ -\sqrt{2}/2 & \sqrt{2}/2 \end{bmatrix}$$

$$\begin{bmatrix} 3 & 1 & 2 \\ 1 & 0'4 & -1 \\ 2 & -1 & 0'5 \end{bmatrix} = \begin{bmatrix} 0'88 & -0'43 & -0'18 \\ 0'12 & 0'57 & -0'81 \\ 0'45 & 0'7 & 0'56 \end{bmatrix} \begin{bmatrix} 4'15 & 0 & 0 \\ 0 & -1'56 & 0 \\ 0 & 0 & 1'31 \end{bmatrix} \begin{bmatrix} 0'88 & 0'12 & 0'45 \\ -0'43 & 0'57 & 0'7 \\ -0'18 & -0'81 & 0'56 \end{bmatrix}$$

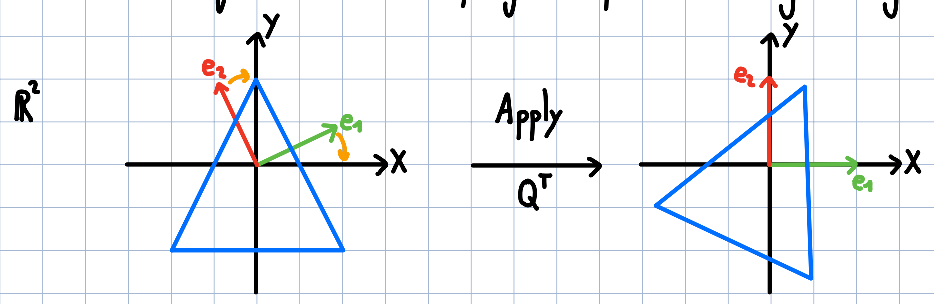
$$\begin{bmatrix} -1 & -2 \\ -2 & -1'8 \end{bmatrix} = \begin{bmatrix} 0'77 & 0'63 \\ -0'63 & 0'77 \end{bmatrix} \begin{bmatrix} 0'64 & 0 \\ 0 & -3'44 \end{bmatrix} \begin{bmatrix} 0'77 & -0'63 \\ 0'63 & 0'77 \end{bmatrix}$$

Visualization of Spectral Decomposition:

It's very important to understand how Q, Λ & Q^T work when we apply them to an object (plane, box, image,...).

1. Apply Q^T:

The idea is to rotate the object the same amout of degrees as if we want to align the eigen vectors with the Standar Basis.



2. Apply Λ:

Scale the object in all the different axis.

3. Apply Q :

Rotate the object back to the original orientation of the eigen vectors.

Pd: This same representation is valid for SVD.

3.- Singular Value Decomposition:

On the surface, it's saying that any matrix regardless of symmetry, rank or shape can unconditionally be decomposed into three very special matrices.

$$A = U\Sigma V^T$$

It's extremely relevant and applicable for other purposes: PCA, low-rank approximation, TLS minimization, pseudoinverse, separable models, optimal rotation,...

Rectangular Matrix:

It's able to convert a vector $\mathbb{R}^n \to \mathbb{R}^m$.

$$\begin{bmatrix} 3 & -2 & 1 \\ 2 & 1 & -1 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$$

This is why we say: matrices apply a linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^m$.

Dimension Eraser:

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 2 \\ 7 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \qquad \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \qquad \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 2 \\ 11 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

Dimension Adder:

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix} \qquad \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} 3 \\ 6 \end{bmatrix} = \begin{bmatrix} 3 \\ 6 \\ 0 \end{bmatrix} \qquad \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} 4 \\ 8 \end{bmatrix} = \begin{bmatrix} 4 \\ 8 \\ 0 \end{bmatrix}$$

Dimension Eraser with Diagonal Matrix:

$$\underset{C}{\begin{bmatrix} 4 & 0 & 0 \\ 0 & 2 & 0 \end{bmatrix}} = \underset{B}{\begin{bmatrix} 4 & 0 \\ 0 & 2 \end{bmatrix}} \underset{A}{\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}}$$

scale x-axis by 4
scale y-axis by 2

Symmetric Matrix:

We can get a symmetric matrix from a rectangular matrix taking the transpose and multiplying by itself:

$$\underset{A_{2\times3}}{\begin{bmatrix}1&2&3\\4&5&6\end{bmatrix}} \underset{A^T_{3\times2}}{\begin{bmatrix}1&4\\2&5\\3&6\end{bmatrix}} = \underset{AA^T_{2\times2}}{\begin{bmatrix}14&32\\32&14\end{bmatrix}}$$

$AA^T_{2\times2} = S_L \rightarrow$ 2 $\perp$ eigen vector in R^2

Left Singular Vectors

$$\underset{A^T_{3\times2}}{\begin{bmatrix}1&4\\2&5\\3&6\end{bmatrix}} \underset{A_{2\times3}}{\begin{bmatrix}1&2&3\\4&5&6\end{bmatrix}} = \underset{A^TA_{3\times3}}{\begin{bmatrix}17&22&27\\22&29&36\\27&36&45\end{bmatrix}}$$

$A^TA_{3\times3} = S_R \rightarrow$ 3 $\perp$ eigen vector in R^3

Right Singular Vectors

S_L & S_R are Positive Semi-Definite (PSD) Matrices.

This implies the eigen value for each eigen vector are non-negative: $\lambda_i > 0$.

Also, S_L & S_R have the same (nonzero) eigen values.

$$\lambda_1^L > \lambda_2^L \qquad\qquad \lambda_1^R > \lambda_2^R > \lambda_3^R$$

$$\lambda_1^L = \lambda_1^R$$

$$\lambda_2^L = \lambda_2^R$$

$$0 = \lambda_3^R$$

Singular values of A, that are equal to the square root of the eigenvalues of A^TA (or AA^T):

$$\sqrt{\lambda_1} = \sigma_1 \qquad\qquad \sqrt{\lambda_2} = \sigma_2$$

SVD Formula Dissection:

Rectangular Diagonal

$$A = U \Sigma V^T$$

Orthogonal Matrix

$$\underset{A_{2\times3}}{\begin{bmatrix}\cdot&\cdot&\cdot\\\cdot&\cdot&\cdot\end{bmatrix}} = \underset{U_{2\times2}}{\begin{bmatrix}|&|\\\vec{u}_1&\vec{u}_2\\|&|\end{bmatrix}} \underset{\Sigma_{2\times3}}{\begin{bmatrix}\sigma_1&0&0\\0&\sigma_2&0\end{bmatrix}} \underset{V^T_{3\times3}}{\begin{bmatrix}—&\vec{v}_1^T&—\\—&\vec{v}_2^T&—\\—&\vec{v}_3^T&—\end{bmatrix}}$$

$U_{2\times2}$: Contains the normalized eigen vector of S_L in descending order of their eigen values

$\Sigma_{2\times3}$: Same dimensions as A

$V^T_{3\times3}$: Contains the normalized eigen vector of S_R in descending order of their eigen values

This generalizes to all kinds of matrix A.

$$\begin{bmatrix} 3 & 2 & 2 \\ 2 & 3 & -2 \end{bmatrix} = \begin{bmatrix} 1/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix} \begin{bmatrix} 5 & 0 & 0 \\ 0 & 3 & 0 \end{bmatrix} \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} & 0 \\ -1/\sqrt{18} & 1/\sqrt{18} & -1/\sqrt{18} \\ 2/3 & -2/3 & -1/3 \end{bmatrix}$$

$$\begin{bmatrix} 2 & 1 \\ 3 & 1 \\ 0 & 1'8 \end{bmatrix} = \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} & 0 \\ -1/\sqrt{18} & 1/\sqrt{18} & -1/\sqrt{18} \\ 2/3 & -2/3 & -1/3 \end{bmatrix} \begin{bmatrix} 3'93 & 0 \\ 0 & 1'67 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 0'9 & 0'44 \\ -0'44 & 0'9 \end{bmatrix}$$

SVD applications:

Principal Component Analysis (PCA):

PCA is a statistical procedure that uses an orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables called principal components.

Relationship Between SVD & PCA:

PCA can be performed via SVD. Given a data matrix X, where each row represents a different observation and each column represents a different variable.

$$X = U \Sigma V^T$$

The column of V (right singular vectors) are the principal directions or axes of the data, and the singular values in Σ are related to the explained variance by each principal component. When performing PCA using SVD, the data matrix X is usually centered (and sometimes also scaled) by subtracting the mean of each variable. This ensures that PCA looks for directions of maximum variance across the dataset.

Steps to Perform PCA Using SVD:

1. Prepare the Data: Center the data by subtracting the mean of each variable. This step is crucial because PCA seeks directions that maximize the variance.
2. Compute SVD: Perform the singular value decomposition of the centered matrix X, yielding U, Σ and V^T.
3. Extract Principal Components: The columns of V (right singular vectors) are the principal components. The singular values in Σ

represent the importance of each principal component.

4. Compute Explained Variance: The square of the singular values (the diagonal of Σ^2) divided by the total variance gives the proportion of variance explained by each principal component.

5. Project Data: Project the original data onto the principal components to reduce its dimensionality. This is achieved by multiplying the original data matrix X by the first K columns of V, where K is the number of principal components to retain.

Pseudoinverse:

It's also Know as the Moore-Penrose inverse, is a generalization of the inverse matrix. While an inverse exists only for square matrices that are non-singular (i.e., matrices with a non-zero determinant), the pseudoinverse can be applied to any matrix, including non-square matrices or matrices that are not full rank.

The pseudoinverse is one way to solve Linear Least Squares problems.

Definition:

Given a matrix A, its pseudoinverse A^+ is defined such that it satisfies the following four Moore-Penrouse conditions:

1. $AA^+A=A$
2. $A^+AA^+=A^+$
3. $(AA^+)^T=AA^+$
4. $(A^+A)^T=A^+A$

These properties ensure that the pseudoinverse behaves as closely as possible to an inverse within the constraints of matrices that are not invertible in the conventional sense.

Computation using SVD:

The pseudoinverse A^+ is calculated as:

$$A^+=V\Sigma^+U^T$$

Here, Σ^+ is obtained by taking the reciprocal of the non-zero elements of Σ, placing them in a diagonal matrix of the same size, and then transposing the matrix. Zeros in Σ remain zeros in Σ^+.