Singular Value Decomposition

- It turns out that every matrix can be written as the combination of such operations.
- Any $m \times n$ real matrix M can be written as

$$M = U\Sigma V^{\top}$$

- where U is an $m \times m$ orthonormal matrix, V is an $n \times n$ orthonormal matrices, and Σ is a $m \times n$ <u>rectangle</u> <u>diagonal matrix</u>
- The values on the diagonal of Σ are <u>singular values</u>
- $\Sigma = \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & & \sigma_n \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & & 0 \end{bmatrix}$

SVD reveals the operations of any matrix. In NN, we will see a similar form called eigendecomposition.

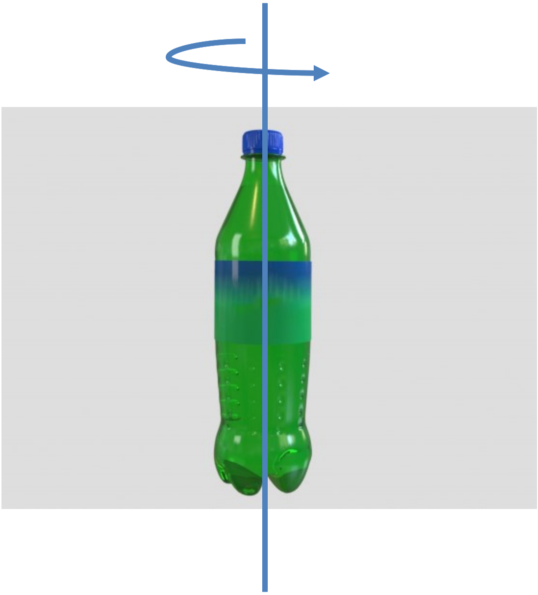
Eigenvalues and Eigenvectors

- A non-zero vector v is an eigenvector of square matrix A if
$$Av = \lambda v$$
- λ is called the eigenvalue associated with v.
- Generically speaking, Av does not change the direction of v (except when $\lambda = 0$).
- This is very special. Most vectors change direction after multiplication with a matrix A.
- Thus, v is an important characteristic of the transformation represented by A.

Examples

- The identity matrix $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ does not change anything, so its eigenvectors can be any non-zero vector and the eigenvalue is always 1.
- The matrix $\begin{bmatrix} 2 & 0 \\ 0 & 0.5 \end{bmatrix}$ stretches the first dimension and compresses the second dimension, so that any vector have non-zero values in both dimension will change direction.
- Thus, the eigenvectors are $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $\begin{bmatrix} 0 \\ 1 \end{bmatrix}$ and their multiples. The eigenvalues are 2 and 0.5 respectively.

Example

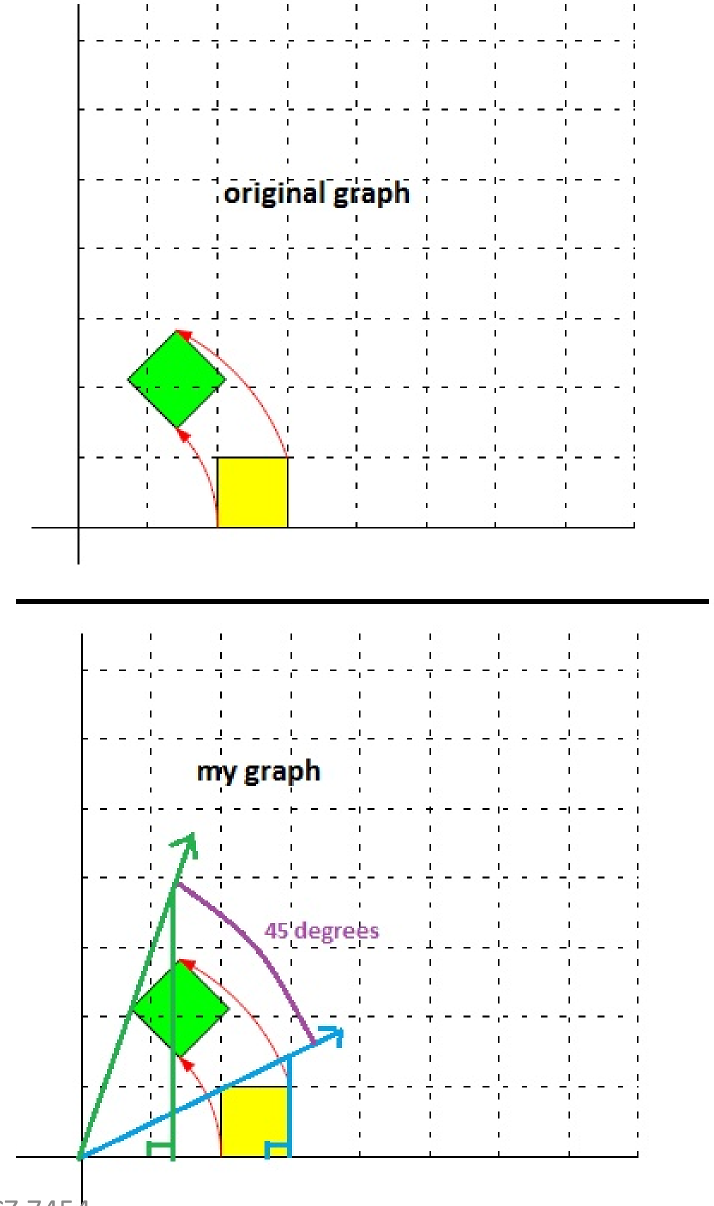


A matrix can represent rotation around an axis.
The only vector that does not change direction is the direction of the axis.

Consider matrix $\begin{bmatrix} \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} & 0 \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} & 0 \\ 0 & 0 & 1 \end{bmatrix}$, which has the eigenvector $\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$ and the eigenvalue 1. The rotation happens around the z-axis.

Other cases may not be so obvious. Rotating 30^o around the z-axis and 45^o around the y-axis yields the matrix

$\begin{bmatrix} 0.6123 & -0.356 & 0.707 \\ 0.5 & 0.866 & 0 \\ -0.6123 & 0.356 & 0.707 \end{bmatrix}$, which has the rotational axis $\begin{bmatrix} 0.22 \\ 0.82 \\ 0.53 \end{bmatrix}$



Example

- If we perform rotation along all N axes in N-dimensional space, the only direction that does not change is the 0 direction.
- However, an eigenvector must not be zero.
- In general, rotational matrices have complex eigenvectors and eigenvalues.

Eigendecomposition

- A square $n \times n$ matrix M can be written as
$$M = Q\Lambda Q^{-1}$$
- Q is a $n \times n$ matrix whose i-th column is the eigenvector q_i. These eigenvectors are usually (but not always) normalized to length 1.
- Λ is a diagonal matrix with the eigenvalues on the diagonal.
- It requires M to be diagonalizable (satisfied by most matrices we care about).

Eigendecomposition

Eigendecomposition reveals inherent characteristics of the matrix. We will see that in the optimization of neural networks.

- A square $n \times n$ matrix M can be written as

$$M = Q\Lambda Q^{-1}$$

- Derivation
- For eigenvector q, we have $Mq = \lambda q$
- For multiple eigenvectors $q_1, \dots, q_n$, we can write

$$M[q_1, \dots, q_n] = [q_1, \dots, q_n] \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix} \text{ or } MQ = Q\Lambda$$

- For a diagonalizable matrix, Q is full rank. Thus, we can multiple its inverse on both side, deriving the equation.

Eigendecomposition

- A square $n \times n$ matrix M can be written as
$$M = Q\Lambda Q^{-1}$$
- For symmetric matrices,
$$M = Q\Lambda Q^{\top}$$
- The eigenvectors are also orthogonal. Q can be normalized to become an orthogonal matrix.

Application

Find the $x \in R^n$ that minimizes $x^\top W x$, subject to $x^\top x = n$. W is a $n \times n$ real symmetric matrix (always diagonalizable).

- The eigenvectors $q_1, \ldots, q_n$ for a complete basis for R^n and are orthogonal to each other.
- For any given $x \in R^n$, we can find $y = \frac{1}{\sqrt{n}} Q^\top x$. That is, $x = \sqrt{n} Q y$
- $x^\top W x = \mathrm{n}(y^\top Q^\top W Q y) = \mathrm{n}(y^\top Q^\top Q \Lambda Q^\top Q y) = n(y^\top \Lambda y)$
- We can set y to have exactly one component equal to 1 and n-1 components being zeros.
- Which one? We will set it to correspond to the minimum eigenvalue.
- The final solution is $x = \sqrt{n} Q [0, \ldots, 1, \ldots, 0]$. The minimum value is $n\lambda_{min}$

Another way

Find the $x \in R^n$ that minimizes $x^\top W x$, subject to $x^\top x = n$. W is a $n \times n$ real symmetric matrix (always diagonalizable).

- The eigenvectors $q_1, \dots, q_n$ for a complete basis for R^n and are orthogonal to each other.
- We can write $x = \alpha_1 q_1 + \alpha_2 q_2 + \cdots + \alpha_n q_n$

$$Wx = \lambda_1 \alpha_1 q_1 + \lambda_2 \alpha_2 q_2 + \cdots + \lambda_n \alpha_n q_n$$

$$x^\top W x = \lambda_1 \alpha_1^2 q_1^2 + \lambda_2 \alpha_2^2 q_n^2 + \cdots + \lambda_n \alpha_n^2 q_n^2 \quad \text{(orthogonality between different } q\text{)}$$

$$x^\top W x = \lambda_1 \alpha_1^2 + \lambda_2 \alpha_2^2 + \cdots + \lambda_n \alpha_n^2 \quad (q\text{'s are normalized})$$

$$x^\top x = \alpha_1^2 + \alpha_2^2 + \cdots + \alpha_n^2 = n$$

- Without loss of generality, assume $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$
- We can pick α such that it is concentrated on α_1
- So the minimum value of $x^\top W x = \mathrm{n}\lambda_1$

Positive Definiteness

- A square matrix A is symmetric iff $A_{ij} = A_{ji}, \forall i, j$

$$\begin{bmatrix} 1 & 3 & 7 \\ 3 & 2 & 6 \\ 7 & 6 & 8 \end{bmatrix} \qquad \begin{bmatrix} 1 & 2 & -4 & 38 \\ 2 & -1 & 71 & 2 \\ -4 & 71 & 9 & 56 \\ 38 & 2 & 56 & 30 \end{bmatrix}$$

- A square matrix is anti-symmetric (or skew-symmetric) iff $A_{ij} = -A_{ji}, \forall i \neq j$

$$\begin{bmatrix} 1 & -3 & -7 \\ 3 & 2 & -6 \\ 7 & 6 & 8 \end{bmatrix}$$

- A matrix A is positive definite (PD) iff for all vector $x \neq 0$, $x^{\top} A x > 0$
- Recall $x^{\top} A x = \sum_i \sum_j A_{ij} x_i x_j$

$$x^{\top} \begin{bmatrix} 2 & -1 & 0 \\ -1 & 3 & -1 \\ 0 & -1 & 1 \end{bmatrix} x = x_1^2 + x_2^2 + x_3^2 + (x_1 - x_2)^2 + (x_3 - x_2)^2 > 0, \forall x \neq 0$$

- Similarly,

$$x^{\top} \begin{bmatrix} 2 & -2 & 0 \\ 0 & 3 & 0 \\ 0 & -2 & 1 \end{bmatrix} x = x_1^2 + x_2^2 + x_3^2 + (x_1 - x_2)^2 + (x_3 - x_2)^2 > 0, \forall x \neq 0$$

Definite Matrices

- A square matrix A is symmetric iff $A_{ij} = A_{ji}, \forall i, j$
- A square matrix is anti-symmetric iff $A_{ij} = -A_{ji}, \forall i \neq j$

- A matrix A is positive definite (PD) iff for all vector $x \neq 0$, $x^\top A x > 0$

Consider an anti-symmetric matrix B

$x^\top B x = \sum_i B_{ii} x_i^2 > 0$ when $B_{ii} > 0$ and $x \neq 0$

$x^\top B x = \sum_i B_{ii} x_i^2 = 0$ when $B_{ii} = 0$

Consider a matrix C that is PD but not symmetric. It can be made symmetric by shifting values between off-diagonal terms.

$$x^\top \begin{bmatrix} 2 & \boxed{-2} & 0 \\ \boxed{0} & 3 & 0 \\ 0 & -2 & 1 \end{bmatrix} x = x^\top \begin{bmatrix} 2 & -1 & 0 \\ -1 & 3 & -1 \\ 0 & -1 & 1 \end{bmatrix} x$$

That is, we can write $C = A + B$ where A is a symmetric PD matrix and B is an anti-symmetric matrix with zero diagonal.

Definite Matrices

- A square matrix A is symmetric iff $A_{ij} = A_{ji}, \forall i, j$
- A square matrix is anti-symmetric iff $A_{ij} = -A_{ji}, \forall i \neq j$

- A matrix A is positive definite (PD) iff for all vector $x \neq 0$, $x^\top A x > 0$
- Non-symmetric PD matrices can be made symmetric by adding an anti-symmetric zero-diagonal matrix to it.

- If A is PD and B is PD, $A + B$ is PD
$$x^\top (A + B)x = x^\top A x + x^\top B x > 0$$
- If A is PD and scalar α is positive, αA is PD
$$x^\top (\alpha A)x = \alpha x^\top A x > 0$$
- If A is symmetric and PD, all eigenvalues of A are positive.

Definite Matrices

- **Positive definite**: for all vector $x \neq 0$, $x^{\top}Ax > 0$
- **Positive semi-definite**: $x^{\top}Ax \geq 0$
- **Negative definite**: for all vector $x \neq 0$, $x^{\top}Ax < 0$
- **Negative semi-definite**: $x^{\top}Ax \leq 0$

- **Positive definite**: All eigenvalues > 0
- **Positive semi-definite**: All eigenvalues ≥ 0
- **Negative definite**: All eigenvalues < 0
- **Negative semi-definite**: All eigenvalues ≤ 0

Another example

$$A = \begin{bmatrix} a^2 & ab & ac \\ ab & b^2 & bc \\ ac & bc & c^2 \end{bmatrix}$$

is positive semi-definite.

- Notice that

$$A = \begin{bmatrix} a \\ b \\ c \end{bmatrix} \begin{bmatrix} a & b & c \end{bmatrix}$$

$$x^\top A x = \left(x^\top \begin{bmatrix} a \\ b \\ c \end{bmatrix} \right) ([a\ b\ c]x) = z^\top z \geq 0$$