Introduction to Data Science

Linear Algebra

Arijit Mondal

Dept. of Computer Science & Engineering

Indian Institute of Technology Patna

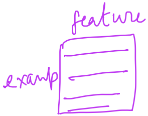
arijit@iitp.ac.in

Matrix representation

- Matrix is every where!!

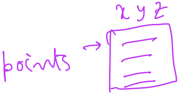
Matrix representation

- Matrix is every where!!
- Data — A data can be represented as $n \times m$ matrix,
  - A row represents an example
  - Each column represent distinct feature / dimension



Matrix representation

- Matrix is every where!!
- Data — A data can be represented as $n \times m$ matrix,
  - A row represents an example
  - Each column represent distinct feature / dimension
- Geometric point set — A $n \times m$ matrix can denote n points in m dimension space



Matrix representation

- Matrix is every where!!
- Data — A data can be represented as $n \times m$ matrix,
  - A row represents an example
  - Each column represent distinct feature / dimension
- Geometric point set — A $n \times m$ matrix can denote n points in m dimension space
- Systems of equations — Equation like $y = c_0 + c_1 x_1 + c_{m-1} x_{m-1}$ can be modeled as $n \times m$ matrix

Matrix representation

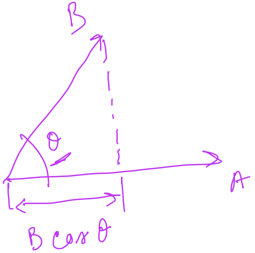
- Matrix is every where!!
- Data — A data can be represented as $n \times m$ matrix,
  - A row represents an example
  - Each column represent distinct feature / dimension
- Geometric point set — A $n \times m$ matrix can denote n points in m dimension space
- Systems of equations — Equation like $y = c_0 + c_1 x_1 + c_{m-1} x_{m-1}$ can be modeled as $n \times m$ matrix
- Graphs & Networks — City network, chemical structure, etc.

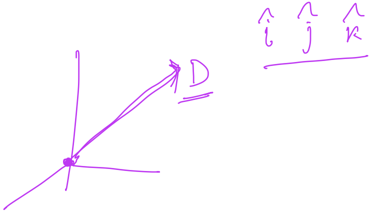
Matrix representation

- Matrix is every where!!
- Data — A data can be represented as $n \times m$ matrix,
  - A row represents an example
  - Each column represent distinct feature / dimension
- Geometric point set — A $n \times m$ matrix can denote n points in m dimension space
- Systems of equations — Equation like $y = c_0 + c_1x_1 + c_{m-1}x_{m-1}$ can be modeled as $n \times m$ matrix
- Graphs & Networks — City network, chemical structure, etc.
- Rearrangements — Permutation of given set of elements

Geometry & Vectors

- Vectors — A $1 \times d$ dimension matrix. In geometry sense a ray from the origin through the given point in d dimension
- Normalization — In many scenarios the vectors are normalized to have unit norm
- Dot Product —
  - Useful to reduce vector to scalar
  - Can be used to measure angle





Matrix operations

- Addition: $\mathrm{C} = \mathrm{A} + \mathrm{B}$, $C_{ij} = A_{ij} + B_{ij}$
- Scalar multiplication: $\mathrm{A}' = c\mathrm{A}$, $A'_{ij} = c \cdot A_{ij}$
- Linear combination: $\alpha\mathrm{A} + (1 - \alpha)\mathrm{B}$

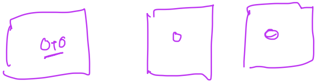



image source: Data Science Design Manual

Matrix Transpose

- Let M be a matrix and M^T be the transpose of M, then $M_{ij} = M'_{ij}$
- $(A^T)^T = A$
- Let $C = A + A^T$ hold, then $C_{ij} = A_{ij} + A_{ji} =$

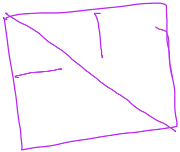


image source: Data Science Design Manual

Matrix Transpose

- Let M be a matrix and M^T be the transpose of M, then $M_{ij} = M'_{ij}$
- $(A^T)^T = A$
- Let $C = A + A^T$ hold, then $C_{ij} = A_{ij} + A_{ji} = C_{ji}$

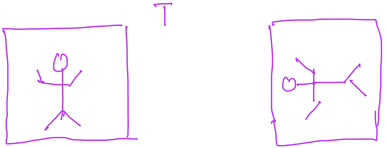


image source: Data Science Design Manual

Matrix Transpose

- Let M be a matrix and M^T be the transpose of M, then $M_{ij} = M'_{ij}$
- $(A^T)^T = A$
- Let $C = A + A^T$ hold, then $C_{ij} = A_{ij} + A_{ji} = C_{ji}$

image source: Data Science Design Manual

Matrix multiplication

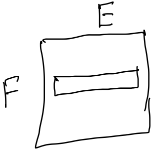
- It is an aggregated version of the vector *dot* or *inner* product
  - $x \cdot y = \sum_i x_i y_i$
- Matrix product XY^T produces 1×1 matrix which contains dot product $X \cdot Y$
- $C = AB, \ C_{ij} = \sum_k A_{ik} B_{kj}$
  - It does not commute, usually $AB \neq BA$
  - It is associative, $A(BC) = (AB)C$
  - Consider the following matrixes: $A_{1\times n}, B_{n\times n}, C_{n\times n}, D_{n\times 1}$. Which of the following is better — $(AB)(CD)$ or $(A(BC))D$?

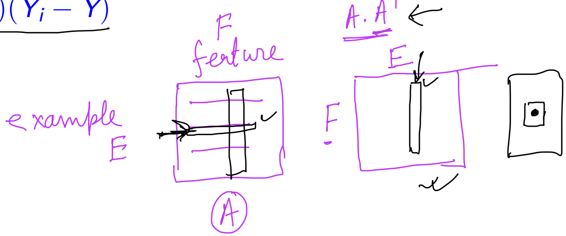
Matrix Chain multiplication

Covariance matrix

- Multiplication by transpose matrix is common ie. $A \cdot A^T$
- Both $A \cdot A^T$ and $A^T \cdot A$ are compatible for multiplications
- Let $A_{n \times d}$ be a feature matrix, each row represents an item and each column denotes a feature
- $C = AA^T$ is a $n \times n$ matrix dot products
  - C_{ij} is a measure how similar item i is to item j (in syncness)
- $D = A^T A$ is a $d \times d$ dot products in syncness among the features
  - D_{ij} represents the similarity between feature i and feature j
- Covariance formula: $Cov(X, Y) = \sum_{i=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y})$

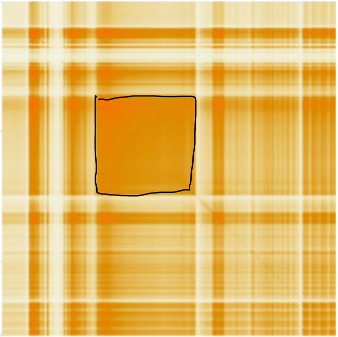
$A^T . A$





Covariance matrix (contd)

- A, $A \cdot A^T$, $A^T \cdot A$



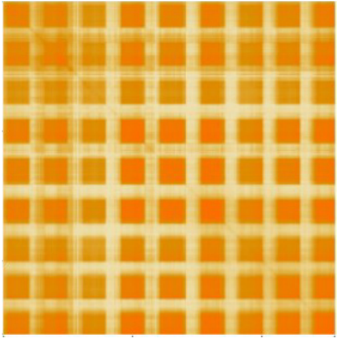


image source: Data Science Design Manual

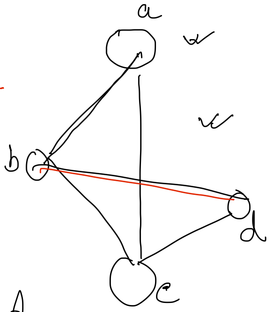
Matrix multiplication & Paths

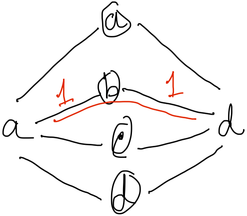
- Square matrix can be multiplied without transposition
- A matrix can represent the connectivity of nodes in a given network
- Let $A_{n \times n}$ can represent adjacency matrix

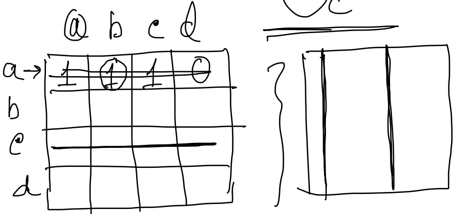
$$A^2_{ij} = \sum_{k=1}^{n} A_{ik} A_{kj}$$

$A^6 \cdot A$

Transition closure







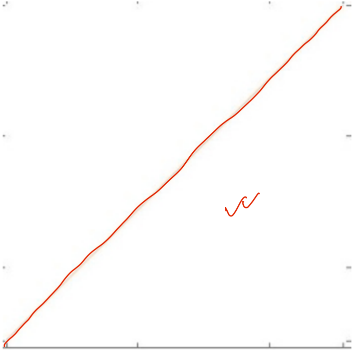
CS244

Matrix multiplication & Permutations

- Multiplication is often used to rearrange the oder of the elements in a particular matrix
- Multiplication with identity matrix (I) does not arrange anything new
- I contains exactly one non-zero element in each row and each column
- Matrix with this property is known as permutation matrix
- For example, multiplication with $P_{(2431)}$

$$P_{(2431)} = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \end{bmatrix}, \quad M = \begin{bmatrix} 11 & 12 & 13 & 14 \\ 21 & 22 & 23 & 24 \\ 31 & 32 & 33 & 34 \\ 41 & 42 & 43 & 44 \end{bmatrix}, \quad PM = \begin{bmatrix} 31 & 32 & 33 & 34 \\ 11 & 12 & 13 & 14 \\ 41 & 42 & 43 & 44 \\ 21 & 22 & 23 & 24 \end{bmatrix}$$

Permutations Example



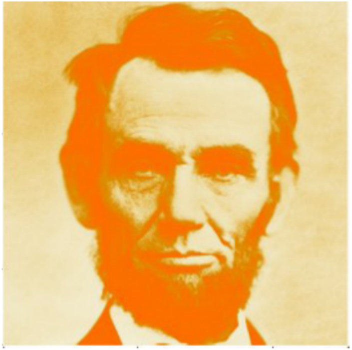
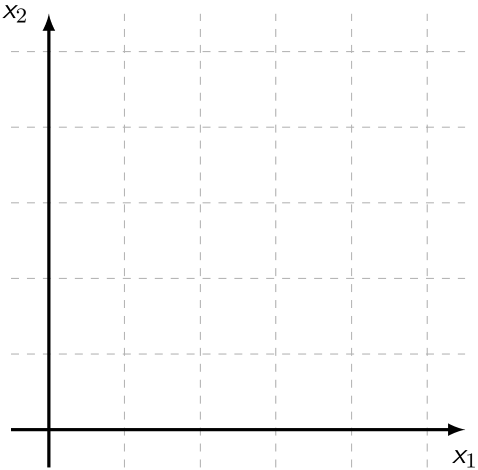


image source: Data Science Design Manual

Linear transformation

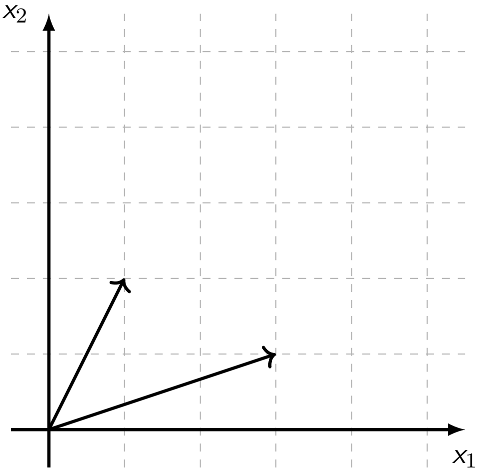


Ax

$$A = \begin{bmatrix} 1 & 3 \\ 2 & 1 \end{bmatrix} \qquad x = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$$

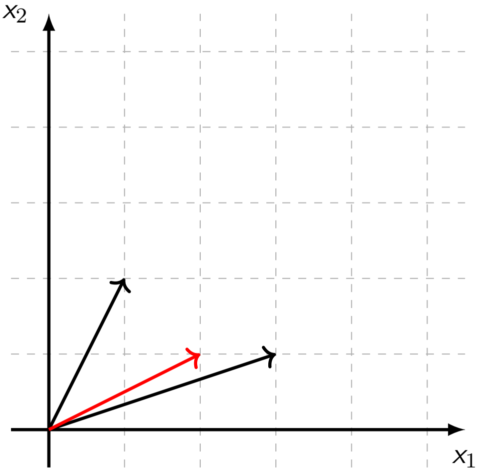
$$A = \left[\begin{bmatrix} 1 \\ 2 \end{bmatrix} \begin{bmatrix} 3 \\ 1 \end{bmatrix} \right]$$

Linear transformation



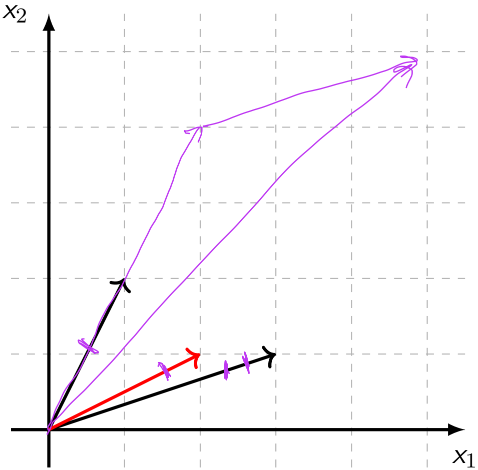
$$A = \begin{bmatrix} 1 & 3 \\ 2 & 1 \end{bmatrix} \quad x = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$$

Linear transformation



$$A = \begin{bmatrix} 1 & 3 \\ 2 & 1 \end{bmatrix} \quad x = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$$

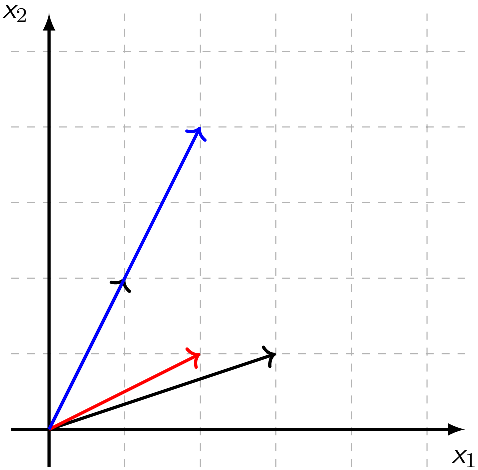
Linear transformation



$$A = \begin{bmatrix} 1 & 3 \\ 2 & 1 \end{bmatrix} \quad x = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$$

$$Ax = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \times 2 + \begin{bmatrix} 3 \\ 1 \end{bmatrix} \times 1 = \begin{bmatrix} 5 \\ 5 \end{bmatrix}$$

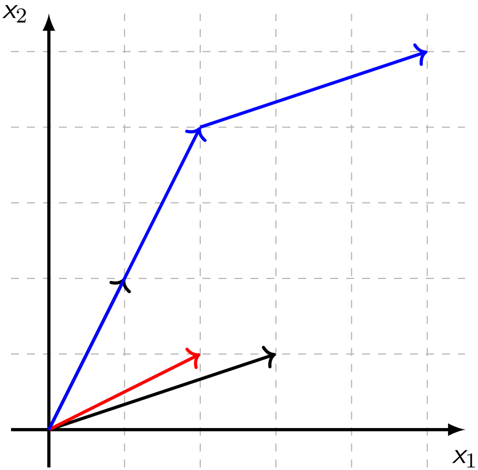
Linear transformation



$$A = \begin{bmatrix} 1 & 3 \\ 2 & 1 \end{bmatrix} \quad x = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$$

$$Ax = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \times 2 + \begin{bmatrix} 3 \\ 1 \end{bmatrix} \times 1 = \begin{bmatrix} 5 \\ 5 \end{bmatrix}$$

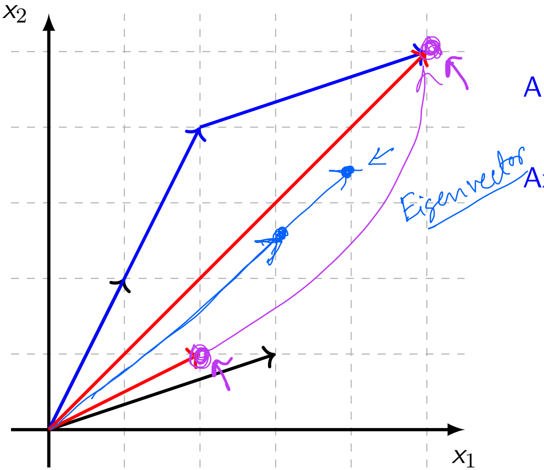
Linear transformation



$$A = \begin{bmatrix} 1 & 3 \\ 2 & 1 \end{bmatrix} \quad x = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$$

$$Ax = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \times 2 + \begin{bmatrix} 3 \\ 1 \end{bmatrix} \times 1 = \begin{bmatrix} 5 \\ 5 \end{bmatrix}$$

Linear transformation



$$A = \begin{bmatrix} 1 & 3 \\ 2 & 1 \end{bmatrix} \quad x = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$$

$$Ax = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \times 2 + \begin{bmatrix} 3 \\ 1 \end{bmatrix} \times 1 = \begin{bmatrix} 5 \\ 5 \end{bmatrix}$$

Ax

Rotating points in space

- Multiplying with the right matrix can rotate a set of points about the origin by angle θ

$$R_\theta = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix}$$

- $\begin{bmatrix} x' \\ y' \end{bmatrix} = R_\theta \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} x\cos(\theta) & -y\sin(\theta) \\ x\sin(\theta) & y\cos(\theta) \end{bmatrix}$

Identity matrix

- Identity plays a big role in algebraic structure

Addition → 0

Multiplication - 1

Identity matrix

- Identity plays a big role in algebraic structure
  - 0 is the identity element for addition operation

Identity matrix

- Identity plays a big role in algebraic structure
  - 0 is the identity element for addition operation
  - 1 is the identity element for multiplication operation

Identity matrix

- Identity plays a big role in algebraic structure
  - 0 is the identity element for addition operation
  - 1 is the identity element for multiplication operation
- Inverse operation is about taking an element x down to its identity
  - For addition operation, inverse of x is $-x$

CS244

Identity matrix

- Identity plays a big role in algebraic structure
  - 0 is the identity element for addition operation
  - 1 is the identity element for multiplication operation
- Inverse operation is about taking an element x down to its identity
  - For addition operation, inverse of x is $-x$
  - For multiplication operation, inverse of x is $\frac{1}{x}$

Identity matrix

- Identity plays a big role in algebraic structure
  - 0 is the identity element for addition operation
  - 1 is the identity element for multiplication operation
- Inverse operation is about taking an element x down to its identity
  - For addition operation, inverse of x is $-x$
  - For multiplication operation, inverse of x is $\frac{1}{x}$
- For matrix, we say A^{-1} is multiplicative inverse if $A \cdot A^{-1} = I$

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

- for 2×2 matrix, $A^{-1} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{ad - bc} \begin{vmatrix} d & -b \\ c & a \end{vmatrix}$

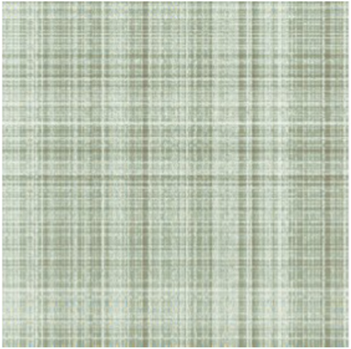
Identity matrix

- Identity plays a big role in algebraic structure
  - 0 is the identity element for addition operation
  - 1 is the identity element for multiplication operation
- Inverse operation is about taking an element x down to its identity
  - For addition operation, inverse of x is $-x$
  - For multiplication operation, inverse of x is $\frac{1}{x}$
- For matrix, we say A^{-1} is multiplicative inverse if $A \cdot A^{-1} = I$
- for 2×2 matrix, $A^{-1} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{ad - bc} \begin{vmatrix} d & -b \\ c & a \end{vmatrix}$
- Matrix that is not invertible is known as *singular* matrix
- Gaussian elimination can be used to find the inverse

CS244

Inversion Example

- Inverse of Lincoln image and $M \cdot M^{-1}$



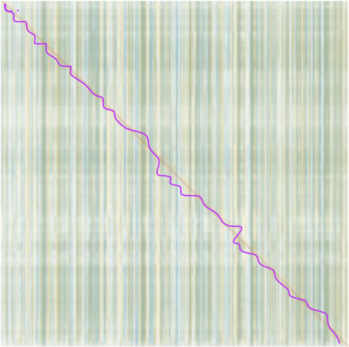
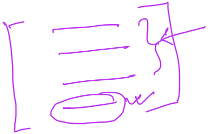


image source: Data Science Design Manual

Linear Systems, Matrix Rank

- Linear systems
  - Consider the following linear equation: $y = c_0 + c_1 x_1 + \cdots + c_{m-1} x_{m-1}$
  - Thus the coefficient of n such linear equations can be represented as a matrix C of size $n \times m$
    $CX = Y \Rightarrow X = C^{-1}Y$
  - What will happen if inverse does not exist?
- Matrix Rank
  - A rank of a matrix measures the number of linearly independent rows
  - Rank can be determined using Gaussian elimination



Factoring matrices

- Factoring matrix A into matrices in B and C represents particular aspect of division
  - Non-singular matrix has an inverse $I = M \cdot M^{-1}$
- Matrix factorization is an important abstraction in data science, leading to feature representation in a compact way
- Suppose matrix A can be factored as $A \approx BC$ where the size of A is $n \times m$, B — $n \times k$, C — $k \times m$ where $k < \min(n, m)$

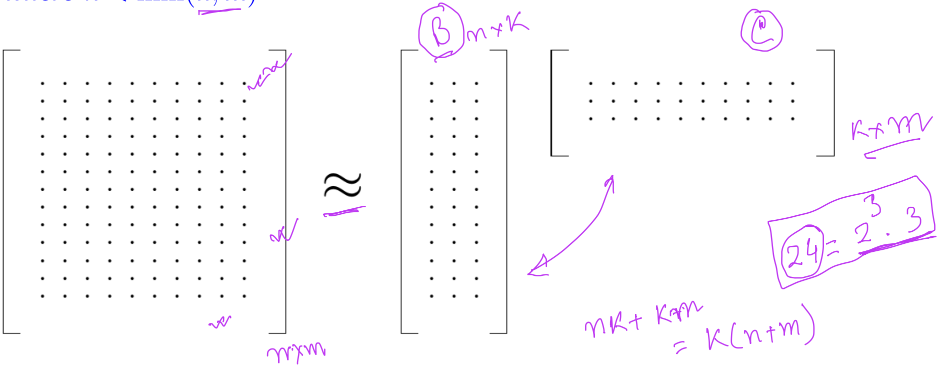


image source: Data Science Design Manual

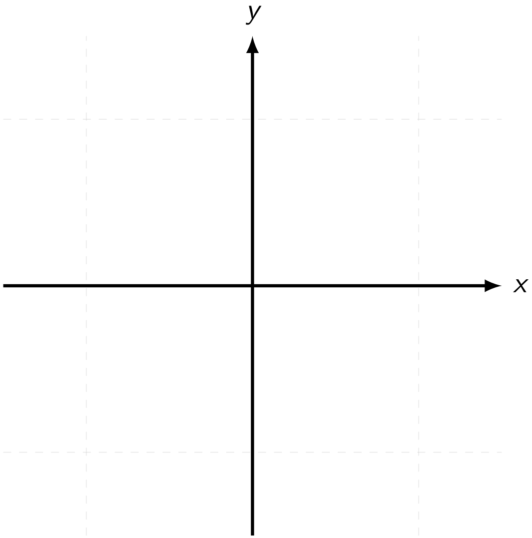
Eigenvalues & Eigenvectors

- Multiplying a vector U by a matrix A can have the same effect as multiplying it by scalar λ

$$\begin{bmatrix} -5 & 2 \\ 2 & -2 \end{bmatrix} \cdot \begin{bmatrix} 2 \\ -1 \end{bmatrix} = -6 \begin{bmatrix} 2 \\ -1 \end{bmatrix}, \begin{bmatrix} -5 & 2 \\ 2 & -2 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 2 \end{bmatrix} = -1 \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

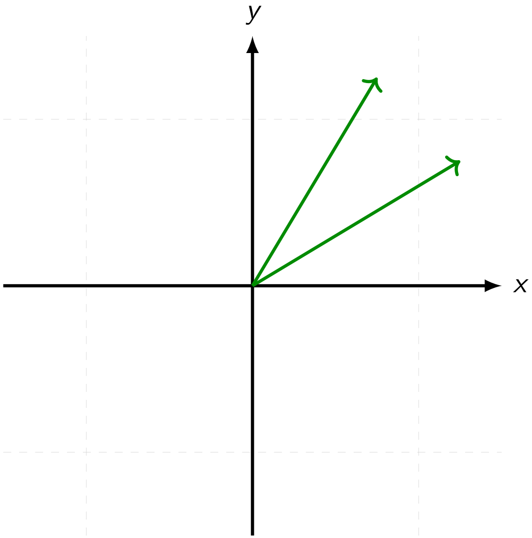
- λ is eigenvalue, U is eigen vector
- Together, the eigenvector and eigenvalue must encode a lot of information about the matrix A
- Properties
  - Each eigenvalue has an associated eigenvector
  - There are in general n eigenvector-eigenvalue pairs for every full rank $n \times n$ matrix
  - Every pair of eigenvectors of symmetric matrix are mutually orthogonal
    - Two vectors are orthogonal if the dot product is 0
  - The eigenvectors can play the role of dimensions or bases in some n dimensional space

Example



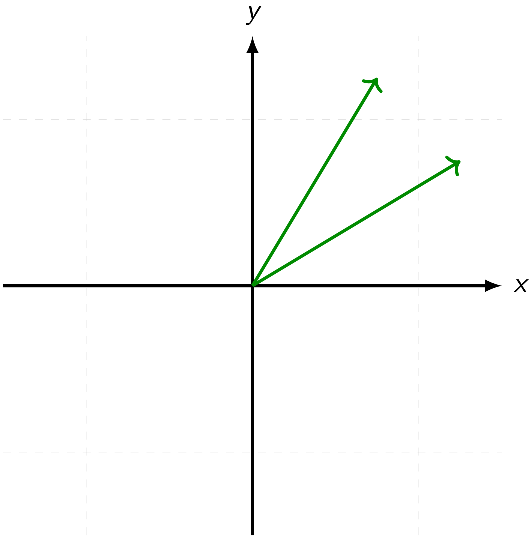
$$A = \begin{bmatrix} 1.25 & 0.75 \\ 0.75 & 1.25 \end{bmatrix}$$

Example



$$A = \begin{bmatrix} 1.25 & 0.75 \\ 0.75 & 1.25 \end{bmatrix}$$

Example

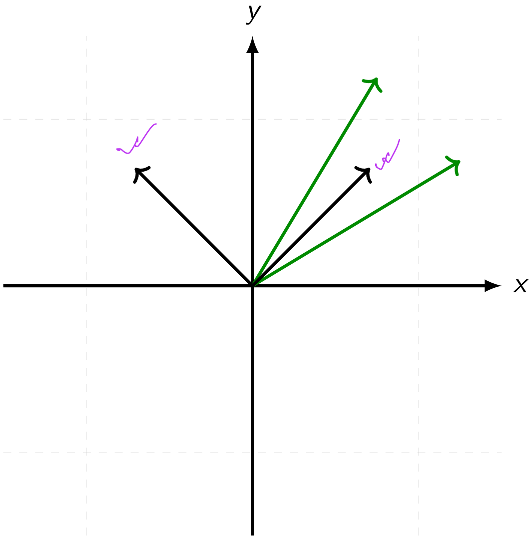


$$A = \begin{bmatrix} 1.25 & 0.75 \\ 0.75 & 1.25 \end{bmatrix}$$

$$v_1 = \begin{bmatrix} 0.707 \\ 0.707 \end{bmatrix}, \lambda_1 = 2.0$$

$$v_2 = \begin{bmatrix} -0.707 \\ 0.707 \end{bmatrix}, \lambda_2 = 0.5$$

Example



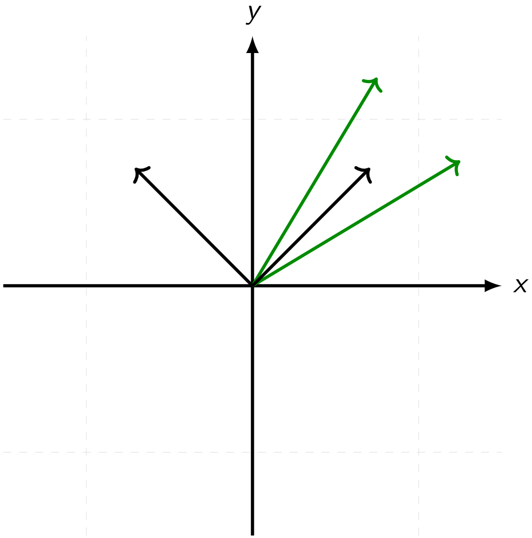
$$\mathsf{A} = \begin{bmatrix} 1.25 & 0.75 \\ 0.75 & 1.25 \end{bmatrix}$$

$$\mathsf{v}_1 = \begin{bmatrix} 0.707 \\ 0.707 \end{bmatrix}, \lambda_1 = 2.0$$

$$\mathsf{v}_2 = \begin{bmatrix} -0.707 \\ 0.707 \end{bmatrix}, \lambda_2 = 0.5$$

CS244

Example



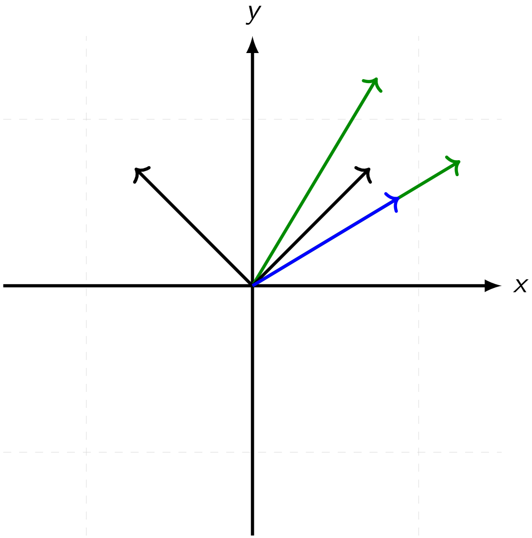
$$A = \begin{bmatrix} 1.25 & 0.75 \\ 0.75 & 1.25 \end{bmatrix}$$

$$v_1 = \begin{bmatrix} 0.707 \\ 0.707 \end{bmatrix}, \lambda_1 = 2.0$$

$$v_2 = \begin{bmatrix} -0.707 \\ 0.707 \end{bmatrix}, \lambda_2 = 0.5$$

$$Av_1 = \begin{bmatrix} 1.414 \\ 1.414 \end{bmatrix}$$

Example



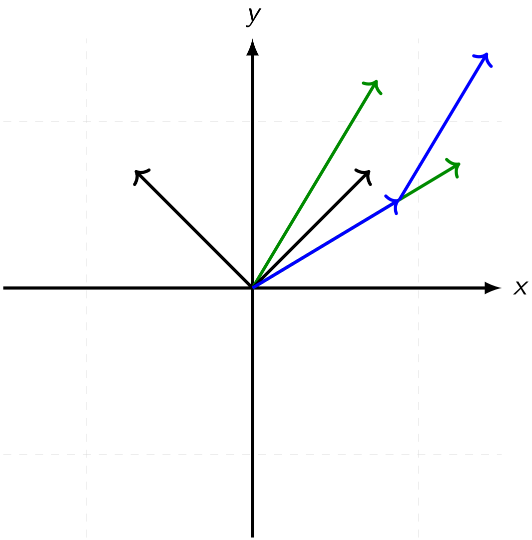
$$A = \begin{bmatrix} 1.25 & 0.75 \\ 0.75 & 1.25 \end{bmatrix}$$

$$v_1 = \begin{bmatrix} 0.707 \\ 0.707 \end{bmatrix}, \lambda_1 = 2.0$$

$$v_2 = \begin{bmatrix} -0.707 \\ 0.707 \end{bmatrix}, \lambda_2 = 0.5$$

$$Av_1 = \begin{bmatrix} 1.414 \\ 1.414 \end{bmatrix}$$

Example



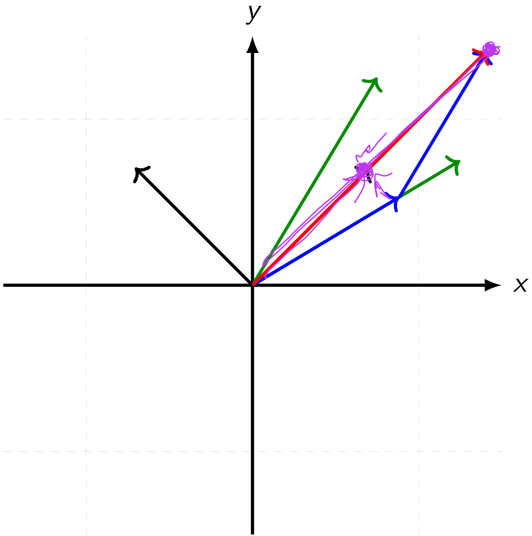
$$A = \begin{bmatrix} 1.25 & 0.75 \\ 0.75 & 1.25 \end{bmatrix}$$

$$v_1 = \begin{bmatrix} 0.707 \\ 0.707 \end{bmatrix}, \lambda_1 = 2.0$$

$$v_2 = \begin{bmatrix} -0.707 \\ 0.707 \end{bmatrix}, \lambda_2 = 0.5$$

$$Av_1 = \begin{bmatrix} 1.414 \\ 1.414 \end{bmatrix}$$

Example



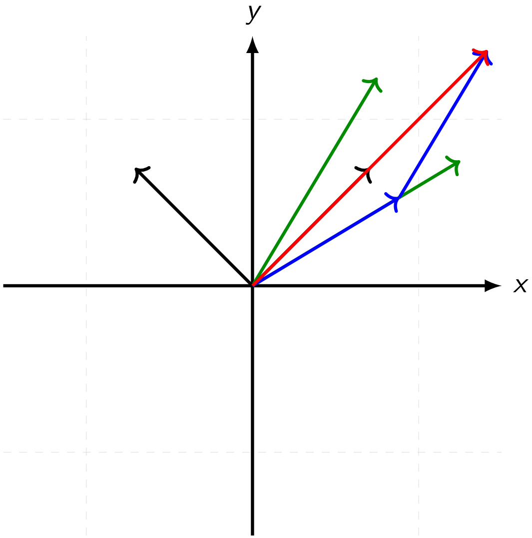
$$A = \begin{bmatrix} 1.25 & 0.75 \\ 0.75 & 1.25 \end{bmatrix}$$

$$v_1 = \begin{bmatrix} 0.707 \\ 0.707 \end{bmatrix}, \lambda_1 = 2.0$$

$$v_2 = \begin{bmatrix} -0.707 \\ 0.707 \end{bmatrix}, \lambda_2 = 0.5$$

$$Av_1 = \begin{bmatrix} 1.414 \\ 1.414 \end{bmatrix}$$

Example



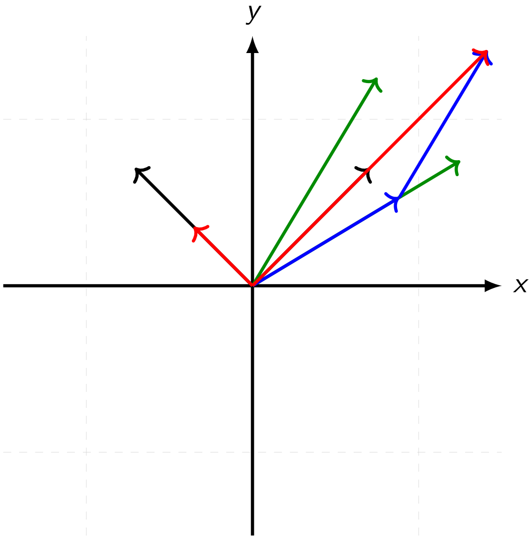
$$A = \begin{bmatrix} 1.25 & 0.75 \\ 0.75 & 1.25 \end{bmatrix}$$

$$v_1 = \begin{bmatrix} 0.707 \\ 0.707 \end{bmatrix}, \lambda_1 = 2.0$$

$$v_2 = \begin{bmatrix} -0.707 \\ 0.707 \end{bmatrix}, \lambda_2 = 0.5$$

$$Av_2 = \begin{bmatrix} -0.354 \\ 0.354 \end{bmatrix}$$

Example



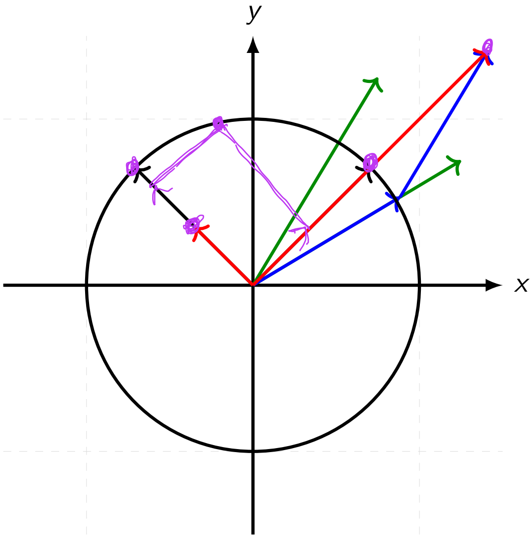
$$A = \begin{bmatrix} 1.25 & 0.75 \\ 0.75 & 1.25 \end{bmatrix}$$

$$v_1 = \begin{bmatrix} 0.707 \\ 0.707 \end{bmatrix}, \lambda_1 = 2.0$$

$$v_2 = \begin{bmatrix} -0.707 \\ 0.707 \end{bmatrix}, \lambda_2 = 0.5$$

$$Av_2 = \begin{bmatrix} -0.354 \\ 0.354 \end{bmatrix}$$

Example



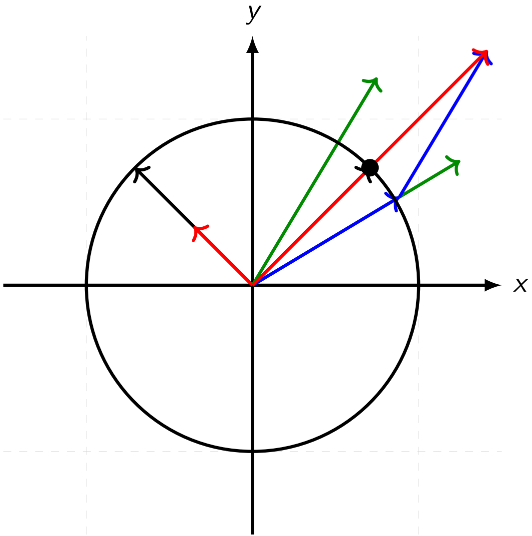
$$A = \begin{bmatrix} 1.25 & 0.75 \\ 0.75 & 1.25 \end{bmatrix}$$

$$v_1 = \begin{bmatrix} 0.707 \\ 0.707 \end{bmatrix}, \lambda_1 = 2.0$$

$$v_2 = \begin{bmatrix} -0.707 \\ 0.707 \end{bmatrix}, \lambda_2 = 0.5$$

Given $\|x\| = 1$, find Ax

Example



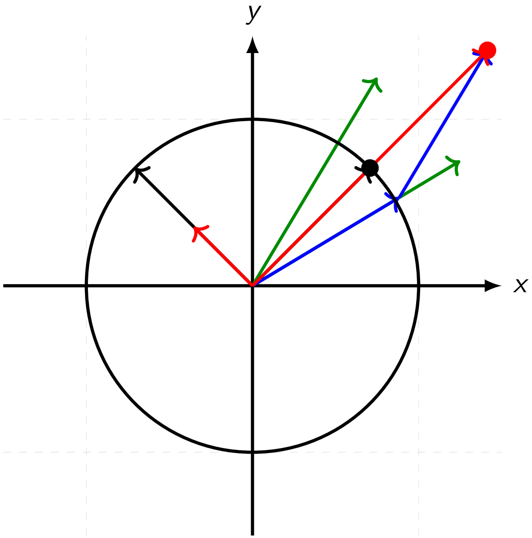
$$A = \begin{bmatrix} 1.25 & 0.75 \\ 0.75 & 1.25 \end{bmatrix}$$

$$v_1 = \begin{bmatrix} 0.707 \\ 0.707 \end{bmatrix}, \lambda_1 = 2.0$$

$$v_2 = \begin{bmatrix} -0.707 \\ 0.707 \end{bmatrix}, \lambda_2 = 0.5$$

Given $\|x\| = 1$, find Ax

Example



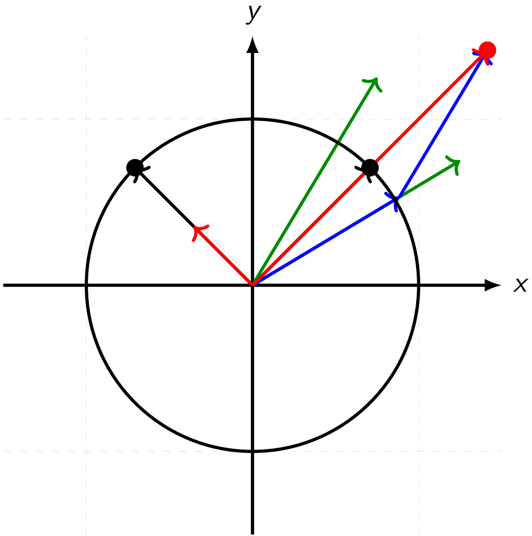
$$A = \begin{bmatrix} 1.25 & 0.75 \\ 0.75 & 1.25 \end{bmatrix}$$

$$v_1 = \begin{bmatrix} 0.707 \\ 0.707 \end{bmatrix}, \lambda_1 = 2.0$$

$$v_2 = \begin{bmatrix} -0.707 \\ 0.707 \end{bmatrix}, \lambda_2 = 0.5$$

Given $\|x\| = 1$, find Ax

Example



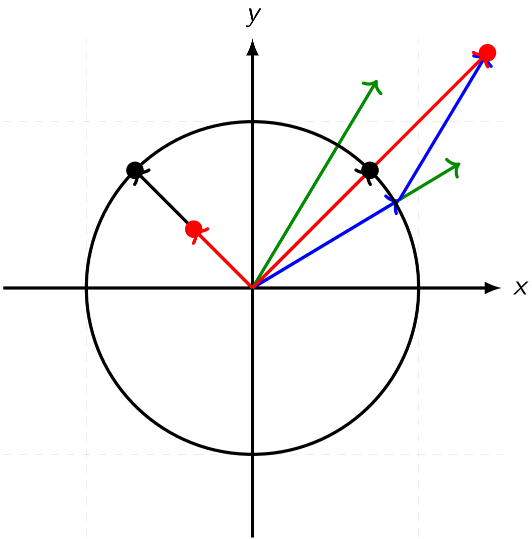
$$A = \begin{bmatrix} 1.25 & 0.75 \\ 0.75 & 1.25 \end{bmatrix}$$

$$v_1 = \begin{bmatrix} 0.707 \\ 0.707 \end{bmatrix}, \lambda_1 = 2.0$$

$$v_2 = \begin{bmatrix} -0.707 \\ 0.707 \end{bmatrix}, \lambda_2 = 0.5$$

Given $\|x\| = 1$, find Ax

Example



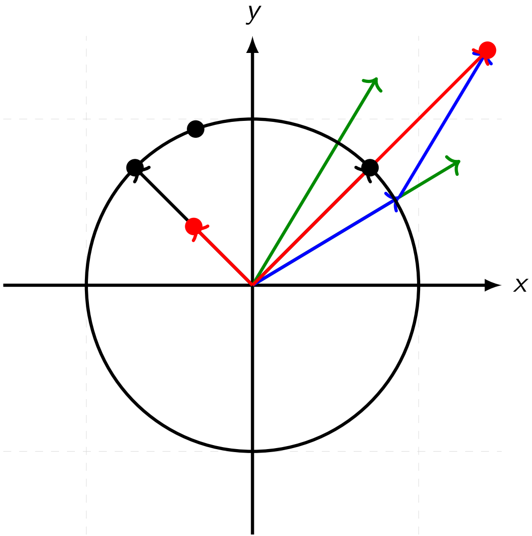
$$A = \begin{bmatrix} 1.25 & 0.75 \\ 0.75 & 1.25 \end{bmatrix}$$

$$v_1 = \begin{bmatrix} 0.707 \\ 0.707 \end{bmatrix}, \lambda_1 = 2.0$$

$$v_2 = \begin{bmatrix} -0.707 \\ 0.707 \end{bmatrix}, \lambda_2 = 0.5$$

Given $\|x\| = 1$, find Ax

Example



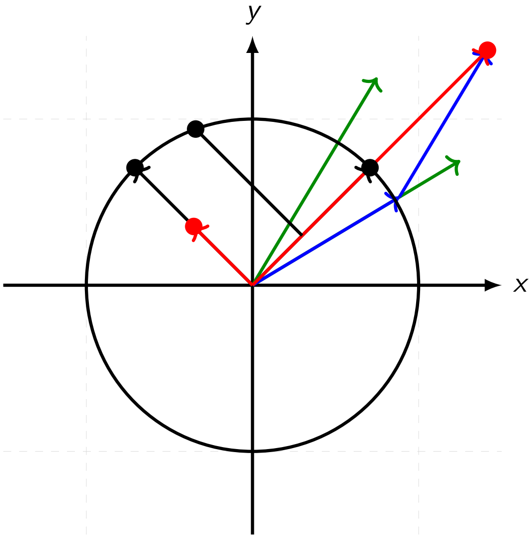
$$A = \begin{bmatrix} 1.25 & 0.75 \\ 0.75 & 1.25 \end{bmatrix}$$

$$v_1 = \begin{bmatrix} 0.707 \\ 0.707 \end{bmatrix}, \lambda_1 = 2.0$$

$$v_2 = \begin{bmatrix} -0.707 \\ 0.707 \end{bmatrix}, \lambda_2 = 0.5$$

Given $\|x\| = 1$, find Ax

Example



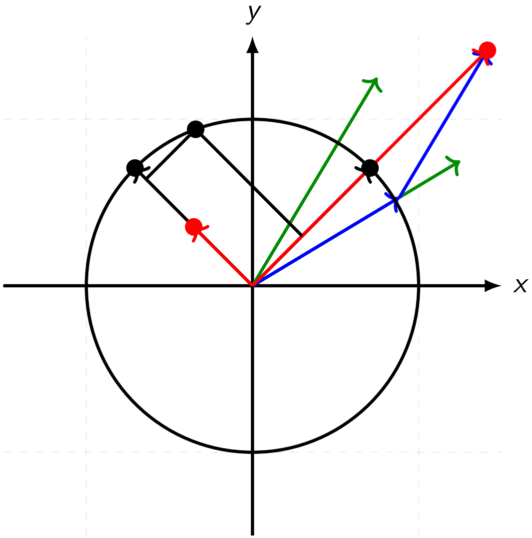
$$A = \begin{bmatrix} 1.25 & 0.75 \\ 0.75 & 1.25 \end{bmatrix}$$

$$v_1 = \begin{bmatrix} 0.707 \\ 0.707 \end{bmatrix}, \lambda_1 = 2.0$$

$$v_2 = \begin{bmatrix} -0.707 \\ 0.707 \end{bmatrix}, \lambda_2 = 0.5$$

Given $\|x\| = 1$, find Ax

Example



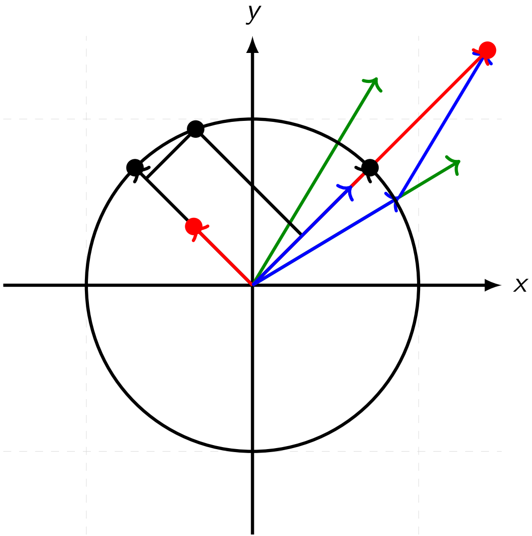
$$\mathsf{A} = \begin{bmatrix} 1.25 & 0.75 \\ 0.75 & 1.25 \end{bmatrix}$$

$$\mathsf{v}_1 = \begin{bmatrix} 0.707 \\ 0.707 \end{bmatrix}, \lambda_1 = 2.0$$

$$\mathsf{v}_2 = \begin{bmatrix} -0.707 \\ 0.707 \end{bmatrix}, \lambda_2 = 0.5$$

Given $\|\mathsf{x}\| = 1$, find Ax

Example



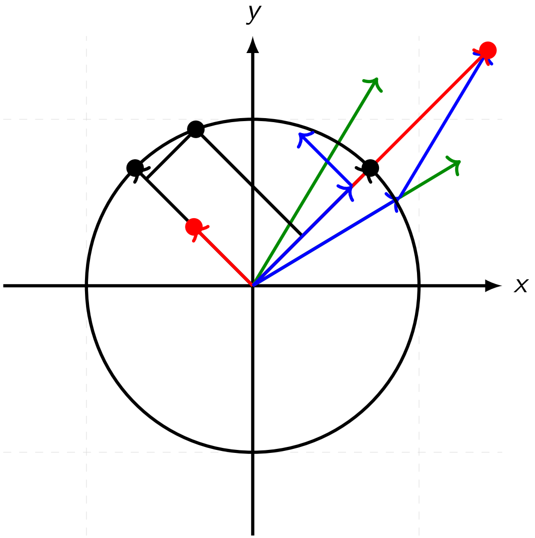
$$A = \begin{bmatrix} 1.25 & 0.75 \\ 0.75 & 1.25 \end{bmatrix}$$

$$v_1 = \begin{bmatrix} 0.707 \\ 0.707 \end{bmatrix}, \lambda_1 = 2.0$$

$$v_2 = \begin{bmatrix} -0.707 \\ 0.707 \end{bmatrix}, \lambda_2 = 0.5$$

Given $\|x\| = 1$, find Ax

Example



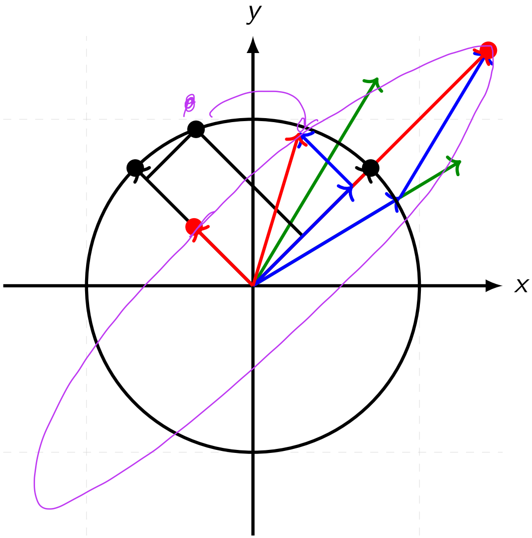
$$A = \begin{bmatrix} 1.25 & 0.75 \\ 0.75 & 1.25 \end{bmatrix}$$

$$v_1 = \begin{bmatrix} 0.707 \\ 0.707 \end{bmatrix}, \lambda_1 = 2.0$$

$$v_2 = \begin{bmatrix} -0.707 \\ 0.707 \end{bmatrix}, \lambda_2 = 0.5$$

Given $\|x\| = 1$, find Ax

Example



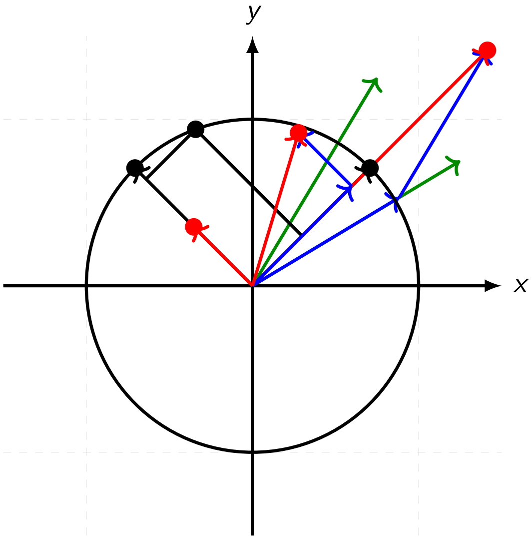
$$A = \begin{bmatrix} 1.25 & 0.75 \\ 0.75 & 1.25 \end{bmatrix}$$

$$v_1 = \begin{bmatrix} 0.707 \\ 0.707 \end{bmatrix}, \lambda_1 = 2.0$$

$$v_2 = \begin{bmatrix} -0.707 \\ 0.707 \end{bmatrix}, \lambda_2 = 0.5$$

Given $\|x\| = 1$, find Ax

Example



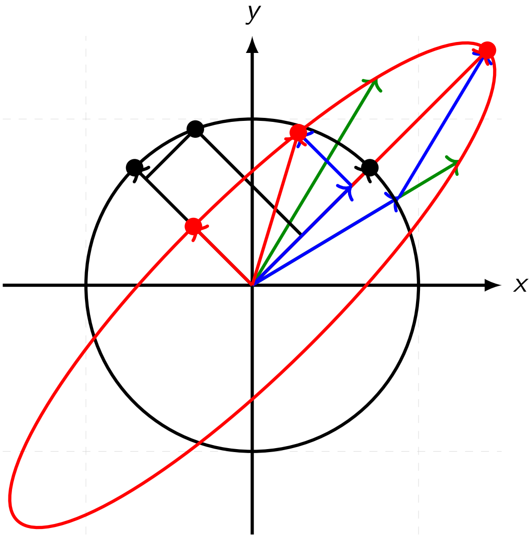
$$A = \begin{bmatrix} 1.25 & 0.75 \\ 0.75 & 1.25 \end{bmatrix}$$

$$v_1 = \begin{bmatrix} 0.707 \\ 0.707 \end{bmatrix}, \lambda_1 = 2.0$$

$$v_2 = \begin{bmatrix} -0.707 \\ 0.707 \end{bmatrix}, \lambda_2 = 0.5$$

Given $\|x\| = 1$, find Ax

Example



$$A = \begin{bmatrix} 1.25 & 0.75 \\ 0.75 & 1.25 \end{bmatrix}$$

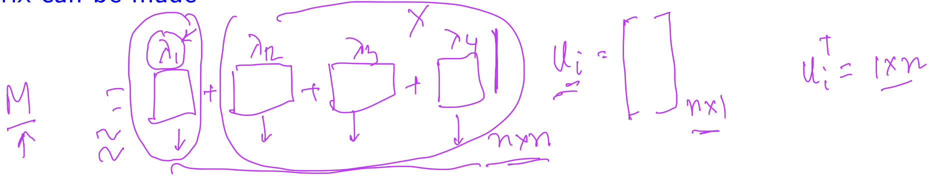
$$v_1 = \begin{bmatrix} 0.707 \\ 0.707 \end{bmatrix}, \lambda_1 = 2.0$$

$$v_2 = \begin{bmatrix} -0.707 \\ 0.707 \end{bmatrix}, \lambda_2 = 0.5$$

Given $\|x\| = 1$, find Ax

Eigenvalue decomposition

- Any $n \times n$ symmetric matrix M can be decomposed into the sum of its n eigenvector products
- Let (λ_i, U_i) be the eigen pairs $i = 1, \ldots, n$ and assume $\lambda_i \geq \lambda_{i+1}$
- Each eigenvector U_i is an $n \times 1$ matrix, multiplying it by its transpose yields an $n \times n$ matrix, product $U_i U_i^T$ same dimension as M
- Linear combination of these matrices weighted by its corresponding eigenvalue gives the original matrix $M = \sum_{i=1}^{n} \lambda_i U_i U_i^T$
  - It holds for symmetric matrices
  - Can be applied on covariance matrix
- Using only the vector associated with the largest eigenvalues, a good approximation of the matrix can be made



Example

- Covariance of Lincoln & $M - U_1 U_1^T$

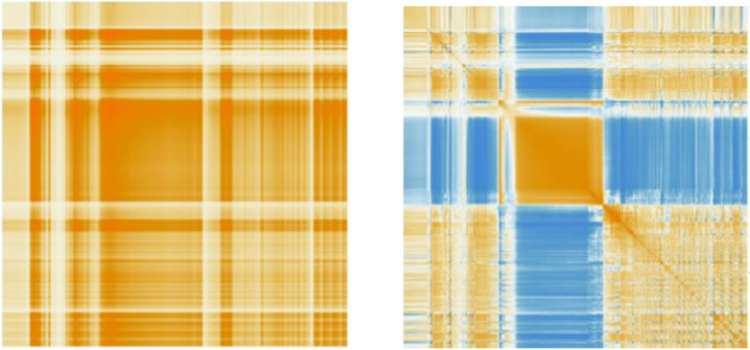


image source: Data Science Design Manual

Error plot

- Reconstructing the Lincoln memorial from the one, five, fifty eigenvectors

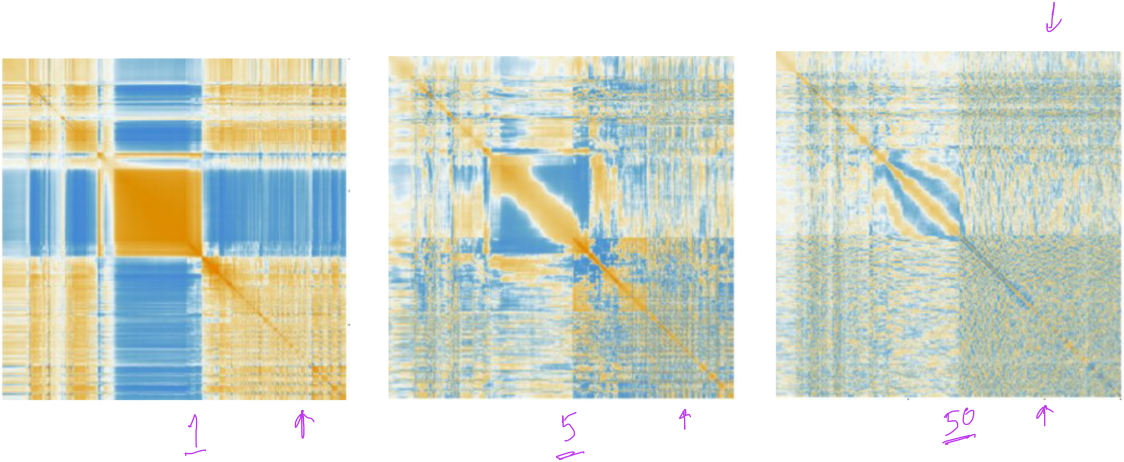


image source: Data Science Design Manual

Singular Value Decomposition

- Eigen value decomposition is good but works for symmetric matrix
- Singular value decomposition is more general matrix factorization approach
- SVD of an $n \times m$ matrix M factors into three matrices $U_{n\times n}, D_{n\times m}, V_{m\times m}$ ie. $M = UDV^T$, D is a diagonal matrix
- The product $U \cdot D$ has the effect of multiplying U_{ij} by D_{ii} D_{jj}
  - Relative importance of each column of U is provided by D
- DV^T provides relative importance of each row of V^T
- The weight of D are known as singular values of M

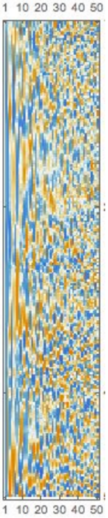
CS244

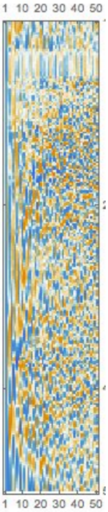
Singular Value Decomposition

- Let X and Y be vectors of size $n \times 1$ and $1 \times m$, then matrix outer product $P = X \otimes Y$ is $n \times m$ matrix, $P_{jk} = X_j Y_K$
- Traditional matrix multiplication can be expressed as $C = A \cdot B = \sum_k A_k \otimes B_k^T$
  - A_k — kth column of A, B_k^T — kth row of B
- M can be expressed as the sum of outer product of vectors resulting from SVD namely $(UD)_k$, V_k^T

Example

- Vectors associated with first 50 singular values, MSE of reconstruction





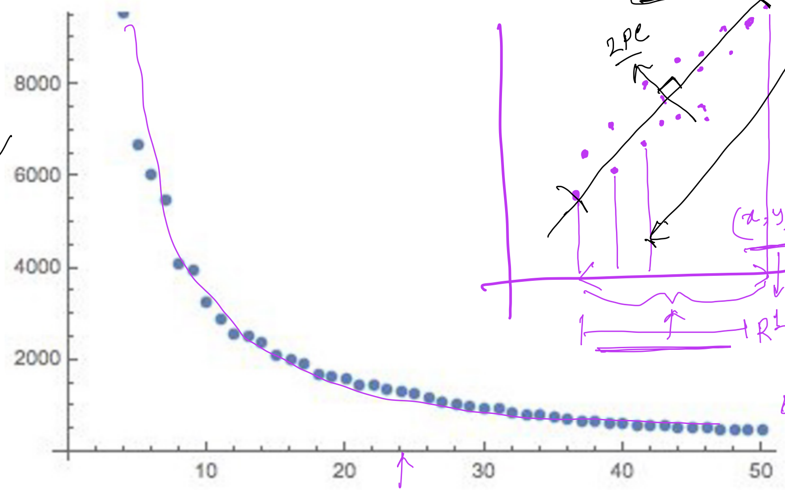


image source: Data Science Design Manual

Example

- Reconstruction with $k = 5, 50$, and error for $k = 50$

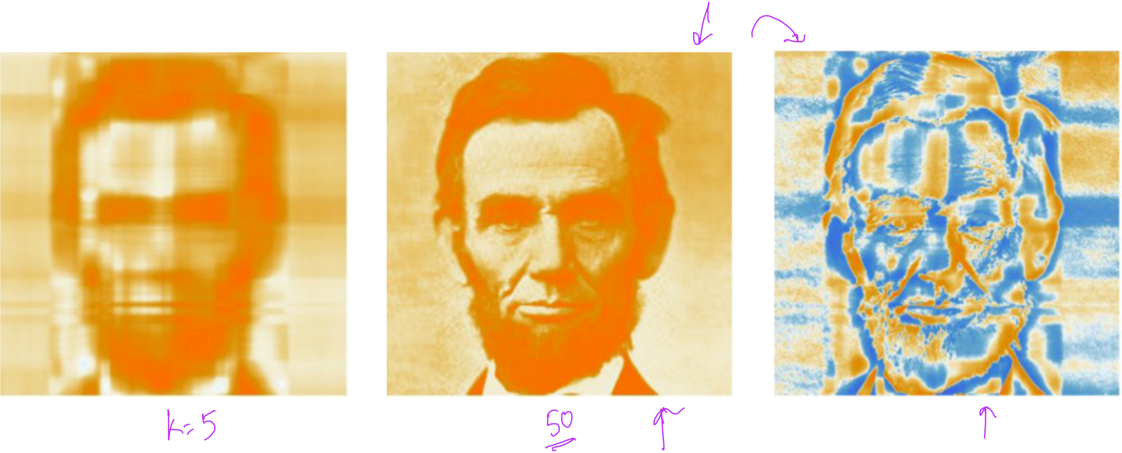


image source: Data Science Design Manual