Linear Algebra Basics

Nakul Gopalan
Georgia Tech

These slides are based on slides from Le Song and Andres Mendez-Vazquez, Chao Zhang, Mahdi Roozbahani, Rodrigo Valente

Some logistics

- Creating team.
- Office hours are started from next week.
- First quiz out this Thursday.
- First assignment out this Thursday (early release).

Outline

- Linear Algebra Basics
- Norms
- Multiplications
- Matrix Inversion
- Trace and Determinant
- Eigen Values and Eigen Vectors
- Singular Value Decomposition
- Matrix Calculus

Why Linear Algebra?

- Linear algebra provides a way of compactly representing and operating on sets of linear equations

  $4x_1 - 5x_2 = -13 \qquad -2x_1 + 3x_2 = 9$

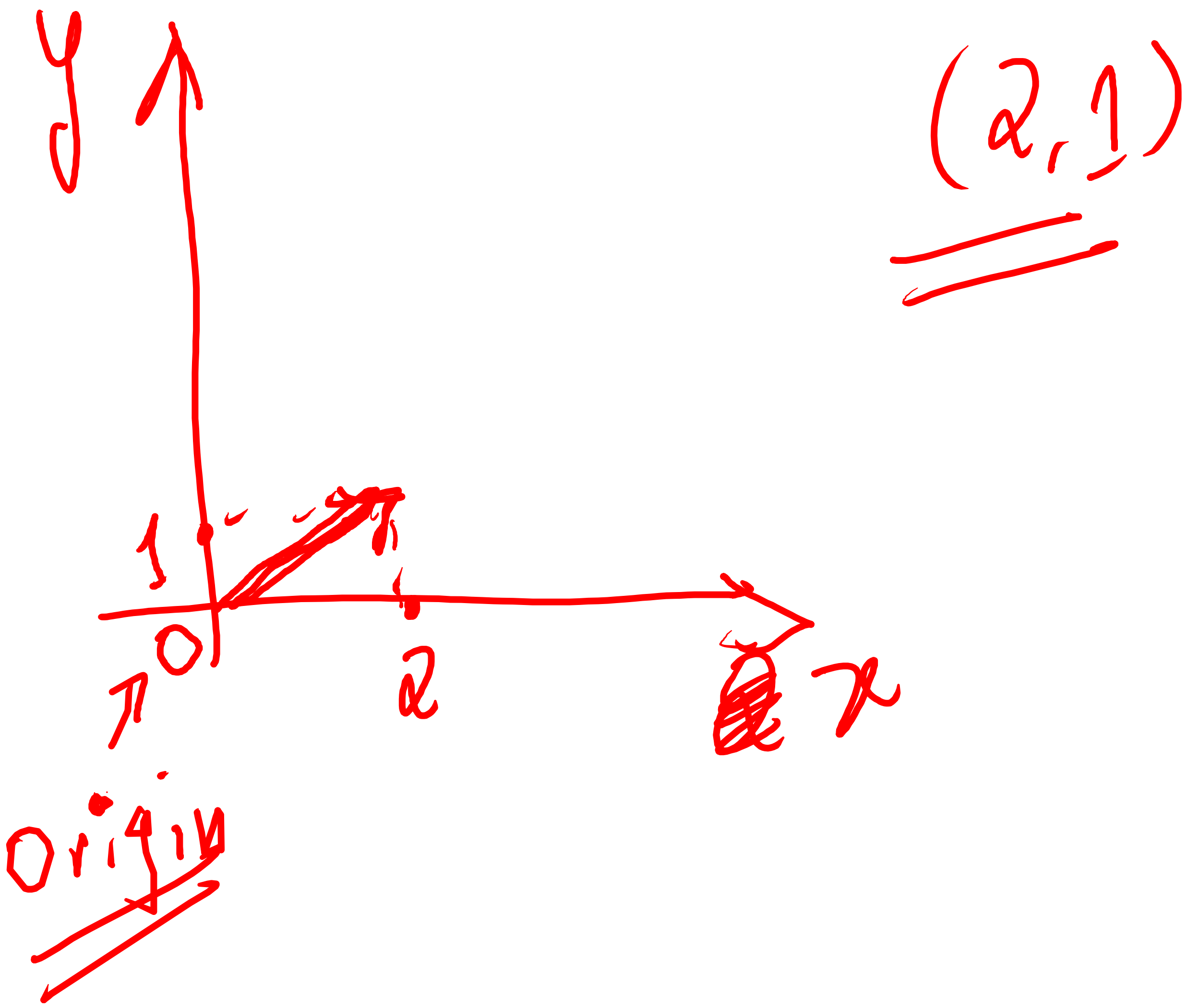
  can be written in the form of $Ax = b$

  $$A = \begin{bmatrix} 4 & -5 \\ -2 & 3 \end{bmatrix} \quad b = \begin{bmatrix} -13 \\ 9 \end{bmatrix}$$

- $A \in \mathbb{R}^{n \times d}$ denotes a matrix with n rows and d columns, where elements belong to real numbers.

- $x \in \mathbb{R}^{d}$ denotes a vector with d real entries. By convention an d dimensional vector is often thought as a matrix with 1 row and d column.

Example: Points



Example: Lines

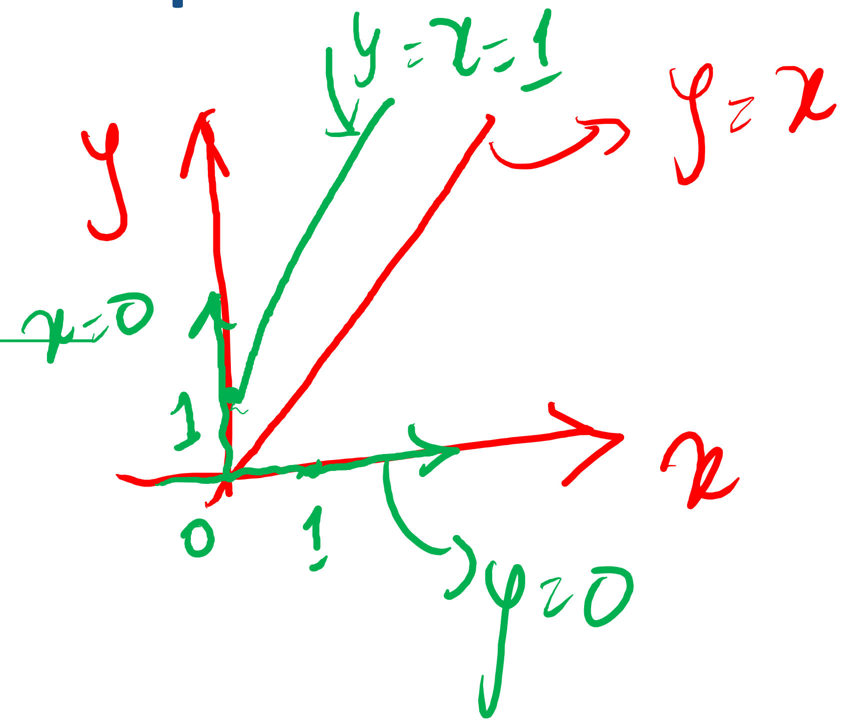
y = mx + c!

y = x

y = x + 1

y = 0

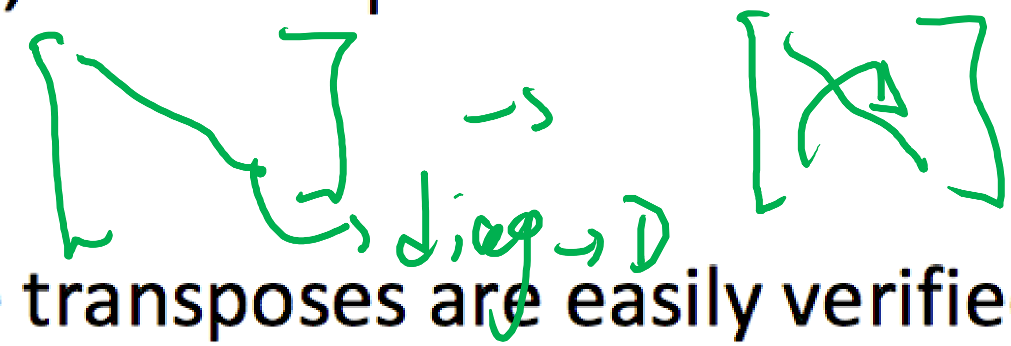
x = 0



Linear Algebra Basics

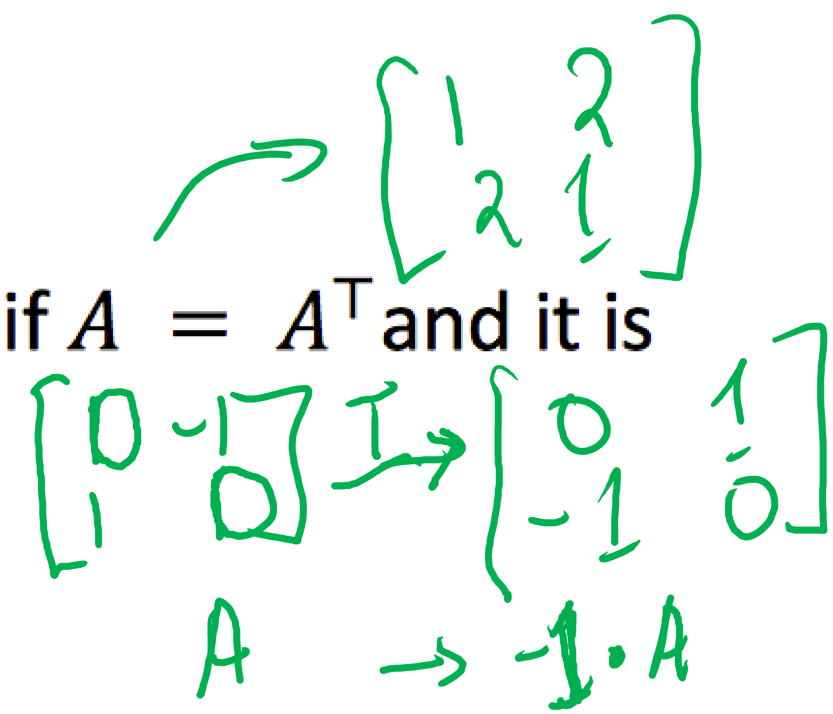
- Transpose of a matrix results from flipping the rows and columns. Given $A \in \mathbb{R}^{n \times d}$, transpose is $A^{\top} \in \mathbb{R}^{d \times n}$

- For each element of the matrix, the transpose can be written as → $A^{\top}_{ij} = A_{ji}$



- The following properties of the transposes are easily verified
  - $(A^{\top})^{\top} = A$
  - $(AB)^{\top} = B^{\top}A^{\top}$
  - $(A + B)^{\top} = A^{\top} + B^{\top}$

- A square matrix $A \in \mathbb{R}^{d \times d}$ is symmetric if $A = A^{\top}$ and it is anti-symmetric if $A = -A^{\top}$.



Outline

- Linear Algebra Basics
- Norms ⬅
- Multiplications
- Matrix Inversion
- Trace and Determinant
- Eigen Values and Eigen Vectors
- Singular Value Decomposition

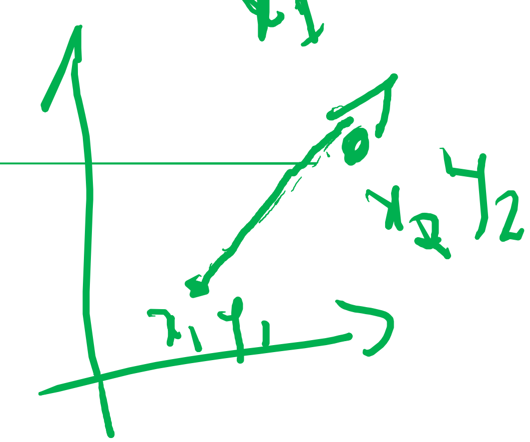
Norms

- Norm of a vector $||x||$ is informally a measure of the "length" of a vector

- More formally, a norm is any function $f: \mathbb{R}^d \rightarrow \mathbb{R}$ that satisfies
  - For all $x \in \mathbb{R}^d$, $f(x) \geq 0$ (non-negativity)
  - $f(x) = 0$ is and only if $x = 0$ (definiteness)
  - For $x \in \mathbb{R}^d$, $t \in \mathbb{R}$, $f(tx) = |t| f(x)$ (homogeneity)
  - For all $x, y \in \mathbb{R}^d$, $f(x + y) \leq f(x) + f(y)$ (triangle inequality)

- Common norms used in machine learning are
  - ℓ_2 norm
    - $||x||_2 = \sqrt{\sum_{i=1}^{d} x_i^2}$

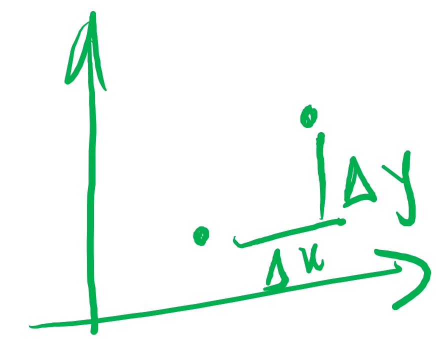
$\sqrt{(x_1-x_2)^2+(y_1-y_2)^2}$



Norms

- ℓ_1 norm
  - $\|x\|_1 = \sum_{i=1}^{d} |x_i|$
- ℓ_∞ norm
  - $\|x\|_\infty = max_i |x_i|$.

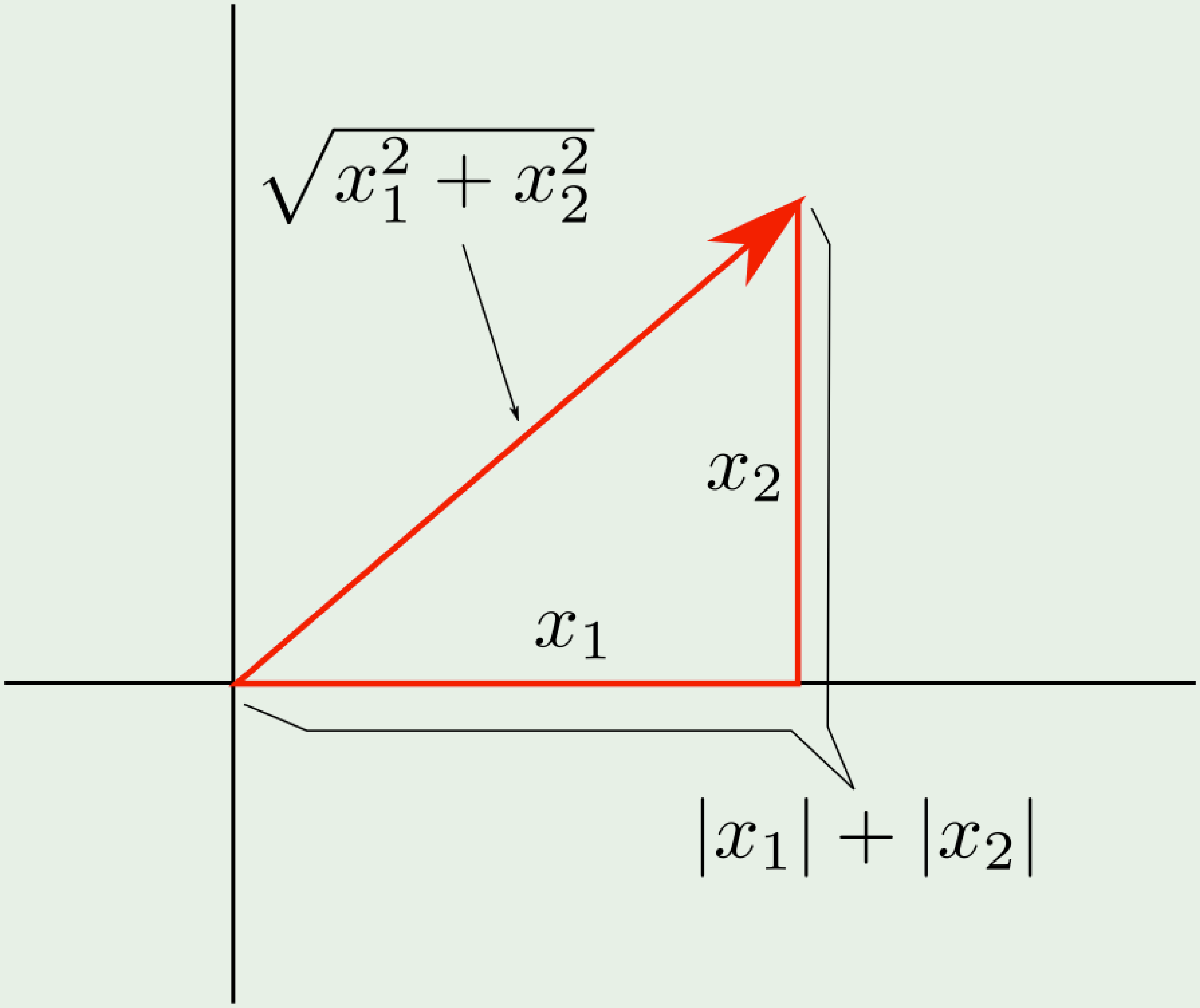
Why????



- All norms presented so far are examples of the family of ℓ_p norms, which are parameterized by a real number $p \geq 1$
  - $\|x\|_p = \left(\sum_{i=1}^{n} |x_i|^p\right)^{\frac{1}{p}}$
- Norms can be defined for matrices, such as the Frobenius norm.
  - $\|A\|_F = \sqrt{\sum_{i=1}^{n} \sum_{j=1}^{d} A_{ij}{}^2} = \sqrt{tr(A^\top A)}$

Vector Norm Examples

Example ℓ_1-norm and ℓ_2-norm



Special Matrices

- The identity matrix, denoted by $I \in \mathbb{R}^{d \times d}$ is a square matrix with ones on the diagonal and zeros everywhere else

- A diagonal matrix is matrix where all non-diagonal matrices are 0. This is typically denoted as $D = diag(d_1, d_2, d_3, \ldots, d_d)$

- Two vectors $x, y \in \mathbb{R}^d$ are orthogonal if $x.y = 0$. A square matrix $U \in \mathbb{R}^{d \times d}$ is orthogonal if all its columns are orthogonal to each other and are normalized

- It follows from orthogonality and normality that
  - $U^T U = I = U U^T$
  - $\|Ux\|_2 = \|x\|_2$

Outline

- Linear Algebra Basics
- Norms
- Multiplications
- Matrix Inversion
- Trace and Determinant
- Eigen Values and Eigen Vectors
- Singular Value Decomposition

Multiplications

- The product of two matrices $A \in \mathbb{R}^{n \times d}$ and $B \in \mathbb{R}^{d \times p}$ is given by $C \in \mathbb{R}^{n \times p}$, where $C_{ij} = \sum_{k=1}^{d} A_{ik} B_{kj}$

- Given two vectors $x, y \in \mathbb{R}^d$, the term xy^T (also $x \cdot y$) is called the **inner product** or **dot product** of the vectors, and is a real number given by $\sum_{i=1}^{d} x_i y_i$. For example,

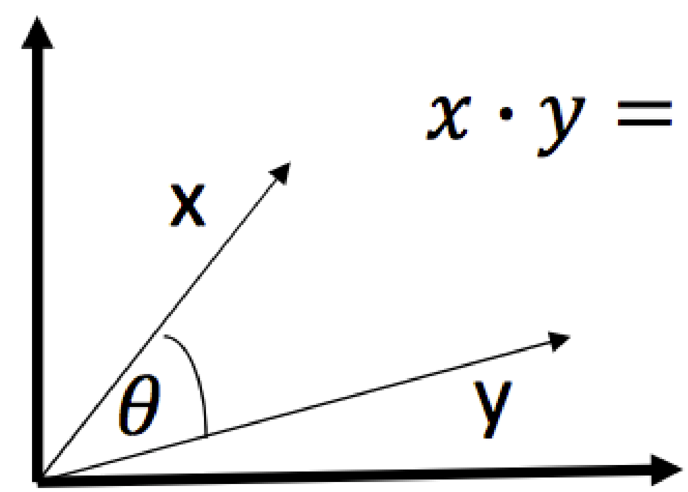
$$xy^T = [x_1 \quad x_2 \quad x_3] \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \sum_{i=1}^{3} x_i y_i$$

- Given two vectors x ∈ $\mathbb{R}^d$, y ∈ $\mathbb{R}^n$, the term $x^T y$ is called the **outer product** of the vectors : $x \otimes y$

Multiplications

$$x \otimes y = xy^{T} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \begin{bmatrix} y_1 & y_2 & y_3 \end{bmatrix} = \begin{bmatrix} x_1y_1 & x_1y_2 & x_1y_3 \\ x_2y_1 & x_2y_2 & x_2y_3 \\ x_3y_1 & x_3y_2 & x_3y_3 \end{bmatrix}$$

- The dot product also has a geometrical interpretation, for vectors in $x, y \in \mathbb{R}^2$ with angle θ between them

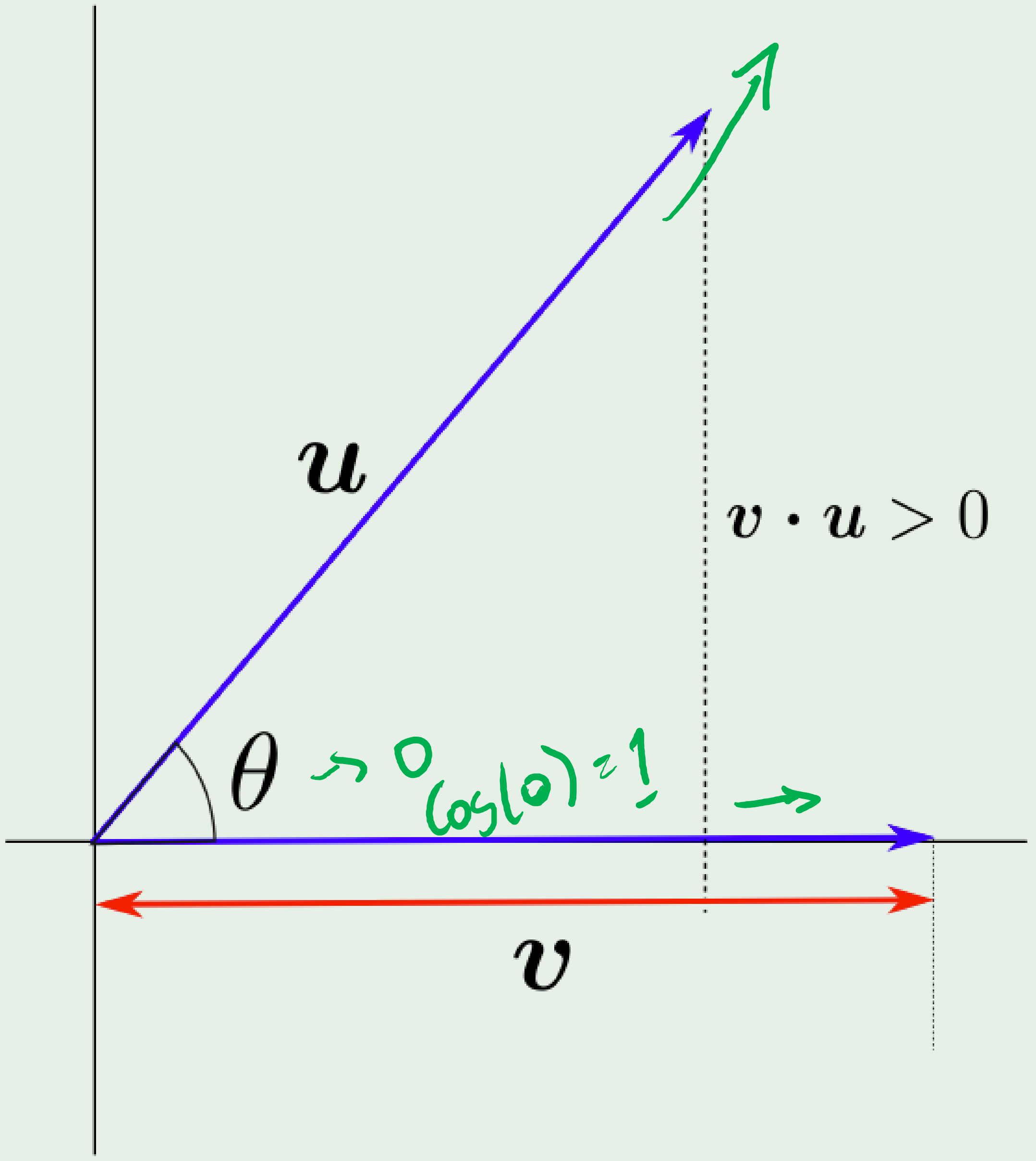


$$x \cdot y = |x||y| \cos\theta$$

which leads to use of dot product for testing orthogonality, getting the Euclidean norm of a vector, and scalar projections.

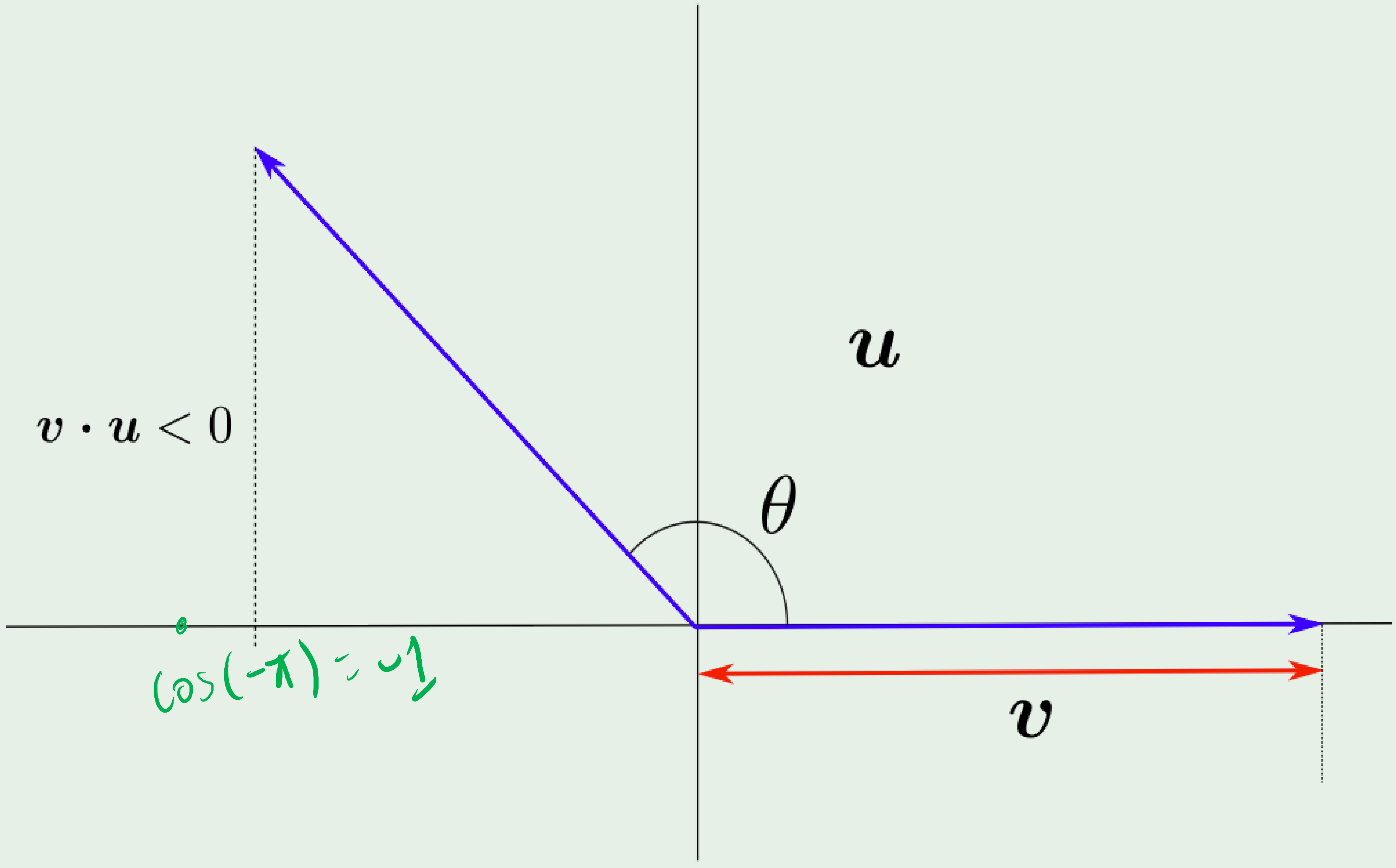
Inner Product Properties

The inner product is a measure of correlation between two vectors, scaled by the norms of the vectors



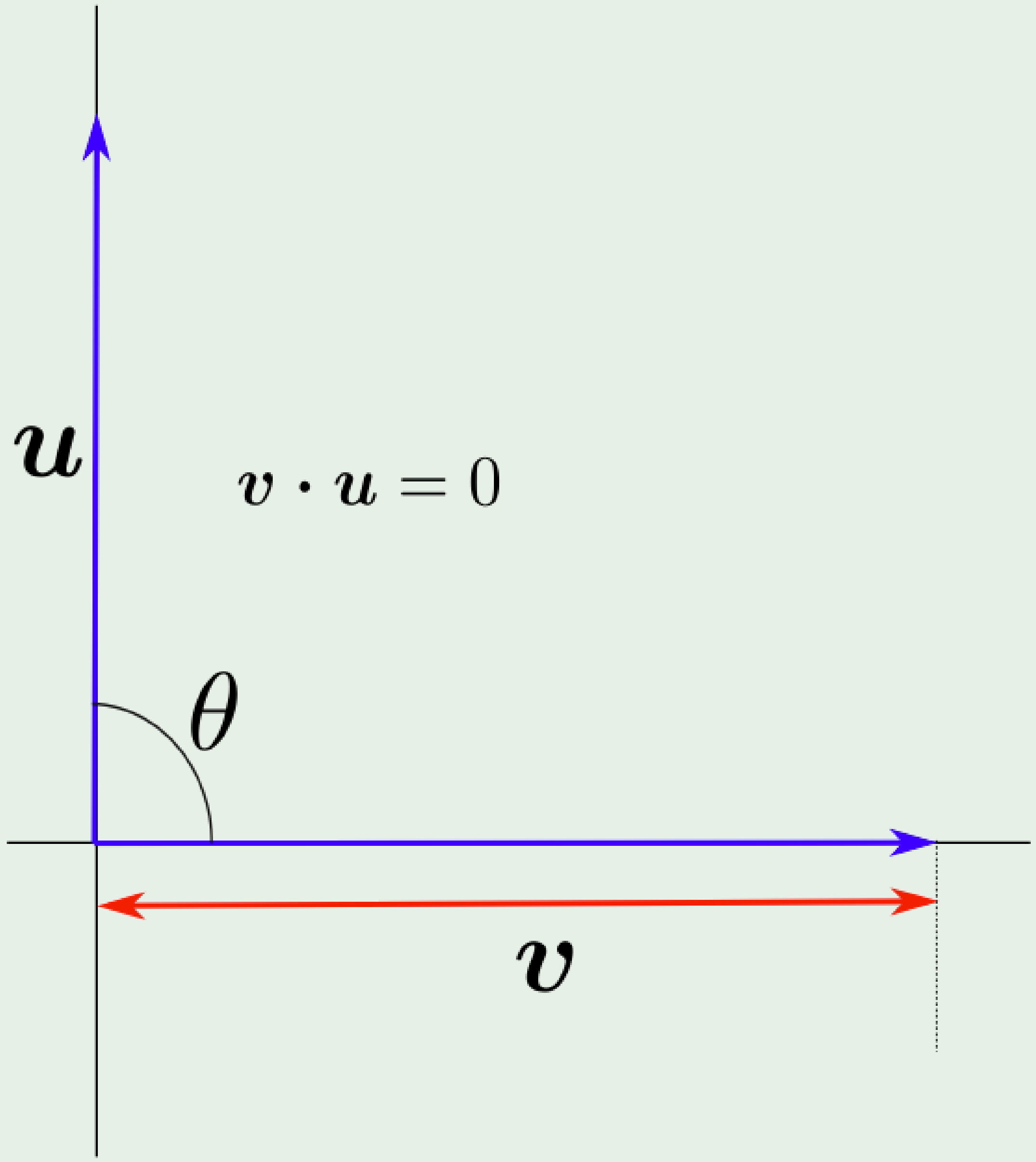
Inner Product Properties

The inner product is a measure of correlation between two vectors, scaled by the norms of the vectors



Inner Product Properties

The inner product is a measure of correlation between two vectors, scaled by the norms of the vectors



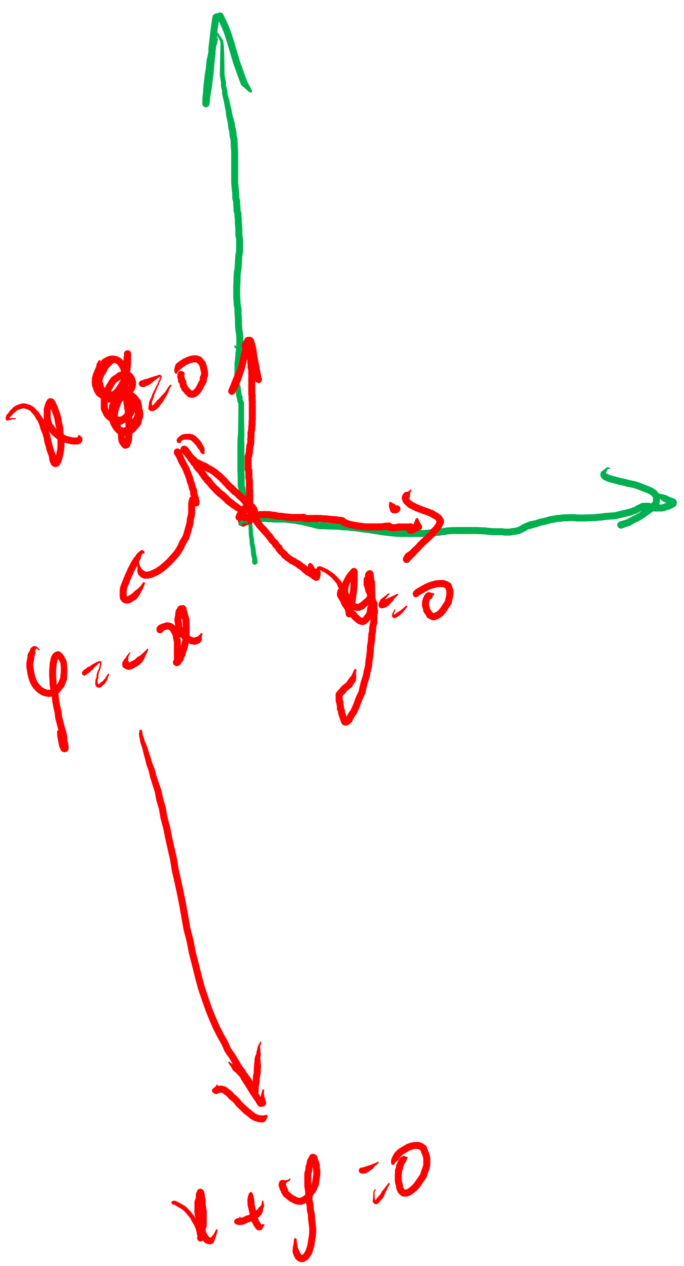
Example dot product

1) $y = 0$

[0 1]

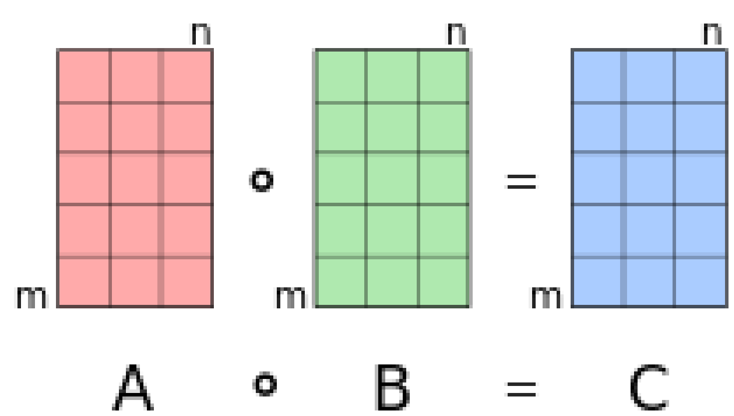
2) $x = 0$

[1 0]



Example Hadamard product or element wise multiplication

- Multiply each element a matrix with index (i,j,…) to another matrix’s element with the same index (i,j,…) to create a new matrix with the same number of elements.
- All matrices involved have the same shape and size.



$$\begin{bmatrix} 1 \\ 2 \end{bmatrix} \odot \begin{bmatrix} 3 \\ 4 \end{bmatrix} = \begin{bmatrix} 1 * 3 \\ 2 * 4 \end{bmatrix} = \begin{bmatrix} 3 \\ 8 \end{bmatrix}$$

Poll

- Identity matrix for Hadamard product?

$$\begin{bmatrix} 1 & 2 \\ 3 & 3 \end{bmatrix} \qquad \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Or

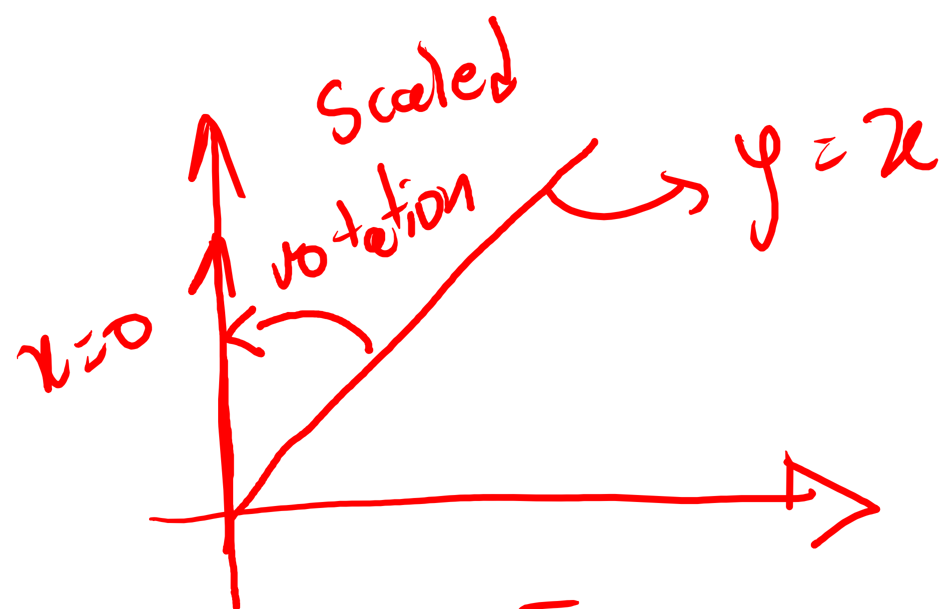
$$\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$$

Matrix multiplication geometric meaning
Affine transformations!!!

$$y = x$$

$$-x + y = 0$$

$$[-1 \quad 1]$$



Identity matrix

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

Random matrix

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

Outline

- Linear Algebra Basics
- Norms
- Multiplications
- Matrix Inversion
- Trace and Determinant
- Eigen Values and Eigen Vectors
- Matrix Decomposition

Linear Independence and Matrix Rank

- A set of vectors $\{x_1,\ x_2, \ldots, x_d\} \subset \mathbb{R}^d$ are said to be ***(linearly) independent*** if no vector can be represented as a linear combination of the remaining vectors. That is if

$$x_d = \sum_{i=1}^{d-1} \alpha_i x_i$$

for some scalar values $\alpha_1, \alpha_2, \ldots \in \mathbb{R}$ then we say that the vectors are linearly ***dependent***; otherwise the vectors are linearly independent

- The **column rank** of a matrix $A \in \mathbb{R}^{n \times d}$ is the size of the largest subset of columns of A that constitute a linearly independent set. **Row rank** of a matrix is defined similarly for rows of a matrix.

Matrix Rank: Examples

What are the ranks for the following matrices?

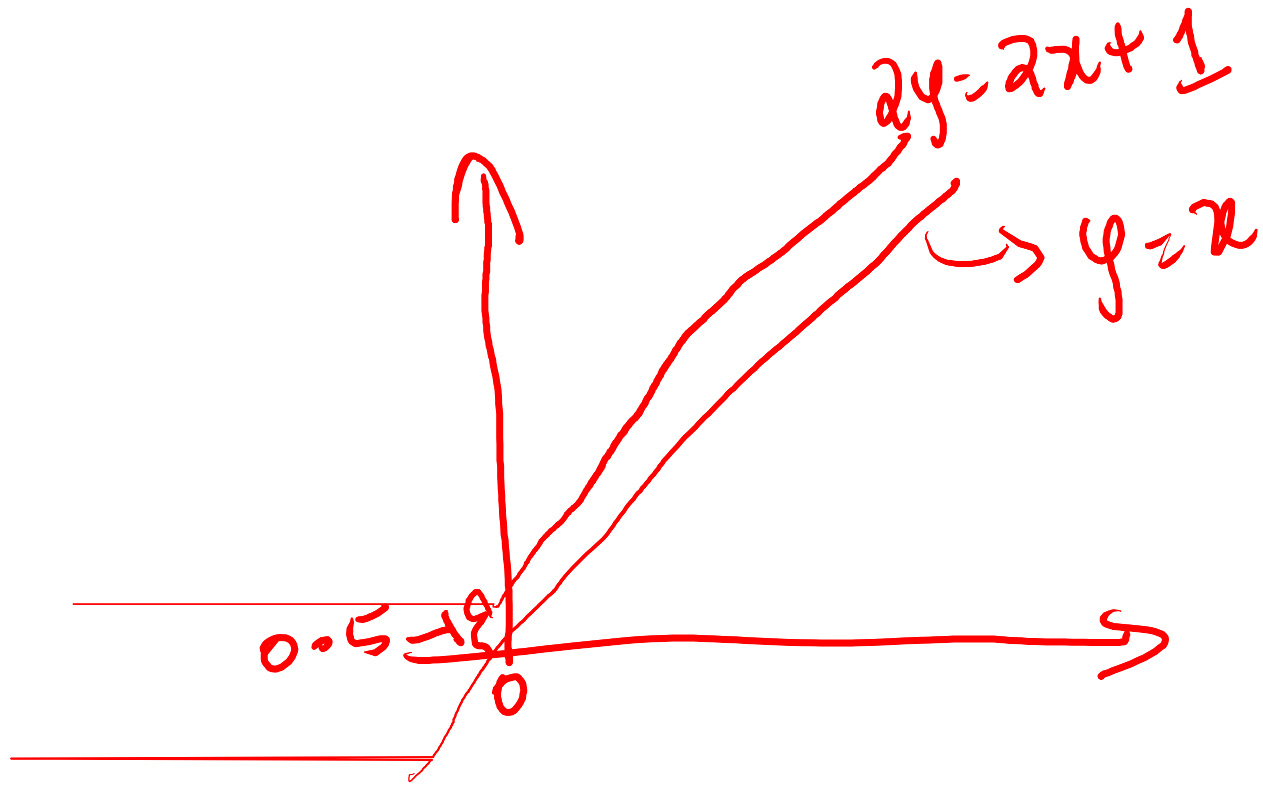
$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 4 & 6 \end{bmatrix}$$

1, 1

$$\mathbf{B} = \begin{bmatrix} 1 & 0 & 2 \\ 2 & 1 & 0 \\ 3 & 2 & 1 \end{bmatrix}$$

Geometric meaning

- Parallel lines: y = x and 2y = 2x + 1



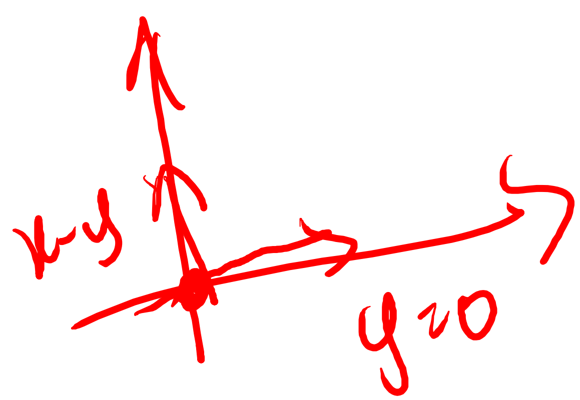
$$A \vec{x} = b$$

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

Rank = 1

Variables = 2

2×2



$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$x=0$ $y=0$

Matrix Inverse

- The inverse of a square matrix $A \in \mathbb{R}^{d \times d}$ is denoted A^{-1} and is the unique matrix such that $A^{-1}A = I = AA^{-1}$

- For some square matrices A^{-1} may not exist, and we say that A is **singular or non-invertible.** In order for A to have an inverse, A must be **full rank.**

- For non-square matrices the inverse, denoted by A^{+} ,is given by $A^{+} = (A^{\top}A)^{-1}A^{\top}$ called the **pseudo inverse**

Outline

- Linear Algebra Basics
- Norms
- Multiplications
- Matrix Inversion
- Trace and Determinant
- Eigen Values and Eigen Vectors
- Singular Value Decomposition

Matrix Trace

- The trace of a matrix $A \in \mathbb{R}^{d \times d}$, denoted as $\boldsymbol{tr}(\boldsymbol{A})$, is the sum of the diagonal elements in the matrix

$$tr(A) = \sum_{i=1}^{d} A_{ii}$$

- The trace has the following properties
  - For $A \in \mathbb{R}^{d \times d}$, $tr(A) = trA^{\top}$
  - For $A, B \in \mathbb{R}^{d \times d}$, $tr(A + B) = tr(A) + tr(B)$
  - For $A \in \mathbb{R}^{d \times d}$, t $\in \mathbb{R}$, $tr(tA) = t \cdot tr(A)$
  - For A, B, C such that ABC is a square matrix $tr(ABC) = tr(BCA) = tr(CAB)$

- The trace of a matrix helps us easily compute norms and eigenvalues of matrices as we will see later

Matrix Determinant

Definition (Determinant)

The determinant of a square matrix A, denoted by $|A|$, is defined as

$$\det(A) = \sum_{j=1}^{d} (-1)^{i+j} a_{ij} M_{ij}$$

where M_{ij} is determinant of matrix A without the row i and column j.

For a 2×2 matrix $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$

$$|A| = ad - bc$$

Properties of Matrix Determinant

Basic Properties

- $|A| = \left|A^T\right|$
- $|AB| = |A|\,|B|$
- $|A| = 0$ if and only if A is not invertible
- If A is invertible, then $\left|A^{-1}\right| = \frac{1}{|A|}$.

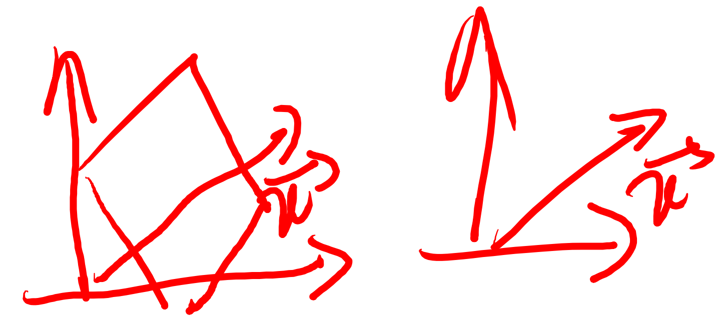
Outline

- Linear Algebra Basics
- Norms
- Multiplications
- Matrix Inversion
- Trace and Determinant
- Eigen Values and Eigen Vectors
- Singular Value Decomposition

Eigenvalues and Eigenvectors

- Given a square matrix $A \in \mathbb{R}^{d \times d}$ we say that $\lambda \in \mathbb{C}$ is an eigenvalue of A and $x \in \mathbb{C}^{d}$ is an eigenvector if

$$Ax = \lambda x, \qquad x \neq 0$$



- Intuitively this means that upon multiplying the matrix A with a vector x, we get the same vector, but scaled by a parameter λ

- Geometrically, we are transforming the matrix A from its original orthonormal basis/co-ordinates to a new set of orthonormal basis x with magnitude as λ

Computing Eigenvalues and Eigenvectors

- We can rewrite the original equation in the following manner

$$Ax = \lambda x, \qquad x \neq 0$$
$$\Rightarrow (A - \lambda I)\, x = 0, \qquad x \neq 0$$

- This is only possible if $(A - \lambda I)$ is singular, that is $|(A - \lambda I)| = 0$.

- Thus, eigenvalues and eigenvectors can be computed.
  - Compute the determinant of $A - \lambda I$.
    - This results in a polynomial of degree d.
  - Find the roots of the polynomial by equating it to zero.
    - The d roots are the d eigenvalues of A. They make $A - \lambda I$ singular.
  - For each eigenvalue λ, solve $(A - \lambda I)\, x$ to find an eigenvector x

Eigenvalue Example

Eigenvalues

$$\mathbf{A} = \begin{bmatrix} 1 & 2 \\ 3 & -4 \end{bmatrix} \quad \begin{matrix} \lambda_1 = -5 \\ \lambda_2 = 2 \end{matrix}$$

$$|A - \lambda I| = 0$$

$$\left| \begin{bmatrix} 1-\lambda & 0 \\ 3 & -6-\lambda \end{bmatrix} \right| = 0$$

$$a\lambda^2 + b\lambda + c = 0$$

Determine eigenvectors: $\mathbf{Ax} = \lambda\mathbf{x}$

$$\begin{matrix} x_1 + 2x_2 = \lambda x_1 \\ 3x_1 - 4x_2 = \lambda x_2 \end{matrix} \Rightarrow \begin{matrix} (1-\lambda)x_1 + 2x_2 = 0 \\ 3x_1 - (4+\lambda)x_2 = 0 \end{matrix}$$

Eigenvector for $\lambda_1 = -5$

$$\begin{matrix} 6x_1 + 2x_2 = 0 \\ 3x_1 + x_2 = 0 \end{matrix} \Rightarrow \mathbf{x}_1 = \begin{bmatrix} -0.3162 \\ 0.9487 \end{bmatrix} \text{or } \mathbf{x}_1 = \begin{bmatrix} 1 \\ -3 \end{bmatrix}$$

Eigenvector for $\lambda_1 = 2$

$$\begin{matrix} -x_1 + 2x_2 = 0 \\ 3x_1 - 6x_2 = 0 \end{matrix} \Rightarrow \mathbf{x}_2 = \begin{bmatrix} 0.8944 \\ 0.4472 \end{bmatrix} \text{or } \mathbf{x}_2 = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$$

Slide credit: Shubham Kumbhar

Matrix Eigen Decomposition

- All the eigenvectors can be written together as $AX = X\Lambda$ where the columns of X are the eigenvectors of A, and Λ is a diagonal matrix whose elements are eigenvalues of A
- If the eigenvectors of A are invertible, then $A = X\Lambda X^{-1}$
- There are several properties of eigenvalues and eigenvectors
  - $Tr(A) = \sum_{i=1}^{d} \lambda_i$
  - $|A| = \prod_{i=1}^{d} \lambda_i$
  - Rank of A is the number of non-zero eigenvalues of A
  - If A is non-singular then $1/\lambda_i$ are the eigenvalues of A^{-1}
  - The eigenvalues of a diagonal matrix are the diagonal elements of the matrix itself!

Outline

- Linear Algebra Basics
- Norms
- Multiplications
- Matrix Inversion
- Trace and Determinant
- Eigen Values and Eigen Vectors
- Singular Value Decomposition

Covariance matrix

- For a dataset $\mathbf{A}$ we can define the covariance matrix as $\mathbf{C} = \frac{\bar{A}^{\mathrm{T}}\bar{A}}{N}$ for large N and $\mathbf{C} = \frac{\bar{A}^{\mathrm{T}}\bar{A}}{N-1}$ for small N. $\bar{\mathbf{A}}$ is the matrix $\mathbf{A}$ centered around its mean

$$\mathbf{A} = \begin{bmatrix} 3 & 2 & 1 \\ 1 & 2 & 2 \\ 4 & 3 & 2 \\ 2 & 1 & 1 \end{bmatrix}, \qquad \boldsymbol{\mu}^T = [2.5 \quad 2.0 \quad 1.5]$$

Covariance matrix

$$\mathbf{A} = \begin{bmatrix} 3 & 2 & 1 \\ 1 & 2 & 2 \\ 4 & 3 & 2 \\ 2 & 1 & 1 \end{bmatrix} \rightarrow \bar{\mathbf{A}} = \begin{bmatrix} 0.5 & 0.0 & -0.5 \\ -1.5 & 0.0 & 0.5 \\ 1.5 & 1.0 & 0.5 \\ -0.5 & -1.0 & -0.5 \end{bmatrix}$$

$$\mathbf{C} = \frac{\bar{A}^{\mathrm{T}} \bar{A}}{N-1} = \frac{1}{4-1} \begin{bmatrix} 0.5 & -1.5 & 1.5 & -0.5 \\ 0.0 & 0.0 & 1.0 & -1.0 \\ -0.5 & 0.5 & 0.5 & -0.5 \end{bmatrix} \begin{bmatrix} 0.5 & 0.0 & -0.5 \\ -1.5 & 0.0 & 0.5 \\ 1.5 & 1.0 & 0.5 \\ -0.5 & -1 & -0.5 \end{bmatrix}$$

$$\mathbf{C} = \begin{bmatrix} 1.7 & 0.7 & 0.0 \\ 0.7 & 0.7 & 0.3 \\ 00. & 0.3 & 0.3 \end{bmatrix}$$

Correlation matrix

- Given that the different features may not be on the same order of magnitude, the covariance matrix can be standardized based on the standard deviation of the individual features to yield the correlation matrix, such that

$$\mathbf{corr} = \frac{covariance(X,Y)}{\sigma_x \sigma_y}$$

Correlation matrix

- Back to our example…

$$\mathbf{corr} = \begin{bmatrix} \frac{1.7}{1.7} & \frac{0.7}{\sqrt{1.7}\sqrt{0.7}} & \frac{0.0}{\sqrt{1.7}\sqrt{0.3}} \\ \frac{0.7}{\sqrt{1.7}\sqrt{0.7}} & \frac{0.7}{0.7} & \frac{0.3}{\sqrt{0.7}\sqrt{0.3}} \\ \frac{0.0}{\sqrt{1.7}\sqrt{0.3}} & \frac{0.3}{\sqrt{0.7}\sqrt{0.3}} & \frac{0.3}{0.3} \end{bmatrix} = \begin{bmatrix} 1.0 & 0.6 & 0.0 \\ 0.6 & 1.0 & 0.7 \\ 00. & 0.7 & 1.0 \end{bmatrix}$$

Singular Value Decomposition

$X_{n \times d}$

n: instances
d: dimensions
X is a centered matrix

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

$$U_{n \times n} \rightarrow unitary\ matrix \rightarrow U \times U^T = I$$

$$\Sigma_{n \times d} \rightarrow diagonal\ matrix$$

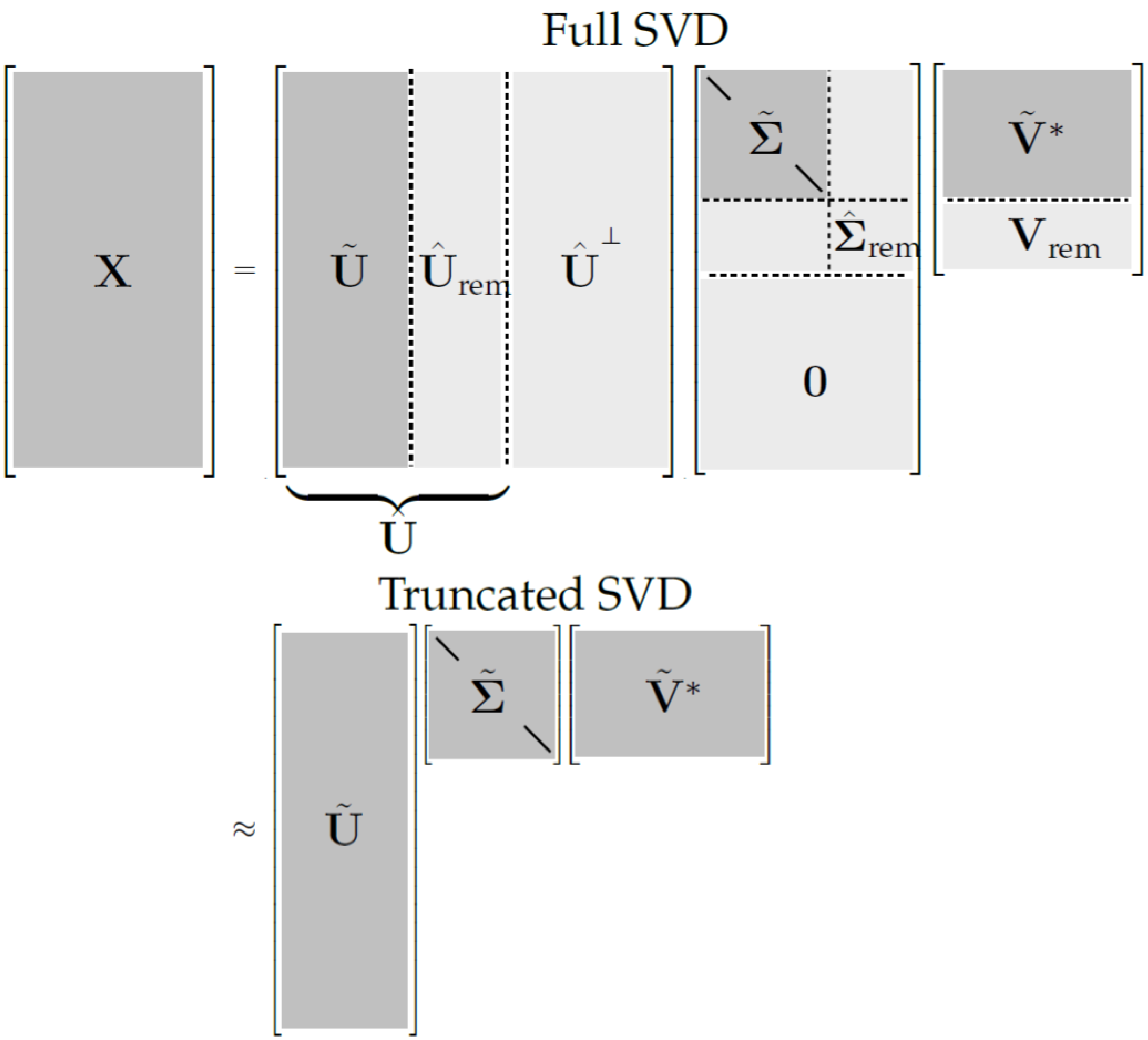
$$\mathrm{V}_{d \times d} \rightarrow unitary\ matrix \rightarrow V \times V^T = I$$

$$X = \begin{bmatrix} u_{1\times1} & \cdots & \cdots & \cdots & u_{1\times n} \\ \vdots & \ddots & \cdots & \cdots & \vdots \\ \vdots & \vdots & \ddots & \cdots & \vdots \\ \vdots & \vdots & \cdots & \ddots & \vdots \\ u_{1\times1} & \cdots & \cdots & \cdots & u_{n\times n} \end{bmatrix} \times \begin{bmatrix} \sum_{1\times1} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \sum_{d\times d} \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \times \begin{bmatrix} v_{1\times1} & \cdots & \cdots & \cdots & v_{1\times d} \\ \vdots & \ddots & \cdots & \cdots & \vdots \\ \vdots & \vdots & \ddots & \cdots & \vdots \\ \vdots & \vdots & \cdots & \ddots & \vdots \\ v_{d\times1} & \cdots & \cdots & \cdots & v_{d\times d} \end{bmatrix}$$

U $\quad\Sigma\quad$ V^T

$$d < n$$

Importance of SVD



Covariance matrix:

$$C_{d\times d} = \frac{X^T X}{n}$$

$$\left.\begin{aligned} X &= U\Sigma V^T \\ C &= \frac{X^T X}{n} \end{aligned}\right\} \quad C = \frac{V\Sigma^T U^T U\Sigma V^T}{n} = \frac{V\Sigma^2 V^T}{n}$$

$$C = \frac{V\Sigma^2 V^T}{n} = V\frac{\Sigma^2}{n}V^T$$

$$CV = V\frac{\Sigma^2}{n}V^T V = V\frac{\Sigma^2}{n}$$

According to Eigen-decomposition definition ➔ $CV = V\Lambda$

$\lambda_i = \frac{\Sigma_i^2}{n}$ ➔ The eigenvalues of covariance matrix

λ_i: Eigenvalue of C or covariance matrix

Σ_i: Singular value of X matrix

So, we can directly calculate eigenvalue of a covariance matrix by having the singular value of matrix X **directly**

Geometric Meaning of SVD

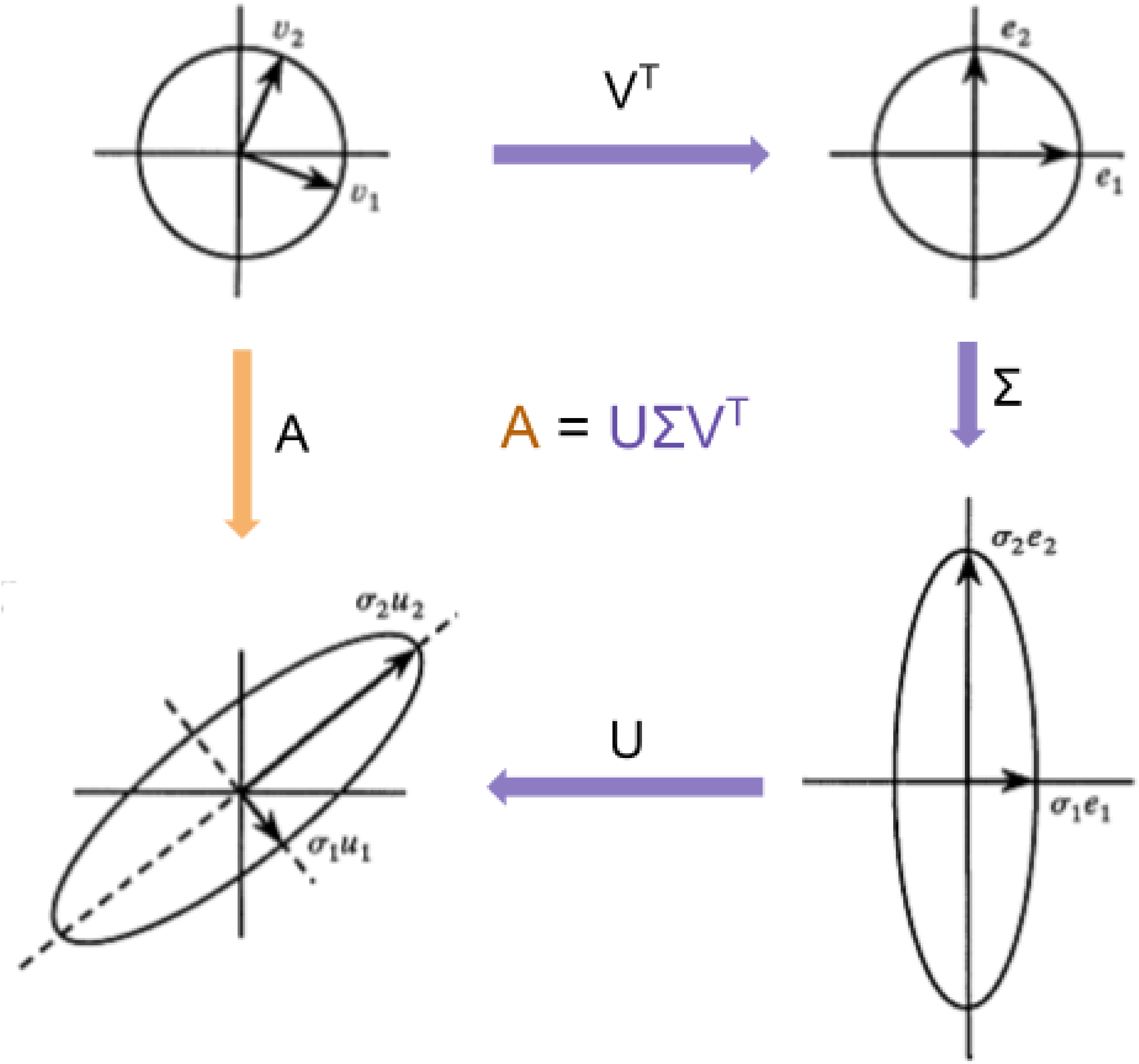


Image Credit: Kevin Binz

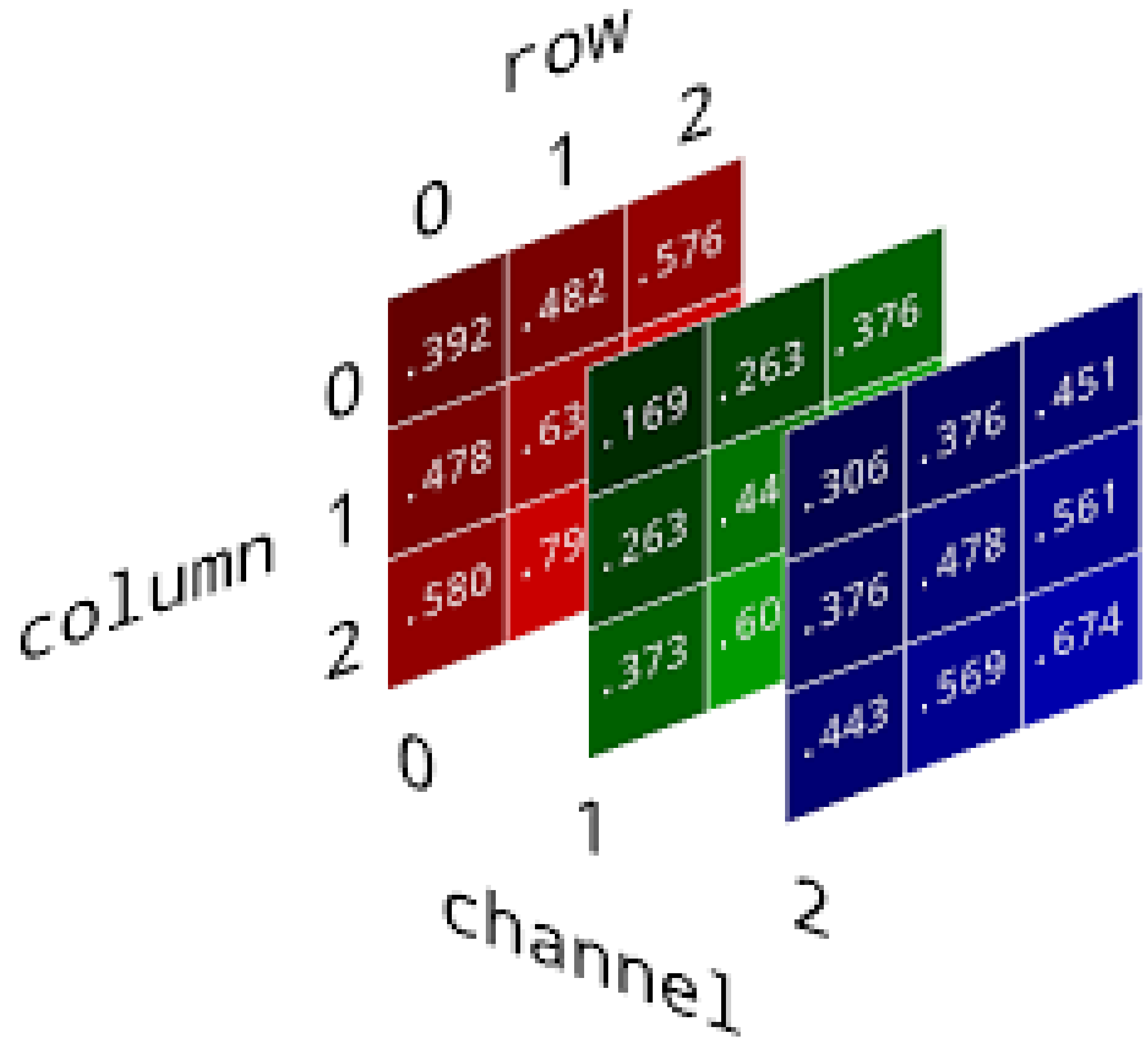
SVD Example

$$\begin{bmatrix} 2 & 0 \\ 0 & -3 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 2 & 0 \\ 0 & 3 \\ 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$\mathbf{M} \qquad \mathbf{U} \qquad \mathbf{\Sigma} \qquad \mathbf{V}^{\mathrm{T}}$$

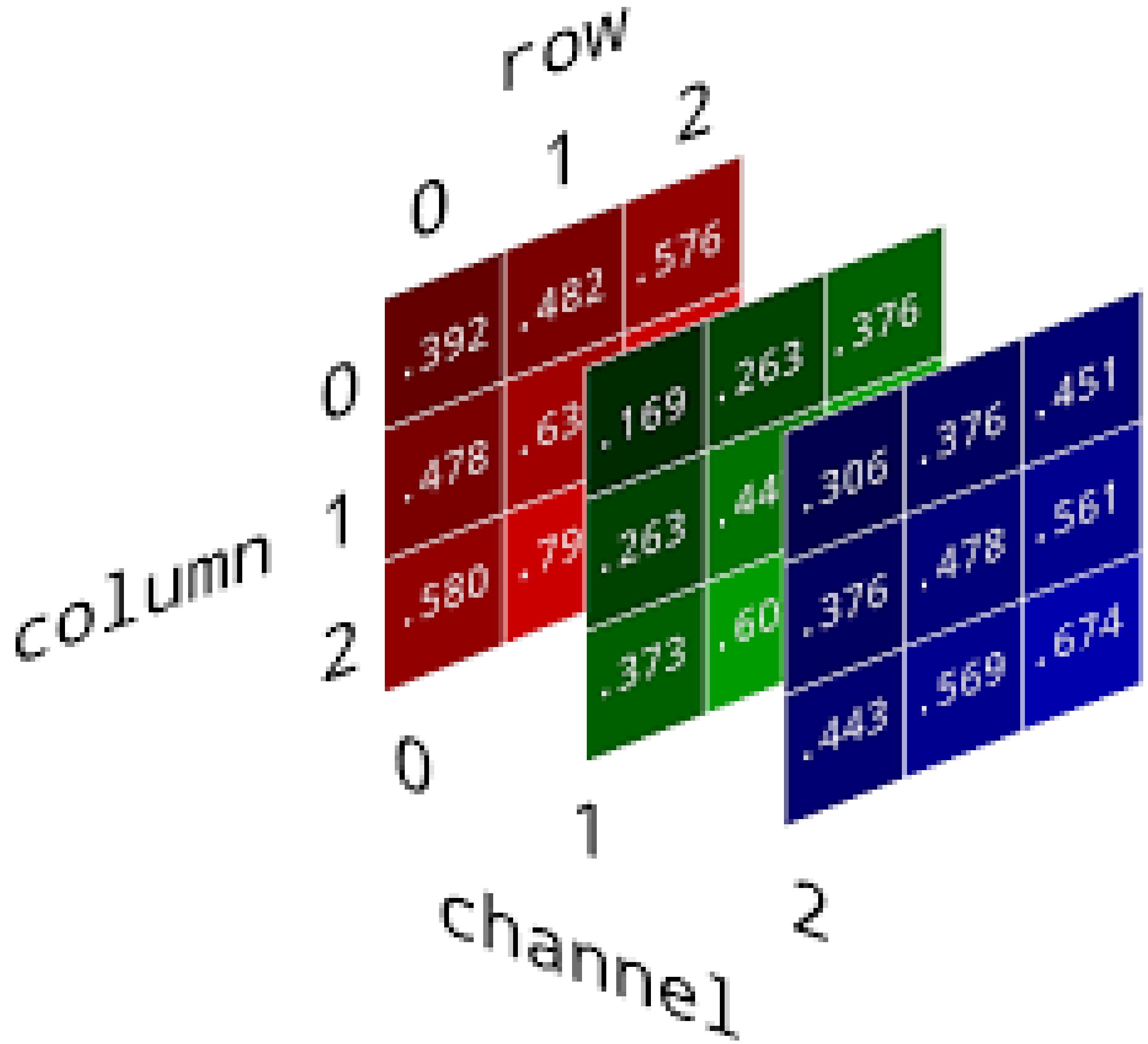
From [Strang]

Example on an image



RGB Image to matrix from kdnuggets.com

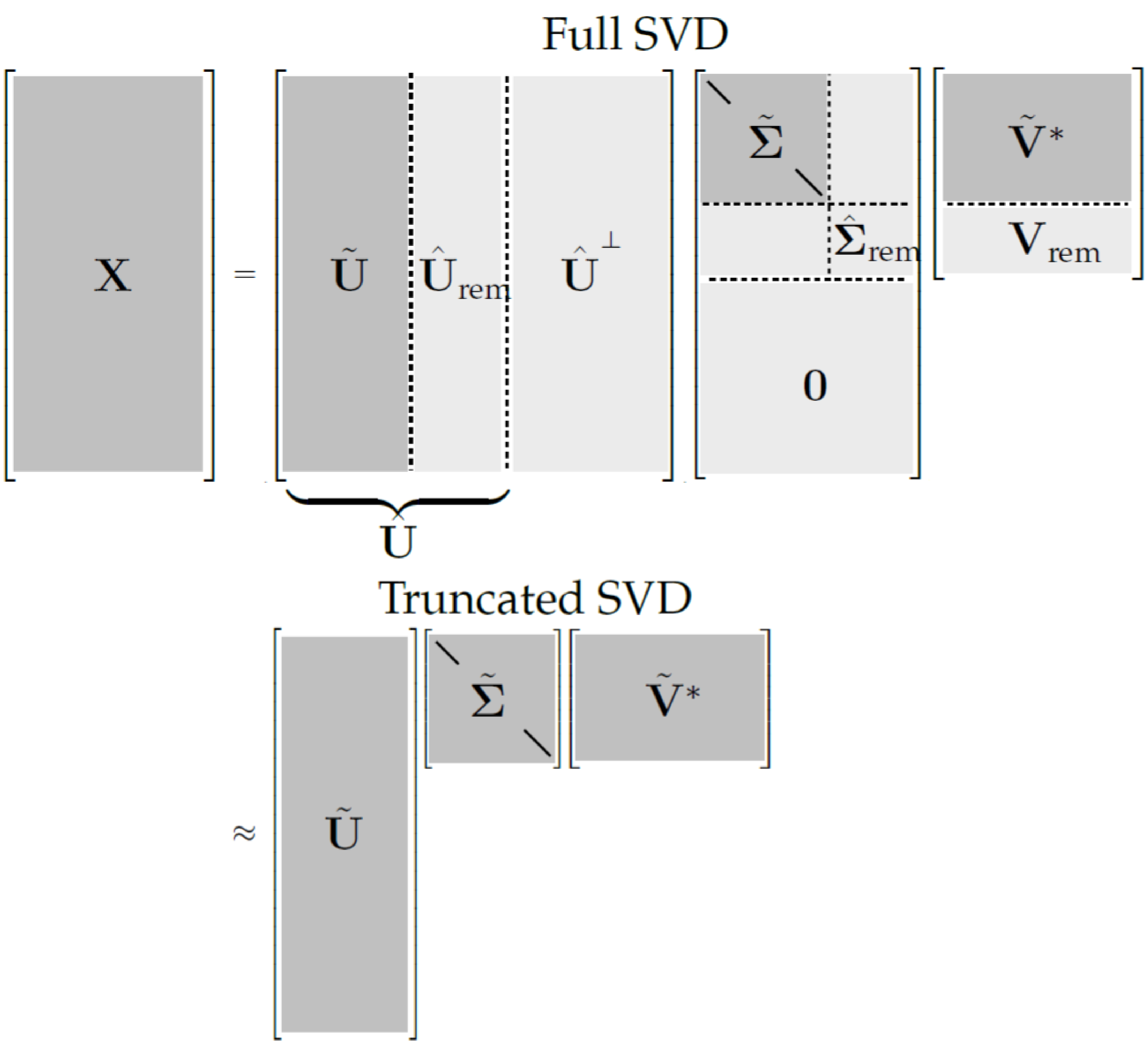
Example on an image



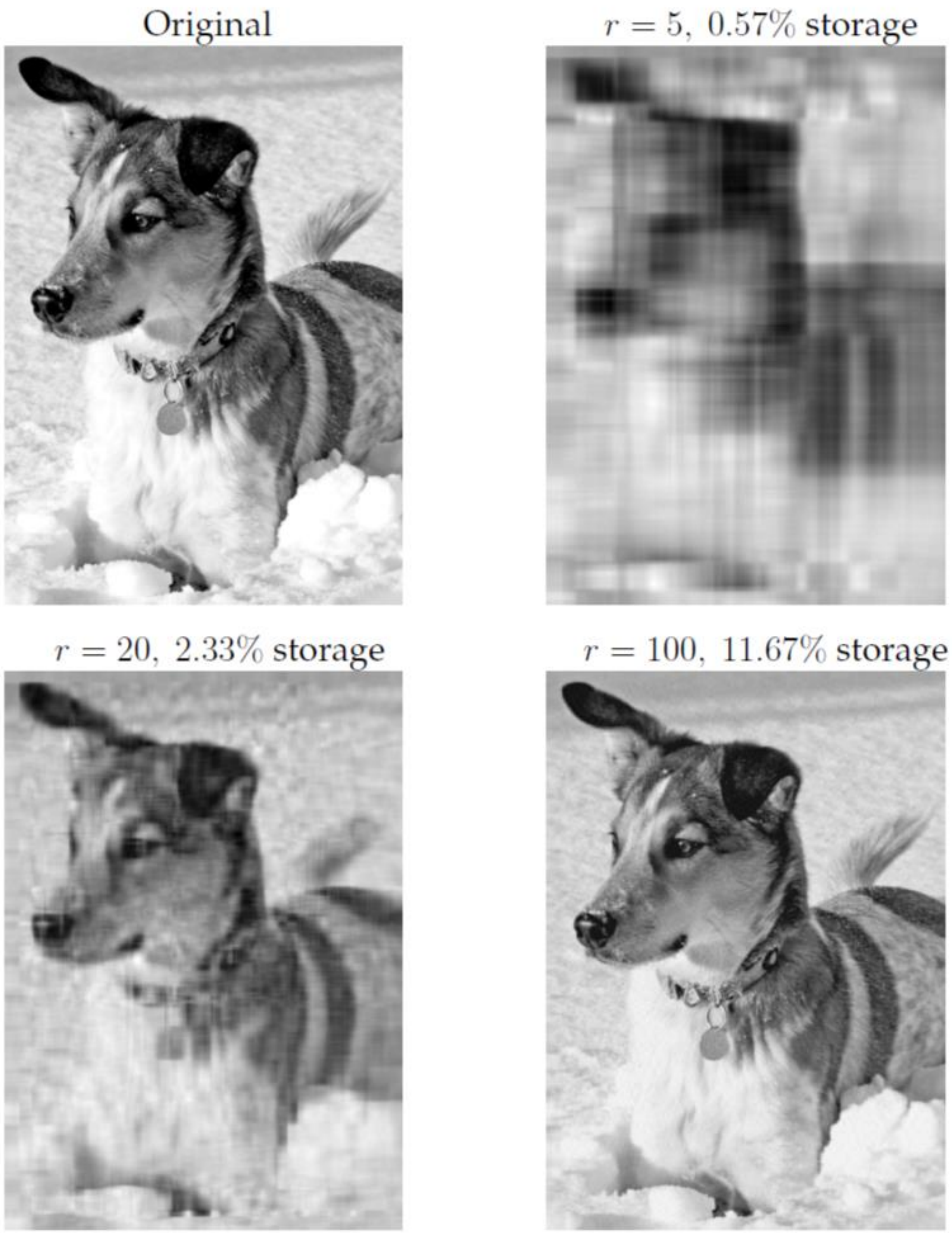
RGB Image to matrix from kdnuggets.com

From SVD example in Data Driven Science & Engineering by Steven L. BruntonAnd J. Nathan Kutz

Example on an image



Example on an image



From SVD example in Data Driven Science & Engineering by Steven L. BruntonAnd J. Nathan Kutz

Summary

- Linear Algebra Basics
- Norms
- Multiplications
- Matrix Inversion
- Trace and Determinant
- Eigen Values and Eigen Vectors
- Singular Value Decomposition