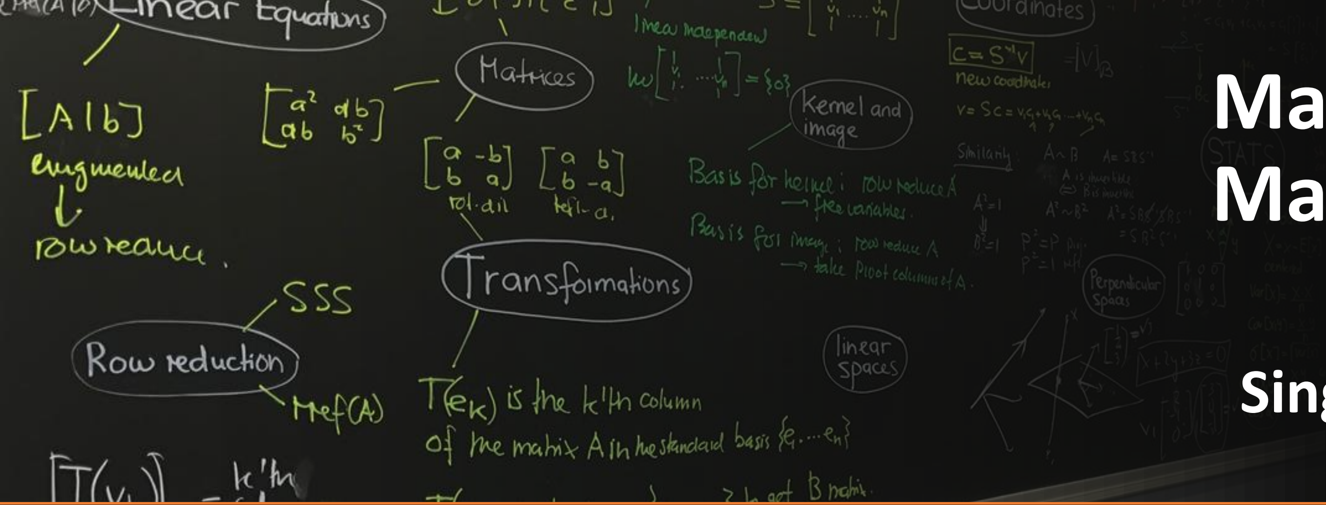


Mathematical Foundations for Machine Learning and Data Science

Singular Value Decomposition, Column Space and Null Space

Dr. Zubair Khalid

Department of Electrical Engineering
School of Science and Engineering
Lahore University of Management Sciences

https://www.zubairkhalid.org/ee212_2021.html

LUMS
A Not-for-Profit University

Outline

- Positive/negative definite and semi-definite matrices
- Singular Value Decomposition (SVD)
  - Formulation
  - Interpretation
  - Application examples
- Column space and Null Space

Positive/Negative Definite/Semi-Definite Matrices

Definition:

For a matrix $A \in \mathbf{R}^{n \times n}$, if

$x^T A x \geq 0$	$\forall\ x \in \mathbf{R}^n$	A is positive semi-definite (PSD)
$x^T A x > 0$	$\forall\ x \in \mathbf{R}^n$	A is positive definite (PD)
$x^T A x \leq 0$	$\forall\ x \in \mathbf{R}^n$	A is negative semi-definite (NSD)
$x^T A x < 0$	$\forall\ x \in \mathbf{R}^n$	A is negative definite (ND)

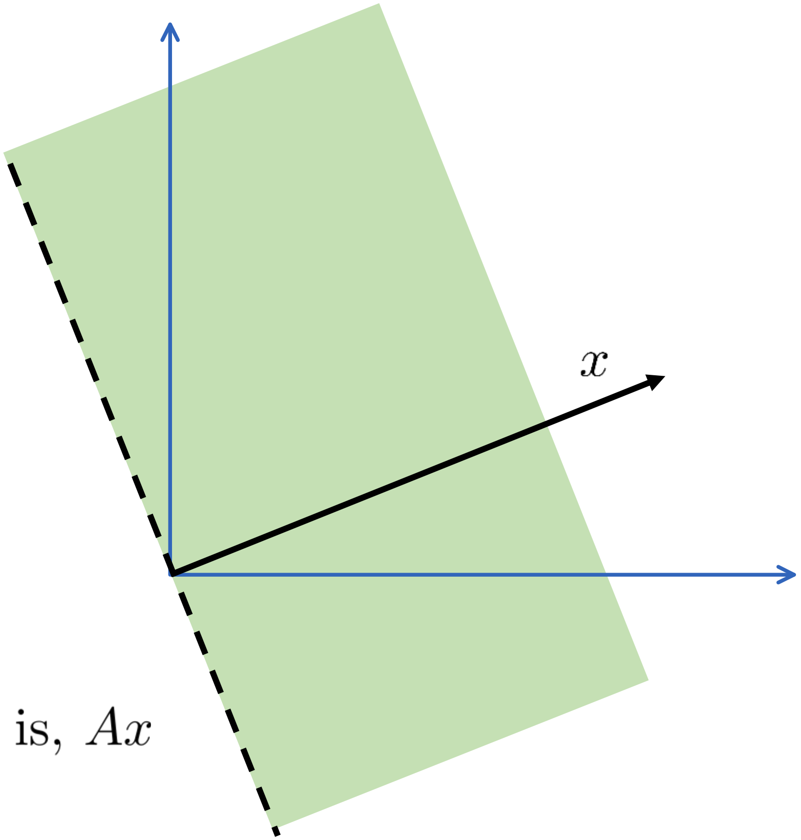
Positive Definite and Semi-Definite Matrices

Interpretation:

A is positive semi-definite (PSD)

$$x^T A x \geq 0$$

- Let $y = Ax$
  - y is a linear transformation defined by the matrix A.
- $x^T y \geq 0$ implies angle between x and y is less than or equal to $\frac{\pi}{2}$.
- $x^T y \geq 0$ implies angle between x and linearly transformed x, that is, Ax is less than or equal to $\frac{\pi}{2}$.



Graphically, a vector x when transformed by a matrix A, that is, Ax can be anywhere in the green region including the dashed boundary where $x^T A x = 0$

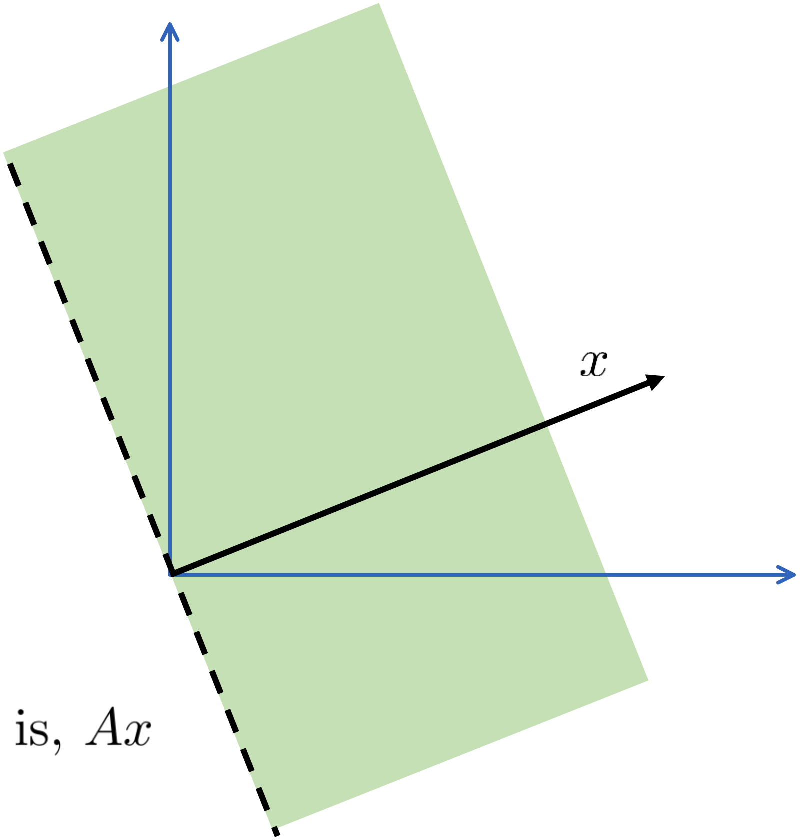
Positive Definite and Semi-Definite Matrices

Interpretation:

A is positive definite (PD)

$$x^T A x > 0$$

- Let $y = Ax$
  - y is a linear transformation defined by the matrix A.
- $x^T y > 0$ implies angle between x and y is less than $\frac{\pi}{2}$.
- $x^T y \geq 0$ implies angle between x and linearly transformed x, that is, Ax is less than $\frac{\pi}{2}$.



Graphically, a vector x when transformed by a matrix A, that is, Ax can be anywhere in the green region excluding the dashed boundary where $x^T A x = 0$

Positive Definite and Semi-Definite Matrices

Eigenvalues of symmetric PSD/PD matrix:

For a symmetric and PD matrix A, eigenvalues are positive.

How?

- We already know that the eigenvalues of a symmetric matrix are real.
- For a PD symmetric, we require $x^T A x > 0$
- If we take $x = q$, where q is an eigenvector with an associated eigenvalue λ

$$q^T A q > 0 \Rightarrow \lambda\, q^T q > 0 \Rightarrow \lambda\, \|q\|_2^2 > 0 \Rightarrow \lambda > 0$$

Similarly, we can show the following:

For a symmetric and PSD matrix A, eigenvalues are non-negative.

For a symmetric and NSD matrix A, eigenvalues are non-positive.

For a symmetric and ND matrix A, eigenvalues are negative.

Outline

- Positive/negative definite and semi-definite matrices
- Singular Value Decomposition (SVD)

Outline

- Positive/negative definite and semi-definite matrices
- Singular Value Decomposition (SVD)
  - Formulation
  - Interpretation
  - Application examples
- Column space and Null Space

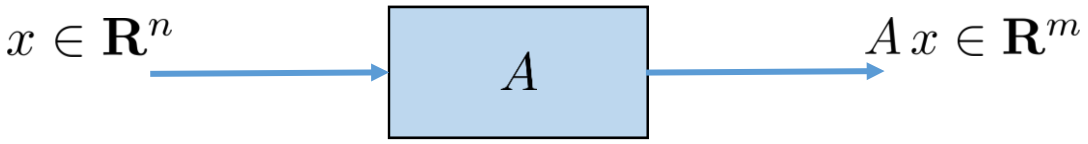
Singular Value Decomposition

Overview:

- The singular value decomposition (SVD) of a matrix is a central matrix decomposition method in linear algebra.

- It has been referred to as the "**fundamental theorem of linear algebra**" (Strang, 1993) because it can be applied to all matrices, not only to square matrices, and it always exists.

- For $A \in \mathbf{R}^{m \times n}$, we have



- SVD explains the underlying geometry of this linear transformation.

Singular Value Decomposition

Formulation:

- For any matrix $A \in \mathbf{R}^{m \times n}$, we have a singular value decomposition (SVD) given by

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

- Matrix $U \in \mathbf{R}^{m \times m}$ is an orthonormal matrix.
- Matrix $V \in \mathbf{R}^{n \times n}$ is an orthonormal matrix.
- Matrix $\Sigma \in \mathbf{R}^{m \times n}$ is a (special) diagonal matrix.

$m < n$

$$\Sigma = \begin{bmatrix} \sigma_1 & 0 & \dots & 0 & 0 & \dots & 0 \\ 0 & \sigma_2 & \dots & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & 0 & \dots & 0 \\ 0 & 0 & \dots & \sigma_m & 0 & \dots & 0 \end{bmatrix}$$

$m = n$

$$\Sigma = \begin{bmatrix} \sigma_1 & 0 & \dots & 0 \\ 0 & \sigma_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_n \end{bmatrix}$$

$m > n$

$$\Sigma = \begin{bmatrix} \sigma_1 & 0 & \dots & 0 \\ 0 & \sigma_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_n \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 \end{bmatrix}$$

Singular Value Decomposition

Formulation:

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

- Columns of U are referred to as left singular vectors of matrix A.
- Columns of V are referred to as right singular vectors of matrix A.
- $\sigma_1, \sigma_2, \ldots, \sigma_{\min(m,n)}$ are singular values of matrix A, which are (usually) indexed such that

$$\sigma_1 \geq \sigma_2 \geq \ \ldots \ \geq \sigma_{\min(m,n)} \geq 0$$

Singular Value Decomposition

How to Compute SVD?

- For a matrix $A \in \mathbf{R}^{m\times n}$, we define a matrix $G = AA^T$.
- Using $A = U\,\Sigma\,V^T$, we can write G as

$$G = U\,\Sigma\,V^T\,V\,\Sigma^T\,U^T = U\,\Sigma\,\Sigma^T\,U^T$$

- What is special about matrix G?
  - G is symmetric by definition.
  - G is positive semi-definite. How? You are fully equipped to show this.
- We note that $\Sigma\Sigma^T$ is a diagonal matrix of size $m \times m$.
- Eigenvalue decomposition of G gives columns of U as eigenvectors and diagonal entries of $\Sigma\Sigma^T$ as eigenvalues.
- In other words, left singular vectors of A are eigenvectors of AA^T and $\sigma^2 = \lambda$ (eigenvalue of AA^T). Furthermore, $\lambda \geq 0$ since $G = AA^T$ is PSD.

Eigenvalue decomposition of AA^T gives m left singular vectors of A and first m singular values.

LUMS
A Not-for-Profit University

Singular Value Decomposition

How to Compute SVD?

- Now we define a matrix $G = A^T A$.
- Using $A = U \Sigma V^T$, we can write G as

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

- What is special about matrix G?
    - G is symmetric by definition.
    - G is positive semi-definite. *How? You are fully equipped to show this.*
- We note that $\Sigma^T \Sigma$ is a diagonal matrix of size $n \times n$.
- Eigenvalue decomposition of G gives columns of V as eigenvectors and diagonal entries of $\Sigma^T \Sigma$ as eigenvalues.
- In other words, right singular vectors of A are eigenvectors of $A^T A$ and $\sigma^2 = \lambda$ (eigenvalue of $A^T A$). Furthermore, $\lambda \geq 0$ since G is PSD.

Eigenvalue decomposition of $A^T A$ gives n right singular vectors of A and first n singular values.

Now you can explain the non-negativity of the singular values.

Singular Value Decomposition

SVD Summary

- Singular value decomposition (SVD) of a matrix $A \in \mathbf{R}^{m\times n}$ is given by

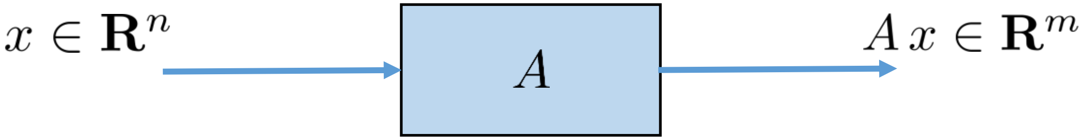
$$A = U\,\Sigma\,V^T$$

- EVD of AA^T gives U and first m singular values.
- EVD of A^TA gives V and first n singular values.
- U and V are always orthogonal.
- SVD always exists.
- Singular values are non-negative, that is,

$$\sigma_1 \geq \sigma_2 \geq \ \ldots \ \geq \sigma_{\min(m,n)} \geq 0$$

Singular Value Decomposition

Geometric Interpretation



Transformation (Change of Basis) | Scaling (In new Basis) | Transformation (Change of Basis)

Change of basis in $\mathbf{R}^n$

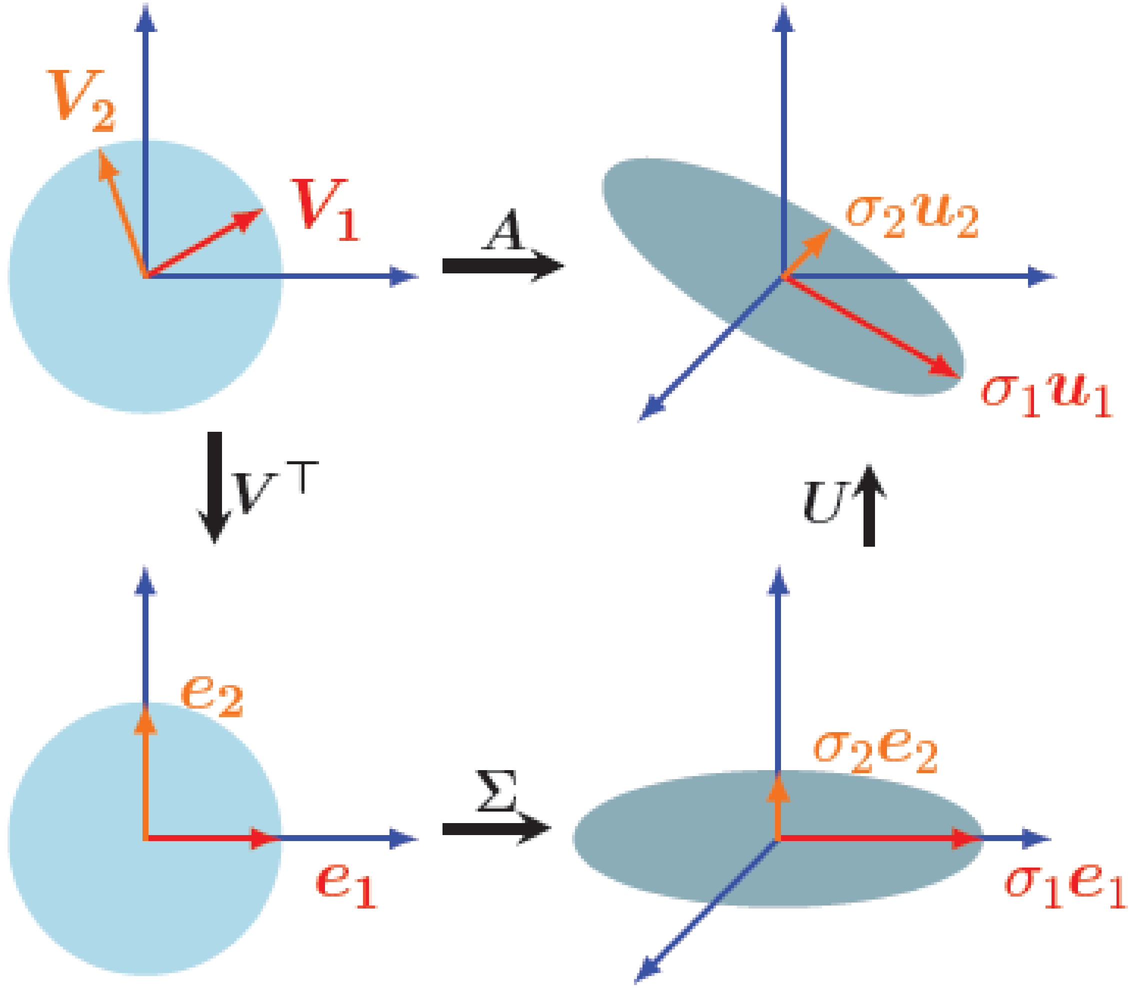
Change of basis in $\mathbf{R}^m$

Scaling along the new basis by singular values.

- $m < n$ - Drop the last $n - m$ basis (impact of columns of zeros in Σ)
- $m > n$ - Append $m - n$ basis (impact of rows of zeros in Σ)

Singular Value Decomposition

Geometric Interpretation

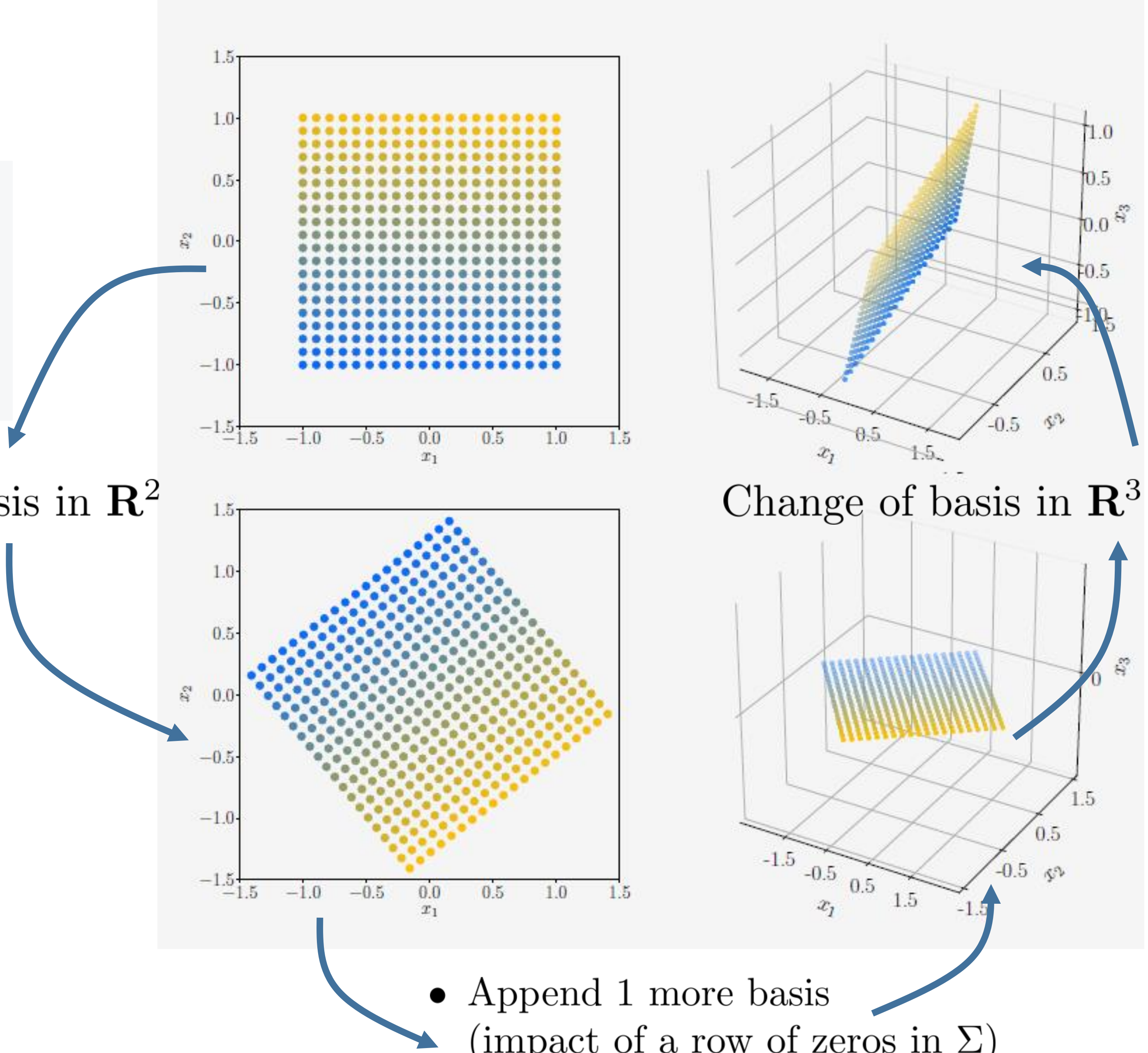


Singular Value Decomposition

Geometric Interpretation - Example

$$A = \begin{bmatrix} 1 & -0.8 \\ 0 & 1 \\ 1 & 0 \end{bmatrix} = U\Sigma V^\top$$

$$= \begin{bmatrix} -0.79 & 0 & -0.62 \\ 0.38 & -0.78 & -0.49 \\ -0.48 & -0.62 & 0.62 \end{bmatrix} \begin{bmatrix} 1.62 & 0 \\ 0 & 1.0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} -0.78 & 0.62 \\ -0.62 & -0.78 \end{bmatrix}$$



- Append 1 more basis (impact of a row of zeros in Σ)

Singular Value Decomposition

Rank of a Matrix:

- The rank of a matrix is equal to the number of non-zero singular values.

How?

- Since $A^T A$ and A have the same rank and we know that the rank of any square matrix equals the number of nonzero eigenvalues.

Application Example – Rank Estimation:

We use SVD for the estimation of rank while analyzing data. **How?**

- Suppose that we have n data points $a_1, a_2, \ldots, a_n$, all of which live in $\mathbf{R}^m$, where n is much larger than m. Let A be the $m \times n$ matrix with columns $a_1, a_2, \ldots, a_n$.
- Assume that the the data points satisfy some linear relations, such that $a_1, a_2, \ldots, a_n$ all lie in an r dimensional subspace of $\mathbf{R}^m$. Then we would expect the matrix A to have rank r.
- If the data points are obtained from measurements with errors, then the matrix A will probably have full rank m. But only r of the singular values of A will be large, and the other singular values will be close to zero.
- Using SVD, we can can estimate an "approximate rank" of A by counting the number of singular values which are much larger than the others.

Singular Value Decomposition

Application: Matrix Approximation

- A matrix $A \in \mathbf{R}^{m \times n}$ can be decomposed using SVD as

$$A = U \Sigma V^T = \sum_{i=1}^{\min(m,n)} u_i \sigma_i v_i^T$$

- If rank of a matrix is $r \leq \min(m, n)$, we can truncate the summationion at r

$$A = \sum_{i=1}^{r} u_i \sigma_i v_i^T$$

- Using SVD formulation, we can define k rank approximation of the matrix A by including first k singular vectors and associated singular values in the representation, that is,

$$A \approx \sum_{i=1}^{k} u_i \sigma_i v_i^T$$

(k-rank approximation)

Outline

- Positive/negative definite and semi-definite matrices
- Singular Value Decomposition (SVD)
  - Formulation
  - Interpretation
  - Application examples
- Column space and Null Space

Column Space and Null Space

Column Space:

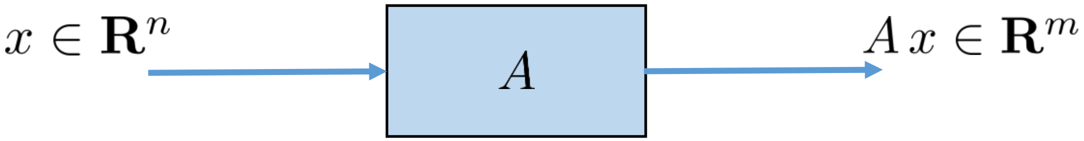
- For a matrix $A \in \mathbf{R}^{m\times n}$, the column space, denoted by $\mathcal{C}(A)$, is the span of the columns of A.
- If $a_1, a_2, \ldots, a_n \in \mathbf{R}^m$ are the columns of A, column space is given by

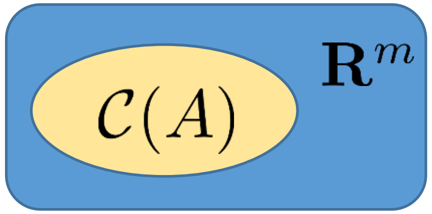
$$\mathcal{C}(A) = \text{span}(a_1, a_2, \ldots, a_n)$$

$$\mathcal{C}(A) = \{Ax | x \in \mathbf{R}^n\}$$

(all possible linear combinations of columns of A)

- In other words, column space is a linear transformation of every point in $\mathbf{R}^n$, that is,





- Consequently, $\mathcal{C}(A)$ is the subspace of $\mathbf{R}^m$.
- What is the dimension of column space $\mathcal{C}(A)$? Number of linearly independent columns of A=rank(A).

Column Space and Null Space

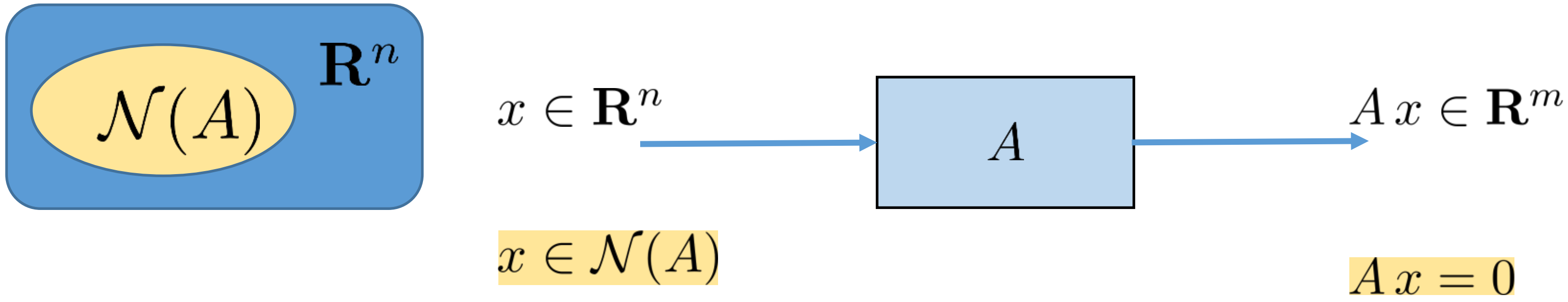
Null Space:

- For a matrix $A \in \mathbf{R}^{m \times n}$, the null space, denoted by $\mathcal{N}(A)$, is the subspace of $\mathbf{R}^n$ such that

$$\mathcal{N}(A) = \{x \in \mathbf{R}^n \mid Ax = 0\}$$

(all points that are mapped to zero by matrix A

- In other words, null space is an inverse linear transformation of $0 \in \mathbf{R}^m$.



- Nullity of the matrix, that is, the dimension of the null-space $\mathcal{N}(A)$ is given by the following rank-nullity theorem (also known as rank+nullity theorem).

$$\text{rank}(A) + \text{nullity}(A) = \text{ number of columns of } A$$

Column Space and Null Space

Example:

$$A = \begin{bmatrix} 1 & 1 & 2 \\ 2 & 1 & 3 \\ 3 & 1 & 4 \\ 4 & 1 & 5 \end{bmatrix}$$

- $m = 4$, $n = 3$
- $\mathcal{C}(A)$ is a subspace of $\mathbf{R}^4$.
- $\mathcal{N}(A)$ is a subspace of $\mathbf{R}^3$.

- Note that a third column is a sum of first two columns and therefore number of linearly independent columns is equal to 2.
- Consequently, $\mathcal{C}(A)$ is a 2-dimensional subspace of $\mathbf{R}^4$.