IHCL Linear Algebra Handout

February 14, 2022

1 Vector Space

1.1 Vector Space

- Definition: Let F be a field. A vector space V over F is defined to be a set of elements, called vectors, with two operations called (vector) addition and scalar multiplication, satisfying the following conditions:
  - Vector addition is closed; if $\boldsymbol{a}, \boldsymbol{b} \in V$, then $\boldsymbol{a} + \boldsymbol{b} \in V$.
  - Vector addition is associative; if $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c} \in V$, then $(\boldsymbol{a} + \boldsymbol{b}) + \boldsymbol{c} = \boldsymbol{a} + (\boldsymbol{b} + \boldsymbol{c})$.
  - Vector addition identity exists; an identity element, denoted by $\mathbf{0}$ (zero vector) exists in V such that $\boldsymbol{a} + \mathbf{0} = \boldsymbol{a}$ for all $\boldsymbol{a} \in V$.
  - Vector inverse exists for each element in V; if $\boldsymbol{a} \in V$, then an element, called $-\boldsymbol{a}$, exists such that $\boldsymbol{a} + (-\boldsymbol{a}) = \mathbf{0}$.
  - Vector addition is commutative; if $\boldsymbol{a}, \boldsymbol{b} \in V$, $\boldsymbol{a} + \boldsymbol{b} = \boldsymbol{b} + \boldsymbol{a}$.
  - Scalar multiplication is closed; if $c \in F$ and $\boldsymbol{a} \in V$, then $c\boldsymbol{a} \in V$.
  - Scalar multiplication is associative; if $c \in F$ and $\boldsymbol{a}, \boldsymbol{b} \in V$, $c(\boldsymbol{a}\boldsymbol{b}) = (c\boldsymbol{a})\boldsymbol{b}$.
  - Multiplication by the scalar 1 is the identity operation; $1\boldsymbol{a} = \boldsymbol{a}$ for all $\boldsymbol{a} \in V$.
  - Scalar multiplication is distributive w.r.t. vector addition; if $c \in F$ and $\boldsymbol{a}, \boldsymbol{b} \in V$, $c(\boldsymbol{a} + \boldsymbol{b}) = c\boldsymbol{a} + c\boldsymbol{b}$.
  - Scalar multiplication is distributive w.r.t. scalar addition; if $c_1, c_2 \in F$ and $\boldsymbol{a} \in V$, $(c_1 + c_2)\boldsymbol{a} = c_1\boldsymbol{a} + c_2\boldsymbol{a}$.

1.2 Subspace

- Definition: If W is a nonempty subset of vector space V over field F, and W satisfies the following conditions:
  - Vector addition is closed; if $\boldsymbol{u}, \boldsymbol{v} \in W$, then $\boldsymbol{u} + \boldsymbol{v} \in W$.
  - Scalar multiplication is closed; if $c \in F$ and $\boldsymbol{u} \in W$, then $c\boldsymbol{u} \in W$.
  - Zero vector $\mathbf{0} \in W$.

1.3 Affine Set

- A set C is called an affine set, if the line passing through any two points $x_1, x_2 \in C$ still belongs to C; precisely,

$$\theta x_1 + (1-\theta)x_2 \in C, \forall\theta \in \mathbb{R}, x_1, x_2 \in C.$$

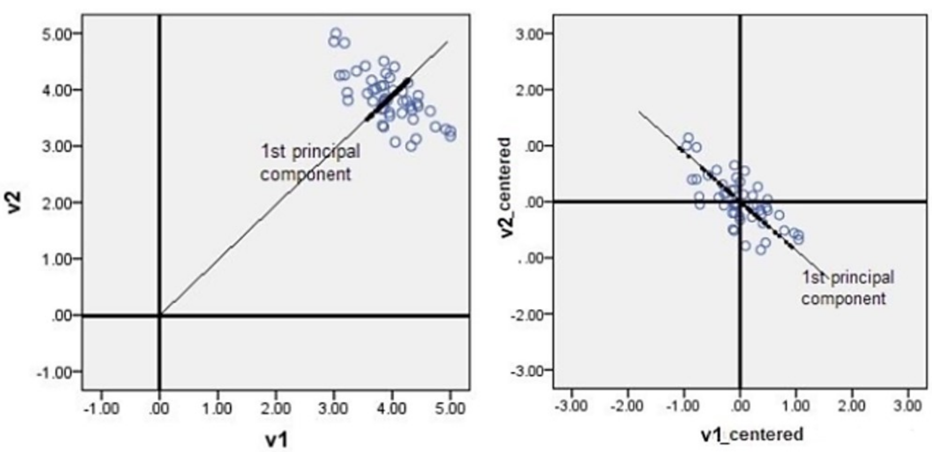
- (Example) Affine sets: a point, a line, a plan, a 3-dimensional subspace,...
- (Example) A ball, a ray and a half-space, are not affine sets.
- It can be extended to more than two points when the cofficients satisfy $\theta_1 + \cdots + \theta_k = 1$.
- An affine set $C \subseteq \mathbb{R}^n$ can also be represented as

$$C = \{\boldsymbol{A}\boldsymbol{x} + \boldsymbol{d} \mid \boldsymbol{x} \in \mathbb{R}^r\},$$

where $\boldsymbol{d} \in C$, $r \leq n$ is the dimension of the subspace $C - \boldsymbol{d}$, and columns of $\boldsymbol{A} \in \mathbb{R}^{n\times r}$ from an orthonormal basis of the subspace $C - \boldsymbol{d}$ (i.e., $\boldsymbol{A}^T\boldsymbol{A} = \boldsymbol{I}_p$).

1.4 Application

- First step of PCA(Principal component analysis)



2 Positive definite/Negative definite Positive Semidefinite/Negative Semidefinite matrix

2.1 Definition

- Let $\boldsymbol{A} \in \mathbb{C}^{n\times n}$, then $\boldsymbol{A}$ is a positive definite matrix ($\boldsymbol{A} \in \mathbb{S}^n_{++}$) if and only if the scalar $\boldsymbol{x}^H\boldsymbol{A}\boldsymbol{x}$ is positive for all $\boldsymbol{x} \in \mathbb{C}^n \backslash \{\boldsymbol{0}\}$, i.e.,

$$\boldsymbol{A} \in \mathbb{S}^n_{++} \iff \boldsymbol{x}^H\boldsymbol{A}\boldsymbol{x} > 0, \forall \boldsymbol{x} \in \mathbb{C}^n \backslash \{\boldsymbol{0}\}.$$

- Let $\boldsymbol{A} \in \mathbb{C}^{n\times n}$, then $\boldsymbol{A}$ is a positive semidefinite matrix ($\boldsymbol{A} \in \mathbb{S}^n_{+}$) if and only if the scalar $\boldsymbol{x}^H\boldsymbol{A}\boldsymbol{x}$ must be non-negative for all $\boldsymbol{x} \in \mathbb{C}^n \backslash \{\boldsymbol{0}\}$, i.e.,

$$\boldsymbol{A} \in \mathbb{S}^n_{+} \iff \boldsymbol{x}^H\boldsymbol{A}\boldsymbol{x} \geq 0, \forall \boldsymbol{x} \in \mathbb{C}^n \backslash \{\boldsymbol{0}\}.$$

- We can change the direction of the inequalities above to get the concepts that $\boldsymbol{A} \in \mathbb{C}^{n\times n}$ is negative definite or negative semidefinite, i.e.,

$$\boldsymbol{A} \in -\mathbb{S}_{++}^{n} \iff \boldsymbol{x}^H \boldsymbol{A}\boldsymbol{x} < 0, \forall \boldsymbol{x} \in \mathbb{C}^n \backslash \{\boldsymbol{0}\},$$

$$\boldsymbol{A} \in -\mathbb{S}_{+}^{n} \iff \boldsymbol{x}^H \boldsymbol{A}\boldsymbol{x} \leq 0, \forall \boldsymbol{x} \in \mathbb{C}^n \backslash \{\boldsymbol{0}\}.$$

  By definitions, we can know that:

  - If $\boldsymbol{A}$ is negative definite, then $-\boldsymbol{A}$ is positive definite.
  - If $\boldsymbol{A}$ is negative simidefinite, then $-\boldsymbol{A}$ is positive simidefinite.

- Let $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ and $\boldsymbol{A}$ is a symmetric matrix, the eigenvalues of $\boldsymbol{A}$ are all positive is a sufficient and necessary condition for $\boldsymbol{A}$ to be a positive definite matrix.

2.2 Significance

For better understanding the meaning of positive definite matrix, in this subsection we only consider the case that $\boldsymbol{A} \in \mathbb{R}^{n\times n}$. The essence of multiplication between matrix and vector is to transform the vector $\boldsymbol{x}$ in the manner specified by matrix $\boldsymbol{A}$. For positive definite matrix $\boldsymbol{A}$, it must satisfy

$$\boldsymbol{x}^T \boldsymbol{A}\boldsymbol{x} > 0.$$

Recall the formula of cosine,

$$\cos\theta = \frac{\boldsymbol{c}^T \boldsymbol{d}}{\|\boldsymbol{c}\| \cdot \|\boldsymbol{d}\|}.$$

Then substitute $\boldsymbol{c} = \boldsymbol{x}$, $\boldsymbol{d} = \boldsymbol{A}\boldsymbol{x}$ into the cosine formula, we can get

$$\cos\theta = \frac{\boldsymbol{x}^T \boldsymbol{A}\boldsymbol{x}}{\|\boldsymbol{x}\| \cdot \|\boldsymbol{A}\boldsymbol{x}\|} > 0 \Longrightarrow 0^\circ < \theta < 90^\circ.$$

The inner product of $\boldsymbol{A}$ and $\boldsymbol{A}\boldsymbol{x}$ is positive represents that the angle between the linearly transformed vector $\boldsymbol{A}\boldsymbol{x}$ and the original vector $\boldsymbol{x}$ is less than 90°. The meaning behind it is the direction is unchanged when $\boldsymbol{A}\boldsymbol{x}$ projecting back to the original vector.

2.3 Application

Positive definite or positive semidefinite matrices are closely related to singular value decomposition (SVD). For an arbitrary matrix $\boldsymbol{A} \in \mathbb{R}^{m\times n}$, the eigenvalues of its interactive product $\boldsymbol{A}^H\boldsymbol{A}$ and $\boldsymbol{A}\boldsymbol{A}^H$ are all not negative, so that $\boldsymbol{A}^H\boldsymbol{A}$ and $\boldsymbol{A}\boldsymbol{A}^H$ are positive semidefinite matrices.

3 Eigenvalue Decompsition (EVD)

3.1 Definiton

A vector $\boldsymbol{v} \in \mathbb{C}^n \backslash \{\mathbf{0}\}$ is an eigenvector of $\boldsymbol{A} \in \mathbb{C}^{n \times n}$ if it satisfies the linear equation:

$$\boldsymbol{A}\boldsymbol{v} = \lambda \boldsymbol{v},$$

where λ is a scalar, termed the eigenvalue corresponding to $\boldsymbol{v}$. The above equation is called the eigenvalue equation or the eigenvalue problem. Because $\boldsymbol{v}$ is a non-zero vector, we can derive the following equation:

$$p(\lambda) = \det(\boldsymbol{A} - \lambda \boldsymbol{I}) = 0.$$

We call $p(\lambda)$ the characteristic polynomial, and call the equation the characteristic equation.

3.2 Concept

As we mentioned before, λ is a scalar in the eigenvalue problem. We can consider $\boldsymbol{A}$ as a linear transform function that elongates or shrinks the eigenvector $\boldsymbol{v}$ by the eigenvalue. We can find the eigenvalue by solving $p(\lambda)$, and find the eigenvector by solving the eigenvalue problem, i.e.,

$$(\boldsymbol{A} - \lambda I)\boldsymbol{v} = \mathbf{0}.$$

After we find the eigenvalue and eigenvector, we can factorize the matrix $\boldsymbol{A}$ as

$$\boldsymbol{A} = \boldsymbol{Q}\boldsymbol{\Lambda}\boldsymbol{Q}^{-1},$$

where $\boldsymbol{Q} \in \mathbb{C}^{n \times n}$ combined by the eigenvector, i.e.,

$$\boldsymbol{Q} = [\boldsymbol{v}_1, ..., \boldsymbol{v}_n],$$

and $\boldsymbol{\Lambda}$ is the diagonal matrix whose diagonal elements are the corresponding eigenvalues, i.e.,

$$\boldsymbol{\Lambda} = \begin{bmatrix} \lambda_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \lambda_n \end{bmatrix}.$$

Note that only diagonalizable matrices can be factorized in this way.

3.3 Special Matrices

For every $\boldsymbol{A} \in \mathbb{S}^n$, the eigenvalues are real and the eigenvectors can be chosen real and orthonormal. We can decomposed $\boldsymbol{A}$ as

$$\boldsymbol{A} = \boldsymbol{Q}\boldsymbol{\Lambda}\boldsymbol{Q}^T,$$

where $\boldsymbol{Q}$ is an orthogonal matrix, and $\boldsymbol{\Lambda}$ is a diagonal matrix.

3.4 Example

Consider the Eigenvalue Decomposition (EVD) of following matrix

$$\boldsymbol{A} = \begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix}.$$

$<$ *Solution* $>$

By adopting the formula $p(\lambda) = det(\boldsymbol{A} - \lambda \boldsymbol{I}) = 0$, we can find the normalized eigenvector respect to the eigenvalue:

$$\lambda = 1,\ \boldsymbol{v}_1 = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \\ 0 \end{bmatrix}$$

$$\lambda = 2,\ \boldsymbol{v}_2 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

$$\lambda = 3,\ \boldsymbol{v}_3 = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \\ 0 \end{bmatrix}.$$

Therefore, the diagonal matrix $\boldsymbol{\Lambda}$ and the unitary matrix $\boldsymbol{U}$ are

$$\boldsymbol{\Lambda} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix},\ \boldsymbol{U} = \begin{bmatrix} \frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & 0 & -\frac{1}{\sqrt{2}} \\ 0 & 1 & 0 \end{bmatrix},$$

where $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^{\mathrm{H}}$.

4 Singular Value Decomposition (SVD)

4.1 Definition

For any matrix $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ can be factorized as

$$\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T.$$

- $\boldsymbol{U} \in \mathbb{R}^{m \times m}$ is an orthogonal matrix whose columns are the eigenvectors of $\boldsymbol{A}\boldsymbol{A}^T$
- $\boldsymbol{V} \in \mathbb{R}^{n \times n}$ is an orthogonal matrix whose columns are the eigenvectors of $\boldsymbol{A}^T\boldsymbol{A}$
- $\boldsymbol{\Sigma} \in \mathbb{R}^{m \times n}$ is an all zero matrix except for the first r diagonal elements $\sigma_i = \Sigma_{ii}, i = 1, ..., r$ (called singular values) that are the square roots of the eigenvalues of $\boldsymbol{A}^T\boldsymbol{A}$ and of $\boldsymbol{A}\boldsymbol{A}^T$, where r denotes the rank of $\boldsymbol{A}$

4.2 Concept

The matrix $\boldsymbol{AA}^T$ and $\boldsymbol{A}^T\boldsymbol{A}$ are very special in linear algebra. These matrices are

- symmetrical
- square
- at least positive semidefinite (eigenvalues are zero or positive)
- both matrices have the same positive eigenvalues
- both have the same rank r as $\boldsymbol{A}$

We name the eigenvectors for $\boldsymbol{AA}^T$ as $\boldsymbol{u}_i$ and $\boldsymbol{A}^T\boldsymbol{A}$ as $\boldsymbol{v}_i$ here and call these sets of eigenvectors $\boldsymbol{u}$ and $\boldsymbol{v}$ the singular vectors of $\boldsymbol{A}$. Both matrices have the same positive eigenvalues. We concatenate vectors $\boldsymbol{u}_i$ into $\boldsymbol{U}$ and $\boldsymbol{v}_i$ into $\boldsymbol{V}$ to form orthogonal matrices.

Comparing to EVD, SVD works on singular and non-square matrices. $\boldsymbol{U}$ and $\boldsymbol{V}$ are invertible for any matrix in SVD and they are orthonormal. SVD reveals useful information about $\boldsymbol{A}$ which is used to solve many linear algebra problems including to obtain low-rank matrix approximations and to find the pseudo-inverse of non-square matrices. A statistical analysis algorithm known as Principal Component Analysis (PCA) relies on SVD.

4.3 Example

Consider the Singular Value Decomposition (SVD) of following matrix

$$\boldsymbol{A}=\begin{bmatrix} 3 & 2 & 2 \\ 2 & 3 & -2 \end{bmatrix}.$$

< *Solution* >

1. $\boldsymbol{AA}^T=\begin{bmatrix} 17 & 8 \\ 8 & 17 \end{bmatrix}$

   $\det(\boldsymbol{AA}^T - \lambda \boldsymbol{I}) = 0$

   eigenvalues: $\lambda_1 = 25$, $\lambda_2 = 9$

   eigenvectors: $\boldsymbol{u}_1 = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}$, $\boldsymbol{u}_2 = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{-1}{\sqrt{2}} \end{bmatrix}$.

2. $\boldsymbol{A}^T\boldsymbol{A}=\begin{bmatrix} 13 & 12 & 2 \\ 12 & 13 & -2 \\ 2 & -2 & 8 \end{bmatrix}$

$\det(\boldsymbol{A}^T\boldsymbol{A} - \lambda\boldsymbol{I}) = 0$

eigenvalues: $\lambda_1 = 25$, $\lambda_2 = 9$, $\lambda_3 = 0$

eigenvectors: $\boldsymbol{v}_1 = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \\ 0 \end{bmatrix}, \ \boldsymbol{v}_2 = \begin{bmatrix} \frac{-1}{\sqrt{18}} \\ \frac{1}{\sqrt{18}} \\ \frac{-4}{\sqrt{18}} \end{bmatrix}, \ \boldsymbol{v}_3 = \begin{bmatrix} \frac{2}{3} \\ \frac{-2}{3} \\ \frac{-1}{3} \end{bmatrix}.$

The singular values are the square root of positive eigenvalues, i.e., $\sqrt{25}$ and $\sqrt{9}$. Therefore, the SVD composition is

$$\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} 5 & 0 & 0 \\ 0 & 3 & 0 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{18}} & \frac{2}{3} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{18}} & \frac{-2}{3} \\ 0 & \frac{-4}{\sqrt{18}} & \frac{-1}{3} \end{bmatrix}.$$

5 Pseudo Inverse Matrix

If $\boldsymbol{A} \in \mathbb{R}^{m \times m}$ is a full rank square matrix, then the inverse of $\boldsymbol{A}$ exists and

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

has a solution

$$\boldsymbol{x} = \boldsymbol{A}^{-1}\boldsymbol{b}.$$

For any given matrix $\boldsymbol{A}$ (i.e., non-square or singular), it is possible to define unique pseudo inverse matrix $\boldsymbol{A}^\dagger$.

5.1 Definition

For $\boldsymbol{A} \in \mathbb{C}^{m \times n}$, pseudo inverse of $\boldsymbol{A}$ is defined as $\boldsymbol{A}^\dagger \in \mathbb{C}^{n \times m}$, satisfying Moore–Penrose conditions:

- $\boldsymbol{A}\boldsymbol{A}^\dagger\boldsymbol{A} = \boldsymbol{A}$
- $\boldsymbol{A}^\dagger\boldsymbol{A}\boldsymbol{A}^\dagger = \boldsymbol{A}^\dagger$
- $(\boldsymbol{A}\boldsymbol{A}^\dagger)^H = \boldsymbol{A}\boldsymbol{A}^\dagger$
- $(\boldsymbol{A}^\dagger\boldsymbol{A})^H = \boldsymbol{A}^\dagger\boldsymbol{A}$

$\boldsymbol{A}^\dagger$ exists for any matrix $\boldsymbol{A}$, when $\boldsymbol{A}$ has full rank (i.e., the rank of $\boldsymbol{A}$ is $\min\{m, n\}$), then $\boldsymbol{A}^\dagger$ can be expressed as a simple algebraic formula, where H is the Hermitian operator (or conjugate transpose).

In particular, when $\boldsymbol{A}$ has linearly independent columns (and thus matrix $\boldsymbol{A}^T\boldsymbol{A}$ is invertible),

$$\boldsymbol{A}^\dagger = (\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^T.$$

This particular pseudo inverse constitutes a left inverse, inducing $\boldsymbol{A}^\dagger\boldsymbol{A} = \boldsymbol{I}$.

When $\boldsymbol{A}$ has linearly independent rows (matrix $\boldsymbol{A}\boldsymbol{A}^T$ is invertible),

$$\boldsymbol{A}^\dagger = \boldsymbol{A}^T(\boldsymbol{A}\boldsymbol{A}^T)^{-1}.$$

This is a right inverse, inducing $\boldsymbol{A}\boldsymbol{A}^\dagger = \boldsymbol{I}$.

5.2 Derivation

Consider a linear equation $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$, $\boldsymbol{A} \in \mathbb{R}^{m\times n}$, if columns of $\boldsymbol{A}$ are linearly independent , i.e., $m>n$, we know that

$$\begin{aligned}
\boldsymbol{A}\boldsymbol{x} &= \boldsymbol{b}\\
\boldsymbol{A}^T\boldsymbol{A}\boldsymbol{x} &= \boldsymbol{A}^T\boldsymbol{b}\\
(\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^T\boldsymbol{A}\boldsymbol{x} &= (\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^T\boldsymbol{b}\\
\boldsymbol{x} &= (\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^T\boldsymbol{b}\\
\boldsymbol{x} &= \boldsymbol{A}^\dagger\boldsymbol{b},\\
\text{where } \boldsymbol{A}^\dagger &= (\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^T.
\end{aligned}$$

Then, $\boldsymbol{A}^\dagger$ is called pseudo inverse matrix of $\boldsymbol{A}$.

5.3 Calculated by SVD

For $\boldsymbol{A} \in \mathbb{R}^{m\times n}$, let $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^T$ is SVD of $\boldsymbol{A}$. We can easily find out $\boldsymbol{A}^\dagger$ by SVD. From the definition of $\boldsymbol{A}^\dagger$, substitute $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^T$ into $\boldsymbol{A}^\dagger = (\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^T$, then

$$\begin{aligned}
\boldsymbol{A}^\dagger &= (\boldsymbol{V}\boldsymbol{S}^T\boldsymbol{U}^T\boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^T)^{-1}(\boldsymbol{V}\boldsymbol{S}^T\boldsymbol{U}^T)\\
&= (\boldsymbol{V}\boldsymbol{S}^T\boldsymbol{S}\boldsymbol{V}^T)^{-1}(\boldsymbol{V}\boldsymbol{S}^T\boldsymbol{U}^T)\\
&= (\boldsymbol{V}^T)^{-1}(\boldsymbol{S}^T\boldsymbol{S})^{-1}\boldsymbol{V}^{-1}\boldsymbol{V}\boldsymbol{S}^T\boldsymbol{U}^T\\
&= \boldsymbol{V}(\boldsymbol{S}^T\boldsymbol{S})^{-1}\boldsymbol{S}^T\boldsymbol{U}^T\\
&= \boldsymbol{V}\boldsymbol{S}^\dagger\boldsymbol{U}^T,
\end{aligned}$$

$$\text{where } \boldsymbol{S}^\dagger = \begin{bmatrix} \sigma_1^{-1} & 0 & \cdots & 0 & \cdots & 0\\ 0 & \sigma_2^{-1} & \cdots & 0 & \cdots & 0\\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & 0 & \sigma_n^{-1} & \cdots & 0 \end{bmatrix} \in \mathbb{R}^{n\times m}, \text{ and } \sigma_i,\ \forall i \in \{1,\dots,n\} \text{ are singular values of } \boldsymbol{S}.$$

5.4 Example

$\boldsymbol{A} = \begin{bmatrix} 63 & 27 & 92 \\ 82 & -81 & 93 \\ -75 & -45 & -69 \\ 83 & 9 & 95 \end{bmatrix}$, find its pseudo inverse matrix $\boldsymbol{A}^\dagger$.

$<$ *Solution* 1 $>$

$$\boldsymbol{A}^\dagger = (\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^T \Rightarrow \boldsymbol{A}^\dagger = \begin{bmatrix} -0.0319 & 0.0064 & -0.0300 & 0.0028 \\ 0.0036 & -0.0084 & -0.0047 & 0.0013 \\ 0.0304 & -0.0027 & 0.0233 & 7.14\times 10^{-4} \end{bmatrix}.$$

$<$ *Solution* 2 $>$

For $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^T$ is SVD of $\boldsymbol{A}$, $\boldsymbol{A}^\dagger = \boldsymbol{V}\boldsymbol{S}^\dagger\boldsymbol{U}^T$.

$$\boldsymbol{A} = \boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^T =$$

$$\begin{bmatrix} -0.4737 & -3.3127 & -0.7524 & -0.3343 \\ -0.5447 & 0.7977 & 0.1145 & -0.2321 \\ 0.4302 & -0.4982 & -0.6481 & 0.3831 \\ -0.5422 & -0.1329 & 0.0281 & 0.8292 \end{bmatrix} \begin{bmatrix} 232.0655 & 0 & 0 \\ 0 & 96.7233 & 0 \\ 0 & 0 & 17.0945 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} -0.6540 & -0.0278 & 0.7560 \\ 0.0306 & -0.9995 & -0.0103 \\ -0.7559 & -0.0164 & -0.6545 \end{bmatrix}^T,$$

Then,

$$\boldsymbol{A}^\dagger = \boldsymbol{V}\boldsymbol{S}^\dagger\boldsymbol{U}^T = \begin{bmatrix} -0.0319 & 0.0064 & -0.0300 & 0.0028 \\ 0.0036 & -0.0084 & -0.0047 & 0.0013 \\ 0.0304 & -0.0027 & 0.0233 & 7.14\times 10^{-4} \end{bmatrix}.$$

$<$ *Solution* 3 $>$

$$\text{Matlab code : AA = pinv(A)} \Rightarrow \boldsymbol{AA} = \begin{bmatrix} -0.0319 & 0.0064 & -0.0300 & 0.0028 \\ 0.0036 & -0.0084 & -0.0047 & 0.0013 \\ 0.0304 & -0.0027 & 0.0233 & 7.14\times 10^{-4} \end{bmatrix}.$$

6 Projections

From Figure 1, the closest point $\boldsymbol{p}$ is at the intersection formed by a line through $\boldsymbol{b}$ that is orthogonal to $\boldsymbol{a}$. If we think of $\boldsymbol{p}$ as an approximation of $\boldsymbol{b}$, then the length of $\boldsymbol{e} = \boldsymbol{b} - \boldsymbol{p}$ is the error in that approximation. We could try to find $\boldsymbol{p}$ using trigonometry or calculus, but it's easier to use linear

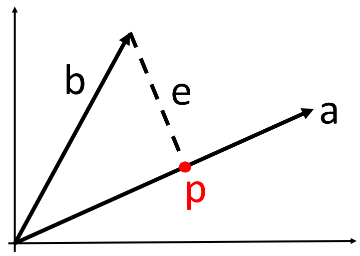


Figure 1. The point closest to $\boldsymbol{b}$ on the line determined by $\boldsymbol{a}$.

algebra. Since $\boldsymbol{p}$ lies on the line through $\boldsymbol{a}$, we know $\boldsymbol{p} = x\boldsymbol{a}$ for some number x. We also know that $\boldsymbol{a}$ is perpendicular to $\boldsymbol{e} = \boldsymbol{b} - x\boldsymbol{a}$:

$$\begin{aligned} \boldsymbol{a}^T(\boldsymbol{b} - x\boldsymbol{a}) &= 0 \\ x\boldsymbol{a}^T\boldsymbol{a} &= \boldsymbol{a}^T\boldsymbol{b} \\ x &= \frac{\boldsymbol{a}^T\boldsymbol{b}}{\boldsymbol{a}^T\boldsymbol{a}}, \end{aligned}$$

and $\boldsymbol{p} = \boldsymbol{a}x = \boldsymbol{a}\frac{\boldsymbol{a}^T\boldsymbol{b}}{\boldsymbol{a}^T\boldsymbol{a}}$.

6.1 Projection matrix

We'd like to write this projection in terms of a projection matrix $\boldsymbol{P} : \boldsymbol{p} = \boldsymbol{P}\boldsymbol{b}$.

$$\boldsymbol{p} = \boldsymbol{a}x = \frac{\boldsymbol{a}\boldsymbol{a}^T\boldsymbol{b}}{\boldsymbol{a}^T\boldsymbol{a}},$$

so the projection matrix $\boldsymbol{P}$ is:

$$\boldsymbol{P} = \frac{\boldsymbol{a}\boldsymbol{a}^T}{\boldsymbol{a}^T\boldsymbol{a}}.$$

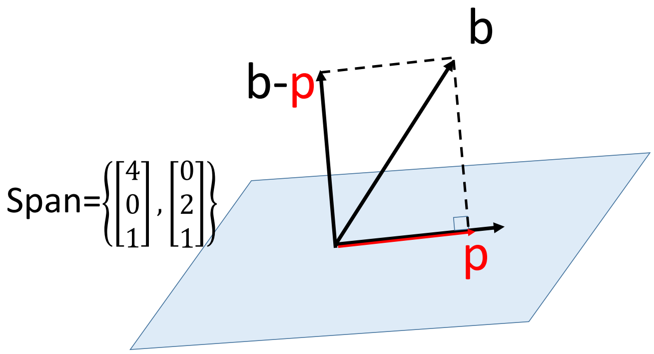
The column space of $\boldsymbol{P}$ is spanned by $\boldsymbol{a}$ because for any $\boldsymbol{b}$, $\boldsymbol{P}\boldsymbol{b}$ lies on the line determined by $\boldsymbol{a}$. $\boldsymbol{P}$ is symmetric. $\boldsymbol{P}^2\boldsymbol{b} = \boldsymbol{P}\boldsymbol{b}$ because the projection of a vector already on the line through $\boldsymbol{a}$ is just that vector. In general, projection matrices have the properties:

$$\boldsymbol{P}^T = \boldsymbol{P} \quad \text{and} \quad \boldsymbol{P}^2 = \boldsymbol{P}.$$

6.2 Example - Projection in higher dimensions

Find the orthogonal projection $\boldsymbol{p}$ of $\boldsymbol{b} = \begin{bmatrix} 2 \\ 0 \\ 11 \end{bmatrix}$ onto $\textbf{Span}\left\{ \begin{bmatrix} 4 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 2 \\ 1 \end{bmatrix} \right\}$.

< *Solution* >



$$\boldsymbol{b}-\boldsymbol{p}=\begin{bmatrix}4\\0\\1\end{bmatrix}\times\begin{bmatrix}0\\2\\1\end{bmatrix}=\begin{bmatrix}-2\\-4\\8\end{bmatrix},\quad \boldsymbol{p}=\begin{bmatrix}2\\0\\11\end{bmatrix}-\begin{bmatrix}-2\\-4\\8\end{bmatrix}=\begin{bmatrix}4\\4\\3\end{bmatrix}.$$

6.3 Application

- SPA (Successive Projection Algorithm).
- Dimension reduction in machine learning.

7 Kronecker Product

7.1 Definition

Assume $\boldsymbol{A}\in\mathbb{R}^{m\times n}$, $\boldsymbol{B}\in\mathbb{R}^{p\times q}$, then Kronecker product $\boldsymbol{A}\otimes\boldsymbol{B}\in\mathbb{R}^{mp\times nq}$ is a block matrix which is defined as

$$\boldsymbol{A}\otimes\boldsymbol{B}=\begin{bmatrix}a_{11}\boldsymbol{B}&\cdots&a_{1n}\boldsymbol{B}\\\vdots&\ddots&\vdots\\a_{m1}\boldsymbol{B}&\cdots&mn\boldsymbol{B}\end{bmatrix},$$

where a_{ij} is the $(i,j)th$ element of $\boldsymbol{A}$.

7.2 Application

When converting an image into mathematical form, Kronecker product will appear naturally, and it can be used to get convenient representations for some matrix equations. For example, consider an equation:

$$\boldsymbol{AXB}=\boldsymbol{C},$$

it can rewrite by Kronecker product as

$$(\boldsymbol{B}^{\mathrm{T}} \otimes \boldsymbol{A})\mathrm{vec}(\boldsymbol{X}) = \mathrm{vec}(\boldsymbol{C}),$$

where $\mathrm{vec}(\boldsymbol{X})$ represents the vectorization of matrix $\boldsymbol{X}$.

7.3 Example - 1

Let $\boldsymbol{A} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$, $\boldsymbol{B} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$, then we get

$$\boldsymbol{A} \otimes \boldsymbol{B} = \begin{bmatrix} 1\times 1 & 1\times 0 & 2\times 1 & 2\times 0 \\ 1\times 0 & 1\times 1 & 2\times 0 & 2\times 1 \\ 3\times 1 & 3\times 0 & 4\times 1 & 4\times 0 \\ 3\times 0 & 3\times 1 & 4\times 0 & 4\times 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 2 & 0 \\ 0 & 1 & 0 & 2 \\ 3 & 0 & 4 & 0 \\ 0 & 3 & 0 & 4 \end{bmatrix}.$$

7.4 Example - 2

Let $\boldsymbol{A} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$, $\boldsymbol{B} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$, $\boldsymbol{C} = \begin{bmatrix} 2 & 1 \\ 4 & 3 \end{bmatrix}$, $\boldsymbol{X} = \begin{bmatrix} x_1 & x_2 \\ x_3 & x_4 \end{bmatrix}$, we get $\begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_3 \\ x_2 \\ x_4 \end{bmatrix} = \begin{bmatrix} 2 \\ 4 \\ 1 \\ 3 \end{bmatrix}$.

Then we can find $\mathrm{vec}(\boldsymbol{X}) = \begin{bmatrix} 1 \\ 3 \\ 2 \\ 4 \end{bmatrix}$, and revert it to matrix $\boldsymbol{X} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$.