Module 3: Supplementary Slides

(These additional materials are optional and intended for students who are interested)

Vectors

More Vector Arithmetic

```
m%%n Modulo operator (gives the remainder of m/n)
%/% Integer division (gives the integer part of m/n)
%*% Matrix multiplication (to be studied later)
%in% Returns TRUE if the left operand occurs in its right operand; FALSE otherwise
```

```
> 14%%5
[1] 4
> 14%%5
[1] 2
> 5%in%14
[1] FALSE
> 5%in%c(5,4)
[1] TRUE
```

General Norm of a Vector

Norm of a Vector

Definition: a norm for a vector $\boldsymbol{x} \in \mathbb{R}^n$ is a function $\|\boldsymbol{x}\|: \mathbb{R}^{\boldsymbol{n}} \to \mathbb{R}^+$ that satisfies the following properties:

1. $\|\boldsymbol{x}\| \geq 0$ and $\|\boldsymbol{x}\| = 0 \iff \boldsymbol{x} = 0$. *(Positive definiteness)*
2. $\|\lambda\boldsymbol{x}\| = |\lambda|\|\boldsymbol{x}\|, \quad \forall\lambda \in \mathbb{R}$ *(Homogeneity)*
3. $\|\boldsymbol{x} + \boldsymbol{y}\| \leq \|\boldsymbol{x}\| + \|\boldsymbol{y}\|$ *(Triangular inequality)*

A norm is a function that assigns a length to a vector. To compute the distance between two vectors, we calculate the norm of the difference between those two vectors. For example, the distance between two column vectors $\boldsymbol{x} \in \mathbb{R}^{\boldsymbol{n}}$ and $\boldsymbol{y} \in \mathbb{R}^{\boldsymbol{n}}$ using the Euclidean norm is

$$\|\boldsymbol{x} - \boldsymbol{y}\| = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \cdots + (x_n - y_n)^2} = (\boldsymbol{x} - \boldsymbol{y})^{\boldsymbol{T}}(\boldsymbol{x} - \boldsymbol{y})$$

Common Norms

$\boldsymbol{L_p}$ *norm* is a family of commonly used norms for vectors $\boldsymbol{x} \in \mathbb{R}^{\boldsymbol{n}}$ that are determined by a scalar $p \geq 1$ as:

$$\|\boldsymbol{x}\|_p = \sqrt[p]{|x_1|^p + |x_2|^p + \cdots + |x_n|^p}$$

Examples:

- $\boldsymbol{L_1}$ *norm:* $\|\boldsymbol{x}\|_1 = |x_1| + |x_2| + \cdots + |x_n|$ *(Manhattan/ City-block norm)*
- $\boldsymbol{L_2}$ *norm:* $\|\boldsymbol{x}\|_2 = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2} = \sqrt{\boldsymbol{x}^{\boldsymbol{T}}\boldsymbol{x}}$ *(Euclidean norm: we use only this)*
- $\boldsymbol{L_\infty}$ *norm:* $\|\boldsymbol{x}\|_\infty = \lim_{p\to\infty} \sqrt[p]{|x_1|^p + |x_2|^p + \cdots + |x_n|^p} = max(|x_1|, |x_2|, \dots, |x_n|)$ *(Maximum norm)*

Angle between Two Vectors

(i) If $\boldsymbol{u}$ and $\boldsymbol{v}$ are two unit vectors, then $\boldsymbol{u} \cdot \boldsymbol{U} = cos\theta$

(ii) Cosine Formula: If $\boldsymbol{a}$ and $\boldsymbol{b}$ are two nonzero vectors then $\frac{\boldsymbol{a} \cdot \boldsymbol{b}}{\|\boldsymbol{a}\|\|\boldsymbol{b}\|} = cos\theta$

(iii) Schwartz Inequality: If $\boldsymbol{a}$ and $\boldsymbol{b}$ are two nonzero vectors then $|\boldsymbol{a} \cdot \boldsymbol{b}| \leq \|\boldsymbol{a}\|\|\boldsymbol{b}\|$

Angle between Two Vectors

Part (i): First, consider $\boldsymbol{u} = (cos\theta, sin\theta)$ and $\boldsymbol{U} = \boldsymbol{i} = (1,0)$. Then, clearly $\boldsymbol{u} \cdot \boldsymbol{U} = cos\theta$. After a rotation through any angle α these are still unit vectors. Call the vectors $\boldsymbol{u} = (cos\beta, sin\beta)$ and $\boldsymbol{U} = (cos\alpha, sin\alpha)$. Their dot product is $cos\alpha cos\beta + sin\alpha sin\beta = cos(\beta - \alpha)$. Since $\beta - \alpha$ equals θ, we have reached the formula $\boldsymbol{u} \cdot \boldsymbol{U} = cos\theta$.

Parts (ii) and (iii) are immediate, following Part (i)

Other Properties of Matrices

Matrix Multiplication

Block Matrices and Block Multiplication

The elements of $\boldsymbol{A}$ can be cut into blocks, which are smaller matrices. If the cuts between columns of $\boldsymbol{A}$ match the cuts between rows of $\boldsymbol{B}$, then block multiplication is allowed.

$$\begin{bmatrix} \boldsymbol{A}_{11} & \boldsymbol{A}_{12} \\ \boldsymbol{A}_{21} & \boldsymbol{A}_{22} \end{bmatrix} \begin{bmatrix} \boldsymbol{B}_{11} & \boldsymbol{B}_{12} \\ \boldsymbol{B}_{21} & \boldsymbol{B}_{22} \end{bmatrix} = \begin{bmatrix} \boldsymbol{A}_{11}\boldsymbol{B}_{11} + \boldsymbol{A}_{12}\boldsymbol{B}_{12} & \boldsymbol{A}_{11}\boldsymbol{B}_{12} + \boldsymbol{A}_{12}\boldsymbol{B}_{22} \\ \boldsymbol{A}_{21}\boldsymbol{B}_{11} + \boldsymbol{A}_{22}\boldsymbol{B}_{21} & \boldsymbol{A}_{21}\boldsymbol{B}_{12} + \boldsymbol{A}_{22}\boldsymbol{B}_{22} \end{bmatrix}$$

Example:

$$\left[\begin{array}{cc:cc} 1 & 3 & 0 & 0 \\ 2 & 3 & 1 & 6 \\ \hdashline -3 & -1 & 1 & 0 \\ 0 & 2 & 0 & 1 \end{array}\right] \left[\begin{array}{cc:cc} 1 & 3 & 0 & 0 \\ 2 & 3 & 1 & 6 \\ \hdashline -3 & -1 & 1 & 0 \\ 0 & 2 & 0 & 1 \end{array}\right] = \left[\begin{array}{cc:cc} 7 & 12 & 3 & 16 \\ 5 & 26 & 4 & 24 \\ \hdashline -8 & -13 & 0 & -6 \\ 4 & 8 & 2 & 13 \end{array}\right]$$

$$\begin{bmatrix} 1 & 3 \\ 2 & 3 \end{bmatrix} \begin{bmatrix} 1 & 3 \\ 2 & 3 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 1 & 6 \end{bmatrix} \begin{bmatrix} -3 & -1 \\ 0 & 2 \end{bmatrix} = \begin{bmatrix} 7 & 12 \\ 5 & 26 \end{bmatrix}$$

Linear Equations

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = b_2 \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n = b_m \end{cases} \Leftrightarrow \boldsymbol{Ax} = \boldsymbol{b}$$

- $\boldsymbol{A} = [a_{ij}]_{m\times n}$ = Coefficient matrix
- $\boldsymbol{x} = [x_j]_{n\times 1}$ = Variable vector
- $\boldsymbol{b} = [b_j]_{m\times 1}$ = Vector of right hand side

The product $\boldsymbol{Ax}$ is the combination of columns of $\boldsymbol{A}$. Hence, the system has solution if $\boldsymbol{b}$ is inside the spanned space of the columns of $\boldsymbol{A}$:

$$x_1 \begin{bmatrix} a_{11} \\ a_{21} \\ \cdot \\ \cdot \\ \cdot \\ a_{m1} \end{bmatrix} + x_2 \begin{bmatrix} a_{12} \\ a_{22} \\ \cdot \\ \cdot \\ \cdot \\ a_{m2} \end{bmatrix} + \cdots + x_n \begin{bmatrix} a_{1n} \\ a_{2n} \\ \cdot \\ \cdot \\ \cdot \\ a_{mn} \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \cdot \\ \cdot \\ \cdot \\ b_m \end{bmatrix}$$

Homogeneous systems

Homogeneous System: The system

$$A_{m\times n}x_{n\times 1} = b_{m\times 1} \Leftrightarrow \begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = b_2 \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n = b_m \end{cases}$$

is called a homogeneous if $b_1 = b_2 \cdots = b_m = 0$. The system is *non-homogeneous* if at least one of the $b_i's$ is not 0.

1. if $k < n$, then the columns of A are linearly dependent, i.e., k columns are independents and $n - k$ columns can be written as a linear combination of the other k columns.
2. if $k = n$, then the columns of A are linearly independent, i.e., no column can be written as a linear combination of other columns.

Moore-Penrose Pseudo Inverse

Moore-Penrose Pseudo Inverse: When $k = n \leq m$ the system of linear equations $\boldsymbol{Ax} = \boldsymbol{b}$ can have no solution. In that case, we can resort to an approximation by using a least square in which we determine the best vector $\boldsymbol{x}$ that minimizes the sum of square of errors $\|\boldsymbol{Ax} - \boldsymbol{b}\|^2 = (\boldsymbol{Ax} - \boldsymbol{b})^T(\boldsymbol{Ax} - \boldsymbol{b})$. The best fit $\boldsymbol{x}$ is obtained as

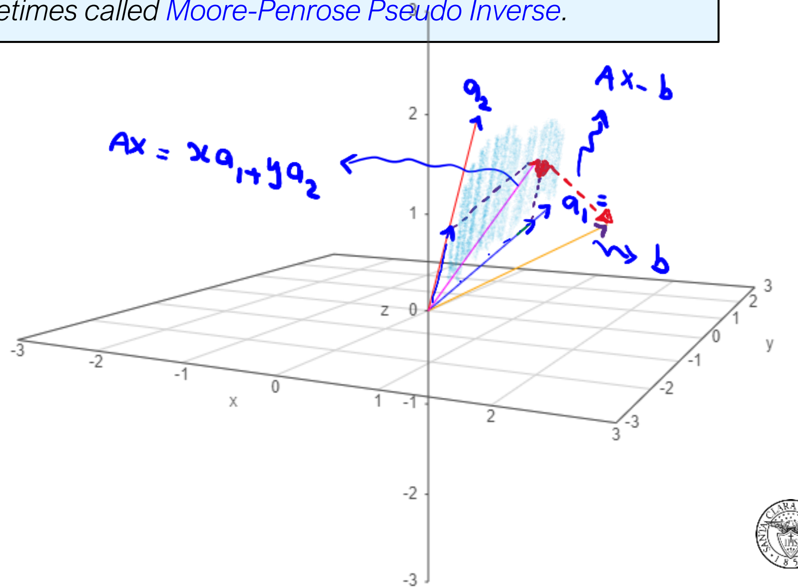
$$\boldsymbol{x} = \left(\boldsymbol{A}^T\boldsymbol{A}\right)^{-1}\boldsymbol{A}^T\boldsymbol{b}$$

Note that $\boldsymbol{A}^T\boldsymbol{A}$ is invertible because it is square matrix. $\left(\boldsymbol{A}^T\boldsymbol{A}\right)^{-1}\boldsymbol{A}^T$ is sometimes called Moore-Penrose Pseudo Inverse.

The minimum Euclidean norm $\|\boldsymbol{Ax} - \boldsymbol{b}\|$ (i.e., the minimum of squares of errors) occurs at a point $\mathbf{x}$ that satisfies: $\boldsymbol{Ax} \perp (\boldsymbol{Ax} - \boldsymbol{b})$:

$$\Rightarrow \boldsymbol{Ax} \cdot (\boldsymbol{Ax} - \boldsymbol{b}) = 0 \Rightarrow (\boldsymbol{Ax})^T \cdot (\boldsymbol{Ax} - \boldsymbol{b}) = 0 \Rightarrow \boldsymbol{x}^T\left(\boldsymbol{A}^T\boldsymbol{Ax} - \boldsymbol{A}^T\boldsymbol{b}\right) = 0$$

$$\Rightarrow \boldsymbol{A}^T\boldsymbol{Ax} - \boldsymbol{A}^T\boldsymbol{b} = 0 \Rightarrow \boldsymbol{x} = \left(\boldsymbol{A}^T\boldsymbol{A}\right)^{-1}\boldsymbol{A}^T\boldsymbol{b}$$



Example

Moore-Penrose Pseudo Inverse

Solve the system by finding the inverse of the coefficient matrix.

$$\begin{cases} x + y = 2 \\ x - y = 0 \\ x + 2y = 1 \end{cases}$$

$$\boldsymbol{A} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \\ 1 & 2 \end{bmatrix}, \boldsymbol{b} = \begin{bmatrix} 2 \\ 0 \\ 1 \end{bmatrix} \Rightarrow \begin{bmatrix} 1 & 1 \\ 1 & -1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 2 \\ 0 \\ 1 \end{bmatrix} \Rightarrow x \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} + y \begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix} = \begin{bmatrix} 2 \\ 0 \\ 1 \end{bmatrix} \Rightarrow \text{no solution}$$

$$\boldsymbol{x} = \begin{bmatrix} x \\ y \end{bmatrix} = \left(\boldsymbol{A}^T \boldsymbol{A}\right)^{-1} \boldsymbol{A}^T \boldsymbol{b} = \left(\begin{bmatrix} 1 & 1 \\ 1 & -1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 \\ 1 & -1 & 2 \end{bmatrix} \right)^{-1} \begin{bmatrix} 1 & 1 & 1 \\ 1 & -1 & 2 \end{bmatrix} \begin{bmatrix} 2 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0.71 \\ 0.43 \end{bmatrix}$$

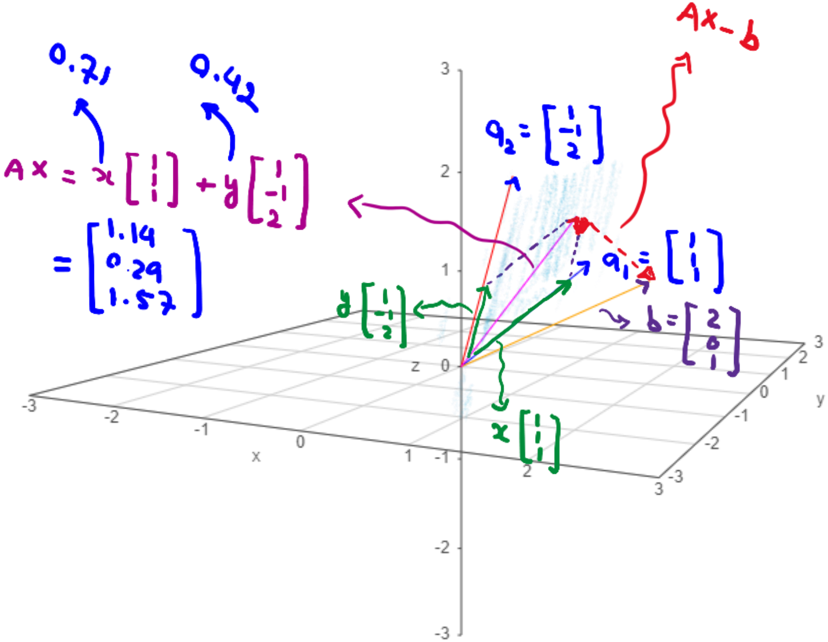
$$\boldsymbol{Ax} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = x \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} + y \begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix} = \begin{bmatrix} 1.14 \\ 0.29 \\ 1.57 \end{bmatrix}$$

Example

Moore-Penrose Pseudo Inverse

```
library(pracma)

A=matrix(c(1,1,1,-1,1,2),3,2,1)
b=matrix(c(2,0,1),3,1)
x=solve(t(A)%*%A)%*%t(A)%*%b
Norm(A%*%x-b)

f= function(x) {
  y=Norm(A%*%x-b);
  return(y);
}

optim(x,f)
```

```
$par
          [,1]
[1,] 0.7142857
[2,] 0.4285714

$value
[1] 1.069045
```

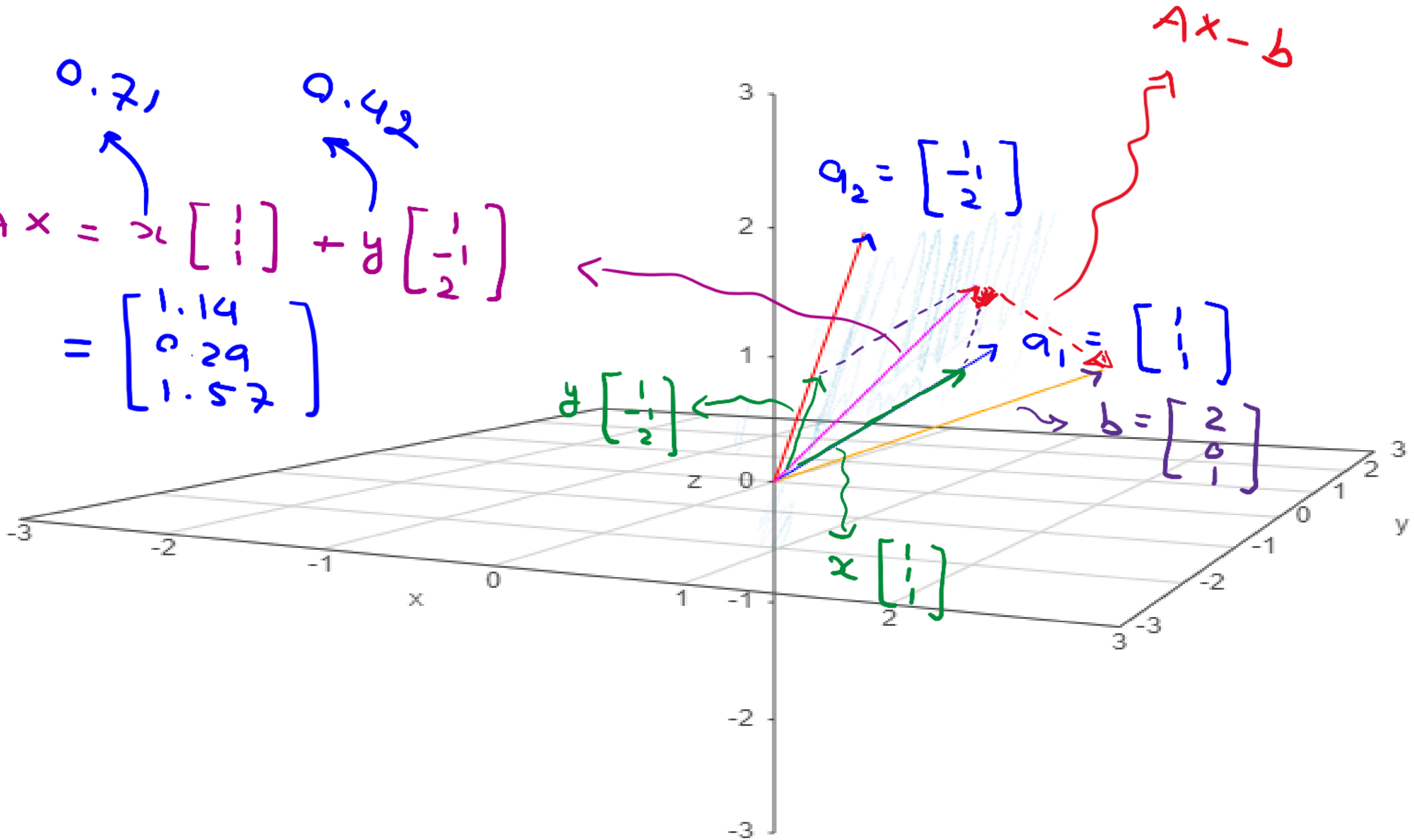


$$Ax \cdot (Ax - b) = (Ax)^T (Ax - b) = 0$$

$$\iff Ax \perp (Ax - b)$$

$$\overset{0.71}{x} \qquad \overset{0.42}{y}$$

$$Ax = x\begin{bmatrix}1\\1\\1\end{bmatrix} + y\begin{bmatrix}1\\-1\\2\end{bmatrix} = \begin{bmatrix}1.14\\0.29\\1.57\end{bmatrix}$$



Basis of a Vector Space

Basis: The linearly set of independent vectors $\boldsymbol{b}_i$, $i = 1,2, \dots , k,$ in the vector space $\boldsymbol{V}$ that every other vector $\boldsymbol{x} \in \boldsymbol{V}$ is a linear combination vectors from the basis and every linear combination is unique.

$$\boldsymbol{x} = \sum_{1}^{k} \lambda_i \boldsymbol{b}_i = \sum_{1}^{k} \beta_i \boldsymbol{b}_i \Rightarrow \lambda_i = \beta_i$$

Determinants

Determinants

Determinant: *The determinant of the symmetric matrix* $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ *is a recursive function that maps* $\boldsymbol{A}$ *into a real number by using Laplace Expansion:*

$$\det(\boldsymbol{A}) = |\boldsymbol{A}| = \begin{vmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{12} & \dots & a_{2n} \\ . & . & \dots & . \\ . & . & \dots & . \\ . & . & \dots & . \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{vmatrix}$$

In R use: `det(A)`

Laplace Expansion: For all $\boldsymbol{j = 1, \dots, n}$

- $|\boldsymbol{A}| = \sum_{k=1}^{n}(-1)^{k+j} a_{kj} \left|\boldsymbol{A}_{\boldsymbol{k,j}}\right|$ *(expansion along column j)*
- $|\boldsymbol{A}| = \sum_{k=1}^{n}(-1)^{k+j} a_{jk} \left|\boldsymbol{A}_{\boldsymbol{j,k}}\right|$ *(expansion along row j)*

$\boldsymbol{A}_{\boldsymbol{k,j}} \in \mathbb{R}^{(n-1)\times(n-1)}$ *is a submatrix of* $\boldsymbol{A}$ *that we obtain by deleting row* k *and column* j.

Remark: Using Laplace expansion along either the first row or the first column, it is not too difficult to verify:

- If $\boldsymbol{A} \in \mathbb{R}^{1\times 1}$ then $|\boldsymbol{A}| = |a_{11}| = a_{11}$
- If $\boldsymbol{A} \in \mathbb{R}^{2\times 2}$ then $|\boldsymbol{A}| = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}a_{12} - a_{21}a_{12}$
- The determinant of a diagonal matrix is the product of the elements on its main diagonal entries.

Determinants

Example

Compute the determinant of $\boldsymbol{A} = \begin{bmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \\ 0 & 0 & 1 \end{bmatrix}$.

Solution. Using Laplace expansion along the first row, we have

$$|A| = (-1)^{1+1}(1)\begin{vmatrix} 1 & 2 \\ 0 & 1 \end{vmatrix} + (-1)^{1+2}(2)\begin{vmatrix} 3 & 2 \\ 0 & 1 \end{vmatrix} + (-1)^{1+3}(3)\begin{vmatrix} 3 & 1 \\ 0 & 0 \end{vmatrix} = 1(1-0) - 2(3-0) + 3(0) = -5$$

Remark: $|A|$ gives n-dimensional volume of a n-dimensional parallelepiped made by the column vectors of A. If $|A| = 0$, then this parallelotope has a zero volume in n dimensions. or it is not n-dimensional, which indicates that the dimension of the image of A is less than n (we say the rank of A is less than n).

Determinants

Properties of Determinant

1. $|\boldsymbol{AB}| = |\boldsymbol{A}||\boldsymbol{B}|$
2. $|\boldsymbol{A}| = |\boldsymbol{A}^T|$
3. $|\boldsymbol{A}^{-1}| = \frac{1}{|\boldsymbol{A}|}$
4. Adding a multiple of a column/row to another does not change $|\boldsymbol{A}|$
5. Multiplication of a column/row with $\lambda \in \mathbb{R}$ scales $|\boldsymbol{A}|$ by λ. In particular $|\lambda \boldsymbol{A}| = \lambda^n |\boldsymbol{A}|$
6. Swapping two rows/columns changes the sign of $|\boldsymbol{A}|$
7. Determinant of any diagonal matrix is the product of the elements on its main diagonal entries.
8. Similar matrices have the same determinant
   - Two matrices $\boldsymbol{A}, \boldsymbol{D} \in \mathbb{R}^{n\times n}$ are similar if there exists an invertible matrix $\boldsymbol{P} \in \mathbb{R}^{n\times n}$ with $\boldsymbol{D} = \boldsymbol{P}^{-1}\boldsymbol{AP}$
   - Using the definition: $|\boldsymbol{D}| = |\boldsymbol{P}^{-1}\boldsymbol{AP}| = |\boldsymbol{P}^{-1}||\boldsymbol{A}||\boldsymbol{P}| = \frac{1}{|\boldsymbol{P}|}|\boldsymbol{A}||\boldsymbol{P}| = |\boldsymbol{A}|$

Theorem: $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ is invertible and full-rank, i.e., $\mathrm{rank}(\boldsymbol{A}) = n$, if and only if $|\boldsymbol{A}| \neq \boldsymbol{0}$

Determinants

Example

Compute the determinant of $\boldsymbol{A} = \begin{bmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \\ 0 & 0 & 1 \end{bmatrix}$ (this time by using determinant properties).

Solution. Our strategy is to use determinant properties to change the first column to $\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$.

To do so, adding -3 times row 1 to row 3 gives:

$$\begin{bmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \\ 0 & 0 & 1 \end{bmatrix} \xrightarrow{-3R_1+R_2} \begin{bmatrix} 1 & 2 & 3 \\ 0 & -5 & -7 \\ 0 & 0 & 1 \end{bmatrix}$$

Now expanding across column 1 is very easy:

$$|A| = (1)^{1+1}(1) \begin{vmatrix} -5 & -7 \\ 0 & 1 \end{vmatrix} + 0 + 0 = -5$$

This approach is especially helpful for obtaining the determinants for higher dimensional matrices.

Eigenvalues and Eigenvectors

Eigenvalues and Eigenvectors

Definition: $\lambda \in \mathbb{R}$ *is an eigenvalue of* $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ *and* $\boldsymbol{x} \in \mathbb{R}^{n}\backslash\{\boldsymbol{0}\}$ *is the corresponding eigenvector of* $\boldsymbol{A}$ *if:*

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

In R use: `eigen(A)`

The above equation is known as the eigenvalue equation.

Remark: The following statements are equivalent:

- $\lambda \in \mathbb{R}$ is an eigenvalue of $\boldsymbol{A} \in \mathbb{R}^{n\times n}$
- There exists $\boldsymbol{x} \in \mathbb{R}^{n}\backslash\{\boldsymbol{0}\}$ with $\boldsymbol{A}\boldsymbol{x} = \lambda\boldsymbol{x}$, or equivalently $(\boldsymbol{A} - \lambda\boldsymbol{I})\boldsymbol{x} = \boldsymbol{0}$, can be solved non-trivially, i.e., $\boldsymbol{x} \neq \boldsymbol{0}$.
- $\mathrm{rank}(\boldsymbol{A} - \lambda\boldsymbol{I}) < n$
- $\det(\boldsymbol{A} - \lambda\boldsymbol{I}) = \boldsymbol{0}$ $(\boldsymbol{A} - \lambda\boldsymbol{I})$ is called singular, i.e., meaning that it is not invertible.

- Remark: $\boldsymbol{p}_{\boldsymbol{A}}(\lambda) \equiv \det(\boldsymbol{A} - \lambda\boldsymbol{I})$ is also known as the Characteristic Polynomial

Eigenvalues and Eigenvectors

Properties of Eigenvalues and Eigenvectors

Theorem (non-uniqueness of eigenvector): If $\boldsymbol{x}$ is an eigenvector of $\boldsymbol{A}$ associated with the eigenvalue λ, then for any $c \neq 0$, $c\boldsymbol{x}$ is also an eigenvector of $\boldsymbol{A}$ with the same eigenvalue.

$$\boldsymbol{A}(c\boldsymbol{x}) = c\boldsymbol{A}\boldsymbol{x} = c\lambda\boldsymbol{x} = \lambda(c\boldsymbol{x})$$

Theorem: $\lambda \in \mathbb{R}$ is an eigenvalue of $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ if and only if λ is a root of the characteristic polynomial of $\boldsymbol{A}$

$$\boldsymbol{p_A}(\lambda) \equiv det(\boldsymbol{A} - \lambda\boldsymbol{I}) = \boldsymbol{0}$$

Other properties:

- $\boldsymbol{A}$ and $\boldsymbol{A^T}$ have the same eigenvalues but not necessarily the same eigenvectors.
- Similar matrices have the same eigenvalues.
- Symmetric positive definite matrices always have positive eigenvalues.
- Determinant of a matrix is equal to the product of its eigenvalues.

Eigenvalues and Eigenvectors

Example

Find the eigenvalues and the eigenvectors of $\boldsymbol{A} = \begin{bmatrix} 4 & 2 \\ 1 & 3 \end{bmatrix}$.

Solution.

Step 1: eigenvalues

$$\boldsymbol{p_A}(\lambda) \equiv \det(\boldsymbol{A} - \lambda \boldsymbol{I}) = \det\left(\begin{bmatrix} 4 & 2 \\ 1 & 3 \end{bmatrix} - \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix}\right) = \begin{vmatrix} 4-\lambda & 2 \\ 1 & 3-\lambda \end{vmatrix} = (4-\lambda)(3-\lambda) - 2 = 0 \Rightarrow \lambda = 2, \lambda = 5$$

Step 2: eigenvectors corresponding to each eigenvalue: $\begin{bmatrix} 4-\lambda & 2 \\ 1 & 3-\lambda \end{bmatrix} \boldsymbol{x} = \boldsymbol{0}$

If $\lambda = 5 \Rightarrow \begin{bmatrix} 4-5 & 2 \\ 1 & 3-5 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \boldsymbol{0} \Rightarrow \begin{bmatrix} -1 & 2 \\ 1 & -2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \boldsymbol{0} \Rightarrow \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$

If $\lambda = 2 \Rightarrow \begin{bmatrix} 4-2 & 2 \\ 1 & 3-2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \boldsymbol{0} \Rightarrow \begin{bmatrix} 2 & 2 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \boldsymbol{0} \Rightarrow \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$

Matrix Decomposition

(Important tool for obtaining complex computations, e.g., $\sqrt{\boldsymbol{A}}$, $\boldsymbol{A}^{-3.456}$, $e^{\boldsymbol{A}}$, and many other results)

Matrix Decomposition

Eigendecomposition and Diagonalization

Similar matrices: *Two matrices* $\boldsymbol{A}, \boldsymbol{D} \in \mathbb{R}^{n\times n}$ *are similar if there exists an invertible matrix* $\boldsymbol{P} \in \mathbb{R}^{n\times n}$ *with* $\boldsymbol{D} = \boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P}$

Diagonal Matrix: *A matrix* $\boldsymbol{D} \in \mathbb{R}^{n\times n}$ *is diagonal if* $d_{ij} = 0,\ \forall\, i \neq j$

Diagonalizable Matrix: *A matrix* $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ *is diagonalizable if it is similar to a diagonal matrix, i.e., if there exists a diagonal matrix* $\boldsymbol{D}$ *and an invertible matrix* $\boldsymbol{P} \in \mathbb{R}^{n\times n}$ *such that* $\boldsymbol{D} = \boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P}.$

Theorem (Eigendecomposition): *A square matrix* $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ *can be factored into*

$$\boldsymbol{A} = \boldsymbol{P}\boldsymbol{D}\boldsymbol{P}^{-1}$$

where $\boldsymbol{P} \in \mathbb{R}^{n\times n}$ *is a matrix whose columns are the eigenvectors of* $\boldsymbol{A}$ *and* $\boldsymbol{D}$ *is a diagonal matrix whose diagonal entries are eigenvalues of* $\boldsymbol{A}$*.*

Matrix Decomposition

Eigendecomposition

Proof.

$\boldsymbol{A} \in \mathbb{R}^{n \times n}$ is diagonalizable if it is similar to a diagonal matrix $\boldsymbol{D}$, i.e., if there exists $\boldsymbol{P} \in \mathbb{R}^{n \times n}$ such that $\boldsymbol{D} = \boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P}$, which is the same as $\boldsymbol{A}\boldsymbol{P} = \boldsymbol{P}\boldsymbol{D}$. Let $\boldsymbol{D}$ be a diagonal matrix with the eigenvalues $\lambda_j, j = 1, \dots, n$ on its main diagonal entries and $\boldsymbol{P} = [\boldsymbol{p_1}, \dots, \boldsymbol{p_n}]$. Then:

$$\boldsymbol{A}\boldsymbol{P} = \boldsymbol{A}[\boldsymbol{p_1}, \dots, \boldsymbol{p_n}] = [\boldsymbol{A}\boldsymbol{p_1}, \dots, \boldsymbol{A}\boldsymbol{p_n}].$$

$$\boldsymbol{P}\boldsymbol{D} = [\boldsymbol{p_1}, \dots, \boldsymbol{p_n}] \begin{bmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_n \end{bmatrix} = [\lambda_1 \boldsymbol{p_1}, \dots, \lambda_n \boldsymbol{p_n}].$$

This implies that $[\boldsymbol{A}\boldsymbol{p_1}, \dots, \boldsymbol{A}\boldsymbol{p_n}] = [\lambda_1 \boldsymbol{p_1}, \dots, \lambda_n \boldsymbol{p_n}]$ or $\boldsymbol{A}\boldsymbol{p_j} = \lambda_1 \boldsymbol{p_j}$. Therefore, $\boldsymbol{p_j}$ must be an eigenvector corresponding to λ_j.

Matrix Decomposition

Real Powers of a Matrix

Remark: For $\boldsymbol{A} \in \mathbb{R}^{n\times n}$, we can see:

$$\boldsymbol{A}^2 = \boldsymbol{A}\times\boldsymbol{A} = (\boldsymbol{PDP}^{-1})(\boldsymbol{PDP}^{-1}) = \boldsymbol{PD}(\boldsymbol{P}^{-1}\boldsymbol{P})\boldsymbol{DP}^{-1} = \boldsymbol{PDIDP}^{-1} = \boldsymbol{PDDP}^{-1} = \boldsymbol{PD}^2\boldsymbol{P}^{-1}$$

$$\boldsymbol{A}^3 = \boldsymbol{A}^2\times\boldsymbol{A} = (\boldsymbol{PD}^2\boldsymbol{P}^{-1})(\boldsymbol{PDP}^{-1}) = \boldsymbol{PD}^2(\boldsymbol{P}^{-1}\boldsymbol{P})\boldsymbol{DP}^{-1} = \boldsymbol{PD}^2\boldsymbol{IDP}^{-1} = \boldsymbol{PD}^2\boldsymbol{DP}^{-1} = \boldsymbol{PD}^3\boldsymbol{P}^{-1}$$

$$\vdots$$

Continuing this way, we can verify that

$$\boldsymbol{A}^k = \boldsymbol{PD}^k\boldsymbol{P}^{-1} = \boldsymbol{P}\begin{bmatrix}\lambda_1^k & \cdots & 0\\ \vdots & \ddots & \vdots\\ 0 & \cdots & \lambda_n^k\end{bmatrix}\boldsymbol{P}^{-1}$$

It can be shown that the above result holds generally for any $k \in \mathbb{R}$ not just integer values. This result, which is based on matrix decomposition, is extremely important in finding $\boldsymbol{A}^k$ when k is a very large number or when it is a real number (e.g., $\sqrt{\boldsymbol{A}}, \boldsymbol{A}^{-3.21}, \ldots$) in which case the direct approach is not applicable.

Matrix Decomposition

Exponential and Logarithm of a Matrix

Definition: *For a matrix* $\boldsymbol{A} \in \mathbb{R}^{n\times n}$*, the exponential of* $\boldsymbol{A}$ *is defined by the Taylor expansion of* e *on* $\boldsymbol{A}$ *as:*

$$e^{\boldsymbol{A}} = \boldsymbol{I} + \boldsymbol{A} + \frac{\boldsymbol{A}^2}{2!} + \frac{\boldsymbol{A}^3}{3!} + \cdots$$

Theorem: For a diagonalizable matrix $\boldsymbol{A} \in \mathbb{R}^{n\times n}$, we have

$$e^{\boldsymbol{A}} = \boldsymbol{P}e^{\boldsymbol{D}}\boldsymbol{P}^{-1} = \boldsymbol{P}\begin{bmatrix} e^{\lambda_1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & e^{\lambda_k} \end{bmatrix}\boldsymbol{P}^{-1}$$

Proof.

$$e^{\boldsymbol{A}} = \boldsymbol{I} + \boldsymbol{A} + \frac{\boldsymbol{A}^2}{2!} + \frac{\boldsymbol{A}^3}{3!} + \cdots$$

$$= \boldsymbol{I} + \boldsymbol{PDP}^{-1} + \frac{\boldsymbol{PD}^2\boldsymbol{P}^{-1}}{2!} + \cdots = \boldsymbol{P}\left(\boldsymbol{I} + \boldsymbol{D} + \frac{\boldsymbol{D}^2}{2!} + \frac{\boldsymbol{D}^3}{3!} + \cdots\right)\boldsymbol{P}^{-1} = \boldsymbol{P}\begin{bmatrix} 1 + \lambda_1 + \frac{\lambda_1^2}{2!} + \cdots & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1 + \lambda_n + \frac{\lambda_n^2}{2!} + \cdots \end{bmatrix}\boldsymbol{P}^{-1} = \boldsymbol{P}\begin{bmatrix} e^{\lambda_1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & e^{\lambda_k} \end{bmatrix}\boldsymbol{P}^{-1}$$

Matrix Decomposition

Exponential and Logarithm of a Matrix

Theorem: For a diagonalizable matrix $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ *, we have*

$$ln\,\boldsymbol{A} = \boldsymbol{P}\begin{bmatrix} ln\,\lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & ln\,\lambda_1 \end{bmatrix}\boldsymbol{P}^{-1}$$

Proof. It is enough to show that the above formula satisfies $e^{ln\,\boldsymbol{A}} = ln\,e^{\boldsymbol{A}} = \boldsymbol{A}$. We show only $e^{ln\,\boldsymbol{A}} = \boldsymbol{I}$ as showing the other is very similar.

$$e^{ln\,\boldsymbol{A}} = \boldsymbol{I} + ln\,\boldsymbol{A} + \frac{(ln\,\boldsymbol{A})^2}{2!} + \frac{(ln\,\boldsymbol{A})^3}{3!} + \cdots = \boldsymbol{I} + \boldsymbol{P}\begin{bmatrix} ln\,\lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & ln\,\lambda_1 \end{bmatrix}\boldsymbol{P}^{-1} + \frac{\mathbf{1}}{2!}\boldsymbol{P}\begin{bmatrix} (ln\,\lambda_1)^2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & (ln\,\lambda_n)^2 \end{bmatrix}\boldsymbol{P}^{-1} + \cdots =$$

$$= \boldsymbol{P}\begin{bmatrix} ln\,\lambda_1 + \frac{1}{2!}(ln\,\lambda_1)^2 + \cdots & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & ln\,\lambda_n + \frac{1}{2!}(ln\,\lambda_n)^2 + \cdots \end{bmatrix}\boldsymbol{P}^{-1} = \boldsymbol{P}\begin{bmatrix} e^{ln\,\lambda_1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & e^{ln\,\lambda_n} \end{bmatrix}\boldsymbol{P}^{-1} = \boldsymbol{P}\begin{bmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_n \end{bmatrix}\boldsymbol{P}^{-1} = \boldsymbol{PDP}^{-1} = \boldsymbol{A}$$

Matrix Decomposition

Example

If $\boldsymbol{A} = \begin{bmatrix} 4 & 2 \\ 1 & 3 \end{bmatrix}$, find the following values:

a. $\sqrt{\boldsymbol{A}}$

b. $\boldsymbol{A}^e$ ($e \approx 2.718$: the Euler's constant).

c. $\boldsymbol{e}^{\boldsymbol{A}}$

Solution. From the previous example's solution, we have $\boldsymbol{D} = \begin{bmatrix} 2 & 0 \\ 0 & 5 \end{bmatrix}, \boldsymbol{P} = \begin{bmatrix} 2 & 1 \\ 1 & -1 \end{bmatrix} \Rightarrow \boldsymbol{A} = \boldsymbol{P}\boldsymbol{D}\boldsymbol{P}^{-1} = \frac{1}{3}\begin{bmatrix} 2 & 1 \\ 1 & -1 \end{bmatrix}\begin{bmatrix} 2 & 0 \\ 0 & 5 \end{bmatrix}\begin{bmatrix} 1 & 1 \\ 1 & -2 \end{bmatrix}$

a. $\sqrt{\boldsymbol{A}} = \boldsymbol{A}^{\frac{1}{2}} = \boldsymbol{P}\boldsymbol{D}^{\frac{1}{2}}\boldsymbol{P}^{-1} = \frac{1}{3}\begin{bmatrix} 2 & 1 \\ 1 & -1 \end{bmatrix}\begin{bmatrix} \sqrt{2} & 0 \\ 0 & \sqrt{5} \end{bmatrix}\begin{bmatrix} 1 & 1 \\ 1 & -2 \end{bmatrix} = \begin{bmatrix} 1.96 & 0.55 \\ 0.27 & 1.69 \end{bmatrix}$.

b. $\boldsymbol{A}^e = \boldsymbol{P}\boldsymbol{D}^e\boldsymbol{P}^{-1} = \frac{1}{3}\begin{bmatrix} 2 & 1 \\ 1 & -1 \end{bmatrix}\begin{bmatrix} 2^e & 0 \\ 0 & 5^e \end{bmatrix}\begin{bmatrix} 1 & 1 \\ 1 & -2 \end{bmatrix} = \begin{bmatrix} 55.15 & 48.57 \\ 24.27 & 30.86 \end{bmatrix}$

c. $\boldsymbol{e}^{\boldsymbol{A}} = \boldsymbol{P}\boldsymbol{e}^{\boldsymbol{D}}\boldsymbol{P}^{-1} = \frac{1}{3}\begin{bmatrix} 2 & 1 \\ 1 & -1 \end{bmatrix}\begin{bmatrix} e^2 & 0 \\ 0 & e^5 \end{bmatrix}\begin{bmatrix} 1 & 1 \\ 1 & -2 \end{bmatrix} = \begin{bmatrix} 101.41 & 94.02 \\ 47.01 & 54.40 \end{bmatrix}$

```
y$vectors%*%diag(exp(y$values))%*%matrix.inverse(y$vectors)
```

Matrix Decomposition

Relationship between Eigenvalues and Determinant

Theorem: *Let* $\lambda_1, \dots, \lambda_n$ *be the eigenvalues of the matrix* $\boldsymbol{A} \in \mathbb{R}^{m \times n}$. *Then*

$$|\boldsymbol{A}| = \prod_{i=1}^{n} \lambda_i = \lambda_1 \times \lambda_2 \times \cdots \times \lambda_n$$

In addition, if $\boldsymbol{A}$ *is singular (i.e.,* $|\boldsymbol{A}| = 0$*) then it has at least an eigenvalue, which is zero.*

Proof.

From the eigendecomposition of $\boldsymbol{A}$, we know that $\boldsymbol{A} = \boldsymbol{P}\boldsymbol{D}\boldsymbol{P}^{-1}$ where $\boldsymbol{P}$ is the matrix of eigenvectors and $\boldsymbol{D}$ is the diagonal matrix whose main diagonal entries are the eigenvalues. Taking the determinant gives:

$$|\boldsymbol{A}| = \left|\boldsymbol{P}\boldsymbol{D}\boldsymbol{P}^{-1}\right| = |\boldsymbol{P}||\boldsymbol{D}|\left|\boldsymbol{P}^{-1}\right| = |\boldsymbol{P}||\boldsymbol{D}||\boldsymbol{P}|^{-1} = |\boldsymbol{D}| = \lambda_1 \times \lambda_2 \times \cdots \times \lambda_n.$$

If $\boldsymbol{A}$ is singular, then $|\boldsymbol{A}| = \prod_{i=1}^{n} \lambda_i = 0$. Hence, one of the eigenvalues is at least zero.

Matrix Norms

Norm of a matrix: *The definition of the corresponding norm for an* $n \times n$ *matrix* $\boldsymbol{A} \in \mathbb{R}^{n\times m}$ *is*

$$\|\boldsymbol{A}\| = \max_{\boldsymbol{x}\neq\boldsymbol{0}} \frac{\|\boldsymbol{Ax}\|}{\|\boldsymbol{x}\|} = \max_{\|\boldsymbol{x}\|=\boldsymbol{1}} \|\boldsymbol{Ax}\| \qquad \forall \boldsymbol{x} \in \mathbb{R}^{n}.$$

Where $\|\cdot\|$ *is the Euclidean norm.*

Remark: From the above, it follows that:

$$\|\boldsymbol{Ax}\| \leq \|\boldsymbol{A}\|\|\boldsymbol{x}\|.$$

Remark: it can be shown that

$$\|\boldsymbol{A}\| = \lambda_{max}$$

where λ_{max} is the maximum eigenvalue of $\boldsymbol{A} \in \mathbb{R}^{n\times m}$.

(The proof needs additional discussion about orthonormal eigenvector bases…)

Positive Definite Matrices

Definition: *a symmetric* $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ *is called positive semidefinite if and only if*

$$\boldsymbol{x^T A x} \geq \boldsymbol{0} \quad \forall \boldsymbol{x} \in \mathbb{R}^n$$

And positive definite if and only if

$$\boldsymbol{x^T A x} > \boldsymbol{0} \quad \forall \boldsymbol{x} \neq \boldsymbol{0} \in \mathbb{R}^n$$

Theorem: $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ *is positive semidefinite if and only if all its eigenvalues are greater than or equal to zero.*

Proof.

By definition, we have $\boldsymbol{x^T A x} \geq \boldsymbol{0} \quad \forall \boldsymbol{x} \in \mathbb{R}^n$. Choose $\boldsymbol{x}$ to be any of the eigenvectors of $\boldsymbol{A}$ with λ the corresponding eigenvalue to $\boldsymbol{x}$.

Hence, we have

$$\boldsymbol{x^T A x} = \boldsymbol{x^T}(\boldsymbol{A x}) = \boldsymbol{x^T}\lambda\boldsymbol{x} = \lambda\boldsymbol{x^T x} = \lambda\|\boldsymbol{x}\|_2{}^2 \geq 0$$

Since $\boldsymbol{x^T x} = \|\boldsymbol{x}\|_2{}^2 > 0$ for any $\boldsymbol{x} \neq \boldsymbol{0}$ (it is the Euclidean or $\boldsymbol{L_2}$ norm), we must have $\lambda \geq 0$.

Remark: From above, it follows that if $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ is positive semidefinite then $|\boldsymbol{A}| \geq 0$ because $|\boldsymbol{A}| = \prod_{i=1}^{n} \lambda_i$.

Positive Definite Matrices

Theorem: For a matrix $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ we can always obtain a symmetric positive semidefinite matrix $\boldsymbol{S} \in \mathbb{R}^{n \times n}$ by defining

$$\boldsymbol{S} = \boldsymbol{A}^T \boldsymbol{A}.$$

Proof.

Symmetry requires that $\boldsymbol{S} = \boldsymbol{S}^T$. We have $\boldsymbol{S}^T = \left(\boldsymbol{A}^T \boldsymbol{A}\right)^T = \boldsymbol{A}^T \left(\boldsymbol{A}^T\right)^T = \boldsymbol{A}^T \boldsymbol{A}.$

By definition, a PSD matrix, we have $\boldsymbol{x}^T \boldsymbol{S} \boldsymbol{x} = \boldsymbol{x}^T \boldsymbol{A}^T \boldsymbol{A} \boldsymbol{x} = (\boldsymbol{x} \boldsymbol{A})^T (\boldsymbol{x} \boldsymbol{A}) = \|\boldsymbol{x} \boldsymbol{A}\|_2{}^2 \geq \boldsymbol{0}$.