Determinant

- 2×2 matrix determinant: $\det\left(\begin{bmatrix} a & b \\ c & d \end{bmatrix}\right) = \begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc$
- $k\times k$ matrix determinant:

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1k} \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{k1} & a_{k2} & a_{k3} & \cdots & a_{kk} \end{vmatrix} = a_{11} \begin{vmatrix} a_{22} & a_{23} & \cdots & a_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ a_{k2} & a_{k3} & \cdots & a_{kk} \end{vmatrix}$$

$$-a_{12} \begin{vmatrix} a_{21} & a_{23} & \cdots & a_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ a_{k1} & a_{k3} & \cdots & a_{kk} \end{vmatrix} + \ldots \pm a_{1k} \begin{vmatrix} a_{21} & a_{22} & \cdots & a_{2(k-1)} \\ \vdots & \vdots & \ddots & \vdots \\ a_{k1} & a_{k2} & \cdots & a_{k(k-1)} \end{vmatrix}.$$

- Properties
  - $\det(A) = \det(A^T)$
  - If $A = diag(a_1, a_2, \ldots, a_n)$, $\det(A) = a_1 a_2 \ldots a_n$
  - $\det(AB) = \det(A)\det(B)$ if A, B are the square matrix and same size
  - The matrix is invertible $\leftrightarrow$ $\det(A) \neq 0$

$\rightarrow \det(A^{-1}) = \dfrac{1}{\det(A)}$

Linear Combination, Span Space

- Given $\mathbf{a}_1, \ldots, \mathbf{a}_n \in \mathbb{R}^m$ and real number $x_1, \ldots, x_n \in \mathbb{R}$, $b = x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + \ldots + x_n\mathbf{a}_n$ is called linear combination of $\mathbf{a}_1, \ldots, \mathbf{a}_n$
- Given $\mathbf{A} = [\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n] \in \mathbb{R}^{m\times n}$, $\mathbf{b} = \mathbf{A}\mathbf{x}$ is called linear combination of rows of A $\rightarrow span(\{\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n\}) \in \mathbb{R}^m$,with m is max rows of A linear independent
- If $0 = x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + \ldots + x_n\mathbf{a}_n$ has one solution $x_1 = x_2 = \ldots = x_n = 0$, $\{\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n\}$ is called linear independence. If $x_i \neq 0$, the system is called linear dependence
- $range(A) = R(A) = span(\{\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n\}) = \left\{\mathbf{b} \in \mathbb{R}^m \mid \mathbf{b} = \mathbf{A}\mathbf{x} = \sum_{i=1}^{n} x_i \mathbf{a}_i\right\}$
- $Null\ Space(A) = N(A) = \{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{A}\mathbf{x} = 0\}$, if rows of A is linear independence, $N(A) = \{0\}$
- $\dim(R(A)) + \dim(N(A)) = n$
- Rank: the dimension of the $range(A)$. For example,

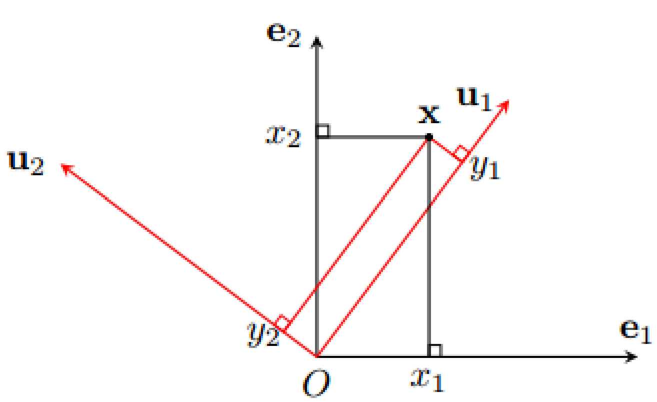
$C = \begin{bmatrix} 1 & 4 & 1 \\ -8 & -2 & 3 \\ 8 & 2 & -2 \end{bmatrix} = [\mathbf{x}_1 \quad \mathbf{x}_2 \quad \mathbf{x}_3] \rightarrow \mathbf{x}_1, \mathbf{x}_2$ are linearly independent, so C has the rank of 2, but $\{\mathbf{x}_1 \quad \mathbf{x}_2 \quad \mathbf{x}_3\}$ are not

- Rank Properties
  - $rank(A) = rank(A^T)$: max rows linear independent = max columns linear independent
  - If $A \in \mathbb{R}^{m\times n}$, $rank(A) \leq \min(m, n)$
  - $rank(AB) \leq \min(rank(A), rank(B))$
  - Sylvester theorem: If $A \in \mathbb{R}^{m\times n}, B \in \mathbb{R}^{n\times k}$, $rank(A) + rank(B) - n \leq rank(AB)$
- Orthogonal: $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_m \in \mathbb{R}^m\}$ is called orthogonal if $\mathbf{u}_i \neq 0; \mathbf{u}_i^T\mathbf{u}_j = 0\ \ \forall 1 \leq i \neq j \leq m$. If $U = [\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_m]$ and $UU^T = I$, U is called orthogonal matrix
- Orthonormal: $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_m \in \mathbb{R}^m\}$ is called orthogonal if it is orthogonal and $\mathbf{u}_i^T\mathbf{u}_j = \begin{cases} 1 & if\ i = j \\ 0 & o.w. \end{cases}$

- Orthogonal Properties
  - If U is orthogonal matrix, $U^T = U^{-1}$
  - If U is orthogonal matrix, $\det(U) = \pm 1$
  - Given $\mathbf{x},\mathbf{y} \in \mathbb{R}^m$ and $U \in \mathbb{R}^{m\times m}$, we use U to transform **x** and **y** to U**x** and U**y**, but this transformation will not change the result of dot product of **x** and **y**: $(U\mathbf{x})^T (U\mathbf{y}) = \mathbf{x}^T U^T U\mathbf{y} = \mathbf{x}^T \mathbf{y}$

Represent vector in different axis

- Set of vector $\mathbf{e}_1, \mathbf{e}_2, ..., \mathbf{e}_n$ in which $\mathbf{e}_i$ has one value $\neq 0$ and $=1$ is called the axis in **m** dimension
- $\mathbf{x} = [x_1, x_2, ..., x_n] \in \mathbb{R}^m$ can be considered as linear combination of **b** and the axis: $\mathbf{x} = x_1\mathbf{e}_1 + x_2\mathbf{e}_2 + ... + x_3\mathbf{e}_3$
- Assume we have another axis $U = [\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_m]$, represent **x** in this axis: $\mathbf{x} = y_1\mathbf{u}_1 + y_2\mathbf{u}_2 + ... + y_n\mathbf{u}_n = Uy \rightarrow y = U^{-1}\mathbf{x}$
- U need to be orthogonal matrix because $U^T = U^{-1}$,so y can be easily calculated by $\mathbf{y} = U^T\mathbf{x}$



Hình 1.1: Chuyển đổi toạ độ trong các hệ cơ sở khác nhau. Trong hệ toạ độ $O\mathbf{e}_1\mathbf{e}_2$, $\mathbf{x}$ có toạ độ là (x_1, x_2). Trong hệ toạ độ $O\mathbf{u}_1\mathbf{u}_2$, $\mathbf{x}$ có toạ độ là (y_2, y_2).

Eigenvalue, Eigenvector

- Given $A \in \mathbb{R}^{n\times n}, \mathbf{x} \in \mathbb{R}^n, \lambda \in \mathbb{R}$, if $A\mathbf{x} = \lambda\mathbf{x}$, $\lambda, \mathbf{x}$ is called eigenvalue and eigenvector respectively of matrix A
- λ is solution of $\det(A - \lambda I) = 0$, spectrum is the set of all λ of matrix A. For example, $A = \begin{bmatrix} 0 & 1 \\ -2 & -3 \end{bmatrix}$, the characteristic equation is $\det(A - \lambda I) = |A - \lambda I| = \begin{vmatrix} -\lambda & 1 \\ -2 & -3-\lambda \end{vmatrix} = 0 \rightarrow \begin{cases} \lambda_1 = -1 \\ \lambda_2 = -2 \end{cases}$
  - If $\lambda_1 = -1$

$$\mathbf{A} \cdot \mathbf{v}_1 = \lambda_1 \cdot \mathbf{v}_1$$

$$(\mathbf{A} - \lambda_1) \cdot \mathbf{v}_1 = 0$$

$$\begin{bmatrix} -\lambda_1 & 1 \\ -2 & -3-\lambda_1 \end{bmatrix} \cdot \mathbf{v}_1 = 0$$

$$\begin{bmatrix} 1 & 1 \\ -2 & -2 \end{bmatrix} \cdot \mathbf{v}_1 = \begin{bmatrix} 1 & 1 \\ -2 & -2 \end{bmatrix} \cdot \begin{bmatrix} \mathrm{v}_{1,1} \\ \mathrm{v}_{1,2} \end{bmatrix} = 0$$

$$\mathbf{v}_1 = \mathrm{k}_1 \begin{bmatrix} +1 \\ -1 \end{bmatrix}$$

  - If $\lambda_2 = -2$

$$\mathbf{A}\cdot\mathbf{v}_2 = \lambda_2\cdot\mathbf{v}_2$$

$$(\mathbf{A}-\lambda_2)\cdot\mathbf{v}_2 = \begin{bmatrix}-\lambda_2 & 1\\ -2 & -3-\lambda_2\end{bmatrix}\cdot\mathbf{v}_2 = \begin{bmatrix}2 & 1\\ -2 & -1\end{bmatrix}\cdot\begin{bmatrix}v_{2,1}\\ v_{2,2}\end{bmatrix} = 0 \quad \text{so}$$

$$2\cdot v_{2,1} + 1\cdot v_{2,2} = 0 \quad (\text{or from bottom line: } -2\cdot v_{2,1} - 1\cdot v_{2,2} = 0)$$

$$2\cdot v_{2,1} = -v_{2,2}$$

$$\mathbf{v}_2 = k_2\begin{bmatrix}+1\\ -2\end{bmatrix}$$

- Properties
  - $\det(A) = \prod_i \lambda_i$ in which λ_i is the eigenvalue of matrix A
  - $\sum_i a_i = \sum_j \lambda_j$ in which $diag(a_1, a_2, ..., a_n)$ in matrix A

Eigendecomposition or Diagonalization

- Assume $\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_n \neq 0$ is the eigenvector of square matrix A having eigenvalue of $\lambda_1, \lambda_2, ..., \lambda_n$
- Let $\Lambda = diag(\lambda_1, \lambda_2, ..., \lambda_n)$, $X = [\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_n]$, we have $AX = X\Lambda$
- If rows of X is linearly independent, X is invertible, so $A = X\Lambda X^{-1}$: eigendecomposition
- Properties
  - Eigendecomposition or Diagonalization is only applicable for **Square Matrix**
  - Square Matrix size n is only diagonalizable iff it has enough n eigenvalues linear independence
  - $A^2 = (X\Lambda X^{-1})(X\Lambda X^{-1}) = X\Lambda^2 X^{-1} \to A^k = X\Lambda^k X^{-1}$
  - $A^{-1} = (X\Lambda X^{-1})^{-1} = X\Lambda^{-1}X^{-1}$. So Diagonalization is also quite useful to solve matrix inverse

Positive Definite

- Symmetric Matrix $A \in \mathbb{R}^{n\times n}$ is called positive definite if $\mathbf{x}^T A\mathbf{x} > 0, \mathbf{x} \in \mathbb{R}^{n\times n}$
- Symmetric Matrix $A \in \mathbb{R}^{n\times n}$ is called positive semidefinite (more usually use **PSD**) if $\mathbf{x}^T A\mathbf{x} \geq 0, \mathbf{x} \in \mathbb{R}^{n\times n}$
- Symbol:
  - $A \succ 0$: positive definite
  - $A \succeq 0$: positive semidefinite
- For example, $A = \begin{bmatrix}1 & -1\\ -1 & 1\end{bmatrix}$ is positive semidefinite with all $\mathbf{x}$ because

  $$\mathbf{x}^T A\mathbf{x} = \begin{bmatrix}u & v\end{bmatrix}\begin{bmatrix}1 & -1\\ -1 & 1\end{bmatrix}\begin{bmatrix}u\\ v\end{bmatrix} = u^2 - 2uv + v^2 = (u-v)^2 \geq 0$$

- Properties
  - All eigenvalues of positive definite matrix are positive

    Proof: $\lambda\mathbf{x} = A\mathbf{x} \to \lambda\mathbf{x}^T\mathbf{x} = \mathbf{x}^T A\mathbf{x} > 0$, $\mathbf{x}^T\mathbf{x} > 0$, so $\lambda > 0$
  - If A is positive definite, A is invertible and $\det(A) > 0$

Norm

$$\|\mathbf{x}\|_p = \left(|x_1|^p + |x_2|^p + ... + |x_n|^p\right)^{\frac{1}{p}}$$

- If $p = 2$, L2 norm: $\|\mathbf{x}\|_2 = \sqrt{x_1^2 + x_2^2 + ... + x_n^2}$

- Frobenius Norm: If $A \in \mathbb{R}^{m\times n}, \|A\|_F = \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij}^2}$
- Trace of Matrix (just for square matrix): sum of diagonal elements
- Trace Properties
  - $trace(A) = trace(A^T)$
  - $trace\left(\sum_{i=1}^{k} A_i\right) = \sum_{i=1}^{k} trace(A_i)$
  - $trace(kA) = k\, trace(A)$
  - $trace(A) = \sum_{i=1}^{D} \lambda_i$ which λ_i is the eigenvalue of matrix A
  - $trace(AB) = trace(BA)$
  - If X is invertible and same size with A: $trace(XAX^{-1}) = trace(XX^{-1}A) = trace(A)$
  - $\|A\|_F^2 = trace(A^T A)$, A is any size