DDA3020 Machine Learning
Lecture 03 Linear Algebra

Baoyuan Wu
School of Data Science, CUHK-SZ

January 17, 2024

Outline

1. Notations, vectors, matrices
2. Matrix inverse, determinant, independence
3. Systems of linear equations

Reference

References for this lecture:

- **[Book1]** Stephen Boyd and Lieven Vandenberghe, "Introduction to Applied Linear Algebra", Cambridge University Press, 2018 (available online).
- **[Book2]** Andreas C. Muller and Sarah Guido, "Introduction to Machine Learning with Python: A Guide for Data Scientists", O'Reilly Media, Inc., 2017.

1. Notations, vectors, matrices

2. Matrix inverse, determinant, independence

3. Systems of linear equations

Notations, vectors, matrices

- A **scalar** is a simple numerical value, like 15 or -3.2.
- **Variables** or **constants** that take scalar values are denoted by an italic letter, like x or a.
- We shall focus on real numbers.
- The summation over a collection $\{x_1, x_2, x_3, \ldots, x_m\}$ is denoted like this:

$$\sum_{i=1}^{m} x_i = x_1 + x_2 + \ldots + x_m$$

- The product over a collection $\{x_1, x_2, x_3, \ldots, x_m\}$ is denoted like this:

$$\prod_{i=1}^{m} x_i = x_1 \cdot x_2 \cdot \ldots \cdot x_m$$

Notations, vectors, matrices

- A **vector** is an ordered list of scalar values, called attributes. We denote a vector as a **bold character**, for example, $\mathbf{x}$ or $\mathbf{w}$.
- Vectors can be visualized as arrows that point to some directions as well as points in a multi-dimensional space.
- In many books, vectors are written column-wise:

$$\mathbf{a} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}, \qquad \mathbf{b} = \begin{bmatrix} -2 \\ 5 \end{bmatrix}, \qquad \mathbf{c} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

Notations, vectors, matrices

Illustrations of three two-dimensional vectors, $\mathbf{a} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$, $\mathbf{b} = \begin{bmatrix} -2 \\ 5 \end{bmatrix}$, $\mathbf{c} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ are given in Figure 1 below.

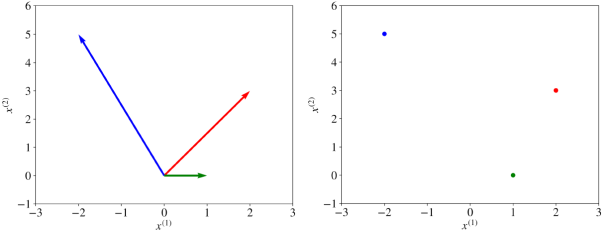


Figure: Three vectors visualized as directions and as points.

Notations, vectors, matrices

- We denote an attribute of a vector as an italic value with an index, like this: $w^{(j)}$ or $x^{(j)}$. The index j denotes a specific **dimension** of the vector, the position of an attribute in the list.
- For instance, in the vector $\mathbf{a}$ shown in red in Figure 1,

$$\mathbf{a} = \begin{bmatrix} a^{(1)} \\ a^{(2)} \end{bmatrix} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}, \text{or more commonly, } \mathbf{a} = \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$$

Note:

- The notation $x^{(j)}$ should not be confused with the power operator, such as the 2 in x^2 (squared) or 3 in x^3 (cubed).
- If we want to apply a power operator, say square, to an indexed attribute of a vector, we write like this: $(x^{(j)})^2$.

Notations, vectors, matrices

- A **matrix** is a rectangular array of numbers arranged in rows and columns. Below is an example of a matrix with two rows and three columns,

$$\mathbf{X} = \begin{bmatrix} 2 & 4 & -3 \\ 21 & -6 & -1 \end{bmatrix}$$

- Matrices are denoted with bold capital letters, such as $\mathbf{X}$ or $\mathbf{W}$.

Note:

- A variable can have two or more indices, like this: $x_i^{(j)}$ or like this $x_{i,j}^{(k)}$.
- For example, in neural networks, we denote as $x_{l,u}^{(j)}$ the input feature j of unit u in layer l.

Operations on Vectors:

$$\mathbf{x} + \mathbf{y} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} x_1 + y_1 \\ x_2 + y_2 \end{bmatrix}$$

$$\mathbf{x} - \mathbf{y} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} - \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} x_1 - y_1 \\ x_2 - y_2 \end{bmatrix}$$

$$a\mathbf{x} = a \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} ax_1 \\ ax_2 \end{bmatrix}$$

$$\frac{1}{a}\mathbf{x} = \frac{1}{a} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} \frac{1}{a}x_1 \\ \frac{1}{a}x_2 \end{bmatrix}$$

- Matrix or Vector Transpose:

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \qquad \mathbf{x}^\top = \begin{bmatrix} x_1 & x_2 \end{bmatrix}$$

$$\mathbf{X} = \begin{bmatrix} \color{red}{x_{1,1}} & \color{red}{x_{1,2}} & \color{red}{x_{1,3}} \\ \color{green}{x_{2,1}} & \color{green}{x_{2,2}} & \color{green}{x_{2,3}} \\ \color{blue}{x_{3,1}} & \color{blue}{x_{3,2}} & \color{blue}{x_{3,3}} \end{bmatrix}, \qquad \mathbf{X}^\top = \begin{bmatrix} \color{red}{x_{1,1}} & \color{green}{x_{2,1}} & \color{blue}{x_{3,1}} \\ \color{red}{x_{1,2}} & \color{green}{x_{2,2}} & \color{blue}{x_{3,2}} \\ \color{red}{x_{1,3}} & \color{green}{x_{2,3}} & \color{blue}{x_{3,3}} \end{bmatrix}$$

- **Dot Product or Inner Product of Vectors**:

$$\begin{aligned} \mathbf{x} \cdot \mathbf{y} &= \mathbf{x}^\top \mathbf{y} \\ &= \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} \\ &= x_1 y_1 + x_2 y_2 \end{aligned}$$

Notations, vectors, matrices

- **Matrix-Vector Product**

$$\mathbf{X}\mathbf{w} = \begin{bmatrix} x_{1,1} & x_{1,2} & x_{1,3} \\ x_{2,1} & x_{2,2} & x_{2,3} \\ x_{3,1} & x_{3,2} & x_{3,3} \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix} = \begin{bmatrix} x_{1,1}w_1 + x_{1,2}w_2 + x_{1,3}w_3 \\ x_{2,1}w_1 + x_{2,2}w_2 + x_{2,3}w_3 \\ x_{3,1}w_1 + x_{3,2}w_2 + x_{3,3}w_3 \end{bmatrix}$$

Matrix-Vector Product

$$\mathbf{x}^\top \mathbf{W} = \begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix} \begin{bmatrix} w_{1,1} & w_{1,2} & w_{1,3} \\ w_{2,1} & w_{2,2} & w_{2,3} \\ w_{3,1} & w_{3,2} & w_{3,3} \end{bmatrix}$$

$$= \begin{bmatrix} (w_{1,1}x_1 + w_{2,1}x_2 + w_{3,1}x_3) & (w_{1,2}x_1 + w_{2,2}x_2 + w_{3,2}x_3) & (w_{1,3}x_1 + w_{2,3}x_2 + w_{3,3}x_3) \end{bmatrix}$$

Matrix-Matrix Product

$$
\mathbf{XW} = \begin{bmatrix} x_{1,1} & \ldots & x_{1,d} \\ \vdots & \ddots & \vdots \\ x_{m,1} & \ldots & x_{m,d} \end{bmatrix} \begin{bmatrix} w_{1,1} & \ldots & w_{1,h} \\ \vdots & \ddots & \vdots \\ w_{d,1} & \ldots & w_{d,h} \end{bmatrix}
$$

$$
= \begin{bmatrix} (x_{1,1}w_{1,1} + \ldots + x_{1,d}w_{d,1}) & \ldots & (x_{1,1}w_{1,h} + \ldots + x_{1,d}w_{d,h}) \\ \vdots & \ddots & \vdots \\ (x_{m,1}w_{1,1} + \ldots + x_{m,d}w_{d,1}) & \ldots & (x_{m,1}w_{1,h} + \ldots + x_{m,d}w_{d,h}) \end{bmatrix}
$$

$$
= \begin{bmatrix} \sum_{i=1}^{d} x_{1,i}w_{i,1} & \ldots & \sum_{i=1}^{d} x_{1,i}w_{i,h} \\ \vdots & \ddots & \vdots \\ \sum_{i=1}^{d} x_{m,i}w_{i,1} & \ldots & \sum_{i=1}^{d} x_{m,i}w_{i,h} \end{bmatrix}
$$

Matrix inverse

Matrix Inverse

- Definition:
  A $d \times d$ square matrix $\mathbf{A}$ is called **invertible** (also **nonsingular**) if there exists a $d \times d$ square matrix $\mathbf{B}$ such that $\mathbf{AB} = \mathbf{BA} = \mathbf{I}$ (Identity matrix) given by

$$\mathbf{I} = \begin{bmatrix} 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & \dots & 1 & 0 \\ 0 & 0 & \dots & 0 & 1 \end{bmatrix} \qquad \text{of } d \text{ by } d \text{ dimension}$$

Matrix Inverse Computation

$$\mathbf{A}^{-1} = \frac{1}{\det(\mathbf{A})}\text{adj}(\mathbf{A})$$

- $\det(\mathbf{A})$ is the determinant of $\mathbf{A}$
- $\text{adj}(\mathbf{A})$ is the adjugate or adjoint of $\mathbf{A}$ which is the transpose of its cofactor matrix $\mathbf{C}$, *i.e.*, $\text{adj}(\mathbf{A}) = \mathbf{C}^\top$
- The cofactor $\mathbf{C}_{i,j}$ of a matrix is the (i,j)-minor $\mathbf{M}_{i,j}$ times a sign factor $(-1)^{i+j}$, *i.e.*, $\mathbf{C}_{i,j} = \mathbf{M}_{i,j} \times (-1)^{i+j}$
- $\mathbf{M}_{i,j}$ is computed through two-steps: removing the i-th row and j-th column from the original matrix to obtain a small matrix; computing the determinant of the small matrix
- For example, $\mathbf{A} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ then $\mathbf{C} = \begin{bmatrix} d & -c \\ -b & a \end{bmatrix}$, $\text{adj}(\mathbf{A}) = \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$

Determinant

Determinant computation

- Example: 2×2 matrix

$$\det(\mathbf{A}) = |\mathbf{A}| = \begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc \tag{1}$$

Determinant

Determinant computation

- Example: 3×3 matrix

$$
\begin{aligned}
|A| = \begin{vmatrix} a & b & c \\ d & e & f \\ g & h & i \end{vmatrix} &= a \begin{vmatrix} \square & \square & \square \\ \square & e & f \\ \square & h & i \end{vmatrix} - b \begin{vmatrix} \square & \square & \square \\ d & \square & f \\ g & \square & i \end{vmatrix} + c \begin{vmatrix} \square & \square & \square \\ d & e & \square \\ g & h & \square \end{vmatrix} \\
&= a \begin{vmatrix} e & f \\ h & i \end{vmatrix} - b \begin{vmatrix} d & f \\ g & i \end{vmatrix} + c \begin{vmatrix} d & e \\ g & h \end{vmatrix} \\
&= a(ei - fh) - b(di - fg) + c(dh - eg)
\end{aligned}
$$

- Each determinant of a 2×2 matrix in this equation is called a minor of the matrix A. This procedure can be extended to give a recursive definition for the determinant of a $n \times n$ matrix, known as **Laplace expansion**.
- Determinant has an elegant **geometric interpretation**. If interested, please refer to `https://spaces.ac.cn/archives/1770`.

Co-factor matrix

Consider a 3×3 matrix

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}.$$

$$\begin{vmatrix} a_{im} & a_{in} \\ a_{jm} & a_{jn} \end{vmatrix} = \det \begin{pmatrix} a_{im} & a_{in} \\ a_{jm} & a_{jn} \end{pmatrix}.$$

Its cofactor matrix is

$$\mathbf{C} = \begin{pmatrix} +\begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} & -\begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} & +\begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix} \\ -\begin{vmatrix} a_{12} & a_{13} \\ a_{32} & a_{33} \end{vmatrix} & +\begin{vmatrix} a_{11} & a_{13} \\ a_{31} & a_{33} \end{vmatrix} & -\begin{vmatrix} a_{11} & a_{12} \\ a_{31} & a_{32} \end{vmatrix} \\ +\begin{vmatrix} a_{12} & a_{13} \\ a_{22} & a_{23} \end{vmatrix} & -\begin{vmatrix} a_{11} & a_{13} \\ a_{21} & a_{23} \end{vmatrix} & +\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} \end{pmatrix},$$

Co-factor matrix

Consider a 3×3 matrix

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}.$$

$$\begin{vmatrix} a_{im} & a_{in} \\ a_{jm} & a_{jn} \end{vmatrix} = \det \begin{pmatrix} a_{im} & a_{in} \\ a_{jm} & a_{jn} \end{pmatrix}.$$

Its cofactor matrix is

$$\mathbf{C} = \begin{pmatrix} +\begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} & -\begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} & +\begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix} \\ -\begin{vmatrix} a_{12} & a_{13} \\ a_{32} & a_{33} \end{vmatrix} & +\begin{vmatrix} a_{11} & a_{13} \\ a_{31} & a_{33} \end{vmatrix} & -\begin{vmatrix} a_{11} & a_{12} \\ a_{31} & a_{32} \end{vmatrix} \\ +\begin{vmatrix} a_{12} & a_{13} \\ a_{22} & a_{23} \end{vmatrix} & -\begin{vmatrix} a_{11} & a_{13} \\ a_{21} & a_{23} \end{vmatrix} & +\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} \end{pmatrix},$$

Co-factor matrix

Consider a 3×3 matrix

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}.$$

$$\begin{vmatrix} a_{im} & a_{in} \\ a_{jm} & a_{jn} \end{vmatrix} = \det \begin{pmatrix} a_{im} & a_{in} \\ a_{jm} & a_{jn} \end{pmatrix}.$$

Its cofactor matrix is

$$\mathbf{C} = \begin{pmatrix} +\begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} & -\begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} & +\begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix} \\ -\begin{vmatrix} a_{12} & a_{13} \\ a_{32} & a_{33} \end{vmatrix} & +\begin{vmatrix} a_{11} & a_{13} \\ a_{31} & a_{33} \end{vmatrix} & -\begin{vmatrix} a_{11} & a_{12} \\ a_{31} & a_{32} \end{vmatrix} \\ +\begin{vmatrix} a_{12} & a_{13} \\ a_{22} & a_{23} \end{vmatrix} & -\begin{vmatrix} a_{11} & a_{13} \\ a_{21} & a_{23} \end{vmatrix} & +\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} \end{pmatrix},$$

Co-factor matrix

Consider a 3×3 matrix

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}.$$

$$\begin{vmatrix} a_{im} & a_{in} \\ a_{jm} & a_{jn} \end{vmatrix} = \det \begin{pmatrix} a_{im} & a_{in} \\ a_{jm} & a_{jn} \end{pmatrix}.$$

Its cofactor matrix is

$$\mathbf{C} = \begin{pmatrix} +\begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} & -\begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} & +\begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix} \\ -\begin{vmatrix} a_{12} & a_{13} \\ a_{32} & a_{33} \end{vmatrix} & +\begin{vmatrix} a_{11} & a_{13} \\ a_{31} & a_{33} \end{vmatrix} & -\begin{vmatrix} a_{11} & a_{12} \\ a_{31} & a_{32} \end{vmatrix} \\ +\begin{vmatrix} a_{12} & a_{13} \\ a_{22} & a_{23} \end{vmatrix} & -\begin{vmatrix} a_{11} & a_{13} \\ a_{21} & a_{23} \end{vmatrix} & +\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} \end{pmatrix},$$

Co-factor matrix

Consider a 3×3 matrix

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}.$$

$$\begin{vmatrix} a_{im} & a_{in} \\ a_{jm} & a_{jn} \end{vmatrix} = \det \begin{pmatrix} a_{im} & a_{in} \\ a_{jm} & a_{jn} \end{pmatrix}.$$

Its cofactor matrix is

$$\mathbf{C} = \begin{pmatrix} +\begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} & -\begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} & +\begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix} \\ -\begin{vmatrix} a_{12} & a_{13} \\ a_{32} & a_{33} \end{vmatrix} & +\begin{vmatrix} a_{11} & a_{13} \\ a_{31} & a_{33} \end{vmatrix} & -\begin{vmatrix} a_{11} & a_{12} \\ a_{31} & a_{32} \end{vmatrix} \\ +\begin{vmatrix} a_{12} & a_{13} \\ a_{22} & a_{23} \end{vmatrix} & -\begin{vmatrix} a_{11} & a_{13} \\ a_{21} & a_{23} \end{vmatrix} & +\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} \end{pmatrix},$$

Linear dependence and independence

Linear dependence and **independence**

- A collection of d-vectors $\mathbf{x}_1, \ldots, \mathbf{x}_m$ (with $m \geq 1$) is called linearly dependent if

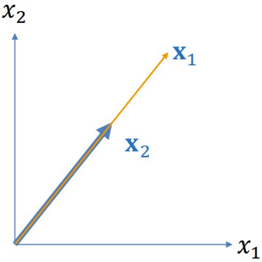
  $$\beta_1 \mathbf{x}_1 + \ldots + \beta_m \mathbf{x}_m = 0$$

  holds for some $\beta_1, \ldots, \beta_m$ that are **not all zero**.
- A collection of d-vectors $\mathbf{x}_1, \ldots, \mathbf{x}_m$ (with $m \geq 1$) is called linearly independent, which means that

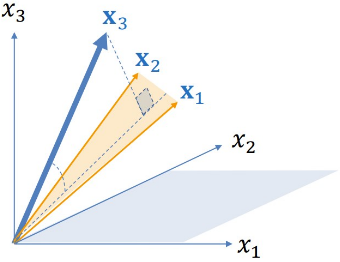
  $$\beta_1 \mathbf{x}_1 + \ldots + \beta_m \mathbf{x}_m = 0$$

  only holds for $\beta_1 = \ldots = \beta_m = 0$.

Geometry of dependency and independency



$$\beta_1 \mathbf{x}_1 + \beta_2 \mathbf{x}_2 = 0$$



$$\beta_1 \mathbf{x}_1 + \beta_2 \mathbf{x}_2 \neq \beta_3 \mathbf{x}_3$$

Systems of linear equations

- Consider a system of m linear equations in d variables $w_1, \ldots, w_d$:

$$x_{1,1}w_1 + x_{1,2}w_2 + \ldots x_{1,d}w_d = y_1$$

$$x_{2,1}w_1 + x_{2,2}w_2 + \ldots x_{2,d}w_d = y_2$$

$$\vdots$$

$$x_{m,1}w_1 + x_{m,2}w_2 + \ldots + x_{m,d}w_d = y_m$$

Systems of linear equations

These equations can be written compactly in matrix-vector notation:

$$\mathbf{X}\mathbf{w} = \mathbf{y},$$

where

$$\mathbf{X} = \begin{bmatrix} x_{1,1} & \dots & x_{1,d} \\ \vdots & \ddots & \vdots \\ x_{m,1} & \dots & x_{m,d} \end{bmatrix}, \qquad \mathbf{w} = \begin{bmatrix} w_1 \\ \vdots \\ w_d \end{bmatrix}, \qquad \mathbf{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix}.$$

Note: $\mathbf{X}$ is of $m \times d$ dimension.

(i) Square of even-determined system: $m = d$ in $\mathbf{Xw} = \mathbf{y}$, $\mathbf{X} \in \mathbb{R}^{m \times d}$ (equal number of equations and unknowns, *i.e.*, $\mathbf{X} \in \mathbb{R}^{d \times d}$)

If $\mathbf{X}$ is invertible (or $\mathbf{X}^{-1}\mathbf{X} = \mathbf{I}$), then pre-multiply both sides by $\mathbf{X}^{-1}$, we have

$$\mathbf{X}^{-1}\mathbf{X}\mathbf{w} = \mathbf{X}^{-1}\mathbf{y}$$

$$\Rightarrow \quad \mathbf{w} = \mathbf{X}^{-1}\mathbf{y}$$

If all rows or columns of $\mathbf{X}$ are **linearly independent,** then $\mathbf{X}$ is **invertible**.

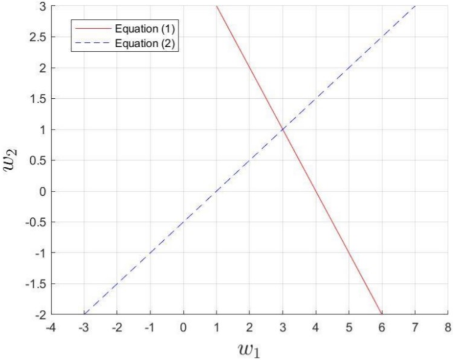
Systems of linear equations

Example $\quad w_1 + w_2 = 4 \quad (1)$

$w_1 - 2w_2 = 1 \quad (2)$

Two unknowns and two equations

$$\underbrace{\begin{bmatrix} 1 & 1 \\ 1 & -2 \end{bmatrix}}_{\mathrm{X}} \underbrace{\begin{bmatrix} w_1 \\ w_2 \end{bmatrix}}_{\mathrm{w}} = \underbrace{\begin{bmatrix} 4 \\ 1 \end{bmatrix}}_{\mathrm{y}}$$



$$\mathbf{w} = \mathbf{X}^{-1}\mathbf{y}$$

$$= \begin{bmatrix} 1 & 1 \\ 1 & -2 \end{bmatrix}^{-1} \begin{bmatrix} 4 \\ 1 \end{bmatrix} = \begin{bmatrix} 3 \\ 1 \end{bmatrix}$$

Here, the rows or columns of **X** are **linearly independent,** hence **X** is **invertible**.

Systems of linear equations

(ii)Over-determined system: $m > d$ in $\mathbf{Xw} = \mathbf{y}$, $\mathbf{X} \in \mathbb{R}^{m \times d}$
(i.e., there are more equations than unknowns)

- This set of linear equations has NO exact solution ($\mathbf{X}$ is non-square and hence not invertible). However, an **approximated solution** is yet available.
- If the left-inverse of $\mathbf{X}$ exists such that $\mathbf{X}^\dagger \mathbf{X} = \mathbf{I}$, then pre-multiply both sides by $\mathbf{X}^\dagger$ results in

$$\mathbf{X}^\dagger \mathbf{X} \mathbf{w} = \mathbf{X}^\dagger \mathbf{y}$$
$$\Rightarrow \mathbf{w} = \mathbf{X}^\dagger \mathbf{y}$$

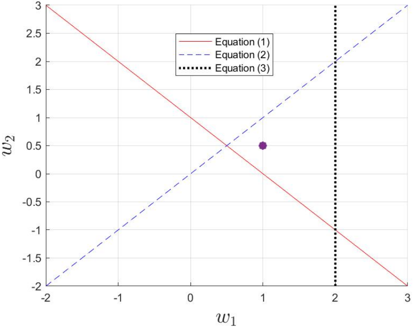
- **Definition**: a matrix **B** that satisfies BA=I (identity matrix) is called a left-inverse of **A**. (Note: **A** is m-by-d and **B** is d-by-m.
- Note: The left-inverse can be computed as $\mathbf{X}^\dagger = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top$ given $\mathbf{X}^\top \mathbf{X}$ is invertible.

Systems of linear equations

Example $$\begin{aligned} w_1 + w_2 &= 1 \quad (1) \\ w_1 - w_2 &= 0 \quad (2) \\ w_1 &= 2 \quad (3) \end{aligned}$$

Two unknowns and three equations

$$\underset{X}{\begin{bmatrix} 1 & 1 \\ 1 & -1 \\ 1 & 0 \end{bmatrix}} \quad \underset{w}{\begin{bmatrix} w_1 \\ w_2 \end{bmatrix}} = \underset{y}{\begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix}}$$



This set of linear equations has NO exact solution.

$$\mathbf{w} = \mathbf{X}^{\dagger}\mathbf{y} = (\mathbf{X}^{\top}\mathbf{X})^{-1}\mathbf{X}^{\top}\mathbf{y}$$

Here $\mathbf{X}^{\top}\mathbf{X}$ is invertible.

$$= \begin{bmatrix} 3 & 0 \\ 0 & 2 \end{bmatrix}^{-1} \begin{bmatrix} 1 & 1 & 1 \\ 1 & -1 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix} = \begin{bmatrix} 1 \\ 0.5 \end{bmatrix}$$

(Approximation)

Systems of linear equations

(iii)Under-determined system: $m < d$ in $\mathbf{X}\mathbf{w} = \mathbf{y}$, $\mathbf{X} \in \mathbb{R}^{m \times d}$
(i.e., there are more unknowns than equations $\Rightarrow$ infinite number of solutions)

- If the right-inverse of $\mathbf{X}$ exists such that $\mathbf{X}\mathbf{X}^{\dagger} = \mathbf{I}$, then the d-vector $\mathbf{w} = \mathbf{X}^{\dagger}\mathbf{y}$ (one of the infinite cases) satisfies the equation $\mathbf{X}\mathbf{w} = \mathbf{y}$, *i.e.*,

$$\mathbf{X}\mathbf{X}^{\dagger}\mathbf{y} = \mathbf{y}$$

$$\Rightarrow \mathbf{y} = \mathbf{y}$$

- **Definition**: a matrix **B** that satisfies AB=I (identity matrix) is called a right-inverse of **A**. (Note: **A** is m-by-d and **B** is d-by-m).
- Note: The right-inverse can be computed as $\mathbf{X}^{\dagger} = \mathbf{X}^{\top}(\mathbf{X}\mathbf{X}^{\top})^{-1}$ given $\mathbf{X}\mathbf{X}^{\top}$ is invertible.

Systems of linear equations

Derivation:

- Under-determined system: $m < d$ in $\mathbf{X}\mathbf{w} = \mathbf{y}$, $\mathbf{X} \in \mathbb{R}^{m \times d}$
  (i.e., there are more unknowns than equations $\Rightarrow$ infinite number of solutions $\Rightarrow$ but a unique solution is yet possible by constraining the search using $\mathbf{w} = \mathbf{X}^\top \mathbf{a}$!)
- If $\mathbf{X}\mathbf{X}^\top$ is invertible, let $\mathbf{w} = \mathbf{X}^\top \mathbf{a}$, then

$$\mathbf{X}\mathbf{X}^\top \mathbf{a} = \mathbf{y}$$

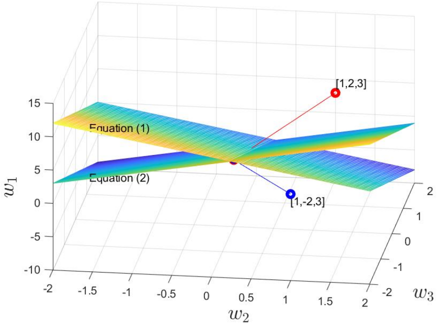
$$\Rightarrow \mathbf{a} = (\mathbf{X}\mathbf{X}^\top)^{-1}\mathbf{y}$$

$$\mathbf{w} = \mathbf{X}^\top \mathbf{a} = \underbrace{\mathbf{X}^\top (\mathbf{X}\mathbf{X}^\top)^{-1}}_{\mathbf{X}^\dagger} \mathbf{y}$$

Systems of linear equations

Example $\quad \begin{aligned} w_1 + 2w_2 + 3w_3 &= 2 \quad (1) \\ w_1 - 2w_2 + 3w_3 &= 1 \quad (2) \end{aligned}$ Three unknowns and two equations

$$\underbrace{\begin{bmatrix} 1 & 2 & 3 \\ 1 & -2 & 3 \end{bmatrix}}_{\mathrm{X}} \underbrace{\begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix}}_{\mathrm{w}} = \underbrace{\begin{bmatrix} 2 \\ 1 \end{bmatrix}}_{\mathrm{y}}$$



This set of linear equations has infinitely many solutions along the intersection line.

$\mathbf{w} = \mathbf{X}^\top(\mathbf{X}\mathbf{X}^\top)^{-1}\mathbf{y}$ Here $\mathbf{X}\mathbf{X}^\top$ is invertible.

$$= \begin{bmatrix} 1 & 1 \\ 2 & -2 \\ 3 & 3 \end{bmatrix} \begin{bmatrix} 14 & 6 \\ 6 & 14 \end{bmatrix}^{-1} \begin{bmatrix} 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 0.15 \\ 0.25 \\ 0.45 \end{bmatrix}$$

(Constrained solution)

Systems of linear equations

Example
$$w_1 + 2w_2 + 3w_3 = 2 \quad (1)$$
$$3w_1 + 6w_2 + 9w_3 = 1 \quad (2)$$

Three unknowns and two equations

$$\underbrace{\begin{bmatrix} 1 & 2 & 3 \\ 3 & 6 & 9 \end{bmatrix}}_{\text{X}} \underbrace{\begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix}}_{\text{w}} = \underbrace{\begin{bmatrix} 2 \\ 1 \end{bmatrix}}_{\text{y}}$$

Here both $\mathbf{X}\mathbf{X}^\top$ and $\mathbf{X}^\top\mathbf{X}$ are NOT invertible!

There is NO solution for the system.