Chapter 10: Linear Systems

Ketan Rajawat

Outline

Introduction

Matrix Inverse

- Properties of Matrix Inverse
- Solving Ax=b

Orthogonal Matrices

- Rotation Matrix
- Columns of Q are orthonormal
- Properties of Orthogonal matrices

Gradients and Minimization

Least Squares

- Fitting a line

Introduction

- Linear system = system of linear equations

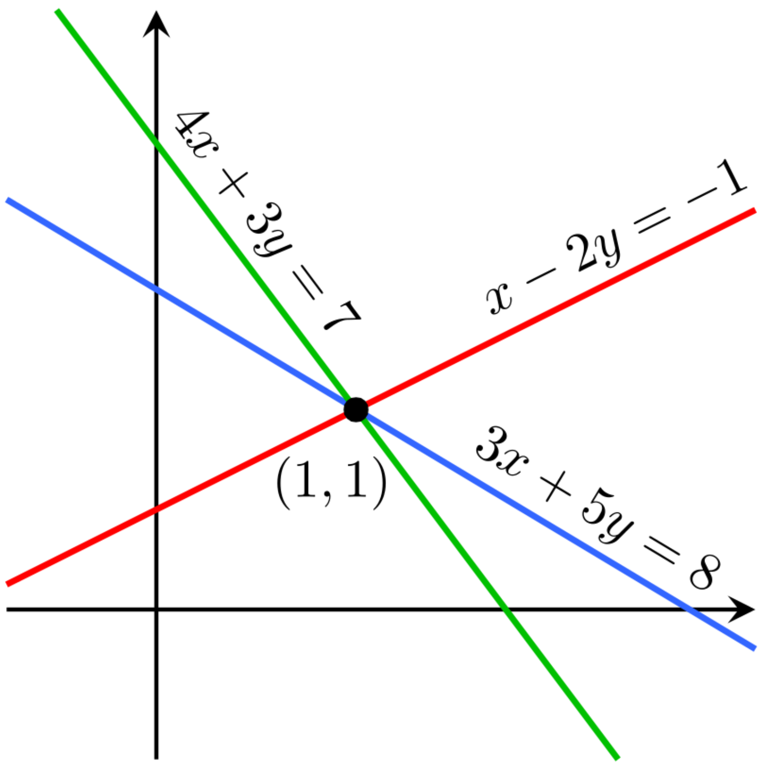
$$\mathbf{A}\mathbf{x} = \mathbf{b}$$

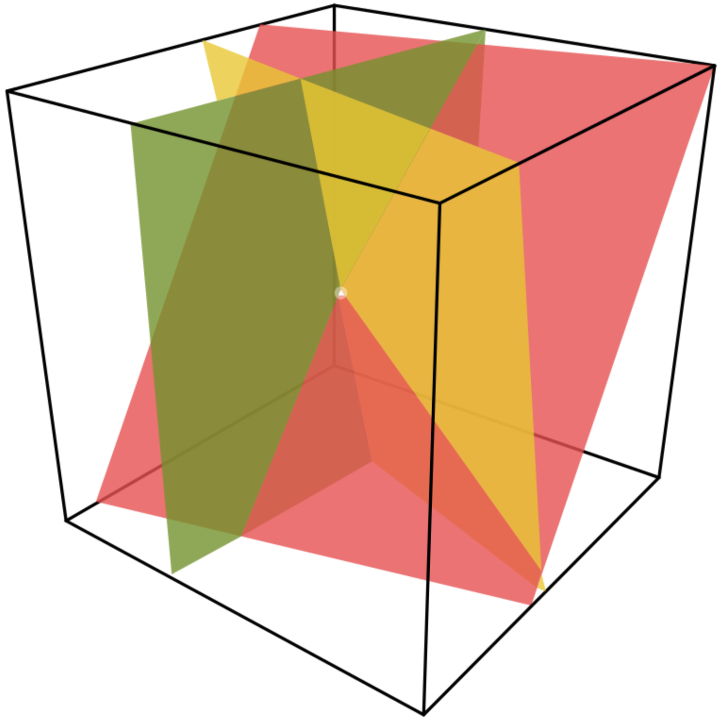
$$\begin{aligned} \mathbf{a}_1^T\mathbf{x} &= b_1 \\ \mathbf{a}_2^T\mathbf{x} &= b_2 \\ &\vdots \\ \mathbf{a}_m^T\mathbf{x} &= b_m \end{aligned}$$

$$\begin{bmatrix} \mathbf{a}_1^T \\ \mathbf{a}_2^T \\ \vdots \\ \mathbf{a}_m^T \end{bmatrix} \mathbf{x} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}$$

- Linear Algebra is all about solving *linear systems*
  - Either exactly
  - Or approximately

Simple examples





Matrix Inverse

- Scalar analogy: $ax = 1 \Rightarrow x = a^{-1}$ when $a \neq 0$
- For *square* matrices, matrix $\mathbf{B}$ is called inverse of $\mathbf{A}$ if $\mathbf{AB} = \mathbf{BA} = \mathbf{I}$
- Inverse is denoted by $\mathbf{A}^{-1}$
- Sometimes $\mathbf{A}^{-1}$ may not exist

Non-singular	Invertible	$\mathbf{A}^{-1}$ exists
Singular	Non-invertible	$\mathbf{A}^{-1}$ does not exist

Matrix Inverse: Properties

- Identity is its own inverse $\mathbf{I}^{-1} = \mathbf{I}$

Result 1: $\mathbf{A}^{-1}$ is unique for each $\mathbf{A}$

- Suppose $\mathbf{B}$ and $\mathbf{C}$ are such that $\mathbf{AB} = \mathbf{BA} = \mathbf{I}$ and $\mathbf{AC} = \mathbf{CA} = \mathbf{I}$
- Then

$$
\begin{aligned}
\mathbf{B} &= \mathbf{BI} && \\
&= \mathbf{B}(\mathbf{AC}) && \text{Since } \mathbf{AC} = \mathbf{I} \\
&= (\mathbf{BA})\mathbf{C} && \text{Associativity of matrix multiplication} \\
&= \mathbf{IC} && \text{Since } \mathbf{BA} = \mathbf{I} \\
&= \mathbf{C} &&
\end{aligned}
$$

Matrix Inverse: Properties

Result 2: For square invertible matrices $(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$

- Let $\mathbf{X} = (\mathbf{AB})^{-1}$ so that: $\mathbf{ABX} = \mathbf{I}$
- Multiplying both sides by $\mathbf{A}^{-1}$: $\mathbf{BX} = \mathbf{A}^{-1}$
- Multiplying both sides by $\mathbf{B}^{-1}$: $\mathbf{X} = \mathbf{B}^{-1}\mathbf{A}^{-1}$

Matrix Inverse: Properties

Result 3: Inverse of inverse: $(\mathbf{A}^{-1})^{-1} = \mathbf{A}$

- Suppose $\mathbf{X} = (\mathbf{A}^{-1})^{-1}$
- By definition: $\mathbf{X}\mathbf{A}^{-1} = \mathbf{I}$
- Multiplying both sides by $\boldsymbol{A}$: $\mathbf{X} = \mathbf{A}$

Matrix Inverse: Properties

Result 4: Inverse of transpose: $(\mathbf{A}^T)^{-1} = (\mathbf{A}^{-1})^T$

- Suppose: $\mathbf{X} = (\mathbf{A}^T)^{-1}$
- By definition: $\mathbf{X}\mathbf{A}^T = \mathbf{I}$
- Taking transpose of both sides: $\mathbf{A}\mathbf{X}^T = \mathbf{I}$ as $(\mathbf{AB})^T = \mathbf{B}^T\mathbf{A}^T$
- Multiplying both sides by $\mathbf{A}^{-1}$: $\mathbf{X}^T = \mathbf{A}^{-1}$
- Taking transpose of both sides: $\mathbf{X} = (\mathbf{A}^{-1})^T$

Matrix Inverse: Examples

- Diagonal Matrix:

$$\mathbf{D} = \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix} \qquad \mathbf{D}^{-1} = \begin{bmatrix} 1/2 & 0 \\ 0 & 1/3 \end{bmatrix}$$

- 2×2 matrix (please verify!)

$$\mathbf{A} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \qquad \mathbf{A}^{-1} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

(not invertible when $ad = bc$)

Solving square system of equations

- When $\mathbf{A} \in \mathbb{R}^{n \times n}$ and $\mathbf{A}$ is invertible (non-singular)
- Then system of equations $\mathbf{A}\mathbf{x} = \mathbf{b}$
- Has solution $\mathbf{x} = \mathbf{A}^{-1}\mathbf{b}$
- Since inverse is unique, the solution is also unique

- However, complexity of calculating $\mathbf{A}^{-1}$ is usually n^3

Orthogonal Matrices

$$\mathbf{Q}\mathbf{Q}^T = \mathbf{Q}^T\mathbf{Q} = \mathbf{I}$$

- Must be square
- Equivalently, $\mathbf{Q}^{-1} = \mathbf{Q}^T$

- Rotation matrix ($2D$ example):

$$\mathbf{R} = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix}$$

- Verify:

$$\mathbf{R}^T\mathbf{R} = \mathbf{R}\mathbf{R}^T = \begin{bmatrix} \cos^2(\theta) + \sin^2(\theta) & \cos(\theta)\sin(\theta) - \sin(\theta)\cos(\theta) \\ \sin(\theta)\cos(\theta) - \cos(\theta)\sin(\theta) & \sin^2(\theta) + \cos^2(\theta) \end{bmatrix} = \mathbf{I}$$

What does a rotation matrix do?

$$\begin{bmatrix} \cos 90^\circ & \sin 90^\circ \\ -\sin 90^\circ & \cos 90^\circ \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} a_2 & a_1 \end{bmatrix}$$

https://xkcd.com/184/

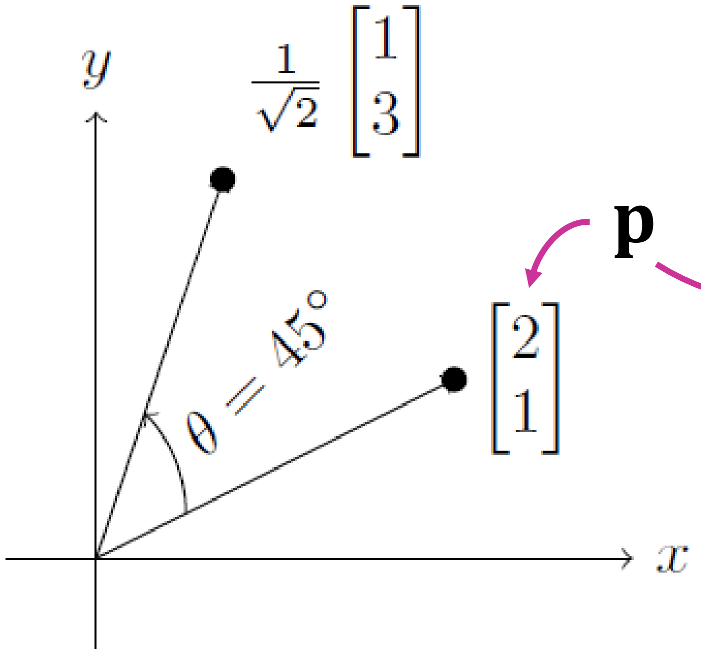
Rotation Matrix

$$\mathbf{R} = \begin{bmatrix} \cos(45°) & -\sin(45°) \\ \sin(45°) & \cos(45°) \end{bmatrix} = \begin{bmatrix} \sqrt{2}/2 & -\sqrt{2}/2 \\ \sqrt{2}/2 & \sqrt{2}/2 \end{bmatrix}$$

- $\theta = 45°$

$$\mathbf{p}' = \begin{bmatrix} \sqrt{2}/2 & -\sqrt{2}/2 \\ \sqrt{2}/2 & \sqrt{2}/2 \end{bmatrix} \begin{bmatrix} 2 \\ 1 \end{bmatrix}$$

$\mathbf{p}$



Rows and columns of $\mathbf{Q}$ are orthonormal

- Can be seen from the Gram-matrix expansion:

$$(\mathbf{Q}^T\mathbf{Q})_{ij} = \langle \mathbf{q}_i, \mathbf{q}_j \rangle = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$$

- Since $\mathbf{Q}\mathbf{Q}^T = \mathbf{I}$, columns of $\mathbf{Q}^T$ are also orthogonal
- In other words, rows of $\mathbf{Q}$ are also orthogonal

Matrices with orthonormal columns

- Converse NOT true: matrix with orthonormal columns may not be orthogonal unless it is square
- Let $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$ be orthogonal vectors with $\mathbf{v}_i \in \mathbb{R}^m$
- Let $\mathbf{A} = [\mathbf{v}_1 \;\; \mathbf{v}_2 \;\; \dots \;\; \mathbf{v}_n]$, then $\mathbf{A}^T\mathbf{A} = \mathbf{I}_n$
- But $\mathbf{A} \in \mathbb{R}^{m \times n}$ not even square ($\mathbf{A}\mathbf{A}^T \in \mathbb{R}^{m \times m} \neq \mathbf{I}_n$)
- Orthogonal when $m = n$

Example

Matrix with orthonormal columns may not be orthogonal

$$\mathbf{A} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}$$

$$\mathbf{A}^T\mathbf{A} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$\mathbf{A}\mathbf{A}^T = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

Orthogonal Matrices: Properties

- Length and inner-product preserving:

$$\langle \mathbf{Qu}, \mathbf{Qv} \rangle = (\mathbf{Qu})^T \mathbf{Qv} = \mathbf{u}^T \mathbf{Q}^T \mathbf{Qv} = \mathbf{u}^T \mathbf{v} = \langle \mathbf{u}, \mathbf{v} \rangle$$

$$\|\mathbf{Qu}\|^2 = \langle \mathbf{Qu}, \mathbf{Qu} \rangle = \langle \mathbf{u}, \mathbf{u} \rangle = \|\mathbf{u}\|^2$$

- Intuition: rotating the displacement vector

Solving $\mathbf{Qx} = \mathbf{b}$

- To solve $\mathbf{Qx} = \mathbf{b}$ we need to find $\mathbf{x} = \mathbf{Q}^T\mathbf{b}$

  (complexity = n^2)

- Compare with finding $\mathbf{x} = \mathbf{A}^{-1}\mathbf{b}$

  (complexity = n^3)

Gradients and Minimization

- If $x^\star$ minimizes $h(x)$ then $\frac{dh(x)}{dx}|_{x=x^\star} = 0$
- Example: $h(x) = \log x - ax$, for $a > 0$

$$\boxed{\frac{dh}{dx} = \frac{1}{x} - a = 0 \Rightarrow x^\star = \frac{1}{a}}$$

- Similar approach for h: $\mathbb{R}^n \to \mathbb{R}$
- If $\mathbf{x}^\star$ minimizes h, then $\frac{\partial h(\mathbf{x})}{\partial x_i}|_{\mathbf{x}=\mathbf{x}^\star} = 0$

- Compact notation: $\boxed{\boldsymbol{\nabla} h(\mathbf{x}^\star) = \mathbf{0}}$

$$[\nabla h(\mathbf{x})]_i = \frac{\partial h(\mathbf{x})}{\partial x_i}$$

Example: minimization of $\|\mathbf{x}-\mathbf{b}\|^2$

$$h(\mathbf{x}) = \|\mathbf{x}-\mathbf{b}\|^2 = \sum_{i=1}^{n}(x_i - b_i)^2$$

$$\frac{\partial h(\mathbf{x})}{\partial x_i} = 2(x_i - b_i) \;\Rightarrow x_i = b_i$$

$$\mathbf{x}^* = \mathbf{b}$$

Least Squares

- Consider rectangular system $\mathbf{Ax} = \mathbf{b}$ where $\mathbf{A} \in \mathbb{R}^{m \times n}$
- What to do when no solution?
- Least-squares problem offers a compromise: $\min_{\mathbf{x}} ||\mathbf{Ax} - \mathbf{b}||^2$

- Minimum possible error: $\min_{\mathbf{x}} ||\mathbf{Ax} - \mathbf{b}||^2$
- Optimum value of $\mathbf{x}$: $\mathbf{x}^\star = \arg\min_{\mathbf{x}} ||\mathbf{Ax} - \mathbf{b}||^2$

Observe notation

"The method of least squares is the automobile of modern statistical analysis … known and valued by nearly all." - Stigler (1981)

Least Squares: invertible $\mathbf{A}^T\mathbf{A}$

- When $\mathbf{A}^T\mathbf{A}$ is invertible (suppose $\mathbf{A} \in \mathbb{R}^{m \times n}$)
  - Only possible when $m > n$ (will see proof later)

- Solution of LS problem is given by

$$\mathbf{x}^\star = \arg\min_{\mathbf{x}} ||\mathbf{A}\mathbf{x} - \mathbf{b}||^2 = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{b}$$

Least Squares: invertible $\mathbf{A}^T\mathbf{A}$ ($n = 1$ case)

- Suppose we want to solve $\mathbf{a}x = \mathbf{b}$, for $\mathbf{a}, \mathbf{b} \in \mathbb{R}^m$ but $x \in \mathbb{R}$
- Generally, no solution unless $\mathbf{b}$ is scalar multiple of $\mathbf{a}$
- Least-squares formulation

$$x^\star = \arg\min_{x} \|\mathbf{a}x - \mathbf{b}\|^2 = \arg\min_{\mathbf{x}} \sum_{i=1}^{m} (a_i x - b_i)^2$$

$$\frac{d}{dx} \sum_{i=1}^{m} (a_i x - b_i)^2 = 2 \sum_{i=1}^{m} a_i (a_i x - b_i) = 0$$

$$\Leftrightarrow x^\star = \frac{\sum_{i=1}^{m} a_i b_i}{\sum_{i=1}^{m} a_i^2} = \frac{\mathbf{a}^T\mathbf{b}}{\mathbf{a}^T\mathbf{a}}$$

Least Squares: invertible $\mathbf{A}^T\mathbf{A}$ (general case)

$$h(\mathbf{x}) = \|\mathbf{A}\mathbf{x} - \mathbf{b}\|^2 = \sum_{i=1}^{m} \left(\mathbf{a}_{i\cdot}^T \mathbf{x} - b_i\right)^2 = \sum_{i=1}^{m} \left(\sum_{k=1}^{n} A_{ik} x_k - b_i \right)^2$$

$$\frac{\partial h(\mathbf{x})}{\partial x_j} = 2 \sum_{i=1}^{m} \left(\sum_{k=1}^{n} A_{ik} x_k - b_i \right) A_{ij} = 2 \sum_{i=1}^{m} \sum_{k=1}^{n} A_{ij} A_{ik} x_k - 2 \sum_{i=1}^{m} A_{ij} b_i$$

$$= 2 \sum_{k=1}^{n} \left(\sum_{i=1}^{m} A_{ij} A_{ik} \right) x_k - 2 \sum_{i=1}^{m} A_{ij} b_i$$

Normal Equations

$$h(\mathbf{x}) = \|\mathbf{A}\mathbf{x} - \mathbf{b}\|^2$$

$$\frac{\partial h(\mathbf{x})}{\partial x_j} = 2\sum_{k=1}^{n} \underbrace{\left(\sum_{i=1}^{m} A_{ij}A_{ik}\right)}_{[\mathbf{A}^T\mathbf{A}]_{jk}} x_k - 2\underbrace{\sum_{i=1}^{m} A_{ij}b_i}_{[\mathbf{A}^T\mathbf{b}]_j}$$

$$\nabla h(\mathbf{x}) = \mathbf{A}^T\mathbf{A}\mathbf{x} - \mathbf{A}^T\mathbf{b}$$

$$\mathbf{A}^T\mathbf{A}\mathbf{x} = \mathbf{A}^T\mathbf{b}$$ Normal Equations

$$\boxed{\mathbf{x}^\star = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{b}}$$

Fitting a line: linear regression

- Given m points in $2D$: $(x_1, y_1), (x_2, y_2), \dots, (x_m, y_m)$
- Find the best-fit line $y = wx + b$
- Least-squares formulation

$$\min_{w,b} \sum_{i=1}^{m} (y_i - wx_i - b)^2$$

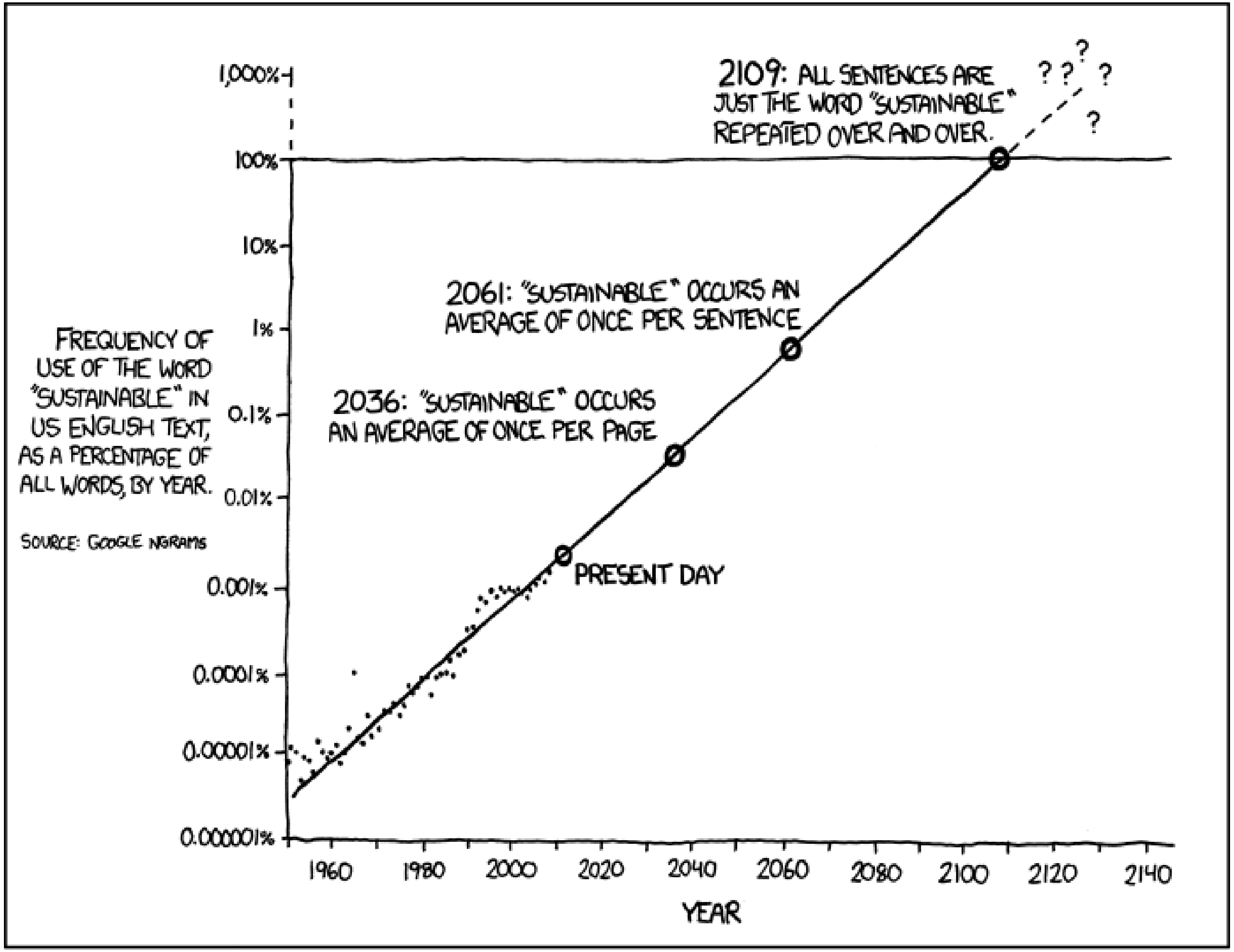
$$\min_{\mathbf{u}} \|\mathbf{y} - \mathbf{X}\mathbf{u}\|^2$$

$$\left\| \underbrace{\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix}}_{\mathbf{y}} - \underbrace{\begin{bmatrix} x_1 & 1 \\ x_2 & 1 \\ \vdots & \vdots \\ x_m & 1 \end{bmatrix}}_{\mathbf{X}} \underbrace{\begin{bmatrix} w \\ b \end{bmatrix}}_{\mathbf{u}} \right\|^2$$

$$w^\star = \frac{\sum_{i=1}^{m} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{m} (x_i - \bar{x})^2}$$

$$b^\star = \bar{y} - w^* \bar{x}$$

Verify



THE WORD "SUSTAINABLE" IS UNSUSTAINABLE.

https://xkcd.com/1007/

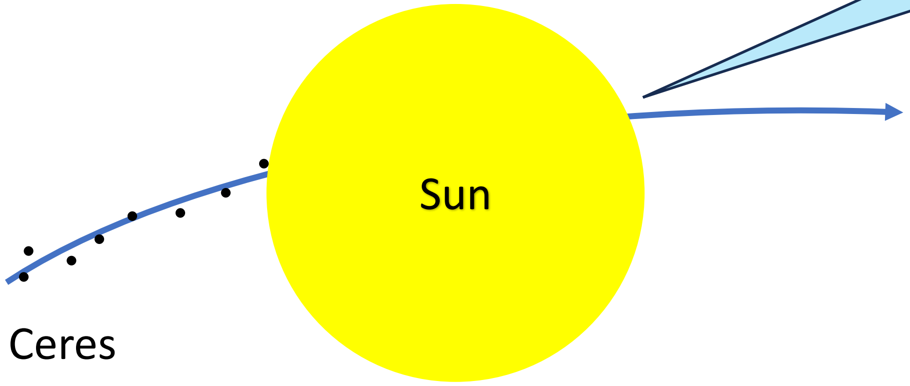
Extension: fitting a parabola

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix} \approx \begin{bmatrix} x_1^2 & x_1 & 1 \\ x_2^2 & x_2 & 1 \\ \vdots & \vdots & \vdots \\ x_m^2 & x_m & 1 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix}$$

- Also a linear system
- Can be solved using Least-Squares

History: Carl Friedrich Gauss

- Gauss first started using in 1795 (when 18)
- Considered obvious, did not publish
- Demonstration in 1801



- Officially Legendre first published in 1805

Thank You

Next: LU & QR