Assignment 2 - Numerical Linear Algebra

Arthur Rabello Oliveira

06/05/2025

Abstract

We derive linear and polynomial regression in subsets of $\mathbb{R}$ and discuss the condition number of the associated matrices, numerical algorithms for the SVD and QR factorization are built and used on an efficiency analysis of the 3 methods to do linear or polynomial regression, stability of these algorirths is mentioned and

Contents

1. Introduction

Given $D \subset \mathbb{R}^2$, a dataset, approximating this set through a *continuous* $f : \mathbb{R} \to \mathbb{R}$ is a very important problem in statistics, we will derive the 2 most important and most used methods to do this: linear and polynomial regression. Both are based on the least squares minimization problem. We will also discuss the conditioning number of the problems shown. A computational approach to regression is shown as well. We discuss how the condition number changes when the matrix is QR or SVD decomposed, and the algorithms for such decompositions are built.

2. Condition of a Problem

A *problem* is usually described as a function $f : X \to Y$ from a **normed** vector space X of data (it has to be normed so qe can *quantify* data) and a *normed* vector space Y of solutions, f is not always a well-behaved continuous function, which is why we are interested in **well-conditioned** problems and not in **ill-conditioned** problems, which we define:

Definition 2.1: (Well-Conditioned Problem) A problem $f : X \to Y$ is *well-conditioned* at $x_0 \in X \Leftrightarrow \forall \varepsilon > 0, \exists \delta > 0 \mid \|x - x_0\| < \delta \Rightarrow \|f(x) - f(x_0)\| < \varepsilon$.

This means that small perturbations in x lead to small changes in $f(x)$, a problem is **ill-conditioned** if $f(x)$ can suffer huge changes with small changes in x.

We usually say f is well-conditioned if it is well-conditioned $\forall x \in X$, if there is at least one x_i in which the problem is ill-conditioned, then we can use that whole problem is ill-conditioned.

2.1. The Condition number of a problem

Conditioning numbers are a tool to quantify how well/ill conditioned a problem is:

Definition 2.1.1: (Absolute Conditioning Number) Let δx be a small pertubation of x, so $\delta f = f(x + \delta x) - f(x)$. The **absolute** conditioning number of f is:

$$\hat{\kappa} = \lim_{\delta \to 0} \sup_{\|\delta x\| \leq \delta} \frac{\|\delta f\|}{\|\delta x\|} \tag{1}$$

The limit of the supremum can be seen as the supremum of all *infinitesimal* perturbations, so this can be rewritten as:

$$\hat{\kappa} = \sup_{\delta x} \frac{\|\delta f\|}{\|\delta x\|} \tag{2}$$

If f is differentiable, we can evaluate the abs.conditioning number using its derivative, if J is the matrix whose $i \times j$ entry is the derivative $\frac{\partial f_i}{\partial x_j}$ (jacobian of f), then we know that $\delta f \approx J(x)\delta x$, with equality in the limit $\|\delta x\| \to 0$. So the absolute conditioning number of f becomes:

$$\hat{\kappa} = \|J(x)\|, \tag{3}$$

2.2. The Relative Condition Number

When, instead of analyzing the whole set X of data, we are interested in *relative* changes, we use the **relative condition number:**

Definition 2.2.1: (Relative Condition Number) Given $f : X \to Y$ a problem, the *relative condition number* $\kappa(x)$ at $x \in X$ is:

$$\kappa(x) = \lim_{\delta \to 0} \sup_{\|\delta x\| \leq \delta} \left(\frac{\|\delta f\|}{\|f(x)\|}\right) \cdot \left(\frac{\|\delta x\|}{\|x\|}\right)^{-1} \tag{4}$$

Or, as we did in Definition 2.1.1, assuming that δf and δx are infinitesimal:

$$\kappa(x) = \sup_{\delta x} \left(\frac{\|\delta f\|}{\|f(x)\|}\right) \cdot \left(\frac{\|\delta x\|}{\|x\|}\right)^{-1} \tag{5}$$

If f is differentiable:

$$\kappa(x) = (\|J(x)\|) \cdot \left(\frac{\|f(x)\|}{\|x\|}\right)^{-1} \tag{6}$$

Relative condition numbers are more useful than absolute conditioning numbers because the **floating point arithmetic** used in many computers produces *relative* errors, the latter is not a highlight of this discussion.

Here are some examples of conditioning:

Example 2.2.1.: Consider the problem of obtaining the scalar $\frac{x}{2}$ from $x \in \mathbb{R}$. The function $f(x) = \frac{x}{2}$ is differentiablle, so by eq. (6):

$$\kappa(x) = (\|J\|) \cdot \left(\frac{\|f(x)\|}{\|x\|}\right)^{-1} = \left(\frac{1}{2}\right) \cdot \left(\frac{\frac{x}{2}}{x}\right)^{-1} = 1. \tag{7}$$

This problem is well-conditioned (κ is small).

Example 2.2.2.: Consider the problem of computing the scalar $x_1 - x_2$ from $(x_1, x_2) \in \mathbb{R}^2$ (Use the ∞-norm in $\mathbb{R}^2$ for simplicity). The function associated is differentiable and the jacobian is:

$$J = \begin{bmatrix} \frac{\partial f}{\partial x_1} & \frac{\partial f}{\partial x_2} \end{bmatrix} = \begin{bmatrix} 1 & -1 \end{bmatrix} \tag{8}$$

With $\|J\|_\infty$ = 2, so the condition number is:

$$\kappa = (\|J\|_\infty) \cdot \left(\frac{\|f(x)\|}{\|x\|}\right)^{-1} = \frac{2}{|x_1 - x_1| \cdot \max\{|x_1|, |x_2|\}} \tag{9}$$

This problem can be ill-conditioned if $|x_1 - x_2| \approx 0$ (κ gets huge), and well-conditioned otherwise

.

2.3. Condition Number of Matrices

We will deduce the conditioning number of a matrix from the conditioning number of *matrix-vector* multiplication:

Consider the problem of obtaining Ax given $A \in \mathbb{C}^{m \times n}$. We will calculate the relative condition number with respect to perturbations on x. Directly from Definition 2.2.1, we have:

$$\kappa = \sup_{\delta x} \frac{\|A(x+\delta x) - Ax\|}{\|Ax\|} \cdot \left(\frac{\|\delta x\|}{\|x\|}\right)^{-1} = \sup_{\delta x} \frac{\|A\delta x\|}{\|\delta x\|} \cdot \left(\frac{\|Ax\|}{\|x\|}\right)^{-1} \tag{10}$$

Since $\sup_{\forall x} \frac{\|A\delta x\|}{\|\delta x\|} = \|A\|$, we have:

$$\kappa = \|A\| \cdot \frac{\|x\|}{\|Ax\|} \tag{11}$$

This is a precise formula as a function of (A, x).

The following theorem will be useful in a near future:

Theorem 2.3.1: $\forall x \in \mathbb{C}^n, A \in \mathbb{C}^{n \times n}, \det(A) \neq 0$, the following holds:

$$\frac{\|x\|}{\|Ax\|} \leq \|A^{-1}\| \tag{12}$$

Proof: Since $\forall A, B \in \mathbb{C}^{n \times n}, \|AB\ \| \leq \|A\|\ \|B\|$, we have:

$$\|AA^{-1}x\ \| \leq \|Ax\|\ \|A^{-1}\| \Leftrightarrow \frac{\|x\|}{\|Ax\|} \leq \|A^{-1}\| \tag{13}$$

□

So using this in eq. (11) , we can write:

$$\kappa \leq \|A\| \cdot \|A^{-1}\| \tag{14}$$

Or:

$$\kappa = \alpha\ \|A\| \cdot \|A^{-1}\| \tag{15}$$

With

$$\alpha = \frac{\|x\|}{\|Ax\|} \cdot \left(\|A^{-1}\|\right)^{-1} \tag{16}$$

From Theorem 2.3.1, we can choose x to make $\alpha = 1$, and therefore $\kappa = \|A\| \cdot \|A^{-1}\|$.

Consider now the problem of calculating $A^{-1}b$ given $A \in \mathbb{C}^{n \times n}$. This is mathematically identical to the problem we just analyzed, so the following theorem has already been proven:

Theorem 2.3.2: Let $A \in \mathbb{C}^{n \times n}, \det(A) \neq 0$, and consider the problem of computing b, from $Ax = b$, by perturbating x. Then the following holds:

$$\kappa = \|A\| \frac{\|x\|}{\|b\|} \leq \|A\| \cdot \|A^{-1}\| \tag{17}$$

Where κ is the condition number of the problem.

Proof: Read from eq. (10) to eq. (16). □

Finally, $\|A\| \cdot \|A^{-1}\|$ is so useful it has a name: **the condition number of A** (relative to the norm $\|\cdot\|$)

If A is singular, usually we write $\kappa(A) = \infty$. Notice that if $\|\cdot\| = \|\cdot\|_2$, then $\|A\| = \sigma_1$ and $\|A^{-1}\| = \frac{1}{\sigma_m}$, so:

$$\kappa(A) = \frac{\sigma_1}{\sigma_m} \tag{18}$$

This is the condition number of A with respect to the 2-norm, which is the most used norm in practice. The condition number of a matrix is a measure of how sensitive the solution of a system of equations is to perturbations in the data. A large condition number indicates that the matrix is ill-conditioned, meaning that small changes in the input can lead to large changes in the output.

3. Linear Regression (1a)

Given a dataset of equally spaced points $D := \{t_i = \frac{i}{m}\}, i = 0, 1, ..., m \in \mathbb{R}$, linear regression consists of finding the best *line* $f(t) = \alpha + \beta t$ that approximates the points $(t_i, b_i) \in \mathbb{R}^2$, where b_i are arbitrary

Approximating 2 points in $\mathbb{R}^2$ by a line is trivial, now approximating more points is a task that requires linear algebra. To see this, we will analyze the following example to build intuition for the general case:

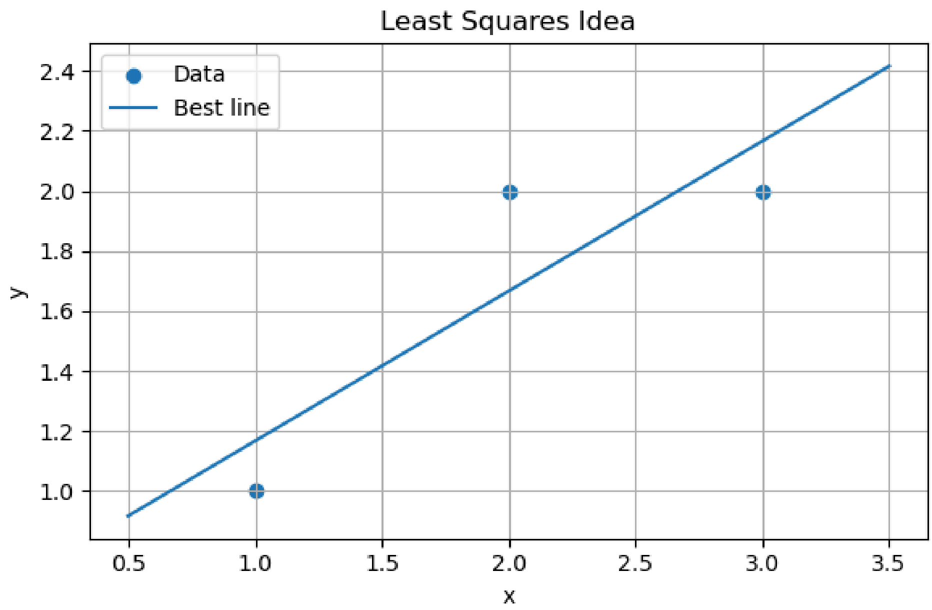

Figure 1: A glimpse into what we want to see

Given the points $(1, 1), (2, 2), (3, 2) \in \mathbb{R}^2$, we have $(t_1, b_1) = (1, 1), (t_2, b_2) = (2, 2), (t_3, b_3) = (3, 2)$ we would like a *line* $f(t) = y(t) = \alpha + \beta t$ that best approximates (t_i, b_i), in other words, since we know that the line does not pass through all 3 points, we would like to find the *closest* line to **each point** of the dataset D, so the system:

$$\begin{aligned} f(1) &= \alpha + \beta = 1 \\ f(2) &= \alpha + 2\beta = 2 \\ f(3) &= \alpha + 3\beta = 2 \end{aligned} \tag{19}$$

Which is:

$$\underbrace{\begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix}}_{A} \cdot \underbrace{\begin{bmatrix} \alpha \\ \beta \end{bmatrix}}_{x} = \underbrace{\begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}}_{b} \tag{20}$$

Clearly has no solution, (the line does not cross the 3 points), but it has a *closest solution*, which we can find through **minimizing** the errors produced by this approximation.

Let $x^* \neq x$ be a solution to the system, let the error produced by approximating the points through a line be $e = Ax - b$, we want the smaller error *square* possible (that is why least squares). We square the error to avoid and detect outliers, so:

$$e_1^2 + e_2^2 + e_3^2 \tag{21}$$

Is what we want to minimize, where e_i is the error (distance) from the ith point to the line:

Figure 2: The errors (distances)

So we will project b into $C(A)$, giving us the closest solution, and the least squares solutions is when $\hat{x}$ minimizes $\|Ax - b\|^2$, this occurs when the residual $e = Ax - b$ is orthogonal to $C(A)$, since $N(A^*) \perp C(A)$ and the dimensions sum up the left dimension of the matrix, so by the well-known projection formula, we have:

$$\begin{aligned} A^* A \hat{x} &= A^* b \\ = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 3 \end{bmatrix} \cdot \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix} \cdot \begin{bmatrix} \alpha \\ \beta \end{bmatrix} &= \begin{bmatrix} 3 & 6 \\ 6 & 14 \end{bmatrix} \cdot \begin{bmatrix} \alpha \\ \beta \end{bmatrix} \\ = \begin{bmatrix} 3 & 6 \\ 6 & 14 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} &= \begin{bmatrix} 5 \\ 11 \end{bmatrix} \end{aligned} \tag{22}$$

So the system to find $\hat{x} = \begin{bmatrix} \hat{\alpha} \\ \hat{\beta} \end{bmatrix}$ becomes:

$$\begin{aligned} 3\alpha + 5\beta &= 5 \\ 6\alpha + 14\beta &= 11 \end{aligned} \tag{23}$$

Notice that with the *errors* e_i^2 as:

$$\begin{aligned} e_1^2 &= (f(t_1) - b_1)^2 = (f(1) - 1)^2 = (\alpha + \beta - 1)^2 \\ e_2^2 &= (f(t_2) - b_2)^2 = (f(2) - 2)^2 = (\alpha + 2\beta - 2)^2 \\ e_3^2 &= (f(t_3) - b_2)^2 = (f(3) - 2)^2 = (\alpha + 3\beta - 2)^2 \end{aligned} \tag{24}$$

The system in eq. (23) is *precisely* what is obtained after using partial derivatives to minimize the erros sum as a function of (α, β):

$$\begin{aligned} f(\alpha, \beta) &= (\alpha + \beta - 1)^2 + (\alpha + 2\beta - 2)^2 + (\alpha + 3\beta - 2)^2 \\ &= 3\alpha^2 + 14\beta^2 + 12\alpha\beta - 10\alpha - 22\beta + 9, \\ \frac{\partial f}{\partial \alpha} = \frac{\partial f}{\partial \beta} = 0 &\Leftrightarrow 6\alpha + 12d - 10 = 28\beta + 12\alpha - 22 = 0 \Leftrightarrow \begin{cases} 3c + 6d = 11 \\ 6c + 14d = 11 \end{cases} \end{aligned} \tag{25}$$

This new system has a solution in $\hat{\alpha} = \frac{2}{3}, \hat{\beta} = \frac{1}{2}$, so the equation of the optimal line, obtained through *linear regression* (or least squares) is:

$$y(t) = \frac{2}{3} + \frac{1}{2}t. \tag{26}$$

If we have $n > 3$ points to approximate through a line, the reasoning is analogous:

Going back to D, we want to find the extended system as we did in eq. (25), so let the best line be:

$$f(t) = \alpha + \beta t \tag{27}$$

That best approximates the points $(0, b_0), \left(\frac{1}{m}, b_1\right), ..., (1, b_m)$. The system is:

$$\begin{aligned} f(0) &= b_0 = \alpha, \\ f\left(\frac{1}{m}\right) &= b_1 = \alpha + \frac{\beta}{m}, \\ f\left(\frac{2}{m}\right) &= b_2 = \alpha + \frac{2}{m}\beta \\ &\dots \\ f(1) &= b_m = \alpha + \beta \end{aligned} \tag{28}$$

Or:

$$\underbrace{\begin{bmatrix} 1 & 0 \\ 1 & \frac{1}{m} \\ \vdots & \vdots \\ 1 & 1 \end{bmatrix}}_{A} \cdot \underbrace{\begin{bmatrix} \alpha \\ \beta \end{bmatrix}}_{x} = \underbrace{\begin{bmatrix} b_0 \\ \vdots \\ b_m \end{bmatrix}}_{b} \tag{29}$$

Projecting into $C(A)$, we have:

$$
\begin{aligned}
A^* A x &= A^* b \\
&= \begin{bmatrix} 1 & 1 & \dots & 1 \\ 0 & \frac{1}{m} & \dots & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 \\ 1 & \frac{1}{m} \\ \vdots & \vdots \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} m+1 & \frac{m+1}{2} \\ \frac{m+1}{2} & \frac{(m+1)(2m+2)}{6m} \end{bmatrix} \cdot \begin{bmatrix} \hat{\alpha} \\ \hat{\beta} \end{bmatrix} \\
&= \begin{bmatrix} 1 & 1 & \dots & 1 \\ 0 & \frac{1}{m} & \dots & 1 \end{bmatrix} \cdot \begin{bmatrix} b_0 \\ \vdots \\ b_m \end{bmatrix} = \begin{bmatrix} b_0 + b_2 + \dots + b_m \\ \frac{1}{m}[b_1 + 2b_2 + \dots + (m-1)b_{m-1} + b_m] \end{bmatrix}
\end{aligned}
\tag{30}
$$

So the system to find the optimal vector $\begin{bmatrix} \hat{\alpha} \\ \hat{\beta} \end{bmatrix}$ is:

$$
\begin{bmatrix} m+1 & \frac{m+1}{2} \\ \frac{m+1}{2} & \frac{(m+1)(2m+2)}{6m} \end{bmatrix} \cdot \begin{bmatrix} \hat{\alpha} \\ \hat{\beta} \end{bmatrix} = \begin{bmatrix} b_0 + b_1 + \dots + b_m \\ \frac{1}{m}[b_1 + 2b_2 + \dots + (m-1)b_{m-1} + b_m] \end{bmatrix} \tag{31}
$$

Or, as a function of t_i, b_i and m:

$$
\underbrace{\begin{bmatrix} m+1 & \sum_{i=1}^{m} t_i \\ \sum_{i=1}^{m} t_i & \sum_{i=1}^{m} t_i^2 \end{bmatrix}}_{\hat{A}} \cdot \begin{bmatrix} \hat{\alpha} \\ \hat{\beta} \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{m} b_i \\ \sum_{i=1}^{m} \frac{i}{m} \cdot b_i \end{bmatrix} \tag{32}
$$

This provides the optimal vector $\hat{x}$ that minimizes the least squares error, which is the solution to the linear regression problem.

4. How the condition number of A changes (1b)

We are interested in the condition number of linear regression, which is the condition number of the matrix A in eq. (32). We will analyze how the condition number of A changes with respect to perturbations m, the number of points in the dataset. A computational approach is appropriate.

Here is a python code that numerically calculates many values of $\kappa(A) = f(m)$ as a function of m:

```python
import numpy as np
import matplotlib.pyplot as plt

def cond_number(m):

    """
    This function computes the condition number of the matrix A(m) in the 2-norm. The matrix A is defined.

    Args:
        m (float): parameter for the matrix A(m)
    Returns:
        float: condition number of A(m)
    Raises:
        ZeroDivisionError: if m = 0
        np.linalg.LinAlgError: if A(m) is not invertible
    """

```

```
    A = np.array([
        [m + 1,          (m + 1) / 2],
        [(m + 1) / 2,  (m + 1)**2 / (3 * m)]
    ])
    A_inv = np.linalg.inv(A)
    return np.linalg.norm(A, 2) * np.linalg.norm(A_inv, 2)

def main():
    M = float(input("Enter maximum m (M > 0): "))
    N = int(input("Enter number of sample points: ")) #however the user wants to plot

    m_vals = np.linspace(0, M, N)
    conds  = []

    for m in m_vals:
        try:
            conds.append(cond_number(m))
        except (ZeroDivisionError, np.linalg.LinAlgError):
            conds.append(np.inf) #if it is not invertible

    plt.figure()
    plt.plot(m_vals, conds)
    plt.xlabel('m')
    plt.ylabel('Condition number κ₂(A)')
    plt.title('Condition number of A(m) over [0, M]')
    plt.grid(True)
    plt.tight_layout()
    plt.show()

if __name__ == "__main__":
    main()
```

Good plots are:

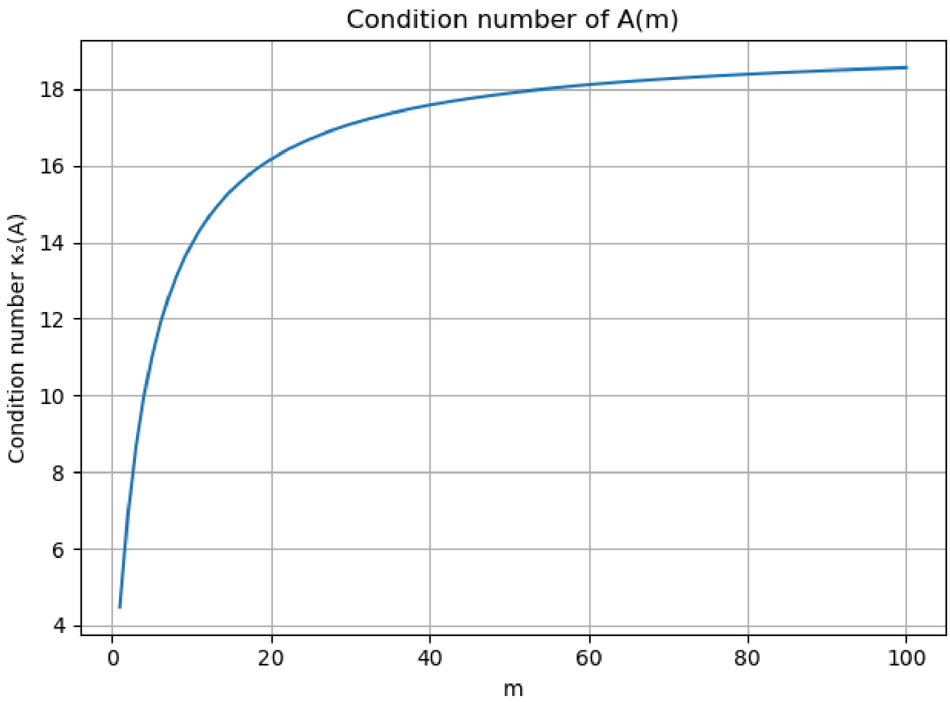

Figure 3: Condition number of A(m) over [0, 100]

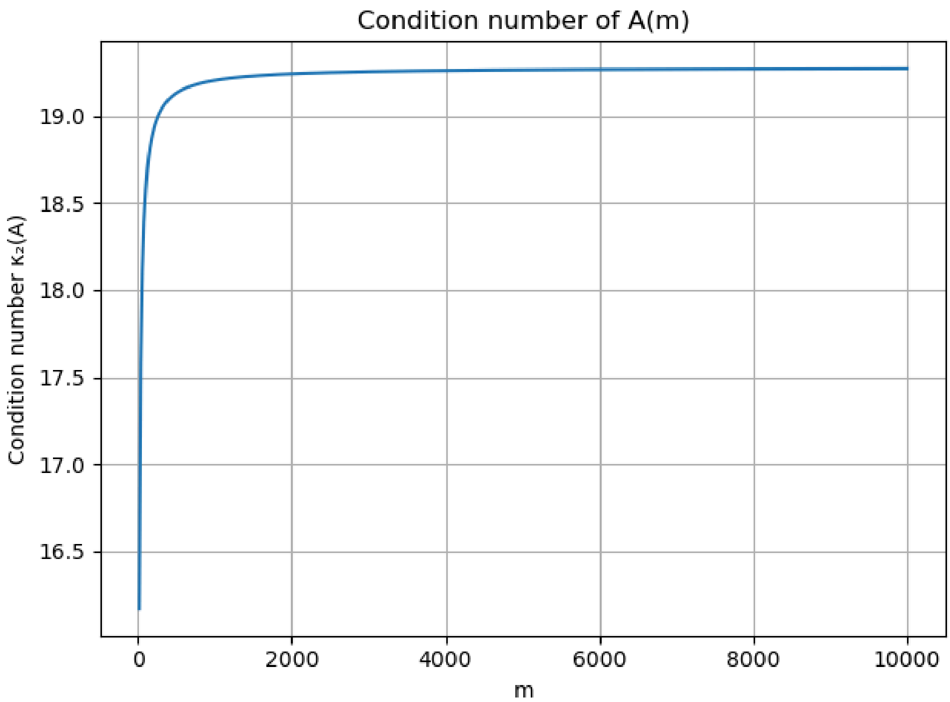

Figure 4: Condition number of A(m) over [0, 10000]

Figure 3 and Figure 4 show us that it looks like $f(m) = \kappa(A_m)$ converges to a real number, we will evaluate this hypothesis below:

Using $\| \cdot \|_2$ the conditioning number of $\hat{A} = A^* A$ in eq. (32) is:

$$\kappa\left(\hat{A}\right) = \left\|\hat{A}\right\|_2 \cdot \left\|\hat{A}^{-1}\right\|_2 = \frac{\sigma_1}{\sigma_m} \quad 33$$

Singular Values are better explored in Section 9.2. Now we will calculate the singular values of $\hat{A}$, which are the square roots of the eigenvalues of $\hat{A}$ (see Theorem 9.2.2). So we have:

$$\begin{aligned} \det\left(\hat{A}-\lambda I\right)=0 \Leftrightarrow \det\left(\begin{bmatrix} m+1-\lambda & \frac{m+1}{2} \\ \frac{m+1}{2} & \frac{(m+1)(2m+1)}{6}-\lambda \end{bmatrix}\right)=0 \\ \Leftrightarrow (m+1-\lambda)\left[\frac{(m+1)(2m+1)}{6}-\lambda\right]-\left(\frac{m+1}{2}\right)^2=0 \\ \Leftrightarrow \lambda^2-\frac{(m+1)(8m+1)}{6m}\lambda+\frac{(m+1)^2(m+2)}{12m}=0 \\ \Leftrightarrow \lambda=\frac{m+1}{12m}\left[(8m+1)\pm\sqrt{52m^2-8m+1}\right] \end{aligned} \tag{34}$$

And the singular values are:

$$\begin{aligned} \sigma_1=\sqrt{\lambda_1}=\sqrt{\frac{m+1}{12m}\left[(8m+1)+\sqrt{52m^2-8m+1}\right]}, \\ \sigma_2=\sqrt{\lambda_2}=\sqrt{\frac{m+1}{12m}\left[(8m+1)-\sqrt{52m^2-8m+1}\right]} \end{aligned} \tag{35}$$

This gives:

$$\begin{aligned} \kappa(A)=\frac{\sigma_1}{\sigma_m}=\frac{\sqrt{\frac{m+1}{12m}\left[(8m+1)+\sqrt{52m^2-8m+1}\right]}}{\sqrt{\frac{m+1}{12m}\left[(8m+1)-\sqrt{52m^2-8m+1}\right]}} \\ =\sqrt{\frac{(8m+1)+\sqrt{52m^2-8m+1}}{(8m+1)-\sqrt{52m^2-8m+1}}} \end{aligned} \tag{36}$$

And the limit as m grows is:

$$\lim_{m\to\infty}\kappa(A)=\lim_{m\to\infty}\sqrt{\frac{(8m+1)+\sqrt{52m^2-8m+1}}{(8m+1)-\sqrt{52m^2-8m+1}}} \tag{37}$$

Multiplying by the conjugate of the denominator and ignoring the square root (it is irelevant for the limit):

$$\begin{aligned} =\lim_{m\to\infty}\left[\frac{(8m+1)+\sqrt{52m^2-8m+1}}{(8m+1)-\sqrt{52m^2-8m+1}}\cdot\frac{(8m+1)+\sqrt{52m^2-8m+1}}{(8m+1)+\sqrt{52m^2-8m+1}}\right] \\ =\lim_{m\to\infty}\frac{\left((8m+1)+\sqrt{52m^2-8m+1}\right)^2}{(8m+1)^2-52m^2-8m+1} \\ =\lim_{m\to\infty}\frac{(8m+1)^2+2(8m+1)\sqrt{52m^2-8m+1}+\left(52m^2-8m+1\right)}{(8m+1)^2-\left(52m^2-8m+1\right)} \\ =\lim_{m\to\infty}\frac{64m^2+16m+1+(16m+1)\sqrt{52m^2-8m+1}+52m^2-8m+1}{64m^2+16m+1-52m^2+8m-1} \end{aligned} \tag{38}$$

Regretting having ignored the square root, and putting it back, we have:

$$
\begin{aligned}
&= \lim_{m\to\infty} \sqrt{\frac{\left((8m+1)+\sqrt{52m^2-8m+1}\right)^2}{12m^2+24m}} \\
&= \lim_{m\to\infty} \frac{(8m+1)+\sqrt{52m^2-8m+1}}{\sqrt{12m^2+24m}} \\
&= \lim_{m\to\infty} \frac{8m+1+m\sqrt{52-\frac{8}{m}+\frac{1}{m^2}}}{m\sqrt{12+\frac{24}{m}}} \\
&= \lim_{m\to\infty} \frac{m\left(8+\frac{1}{m}+\sqrt{52-\frac{8}{m}+\frac{1}{m^2}}\right)}{m\sqrt{12+\frac{24}{m}}} \\
&= \lim_{m\to\infty} \frac{8+\frac{1}{m}+\sqrt{52-\frac{8}{m}+\frac{1}{m^2}}}{\sqrt{12+\frac{24}{m}}}
\end{aligned} \tag{39}
$$

And finally:

$$
\lim_{m\to\infty} \kappa(A) = \frac{8+\sqrt{52}}{\sqrt{12}} = \frac{4+\sqrt{13}}{\sqrt{3}} \tag{40}
$$

A very good visualization of this is:

Figure 5: The purple line is the limit and the red is the function eq. (36)

Figure 5 shows the function approaching the limit. One could say that this problem is well conditioned, for $\kappa(A_m) < \frac{4+\sqrt{13}}{\sqrt{3}}, \forall m > 0$, and $\frac{4+\sqrt{13}}{\sqrt{3}}$ is not a very big number. We will not go deep into the discussion of how well-condition this problem is, but we can say that the condition number of A is not a problem for the linear regression algorithm.

5. Polynomial Regression (1c)

In this section we will discuss what changes when we decide to use **polynomials** instead of **lines** to approximate our dataset:

$$f(t) = \alpha + \beta t \to p(t) = \varphi_0 + \varphi_1 t + \ldots + \varphi_n t^n \tag{41}$$

From a first perspective, it seems way more efficient to describe a dataset with many variables then to do so with a simple line $\alpha + \beta t$, so let's use the same dataset $S := \left\{(t_i, b_i), t_i = \frac{i}{m}\right\}, i = 0, 1, \ldots, m$. Where b_i is arbitrary. As we did in Section 3, finding the new system to be solved gives us:

$$\begin{gathered} p(t_0 = 0) = b_0 = \varphi_0, \\ p\left(t_1 = \frac{1}{m}\right) = b_1 = \varphi_0 + \varphi_1 \frac{1}{m} + \ldots + \varphi_n \left(\frac{1}{m}\right)^n \\ p\left(t_2 = \frac{2}{m}\right) = b_2 = \varphi_0 + \varphi_1 \frac{2}{m} + \varphi_2 \left(\frac{2}{m}\right)^2 + \ldots + \varphi_n \left(\frac{2}{m}\right)^n \\ \vdots \\ p(t_m = 1) = b_m = \varphi_0 + \ldots + \varphi_n \end{gathered} \tag{42}$$

Or:

$$\underbrace{\begin{bmatrix} 1 & 0 & 0 & \ldots & 0 \\ 1 & \frac{1}{m} & \left(\frac{1}{m}\right)^2 & \ldots & \left(\frac{1}{m}\right)^n \\ 1 & \frac{2}{m} & \left(\frac{2}{m}\right)^2 & \ldots & \left(\frac{2}{m}\right)^n \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 1 & 1 & \ldots & 1 \end{bmatrix}}_{A_{m+1 \times n+1}} \cdot \underbrace{\begin{bmatrix} \varphi_0 \\ \varphi_1 \\ \varphi_2 \\ \vdots \\ \varphi_n \end{bmatrix}}_{\Phi_{n+1 \times 1}} = \underbrace{\begin{bmatrix} b_0 \\ b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}}_{b_{m+1 \times 1}} \tag{43}$$

Projecting into $C(A)$:

$$\begin{aligned} A^* A \hat{\Phi} &= \begin{bmatrix} 1 & 1 & 1 & \ldots & 1 \\ 0 & \frac{1}{m} & \frac{2}{m} & \ldots & 1 \\ 0 & \left(\frac{1}{m}\right)^2 & \left(\frac{2}{m}\right)^2 & \ldots & 1 \\ \vdots & \vdots & \vdots & \ldots & \vdots \\ 0 & \left(\frac{1}{m}\right)^n & \left(\frac{2}{m}\right)^n & \ldots & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 & 0 & \ldots & 0 \\ 1 & \frac{1}{m} & \left(\frac{1}{m}\right)^2 & \ldots & \left(\frac{1}{m}\right)^n \\ 1 & \frac{2}{m} & \left(\frac{2}{m}\right)^2 & \ldots & \left(\frac{2}{m}\right)^n \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 1 & 1 & \ldots & 1 \end{bmatrix} \cdot \begin{bmatrix} \widehat{\varphi_0} \\ \widehat{\varphi_1} \\ \widehat{\varphi_2} \\ \vdots \\ \widehat{\varphi_n} \end{bmatrix} \\ &= \begin{bmatrix} m+1 & \sum_{i=1}^m \frac{i}{m} & \sum_{i=1}^m \left(\frac{i}{m}\right)^2 & \ldots & \sum_{i=1}^m \left(\frac{i}{m}\right)^n \\ \sum_{i=1}^m \frac{i}{m} & \sum_{i=1}^m \left(\frac{i}{m}\right)^2 & \sum_{i=1}^m \left(\frac{i}{m}\right)^3 & \ldots & \sum_{i=1}^m \left(\frac{i}{m}\right)^{n+1} \\ \sum_{i=1}^m \left(\frac{i}{m}\right)^2 & \sum_{i=1}^m \left(\frac{i}{m}\right)^3 & \sum_{i=1}^m \left(\frac{i}{m}\right)^4 & \ldots & \sum_{i=1}^m \left(\frac{i}{m}\right)^{n+2} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \sum_{i=1}^m \left(\frac{i}{m}\right)^n & \sum_{i=1}^m \left(\frac{i}{m}\right)^{n+1} & \sum_{i=1}^m \left(\frac{i}{m}\right)^{n+2} & \ldots & \sum_{i=1}^m \left(\frac{i}{m}\right)^{2n} \end{bmatrix} \cdot \begin{bmatrix} \widehat{\varphi_0} \\ \widehat{\varphi_1} \\ \widehat{\varphi_2} \\ \vdots \\ \widehat{\varphi_n} \end{bmatrix} \\ &= \begin{bmatrix} 1 & 1 & 1 & \ldots & 1 \\ 0 & \frac{1}{m} & \frac{2}{m} & \ldots & 1 \\ 0 & \left(\frac{1}{m}\right)^2 & \left(\frac{2}{m}\right)^2 & \ldots & 1 \\ \vdots & \vdots & \vdots & \ldots & \vdots \\ 0 & \left(\frac{1}{m}\right)^n & \left(\frac{2}{m}\right)^n & \ldots & 1 \end{bmatrix} \cdot \begin{bmatrix} b_0 \\ b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix} = \begin{bmatrix} \sum_{i=0}^m b_i \\ \sum_{i=0}^m \frac{i b_i}{m} \\ \sum_{i=0}^m \left(\frac{i}{m}\right)^2 m \\ \vdots \\ \sum_{i=0}^m \left(\frac{i}{m}\right)^n b_i \end{bmatrix} \end{aligned} \tag{44}$$

So the system to be solved is:

$$\underbrace{\begin{bmatrix} m+1 & \sum_{i=1}^{m} \frac{i}{m} & \dots & \sum_{i=1}^{m} \left(\frac{i}{m}\right)^{n} \\ \sum_{i=1}^{m} \frac{i}{m} & \sum_{i=1}^{m} \left(\frac{i}{m}\right)^{2} & \dots & \sum_{i=1}^{m} \left(\frac{i}{m}\right)^{n+1} \\ \sum_{i=1}^{m} \left(\frac{i}{m}\right)^{2} & \sum_{i=1}^{m} \left(\frac{i}{m}\right)^{3} & \dots & \sum_{i=1}^{m} \left(\frac{i}{m}\right)^{n+2} \\ \vdots & \vdots & \vdots & \vdots \\ \sum_{i=1}^{m} \left(\frac{i}{m}\right)^{n} & \sum_{i=1}^{m} \left(\frac{i}{m}\right)^{n+1} & \dots & \sum_{i=1}^{m} \left(\frac{i}{m}\right)^{2n} \end{bmatrix}}_{\hat{A}} \cdot \begin{bmatrix} \widehat{\varphi}_0 \\ \widehat{\varphi}_1 \\ \widehat{\varphi}_2 \\ \vdots \\ \widehat{\varphi}_n \end{bmatrix} = \begin{bmatrix} \sum_{i=0}^{m} b_i \\ \sum_{i=0}^{m} \frac{ib_i}{m} \\ \sum_{i=0}^{m} \left(\frac{i}{m}\right)^{2} m \\ \vdots \\ \sum_{i=0}^{m} \left(\frac{i}{m}\right)^{n} b_i \end{bmatrix} \tag{45}$$

This gives the optimal vector $\hat{\Phi}$ that minimizes the least squares error. We will use computational methods to analyze this system in some of the next sections.

6. Computing the polynomial regression matrix A given (m,n) (1d)

Here is a python function that calculates the polynomial regression matrix from eq. (45), given the dimensions (m, n):

```python
import numpy as np

def poly_ls(m, n):

    """
    Build the (n+1) x (n+1) matrix A for least-squares polynomial fitting.

    Args:
        m (int): number of subintervals (m >= 0)
        n (int): polynomial degree (n >= 0)
    Returns:
        np.ndarray: shape (n+1, n+1) Gram matrix
    Raises:
        ValueError: if m or n is negative or not integer
    """

    if not isinstance(m, int) or not isinstance(n, int):
        raise ValueError("m and n must be integers")
    if m < 0 or n < 0:
        raise ValueError("m and n must be non-negative")

    x = np.linspace(0, 1, m+1) #sample space

    A = np.zeros((n+1, n+1), dtype=float) #intializes 0 matrix to be filled
    np.set_printoptions(precision=3, suppress=True)
    for j in range(n+1):
        for k in range(n+1):
            A[j, k] = np.sum(x**(j + k)) #fills each entry

```

```
    return A

for m, n in [(1, 1), (2, 2), (2, 3)]: #trivial examples
    M = poly_ls(m, n)
    print(f"m = {m}, n = {n}:")
    print(M, end="\n\n")
```

Some simple cases are:

$$A(1,1) = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}$$

$$A(2,2) = \begin{bmatrix} 3 & 1.5 & 1.25 \\ 1.5 & 1.25 & 1.125 \\ 1.25 & 1.125 & 1.062 \end{bmatrix} \tag{46}$$

$$A(2,3) = \begin{bmatrix} 3 & 1.5 & 1.25 & 1.125 \\ 1.5 & 1.25 & 1.125 & 1.062 \\ 1.25 & 1.125 & 1.062 & 1.031 \\ 1.125 & 1.062 & 1.031 & 1.016 \end{bmatrix}$$

7. How Perturbations Affect The Condition Number of A (1e)

Still on polynomial regression, in this section we analyze what happens to $\kappa(A)$, when A is perturbated with $m = 100$ and $n = 1, ..., 20$.

We will run *poly_ls(m, n)* built in Section 6 for $m = 100$ and $n = 1, ..., 20$ and then numerically calculate the condition number of the matrices, the following code is used:

```python
import numpy as np
import matplotlib.pyplot as plt

def format_scientific(x, sig=3):

    """
    Formats a number in scientific notation with a specified number of significant digits.

    Args:
        x (float): number to format
        sig (int): number of significant digits (default: 3)
    Returns:
        str: formatted string in scientific notation
    """

    if x == 0:
        return "0"
    exp = int(np.floor(np.log10(abs(x))))
    mant = x / 10**exp
    return f"{mant:.{sig}f} * 10^{exp}"
```

```

def compute_condition_numbers(m, max_n):

    """
    Returns a list of the condition numbers of the polynomial least-squares matrix A(m) for degrees n = 1 to max_n.

    Args:
        m (int): number of subintervals (m >= 0)
        max_n (int): maximum polynomial degree (max_n >= 0)
    Returns:
        list: condition numbers of A(m) for degrees n = 1 to max_n
    """

    conds = []
    for n in range(1, max_n + 1):
        A = poly_ls(m, n)
        sv = np.linalg.svd(A, compute_uv=False) #computes singular values
        conds.append(sv[0] / sv[-1]) #condition number is the ratio of the largest to smallest singular value.
    return conds

if __name__ == "__main__":
    m = 100
    max_n = 20

    cond_nums = compute_condition_numbers(m, max_n)
    n_values = np.arange(1, max_n + 1)

    print(f"Condition numbers of A (m={m}) for degree n:")
    for n, c in zip(n_values, cond_nums):
        print(f"  n = {n:2d} → κ₂(A) = {format_scientific(c)}")

    plt.figure()
    plt.semilogy(n_values, cond_nums, marker="o", linestyle="-")
    plt.xlabel("Polynomial degree $n$")
    plt.ylabel("Condition number $\\kappa(A)$")
    plt.title(f"Growth of Condition Number, $m={m}$")
    plt.grid(True, which="both", ls="--")
    plt.tight_layout()
    plt.show()
```

A good plot of the growth of the condition number is:

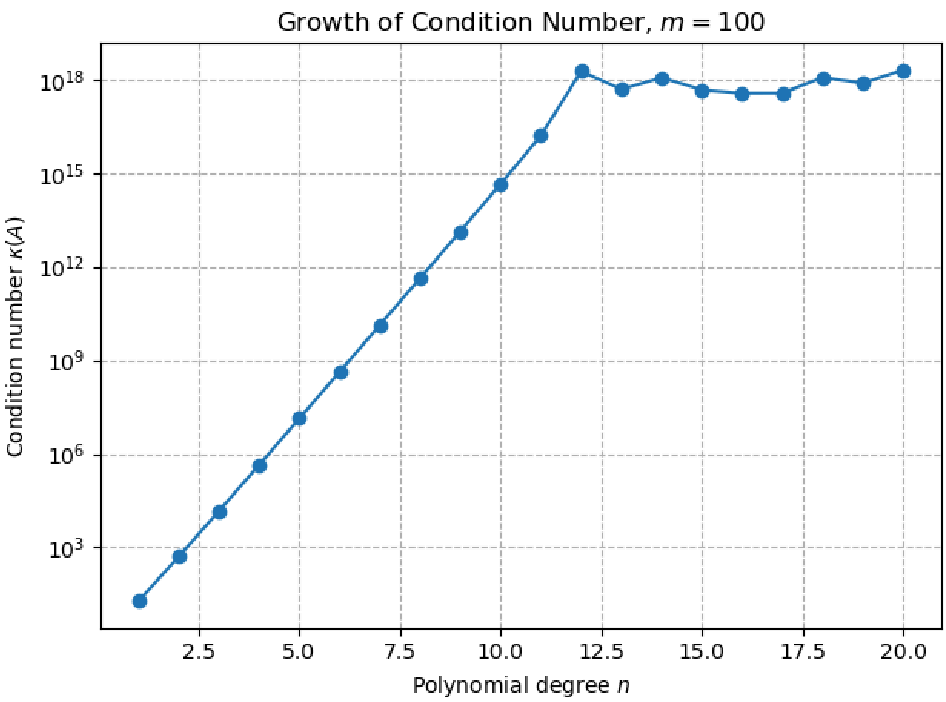

Figure 6: Growth of the condition number of A(m) for polynomial regression

Figure 6 Shows that *magic* happens

8. Polynomial Regression with a Different Dataset

8.1. A Different Dataset

If we change $S := \left\{(t_i, b_i) \mid t_i = \frac{i}{m}, i = 0, 1, ..., m\right\}$ to $\hat{S} = \left\{(t_i, b_i) \mid t_i = \frac{i}{m} - \frac{1}{2}\right\}$, the polynomial regression becomes:

$$\begin{gathered} p\left(t_0 = -\frac{1}{2}\right) = \varphi_0 + \varphi_1\left(-\frac{1}{2}\right) + ... + \varphi_n\left(-\frac{1}{2}\right)^n = b_0 \\ p\left(t_1 = \frac{1}{m} - \frac{1}{2}\right) = \varphi_0 + \varphi_1\left(\frac{1}{m} - \frac{1}{2}\right) + \varphi_2\left(\frac{1}{m} - \frac{1}{2}\right)^2 + ... + \varphi_n\left(\frac{1}{m} - \frac{1}{2}\right)^n \\ \vdots \\ p\left(t_m = 1 - \frac{1}{2}\right) = \varphi_0 + \varphi_1\left(1 - \frac{1}{2}\right) + ... + \varphi_n\left(1 - \frac{1}{2}\right)^n \end{gathered} \tag{47}$$

So:

$$\underbrace{\begin{bmatrix} 1 & -\frac{1}{2} & \left(-\frac{1}{2}\right)^2 & \cdots & \left(-\frac{1}{2}\right)^n \\ 1 & \left(\frac{1}{m} - \frac{1}{2}\right) & \left(\frac{1}{m} - \frac{1}{2}\right)^2 & \cdots & \left(\frac{1}{m} - \frac{1}{2}\right)^n \\ 1 & \left(\frac{2}{m} - \frac{1}{2}\right) & \left(\frac{2}{m} - \frac{1}{2}\right)^2 & \cdots & \left(\frac{2}{m} - \frac{1}{2}\right)^n \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & \left(-\frac{1}{2}\right) & \left(-\frac{1}{2}\right)^2 & \cdots & \left(-\frac{1}{2}\right)^n \end{bmatrix}}_{A} \cdot \underbrace{\begin{bmatrix} \varphi_0 \\ \varphi_1 \\ \vdots \\ \varphi_n \end{bmatrix}}_{\Phi} = \underbrace{\begin{bmatrix} b_0 \\ b_1 \\ \vdots \\ b_m \end{bmatrix}}_{b} \tag{48}$$

Projecting onto $C(A)$:

$$\underbrace{\begin{bmatrix} 1 & 1 & 1 & \dots & 1 \\ -\frac{1}{2} & \left(\frac{1}{m}-\frac{1}{2}\right) & \left(\frac{2}{m}-\frac{1}{2}\right) & \dots & -\frac{1}{2} \\ \left(-\frac{1}{2}\right)^2 & \left(\frac{1}{m},-\frac{1}{2}\right)^2 & \left(\frac{2}{m}-\frac{1}{2}\right)^2 & \dots & \left(-\frac{1}{2}\right)^2 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \left(-\frac{1}{2}\right)^n & \left(\frac{1}{m}-\frac{1}{2}\right)^n & \left(\frac{2}{m}-\frac{1}{2}\right)^n & \dots & \left(-\frac{1}{2}\right)^n \end{bmatrix}}_{A^*} \cdot \underbrace{\begin{bmatrix} 1 & -\frac{1}{2} & \left(-\frac{1}{2}\right)^2 & \dots & \left(-\frac{1}{2}\right)^n \\ 1 & \left(\frac{1}{m}-\frac{1}{2}\right) & \left(\frac{1}{m}-\frac{1}{2}\right)^2 & \dots & \left(\frac{1}{m}-\frac{1}{2}\right)^n \\ 1 & \left(\frac{2}{m}-\frac{1}{2}\right) & \left(\frac{2}{m}-\frac{1}{2}\right)^2 & \dots & \left(\frac{2}{m}-\frac{1}{2}\right)^n \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & -\frac{1}{2} & \left(-\frac{1}{2}\right)^2 & \dots & \left(-\frac{1}{2}\right)^n \end{bmatrix}}_{A} \cdot \underbrace{\begin{bmatrix} \widehat{\varphi_0} \\ \widehat{\varphi_1} \\ \vdots \\ \widehat{\varphi_n} \end{bmatrix}}_{\hat{\Phi}}$$

$$= \begin{bmatrix} n+1 & \sum_{i=0}^n \left(\frac{i}{m}-\frac{1}{2}\right) & \sum_{i=0}^n \left(\frac{i}{m}-\frac{1}{2}\right)^2 & \dots & \sum_{i=0}^n \left(\frac{i}{m}-\frac{1}{2}\right)^n \\ \sum_{i=0}^n \frac{i}{m}-\frac{1}{2} & \sum_{i=0}^n \left(\frac{i}{m}-\frac{1}{2}\right)^2 & \sum_{i=0}^n \left(\frac{i}{m}-\frac{1}{2}\right)^3 & \dots & \sum_{i=0}^n \left(\frac{i}{m}-\frac{1}{2}\right)^{n+1} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \sum_{i=0}^n \left(\frac{i}{m}-\frac{1}{2}\right)^n & \sum_{i=0}^n \left(\frac{i}{m}-\frac{1}{2}\right)^{n+1} & \sum_{i=0}^n \left(\frac{i}{m}-\frac{1}{2}\right)^{n+2} & \dots & \sum_{i=0}^n \left(\frac{i}{m}-\frac{1}{2}\right)^{2n} \end{bmatrix} \cdot \begin{bmatrix} \widehat{\varphi_0} \\ \widehat{\varphi_1} \\ \vdots \\ \widehat{\varphi_n} \end{bmatrix} \quad 49$$

$$= \begin{bmatrix} 1 & 1 & 1 & \dots & 1 \\ -\frac{1}{2} & \left(\frac{1}{m}-\frac{1}{2}\right) & \left(\frac{2}{m}-\frac{1}{2}\right) & \dots & -\frac{1}{2} \\ \left(-\frac{1}{2}\right)^2 & \left(\frac{1}{m},-\frac{1}{2}\right)^2 & \left(\frac{2}{m}-\frac{1}{2}\right)^2 & \dots & \left(-\frac{1}{2}\right)^2 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \left(-\frac{1}{2}\right)^n & \left(\frac{1}{m}-\frac{1}{2}\right)^n & \left(\frac{2}{m}-\frac{1}{2}\right)^n & \dots & \left(-\frac{1}{2}\right)^n \end{bmatrix} \cdot \begin{bmatrix} b_0 \\ b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix} = \begin{bmatrix} \sum_{i=0}^n b_i \\ \sum_{i=0}^n \left(\frac{i}{m}-\frac{1}{2}\right) b_i \\ \sum_{i=0}^n \left(\frac{i}{m}-\frac{1}{2}\right)^2 b_i \\ \vdots \\ \sum_{i=0}^n \left(\frac{i}{m}-\frac{1}{2}\right)^n b_i \end{bmatrix}$$

$\therefore$ The least squares optimal solution $(\widehat{\varphi_0}, \widehat{\varphi_1}, \dots, \widehat{\varphi_n})$ is:

$$\begin{bmatrix} n+1 & \sum_{i=0}^n \left(\frac{i}{m}-\frac{1}{2}\right) & \dots & \sum_{i=0}^n \left(\frac{i}{m}-\frac{1}{2}\right)^n \\ \sum_{i=0}^n \frac{i}{m}-\frac{1}{2} & \sum_{i=0}^n \left(\frac{i}{m}-\frac{1}{2}\right)^2 & \dots & \sum_{i=0}^n \left(\frac{i}{m}-\frac{1}{2}\right)^{n+1} \\ \vdots & \vdots & \vdots & \vdots \\ \sum_{i=0}^n \left(\frac{i}{m}-\frac{1}{2}\right)^n & \sum_{i=0}^n \left(\frac{i}{m}-\frac{1}{2}\right)^{n+1} & \dots & \sum_{i=0}^n \left(\frac{i}{m}-\frac{1}{2}\right)^{2n} \end{bmatrix}^{-1} \cdot \begin{bmatrix} \sum_{i=0}^n b_i \\ \sum_{i=0}^n \left(\frac{i}{m}-\frac{1}{2}\right) b_i \\ \sum_{i=0}^n \left(\frac{i}{m}-\frac{1}{2}\right)^2 b_i \\ \vdots \\ \sum_{i=0}^n \left(\frac{i}{m}-\frac{1}{2}\right)^n b_i \end{bmatrix} \quad 50$$

We provide numerical examples in the next section for a better visualization of eq. (50).

8.2. How Conditioning changes (1f)

BOTA AS PORRA DOS GRAFICO

9. Least Squares with QR and SVD decompositions

We have shown the solutions to the least squares problem $Ax = b$, but this problem could be solved with factorizations of A, such as the QR and SVD, in the following sections we will show these factorizations and use them to solve the least squares problem.

9.1. QR

The QR factorization of a full-rank $A \in \mathbb{C}^{m \times n}$, $m \geq n$ an consists of finding orthonormal vectors $q_1, \dots, q_n$ such that $q_1, \dots, q_i$ spans $a_1, \dots, q_1$, where a_i is the ith-column of A. So we want:

$$\begin{gathered} \operatorname{span}(a_1) = \operatorname{span}(q_1) \\ \operatorname{span}(a_1, a_2) = \operatorname{span}(q_1, q_2) \\ \vdots \\ \operatorname{span}(a_1, \dots, a_n) = \operatorname{span}(q_1, \dots, q_n) \end{gathered} \quad 51$$

This is equivalent to:

$$A = \begin{bmatrix} q_1 & \dots & q_n \end{bmatrix} \cdot \begin{bmatrix} r_{11} & r_{12} & \dots & r_{1n} \\ & r_{22} & \dots & r_{2n} \\ & & \vdots & \vdots \\ & & & r_{nn} \end{bmatrix} \quad (52)$$

Where $r_{ii} \neq 0$, because a_i will be expressed as a linear combination of q_i, and since the triangular matrix is invertible, q_i can be expressed as a linear combination of a_i. Therefore eq. (52) is:

$$\begin{aligned} a_1 &= q_1 r_{11}, \\ a_2 &= r_{12} q_1 + r_{22} q_2, \\ &\vdots \\ a_n &= r_{1n} q_1 + r_{2n} q_2 + \dots + r_{nn} q_n. \end{aligned} \quad (53)$$

Or:

$$A = \hat{Q} \hat{R} \quad (54)$$

Is the *reduced* QR decomposition of A.

The *full* QR decomposition of $A \in \mathbb{C}^{m \times n}$ not of full-rank is analogous to the reduced, but $|m - n|$ 0-columns are appended to $\hat{Q}$ to make it a unitary $m \times m$ matrix Q, and 0-rows are aded to $\hat{R}$ to make it a $m \times n$ still triangular matrix:

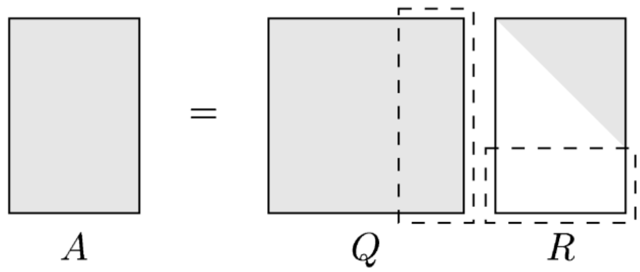

Figure 7: Full QR factorization

And the decomposition becomes:

$$A = QR \quad (55)$$

Here are some examples:

Example 9.1.1.:

$$A = \begin{bmatrix} 3 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 5 \end{bmatrix} \quad (56)$$

This is a diagonal matrix, so its QR factorization is particularly simple:

$$Q = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, R = \begin{bmatrix} 3 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 5 \end{bmatrix} \quad (57)$$

With diagonal matrices, Q is the identity matrix and $R = A$.

Example 9.1.2.:

$$A = \begin{bmatrix} 1 & 1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \quad (58)$$

For this 3×2 matrix, we compute the reduced QR factorization:

$$\hat{Q} = \begin{bmatrix} \frac{1}{\sqrt{2}} & 0 \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ 0 & \frac{1}{\sqrt{2}} \end{bmatrix}, \hat{R} = \begin{bmatrix} \sqrt{2} & \frac{1}{\sqrt{2}} \\ 0 & \frac{\sqrt{2}}{2} \end{bmatrix} \quad (59)$$

This is a reduced QR factorization where $\hat{Q}$ is 3×2. The full QR factorization would require extending $\hat{Q}$ to a 3×3 orthogonal matrix and adding a row of zeros to $\hat{R}$ as shown in Figure 7.

9.2. SVD

The *singular value decomposition* of a matrix is based on the fact that the image of the unit sphere under a $m \times n$ matrix is a **hyperellipse:**

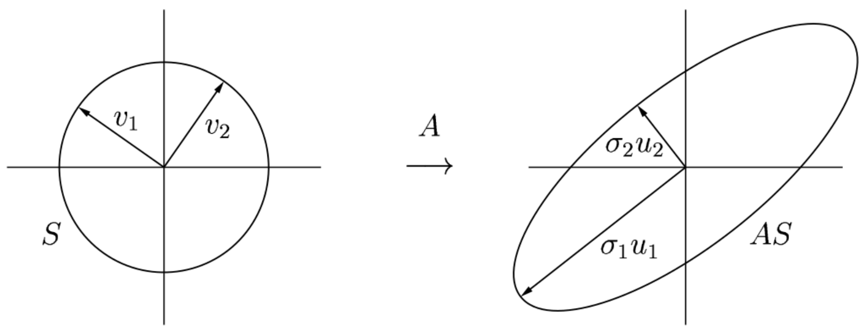

Figure 8: SVD of a 2×2 matrix

So the independent directions v_1, v_2 have been mapped to another set of orthogonal directions $\sigma_1 v_1, \sigma_2 v_2$, so with $S := \{v \in \mathbb{C}^n \mid \|v\| = 1\}$ as the unit ball, let's define:

Definition 9.2.1: (Singular Values) The n *singular values* σ_i of $A \in \mathbb{C}(m \times n)$ are the lengths of the n new axes of AS, written in non-crescent order $\sigma_1 \geq \ldots \geq \sigma_n$.

Definition 9.2.2: (Left Singular Vectors) The n **left** singular vectors of A are the unit vectors u_i laying in AS, oriented to correspond and number the singular values σ_i, respectively

Definition 9.2.3: (Right Singular Vectors) The **right** singular vectors of A are the v_i in S that are the preimages of $\sigma_i u_i \in AS$, such that $A v_i = \sigma_i u_i$

The equation $A v_i = \sigma_i u_i$ is equivalent to:

$$A \cdot \begin{bmatrix} v_1 & v_2 & \ldots & v_n \end{bmatrix} = \begin{bmatrix} \sigma_1 u_1 & \sigma_2 u_2 & \ldots & \sigma_n u_n \end{bmatrix} \quad (60)$$

Better:

$$A \cdot \begin{bmatrix} v_1 & v_2 & \dots & v_n \end{bmatrix} = \begin{bmatrix} u_1 & u_2 & \dots & u_n \end{bmatrix} \cdot \begin{bmatrix} \sigma_1 & 0 & \dots & 0 \\ 0 & \sigma_2 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & \sigma_n \end{bmatrix} \quad (61)$$

Or simple $AV = U\Sigma$, but since V has orthonormal columns:

$$A = U\Sigma V^* \quad (62)$$

The SVD is a very particular factorization for matrices, as the following theorem states:

Theorem 9.2.1: (Existence of SVD) *Every* matrix $A \in \mathbb{C}^{m \times n}$ has a singular value decomposition

Proof: We prove the existane by fixing the largest image of A and using induction on the dimension of A:

Let $\sigma_1 = \|A\|_2$. There must exist unitary vectors $u_1, v_1 \in \mathbb{C}^n$ such that $Av_1 = \sigma_1 u_1$, with $\|v_1\|_2 = \|u_1\|_2 = 1$. Let $\{v_j\}$ and $\{u_j\}$ be 2 orthonormal bases of $\mathbb{C}^n$. These column vectors form the unitary matrices V_1 and U_1. We will compute:

$$\Phi = U_1^* A V_1 \quad (63)$$

Notice that the first column of Φ is $U_1^* A v_1 = \sigma_1 U_1^* v_1 = \sigma_1 e_1$, since u_1 is the first column of U_1. So Φ looks like:

$$\Phi = \begin{bmatrix} \sigma_1 & w^* \\ 0 & B \end{bmatrix} \quad (64)$$

Where w^* is the rest of the first row, the action of A onto the remaining columns v_j. B acts on the subspace orthogonal to v_1.

We want $w = 0$, we can force this by using the norm. We know that:

$$\left\| \begin{bmatrix} \sigma_1 & w^* \\ 0 & B \end{bmatrix} \cdot \begin{bmatrix} \sigma_1 \\ w \end{bmatrix} \right\|_2 = \left\| \begin{bmatrix} \sigma_1^2 + w^* w \\ Bw \end{bmatrix} \right\|_2 = \sqrt{|\sigma_1^2 + w^* w|^2 + \|Bw\|_2^2} \quad (65)$$

And:

$$\sqrt{|\sigma_1^2 + w^* w|^2 + \|Bw\|_2^2} \geq \sigma_1^2 + w^* w \quad (66)$$

We also know:

$$\|\Phi\|_2 = \sup_{\|y\|=1} \|\Phi y\|_2 \quad (67)$$

For the specific $x = [\sigma_1, w]$ scaled to the unit ball, and knowing $\|\Phi\|_2 = \sigma_1$, we have:

$$\|\Phi\|_2 \geq \frac{\|\Phi x\|_2}{\|x\|_2} \geq \frac{\sigma_1^2 + w^* w}{\sqrt{\sigma_1^2 + w^* w}} = \sqrt{\sigma_1^2 + w^* w} \Leftrightarrow \sigma_1 \geq \sqrt{\sigma_1^2 + w^* w}$$
$$\Leftrightarrow \sigma_1^2 \geq \sigma_1^2 + w^* w \Leftrightarrow w^* w = 0 \Leftrightarrow w = 0. \quad (68)$$

If $m = 1$ or $n = 1$, we are done, If not, B has an SVD decomposition $B = U_2 \Sigma_2 V_2^*$ by the induction hypothesis, so from eq. (63) we have that the following is a SVD decomposition of A, completing the proof:

$$A = U_1 \begin{bmatrix} 1 & 0 \\ 0 & U_2 \end{bmatrix} \begin{bmatrix} \sigma_1 & 0 \\ 0 & \Sigma_2 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & V_2 \end{bmatrix}^* V_1^* \qquad (69)$$

□

Is the SVD factorization of A. There are more about the SVD on computing U, Σ, V^*, as we will show below:

Theorem 9.2.2: $\forall A \in \mathbb{C}^{m \times n}$, the following holds:

- The eigenvalues of $A^* A$ are the singular values *squared* of A, and the column-eigenvectors of $A^* A$ form the matrix V.
- The eigenvalues of AA^* are the singular values *squared* of A, and the column-eigenvectors of AA^* form the matrix U.

Proof: Let $U \Sigma V^* = A$ be the SVD of A, then computing $A^* A$, knowing U, V are unitary matrices, we have:

$$A^* A = (U \Sigma V^*)^* (U \Sigma V^*) = V \Sigma^* U^* U \Sigma V^* = V \Sigma^* \Sigma V^* = V \Sigma^2 V^* \qquad (70)$$

This is an *eigenvalue* decomposition of $A^* A$, where the eigenvalues are the entries of Σ^2, which are the singular values of A **squared**, and the eigenvectors are the columns of V.

For AA^*, we have:

$$AA^* = (U \Sigma V^*)(U \Sigma V^*)^* = U \Sigma V^* V \Sigma^* U^* = U \Sigma \Sigma^* U^* = U \Sigma^2 U^* \qquad (71)$$

The reasoning here is analogous. So the proof is complete. □

By Theorem 9.2.2, calculating the SVD of A has been reduced to calculating the eigenvalues and eigenvectors of $A^* A$ and AA^*, here are some examples of singular value decompositions:

Example 9.2.1.: Consider $A = \begin{bmatrix} 3 & 2 \\ 2 & 3 \end{bmatrix}$. Computing the SVD:

First, find $A^* A = \begin{bmatrix} 13 & 12 \\ 12 & 13 \end{bmatrix}$ and calculate its eigenvalues: $\lambda_1 = 25, \lambda_2 = 1$

The singular values are $\sigma_1 = 5, \sigma_2 = 1$.

The right singular vectors (eigenvectors of $A^* A$): $V = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix}$

The left singular vectors (obtained from $A v_i = \sigma_i u_i$): $U = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix}$

Therefore, the SVD is: $A = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix} \cdot \begin{bmatrix} 5 & 0 \\ 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix}^*$

Example 9.2.2.: Consider a non-square matrix $A = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix}$. For this 2×3 matrix, for the SVD we do:

$$A^*A = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 1 & 1 \end{bmatrix} \quad (72)$$

The eigenvalues of A^*A are $\lambda_1 = 3, \lambda_2 = 1, \lambda_3 = 0$, so the singular values are $\sigma_1 = \sqrt{3}, \sigma_2 = 1, \sigma_3 = 0$

The right singular vectors (eigenvectors of A^*A) are:

$$V = \begin{bmatrix} \frac{1}{2} & -\frac{1}{\sqrt{2}} & \frac{1}{2} \\ \frac{1}{\sqrt{2}} & 0 & -\frac{1}{\sqrt{2}} \\ \frac{1}{2} & \frac{1}{\sqrt{2}} & \frac{1}{2} \end{bmatrix} \quad (73)$$

And now for AA^*:

$$AA^* = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \quad (74)$$

The eigenvalues are $\lambda_1 = 3, \lambda_2 = 1$, so the singular values are $\sigma_1 = \sqrt{3}, \sigma_2 = 1$. The eigenvectors are:

$$U = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix} \quad (75)$$

Therefore, the full SVD is:

$$A = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix} \cdot \begin{bmatrix} \sqrt{3} & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} \frac{1}{2} & -\frac{1}{\sqrt{2}} & \frac{1}{2} \\ \frac{1}{\sqrt{2}} & 0 & -\frac{1}{\sqrt{2}} \\ \frac{1}{2} & \frac{1}{\sqrt{2}} & \frac{1}{2} \end{bmatrix}^* \quad (76)$$

9.3. Least Squares with QR and SVD

Here we will write code that solves the least squares problem usig the 2 factorizations shown in Section 9.1 and Section 9.2, as well as the ordinary approach to least squares shown in Section 3.

9.4. Examples (2b)

We will also use these algorithms to do linear regression on the simple functions $f, g, h : \mathbb{R} \to \mathbb{R}$ defined as:

$$\begin{aligned} f(t) &= \sin(t) \\ g(t) &= e^t \\ h(t) &= \cos(3t) \end{aligned} \quad (77)$$

BOTA OS CODIGO

9.5. How good are the approximations? (2c)