Least Squares and Data Fitting
Data fitting

How do we best fit a set of data points?
Linear Least Squares

1) Fitting with a line

Given $m$ data points $\{(t_1, y_1), \ldots, (t_m, y_m)\}$, we want to find the function

$$y = x_0 + x_1 t$$

that best fit the data (or better, we want to find the coefficients $x_0, x_1$).

Thinking geometrically, we can think “what is the line that most nearly passes through all the points?”

![Graph showing data points and a line of best fit.](image-url)
Given $m$ data points $\{\{t_1, y_1\}, \ldots, \{t_m, y_m\}\}$, we want to find $x_0$ and $x_1$ such that

$$y_i = x_0 + x_1 t_i \quad \forall i \in [1, m]$$

or in matrix form:

$$
\begin{pmatrix}
1 & t_1 \\
\vdots & \vdots \\
1 & t_m
\end{pmatrix}
\begin{bmatrix}
x_0 \\
x_1
\end{bmatrix}
= 
\begin{bmatrix}
y_1 \\
\vdots \\
y_m
\end{bmatrix}
\Rightarrow A x = b
$$

Note that this system of linear equations has more equations than unknowns – OVERDETERMINED SYSTEMS

We want to find the appropriate linear combination of the columns of $A$ that makes up the vector $b$.

If a solution exists that satisfies $A x = b$ then $b \in \text{range}(A)$
Linear Least Squares

- In most cases, $b \not\in \text{range}(A)$ and $A \, x = b$ does not have an exact solution!

- Therefore, an overdetermined system is better expressed as
  \[ A \, x \approx b \]
Linear Least Squares

• **Least Squares**: find the solution $\mathbf{x}$ that minimizes the residual

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

• Let’s define the function $\phi$ as the square of the 2-norm of the residual

$$ \phi(\mathbf{x}) = \| \mathbf{b} - A \mathbf{x} \|_2^2 $$
Linear Least Squares

- **Least Squares**: find the solution $\boldsymbol{x}$ that minimizes the residual

  $$r = \boldsymbol{b} - A \boldsymbol{x}$$

- Let’s define the function $\phi$ as the square of the 2-norm of the residual

  $$\phi(\boldsymbol{x}) = \|\boldsymbol{b} - A \boldsymbol{x}\|_2^2$$

- Then the least squares problem becomes

  $$\min_{\boldsymbol{x}} \phi(\boldsymbol{x})$$

- Suppose $\phi: \mathbb{R}^m \to \mathbb{R}$ is a smooth function, then $\phi(\boldsymbol{x})$ reaches a (local) maximum or minimum at a point $\boldsymbol{x}^* \in \mathbb{R}^m$ only if

  $$\nabla \phi(\boldsymbol{x}^*) = 0$$
How to find the minimizer?

- To minimize the 2-norm of the residual vector

\[
\min_{\mathbf{x}} \phi(\mathbf{x}) = \| \mathbf{b} - A \mathbf{x} \|^2_2 
\]

\[
\phi(\mathbf{x}) = (\mathbf{b} - A \mathbf{x})^T (\mathbf{b} - A \mathbf{x}) 
\]

\[
\nabla \phi(\mathbf{x}) = 2(A^T \mathbf{b} - A^T A \mathbf{x}) \]

**Normal Equations** – solve a linear system of equations

First order necessary condition:
\[
\nabla \phi(\mathbf{x}) = 0 \rightarrow A^T \mathbf{b} - A^T A \mathbf{x} = 0 \rightarrow A^T A \mathbf{x} = A^T \mathbf{b} 
\]

Second order sufficient condition:
\[
D^2 \phi(\mathbf{x}) = 2A^T A 
\]

\(2A^T A\) is a positive semi-definite matrix → the solution is a minimum
Linear Least Squares (another approach)

- Find \( y = Ax \) which is closest to the vector \( b \)
- What is the vector \( y = Ax \in \text{range}(A) \) that is closest to vector \( y \) in the Euclidean norm?

When \( r = b - y = b - Ax \) is orthogonal to all columns of \( A \), then \( y \) is closest to \( b \)

\[
A^T r = A^T (b - Ax) = 0 \quad \rightarrow \quad A^T Ax = A^T b
\]
Summary:

- $A$ is a $m \times n$ matrix, where $m > n$.
- $m$ is the number of data pair points. $n$ is the number of parameters of the “best fit” function.

- Linear Least Squares problem $A \, x \cong b$ always has solution.

- The Linear Least Squares solution $x$ minimizes the square of the 2-norm of the residual:
  \[
  \min_x \|b - A \, x\|^2_2
  \]

- One method to solve the minimization problem is to solve the system of Normal Equations
  \[
  A^T A \, x = A^T \, b
  \]

- Let’s see some examples and discuss the limitations of this method.
Example:

\[
\begin{align*}
\text{Solve: } & A^T A \, x = A^T \, b \\
& x = \begin{bmatrix} 2.81441707 \\ 1.24048133 \end{bmatrix}
\end{align*}
\]
Data fitting - not always a line fit!

- Does not need to be a line! For example, here we are fitting the data using a quadratic curve.

**Linear Least Squares**: The problem is linear in its coefficients!
Another examples

We want to find the coefficients of the quadratic function that best fits the data points:

\[ y = x_0 + x_1 t + x_2 t^2 \]

We would not want our “fit” curve to pass through the data points exactly as we are looking to model the general trend and not capture the noise.
Data fitting

\[
\begin{bmatrix}
1 & t_1 & t_1^2 \\
\vdots & \vdots & \vdots \\
1 & t_m & t_m^2
\end{bmatrix}
\begin{bmatrix}
x_0 \\
x_1 \\
x_2
\end{bmatrix}
= 
\begin{bmatrix}
y_1 \\
\vdots \\
y_m
\end{bmatrix}
\]

Solve: \( A^T A \mathbf{x} = A^T \mathbf{b} \)
Which function is not suitable for linear least squares?

A) $y = a + b \cdot x + c \cdot x^2 + d \cdot x^3$
B) $y = x(a + b \cdot x + c \cdot x^2 + d \cdot x^3)$
C) $y = a \sin(x) + b / \cos(x)$
D) $y = a \sin(x) + x / \cos(bx)$
E) $y = a e^{-2x} + b e^{2x}$
Computational Cost

\[ A^T A x = A^T b \]

- Compute \( A^T A \): \( O(mn^2) \)
- Factorize \( A^T A \): LU \( \rightarrow O \left( \frac{2}{3}n^3 \right) \), Cholesky \( \rightarrow O \left( \frac{1}{3}n^3 \right) \)
- Solve \( O(n^2) \)
- Since \( m > n \) the overall cost is \( O(mn^2) \)
Short questions

Given the data in the table below, which of the plots shows the line of best fit in terms of least squares?

<table>
<thead>
<tr>
<th>x</th>
<th>1</th>
<th>2</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>y</td>
<td>2</td>
<td>18</td>
<td>12</td>
</tr>
</tbody>
</table>

A) ![Graph A](image)
B) ![Graph B](image)
C) ![Graph C](image)
D) ![Graph D](image)
Short questions

Given the data in the table below, and the least squares model

\[ y = c_1 + c_2 \sin(t\pi) + c_3 \sin(t\pi/2) + c_4 \sin(t\pi/4) \]

written in matrix form as

\[
A \begin{bmatrix}
  c_1 \\
  c_2 \\
  c_3 \\
  c_4 \\
\end{bmatrix} \cong y
\]

determine the entry \( A_{23} \) of the matrix \( A \).

Note that indices start with 1.

A) \(-1.0\)
B) \(1.0\)
C) \(-0.7\)
D) \(0.7\)
E) \(0.0\)
Solving Linear Least Squares with SVD
What we have learned so far...

\( A \) is a \( m \times n \) matrix where \( m > n \)
(more points to fit than coefficient to be determined)

Normal Equations: \( A^T A \mathbf{x} = A^T \mathbf{b} \)

• The solution \( A \mathbf{x} \cong \mathbf{b} \) is unique if and only if \( \text{rank}(A) = n \)
\( (A \) is full column rank)\)

• \( \text{rank}(A) = n \) \( \rightarrow \) columns of \( A \) are \textit{linearly independent} \( \rightarrow n \) non-zero
    singular values \( \rightarrow A^T A \) has only positive eigenvalues \( \rightarrow A^T A \) is a symmetric
    and positive definite matrix \( \rightarrow A^T A \) is invertible

\[
\mathbf{x} = (A^T A)^{-1} A^T \mathbf{b}
\]

• If \( \text{rank}(A) < n \), then \( A \) is rank-deficient, and solution of linear least squares
  problem is \textit{not unique}.
Finding the least square solution of \( A x \cong b \) (where \( A \) is full rank matrix) using the Normal Equations

\[
A^T A x = A^T b
\]

has some advantages, since we are solving a square system of linear equations with a symmetric matrix (and hence it is possible to use decompositions such as Cholesky Factorization)

However, the normal equations tend to worsen the conditioning of the matrix.

\[
\text{cond}(A^T A) = (\text{cond}(A))^2
\]

How can we solve the least square problem without squaring the condition of the matrix?
SVD to solve linear least squares problems

$A$ is a $m \times n$ rectangular matrix where $m > n$, and hence the SVD decomposition is given by:

$$A = 
\begin{pmatrix}
\vdots & \cdots & \vdots \\
\mathbf{u}_1 & \cdots & \mathbf{u}_m \\
\vdots & \cdots & \vdots
\end{pmatrix}
\begin{pmatrix}
\sigma_1 \\
\vdots \\
\sigma_n \\
0 \\
\vdots \\
0
\end{pmatrix}
\begin{pmatrix}
\cdots & \mathbf{v}_1^T & \cdots \\
\vdots & \cdots & \vdots \\
\cdots & \mathbf{v}_n^T & \cdots
\end{pmatrix}
$$

We want to find the least square solution of $A \mathbf{x} \cong \mathbf{b}$, where $A = U \Sigma V^T$

or better expressed in reduced form: $A = U_R \Sigma_R V^T$
Recall Reduced SVD  \( m > n \)

\[
A = U_R \Sigma_R V^T
\]

- **\( A \)**: \( m \times n \)
- **\( U_R \)**: \( m \times n \)
- **\( \Sigma_R \)**: \( n \times n \)
- **\( V^T \)**: \( n \times n \)
Shapes of the Reduced SVD

Suppose you compute a reduced SVD $A = U \Sigma V^T$ of a $10 \times 14$ matrix $A$. What will the shapes of $U$, $\Sigma$, and $V$ be?

**Hint:** Remember the transpose on $V$!

- The shape of $U$ will be $\times$.
- The shape of $\Sigma$ will be $\times$.
- The shape of $V$ will be $\times$. 
SVD to solve linear least squares problems

\[ A = U_R \Sigma_R V^T \]

\[ A = \begin{pmatrix} u_1 & \cdots & u_n \end{pmatrix} \begin{pmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_n \end{pmatrix} \begin{pmatrix} v_1^T \\ \vdots \\ v_n^T \end{pmatrix} \]

We want to find the least square solution of \( Ax \approx b \), where \( A = U_R \Sigma_R V^T \)

Normal equations: \( A^T A x = A^T b \rightarrow (U_R \Sigma_R V^T)^T (U_R \Sigma_R V^T) x = (U_R \Sigma_R V^T)^T b \)

\[ V \Sigma_R U_R^T (U_R \Sigma_R V^T) x = V \Sigma_R U_R^T b \]

\[ V \Sigma_R V^T x = V \Sigma_R U_R^T b \]

\[ (\Sigma_R)^2 V^T x = \Sigma_R U_R^T b \]

When can we take the inverse of the singular matrix?
\[(\Sigma_R)^2 V^T x = \Sigma_R U_R^T b\]

1) **Full rank matrix** \((\sigma_i \neq 0 \ \forall i)\):

\[
\text{rank}(A) = n
\]

\[
V^T x = (\Sigma_R)^{-1} U_R^T b
\]

Unique solution:

\[
x = V (\Sigma_R)^{-1} U_R^T b
\]

2) **Rank deficient matrix** \((\text{rank}(A) = r < n)\)

\[
(\Sigma_R)^2 V^T x = \Sigma_R U_R^T b
\]

Solution is not unique!!

Find solution \(x\) such that

\[
\min_x \phi(x) = \|b - A x\|_2^2
\]

and also

\[
\min_x \|x\|_2
\]
2) Rank deficient matrix (continue)

We want to find the solution $\mathbf{x}$ that satisfies $(\Sigma_R)^2 V^T \mathbf{x} = \Sigma_R U_R^T \mathbf{b}$ and also satisfies
\[
\min_{\mathbf{x}} \|\mathbf{x}\|_2
\]

Change of variables: Set $V^T \mathbf{x} = \mathbf{y}$ and then solve $\Sigma_R \mathbf{y} = U_R^T \mathbf{b}$ for the variable $\mathbf{y}$

\[
\begin{pmatrix}
\sigma_1 \\
\vdots \\
\sigma_r \\
0 \\
\vdots \\
0
\end{pmatrix}
\begin{pmatrix}
y_1 \\
\vdots \\
y_r \\
y_{r+1} \\
\vdots \\
y_n
\end{pmatrix}
= \begin{pmatrix}
\mathbf{u}_1^T \mathbf{b} \\
\vdots \\
\mathbf{u}_r^T \mathbf{b} \\
\vdots \\
\mathbf{u}_n^T \mathbf{b}
\end{pmatrix}
\]

\[
y_i = \frac{\mathbf{u}_i^T \mathbf{b}}{\sigma_i} \quad i = 1, 2, \ldots, r
\]

What do we do when $i > r$?
Which choice of $y_i$ will minimize
\[
\|\mathbf{x}\|_2 = \|V \mathbf{y}\|_2?
\]

Set $y_i = 0, \quad i = r + 1, \ldots, n$

Evaluate
\[
\mathbf{x} = V \mathbf{y} = \begin{pmatrix}
\vdots & \cdots & \vdots \\
\mathbf{v}_1 & \cdots & \mathbf{v}_n
\end{pmatrix}
\begin{pmatrix}
y_1 \\
\vdots \\
y_n
\end{pmatrix}
\]

\[
\mathbf{x} = \sum_{i=1}^{n} y_i \mathbf{v}_i = \sum_{\sigma_i \neq 0}^{n} \frac{\mathbf{u}_i^T \mathbf{b}}{\sigma_i} \mathbf{v}_i
\]
Solving Least Squares Problem with SVD (summary)

Find $\mathbf{x}$ that satisfies $\min_{\mathbf{x}} \| \mathbf{b} - \mathbf{A} \mathbf{x} \|^2_2$

Find $\mathbf{y}$ that satisfies $\min_{\mathbf{y}} \| \mathbf{\Sigma} \mathbf{y} - \mathbf{U}_R^T \mathbf{b} \|^2_2$

Propose $\mathbf{y}$ that is solution of $\mathbf{\Sigma} \mathbf{y} = \mathbf{U}_R^T \mathbf{b}$

Evaluate: $\mathbf{z} = \mathbf{U}_R^T \mathbf{b}$

Set: $y_i = \begin{cases} \frac{z_i}{\sigma_i}, & \text{if } \sigma_i \neq 0 \\ 0, & \text{otherwise} \end{cases}$ $i = 1, \ldots, n$

Then compute $\mathbf{x} = \mathbf{V} \mathbf{y}$

Cost of SVD: $O(m \, n^2)$

Cost: $m \, n$

$n$

$n^2$
Solving Least Squares Problem with SVD (summary)

• If $\sigma_i \neq 0$ for $\forall i = 1, ..., n$, then the solution $y = V (\Sigma_R)^{-1} U_R^T b$ is unique (and not a “choice”).

• If at least one of the singular values is zero, then the proposed solution $y$ is the one with the smallest 2-norm ($\|y\|_2$ is minimal) that minimizes the 2-norm of the residual $\| \Sigma_R y - U_R^T b \|_2$

• Since $\|x\|_2 = \|V y\|_2 = \|y\|_2$, then the solution $x$ is also the one with the smallest 2-norm ($\|x\|_2$ is minimal) for all possible $x$ for which $\|Ax - b\|_2$ is minimal.
Solving Least Squares Problem with SVD (summary)

Solve $Ax \approx b$ or $U_R \Sigma_R V^T x \approx b$

$x \approx V (\Sigma_R)^+ U_R^T b$
Example:

Consider solving the least squares problem $A \mathbf{x} \cong \mathbf{b}$, where the singular value decomposition of the matrix $A = U \Sigma V^T \mathbf{x}$ is:

$$
\begin{bmatrix}
\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 & 0 \\
\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 & 0 \\
0 & 0 & 0 & 1 \\
0 & 0 & 1 & 0
\end{bmatrix}
\begin{bmatrix}
14 & 0 & 0 \\
0 & 14 & 0 \\
0 & 0 & 0 \\
0 & 0 & 0
\end{bmatrix}
\begin{bmatrix}
1 & 0 & 0 \\
0 & 1 & 0 \\
0 & 0 & 1
\end{bmatrix}
\mathbf{x} \cong
\begin{bmatrix}
12 \\
9 \\
9 \\
10
\end{bmatrix}
$$

Determine $\| \mathbf{b} - A \mathbf{x} \|_2$
Example

Suppose you have $A = U \Sigma V^T x$ calculated. What is the cost of solving

$$\min_x ||b - Ax||_2^2$$

A) $O(n)$
B) $O(n^2)$
C) $O(mn)$
D) $O(m)$
E) $O(m^2)$