Lecture Least Squares Approximation

Motivation: Hooke’s Law and Data Fitting

In physics, **Hooke’s Law** tells us that (within certain limits), the force y exerted by a spring is linearly related to its length x:

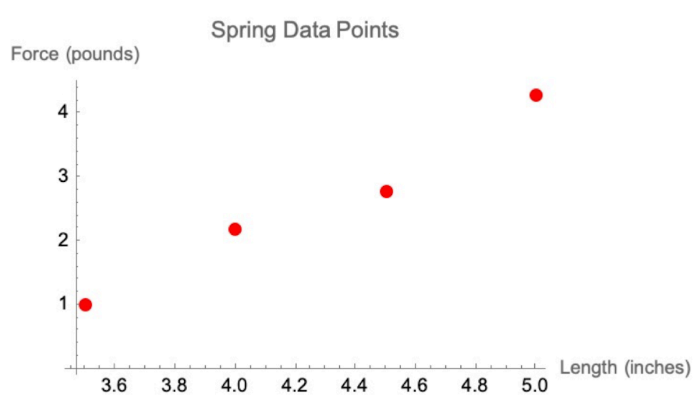
$$y = cx + d$$

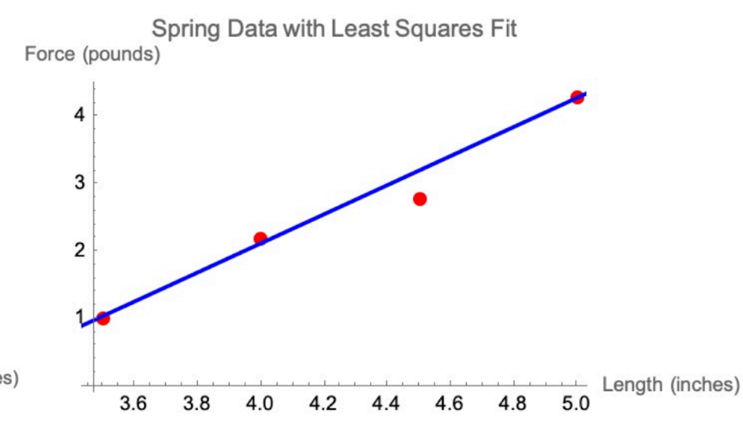
where:

- y = force (in pounds),
- x = spring length (in inches),
- c = **spring constant**,

Suppose we collect some data from an experiment:

Length x (inches)	Force y (pounds)
3.5	1.0
4.0	2.2
4.5	2.8
5.0	4.3





We can see the data *approximately* follows a line — but not perfectly. Measurement errors or other factors may cause deviations.

Question:

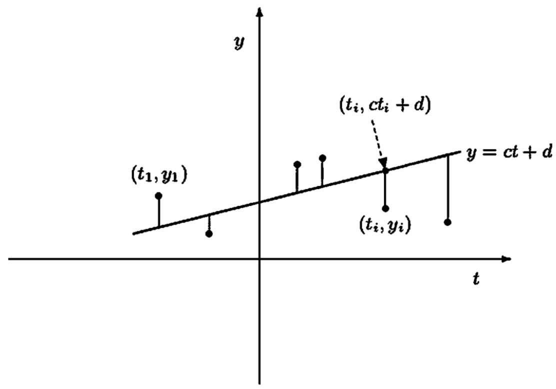
How do we find the **best linear fit** of the form $y = cx + d$ that describes the relationship between x and y?

General Linear Least Squares Problem

Suppose an experimenter collects data points $(t_1, y_1), (t_2, y_2), \ldots, (t_m, y_m)$ by measuring a quantity y at different times t. The data is plotted as points in the t-y plane.

If we suspect that y and t are related **approximately linearly**, we seek a function of the form: $y = ct + d$ to best fit the data.

Goal: Find constants c and d such that the line $y = ct + d$ **minimizes the total squared vertical distance** from the data points to the line.



This total squared error is defined as:

$$\hat{y}_i = ct_i + d$$

$$E = \sum_{i=1}^{m} (y_i - (ct_i + d))^2$$

This is called the **least squares error**.

Matrix Formulation

We write the system as:

$$A = \begin{bmatrix} t_1 & 1 \\ t_2 & 1 \\ \vdots & \vdots \\ t_m & 1 \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} c \\ d \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix}$$

$$\vec{y} - A\vec{x} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix} - \begin{bmatrix} ct_1 + d \\ ct_2 + d \\ \vdots \\ ct_m + d \end{bmatrix} = \begin{bmatrix} y_1 - (ct_1 + d) \\ \vdots \\ y_m - (ct_m + d) \end{bmatrix}$$

Then the model becomes:

$$A\mathbf{x} \approx \mathbf{y}$$

The **least squares solution** minimizes the error:

$$E = \|\mathbf{y} - A\mathbf{x}\|^2$$

For the rest of this lecture, we will develop the general method for finding a vector $\mathbf{x}_0 \in \mathrm{F}^n$ that minimizes the error:

$$\|\mathbf{y} - A\mathbf{x}_0\| \leq \|\mathbf{y} - A\mathbf{x}\| \quad \text{for all } \mathbf{x} \in \mathrm{F}^n$$

where A is an $m \times n$ matrix and $\mathbf{y} \in \mathrm{F}^m$ is the data vector.

This method finds the **best linear fit** to the data and can be extended to fit **polynomials of degree ≤ k**.

For $\mathbf{x}, \mathbf{y} \in \mathrm{F}^n$, recall the standard inner product: $\langle \mathbf{x}, \mathbf{y} \rangle_n = \mathbf{y}^* \mathbf{x}$ with $\mathbf{x}$ and $\mathbf{y}$ treated as column vectors. $= \sum_{i=1}^{n} \bar{y}_i x_i$

Lemma 1.
Let $A \in M_{m\times n}(\mathrm{F})$, $\mathbf{x} \in \mathrm{F}^n$, and $\mathbf{y} \in \mathrm{F}^m$. Then

$$\langle A\mathbf{x}, \mathbf{y} \rangle_m = \langle \mathbf{x}, A^*\mathbf{y} \rangle_n.$$

$A = A^{**}$ by Corallary (d) of Thm 6.11

Proof: $\langle Ax, y \rangle_m \overset{\text{by def}}{=} y^*(Ax) \underset{\text{by associativity of multiplying matrices}}{=} (y^*A)x = y^*A^{**}x \underset{(AB)^* = B^*A^*}{=} (A^*y)^*x \underset{\text{by def.}}{=} \langle x, A^*y \rangle_n$

Lemma 2.
Let $A \in M_{m\times n}(\mathrm{F})$. Then

$$\operatorname{rank}(A^*A) = \operatorname{rank}(A).$$

Proof: Let $L_A : V^n \to W^m$ be the linear map defined by A.

where $\dim(V) = n$ and $\dim(W) = m$. $\mathrm{Ker}(A) = \{\vec{x} \in V \mid A\vec{x} = \vec{0}\}$.

By the rank-nullity theorem, $\dim(\mathrm{Ker}(A))$.

$n = \dim(V) = \operatorname{rank}(A) + \operatorname{nullity}(A)$

$= \operatorname{rank}(A^*A) + \operatorname{nullity}(A^*A)$

Thus we need to show $\operatorname{nullity}(A) = \operatorname{nullity}(A^*A)$.

Then it suffices to show

$\ker(A) = \ker(A^*A)$.

$(\{\vec{x} \in V \mid A\vec{x} = \vec{0}\} = \{x \in V \mid A^*Ax = 0\})$

First, if $A\vec{x} = \vec{0}$, then clearly $A^*A\vec{x} = \vec{0}$ $(\ker A \subseteq \ker(A^*A))$

Conversely, suppose $A^*A\vec{x} = \vec{0}$,

$$0 = \langle \underbrace{A^*A\vec{x}}_{=\vec{0}}, \vec{x} \rangle_n = \langle A\vec{x}, A\vec{x} \rangle_m = \|A\vec{x}\|^2$$

by Lemma 1, and $A^{**} = A$

$\Rightarrow A\vec{x} = \vec{0}$. $(\ker(A^*A) \subseteq \ker(A))$.

Thus the result follows.

As a consequence of **Lemma 2.**, we have

Corollary.
If A is an $m \times n$ matrix with $\operatorname{rank}(A) = n$, then A^*A is invertible.

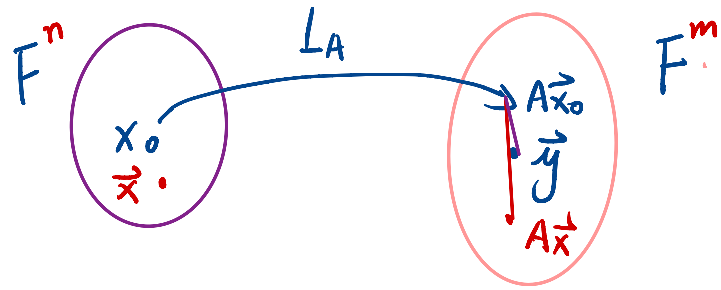
Theorem 6.12.

Let $A \in M_{m\times n}(\mathrm{F})$ and $\mathbf{y} \in \mathrm{F}^m$. Then there exists $\mathbf{x}_0 \in \mathrm{F}^n$ such that

$$(A^*A)\mathbf{x}_0 = A^*\mathbf{y} \quad \text{and} \quad \|A\mathbf{x}_0 - \mathbf{y}\| \le \|A\mathbf{x} - \mathbf{y}\| \text{ for all } \mathbf{x} \in \mathrm{F}^n.$$

Furthermore, if $\operatorname{rank}(A) = n$, then

$$\mathbf{x}_0 = (A^*A)^{-1}A^*\mathbf{y}.$$



subspace of F^m.

Proof: Define $W = \{A\vec{x} : \vec{x} \in F^n\}$, i.e. $W = \text{Range}(L_A)$

By Corollary to Thm 6.5 in Lecture 5, there exists a unique vector in W such that is closest to $\vec{y}$.

Denote this vector $A\vec{x}_0$, where $\vec{x}_0 \in F^n$

Then $\quad \|A\vec{x}_0 - \vec{y}\| \le \|A\vec{x} - \vec{y}\|$

So $\vec{x}_0$ has the property that $\|A\vec{x}_0 - \vec{y}\| \in F$ is minimal as desired.

To find such $\vec{x}_0 \in V$ we know from Thm 6.6 and its Corollary that $A\vec{x}_0 - \vec{y} \in W^\perp$.

Thus $\quad \langle \underbrace{A\vec{x}}_{\in W}, A\vec{x}_0 - \vec{y} \rangle_m = 0$, for all $x \in F^n$.

Thus by lemma 1,

$$0 = \langle A\vec{x}, A\vec{x}_0 - \vec{y} \rangle_m = \langle \vec{x}, A^*(A\vec{x}_0 - \vec{y}) \rangle_n = 0.$$

for all $x \in F^n$.

This means $A^*(A\vec{x}_0 - \vec{y}) = \vec{0}$

So we need to find $\vec{x}_0$ s.t.

$$A^*A\vec{x}_0 = A^*\vec{y}$$

If, in addition, we assume $\operatorname{rank}(A) = n$, then

by Lemma 2. $\operatorname{rank}(\underbrace{A^*}_{n\times m}\underbrace{A}_{m\times n}) = n$. ($(A^*A)^{-1}$ exists)

Then we have. $\vec{x}_0 = (A^*A)^{-1}A^*\vec{y}$

Now we return to the **Hooke's Law and Data Fitting** problem:

We rewrite the model as:

$$y_i = cx_i + d \quad \text{for each } i$$

coeff of x from the data.

This is a linear model in c and d. Let's write this in matrix form:

$$A = \begin{bmatrix} 3.5 & 1 \\ 4.0 & 1 \\ 4.5 & 1 \\ 5.0 & 1 \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} c \\ d \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} 1.0 \\ 2.2 \\ 2.8 \\ 4.3 \end{bmatrix},$$

ANS: We want to solve the normal equation:

$$A^*A\vec{x} = A^*\vec{y}$$

by applying Thm 6.12.

Step 1. Compute A^*A and $A^*\vec{y}$

$$A^*A = \begin{bmatrix} 3.5 & 4.0 & 4.5 & 5.0 \\ 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 3.5 & 1 \\ 4.0 & 1 \\ 4.5 & 1 \\ 5.0 & 1 \end{bmatrix} = \begin{bmatrix} 73.5 & 17.0 \\ 17.0 & 4 \end{bmatrix}$$

$$A^*\vec{y} = \begin{bmatrix} 3.5 & 4.0 & 4.5 & 5.0 \\ 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1.0 \\ 2.2 \\ 2.8 \\ 4.3 \end{bmatrix} = \begin{bmatrix} 46.4 \\ 10.3 \end{bmatrix}$$

Step 2. Solve the system.

Solve $\begin{bmatrix} 73.5 & 17.0 \\ 17.0 & 4 \end{bmatrix} \begin{bmatrix} c \\ d \end{bmatrix} = \begin{bmatrix} 46.4 \\ 10.3 \end{bmatrix}$

Recall $\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{(ad-bc)} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$

Thus $\begin{bmatrix} 73.5 & 17.0 \\ 17.0 & 4 \end{bmatrix}^{-1} = \begin{bmatrix} 0.8 & -3.4 \\ -3.4 & 14.7 \end{bmatrix}$

So $\vec{x}_0 = \begin{bmatrix} c \\ d \end{bmatrix} = (A^*A)^{-1} A^* \vec{y} = \begin{bmatrix} 0.8 & -3.4 \\ -3.4 & 14.7 \end{bmatrix} \begin{bmatrix} 46.4 \\ 10.3 \end{bmatrix} = \begin{bmatrix} 2.1 \\ -6.35 \end{bmatrix}$

Thus $y = 2.1x - 6.35$.