# Basic Maths for Non-mathematicians

**Peleg Bar Sapir**

$$\int_a^b f(x)\,\mathrm{d}x = \lim_{\Delta x \to 0} \sum_{k=1}^{N} f(x_k)\Delta x$$

$$(AB)^\top = B^\top A^\top$$

$$\mathbb{R}^n \xrightarrow{T} \mathbb{R}^m$$

$$\vec{v} = \sum_{i=1}^{n} \alpha_i \hat{e}_i$$

$$\frac{\mathrm{d}f}{\mathrm{d}x} = \lim_{\Delta x \to 0} \frac{f(x+\Delta x)-f(x)}{\Delta x}$$

$$A = Q\Lambda Q^{-1}$$

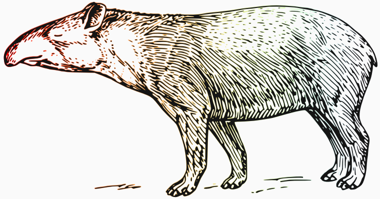
$$A\vec{v} = \lambda\vec{v}$$

$$\mathrm{Rot}(\theta) = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix}$$

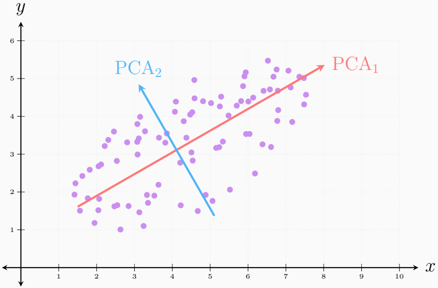
$$\int_a^b f(x)\,\mathrm{d}x = F(b) - F(a)$$

$$\langle \hat{e}_i, \hat{e}_j \rangle = \delta_{ij}$$

$$T(\alpha\vec{u} + \beta\vec{v}) = \alpha T(\vec{u}) + \beta T(\vec{v})$$

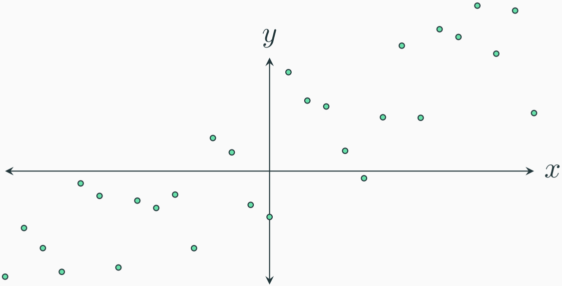


# Chapter 7: Some Real-World Uses of Linear Algebra



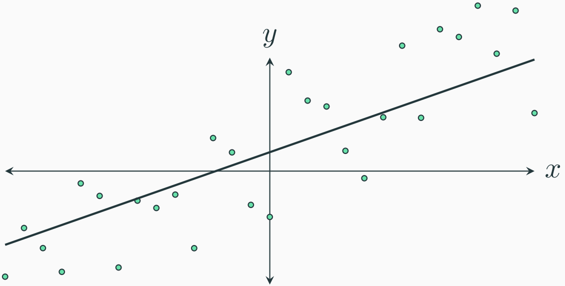
# Least Squares Approximation

What is the best linear approximation to a set of measurements?



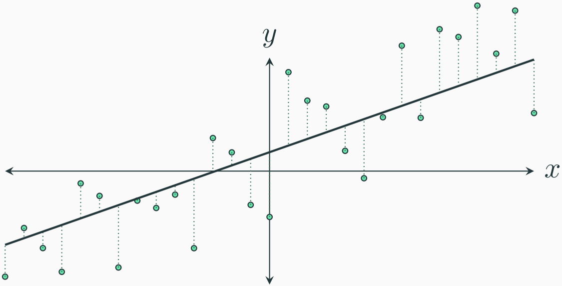
# Least Squares Approximation

What is the best linear approximation to a set of measurements?



# Least Squares Approximation

What is the best linear approximation to a set of measurements?




A good approximation is the line $f(x) = ax + b$ for which the sum of the distances from the line to each point $(x_i, y_i)$ is minimal, i.e.

$$S = \min\left(\sum_{i=1}^{n}\left[f\left(x_i\right) - y_i\right]\right).$$

# Least Squares Approximation

We can collect all the $y$ values of our measurement points to a vector:

$$\vec{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix},$$

and similarly collect all the $y = f(x)$ values of the line:

$$\vec{f} = \begin{pmatrix} f(x_1) \\ f(x_2) \\ \vdots \\ f(x_n) \end{pmatrix}.$$

# Least Squares Approximation

The sum $s = \sum_{i=1}^{n} \left[f\left(x_i\right) - y_i\right]$ then becomes:

$$s = \sum_{i=1}^{n} \left[f\left(x_i\right) - y_i\right]$$
$$= \sum_{i=1}^{n} \left[\vec{f_i} - \vec{y_i}\right].$$

However, $s$ is a bit problematic, as some elements $\vec{f_i} - \vec{y_i}$ can be negative. Instead, we can minimize the following expression:

$$s^* = \sum_{i=1}^{n} \left[\vec{f_i} - \vec{y_i}\right]^2.$$

# Least Squares Approximation

...and the expression

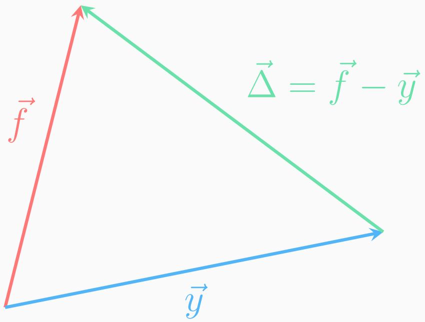
$$s^* = \sum_{i=1}^{n} \left[\vec{f_i} - \vec{y_i}\right]^2$$

is exactly the square norm of the vector

$$\vec{\Delta} = \begin{pmatrix} f_1 - y_1 \\ f_2 - y_2 \\ \vdots \\ f_n - y_n \end{pmatrix} = \vec{f} - \vec{y}.$$

## Least Squares Approximation

Drawing the a 2-dimentional scheme of the vectors $\vec{v}, \vec{f}$ and their difference $\vec{\Delta} = \vec{f} - \vec{v}$:



## Least Squares Approximation

The norm of the vector $\vec{\Delta} = \vec{f} - \vec{y}$ is minimal when $\vec{f} \perp \vec{\Delta}$, i.e. when

$$\vec{f} \cdot \vec{\Delta} = \vec{f} \cdot \left(\vec{f} - \vec{y}\right) = 0.$$

Let's find what condition on $\vec{f}$ yields this.

# Least Squares Approximation

First, we note that the vector $\vec{f}$ can be written as a matrix-vector product:

$$\vec{f} = \begin{pmatrix} f(x_1) \\ f(x_2) \\ \vdots \\ f(x_n) \end{pmatrix} = \begin{pmatrix} ax_1 + b \\ ax_2 + b \\ \vdots \\ ax_n + b \end{pmatrix} = \begin{pmatrix} x_1 & 1 \\ x_2 & 1 \\ \vdots & \\ x_n & 1 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix}.$$

Thus, the condition $\vec{f} \cdot \left( \vec{f} - \vec{y} \right) = 0$ becomes

$v$

$$A\vec{v} \cdot (A\vec{v} - \vec{y}) = 0.$$

# Least Squares Approximation

Some algebra:

$$A\vec{v} \cdot (A\vec{v} - \vec{y}) = 0.$$

# Least Squares Approximation

Some algebra:

$$A\vec{v} \cdot A\vec{v} - A\vec{v} \cdot \vec{y} = 0.$$

# Least Squares Approximation

Some algebra:

$$A\vec{v} \cdot A\vec{v} = A\vec{v} \cdot \vec{y}.$$

# Least Squares Approximation

Some algebra:

$$A\vec{v} \cdot A\vec{v} = A\vec{v} \cdot \vec{y}.$$

Since $A\vec{v}$ is a vector, it can be dotted with either itself or $\vec{y}$. However, we can consider $A\vec{v}$ as an $n \times 1$ matrix, and to keep the product defined we transpose it, i.e.

$$(A\vec{v})^\top \cdot A\vec{v} = (A\vec{v})^\top \cdot \vec{y}.$$

This doesn't change the truthness of the equation.

# Least Squares Approximation

Expanding the transposed product $(A\vec{v})^\top$ yields

$$\vec{v}^\top A^\top A\vec{v} = \vec{v}^\top A^\top \vec{y},$$

where $\vec{v}^\top$ is a row vector.

We can remove $\vec{v}^\top$ from both sides, leaving us with

$$A^\top A\vec{v} = A^\top \vec{y}.$$

This linear system is surprisingly easy to solve!

# Least Squares Approximation

## Example

Let's look at 6 points:

$$p_1 = (-2, -7.3)$$
$$p_2 = (-1, -3.9)$$
$$p_3 = (0, -1.2)$$
$$p_4 = (1, 2.4)$$
$$p_5 = (2, 4.7)$$
$$p_6 = (3, 7.7)$$

### Example

The linear system we need to solve is thus

$$\begin{pmatrix} -2 & -1 & 0 & 1 & 2 & 3 \\ 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} -2 & 1 \\ -1 & 1 \\ 0 & 1 \\ 1 & 1 \\ 2 & 1 \\ 3 & 1 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} -2 & -1 & 0 & 1 & 2 & 3 \\ 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} -7.3 \\ -3.9 \\ -1.2 \\ 2.4 \\ 4.7 \\ 7.7 \end{pmatrix}.$$

Multiplying both matrix-matrix products yields

$$\begin{pmatrix} 19 & 3 \\ 3 & 6 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} 53.4 & 2.4 \end{pmatrix},$$

which when solved for $a$ and $b$ yields

$$a = 2.98 \quad b = -1.09.$$

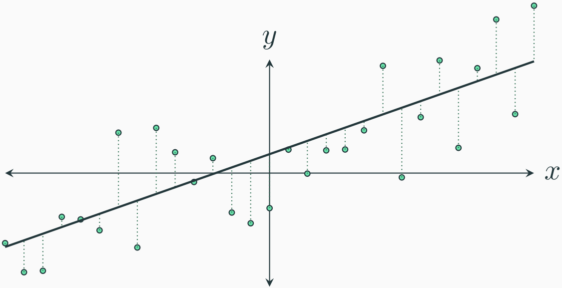
## Least Squares Approximation

How can we quantify the "goodness" of fit between the proposed approximation and out data points?

# Least Squares Approximation

We can first look at the average difference between $y_i$ and the linear approximation (the **variance** in the $y$-values in respect to the line):

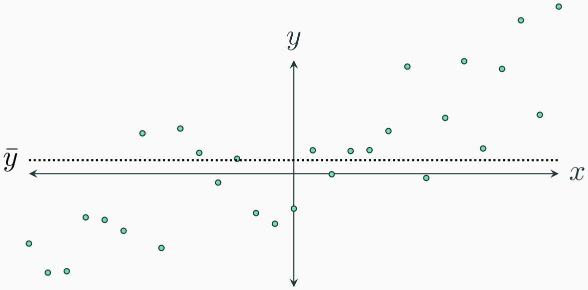
$$\sigma_{\text{line}} = \frac{1}{n} \sum_{i=1}^{n} \left[ f\left(x_i\right) - y_i \right]^2 .e$$



# Least Squares Approximation

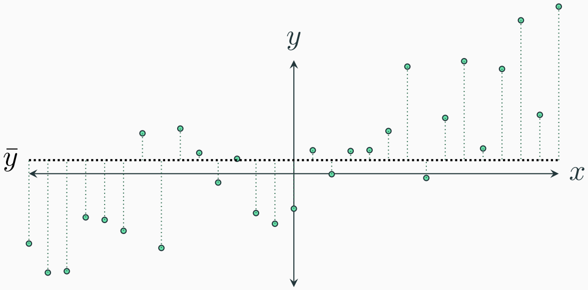
Then we look at the average $y$ value of our data points:

$$\bar{y} = \frac{1}{n} \sum_{i=1}^{n} y_i.e$$



We can calculate the total distance of out data points to $\bar{y}$:

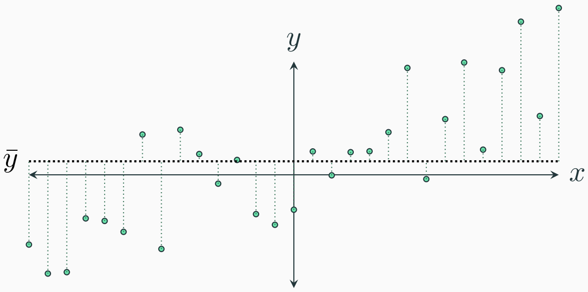
$$\mathsf{SE}_{\bar{y}} = (y_1 - \bar{y})^2 + (y_2 - \bar{y})^2 + \cdots + (y_n - \bar{y})^2 = \sum_{i=1}^{n} (y_i - \bar{y})^2 .e$$



## Least Squares Approximation

The average of $\mathsf{SE}_{\bar{y}}$ is the variance in the $y$-values:

$$\sigma_y = \frac{1}{n}\mathsf{SE}_{\bar{y}} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2 .e$$



# Least Squares Approximation

The ratio of the two variances

$$\rho = \frac{\sigma_{\text{line}}}{\sigma_{\bar{y}}}$$

is a measurement of what percentage of the total variation is **NOT** described by the linear approximaion. It is in the range

$$0 \leq \rho \leq 1.$$

Thus,

$$r^2 \equiv 1 - \rho = 1 - \frac{\sigma_{\text{line}}}{\sigma_{\bar{y}}}$$

describes how much of the total variation is described by the linear approximation.

# Least Squares Approximation

An $r^2$ close to $1$ means that $\rho$ is close to $0$, i.e. the variation of $y_i$ from the line, $\sigma_{\text{line}}$, is small compared to the total variance of the points.

### Example

The average $y$ value of the points in the previous example is

$$\bar{y} = \frac{1}{6}\left(-7.3 - 3.9 - 1.2 + 2.4 + 4.7 + 7.7\right) = \frac{2.4}{6} = 0.4.$$

Their total variance is thus

$$\begin{aligned}
\sigma_{\bar{y}} = \frac{1}{6}\Big[&(-7.3-0.4)^2 + (-3.9-0.4)^2 + (-1.2-0.4)^2 \\
&+ (2.4-0.4)^2 + (4.7-0.4)^2 + (7.7-0.4)^2\Big] \\
&= \frac{1}{6}\left[59.29 + 18.49 + 2.56 + 4 + 18.49 + 53.29\right] \\
&= 26.02.
\end{aligned}$$

# Least Squares Approximation

## Example

The linear approximation was calculated as $f(x) = 2.98x - 1.09$, and so the variance to the linear approximation is

$$\begin{aligned}\sigma_{\text{line}} &= \frac{1}{6}\Big[(-7.05+7.3)^2 + (-4.07+3.9)^2 + (-1.09+1.2)^2 \\ &\quad +(1.89-2.4)^2 + (4.87-4.7)^2 + (7.85-7.7)^2\Big] \\ &= \frac{1}{6}\left[0.06+0.03+0.01+0.26+0.03+0.02\right] \\ &= 0.0692.\end{aligned}$$

# Least Squares Approximation

**Example**

Thus,

$$r^2 = 1 - \frac{\sigma_{\text{line}}}{\sigma_{\bar{y}}} = 1 - \frac{0.0692}{26.02} = 1 - 0.0027 = 0.9973,$$

which means that the linear approximation given by the least squares method for this set of points is an exceptionally good approximation.