By taking measurements, we find points $\begin{bmatrix} x_1 \\ y_1 \end{bmatrix}, \dots, \begin{bmatrix} x_N \\ y_N \end{bmatrix}$ and scalars $c_1, \dots, c_N \in \mathbb{R}$ with $f\begin{bmatrix} x_i \\ y_i \end{bmatrix} = ax_i + by_i \approx c_i$ for $i = 1, \dots, N$.

some measurement error

Due to measurement errors, there might not be **any** pair a, b with $ax_i + by_i = c_i$ for all $i = 1, \dots, N$; the system is usually **overdetermined**.

Q: What can we do?

Soln: If we can't find an exact solution, try to find a "good" approximate solution.

Defn Given an $m \times n$ matrix A and a vector $b \in \mathbb{R}^m$, the associated **linear least-squares problem** is to minimise the approximation error $\|Ax - b\|$ to an exact solution "$Ax = b$" among all vectors $x \in \mathbb{R}^n$.

More formally: A **least-squares solution** of "$Ax = b$" is any

$$\hat{x} \in \mathbb{R}^n \text{ with } \|A\hat{x} - b\| \leq \|Ax - b\| \text{ for all } x \in \mathbb{R}^n.$$

Q: ① Is there always a choice of $\hat{x}$ which minimizes $\|A\hat{x} - b\|$?

② Why is this called "least-squares"?

The second question is easy: minimizing $\|A\hat{x}-b\|$ minimizes $y_1^2+\dots+y_m^2$,

where $y = A\hat{x}-b \in \mathbb{R}^m$.

Geometric interpretation

We wish to find $\hat{x} \in \mathbb{R}^n$ s.t. $A\hat{x}$ is the closest point to b within all of Col A. By the Best Approximation Thm, this just means that

$$\boxed{A\hat{x} = \text{orthogonal projection of } b \text{ onto Col} A.}$$

It is clear that such an $\hat{x}$ exists (Why?)... but how can we find it?

Write $A = [a_1 \dots a_n]$ for $a_1, \dots, a_n \in \mathbb{R}^m$.

Then: $\hat{x}$ is a least-squares solution of "$Ax=b$"

$\iff$ $A\hat{x}$ is the orthogonal projection of b onto Col A

$\iff$ $A\hat{x}-b$ is orthogonal to each vector in Col A

$\iff$ $a_j \cdot (A\hat{x}-b) = 0$ for $j = 1, \dots, n$

$\iff$ $a_j^T (A\hat{x}-b) = 0$ for $j = 1, \dots, n$

$\iff$ $A^T(A\hat{x}-b) = 0 \iff \boxed{A^TA\hat{x} = A^Tb}$

This is called the normal equation for "$Ax=b$"

We have established the following:

Thm ① Given A, b, there always exists a least-squares solution of "$Ax = b$".

② The least-squares solutions of "$Ax = b$" are precisely the exact solutions of the normal equation $\boxed{A^T A \hat{x} = A^T b.}$

③ The latter can (e.g.) be found using Gaussian elimination.

Ex: let $A = \begin{bmatrix} 4 & 0 \\ 0 & 2 \\ 1 & 1 \end{bmatrix}$ and $b = \begin{bmatrix} 2 \\ 0 \\ 11 \end{bmatrix}$. Easy: We can't solve "$Ax = b$" exactly.

$$A^T A = \begin{bmatrix} 4 & 0 & 1 \\ 0 & 2 & 1 \end{bmatrix} \begin{bmatrix} 4 & 0 \\ 0 & 2 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 17 & 1 \\ 1 & 5 \end{bmatrix}, \quad A^T b = \begin{bmatrix} 4 & 0 & 1 \\ 0 & 2 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ 0 \\ 11 \end{bmatrix} = \begin{bmatrix} 19 \\ 11 \end{bmatrix}.$$

So we need to solve $\begin{bmatrix} 17 & 1 \\ 1 & 5 \end{bmatrix} \hat{x} = \begin{bmatrix} 19 \\ 11 \end{bmatrix}$.

Let's do that: $\left[\begin{array}{cc|c} 17 & 1 & 19 \\ 1 & 5 & 11 \end{array}\right] \rightsquigarrow \left[\begin{array}{cc|c} 1 & 5 & 11 \\ 0 & -84 & -168 \end{array}\right] \rightsquigarrow \left[\begin{array}{cc|c} 1 & 5 & 11 \\ 0 & 1 & 2 \end{array}\right]$

$\rightsquigarrow \left[\begin{array}{cc|c} 1 & 0 & 1 \\ 0 & 1 & 2 \end{array}\right]$ so $\hat{x} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$ is the (unique!) least-squares solution.

Approximation error: $\|A\hat{x} - b\| = \ldots = 2\sqrt{21}$