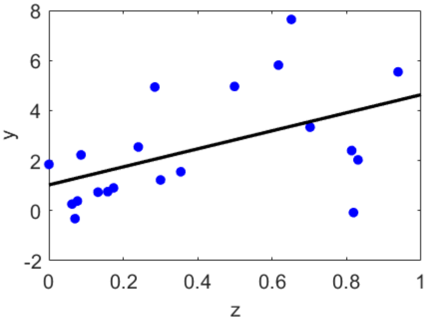
2. Linear Least Squares

- Least squares for data fitting
- Linear systems
- Least squares properties
- QR factorization

Least squares for data fitting

Fitting a line to data

Fit a line to observations y_i given input z_i, $i = 1, \ldots, n$



$$\begin{array}{ll} \underset{s,c}{\text{minimize}} & \frac{1}{2} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \\ \text{subject to} & s z_i + c = \hat{y}_i \end{array}$$

Fitting a line to data

$$\underset{s,c}{\text{minimize}} \quad \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \qquad \text{subject to} \quad sz_i + c = \hat{y}_i$$

Reframe as **least squares problem**

$$\underset{x}{\text{minimize}} \quad \|Ax - b\|_2^2 = \sum_{i=1}^{m} (a_i^T x - b_i)^2$$

where

$$A = \begin{bmatrix} z_1 & 1 \\ z_2 & 1 \\ \vdots & \vdots \\ z_m & 1 \end{bmatrix}, \quad b = \begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \vdots \\ \hat{y}_m \end{bmatrix}$$

Solution $x \in \mathbf{R}^2$ contains slope s and intercept c:

$$x = \begin{bmatrix} s \\ c \end{bmatrix}$$

Example: Polynomial data fitting

Given m distinct points t_i (eg, measurement times) and values (eg, measurement values):

$$(t_1, y_1),\ (t_2, y_2), \ldots, (t_m, y_m)$$

Goal: fit a polynomial of degree n to the m points:

$$p(t) = x_0 + x_1 t + x_2 t^2 + \ldots + x_n t^n \quad (x_i = \text{coeff's})$$

If $n = m - 1$, then can fit perfectly:

$$\begin{aligned} p(t_1) &= y_1 \\ p(t_2) &= y_2 \\ &\vdots \\ p(t_m) &= y_m \end{aligned} \qquad \text{or} \qquad \begin{bmatrix} 1 & t_1 & t_1^2 \cdots t_1^{n} \\ 1 & t_2 & t_2^2 \cdots t_2^{n} \\ \vdots & & \\ 1 & t_m & t_m^2 \cdots t_m^{n} \end{bmatrix} \begin{bmatrix} x_0 \\ x_1 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix}$$

Often better to approximate with a lower-order polynomial, eg, $n \ll m$. Must then settle for an approximate fit:

$$
\begin{aligned}
p(t_1) &= y_1 \\
p(t_2) &= y_2 \\
&\vdots \\
p(t_m) &= y_m
\end{aligned}
\qquad \text{or} \qquad
\begin{bmatrix} 1 & t_1 & t_1^2 \cdots t_1^{n} \\ 1 & t_2 & t_2^2 \cdots t_2^{n} \\ \vdots & & \\ 1 & t_m & t_m^2 \cdots t_m^{n} \end{bmatrix}
\begin{bmatrix} x_0 \\ x_1 \\ \vdots \\ x_n \end{bmatrix}
\approx
\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix}
$$

↗ Vandermonde matrix.

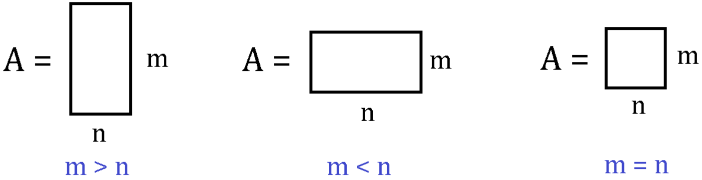
There are fewer unknowns than equations, and so we fit in the least-squares sense:

$$\underset{x_0,\ldots,x_{n-1}}{\text{minimize}} \quad \tfrac{1}{2}\sum_{i=1}^{m}[p(t_i) - y)i]^2$$

Linear systems

Solving linear systems

Find x where $Ax \approx b$.



- $m > n$ overdetermined (possibly no solution)
- $m < n$ underdetermined (possibly infinite solutions)
- $m = n$ might be invertible (possibly 1 solution)

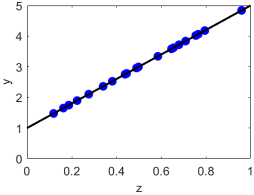
Least-squares problem finds "best" (in 2-norm sense) solution to $Ax \approx b$:

- residual vector $r = Ax - b$
- find x^* such that $\|Ax^* - b\|_2 \leq \|Ax - b\|_2$ for all x
  ie, $\|r^*\| \leq \|r\|$

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

Solving linear systems

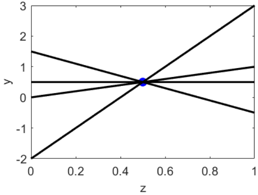
Find x where $Ax = b$.



1 solution

overdetermined

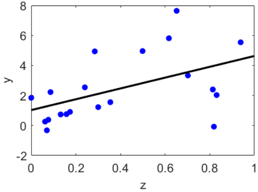
$b \in \mathbf{range}(A)$



many solutions

underdetermined

$b \in \mathbf{range}(A)$



no solution

overdetermined

$b \notin \mathbf{range}(A)$

Least squares properties

Back to least squares

$$x^* = \underset{x}{\operatorname{argmin}} \|Ax - b\|_2^2$$

- Normal equations

$$A^T A x^* = A^T b$$

- Residual

$$r^* = Ax^* - b, \qquad A^T r^* = 0$$

- x^* satisfying normal equations is minimizer, may not be unique
- $y^* = Ax^*$ is unique

  (Why? Hint: reformulate as quadratic over y.)

- In MATLAB or Julia

```
x = A \ b
```

Geometry

Recall

- $\mathbf{range}(A) = \{y : y = Ax \quad \text{for some} \quad x\}$
- $\mathbf{null}(A^T) = \{z : A^T z = 0\}$

$= \{x + y \mid x \in R(A), y \in N(A^T)\}$

Orthogonal complement:

$$\mathbf{range}(A) \oplus \mathbf{null}(A^T) = \mathbf{R}^m.$$

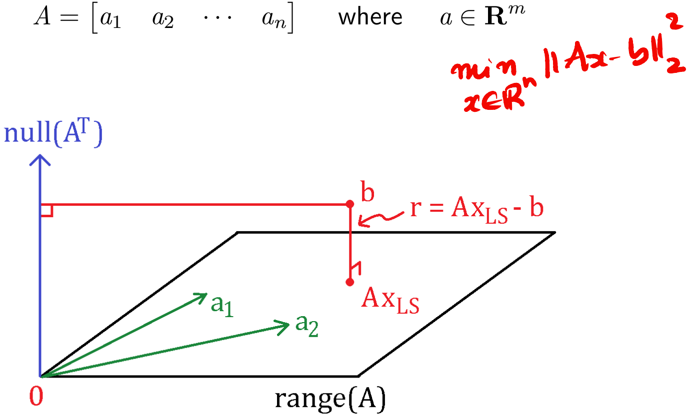
- for all $x \in \mathbf{R}^m$,

$$x = u + v, \quad u \in \mathbf{range}(A), \quad v \in \mathbf{null}(A^T), \quad u^T v = 0$$

- u and v are **uniquely determined**.

Geometry

$$A = \begin{bmatrix} a_1 & a_2 & \cdots & a_n \end{bmatrix} \quad \text{where} \quad a \in \mathbf{R}^m$$



$$b = Ax_{LS} + r, \qquad Ax_{LS} \in \mathbf{range}(A), \qquad r \in \mathbf{null}(A^T)$$

Least-squares optimality

- orthogonality of residual $r = Ax - b$ with columns of A

$$
\begin{array}{c} a_1^T r = 0 \\ \vdots \\ a_n^T r = 0 \end{array} \quad \text{or} \quad \underbrace{\begin{bmatrix} a_1^T \\ a_2^T \\ \vdots \\ a_n^T \end{bmatrix}}_{A^T} r = 0 \quad \text{or} \quad A^T r = 0
$$

- equivalent conditions:

$$
A^T r = 0 \quad (r = Ax - b) \qquad \text{and} \qquad \boxed{A^T A x = A^T b}
$$

- projection $y := Ax_{\text{LS}}$ is **unique**: suppose $z \neq y$ and $z \in \mathbf{range}(A)$. Then $z - y \perp b - y$ and

orthogonality

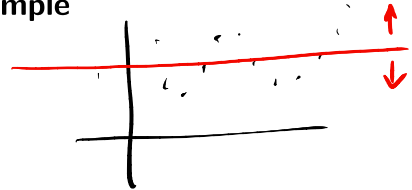
$$
\|b - z\|^2 = \|b - y + y - z\|^2 = \|b - y\|^2 + \|y - z\|^2 > \|b - y\|^2
$$

Thus, no vector other than y is optimal.

- x_{LS} is unique if and only if A is full rank

$N(A) = N(A^T A)$

Example



$$\underset{x}{\text{minimize}} \ \frac{1}{2}\|Ax - b\|_2^2$$

where

$$A = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} = e \in \mathbf{R}^m \quad \text{and} \quad b = \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix}$$

$\mathbb{1}$

$$e_1 = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

What is x^*?

$$A^T A x = A^T b$$

$$\Rightarrow \quad m x = \sum b$$

$$\Rightarrow \quad x_{LS} = \frac{1}{m} \sum b$$

Least squares

$$\underset{x}{\text{minimize}} \quad f(x) := \frac{1}{2}\|Ax - b\|_2^2 = \frac{1}{2}\sum_{i=1}^{m}(a_i^T x - b_i)^2$$

A matrix $A \in \mathbb{R}^{n \times n}$ is PSD if $x^T A x \geq 0$ for all $x \in \mathbb{R}^n$.

- Gradient

$$\nabla f(x) = A^T(Ax - b)$$

- Hessian

$$\nabla^2 f(x) = A^T A$$

- Is the Hessian positive definite? positive semidefinite?

  Ans: Always positive semidefinite. Positive definite if A has full row rank.

- Conditions for $x = x^*$ to be a global minimizer?

$$A^T(Ax^* - b) = 0$$

These are the **normal equations** of the least squares problem.