Professional program in AI

Least squares and optimization

May 30, 2019

Solving Ax=b

If A is square invertible, then $A\underline{x} = \underline{b}$ has a (unique) solution $\underline{x} = A^{-1}\underline{b}$.

In general, $A\underline{x} = \underline{b}$ could have zero, one or many (infinite) solutions.

Suppose A is tall, then $A\underline{x} = \underline{b}$ may not have any solution.

Least squares

Find $\underline{x}$ that minimizes the residual $A\underline{x} - \underline{b}$

minimize $\|A\underline{x} - \underline{b}\|^2$

A: $m \times n$

b: $m \times 1$

A, $\underline{b}$ are given

$m \geq n$.

Examples

Calculus based solution

Find the gradient and set to zero

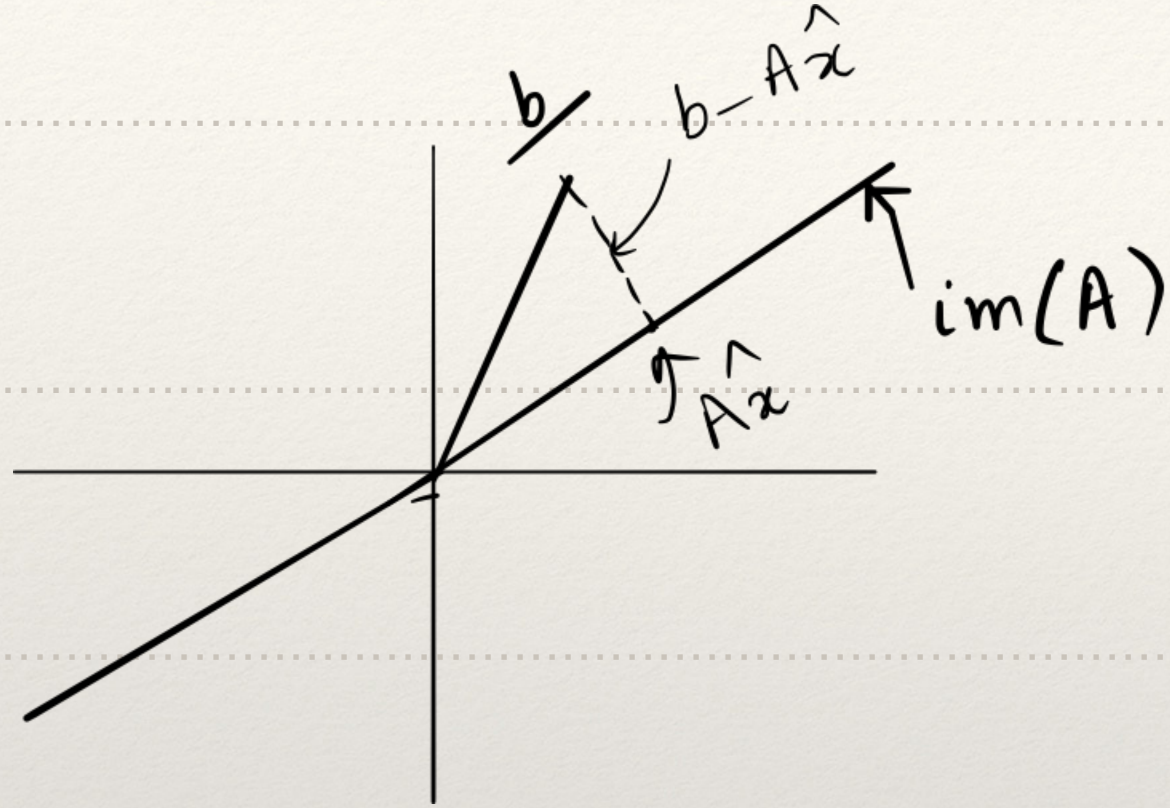
$$\nabla \|A\underline{x} - \underline{b}\|^2 = 2A^T(A\underline{x} - \underline{b})$$

Using chain rule.

$$\underline{x} = \underbrace{(A^TA)^{-1}A^T}_{\text{pseudo-inverse of } A}\underline{b}$$

Check second-derivative?

Projection interpretation



Error $b - A\hat{x}$ is orthogonal to $\text{im}(A)$.

$$A^T(b - A\hat{x}) = 0$$

Solving least squares

- Computing inverse

- QR decomposition

$$R\hat{x} = Q^T b$$

$$\underset{m\times n}{A} = \underset{m\times n}{Q}\ \underset{n\times n}{R}$$

- Gradient descent

Matrix least squares

Solve $\|AX - B\|^2$ where X is a matrix

$A: m \times n$, $X: n \times k$, $B: m \times k$

$$\|AX - B\|^2 = \sum_{i=1}^{m} \|A\underline{x_i} - \underline{b_i}\|^2$$

Solved by $\underline{x_i} = (A^T A)^{-1} A^T \underline{b_i}$.

Optimization basics

$$\min \quad f_0(\underline{x})$$

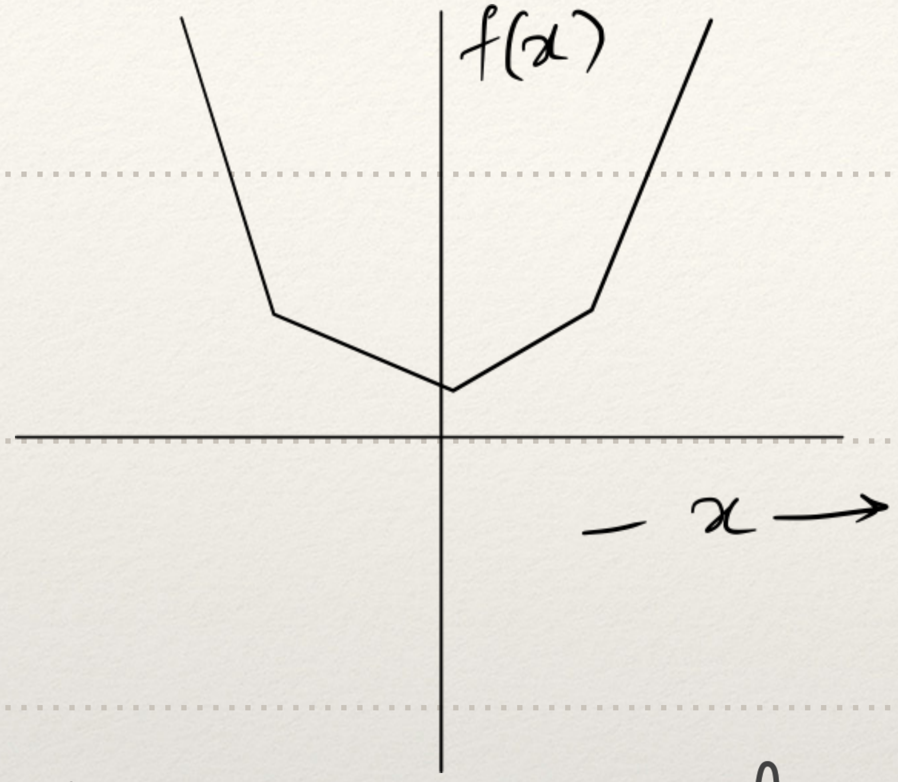
$$\text{subject to} \quad f_i(\underline{x}) \leq 0 \quad \} \text{ inequality constraints}$$

$$h_i(\underline{x}) = 0 \quad \} \text{ equality constraints}$$

If f_i are convex and h_i linear (affine)

efficient algorithms exist to solve the problem

Convexity



Example convex function.
in one variable

An important class of convex functions are defined via norms

$f(\underline{x}) = \|\underline{x}\|$, $\|A\underline{x} - \underline{b}\|$, etc are convex.

Unconstrained optimization

min $f_0(x)$ No constraints

- obtained by setting Gradient to zero
  assume differentiable / no-corners

- If $f_0(x)$ is convex, no need to check
  second order conditions

Constrained optimization

$$\min \quad f_0(x)$$
$$\text{s.t} \quad f_i(x) \leq 0$$
$$h_i(x) = 0$$

Assume f_i convex and h_i: linear / affine.

Can be solved numerically

To obtain explicit solution, formulate the KKT conditions

KKT conditions

$$f_0(x) + \sum_{i=1}^{p} \lambda_i f_i(x) + \sum_{i=1}^{r} \nu_i h_i(x)$$

Find the gradient, set to zero.