Introduction to Nonlinear Optimization

(Evans Note, Ch. 1)

Unconstrained Optimization

$$\min_{x \in \mathbb{R}^n} f(x)$$

First derivative test: $\nabla f(x) = 0$

x is a critical pt.

Gradient: $\nabla f(x)$

(1D: $f'(x) = 0$)

A bit of notation:

$$\nabla f(x) = \begin{pmatrix} \partial_{x_1} f(x) \\ \partial_{x_2} f(x) \\ \vdots \\ \partial_{x_n} f(x) \end{pmatrix}$$ column vector

$$Df(x) = (\partial_{x_1} f, \partial_{x_2} f, \cdots \partial_{x_n} f)$$ row vector

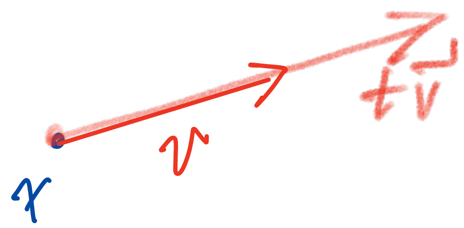
$$\left(Df = (\nabla f)^T \text{ or } \nabla f = (Df)^T \right)$$

(1) $\nabla f(x)$: direction of steepest ascent

(2) $-\nabla f(x)$: direction of steepest descent.

(3) Directional derivative:

$$\frac{d}{dt} f(x + t\vec{v}) \Big|_{t=0} = \langle \nabla f(x), \vec{v} \rangle$$



Second Derivative test: $[D^2 f(x)] > 0$

(Stability / local minimizer) (1D : $f''(x) > 0$)

Hessian $H = [D^2 f(x)] = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1 \partial x_1} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots \\ \cdots & \frac{\partial^2 f}{\partial x_i \partial x_j} & \cdots \\ \cdots & \cdots & \cdots \end{bmatrix}$

$H^T = H$ (Symmetric $n \times n$ matrix) $\left(\frac{\partial^2 f}{\partial x_i \partial x_j} = \frac{\partial^2 f}{\partial x_j \partial x_i} \right)$

$H > 0$ if $\langle Hv, v\rangle > 0$ for $v \neq 0$
$(\geq)$ $(\geq)$

$H > 0$: positive definite

$H \geq 0$: semi-positive definite

A bit of notation: $H = (H_{ij})$, $V = (V_i)$

$$\langle HV, V\rangle = (HV)^T V$$
$$= \sum_i (HV)_i V_i$$
$$= \sum_i \Big(\sum_j H_{ij} V_j\Big) V_i$$
$$= \sum_i \sum_j H_{ij} V_i V_j$$

$H > 0$ iff all eigenvalues of H are > 0
$(\geq)$ (≥ 0)

Taylor Series Expansion

(1D) $$f(x) = f(x_0) + f'(x_0)(x - x_0) + \frac{1}{2} f''(x_0)(x - x_0)^2 + \cdots$$

(nD) $$f(x) = f(x_0) + \langle \nabla f(x_0), x - x_0 \rangle + \frac{1}{2} \langle D^2 f(x_0)(x - x_0), (x - x_0) \rangle + \cdots$$

$$= f(x_0) + \sum_i \partial_{x_i} f(x_0)(x_i - x_{0_i}) + \frac{1}{2} \sum_{i,j} \partial^2_{x_i x_j} f(x_0)(x_i - x_{0_i})(x_j - x_{0_j})$$

Examples (Evans, Ch.)

(1) Fermat's Principle / Snell's Law

(2) Reflection

(3) Steiner Tree

(4) Electric Circuit.

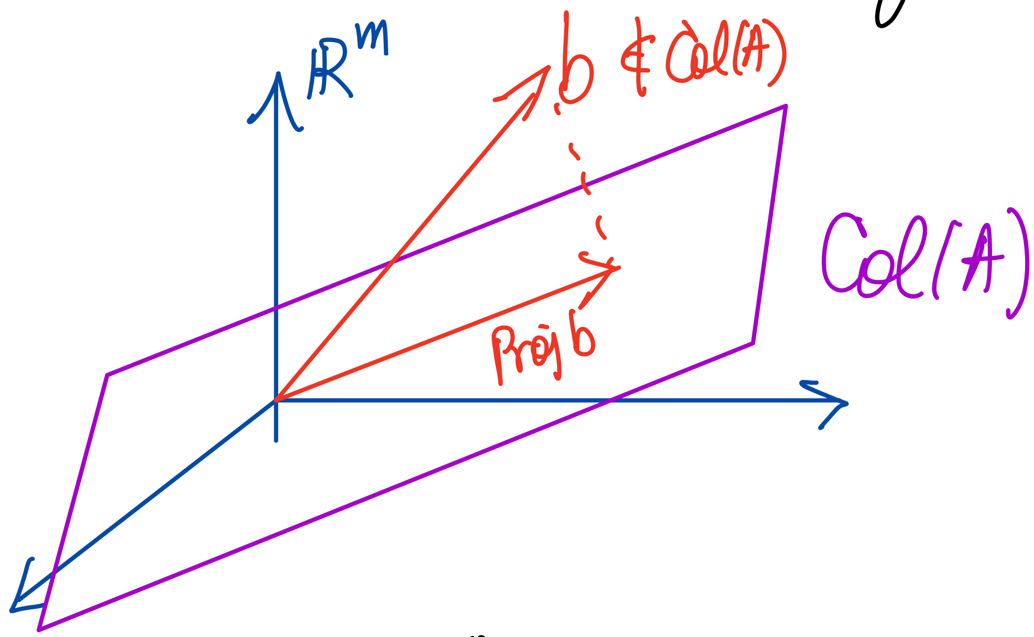
Least Square Approximation

$AX = b$ (might not be solvable)
$b \notin Col(A)$

$AX = Proj\, b$ (Project b onto $Col(A)$
— must be solvable)

$\Updownarrow$

$A^T A X = A^T b$ (normal equation
—always solvable)



If $Rank(A^{m\times n}) = n$ (A has linearly independent columns)

$\Rightarrow A^T A$ is invertible

$\Rightarrow X = (A^T A)^{-1} A^T b$ (least square soln)

$$\min f(x) = \|Ax - b\|^2$$

usual distance

$$f(x) = \langle Ax - b, Ax - b \rangle$$

$$= (Ax - b)^T (Ax - b)$$

$$= (x^T A^T - b^T)(Ax - b)$$

$$= x^T A^T A x - x^T A^T b - b^T A x + b^T b$$

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

a row vector

$$\nabla f(x) = 0 \iff \underline{A^T A x = A^T b}$$

"Component form"

$A^{m \times n}$

$$f(x) = \sum_{i=1}^{m} \left((Ax)_i - b_i \right)^2$$

$$= \sum_{i=1}^{m} \left(\sum_{j=1}^{n} A_{ij} x_j - b_i \right)^2$$

$$\partial_{x_k} f(x) = 2\sum_i \left(\sum_j A_{ij} x_j - b_i\right) A_{ik}$$

$$= 2\left(\sum_i \sum_j A_{ij} A_{ik} x_j - \sum_i A_{ik} b_i\right)$$

$$= 2\left(\sum_j \left(\sum_i A_{ik} A_{ij}\right) x_j - \sum_i A_{ik} b_i\right)$$

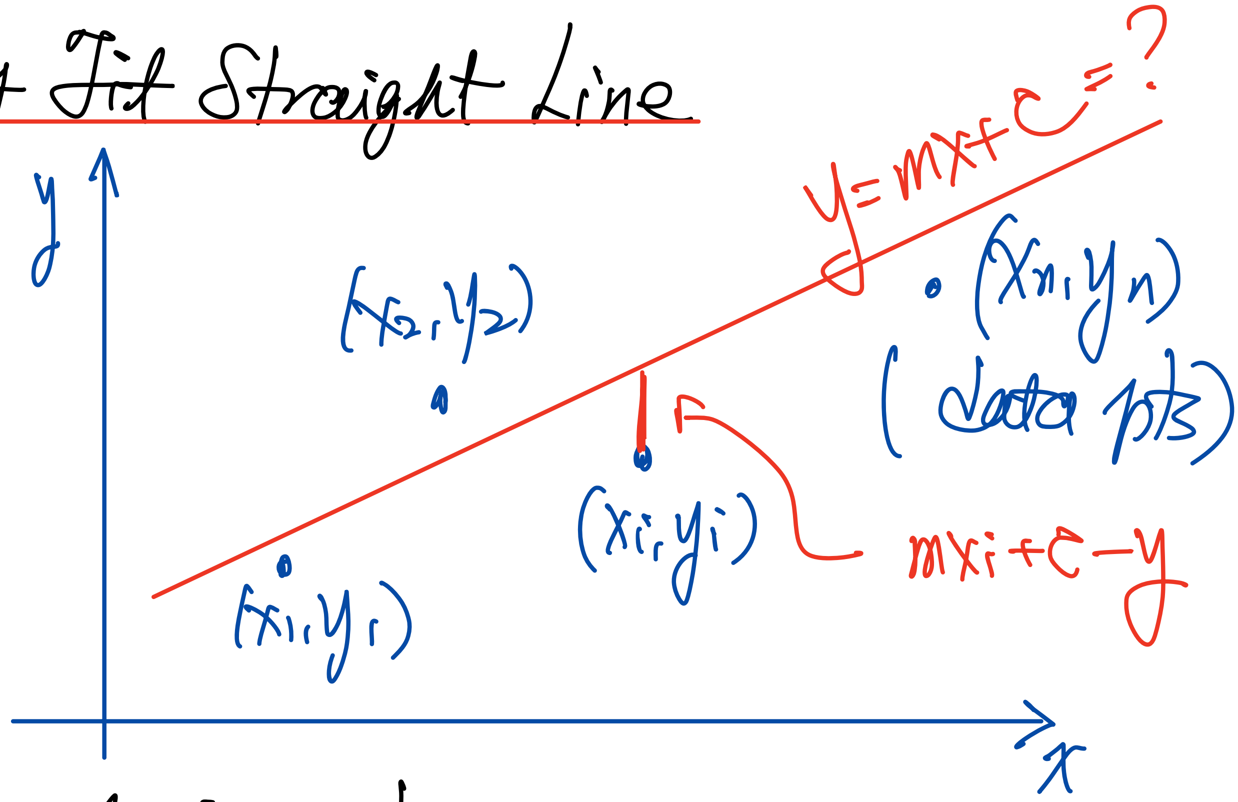
$$= 2\left(\underbrace{\sum_j \underbrace{\left(\sum_i A^T_{ki} A_{ij}\right)}_{(A^TA)_{kj}} x_j}_{(A^TAX)_k} - \underbrace{\sum_i A^T_{ki} b_i}_{(A^Tb)_k}\right)$$

$$= 2\left(A^TAX - A^Tb\right)_k$$

Hence $\nabla f(x) = 0 \iff A^TAX - A^Tb = 0$

i.e. $\underline{A^TAX = A^Tb}$

Best Fit Straight Line



Find m, c s.t

$$\begin{cases} mx_1 + c = y_1 \\ mx_2 + c = y_2 \\ \vdots \quad \vdots \\ mx_n + c = y_n \end{cases} \iff \begin{bmatrix} x_1 & 1 \\ x_2 & 1 \\ \vdots & \\ x_n & 1 \end{bmatrix} \begin{bmatrix} m \\ c \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

normal equation

$$\begin{bmatrix} x_1 & x_2 & \cdots & x_n \\ 1 & 1 & \cdots & 1 \end{bmatrix} \begin{bmatrix} x_1 & 1 \\ x_2 & 1 \\ \vdots & \vdots \\ x_n & 1 \end{bmatrix} \begin{bmatrix} m \\ c \end{bmatrix} = \begin{bmatrix} x_1 & \cdots & x_n \\ 1 & \cdots & 1 \end{bmatrix} \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$$

$$\boxed{\begin{bmatrix} \sum x_i^2 & \sum x_i \\ \sum x_i & n \end{bmatrix} \begin{bmatrix} m \\ c \end{bmatrix} = \begin{bmatrix} \sum x_i y_i \\ \sum y_i \end{bmatrix}}$$

$$e(m,c) = \sum_{i=1}^{n} (m x_i + c - y_i)^2$$
(error)

mean square error

$$\frac{\partial e}{\partial m} = 2\sum_{i=1}^{n} (m x_i + c - y_i)\, x_i$$
$$= 2\left[m\left(\sum_{i=1}^{n} x_i^2\right) + c\left(\sum_{i=1}^{n} x_i\right) - \left(\sum_{i} x_i y_i\right)\right]$$

$$\frac{\partial e}{\partial c} = 2\sum_{i=1}^{n} (m x_i + c - y_i)$$
$$= 2\left[m\left(\sum_{i=1}^{n} x_i\right) + nc - \sum_{i=1}^{n} y_i\right]$$

$$\left(\frac{\partial e}{\partial m} = 0,\ \frac{\partial e}{\partial c} = 0\right) \Rightarrow$$

$$m\left(\sum_{i=1}^{n} x_i^2\right) + c\left(\sum_{i=1}^{n} x_i\right) = \left(\sum_{i} x_i y_i\right)$$
$$m\left(\sum_{i=1}^{n} x_i\right) + nc = \left(\sum_{i=1}^{n} y_i\right)$$

$$\boxed{\begin{bmatrix} \sum_i x_i^2 & \sum_i x_i \\ \sum_i x_i & n \end{bmatrix} \begin{bmatrix} m \\ c \end{bmatrix} = \begin{bmatrix} \sum_i x_i y_i \\ \sum y_i \end{bmatrix}}$$

i.e. $$\boxed{\begin{bmatrix} \|X\|_2^2 & \bar{X} \\ \bar{X} & n \end{bmatrix} \begin{bmatrix} m \\ c \end{bmatrix} = \begin{bmatrix} \langle X, Y \rangle \\ \bar{Y} \end{bmatrix}}$$

Other error measures

L^1 : $$e_1 = \|AX - b\|_1 = \sum_i |(AX)_i - b_i|$$

L^∞ : $$e_\infty = \max_i |(AX)_i - b_i|$$

(Chebyshev Approximation)