What is optimization?

$\{ s_1, s_2, \dots \}$, $f(s)$

set of possible solutions S ; objective function

Find s^* such that $\underbrace{f(s^*) \leq f(s_i)}_{\text{Minimization (cf. maximization)}}$ $s_i \neq s^*$

Optimization is also known as "programming"

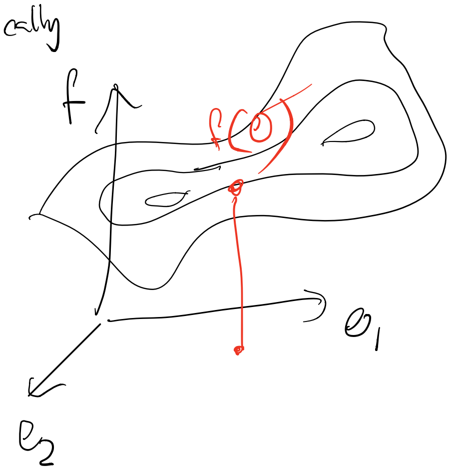
↙ linear programming

↓ dynamic programming

↘ stochastic programming

$\underline{f(x_1, x_2, \dots, x_n)} = \underline{\underline{S}}$

Geometrically



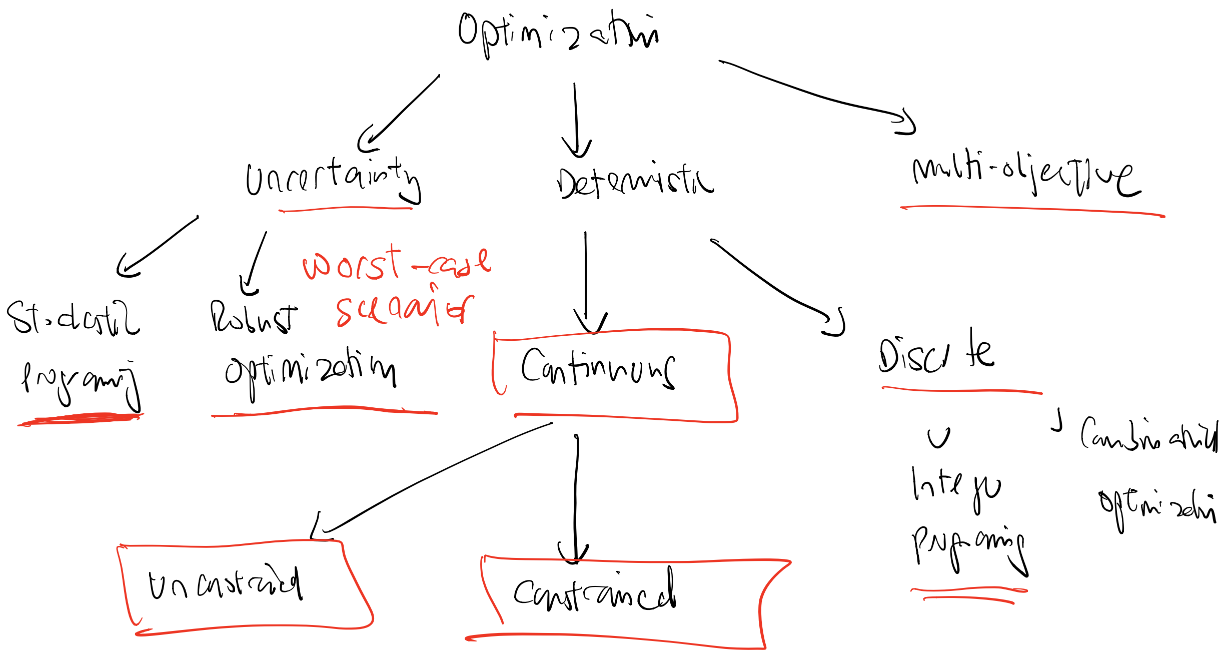
f is a surface

$f : \mathbb{R}^n \mapsto \mathbb{R}$

Applications → optimization occurs everywhere you want to find the "best"

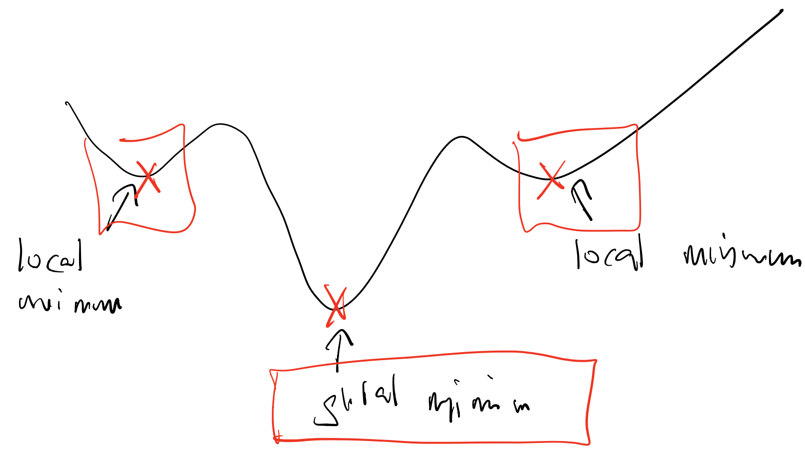
In statistics ⟶ "Best fit" model

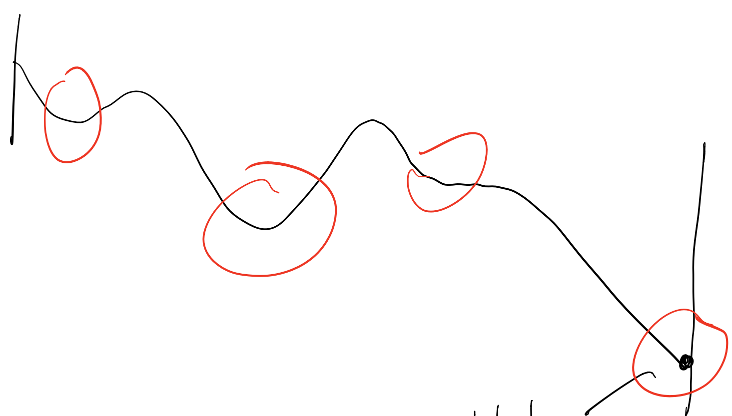
Optimization Tree



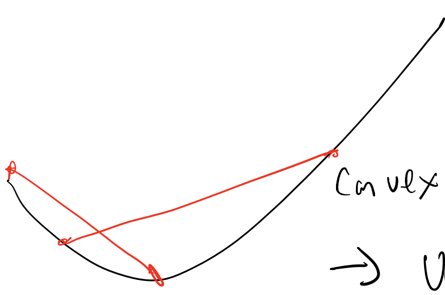
Concepts

Global vs local solution





global minimum can be an boundary in constrained problems

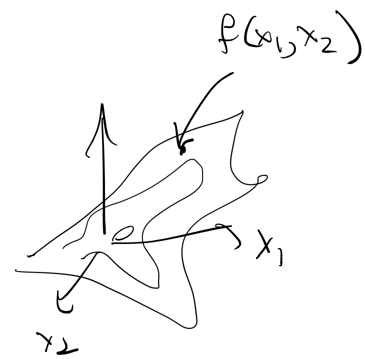


Convex objective function

→ <u>Unique</u> global minimum

Calculus review

$$\underline{\text{Min} \quad f(x_1, x_2, \dots, x_k)}$$



Consider 2 variables

$$\text{Min} \quad f(x_1, x_2)$$

$$\nabla f = \left[\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2} \right]^T \longrightarrow \text{1st derivatives}$$

nabla "del"

$$H = \nabla^2 f = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} \end{bmatrix} \qquad \frac{\partial^2 f}{\partial x_1 x_2} = \frac{\partial^2 f}{\partial x_2 \partial x_1}$$

Hessian

$$\Rightarrow \underline{H \text{ is symmetric}}$$

$$f(x,y) = 3x^3 + 4x^2y + y^3 + 7 \qquad H \sim D.$$

$$\nabla f = [\underline{9x^2 + 8xy}, \; 4x^2 + 3y^2]^T$$

$$H = \nabla^2 f = \begin{bmatrix} 18x + 8y & 8x \\ 8x & 6y \end{bmatrix}$$

$\boxed{\text{Jacobian}}$ → First derivatives for set of functions $f_1, f_2, \dots$

E.g. Hessian is Jacobian of gradient.

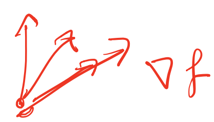
Gradient

Directional derivative

let v be a unit vector

$\nabla f \cdot v$ = directional derivative $\nabla_v f$

$\nabla f \cdot v = |\nabla f|\,|v| \cos\theta$

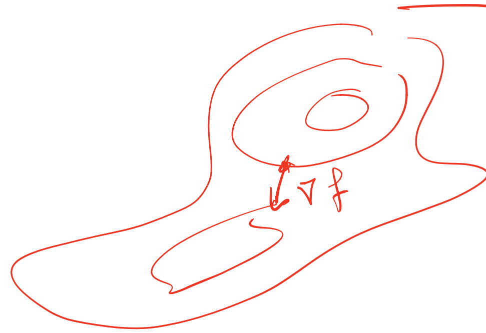


max when $\cos\theta = 1$

$\Rightarrow$ v is in same direction as ∇f

∇f is direction of steepest ascent

$-\nabla f$ is direction of steepest descent



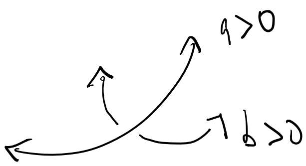
Hessian $\longrightarrow$ local curvature

$$f(x,y) = \frac{1}{2}(ax^2 + by^2)$$

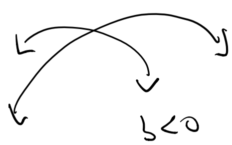
$$\nabla f = \begin{bmatrix} ax \\ by \end{bmatrix}$$

$$H = \begin{bmatrix} a & 0 \\ 0 & b \end{bmatrix}$$

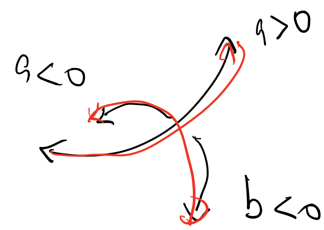
eigenvalues are a, b
with eigenvectors e_1, e_2



$\lambda_1, \lambda_2 > 0$
Positive definite
local minima



$\lambda_1, \lambda_2 < 0$
Negative definite
local maxima



$\lambda_1 > 0, \lambda_2 < 0$
Saddle

2nd derivative test

Taylor

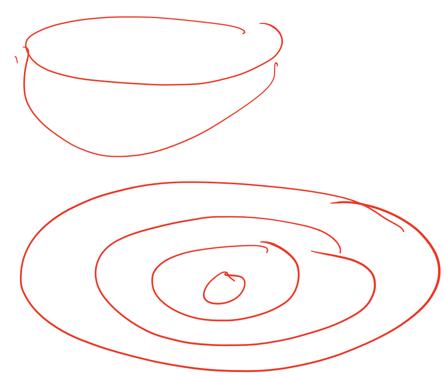
$x, h \in \mathbb{R}$

$$f(x+h) = f(x) + f'(x)\,h + \frac{f''(x)}{2} h^2 + \dots$$

$x, h \in \mathbb{R}^n$

Hessian

$$f(x+h) = f(x) + \nabla f^T h + \frac{1}{2} h^T \nabla^2 f\, h + \dots$$



Root finding

Root finding is equivalent to solving algebraic equations

$$f(x) = g(x)$$

x — roots, zeros

$$\Rightarrow \quad h(x) = f(x) - g(x) = 0$$

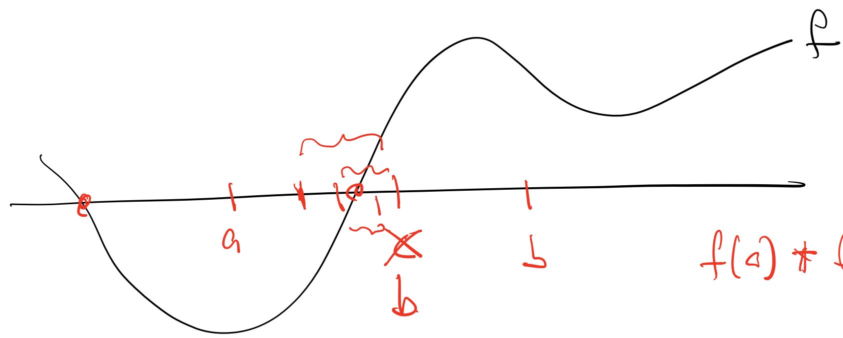
Solutions are roots or zeros of $h(x)$

Connection to optimization

For optimal solution $\boxed{f'(x) = 0}$

So looking for roots of $f'(x)$.

Bisection



$f(a) * f(b) < 0$

Brackets (a, b) → Guarantee (safety)

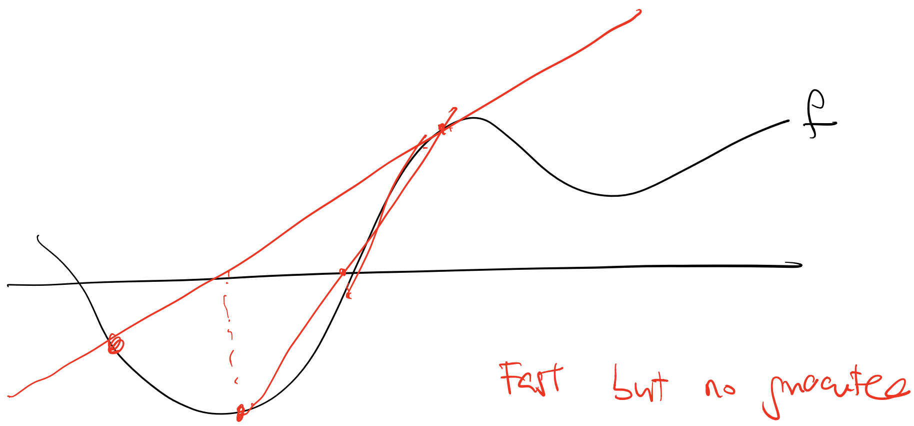
stopping
- x doesn't change
- $f(x)$ is close to 0

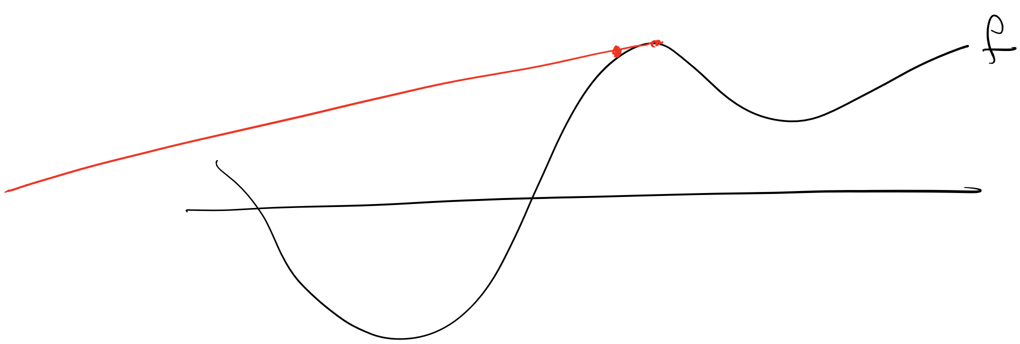
} tolerance

$f(x) < 1e^{-6}$ stop.

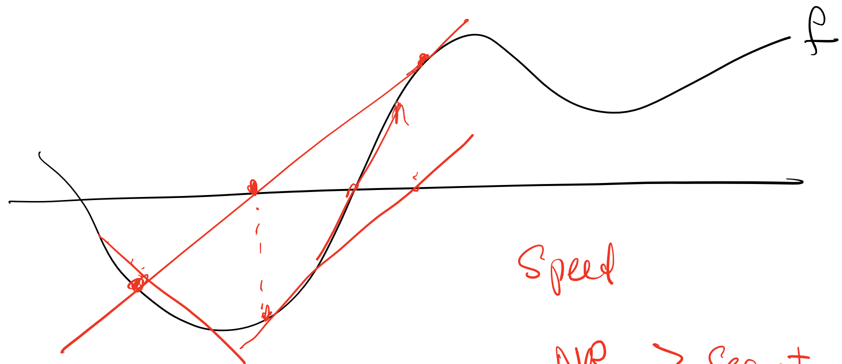
$x_{k+1} - x_k < 1e^{-6}$

Secant ? False position





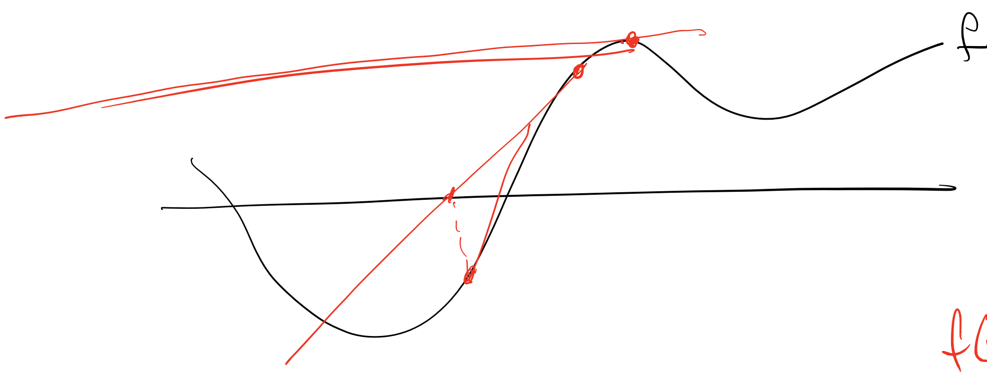
Newton – Raphson



Speed

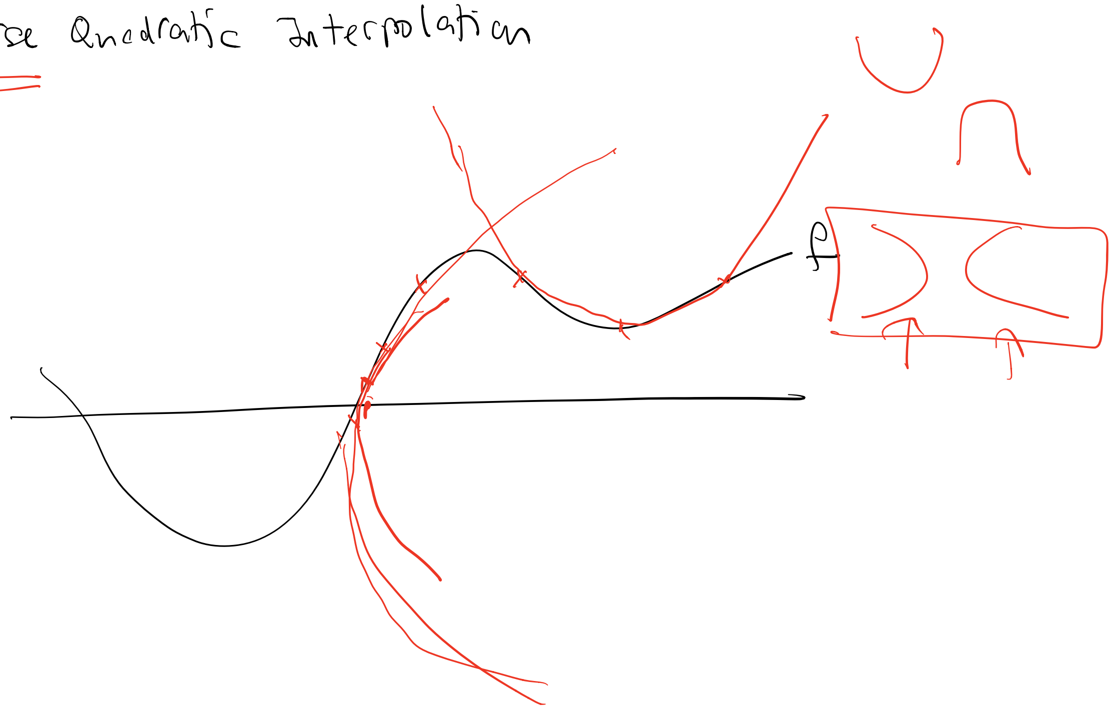
NR > Secant > Bisection

Quadratic 2 | 1.618... | linear 1



$$x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)}$$

Inverse Quadratic Interpolation



Brent's method

Combines → secant
→ Inverse quadratic interpolation
→ Bisection

⇒ speed + safety

scipy.optimize

Gauss-Newton

Newton - Raphson

$$x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)}$$

Gauss - Newton

$$x_{k+1} \approx x_k - \underbrace{J^{-1}}_{\text{Not square}} f(x_k)$$

$$x_{k+1} = x_k - \underbrace{(J^T J)^{-1} J^T}_{\text{" } \overline{f'(x)} \text{ "}} f(x_k)$$

J = Jacobian.

Roots of Polynomials

$$p(x) = a_0 + a_1 x + a_2 x^2 + \dots + a_{n-1} x^{n-1} + x_n$$

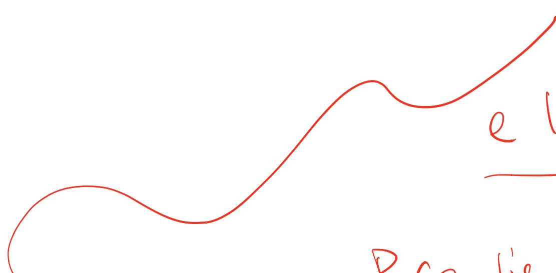
Companion matrix

$$A = \begin{bmatrix} 0 & 0 & \cdots & 0 & -a_0 \\ 1 & 0 & \cdots & 0 & -a_1 \\ 0 & 1 & \cdots & 0 & -a_2 \\ \vdots & & & & \\ 0 & \cdots & & 1 & -a_{n-1} \end{bmatrix}$$

Characteristic Polynomial

$$|A - \lambda I| = a_0 + a_1 \lambda + a_2 \lambda^2 + \dots + \lambda^n$$

$\Rightarrow$ Eigenvalue of companion matrix are roots of polynomial function.



e ly accesib h

Brent's method
- Secant
- IQI
- Bis

Companion Matrix