Unconstrained Optimization. Ch 2 of N.W.

Identify a local minima → Use Taylor's thm

1st order or 2nd order conditions

↓

Overview of Algo's

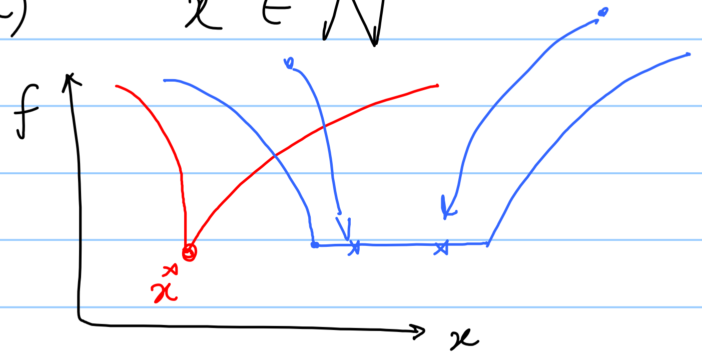
1) Line search
2) trust region

Identifying a local minima $\rightarrow x^*$

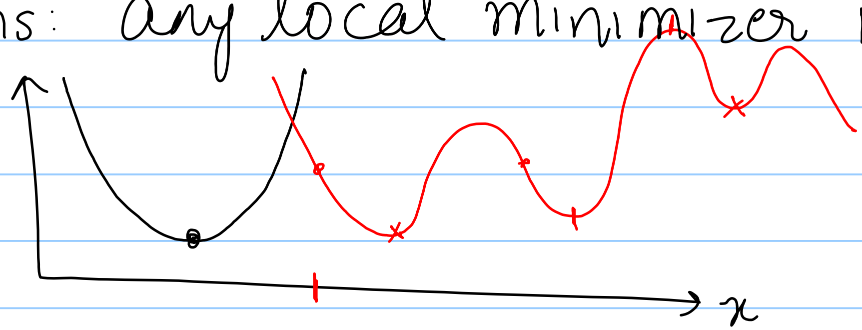
① Global minimizer : $f(x^*) \leq f(x) \quad \forall\, x \in dom(f)$

② Local minimizer : $f(x^*) \leq f(x) \quad x \in N$

adjectives: weak " $\leq$ ", strong " $<$ "



Convex fns: any local minimizer is also a global minimizer.



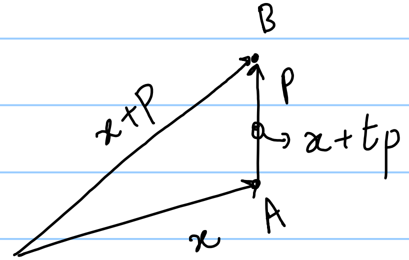
Taylors thm: $f: \mathbb{R}^n \to \mathbb{R}$, $x, p \in \mathbb{R}^n$, $t \in (0,1)$.

① Continuously differentiable (i.e. 1st order)

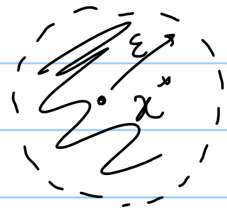
$$f(x+p) = f(x) + \nabla f(x+tp)^T p \qquad \text{(MVT)}$$

② Twice continuously differentiable fn (i.e. 2nd order)

$$f(x+p) = \underbrace{f(x) + \nabla f(x)^T p}_{\text{Taylor series + 1st order}} + \frac{1}{2} \underbrace{p^T \nabla^2 f(x+tp)\, p}_{\text{remainder term}}$$



↳ Open nbd of a point



$$\|x - x^*\| < \varepsilon$$

↳ 1^{st} order condition:

If x^* is a local minimizer & f is continuously differenteable in a open nbd of x^*, then $\nabla f(x^*) = 0$. [necessary condition].

Further, if $\nabla f(x^*) = 0$, then x^* is called a stationary point.

2^{nd} order Conditions:

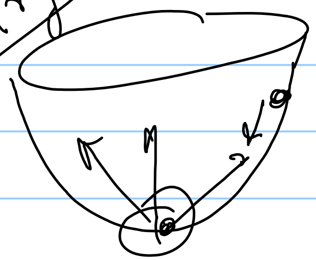
Require: $\nabla^2 f$ exists & is continuous on an open nbd of x^*

Necessary: If x^* is a minimizer, then:

1) $\nabla f(x^*) = 0$
2) $\nabla^2 f(x^*)$ is positive semi definite.

Sufficient: If both $\nabla f(x^*) = 0$ and $\nabla^2 f(x^*)$ is PSD then x^* is a local minimizer.

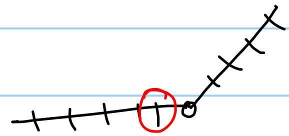
Geometry



weak → PSD

strong → PD

$$\begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2}, & \frac{\partial^2 f}{\partial x_2^2} \\ \frac{\partial^2 f}{\partial x_1 x_2} & \end{pmatrix}$$



Proof of 1^{st} order condition. $\rightarrow$ By contradiction.

$$f(x^* + p) = f(x^*) + \nabla f(x^* + tp)^T p$$

Assume x^* is a minimizer, but $\nabla f(x^*) \neq 0$.

① $g(t) = p^T \nabla f(x^* + tp)$

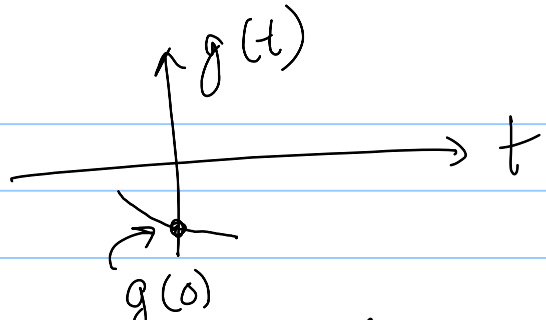
② We can choose any p. Say $p = -\nabla f(x^*)$

$$g(t) = -\nabla f(x^*)^T \nabla f(x^* + tp) \leftarrow$$

$$g(0) = -\|\nabla f(x^*)\|^2 < 0$$

③ $g(t)$ is a conts fn

$\therefore$



$\Rightarrow$ $\underline{\underline{g(t) < 0}}$ for some small t

(4) $$f(x^* + tp) = \underline{f(x^*)} + \underbrace{\nabla f(x^* + tp)^T p}_{g(t) < 0}$$

$$\Rightarrow f(x^* + tp) < f(x^*)$$

i.e. ~~$x^* + tp$ is a "better" minimizer~~

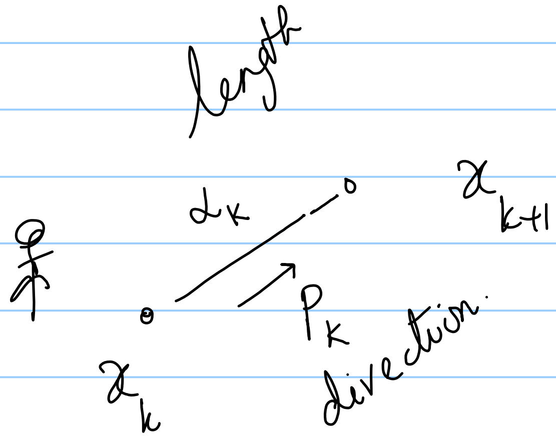
or x^* is not a minimizer.

Overview of Algorithms

Start from x_0 $\longrightarrow$ iterations $\{x_k\}$ $\longrightarrow$ Soln x^*

$$\{x_k\} \longrightarrow x_{k+1}$$

line search ← x_{k+1} → trust region



$$\boxed{\{\alpha^*, p^*\} = \underset{\alpha_k, p_k}{\arg\min}\ f(x_k + \alpha_k p_k)}$$

Line Search Algorithms $\rightarrow$ different choices of $\underline{p_k}$.

① Steepest descent / gradient descent

$$p_k = -\nabla f_k$$

② Newton method, $p_k = -(\nabla^2 f_k)^{-1} \nabla f_k$ AND

(quadratic) $\nabla^2 f$ to be P.D.

③ Quasi Newton, $p_k = -(B_k) \nabla f_k$

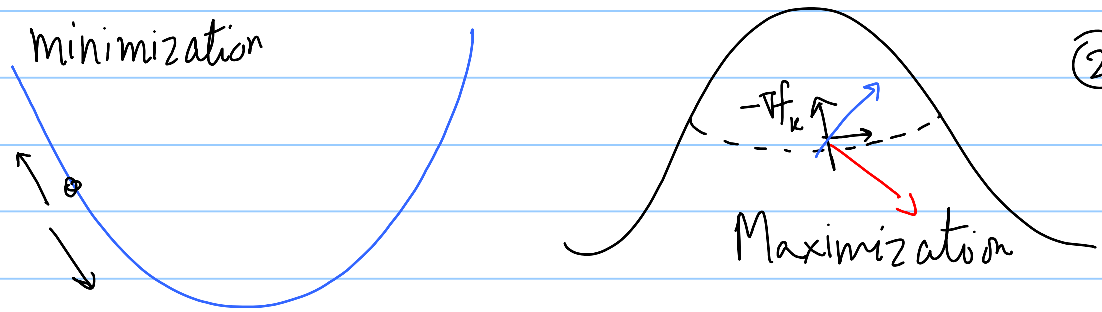
$\rightarrow$ Cheaper approx of $(\nabla^2 f_k)^{-1}$

④ Conjugate Gradient Method →

$$p_k = -\nabla f_k + \beta_k p_{k-1}$$

Descent directions → requirements.

① The direction of steepest descent is $-\nabla f_k$



② A descent direction is legitimate if it makes an angle strictly less than $\pi/2$ with $-\nabla f_k$.

How to choose x_0?

- ↳ Choose at random
- ↳ Start from $\vec{0}$
- ↳ Solve a simpler problem.
  - ↳ use that soln. to get to x_0
- ↳ Use ML to give a guess.

↳ Pick random $\{x_k\}$

choose $\max\{\|\nabla f(x_k)\|\}$