What if the Hessian is not PD?

→ Hessian Modification.

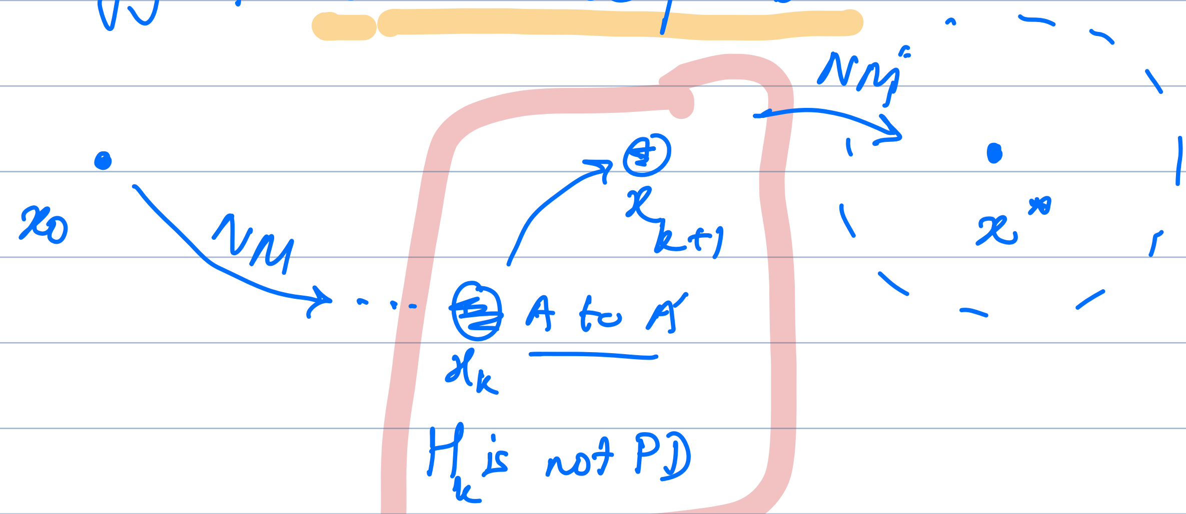
As I get close x^*; H is PD

When far way, H can be ~~PD~~.

→ Lin Alg: Cholesky decomp: → If A is a real sym PD matrix then $A = LL^T$, where L is lower triangular, diagonal entries are +ve. The converse also holds: If A can be exp as LL^T, then A is sym PD (L must low. triang. invertible).

We are x_k → $\underbrace{\nabla^2 f_k}_{A}$ is not PD. (H is sym)

Modify A as little as possible.



$A \longrightarrow A + \Delta A \implies$ as low norm as possible.

$\Delta A = \tau I \qquad \|\Delta A\|_2 = \tau$? $\qquad \underline{\tau > 0}$.

$A = Q \Lambda Q^T \qquad \Delta A = \tau \, Q I Q^T$

sym.

$\implies \underbrace{A + \Delta A}$ $= Q[\Lambda + \tau I] Q^T$

to become PD ?

Say I want $\lambda_{min}(A + \Delta A) > \delta$

$\implies \underline{\tau = \max(0, \delta - \lambda_{min}(A))}$

$O(n^3)$ op.

Avoid an EVD for practical reasons

→ Start with small value of τ_0

→ Increment it and ? $A + \tau I \rightarrow$

Check if I can do Cholesky decomp of A

If yes, ✓

If no

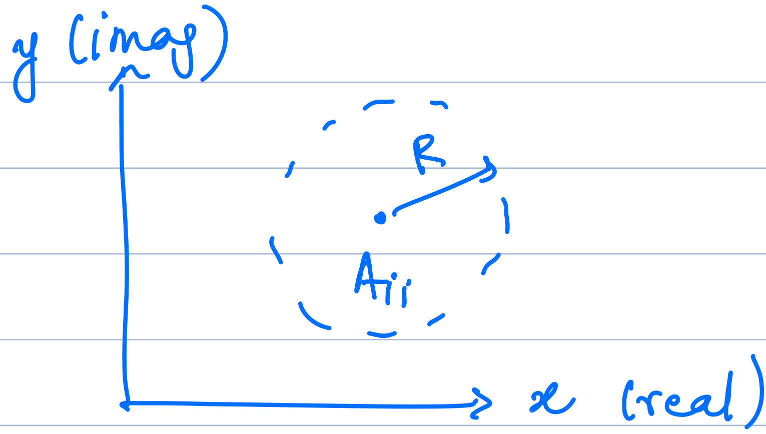
<u>3.3 of NW.</u>

↳ <u>Gershgorin Circle Theorem</u>.

Given a complex $n \times n$ matrix A.

Disc D_i = centre $\rightarrow A_{ii}$

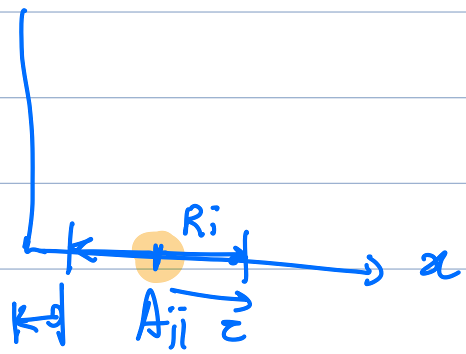
radius $\rightarrow \sum_{\substack{j \\ j \neq i}} |A_{ij}|$



Thm: Every eigenvalue of A lies within atleast one Disc D_i.

↳ In our case A is real

A



$A_{ii} - R_i > 0 \quad \forall i$

$A + \tau I$

A cheap way of computing τ.

— x —

Quasi Newton Method

SD $\rightarrow$ linear

QN $\rightarrow$ superlinear

$$\lim_{k \to \infty} \frac{\| x_{k+1} - x^* \|}{\| x_k - x^* \|} = \mu \qquad 0 < \mu < 1$$

$$\lim_{k \to \infty} \frac{\| x_{k+1} - x^* \|}{\| x_k - x^* \|} = 0$$

$$p_k = -(\nabla^2 f_k)^{-1} \nabla f_k$$

Sub $\nabla^2 f_k$

Intuition: quadratic model

$$m_k(p) = f_k + \nabla f_k^T p + \frac{1}{2} p^T B_k p$$

sym PD matrix updated at each step.

f_k

m_k? → agrees with f_k and ∇f_k at $p = 0$

$$\nabla m_k = \nabla f_k + B_k p$$

Take m_{k+1} : must match the gradient of f at both $k+1$ & k.

$$x_{k+1} = x_k + \underbrace{\alpha_k p_k}.$$ update.

How do I get ∇f_k from $m_{k+1}(p)$

$p = 0$, $m_{k+1} \Rightarrow x_{k+1}$

p? will give $m_{k+1} \to x_k$

$x_k = x_{k+1} - \alpha_k p_k$.

$$\underbrace{\nabla m_{k+1}(-\alpha_k p_k) = \nabla f_{k+1}}_{\text{by defn}} - \alpha_k B_{k+1} \underbrace{p_k = \nabla f_k}_{\text{by our demand.}}$$

$$B_{k+1}[\alpha_k p_k] = \underbrace{\nabla f_{k+1} - \nabla f_k}_{= y_k}$$

$$\underline{s_k = x_{k+1} - x_k}$$

$$\left[\underline{B_{k+1}} s_k = y_k.\right] \to \text{secant eqn.}$$

$\to$ n eqns.

$\to$ I want sym PD

$\Rightarrow$ how many vars $= \frac{n(n+1)}{2}$

$\downarrow$ PD

$$s_k^T B_{k+1} s_k = \underbrace{s_k^T y_k > 0}_{\text{we want.}} \quad \text{if PD.}$$

$\Rightarrow$ w/o proof: If you use Wolfe cond^{n}s

(6.1) the above is satisfied always. (α_k)

$$s_k = B_{k+1}^{-1} y_k = H_{k+1} y_k.$$

L-BFGS relation: $\quad p_k = -H_k \nabla f_k.$

$$H_{k+1} = (I - \rho_k s_k y_k^T) H_k (I - \rho_k s_k y_k^T) + \rho_k s_k s_k^T$$

where $\rho_k = \dfrac{1}{y_k^T s_k}$.

1) start approx H_0
2) line search w/ wolfe condn α_k.
3) $p_k = - H_k \nabla f_k$, using BFGS reln.

—x—