Non linear CG (cont...)

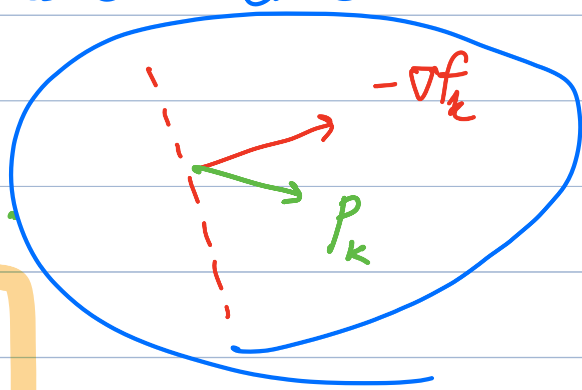
No matrix anymore, so conjugacy ?

p_k legitimate? $\qquad x_{k+1} = x_k + \alpha_k p_k$.

$\Rightarrow$ make sure p_k is a descent direction.

If so:

$$\underline{\nabla f_k^T p_k < 0}$$



$$\left[\; p_k = -\nabla f_k + \beta_k p_{k-1} \;\right]$$

I want p_k to be a descent dir.

$$\left[\; \nabla f_k^T p_k = -\|\nabla f_k\|^2 + \underbrace{\beta_k \nabla f_k^T p_{k-1}}_{} < 0 \;\right]$$

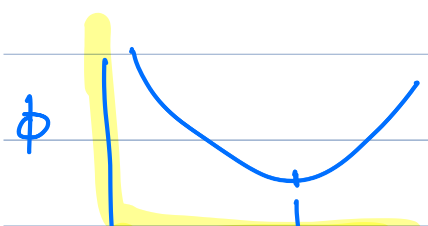
Need to focus here

Need to make sure $\nabla f_k^T p_{k-1} \leq 0$.

Say I was doing exact line search; ie α_{k-1} is exact. Sitting at $x_{k-1} \to$ wanting to go to x_k.

$$\phi(\alpha) = f(x_{k-1} + \alpha p_{k-1})$$

$$\frac{d\phi(\alpha)}{d\alpha} = \nabla f(\underbrace{\text{'' ''}}_{x_k})^T p_{k-1}$$



$= 0$ when exact LS.

α_{k-1} $\qquad \nabla f_k^T p_{k-1} = 0$

$\Rightarrow$ When exact LS, then p_k is also a descent dir.

exact LS $\to$ not easy.

$\Rightarrow$ $\nabla f_k^T p_k$ may be > 0.

The fix: Impose strong Wolf condns with $0 < c_1 < c_2 < 0.5$ then all directions are descent directions.

$\hookrightarrow$ when p_k becomes "bad" $\theta \approx 90^\circ$

$\Rightarrow$ Not much progress.

You can show $p_{k+1} \approx p_k$.

$\hookrightarrow$ Should reset, $\beta = 0$

$\to$ "Polak-Ribiere" NLCG

Newton & Newton-like Methods

$\hookrightarrow$ Ch 3, 6, 7 of NW.

Motivation: quadratic convergence

price : second order info.

It is a line search $\Rightarrow x_{k+1} = x_k + \alpha_k p_k$.

↪ α : (in)exact line search algo.

↪ descent dir : $p_k^N = -(\nabla^2 f_k)^{-1} \nabla f_k$

$p_k^{SD} = -(I) \nabla f_k$.

↪ Hessian needed to be PD. Can show this $\Rightarrow$ p_k is descent dir.

How?

$$\cos\theta = \frac{-\nabla f_k^T p_k}{\|\nabla f_k\| \, \|p_k\|}$$

$$= \frac{+\nabla f_k^T (\nabla^2 f_k)^{-1} \nabla f_k}{(\quad)(\quad)} > 0$$

for descent dir.

If $(\nabla^2 f_k)^{-1}$ is PD then $y^T A y > 0$

Requirement.

Terminology.

$$p_k = -B_k^{-1} \nabla f_k$$

① If $B_k = \nabla^2 f_k \longrightarrow$ Newton method

② If B_k is an approx $\rightarrow$ quasi Newton method.

$$\left(\begin{array}{ll} B_k = Q \Lambda Q^T & \Lambda = \mathrm{diag}(\lambda_i) \\ B_k^{-1} = Q \Lambda^{-1} Q^T & Q \to \text{orthogonal} \end{array}\right)$$

Convergence.

Zoutendijk's condn

$$f_{k+1} \leq f_0 - c \underbrace{\sum_{j=0}^{k} \cos^2\theta_j \|\nabla f_j\|^2}$$

$$\Rightarrow \quad \sum \cos^2\theta_j \|\nabla f_j\|^2 < \infty$$

$$\lim_{j\to\infty} \cos\theta_j \|\nabla f_j\| \to 0$$

either $\cos \to 0$ or $\nabla f \to 0$

$\cos\theta > 0 \quad \Leftarrow$ We want this

↳ Proof uses an assumption that

$\kappa(B_k) \leq M$ Condn no.

$\Rightarrow \cos\theta \geq \dfrac{1}{M}$.

Proof:

$$\cos\theta = \frac{-\nabla f^T p}{\|\nabla f\| \|p\|}$$

① If B is PD $\Rightarrow$ B^{-1} exists

$B^{1/2}$, $B^{-1/2}$ exists, $\|B^{1/2}\| = \|B\|^{1/2}$.

② Euclidean norm: $\|AB\| \leq \|A\| \|B\|$.

$$p = -B^{-1}\nabla f \Rightarrow \nabla f = Bp$$

$$\cos\theta = \frac{p^T Bp}{\|Bp\| \|p\|} \qquad B = B^{1/2} B^{1/2}$$

$$\boxed{\kappa(B) = \|B\| \|B^{-1}\| \leq M.}$$

$$\cos\theta \geq \frac{p^T Bp}{\|B\| \|p\|^2} = \frac{\|B^{1/2}p\|^2}{\|B\| \|p\|^2}$$

$$p = B^{-1/2} B^{1/2} p$$

$$\geq \frac{\|B^{1/2}p\|^2}{\|B\| \left(\|B^{-1/2}\| \|B^{1/2}p\|\right)^2}$$

$$= \frac{1}{\|B\| \cdot \|B^{-1}\|} \geq \frac{1}{M}$$

↳ Rate: Ref sec 3.5 of NW. for proof.

Quadratic conv..

$$\| x_{k+1} - x^* \| \leq L \, \| \nabla^2 f(x^*)^{-1} \| \, \| x_k - x^* \|^2$$

$\Rightarrow$ True in a nbd of x^* when H is guaranteed to be PD.