Line Search Algorithms.

$$\{\alpha_k^*, p_k^*\} = \underset{\alpha, p_k}{\operatorname{argmin}}\ f(x_k + \alpha_k p_k) \qquad \alpha \in [0, A]$$

leads to $\quad x_{k+1} = x_k + \alpha_k^* p_k^*$.

choices of $p_k \to$ 4 different families.

1) Steepest descent, $p_k = -\nabla f_k$
2) Newton method, $\boxed{p_k = -(\nabla^2 f_k)^{-1} \nabla f_k}$ $*$
3) Quasi Newton, like (2) but cheaper approx of $\nabla^2 f$.
4) Conjugate gradient, $p_k = \underbrace{-\nabla f_k}_{} + \underbrace{\beta_k p_{k-1}}_{}$

$\left((\nabla^2 f)^{-1} \text{ is pos def}\right)$

Descent Directions

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

$\to$ Proof:

$$x_k \xrightarrow{\varepsilon} p_k$$

$$f(x_k + \varepsilon p_k) = f(x_k) + \varepsilon p_k^T \nabla f_k + O(\varepsilon^2) \leftarrow$$

vector

scalar

1st order

ε be very small, the $O(\varepsilon^2) \to 0$

$$p_k^T \nabla f_k = \underbrace{\| p_k \|}_{>0} \; \underbrace{\| \nabla f_k \|}_{\geq 0} \cos \theta_k$$

I want $f(x_k + \varepsilon p_k)$ to be $< f(x_k)$

$\Rightarrow$ Max decrease when $\theta_k = \pi$

$\Rightarrow$ $p_k = -\nabla f_k$. $\Rightarrow$ Direction of max decrease.

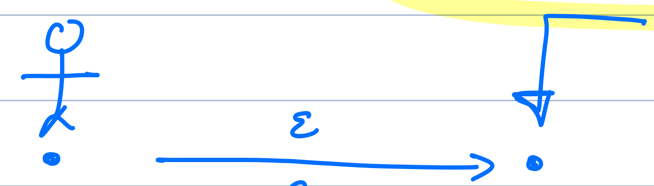
$\Rightarrow$ when $|\theta_k| > \pi/2$ $\cos \theta_k < 0$

$\Rightarrow$ angle with $-\nabla f_k$ is $< \pi/2$.

As long as this is true p_k will lead to a decrease in f.

3) The Newton direction is given by $-(\nabla^2 f_k)^{-1} \nabla f_k$.

Talyor series: $f(x_k + \varepsilon p) = f_k + \varepsilon p^T \nabla f_k + \frac{\varepsilon^2}{2} p^T \nabla^2 f_k \, p + O(\varepsilon^3)$



I want it to be a stationary point

$x_k \quad p \quad (x_k + \varepsilon p)$

If true $\nabla_p f(x_k + \varepsilon p) = 0$.

gradient w.r.t. p.

$$\nabla_p f(x_k + \varepsilon p) = \nabla_p \left(f_k + \varepsilon (\ ) + \varepsilon^2 (\ )\right) \ldots$$

$$\underbrace{0 = 0 + \varepsilon \nabla f_k} + \varepsilon^2 (\nabla^2 f) p$$

$$\Rightarrow \boxed{p = \frac{-1}{\varepsilon} (\nabla^2 f_k)^{-1} \nabla f_k}$$

You can reach minima in one shot

$$\rightarrow p \propto -(\nabla^2 f_k)^{-1} \nabla f_k.$$

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

$$\underbrace{f(x_k + p) - f(x_k)} = \nabla f_k^T p + \ldots$$

$$= \underbrace{-\nabla f_k^T (\nabla^2 f_k)^{-1} \nabla f_k} + \ldots$$

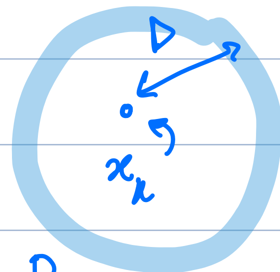
$(\nabla^2 f_k)^{-1}$ pos. def $\Rightarrow$ > 0

$$\Rightarrow f(x_k + p) < f(x_k).$$

↪ Trust region Methods

1) Construct a model m_k whose behavior matches that of f around x_k
2) Search for a minimiza in a region around x_k

$$\min_p \boxed{m_k(x_k + p)} \quad \text{where} \quad x_k + p \in T$$



Most common:

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

Either the Hessian or an approximation to it.