Line Search

Less computation per iteration ↑ ↓ Fewer Iterations

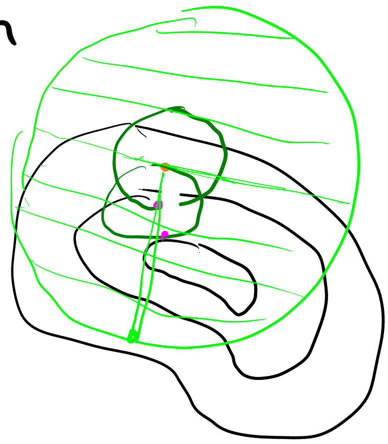
- Gradient
- Conj Grad
- Quasi-Newton
- Newton

Momentum

Adam

Trust Region

Trust Region



initialize δ, $\vec{x}^0$

loop

What is $\hat{f}$

how do we solve approximately

$$\vec{x}^{k+1} \leftarrow \underset{\vec{x}}{\arg\min}\ \hat{f}(\vec{x})$$

Linear

$$\text{subject to } \|\vec{x} - \vec{x}^k\| \le \delta$$

How do we adjust trust region

if $f(\vec{x}^{k+1})$ not near $\hat{f}(\vec{x}^{k+1})$

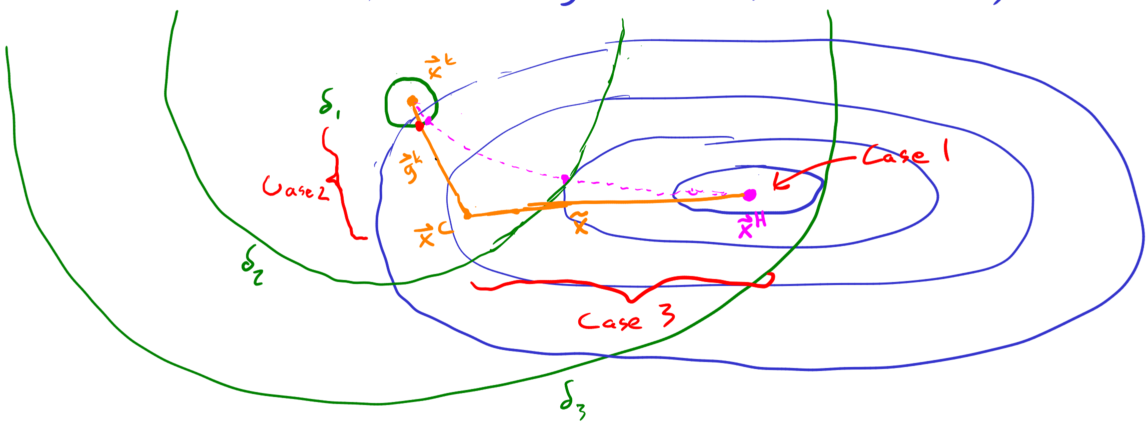
shrink δ

else

expand δ

end

$$\hat{f}^k(\vec{x}) = f^k + \vec{g}^k(\vec{x} - \vec{x}^k) + \frac{1}{2}(\vec{x} - \vec{x}^k)^T H^k (\vec{x} - \vec{x}^k)$$



$$\vec{x}^H \equiv \vec{x}^k - (H^k)^{-1}\vec{g}^k$$

Case 1 δ_3 →

If $\|\vec{x}^H - \vec{x}^k\| < \delta$

$$\vec{x}^{k+1} \leftarrow \vec{x}^H$$

$$\tau^c = \underset{\tau \in [0,\delta]}{\arg\min}\ \hat{f}(\vec{x}^k - \tau \vec{g}^k)$$

$$\vec{x}^c = x^k + \tau^c \vec{g}^k$$

$$\nabla_{\vec{g}^k}\hat{f}(\vec{x}) = \vec{g}^{k\,T}\nabla\hat{f}(x) = \vec{g}^{k\,T}\big(\vec{g}^k + H^k \underbrace{(\vec{x} - \vec{x}^k)}_{-\tau^c\vec{g}^k}\big) = 0$$

$$-\tau^c \vec{g}^{kT} H^k \vec{g}^k = -\vec{g}^{kT}\vec{g}^k$$

$$\tau^c = \frac{\vec{g}^{kT}\vec{g}^k}{\vec{g}^{kT} H^k \vec{g}^k}$$

$$\boxed{\vec{x}^c = \vec{x}^k - \frac{\vec{g}^{kT}\vec{g}^k}{\vec{g}^{kT} H^k \vec{g}^k}\vec{g}^k}$$

→ If $\|\vec{x}^c - \vec{x}\| \geq \delta$

Case 2 δ_1

$$\vec{x}^{k+1} \leftarrow \vec{x}^k + \frac{\delta}{\|\vec{x}^c - \vec{x}\|}(\vec{x}^c - \vec{x})$$

In between $\vec{x}^c$ and $\vec{x}^H$

First note that if $H \succ 0$, the function

$$\tilde{x}(\tau) = \vec{x}^c + \tau(\vec{x}^H - \vec{x}^c)$$

has the following properties

1) $\|\tilde{x}(\tau) - \vec{x}^k\|$ is an increasing function of τ for $\tau \in [0,1]$
2) $\hat{f}(\tilde{x}(\tau))$ is a decreasing function of τ for $\tau \in [0,1]$

N+W Lemma 4.2

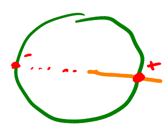
$\Rightarrow$ $\tilde{x}$ crosses constraint at only 1 point and that point is the minimum along $\tilde{x}$ subject to the constraint

$$\overbrace{\|\vec{x}^c + \tau(\vec{x}^H - \vec{x}^c) - \vec{x}^k\|}^{\|\tilde{x}(\tau) - \vec{x}^k\|}{}^2 = \delta^2 \longleftarrow \text{How to solve for } \tau$$

$$\sum_i \left(x_i^c + \tau(x_i^H - x_i^c) - x_i^k\right)^2 = \delta^2$$

$$\underbrace{\tau^2 \sum_i (x_i^H - x_i^c)^2}_{a} + \underbrace{2\tau \sum_i (x_i^c - x_i^k)(x_i^H - x_i^c)}_{b} + \underbrace{\sum_i (x_i^c - x_i)^2 - \delta^2}_{c} = 0$$

$$\tau = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$



δ_2

Case 3

$$\vec{x}^{k+1} \leftarrow \vec{x}^c + \tau(\vec{x}^H - \vec{x}^c)$$

How to adjust trust region

$$\eta = \frac{\text{actual improvement}}{\text{predicted improvement}} = \frac{f(\vec{x}^k) - f(\vec{x}^{k+1})}{f(\vec{x}^k) - \hat{f}(\vec{x}^{k+1})}$$

If $\eta < \eta_1$ (0.25) or $\eta < 0$

$\vec{x}^{k+1} \leftarrow \vec{x}^k$

$\delta \leftarrow \gamma_1 \delta$ $\quad\quad \gamma_1 \in (0,1)$

If $\eta > \eta_2$ (0.5)

$\delta \leftarrow \gamma_2 \delta$ $\quad\quad \gamma_2 \in (1,\infty)$

Theory

Global Convergence

Roughly, Better than Cauchy

$\eta_1 = 0 \Rightarrow \liminf_{k\to\infty} \|\vec{g}^k\| = 0$ — Theorem 4.5

$\eta_1 \in (0, \frac{1}{4}) \Rightarrow \lim \|\vec{g}^k\| = 0$ — Thm 4.6

Local Convergence Rate

Roughly, Better than Cauchy and steps are asymptotically similar to Newton

Thm 4.9

$\{\vec{x}^k\}$ converges superlinearly to $\vec{x}^*$