April 19, 2021

Gauss-Newton

$h : \mathbb{R}^n \to \mathbb{R}^m$

- $f(x) = \frac{1}{2}\|h(x)\|_2^2 \leftarrow \frac{1}{2}\sum h_i(x)^2$

multi-objective

$h_i(x)$ = error on i^{th} example

Taylor: $h(x+p) \approx \boxed{h(x) + J(x)p}$

$$J_{il}(x) = \frac{\partial h_i(x)}{\partial x_l} \qquad J \in \mathbb{R}^{m\times n}$$

$$x_{k+1} = x_k + p_k$$

$$p_k = \operatorname*{argmin}_{p\in\mathbb{R}^n} \|J_k p + h_k\|_2^2$$

$(m>n)$

Claim: p_k is descent dir (J_k full rank)

Proof: $$p_k = -\underbrace{(J_k^T J_k)^{-1}}_{\text{SPD}} \underbrace{J_k^T h_k}_{=\nabla f_k}$$

$$\frac{\partial f(x)}{\partial x_l} = \sum_{i=1}^m h_i(x) \frac{\partial h_i(x)}{\partial x_l}$$

$J_{il}(x)$

$$= \sum h_i(x) J_{il}(x)$$
$$= [J^T h(x)]_l$$

$$f(x_k+p_k) \approx f(x_k) \underbrace{- \nabla f(x_k)^T (J_k^T J_k)^{-1} \nabla f(x_k)}_{<0}$$

Is this just Newton?

$$p_k = -(J_k^T J_k) J_k^T h_k \quad \text{vs.} \quad p_k = -H_k^{-1} \nabla f_k$$

$\partial/\partial x_t$

$$= \sum_{i=1}^{m} \underbrace{\frac{\partial h_i(x)}{\partial x_t}}_{J_{it}} \underbrace{\frac{\partial h_i(x)}{\partial x_\ell}}_{J_{i\ell}} + h_i(x) \frac{\partial^2 h_i(x)}{\partial x_t \partial x_\ell}$$

$(J_k^T J_k)_{t\ell}$

- $f(x) = \frac{1}{2} \|h(x)\|_2^2$ near min $\Rightarrow h(x) \approx 0$
- $\Rightarrow J_k^T J_k \approx H_k$

$$s \cdot p_k'(0) + (s-1)\left(-H_k^{-1} p_k - p_k'(0)\right) \qquad 1 \leq s$$