CH 8 [CONT]

$$\boxed{\min_{\theta} \underbrace{\frac{1}{2}\|M(\theta)-y\|_2^2}_{=: f(\theta)}}$$

$M : \mathbb{R}^d \to \mathbb{R}^N$

$f : \mathbb{R}^d \to \mathbb{R}$

NEC. OPT. COND:

θ^* OPT $\Rightarrow$ $\nabla f(\theta^*) = 0$

$\nabla^2 f(\theta^*) \succeq 0$

SUFFICIENT OPTIMALITY CONDITIONS

θ^* LOCAL MINIMIZER $\Leftarrow$ $\nabla f(\theta^*) = 0$

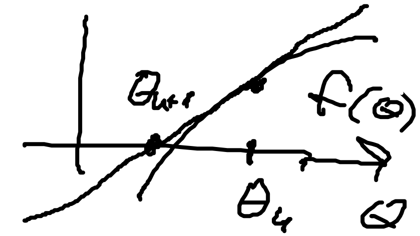
$\nabla^2 f(\theta^*) \succ 0$

---

f CONVEX $\Leftrightarrow$ $\nabla^2 f(\theta) \succeq 0$ FOR ALL θ

IF f CONVEX: $\boxed{\theta^* \text{ GLOBAL OPTIMIZER} \Leftrightarrow \nabla f(\theta^*) = 0}$

8.2 NEWTON'S METHOD



SOLVE $\boxed{\nabla f(\theta) = 0}$ ITERATIVELY.

$\theta_0, \theta_1, \ldots$ $\quad (\theta_i \in \mathbb{R}^d)$

θ_{k+1} FROM SOLUTION OF $\boxed{\underline{\nabla f(\theta_k)} + \underline{\nabla^2 f(\theta_k)} \underbrace{(\theta - \theta_k)}_{p_k} = 0}$

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

$$\theta_{k+1} = \theta_k + p_k$$

PROBLEMS, AMONG OTHER: ATTRACTED BY SADDLE ETC.

GOOD: QUADRATIC CONVERGENCE

COMPUTE $\nabla f(\theta)$ FOR $f(\theta) = \frac{1}{2} \| M(\theta) - y \|_2^2$

$$\frac{\partial}{\partial \theta_i} \left[\frac{1}{2} \left(M_k(\theta) - y_k \right)^2 \right] = \left(M_k(\theta) - y_k \right) \frac{\partial M_k(\theta)}{\partial \theta_i}$$

$$\nabla_\theta \left[\quad '' \quad \right] = \left(\quad '' \quad \right) \left(\frac{\partial M_k(\theta)}{\partial \theta} \right)^T$$

$$\nabla_\theta f(\theta) = \nabla_\theta \sum_{k=1}^{N} \frac{1}{2} \left(M_k(\theta) - y_k \right)^2 = \sum_{k=1}^{N} \left(\underline{M_k(\theta) - y_k} \right) \left(\frac{\partial M_k(\theta)}{\partial \theta} \right)^T$$

$$\boxed{= \underbrace{\frac{\partial M(\theta)}{\partial \theta}^T}_{\in \mathbb{R}^{d \times N}} \underbrace{\left(M(\theta) - y \right)}_{\in \mathbb{R}^N} = \underbrace{\nabla f(\theta)}_{\in \mathbb{R}^d}}$$

8.3 SPECIAL CASE: AFFINE MODEL

$$M(\theta) = b + A \cdot \theta \qquad A \in \mathbb{R}^{N \times d}$$

$$\nabla f(\theta) = A^T (A\theta + b - y) \qquad \frac{\partial M}{\partial \theta} = A$$

$$\nabla f(\theta^*) = 0 \iff (A^T A)\,\theta^* = A^T(y - b) \iff \theta^* = A^\dagger (y - b)$$

SECOND DERIVATIVES ?

$$\frac{\partial}{\partial \theta}\left(\nabla f(\theta)\right) = A^T A \succeq 0$$

$\forall z \in \mathbb{R}^d$

$z^T A^T A z$

$= \|Az\|_2^2 \geq 0$

f CONVEX : $\nabla f(\theta^*)$ DEFINES GLOBAL OPTIMIZER.

8.4 SECOND DERIVATIVE FOR NONLINEAR LEAST SQUARES

$$\frac{\partial}{\partial\theta}\nabla f(\theta) = \frac{\partial}{\partial\theta}\sum_{k=1}^{N}\left(M_k(\theta)-y_k\right)\left(\frac{\partial M_k(\theta)}{\partial\theta}\right)^T$$

$$= \sum_{k=1}^{N}\underbrace{\frac{\partial M_k}{\partial\theta}^T\frac{\partial M_k}{\partial\theta}}_{\in\mathbb{R}^{d\times d}} + \sum_{k=1}^{N}\underbrace{\left(M_k(\theta)-y_k\right)}_{\substack{\text{RESIDUALS}\\ \in\mathbb{R}}}\cdot\underbrace{\nabla_\theta^2 M_k(\theta)}_{\in\mathbb{R}^{d\times d}}$$

$$= \underbrace{\left(\frac{\partial M}{\partial\theta}(\theta)\right)^T\frac{\partial M}{\partial\theta}(\theta)}_{\succeq 0 \;\; (A^TA)} + \underbrace{E}_{\substack{\text{"ERROR MATRIX"}\\ \text{(ZERO FOR } y=M(\theta))}} = \nabla^2 f(\theta)$$

8.5 GAUSS-NEWTON METHOD

USE $\frac{\partial M}{\partial \theta}^T \frac{\partial M}{\partial \theta}$ INSTEAD OF $\nabla^2 f(\theta)$

$$\theta_{k+1} = \theta_k - \left(\frac{\partial M}{\partial \theta}(\theta_k)^T \underbrace{\frac{\partial M}{\partial \theta}(\theta_k)}_{=:A}\right)^{-1} \nabla f(\theta_k)$$

$$= \underline{\theta_k - (A^T A)^{-1} A^T \left(M(\theta_k) - y\right)}$$

$$\Longleftrightarrow$$

$$\boxed{\theta_{k+1} = \arg\min_{\theta} \frac{1}{2} \left\| M(\theta_k) - y + \frac{\partial M}{\partial \theta}(\theta_k)(\theta - \theta_k) \right\|_2^2}$$

PHILOSOPHY: LINEARIZE INSIDE NORM,
i.e. DISREGARD "ERROR MATRIX" IN NEWTON

$\boxed{\text{lsqnonlin}}$

IMPLEMENT VARIANT OF GAUSS-NEWTON.

LLS : $\arg\min_{\theta} \frac{1}{2} \|A\theta - y\|_2^2 = \underline{\underline{A^{\dagger} y}}$

$A^{\dagger} = (A^T A)^{-1} A^T$

NLS : USE G-N. $\Leftrightarrow$ ITERATIVE REFINEMENT WITH LLS

<u>USER</u> NEEDS TO GIVE Θ_0 "INITIAL GUESS"

8.6 SENSITIVITY W.R.T. MEASUREMENTS

FOR LLS: $\hat{\Theta}_{ML} = \Theta^* = A^+ y$

$$\mathrm{Cov}\,\hat{\Theta}_{ML} = A^+ \,\mathrm{Cov}\, y \,(A^+)^T$$

FOR NLS: $\hat{\Theta}_{ML} = \Theta^* = \arg\min_{\Theta} \frac{1}{2}\|M(\theta) - y\|_2^2$
$= \hat{\Theta}(y)$

FORTUNATELY: FOR $y + \Delta y$ WE GET

$$\hat{\Theta}(y+\Delta y) \approx \arg\min \frac{1}{2}\left\| \frac{\partial M}{\partial \theta}(\theta^*)(\theta - \theta^*) + M(\theta^*) - (y + \Delta y)\right\|_2^2$$

$$= \left(\frac{\partial M}{\partial \theta}(\theta^*)\right)^+ \Delta y$$

$$(\mathrm{Var}\ \hat{\theta}) \approx \frac{\partial m}{\partial \theta}(\theta^*)^{+} (\mathrm{Var}\ y) \left(\frac{\partial m}{\partial \theta}(\theta^*)\right)^{+T}$$

$$\frac{\partial m}{\partial \theta}^{+} = A^{+} = \underline{\underline{(A^T A)}}^{-1} A^T$$