CH 3 NEWTON-TYPE OPT. METHODS

$$\min \; f(x) \quad \text{s.t.} \quad g(x) = 0$$

$$\nabla \mathcal{L} = 0$$
$$g = 0$$

$$\begin{bmatrix} x_{k+1} \\ \lambda_{k+1} \end{bmatrix} = -\left(\begin{array}{c|c} B_k & \nabla g \\ \hline \nabla g^T & 0 \end{array}\right)^{-1} \begin{bmatrix} \nabla \mathcal{L} \\ g \end{bmatrix} + \begin{bmatrix} x_k \\ \lambda_k \end{bmatrix}$$

B_k

x_k, λ_k

$$\min_x \; \nabla f(x_k)^T x + \frac{1}{2}(x-x_k)^T \underbrace{\nabla^2 \mathcal{L}(\cdot)}_{B_k} (x-x_k)$$
$$\text{s.t.} \quad g(x_k) + \nabla g(x_k)^T (x - x_k) = 0$$

x_{k+1}, λ_{k+1}

"QUADRATIC MODEL INTERPRETATION"

1) EXACT NEWTON $B_k = \nabla^2 \mathcal{L}$: (QUADRATIC)

2) GAUSS-NEWTON (ONLY IF $f(x) = \frac{1}{2}\|R(x)\|_2^2$) : $B_k = \nabla R \, \nabla R^T \approx \nabla^2 \mathcal{L}$
(LINEAR)
(BETTER IF $\|R\|$ IS SMALLER)

3) ... BFGS (SUPERLINEAR)

3.1.4 HESSIAN UPDATES BY BFGS

IDEA: USE $\nabla_x \mathcal{L}$ FROM DIFFERENT ITERATIONS TO LEARN ABOUT $\nabla_x^2 \mathcal{L}$

$$\nabla_x^2 \mathcal{L}(\cdot)\cdot \underbrace{(x_{k+1}-x_k)}_{=:s_k} \approx \underbrace{\nabla_x \mathcal{L}(x_{k+1},\lambda_{k+1}) - \nabla_x \mathcal{L}(x_k,\underline{\underline{\lambda_{k+1}}})}_{=:y_k}$$

(AFTER SOLVING QP)

REQUIRE:

$$\boxed{B_{k+1}\cdot s_k = y_k}$$

SECANT EQUATION

AND THAT $B_{k+1} \approx B_k$ (NOT THROW OLD INFORMATION AWAY)

$$\boxed{B_{k+1} = B_k - \frac{(B_k s_k)(s_k^T B_k)}{s_k^T B_k s_k} + \frac{y_k y_k^T}{s_k^T y_k}}$$

BFGS-FORMULA

~~FULL RAN~~

UPDATE OF "RANK 2"

$$B_{k+1} s_k = B_k \cdot s_k - B_k s_k + y_k = y_k$$

$$y_k = \nabla f(x_{k+1}) + \nabla g(x_{k+1})\lambda_{k+1} - \nabla f(x_k) - \nabla g(x_k)\cdot\lambda_{k+1}$$

BFGS

$\boxed{D_{k+1}}$

NEW QP

$$B_k \approx \nabla^2 \mathcal{L}$$

$$\|w_{k+1} - w^*\| \leq \kappa_k \cdot \|w_k - w^*\| \quad \text{WITH } \kappa_k \to 0$$

"SUPERLINEAR"

3.2 LOCAL CONVERGENCE OF NEWTON-TYPE METHODS

THM 1:

$$F(w)=0 \qquad F:\mathbb{R}^n \to \mathbb{R}^n \qquad F(w^*)=0$$

$$\boxed{\boxed{w_{k+1} = w_k - M_k^{-1} F(w_k)}}$$

WITH $M_k \approx \frac{\partial F}{\partial w} =: J(w_k)$

IF

1. $$\boxed{\left\| M_k^{-1}\left(J(w_k)-J(w)\right)\right\| \leq \omega \|w_k - w\|}$$ WITH $\omega < \infty$

2. $$\boxed{\left\| M_k^{-1}\left(M_k - J(w_k)\right)\right\| \leq \kappa_k}$$ WITH $\kappa_k \leq \kappa < 1$

3. $$\boxed{\|w_0 - w^*\| \leq \frac{2}{\omega}(1-\kappa)}$$

THEN:

$$\boxed{\|w_{k+1}-w^*\| \leq \underbrace{\left(\kappa_k + \frac{\omega}{2}\|w_k - w^*\|\right)}_{\leq \kappa + \frac{\omega}{2}\|w_k - w^*\| < 1} \cdot \|w_k - w^*\|}$$

1) EXACT NEWTON ~~...~~ $K=0$

$$\|w_{k+1}-w^*\| \le \left(\|w_k-w^*\|\right)\|w_k-w^*\|$$

2) ~~B~~ GAUSS~~-N~~ NEWTON $K_k = K > 0$ $(K<1)$:

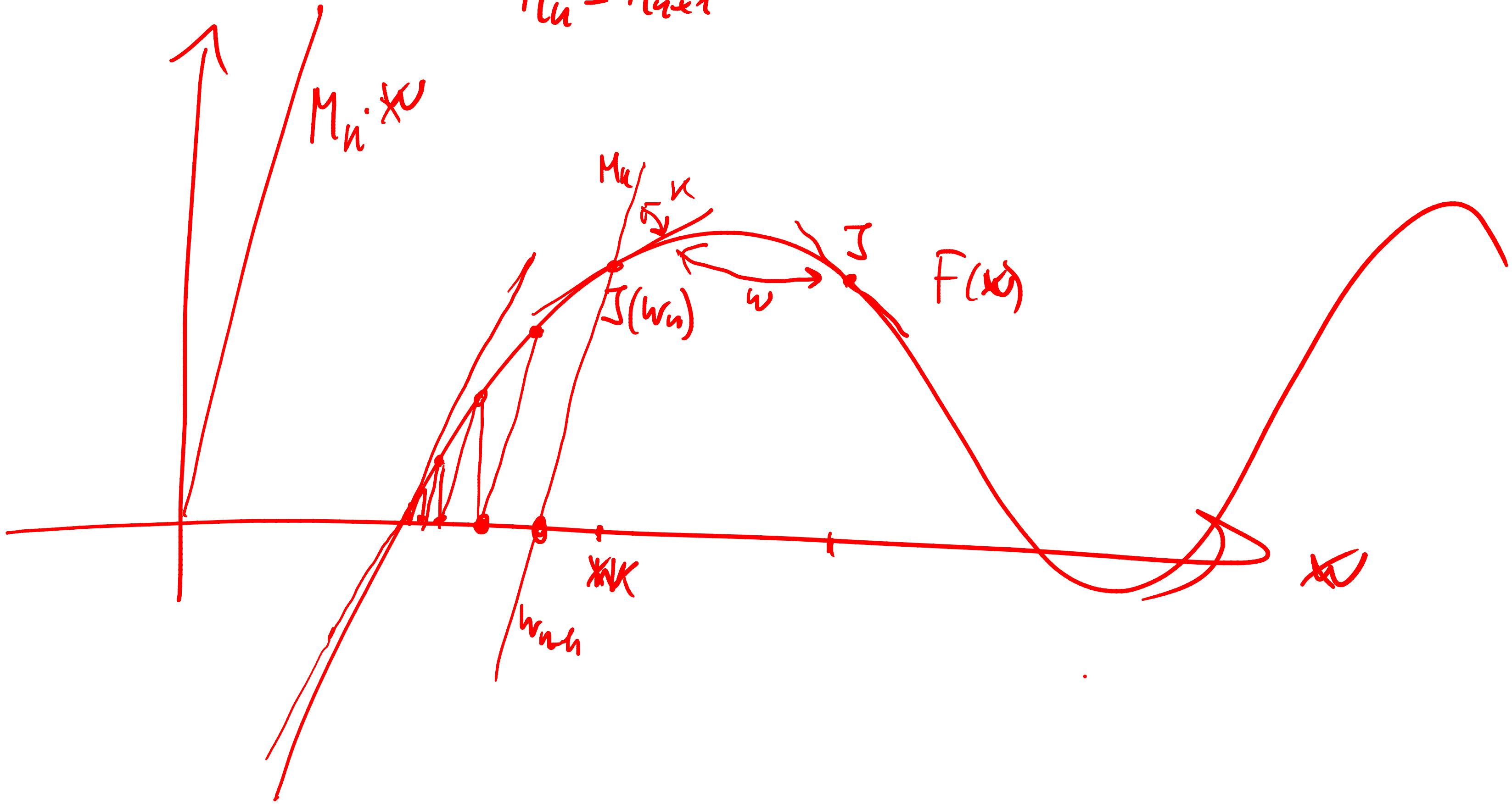
$$\underbrace{\|w_{k+1}-w^*\| \le K \cdot \|w_k-w^*\|}_{\text{LINEAR CONVERGENCE WITH RATE } K} + \underbrace{\frac{\omega}{2}\|w_k-w^*\|_2^2}_{0}$$

3) BFGS : $K_k \longrightarrow 0$ "SUPERLINEAR"

$$J(w_k) = \begin{bmatrix} \nabla^2 \mathcal{L} & \nabla g \\ \nabla g^T & 0 \end{bmatrix}$$

$$M_k = \begin{bmatrix} B_k & \nabla g \\ \nabla g^T & 0 \end{bmatrix} \qquad J - M_k = \left[\begin{array}{c|c} \nabla^2\mathcal{L} - B_k & 0 \\ \hline 0 & 0 \end{array}\right]$$

$$M_n = M_{n+1}$$



3.3 INEQUALITY CONSTR. OPTIMIZATION

$$\boxed{\begin{aligned} \min_x \quad & f(x) \\ \text{s.t.} \quad & g(x) = 0 \\ & h(x) \le 0 \end{aligned}}$$

NEWTON-TYPE (SOLVE KKT-CONDITION)

- NONLINEAR INTERIOR POINT (IP) (IPOPT)
- SEQUENTIAL QUADRATIC PROGRAMMING (SQP) (fmincon)

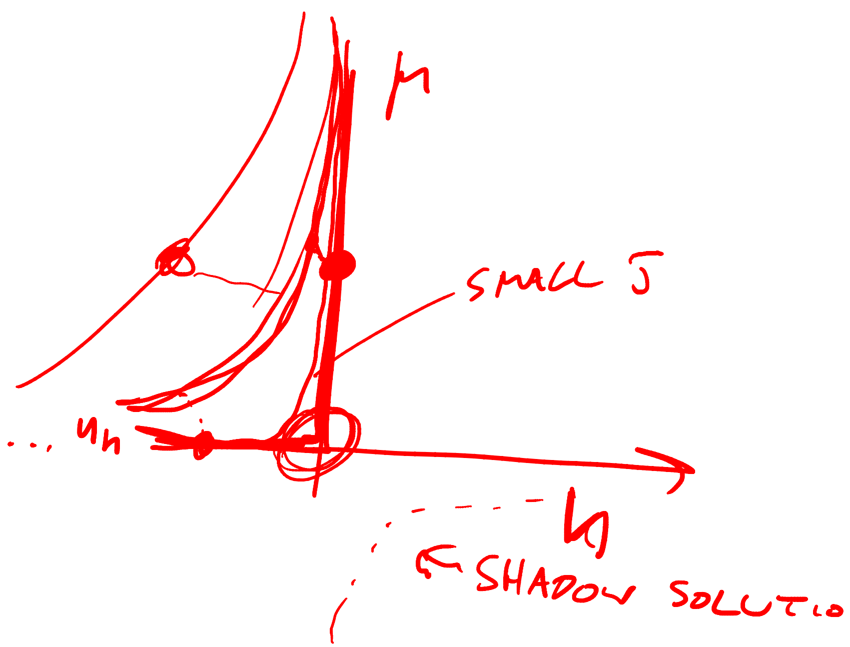
3.3.1 NONLIN. IP

$$\nabla \mathcal{L}(x^*, \lambda^*, \mu^*) = 0$$

$$g(x^*) = 0$$

$$\left.\begin{aligned} h_i(x^*) &\le 0 \\ \mu_i^* &\ge 0 \\ \mu_i^* h_i(x^*) &= 0 \end{aligned}\right\} \quad i = 1, \dots, n_h$$

$$\boxed{\mu_i^* h_i(x^*) + \tau = 0}$$



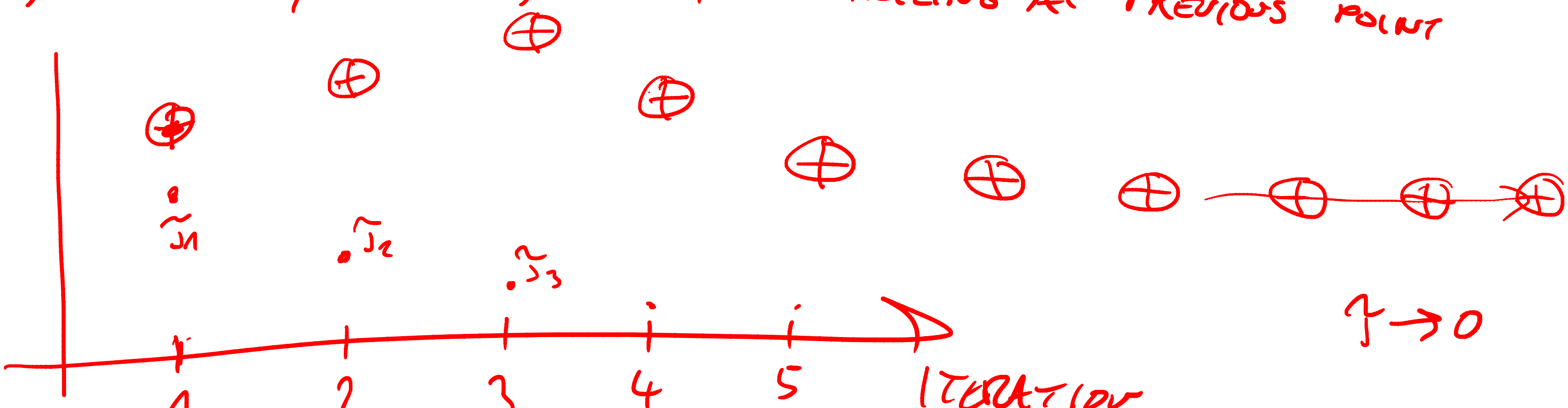
N-IP:

1) CHOOSE BIG $\tilde{s}$

2) SOLVE ROOT FINDING PROBLEM, GET (NEWTON-TYPE)

$$\begin{bmatrix} x^*(\tilde{s}) \\ \lambda^*(\tilde{s}) \\ \mu^*(\tilde{s}) \end{bmatrix}$$

3) REDUCE $\tilde{s}$, SOLVE 2) AGAIN, INITIALIZING AT PREVIOUS POINT



$\tilde{s} \to 0$

$$\mu_i \cdot h_i(x) + \tilde{\jmath} = 0 \iff \mu_i = -\frac{\tilde{\jmath}}{h_i(x)}$$

$$0 = \nabla \mathcal{L} = \nabla f + \nabla g \cdot \lambda + \underbrace{\sum \nabla h_i(x) \cdot \mu_i}_{\underbrace{-\sum \nabla h_i(x) \cdot \frac{\tilde{\jmath}}{h_i(x)}}_{= \sum \nabla_x \left(-\tilde{\jmath} \log\left(-h_i(x)\right)\right)}}$$

$$\frac{\partial \log(-h)}{\partial h} = -\frac{1}{h}$$

$$\frac{\partial \log(-h)}{\partial x} = \frac{\partial \log(-h)}{\partial h} \cdot \frac{\partial h}{\partial x} = -\frac{1}{h} \cdot \nabla h$$

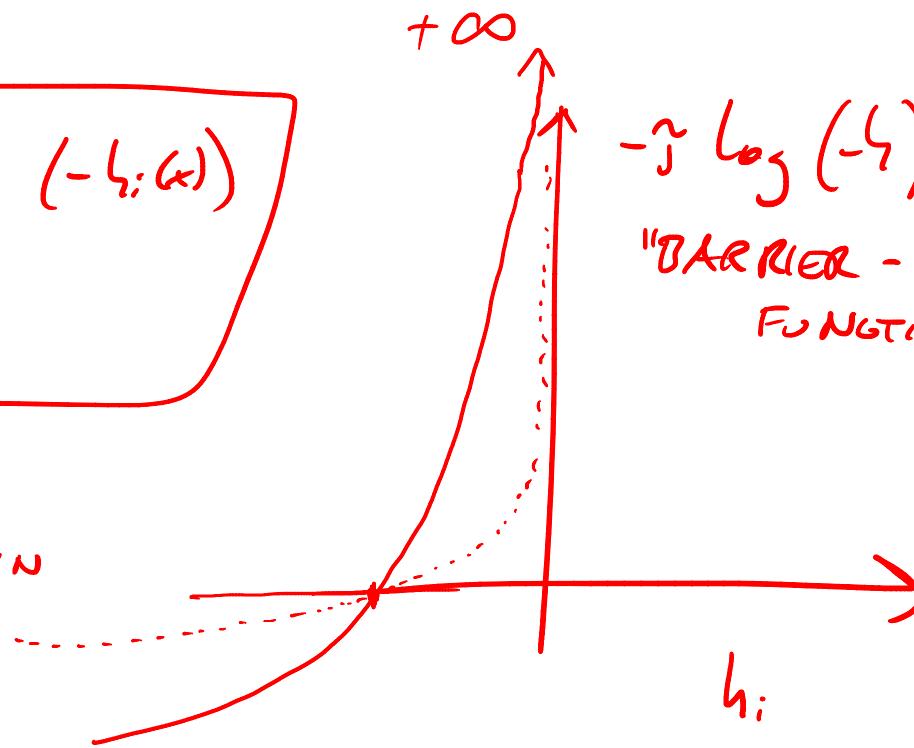
KKT-(μ)-

$$\underline{\nabla f(x)} + \nabla g(x)\cdot\lambda + \underline{\nabla_x\left(-\mu\sum\log\left(-h_i(x)\right)\right)} = 0$$

$= 0$

$g(x)$

$$\boxed{\begin{aligned} \min_x \quad & f(x) - \mu\sum\log\left(-h_i(x)\right) \\ \text{s.t.} \quad & g(x) = 0 \end{aligned}}$$



$-\mu\log(-h)$

"BARRIER - FUNCTIONS"

ITERATES WILL STAY IN <u>INTERIOR</u> OF FEASIBLE SET

3.3.2 SEQUENTIAL QUADRATIC PROGRAMMING (SQP)

USE QUADRATIC MODEL INTERPRETATION:

$$x_k,\ \lambda_k,\ \mu_k \longrightarrow \boxed{\begin{aligned} \min_x \quad & \nabla f(x_k)^T x + \frac{1}{2}(x-x_k)^T B_k (x-x_k) \\ \text{s.t.} \quad & g(x_k) + \nabla g(x_k)^T (x-x_k) = 0 \\ & h(x_k) + \nabla h(x_k)^T (x-x_k) \leq 0 \end{aligned}} \longrightarrow x_{k+1},\ \lambda_{k+1},\ \mu_{k+1}$$

→ SOLVE BY "QP-SOLVER"

"HARD WORK" DONE BY QP-SOLVER

(ACTIVESET / IP)

GRADIENT, ...

3.3.3 ✓

fmincon