Teorem 4.9 $\vec{x}_{k+1} = \vec{x}_k - \alpha_k \nabla f(x_k)$ (steepest descent)

Anta at vi bruker eksakt linjesøk på $f(\vec{x}) = \vec{x}^T A \vec{x}$, der A er positiv definit.

Da er $f(\vec{x}_{k+1}) \le m_A f(\vec{x}_k)$, der $m_A = \left(\frac{\lambda_1 - \lambda_n}{\lambda_1 + \lambda_n}\right)^2$

$\left(\text{cond}(A) = \frac{\lambda_1}{\lambda_n}\right)$

Lemma 4.10 Anta f konveks, og $\lambda_{min}(\nabla^2 f(\vec{x})) \ge m$
(anta f to ganger deriverbar)

(f^*: minimumsverdi) Da er $f(x) - f^* \le \frac{1}{2m} \|\nabla f(\vec{x})\|^2$

(x^* min: $f(x^*) = f^*$)

__Bevis__: Vi kan skrive:

$$f(\overset{\vec{x}+\vec{h}}{\vec{y}}) = f(\vec{x}) + \nabla f(\vec{x})^T \overbrace{(\vec{y} - \vec{x})}^{\vec{h}} + \frac{1}{2}(\vec{y} - \vec{x})^T \nabla^2 f(\overset{x+t\vec{h}}{\vec{z}})(\vec{y} - \vec{x})$$

← teorem 4.a

$$\ge \underbrace{f(\vec{x}) + \nabla f(\vec{x})^T(\vec{y} - \vec{x}) + \frac{1}{2} m \|\vec{y} - \vec{x}\|^2}_{\text{minimer i } \vec{y}\text{, ved å sette gradienten til } 0.}$$

(grad. til $x^T A x - b^T x$ er $A\vec{x} - \vec{b}$)

$$\nabla f(\vec{x}) + m(\vec{y} - \vec{x}) = \vec{0}$$

$$\vec{y} - x = -\frac{1}{m} \nabla f(\vec{x})$$

$$\vec{y} = y^* = \vec{x} - \frac{1}{m} \nabla f(\vec{x})$$

$$f(\vec{y}) \ge f(y^*) = f(x) + \nabla f(x)^T \left(-\frac{1}{m} \nabla f(x)\right) + \frac{1}{2} m \left\|-\frac{1}{m} \nabla f(x)\right\|^2$$

$$= f(x) - \frac{1}{m} \|\nabla f(x)\|^2 + \frac{1}{2m} \|\nabla f(x)\|^2$$

$$= f(x) - \frac{1}{2m} \|\nabla f(x)\|^2$$

$\vec{y} = x^*$

$$f^* \ge f(x) - \frac{1}{2m} \|\nabla f(x)\|^2$$

$$f(\vec{x}) - f^* \le \frac{1}{2m} \|\nabla f(x)\|^2$$

$$\|A\|_2 = \max_{\|x\|=1} \|Ax\|$$

Teorem 4.11 Anta f konveks, to ganger kont. deriverbar, og

(i) $\lambda_{min}(D^2 f(\vec{x})) \geq m$, alle $\vec{x}$

(ii) $\|D^2 f(x) - D^2 f(y)\|_2 \leq L\|x-y\|$ (finnes en slik L)

Anta også f har min x^*.

Da vil Newtons metode konvergere, og $x_k \to x^*$ med kvadratisk konvergenshastighet.

Bevis: Vi skriver $f^* = f(x^*)$

Første iterasjoner av Newton (damped Newton phase). Viser

Lemma 4.12: For enhver η finnes det en $\gamma > 0$ s.a. hvis $\|Df(x_k)\| \geq \eta$ så er $f(\vec{x}_{k+1}) \leq f(\vec{x}_k) - \gamma$

Lemma 4.13 Finnes η med $0 < \eta < \frac{m^2}{L}$ s.a. hvis $\frac{L}{m^2} < \frac{1}{\eta}$

$\|Df(x_k)\| < \eta$, så vil $\alpha_k = 1$ tilfredsstille stoppbetingelsen i backtracking line search,

og $\frac{L}{2m^2}\|Df(x_{k+1})\| \leq \left(\frac{L}{2m^2}\|Df(x_k)\|\right)^2$ $(\mu_{l+1} \leq \mu_l^2)$

Bevis teorem 4.11: Finn η fra lemma 4.13

Finn så γ for denne η (lemma 4.12), kjør iterasjoner helt til $\|Df(x_k)\| < \eta$ (endelig mange)

kan nå anta $\|Df(x_k)\| \leq \eta$

Lemma 4.13: $\|Df(x_{k+1})\| \leq \frac{2m^2}{L}\left(\frac{L}{2m^2}\|Df(x_k)\|\right)^2 = \frac{L}{2m^2}\|Df(x_k)\|^2$

$$\leq \frac{L}{2m^2}\eta^2 < \frac{1}{2}\frac{1}{\eta}\eta^2 = \frac{\eta}{2} < \eta$$

$\Rightarrow$ alle senere gradienter er også $\leq \eta$ (i lengde)

Sett $\mu_l = \frac{L}{2m^2}\|Df(x_l)\|$, $l \geq k$ $\qquad \mu_{l+1} \leq \mu_l^2 \leq (\mu_{l-1}^2)^2 = \mu_{l-1}^{2^2}$

$\leq \frac{1}{2}\frac{1}{\eta}\eta = \frac{1}{2}$ $\qquad \leq \dots \leq \mu_k^{2^{l-k}} \leq \left(\frac{1}{2}\right)^{2^{l-k}}$

Lemma 4.10: $f(x_l) - f^* \leq \frac{1}{2m}\|Df(x_l)\|^2 = \frac{1}{2m}\frac{4m^4}{L^2}\left(\frac{L}{2m^2}\|Df(x_l)\|\right)^2$

$$= \frac{2m^3}{L^2}\mu_l^2 \leq \frac{2m^3}{L^2}\mu_k^{2^{l-k-1}} \to 0$$

Kap. 5 $\min f(x)$

$$h_i(x) = 0 \quad (1 \le i \le m)$$
$$g_j(x) \le 0 \quad (1 \le j \le r)$$

5.1 Kun likhetsbetingelser: $\min f(x)$, $h_i(x) = 0$ $(1 \le i \le m)$ $\nearrow$ $\vec{H}(\vec{x}) = \vec{0}$

$$\vec{H}(\vec{x}) = (h_1(\vec{x}), \ldots, h_m(\vec{x})) \qquad H: \mathbb{R}^n \to \mathbb{R}^m$$

$\vec{x}^*$ regulær: alle $\nabla h_i(\vec{x})$ er lineært uavhengige.

Antar alltid f, h_i, g_j kont. deriverbare.

Teorem 5.1 Anta at $\vec{x}^*$ lokalt min, og regulært.

1. Da finnes det en vektor $\lambda^* = (\lambda_1, \ldots, \lambda_m)$ s.a.
$$\nabla f(x^*) + \sum_{i=1}^m \lambda_i \nabla h_i(x^*) = \vec{0}$$
2. Hvis f, h_i er to ganger kont. deriverbare. Da er
$$\vec{h}^T \left(\nabla^2 f(x^*) + \sum_{i=1}^m \lambda_i^* \nabla^2 h_i(x^*)\right)\vec{h} \ge 0, \text{ alle } h \in T(x^*),$$
der $T(x^*) = \{h \in \mathbb{R}^n : \underbrace{\nabla h_i(\vec{x}^*) \cdot \vec{h} = 0}_{\text{retningsderivert}}, \text{ alle } 1 \le i \le m\}$

λ_i^*: Lagrange multiplikatorer

Lagrangefunksjonen: $L(x, \lambda) = f(\vec{x}) + \sum_{i=1}^m \lambda_i h_i(\vec{x}) = f(\vec{x}) + \lambda^T H(\vec{x})$

Med denne får vi $\nabla_x L(x, \lambda) = \nabla f(x) + \sum_{i=1}^m \lambda_i \nabla h_i(\vec{x}) = \vec{0}$

$$\nabla_\lambda L(x, \lambda) = H(\vec{x}) = \vec{0}$$

Betingelsene i teoremet $\Longleftrightarrow$ $\nabla L(x, \lambda) = \vec{0}$

Tolkning $T(x^*)$: retninger $\vec{h}$ som vi kan gå i som ikke bryter med betingelsene $h_i(\vec{x}) = 0$.

Bevis: Definer $F^k(\vec{x}) = f(\vec{x}) + \underbrace{\frac{k}{2}\|H(\vec{x})\|^2} + \frac{\alpha}{2}\|x - x^*\|^2$

La x^k være min. for F^k på $\bar{B}(x^*, \varepsilon)$ (problem uten betingelser)

$F^k(x^k) = f(x^k) + \frac{k}{2}\|H(x^k)\|^2 + \frac{\alpha}{2}\|x^k - x^*\|^2$

$\leq F^k(x^*) = f(x^*)$ $(H(x^*) = 0,\ x^* - x^* = 0)$

La $k \to \infty$: ser at $\lim_{k\to\infty} \|H(x^k)\| = 0$

For enhver grense $\bar{x}$ for x^k, så vil $H(\bar{x}) = \vec{0}$

sett inn $\bar{x}$

$F^k(\bar{x}) = f(\bar{x}) + \frac{\alpha}{2}\|\bar{x} - x^*\|^2 \leq f(x^*) \leq f(\bar{x})$

Derfor: $\bar{x} = x^*$, slik at $x^k \to x^*$

x^k er derfor et ubetinget minimum for F^k.

Teorem 4.1 sier da at $\nabla F^k(x^k) = \vec{0}$.

$$\nabla f(x^k) + k\, H'(x^k)^T H(x^k) + \alpha(x^k - x^*) = 0$$

$\Rightarrow (H'(x^*)(H'(x^*))^T)$ inverterbar

x^* regulær $\Rightarrow H'(x^*)$ har full rang (alle rader er lin. uavh.)

$\Rightarrow H'(x^k) H'(x^k)^T$ inverterbar for store k

$H'(x^k) \to H'(x^*)$

ganger med $(H'(x_k)H'(x_k)^T)^{-1} H'(x_k)$ på begge sider..

$$\underbrace{k H(\vec{x})} = -(H'(x_k)H'(x_k)^T)^{-1} H'(x_k)(\nabla f + \alpha(x_k - x^*))$$

$k \to \infty$

går mot en grense λ^*. setter inn $\lambda^* = \lim_{k\to\infty} k H(\vec{x})$

i ligningen over: $\nabla f(x^*) + H'(x^*)^T \lambda^* = \vec{0}$.