Kap. 4 i del III

Ubetinget optimering.

Teorem 4.1

Anta $f: \mathbb{R}^n \to \mathbb{R}$ kont. part. der. og anta at x^* er lokalt min.

1. Da er $\nabla f(x^*) = \vec{0}$ (førsteordens betingelse)

2. Hvis f også har 2. ordens kont. part. der., så er $\nabla^2 f(x^*)$ pos. semidefinit. (andreordens betingelse)

Bevis: 1. Anta $\nabla f(x^*) \neq \vec{0}$. Sett $h = -\alpha \nabla f(x^*)$, $\alpha > 0$

$\Rightarrow \nabla f(x^*)^T \vec{h} = -\alpha \nabla f(x^*)^T \nabla f(x^*) = -\alpha \|\nabla f(x^*)\|^2 < 0$

siden $\nabla f(x^*)$ er kontinuerlig så vil da $\nabla f(\vec{x})^T \vec{h} < 0$ når $\vec{x}$ er nær $\vec{x}^*$

Taylor: $f(\vec{x}^* + \vec{h}) = f(\vec{x}^*) + \underbrace{\nabla f(x^* + th)^T}_{\text{velg } t \text{ så liten at ligger i omegnen over.}} \vec{h} < f(x^*)$ for en $t \in [0,1]$

$\Rightarrow x^*$ ikke lokalt min, motsigelse.

2. $f(x^* + \vec{h}) = f(x^*) + \underbrace{\nabla f(x^*)^T}_{0} \vec{h} + \frac{1}{2} \vec{h}^T \nabla^2 f(x^* + t\vec{h}) \vec{h}$, $t \in [0,1]$

$= f(x^*) + \frac{1}{2} \vec{h}^T \nabla^2 f(x^* + th) \vec{h}$

hvis $\nabla^2 f(x^*)$ ikke er pos semidef., så finnes $\vec{h}$ s.a. $\vec{h}^T \nabla^2 f(x^*) \vec{h} < 0$

$\Rightarrow$ pga kontinuitet: $\vec{h}^T \nabla^2 f(x) \vec{h} < 0$ for x nær x^*.

for $\vec{h}$ liten $\Rightarrow x^* + th$ nær $\vec{x}^*$

$f(x^*) + \frac{1}{2} h^T \nabla^2 f(x^* + th) \vec{h} < f(x^*)$

$\Rightarrow x^*$ ikke lokalt min.

Sadelpunkt: stasjonært ($\nabla f(x^*) = \vec{0}$), men hverken min/maks.

Eks. $f(\vec{x}) = \frac{1}{2} \vec{x}^T A \vec{x} - b^T \vec{x}$

$\nabla f = \vec{0} \Leftrightarrow A\vec{x} - \vec{b} = 0 \Leftrightarrow A\vec{x} = \vec{b}$

$\nabla^2 f = A$

Lemma 4.3 Anta at A er symmetrisk, og at λ_n = minste egenverdi i A er >0

Da er $\vec{h}^T A \vec{h} \geq \lambda_n \|\vec{h}\|^2$

Bevis: A symmetrisk $\Rightarrow$ A ortogonaldiagonaliserbar

$A = PDP^T$ $\|y\| = \|P^T h\| = \|\vec{h}\|$ siden $\vec{h}$ er ortogonal.

$$\vec{h}^T A \vec{h} = \underbrace{\vec{h}^T P}_{\vec{y}^T} D \underbrace{P^T \vec{h}}_{\vec{y}} = \vec{y}^T D \vec{y} = \sum_{j=1}^n \lambda_j y_j^2 \geq \sum_{j=1}^n \lambda_n y_j^2$$

$$= \lambda_n \|y\|^2 = \lambda_n \|\vec{h}\|^2$$

Teorem 4.4 motsatt vei:

Anta f har kont. 2 ordens. part.-der

Hvis $\nabla f(x^*) = \vec{0}$, og $\nabla^2 f(x^*)$ pos definitt:

Da er x^* lokalt min.

Bevis: $f(x^* + \vec{h}) = f(x^*) + \underbrace{\nabla f(x^*)}_{0}{}^T \vec{h} + \frac{1}{2} \vec{h}^T \nabla^2 f(x^*) \vec{h} + \varepsilon(\vec{h}) \|\vec{h}\|^2$

der $\varepsilon(\vec{h}) \to 0$ når $\vec{h} \to 0$

$$\geq f(x^*) + \frac{1}{2}\lambda_n \|\vec{h}\|^2 + \varepsilon(\vec{h}) \|\vec{h}\|^2$$

Velg $\vec{h}$ så liten at $|\varepsilon(\vec{h})| < \frac{\lambda_n}{4}$ (kan gjøres siden $\varepsilon(\vec{h}) \to 0$)

$$\frac{f(x^* + \vec{h}) - f(x^*)}{\|\vec{h}\|^2} \geq \frac{1}{2}\lambda_n + \varepsilon(\vec{h}) \geq \frac{1}{4}\lambda_n > 0 \Rightarrow f(x^* + \vec{h}) > f(x^*).$$

$\Rightarrow x^*$ lokalt min.

Teorem 4,5

1. Anta f konveks lokalt min $\Leftrightarrow$ globalt min.
2. Hvis f også er deriverbar lokalt min $\Leftrightarrow$ stasjonært punkt.

<u>Bevis</u>: 1. Anta $\vec{x}_1$ lokalt min.

Hvis $\vec{x}_1$ <u>ikke</u> er globalt min så fennes en $\vec{x}_2$ slik at $f(\vec{x}_2)<f(\vec{x}_1)$

$$f((1-\lambda)\vec{x}_1+\lambda\vec{x}_2) \le (1-\lambda)f(\vec{x}_1)+\lambda f(\vec{x}_2)$$
$$< (1-\lambda)f(\vec{x}_1)+\lambda f(\vec{x}_1) = f(\vec{x}_1). \quad (\text{alle } \lambda\in(0,1))$$

velg λ nær 0 : $(1-\lambda)\vec{x}_1+\lambda\vec{x}_2$ nær $\vec{x}_1$, og $f((1-\lambda)\vec{x}_1+\lambda\vec{x}_2)<f(\vec{x}_1)$

$\Rightarrow \vec{x}_1$ kan umulig være lokalt min. $\Rightarrow \vec{x}_1$ er globalt min.

2. Anta f konveks og deriverbar.
Anta at x^* er stasjonært, så $\nabla f(x^*)=\vec{0}$

Teorem 2.10 : $f(\vec{x}) \geq f(x^*) + \nabla f(x^*)(x-x^*) = f(x^*)$

$\Rightarrow x^*$ lokalt min.

4.2 Metoder

Iterative metoder for å finne min.

Linjesøkemetoder: Velg søkeretning d_k, deretter en steglengde α_k, som sier hvor langt vi skal følge søkeretningen.

$$X_{k+1} = X_k + \alpha_k d_k$$

metoder for å finne steglengde:

1. eksakte linjesøk: Finn den $\alpha_k > 0$ som gir minst verdi for $g(\alpha_k) = f(X_k + \alpha_k d_k)$

Valg av søkeretning: $d_k = -\nabla f(\vec{X}_k)$. Kalles steepest descent.

$$\Updownarrow$$

$$\vec{X}_{k+1} = \vec{X}_k - \alpha_k \nabla f(\vec{X}_k)$$

f avtar raskest i denne retningen:

$$f(\vec{x}+\vec{h}) = f(\vec{x}) + \underbrace{\nabla f(\vec{x})^T \vec{h}} + \dots \overset{\vec{h} = -\alpha_k \nabla f(x_k)}{\geq} f(\vec{x}) - \alpha_k \|\nabla f(\vec{x})\|^2$$

Vi har at $\nabla f(x)^T \vec{h} \geq -\|\nabla f(x)\| \, \|\vec{h}\|$ $\quad (|\nabla f(\vec{x}) \cdot \vec{h}| \leq \|\nabla f(\vec{x})\| \, \|\vec{h}\|$.

(Cauchy Schwarz)

med $\vec{h} = -\alpha_k \nabla f(x)$ fås her likhet, dette gir største mulige reduksjon i funksjonsverdi

$\vec{h}$ kalles en "descent direction" hvis $\nabla f(\vec{x})^T \vec{h} < 0$ $\quad$ (i $\vec{x}$).

Newtons metode $\quad$ min $f \iff \nabla f = \vec{0}$

Newtons metode: $X_{k+1} = X_k - \alpha_k \left(\nabla^2 f(X_k)\right)^{-1} \nabla f(X_k)$. $\quad \alpha_k = 1$: Fullt Newtonsteg.

(Newtons metode utledes ved å minimere Taylortilnærming

$$f(X_k + h) \approx \underbrace{f(X_k) + \nabla f(X_k)^T \vec{h} + \tfrac{1}{2} \vec{h}^T \nabla^2 f(X_k) \vec{h}}$$

gradient: $\quad \nabla^2 f(X_k) \vec{h} + \nabla f(X_k) = 0$

$$\vec{h} = -\left(\nabla^2 f(X_k)\right)^{-1} \nabla f(X_k).$$

andre regel for valg av steglengde:

Def. 4.6 (backtracking line search)

Anta gitt $s \le 1$, $0 < \beta < 1$, $0 < \overset{10^{-3}}{\overset{\parallel}{\sigma}} < 1$

Vi setter $m_k = \min \{ m : m \geq 0,\ f(\vec{x}_k) - f(\vec{x}_k + \underbrace{\beta^m s}_{\text{steglengde som vi prøver}} \vec{d}_k) \geq -\sigma \beta^m s \underbrace{\nabla f(x_k)^T \vec{d}_k} \}$

og vi velger steglengde $\alpha_k = \beta^{m_k} s$

prøver steglengde $\beta^0 s$, $\beta^1 s$, $\beta^2 s$, $\vdots$

stoppbetingelse

reduksjon vi får til i funksjonsverdi

$\boxed{\nabla f(x_k)^T \vec{h}}$

$\vec{h} = \beta^m s\, d_k$

Teorem 4.8 Anta $x_{k+1} = x_k + \alpha_k d_k$, og at vi velger α_k ved backtracking linesearch.

Anta videre at $\limsup\limits_{k \to 0} \nabla f(x_{s_k})^T d_{k_s} < 0$ for alle delfølger x_{s_k} som konv. mot et ikke stasjonært punkt.

Da vil x_k konv mot et stasjonært punkt.