## Föreläsning 8: Olinjär optimering utan bivillkor

1. Olinjära optimeringsproblem utan bivillkor
2. Optimalitetsvillkor
3. Optimeringsalgoritmer
   - Gradientmetoden
   - Newtons metod
4. Olinjär minsta kvadratskattning



## Lokala och globala minpunkter

$$\begin{aligned} \text{minimera} \quad & f(\mathbf{x}) \\ \text{då} \quad & \mathbf{x} \in \mathbf{R}^n \end{aligned}$$

**Definition 1.** *$\hat{\mathbf{x}} \in \mathbf{R}^n$ är en lokal minpunkt till funktionen $f$ om det finns $\delta > 0$ så att*

$$f(\hat{\mathbf{x}}) \leq f(\mathbf{x}), \ \forall \mathbf{x} \in \mathbf{R}^n \text{ så att } \|\mathbf{x} - \hat{\mathbf{x}}\| \leq \delta.$$

**Definition 2.** *$\hat{\mathbf{x}} \in \mathbf{R}^n$ är en global minpunkt till funktionen $f$ om*

$$f(\hat{\mathbf{x}}) \leq f(\mathbf{x}), \ \forall \mathbf{x} \in \mathbf{R}^n$$



## Derivator

Gradienten till $f$ i punkten $\mathbf{x}$ definieras som radvektorn

$$\nabla f(\mathbf{x}) = \begin{bmatrix} \frac{\partial f}{\partial x_1}(\mathbf{x}) & \dots & \frac{\partial f}{\partial x_n}(\mathbf{x}) \end{bmatrix}$$

Hessianen till $f$ i punkten $\mathbf{x}$ definieras som den symmetriska $n \times n$ matrisen

$$\mathbf{F}(\mathbf{x}) = \nabla^2 f(\mathbf{x}) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2}(\mathbf{x}) & \dots & \frac{\partial^2 f}{\partial x_1 \partial x_n}(\mathbf{x}) \\ \vdots & & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1}(\mathbf{x}) & \dots & \frac{\partial^2 f}{\partial x_n^2}(\mathbf{x}) \end{bmatrix}$$

Vi antar nedan att $f$ är två gånger kontinuerligt deriverbar.



## Riktningsderivatan

Betrakta funktionen $f$ i punkten $\mathbf{x}$ och riktningen $\mathbf{d}$, och låt $F_{\mathbf{d}}(\alpha) = f(\mathbf{x} + \alpha\mathbf{d})$. Detta är en funktion av <u>en</u> variabel; skalären $\alpha$.

$$\begin{aligned} F'_{\mathbf{d}}(\alpha) &= \lim_{h \to 0} \frac{F_{\mathbf{d}}(\alpha + h) - F_{\mathbf{d}}(\alpha)}{h} = \lim_{h \to 0} \frac{f(\mathbf{x} + \alpha\mathbf{d} + h\mathbf{d}) - f(\mathbf{x} + \alpha\mathbf{d})}{h} \\ &= \lim_{h \to 0} \frac{f(\mathbf{x} + \alpha\mathbf{d}) + h\nabla f(\mathbf{x} + \alpha\mathbf{d})\mathbf{d} + \frac{1}{2}h^2\mathbf{d}^{\mathsf{T}}\nabla^2 f(\xi)\mathbf{d} - f(\mathbf{x} + \alpha\mathbf{d})}{h} \\ &= \nabla f(\mathbf{x} + \alpha\mathbf{d})\mathbf{d} \end{aligned}$$

Speciellt gäller att $F'_{\mathbf{d}}(0) = \nabla f(\mathbf{x})\mathbf{d}$ är riktningsderivatan för $f$ i punkten $\mathbf{x}$ och riktningen $\mathbf{d}$.

## Avtaganderiktningar och riktningsderivator

**Definition 3.** $\mathbf{d}$ *är en avtaganderiktning till* $f$ *i punkten* $\mathbf{x}$ *om det finns ett* $\epsilon > 0$ *sådant att* $f(\mathbf{x} + t\mathbf{d}) < f(\mathbf{x})$ *för alla* $t \in (0, \epsilon)$.

Avtaganderiktningar kan karaktäriseras med hjälp av riktningsderivatan:

**Lemma 1.** *Om* $\nabla f(\mathbf{x})\mathbf{d} < 0$ *så är* $\mathbf{d}$ *en avtaganderiktning till* $f$ *i* $\mathbf{x}$.

Om det inte finns några avtaganderiktningar till $f$ i punkten $\mathbf{x}$ så måste det alltså gälla att $\nabla f(\mathbf{x})\mathbf{d} \geq 0$ för alla $\mathbf{d}$, d.v.s. att $\nabla f(\mathbf{x}) = 0$.



## Optimalitetsvillkor

**Sats 1 (Första ordningens nödvändiga villkor).**
*Om* $\hat{\mathbf{x}}$ *är en lokal minpunkt till* $f$ *så är* $\nabla f(\hat{\mathbf{x}}) = \mathbf{0}^\mathsf{T}$.

**Sats 2 (Andra ordningens nödvändiga villkor).**
*Om* $\hat{\mathbf{x}}$ *är en lokal minpunkt till* $f$ *så är* $\nabla f(\hat{\mathbf{x}}) = \mathbf{0}^\mathsf{T}$ *och* $\mathbf{F}(\hat{\mathbf{x}}) \geq \mathbf{0}$ *(positivt semidefinit),*

**Sats 3 (Andra ordningens tillräckliga villkor).**
*Om* $\nabla f(\hat{\mathbf{x}}) = \mathbf{0}^\mathsf{T}$ *och* $\mathbf{F}(\hat{\mathbf{x}}) > \mathbf{0}$ *(positivt definit) då är* $\hat{\mathbf{x}}$ *en lokal minpunkt.*



## Optimeringsalgoritmer

Vi betraktar två iterativa metoder för minimering av flervariabla funktioner.

1. Gradient metoden (steepest descent/brantaste lutningsmetoden)
   - Sökriktningen bestäms med hjälp av gradienten, dvs första ordningens information.
2. Newtons metod.
   - Använder andra ordningens information, dvs gradienten och Hessianen, för att bestämma sökriktningen.



## Gradientmetoden

Idé: Sök i riktningen där funktionen avtar mest $\Rightarrow$ sökriktningen ges av $\mathbf{d}^{(k)} = -\nabla f(\mathbf{x}^{(k)})^\mathsf{T}$.

**Algoritm:**

(0) Bestäm startpunkt $\mathbf{x}^{(0)}$ och låt $k = 0$.

$(i)$ Låt $\mathbf{d}^{(k)} = -\nabla f(\mathbf{x}^{(k)})^\mathsf{T}$.

$(ii)$ Kontrollera avbrottskriterium: Om $|\nabla f(\mathbf{x}^{(k)})| \leq \epsilon$ så avslutas sökningen.

$(iii)$ Utför linjesökningen

$$t^{(k)} = \arg\min_{t \geq 0} f(\mathbf{x}^{(k)} + t\mathbf{d}^{(k)})$$

och låt $\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + t^{(k)}\mathbf{d}^{(k)}$

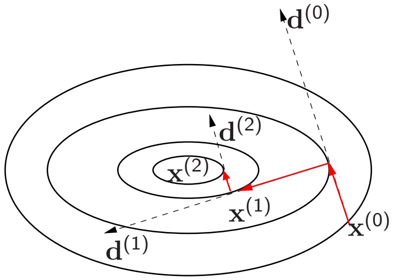
$(iv)$ Uppdatera $k := k + 1$ och gå till steg $(i)$.

## Kommentarer till gradientmetoden

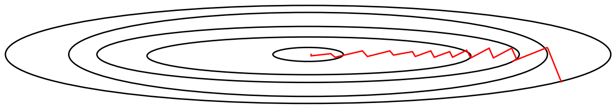
Om exakt linjesökning utförs får man $\mathbf{d}^{(k+1)} \perp \mathbf{d}^{(k)}$.

**Bevis:** Om $\varphi(t) = f(\mathbf{x}^{(k)} + t\mathbf{d}^{(k)})$ så får man

$$0 = \varphi'(t^{(k)}) = \nabla f(\mathbf{x}^{(k)} + t^{(k)}\mathbf{d}^{(k)})^{\mathsf{T}}\mathbf{d}^{(k)} = -(\mathbf{d}^{(k+1)})^{\mathsf{T}}\mathbf{d}^{(k)}$$




Ortogonala sökriktningar



Detta kan leda till långsam konvergens



## Linjesökning

Vi låter $\varphi(t) = f(\mathbf{x}^{(k)} + t\mathbf{d}^{(k)})$. Linjesökningen svarar mot att lösa

$$\min_{t \geq 0} \varphi(t).$$

Linjesökningen måste i regel göras approximativt. Vi presenterar två metoder

1. Intervallhalvering
2. Newtons metod



### Intervallhalvering

Vid intervallhalvering används första ordningens information för att söka efter en punkt där $\varphi'(t^{(k)}) \approx 0$.

**Algoritm:**

(0) Låt $\alpha_0 = 0$ och $\beta_0 = t_{\max}$, där $t_{\max}$ är en övre gräns sådan att $\varphi'(t_{\max}) > 0$.

$(i)$ $t_k = \dfrac{\alpha_k + \beta_k}{2}$.

$(ii)$ Om $|\varphi'(t_k)| \leq \epsilon$ så $t^{(k)} = t_k$. Klart!

$(iii)$ Om $\varphi'(t_k) < 0$ så $\alpha_{k+1} = t_k$ och $\beta_{k+1} = \beta_k$.
Om $\varphi'(t_k) \geq 0$ så $\alpha_{k+1} = \alpha_k$ och $\beta_{k+1} = t_k$.

$(iv)$ $k = k + 1$. Gå till $(i)$.



## Newtons metod (för linjesökning)

Newtons metod använder första och andra ordningens information för att bestämma en punkt sådan att $\varphi'(t^{(k)}) \approx 0$.

Låt $t_0 = 0$ och och utför iterationen

$$t_{k+1} = t_k - \frac{\varphi'(t_k)}{\varphi''(t_k)}$$

till dess att $|\varphi'(t_k)| \leq \epsilon$. Optimala punkten är approximativt $t^{(k)} = t_k$.

Denna metod beskrivs i ett mera generellt fall härnäst.

## Newtons metod

Idén bakom Newtons metod är att att approximera $f(\mathbf{x})$ med en andra ordningens Taylorutveckling.

Låt $\mathbf{x} = \mathbf{x}^{(k)} + \mathbf{d}$. Newtons metod utnyttjar approximationen

$$\min_{\mathbf{x}} f(\mathbf{x}) \approx \min_{\mathbf{d}} f(\mathbf{x}^{(k)}) + \nabla f(\mathbf{x}^{(k)})\mathbf{d} + \frac{1}{2}\mathbf{d}^{\mathsf{T}}\mathbf{F}(\mathbf{x}^{(k)})\mathbf{d}$$

Om $\mathbf{F}(\mathbf{x}^{(k)}) > 0$ (positivt definit) så uppfyller minpunkten $\mathbf{d}^{(k)}$

$$\mathbf{F}(\mathbf{x}^{(k)})\mathbf{d}^{(k)} = -\nabla f(\mathbf{x}^{(k)})^{\mathsf{T}}$$

I annat fall kan man låta $\mathbf{H}(\mathbf{x}^{(k)}) = \mathbf{F}(\mathbf{x}^{(k)}) + \mu I$, där $\mu > 0$ är tillräckligt stor så att $\mathbf{H}(\mathbf{x}^{(k)}) > 0$ och då använda sökriktningen $\mathbf{d}^{(k)}$ som uppfyller

$$\mathbf{H}(\mathbf{x}^{(k)})\mathbf{d}^{(k)} = -\nabla f(\mathbf{x}^{(k)})^{\mathsf{T}}$$



**Newtons algoritm:**

(0) Bestäm startpunkt $\mathbf{x}^{(0)}$ och låt $k = 0$.

$(i)$ Kontrollera avbrottskriteriet $\|\nabla f(\mathbf{x}^{(k)})\| \leq \epsilon$

$(ii)$ Bestäm sökriktning

$$\mathbf{F}(\mathbf{x}^{(k)})\mathbf{d}^{(k)} = -\nabla f(\mathbf{x}^{(k)})^{\mathsf{T}}$$

$(iii)$ Låt $\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + t^{(k)}\mathbf{d}^{(k)}$, där $t^{(k)}$ är det största talet av $1,\ 1/2,\ 1/4, \ldots$, så att $f(\mathbf{x}^{(k)} + t^{(k)}\mathbf{d}^{(k)}) < f(\mathbf{x}^{(k)})$

$(iv)$ $k = k + 1$. Gå till $(i)$.

**Kommentar 1.** $(iii)$ *kan ersättas med linjesökning. Detta är speciellt lämpligt om* $f(\cdot)$ *inte är en konvex funktion.*



## Ickelinjär minstakvadratskattning

**Problem** Finn $\mathbf{x}$ så att (approximativt)

$$\begin{aligned} h_1(\mathbf{x}) &= 0 \\ h_2(\mathbf{x}) &= 0 \\ &\vdots \\ h_m(\mathbf{x}) &= 0 \end{aligned}$$

Idé: Lös det olinjära minsta kvadratproblemet

$$\min f(\mathbf{x}) = \frac{1}{2}\sum_{i=1}^{m} h_i(\mathbf{x})^2 = \frac{1}{2}\mathbf{h}(\mathbf{x})^{\mathsf{T}}\mathbf{h}(\mathbf{x}) \quad \text{där} \quad \mathbf{h}(\mathbf{x}) = \begin{bmatrix} h_1(\mathbf{x}) \\ \vdots \\ h_m(\mathbf{x}) \end{bmatrix} \qquad (1)$$

Om $f(\hat{\mathbf{x}}) \approx 0$ så gäller att $h_i(\hat{\mathbf{x}}) \approx 0$, $i = 1, \ldots, m$.



## Gauss-Newtons metod

Vi betraktar två härledningar av Gauss-Newtons metod för att lösa (1).

**Metod 1:** Om vi använder att

$$\mathbf{h}(\mathbf{x}^{(k)} + \mathbf{d}) \approx \mathbf{h}(\mathbf{x}^{(k)}) + \nabla\mathbf{h}(\mathbf{x}^{(k)})\mathbf{d}$$

så får vi approximationen

$$\min_{\mathbf{x}} \frac{1}{2}\mathbf{h}(\mathbf{x})^{\mathsf{T}}\mathbf{h}(\mathbf{x}) \approx \min_{\mathbf{d}} \frac{1}{2}|\nabla h(\mathbf{x}^{(k)})\mathbf{d} + \mathbf{h}(\mathbf{x}^{(k)})|^2$$

Detta är ett minsta kvadratproblem på standardform vars lösning ges av normalekvationen

$$\nabla\mathbf{h}(\mathbf{x}^{(k)})^{\mathsf{T}}\nabla\mathbf{h}(\mathbf{x}^{(k)})\mathbf{d}^{(k)} = -\nabla\mathbf{h}(\mathbf{x}^{(k)})^{\mathsf{T}}\mathbf{h}(\mathbf{x}^{(k)})$$

Nästa iterationspunkt ges således av $\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + t^{(k)}\mathbf{d}^{(k)}$ där $t^{(k)}$ t.ex. bestäms med linjesökning.

**Metod 2:** Använd Newtonriktningen. För $f(\mathbf{x}) = \frac{1}{2}\mathbf{h}(\mathbf{x})^{\mathsf{T}}\mathbf{h}(\mathbf{x})$ får vi

$$\nabla f(\mathbf{x}) = \sum_{i=1}^{m} h_i(\mathbf{x})\nabla h_i(\mathbf{x}) = \mathbf{h}(\mathbf{x})^{\mathsf{T}}\nabla\mathbf{h}(\mathbf{x})$$

$$\begin{aligned}\mathbf{F}(\mathbf{x}) = \nabla^2 f(\mathbf{x}) &= \sum_{i=1}^{m}(\nabla h_i(\mathbf{x})^{\mathsf{T}}\nabla h_i(\mathbf{x}) + h_i(\mathbf{x})\nabla^2 h_i(\mathbf{x})) \\ &= \nabla\mathbf{h}(\mathbf{x})^{\mathsf{T}}\nabla\mathbf{h}(\mathbf{x}) + \sum_{i=1}^{m} h_i(\mathbf{x})\nabla^2 h_i(\mathbf{x}))\end{aligned}$$

Newtonriktningen ges av

$$(\nabla\mathbf{h}(\mathbf{x}^{(k)})^{\mathsf{T}}\nabla\mathbf{h}(\mathbf{x}^{(k)}) + \sum_{i=1}^{m} h_i(\mathbf{x}^{(k)})\nabla^2 h_i(\mathbf{x})^{(k)})\mathbf{d}^{(k)} = -\mathbf{h}(\mathbf{x}^{(k)})^{\mathsf{T}}\nabla\mathbf{h}(\mathbf{x}^{(k)})$$



Om vi gör approximationen $h_i(\mathbf{x}^{(k)}) \approx 0$ så får vi

$$\nabla\mathbf{h}(\mathbf{x}^{(k)})^{\mathsf{T}}\nabla\mathbf{h}(\mathbf{x}^{(k)})\mathbf{d}^{(k)} = -\mathbf{h}(\mathbf{x}^{(k)})^{\mathsf{T}}\nabla\mathbf{h}(\mathbf{x}^{(k)})$$

vilket överenstämmer med **Metod 1**.

Nästa iterationspunkt ges av $\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + t^{(k)}\mathbf{d}^{(k)}$ där $t^{(k)}$ t.ex. bestäms med linjesökning.



## Läsanvisningar

- Ickelinjär optimering, Grå häftet, sidan 1-17.