Föreläsning 10: Lagrangerelaxering och KKT villkoren

1. Lagrangerelaxering
   - Globala optimalitetsvillkoren
   - KKT villkoren för konvexa problem
   - Tillämpningar
2. KKT villkoren för allmänna ickelinjära optimeringsproblem

Lagrangerelaxering

Vi betraktar optimeringsproblemet

$$\begin{aligned} \text{minimera} \quad & f(\mathbf{x}) \\ \text{då} \quad & g_i(\mathbf{x}) \le 0, \; i = 1, \dots, m \end{aligned} \tag{1}$$

där $f(\mathbf{x})$ och $g_i(\mathbf{x})$ är reellvärda funktioner.

Om $\mathbf{g}(\mathbf{x}) = \begin{bmatrix} g_1(\mathbf{x}) & \dots & g_m(\mathbf{x}) \end{bmatrix}^\mathsf{T}$ så kan (1) skrivas

$$\begin{aligned} \text{minimera} \quad & f(\mathbf{x}) \\ \text{då} \quad & \mathbf{g}(\mathbf{x}) \le \mathbf{0}. \end{aligned} \tag{2}$$

Idéen bakom Lagrangerelaxering är att sätta icke-negativa priser, $y_i \ge 0$, på bivillkoren och sedan addera dessa till målfunktionen. Detta ger optimeringsproblemet (utan bivillkor)

$$\text{minimera} \quad f(\mathbf{x}) + \sum_{i=1}^{m} y_i g_i(\mathbf{x}) \tag{3}$$

vilket om $\mathbf{y} = \begin{bmatrix} y_1 & \dots & y_m \end{bmatrix}^\mathsf{T}$ kan skrivas

$$\text{minimera} \quad f(\mathbf{x}) + \mathbf{y}^\mathsf{T}\mathbf{g}(\mathbf{x})$$

"Priset" y_i kallas för en *Lagrangemultiplikator.*

Definition 1. *Funktionen* $L : \mathbf{R}^n \times \mathbf{R}^m \to \mathbf{R}$ *definierad av* $L(\mathbf{x}, \mathbf{y}) = f(\mathbf{x}) + \mathbf{y}^\mathsf{T}\mathbf{g}(\mathbf{x})$ *kallas Lagrangefunktionen till (2).*

Svag dualitet

Sats 1 (Svag dualitet). *För godtyckligt* $\mathbf{y} \ge \mathbf{0}$ *gäller att* $\min_{\mathbf{x}} L(\mathbf{x}, \mathbf{y}) \le f(\hat{\mathbf{x}})$, *där* $\hat{\mathbf{x}}$ *är en optimallösning till (2).*

Bevis: *Eftersom* $\hat{\mathbf{x}}$ *är tillåten lösning till (2) gäller* $\mathbf{g}(\hat{\mathbf{x}}) \le \mathbf{0}$. *Vi får*

$$\min_{\mathbf{x}} L(\mathbf{x}, \mathbf{y}) \le L(\hat{\mathbf{x}}, \mathbf{y}) = f(\hat{\mathbf{x}}) + \underbrace{\mathbf{y}^\mathsf{T}}_{\ge 0} \underbrace{\mathbf{g}(\hat{\mathbf{x}})}_{\le 0} \le f(\hat{\mathbf{x}}). \tag{4}$$

- minimering av Lagrangefunktionen ger underskattningar till optimeringsproblemet (2).
- Genom lämpligt val av $\mathbf{y}$ försöker man på detta sätt att få en bra approximation till optimallösningen för (2). I praktiska algoritmer försöker man lösa $\max_{\mathbf{y} \ge \mathbf{0}} \min_{\mathbf{x}} L(\mathbf{x}, \mathbf{y})$. Nästa sats ger villkor på Lagrangemultiplikatorn som ger likhet i (4).

Globala optimalitetsvillkoren

Sats 2. *Om* $(\hat{\mathbf{x}}, \hat{\mathbf{y}}) \in \mathbf{R}^n \times \mathbf{R}^m$ *uppfyller villkoren*

(1) $L(\hat{\mathbf{x}}, \hat{\mathbf{y}}) = \min_x L(\mathbf{x}, \hat{\mathbf{y}})$,

(2) $\mathbf{g}(\hat{\mathbf{x}}) \leq \mathbf{0}$,

(3) $\hat{\mathbf{y}} \geq 0$,

(4) $\hat{\mathbf{y}}^\mathsf{T}\mathbf{g}(\hat{\mathbf{x}}) = 0$.

så är $\hat{\mathbf{x}}$ *en optimallösning till (2).*

Bevis: *Om* $\mathbf{x}$ *är en godtycklig tillåten lösning till (2) så gäller* $\mathbf{g}(\mathbf{x}) \leq \mathbf{0}$, *vilket ger att*

$$f(\mathbf{x}) \geq f(\mathbf{x}) + \hat{\mathbf{y}}^\mathsf{T}\mathbf{g}(\mathbf{x}) = L(\mathbf{x}, \hat{\mathbf{y}}) \geq L(\hat{\mathbf{x}}, \hat{\mathbf{y}}) = f(\hat{\mathbf{x}})$$

där första olikheten följer av (3) *och* $g(\mathbf{x}) \leq \mathbf{0}$, *andra olikheten följer av* (1), *och sista likheten följer av* (4).

Konvexa optimeringsproblem

- Om funktionerna f och $g_1, \ldots, g_m$ är konvexa och kontinuerligt deriverbara så är villkoret (1) i Sats 2 ekvivalent med villkoret

$$\nabla f(\hat{\mathbf{x}}) + \sum_{i=1}^{m} \hat{y}_i \nabla g_i(\hat{\mathbf{x}}) = \mathbf{0}^\mathsf{T} \qquad (5)$$

  Detta följer eftersom $L(\mathbf{x}, \hat{\mathbf{y}})$ är konvex då $\hat{\mathbf{y}} \geq 0$ och då gäller att $\hat{\mathbf{x}}$ är en minpunkt för $L(\mathbf{x}, \hat{\mathbf{y}})$ om och endast om $\nabla_\mathbf{x} L(\hat{\mathbf{x}}, \hat{\mathbf{y}}) = \mathbf{0}^\mathsf{T}$, d.v.s om och endast om (5) är uppfylld.

- De globala optimalitetsvillkoren i Sats 2 är tillräckliga villkor för optimalitet men i allmänhet inte nödvändiga. Nästa sats visar att de ofta även är nödvändiga villkor för konvexa optimeringsproblem.

Definition 2. *Optimeringsproblemet (1) är ett reguljärt konvext problem om funktionerna* f *och* $g_1, \ldots, g_m$ *är konvexa och kontinuerligt deriverbara och det finns en punkt* $\mathbf{x}_0 \in \mathbf{R}^n$ *sådan att* $g_i(\mathbf{x}_0) < 0$, $i = 1, \ldots, m$.

Sats 3 (KKT för konvexa problem). *Antag att (1) är ett reguljärt konvext problem. Då är* $\hat{\mathbf{x}}$ *en (globalt) optimal lösning om och endast om det finns en vektor* $\hat{\mathbf{y}} \in \mathbf{R}^m$ *så att*

(1) $\nabla f(\hat{\mathbf{x}}) + \sum_{i=1}^m \hat{y}_i \nabla g_i(\hat{\mathbf{x}}) = \mathbf{0}^\mathsf{T}$

(2) $\mathbf{g}(\hat{\mathbf{x}}) \leq \mathbf{0}$,

(3) $\hat{\mathbf{y}} \geq \mathbf{0}$,

(4) $\hat{\mathbf{y}}^\mathsf{T}\mathbf{g}(\hat{\mathbf{x}}) = 0$.

Bevis: *Tillräckligheten bevisades ovan. Nödvändigheten bevisas i Sektion 12 i grå häftet.*

Villkoren (2) – (4) kan göras mer explicita. Vi har att

$$\hat{\mathbf{y}}^\mathsf{T}\mathbf{g}(\hat{\mathbf{x}}) = \sum_{i=1}^{m} \hat{y}_i g_i(\hat{\mathbf{x}}) = 0$$

Eftersom $g_i(\hat{\mathbf{x}}) \leq 0$ och $\hat{y}_i \geq 0$ följer att $\hat{y}_i g_i(\hat{\mathbf{x}}) = 0$, $i = 1, \ldots, m$. Vi får således följande ekvivalenta villkor

(2′) $g_i(\hat{\mathbf{x}}) \leq 0$, $i = 1, \ldots, m$,

(3′) $\hat{y}_i \geq 0$, $i = 1, \ldots, m$,

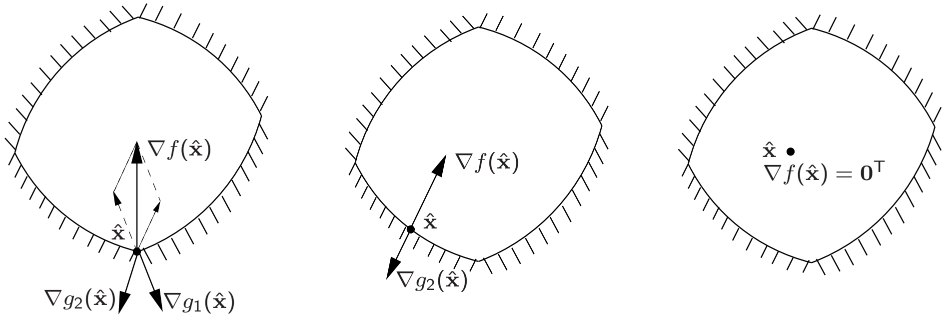
(4′) $\hat{y}_i \cdot g_i(\hat{\mathbf{x}}) = 0$, $i = 1, \ldots, m$.

Geometrisk tolkning

Komplementaritetsvillkoret (4′) implicerar att om $g_i(\hat{\mathbf{x}}) < 0$ så är $\hat{y}_i = 0$. Således kan villkor (1) skrivas

$$\nabla f(\hat{\mathbf{x}}) = -\sum_{i:g_i(\hat{\mathbf{x}})=0} \hat{y}_i \nabla g_i(\hat{\mathbf{x}})$$

Detta betyder att gradienten är en negativ linjärkombination av de bindande (aktiva) bivillkorens gradienter.



Kvadratisk optimering med olikhetsbivillkor

$$\begin{aligned} \text{minimera} \quad & \frac{1}{2}\mathbf{x}^\mathsf{T}\mathbf{H}\mathbf{x} + \mathbf{c}^\mathsf{T}\mathbf{x} + c_0 \\ \text{då} \quad & \mathbf{A}\mathbf{x} \geq \mathbf{b}. \end{aligned} \tag{6}$$

Om $\mathbf{H}$ är positivt semidefinit är detta ett konvext optimeringsproblem och vi kan applicera Sats 3.

Sats 4. *$\hat{\mathbf{x}}$ är en (globalt) optimal lösning till (6) om och endast om det finns en vektor $\hat{\mathbf{y}} \in \mathbf{R}^m$ så att*

(1) $\mathbf{H}\hat{\mathbf{x}} + \mathbf{c} = \mathbf{A}^\mathsf{T}\hat{\mathbf{y}}$

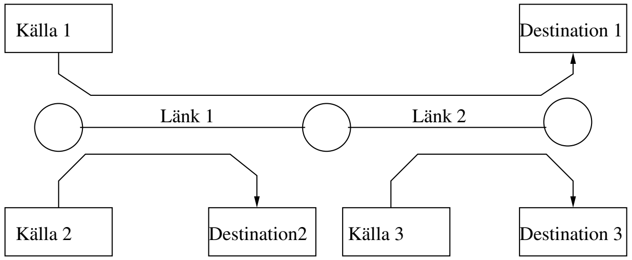
(2) $\mathbf{A}\hat{\mathbf{x}} \geq \mathbf{b}$,

(3) $\hat{\mathbf{y}} \geq 0$,

(4) $\hat{\mathbf{y}}^\mathsf{T}(\mathbf{A}\hat{\mathbf{x}} - \mathbf{b}) = 0$.

Trafikreglering i kommunikationssystem

Vi betraktar ett kommunikationsnätverk bestående av två länkar. Tre källor skickar data över nätverket till tre olika destinationer.



- Källa 1 använder båda länkarna.
- Källa 2 använder länk 1.
- Källa 3 använder länk 2.

- Länk 1 har kapaciteten 2 (normaliserad storhet [datamängd/sek])
- Länk 2 har kapaciteten 1
- De tre källorna sänder med datahastigheterna x_r, $r = 1, 2, 3$.
- De tre källorna har varsin nyttofunktion $U_r(x)$, $r = 1, 2, 3$. Ett vanligt val av nyttofunktion är $U_r(x_r) = w_r \log(x_r)$.

För effektivt och rättvist utnyttjande av den tillgängliga kapaciteten väljer man datahastigheterna enligt följande optimeringskriterium

$$\begin{aligned} \text{maximera} \quad & U_1(x_1) + U_2(x_2) + U_3(x_3) \\ \text{då} \quad & x_1 + x_2 \leq 2 \\ & x_1 + x_3 \leq 1 \\ & x_1 \geq 0,\ x_2 \geq 0,\ x_3 \geq 0 \end{aligned}$$

Antag $U_k(x) = \log(x_k)$, $k = 1, 2, 3$. Optimeringsproblemet kan skrivas

$$\begin{aligned} -\text{minimera} \quad & -\log(x_1) - \log(x_2) - \log(x_3) \\ \text{då} \quad & x_1 + x_2 \leq 2 \\ & x_1 + x_3 \leq 1 \end{aligned}$$

Vi strök bivillkoren $x_k \geq 0$, $k = 1, \ldots, 3$ eftersom de automatiskt blir uppfyllda ty $-\log(x) \to \infty$ $x \to 0$.

Optimeringsproblemet är konvext eftersom bivillkoren består av linjära olikheter och målfunktionen är en summa av konvexa funktioner och är därmed konvex.

Optimalitetsvillkoren i Sats 3 blir

$$\begin{aligned}
(1) \quad & -\tfrac{1}{x_1} + y_1 + y_2 = 0 \\
& -\tfrac{1}{x_2} + y_1 = 0 \qquad \nabla f(x) + y_1 \nabla g_1(x) + y_2 \nabla g_2(x) = 0 \\
& -\tfrac{1}{x_3} + y_2 = 0 \\
(2) \quad & x_1 + x_2 - 2 \leq 0 \qquad g_i(x) \leq 0 \\
& x_1 + x_3 - 1 \leq 0 \\
(3) \quad & y_1 \geq 0 \qquad y_i \geq 0 \\
& y_2 \geq 0 \\
(4) \quad & y_1(x_1 + x_2 - 2) = 0 \qquad y_i g_i(x) = 0 \\
& y_2(x_1 + x_3 - 1) = 0
\end{aligned}$$

Ur (1) får vi

$$x_1 = \frac{1}{y_1 + y_2} \qquad x_2 = \frac{1}{y_1} \qquad x_3 = \frac{1}{y_2}$$

Detta medför att $y_1 > 0$ och $y_2 > 0$ varvid komplementaritetsvillkoret (4) ger att (2) är uppfylld med likhet. Vi får

$$\begin{matrix} x_1+x_2-2=0 \\ x_1+x_3-1=0 \end{matrix} \quad \Rightarrow \quad \begin{matrix} \frac{1}{y_1+y_2} + \frac{1}{y_1} = 2 \\ \frac{1}{y_1+y_2} + \frac{1}{y_2} = 1 \end{matrix} \quad \Rightarrow \quad \begin{matrix} y_1 = \frac{\sqrt{3}}{\sqrt{3}+1} \\ y_2 = \sqrt{3} \end{matrix}$$

vilket i sin tur ger de optimala datahastigheterna

$$\hat{x}_1 = \frac{\sqrt{3}+1}{3+2\sqrt{3}} \qquad \hat{x}_2 = \frac{\sqrt{3}}{\sqrt{3}+1}, \qquad \hat{x}_3 = \frac{1}{\sqrt{3}}$$

Allmänna icke-linjära problem under likhetsbivillkor

$$\begin{aligned} \text{minimera} \quad & f(\mathbf{x}) \\ \text{då} \quad & h_i(\mathbf{x}) = 0, \; i = 1, \ldots, m \end{aligned} \tag{7}$$

Det tillåtna området $\mathcal{F} = \{\mathbf{x} \in \mathbf{R}^n : h_i(\mathbf{x}) = 0, \; i = 1, \ldots, m\}$ är inte konvext såvida inte funktionerna h_i är affina, d.v.s. $h_i(\mathbf{x}) = \mathbf{a}_i^\mathsf{T}\mathbf{x} + b_i$. Vi antar att $n > m$.

Vi behöver följande tekniska antagande:

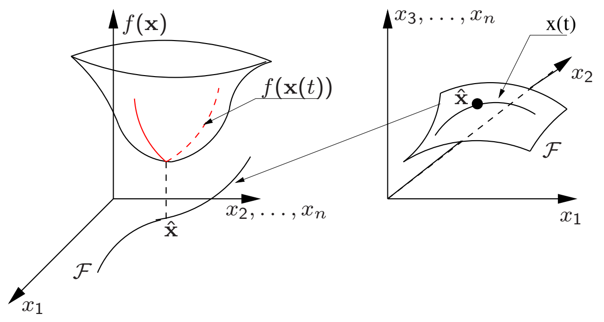
Definition 3. *En tillåten lösning $\mathbf{x} \in \mathcal{F}$ är en reguljär punkt till (7) om $\nabla h_i(\mathbf{x})$, $i = 1, \ldots, m$ är linjärt oberoende.*

Sats 5 (KKT för problem med likhetsbivillkor). *Antag att $\hat{\mathbf{x}} \in \mathcal{F}$ är en reguljär punkt och en lokal optimallösning till (7). Då finns $\hat{\mathbf{u}} \in \mathbf{R}^m$ så att*

(1) $\nabla f(\hat{\mathbf{x}}) + \sum_{i=1}^{m} \hat{u}_i \nabla h_i(\hat{\mathbf{x}}) = \mathbf{0}^\mathsf{T}$,

(2) $h_i(\hat{\mathbf{x}}) = 0$, $i = 1, \ldots, m$.

Bevisskiss: *Låt $\mathbf{x}(t)$ vara en godtycklig parametriserad kurva i tillåtna mängden $\mathcal{F} = \{\mathbf{x} \in \mathbf{R}^n : h_i(\mathbf{x}) = 0,\ i = 1, \ldots, m\}$ sådan att $\mathbf{x}(0) = \hat{\mathbf{x}}$. Figuren på nästa sida illusterar hur denna parametriserade kurva avbildas på en kurva $f(\mathbf{x}(t))$ på målfunktionsytan. Den tillåtna mängden $\mathcal{F}$ är i allmänhet av högre dimension än 1 vilket illustreras med hjälp av högra figuren.*



Eftersom $\mathbf{x}(0) = \hat{\mathbf{x}}$ är en lokal optimal lösning gäller att

$$\frac{d}{dt} f(\mathbf{x}(t))|_{t=0} = \nabla f(\hat{\mathbf{x}}) \cdot \mathbf{x}'(0) = 0$$

Vidare gäller att $\mathbf{x}(t) \in \mathcal{F}$ vilket medför att

$$h_i(\mathbf{x}(t)) = 0, \quad i = 1, \ldots, m, \quad \forall t \in (-\epsilon, \epsilon)$$

för något $\epsilon > 0$.

Detta innebär att

$$\frac{d}{dt} h_i(\mathbf{x}(t))|_{t=0} = \nabla h_i(\hat{\mathbf{x}}) \cdot \mathbf{x}'(0) = 0, \quad i = 1, \ldots, m$$

vilket i sin tur innebär att $\mathbf{x}'(0) \in \mathcal{N}(\mathbf{A})$, där

$$\mathbf{A} = \begin{bmatrix} \nabla h_1(\hat{\mathbf{x}}) \\ \vdots \\ \nabla h_m(\hat{\mathbf{x}}) \end{bmatrix}$$

Omvänt kan man med hjälp av implicita funktionssatsen visa att om $\mathbf{p} \in \mathcal{N}(\mathbf{A})$ så finns en parametriserad kurva $\mathbf{x}(t) \in \mathcal{F}$ med $\mathbf{x}(0) = \hat{\mathbf{x}}$ och $\mathbf{x}'(0) = \mathbf{p}$.

Sammantaget betyder ovanstående resonemang att

$$\begin{aligned} & \nabla f(\hat{\mathbf{x}})\mathbf{p} = 0, \qquad \forall \mathbf{p} \in \mathcal{N}(\mathbf{A}) \\ \Leftrightarrow \quad & \nabla f(\hat{\mathbf{x}})^\mathsf{T} \in \mathcal{N}(\mathbf{A})^\perp = \mathcal{R}(\mathbf{A}^\mathsf{T}) \\ \Leftrightarrow \quad & \nabla f(\hat{\mathbf{x}})^\mathsf{T} = \mathbf{A}^\mathsf{T}\hat{\mathbf{v}}, \end{aligned}$$

för något $\hat{\mathbf{v}} \in \mathbf{R}^m$.
Om vi låter $\hat{\mathbf{u}} = -\hat{\mathbf{v}} \in \mathbf{R}^m$ kan det sista uttrycket skrivas

$$\nabla f(\hat{\mathbf{x}}) + \sum_{i=1}^{m} \hat{u}_i \nabla h_i(\hat{\mathbf{x}}) = \mathbf{0}^\mathsf{T}$$

vilket skulle visas.

Allmänna ickelinjära optimeringsproblem med olikhetsbivillkor

$$\begin{aligned} \text{minimera} \quad & f(\mathbf{x}) \\ \text{då} \quad & g_i(\mathbf{x}) \le 0, \; i = 1, \dots, m \end{aligned} \tag{8}$$

där $f(\mathbf{x})$ och $g_i(\mathbf{x})$ är reellvärda funktioner. Om problemet inte är konvext (dvs om $\mathcal{F} = \{\mathbf{x} \in \mathbf{R}^n : g_i(\mathbf{x}) \le 0, \; i = 1, \dots, m\}$ inte är konvext och/eller f inte är konvex på $\mathcal{F}$) kan man i allmänhet endast härleda nödvändiga optimalitetsvillkor. Regularitetsvillkoret i Definition 2 måste ersättas med ett starkare villkor.

Definition 4. *För $\mathbf{x} \in \mathcal{F}$ låter vi $\mathcal{I}_a(\mathbf{x})$ beteckna indexmängden för de aktiva bivillkoren i punkten $\mathbf{x}$, dvs $\mathcal{I}_a(\mathbf{x}) = \{i \in \{1, \dots, m\} : g_i(\mathbf{x}) = 0\}$.*

Definition 5. *En tillåten lösning $\mathbf{x} \in \mathcal{F}$ är en reguljär punkt till (8) om $\nabla g_i(\mathbf{x})$ för $i \in \mathcal{I}_a(\mathbf{x})$ är linjärt oberoende.*

Sats 6 (KKT för allmänna problem med olikhetsbivillkor). *Antag att (8) är ett reguljärt problem. Om $\hat{\mathbf{x}}$ är en lokal optimal lösning så finns en vektor $\hat{\mathbf{y}} \in \mathbf{R}^m$ så att*

(1) $\nabla f(\hat{\mathbf{x}}) + \sum_{i=1}^m \hat{y}_i \nabla g_i(\hat{\mathbf{x}}) = \mathbf{0}^\mathsf{T}$

(2) $g_i(\hat{\mathbf{x}}) \le 0$, $i = 1, \dots, m$,

(3) $\hat{\mathbf{y}}_i \ge 0$, $i = 1, \dots, m$,

(4) $\hat{\mathbf{y}}_i \cdot g_i(\hat{\mathbf{x}}) = 0$, $i = 1, \dots, m$.

Bevis: *Satsen bevisas i Sektion 12 i grå häftet (ganska avancerat). Det är samma bevis som beviset för nödvändighet i Sats 3.*

Ett exempel

$$\begin{aligned} \text{minimera} \quad & (x_1 - 3)^2 + (x_2 - 2)^2 \\ \text{då} \quad & 2x_1 + x_2 - 6 \le 0, \\ & x_1 + 2x_2 - 6 \le 0 \end{aligned}$$

Här är $f(\mathbf{x}) = (x_1 - 3)^2 + (x_2 - 2)^2$,
$g_1(\mathbf{x}) = 2x_1 + x_2 - 6$ och $g_2(\mathbf{x}) = x_1 + 2x_2 - 6$, så då blir

$$\nabla f(\mathbf{x}) = \begin{bmatrix} 2(x_1 - 3) & 2(x_2 - 2) \end{bmatrix},$$

$$\nabla g_1(\mathbf{x}) = \begin{bmatrix} 2 & 1 \end{bmatrix}, \quad \nabla g_2(\mathbf{x}) = \begin{bmatrix} 1 & 2 \end{bmatrix}.$$

Bestäm lösningar till KKT

Vi kan bestämma alla lösningar till KKT-villkoren genom att betrakta alla kombinationer av aktiva och inaktiva bivillkor.
(funkar på små problem)

Fyra fall uppstår:

1. Inga bivillkor aktiva $\mathcal{I}_a(\mathbf{x}) = \emptyset$,
   dvs $g_1(\mathbf{x}) < 0$ och $g_2(\mathbf{x}) < 0$.
2. Första aktivt och andra inaktivt $\mathcal{I}_a(\mathbf{x}) = \{1\}$,
   dvs $g_1(\mathbf{x}) = 0$ och $g_2(\mathbf{x}) < 0$.
3. Första inaktivt och andra aktivt $\mathcal{I}_a(\mathbf{x}) = \{2\}$,
   dvs $g_1(\mathbf{x}) < 0$ och $g_2(\mathbf{x}) = 0$.
4. Båda bivillkoren aktiva $\mathcal{I}_a(\mathbf{x}) = \{1, 2\}$,
   dvs $g_1(\mathbf{x}) = 0$ och $g_2(\mathbf{x}) = 0$.

Fall 1: $\mathcal{I}_a(\mathbf{x}) = \emptyset$

Från KKT(4) får vi att både $y_1 = 0$ och $y_2 = 0$.

Då säger KKT(1) att

$$\nabla f(\mathbf{x}) = \begin{bmatrix} 2(x_1 - 3) & 2(x_2 - 2) \end{bmatrix} = 0,$$

vilket ger att $x_1 = 3$ och $x_2 = 2$.

Men eftersom KKT(2) inte är uppfyllt (dvs $\mathbf{x}$ är inte tillåten) för dessa värden på $\mathbf{x}$ så kan inte detta vara en lokal minpunkt.

Fall 2: $\mathcal{I}_a(\mathbf{x}) = \{1\}$

Från KKT(4) får vi att $y_2 = 0$. Då säger KKT(1) att

$$\nabla f(\mathbf{x}) + y_1 \nabla g_1(\mathbf{x}) = \begin{bmatrix} 2(x_1 - 3) & 2(x_2 - 2) \end{bmatrix} + y_1 \begin{bmatrix} 2 & 1 \end{bmatrix} = 0,$$

vilket ger att $x_1 = 3 - y_1$ och $x_2 = 2 - y_1/2$.
Antagandet att $g_1(\mathbf{x}) = 0$ ger nu med dessa x_1 och x_2

$$2(3 - y_1) + (2 - y_1/2) - 6 = 0 \qquad \Rightarrow \qquad y_1 = 4/5$$

Detta y_1 (och y_2) uppfyller KKT(3), och då blir
$x_1 = 3 - 4/5 = 11/5$ och $x_2 = 2 - 1/2(4/5) = 8/5$.

Slutligen, $g_2(\mathbf{x}) = 11/5 + 2(8/5) - 6 = -3/5 \leq 0$ så KKT(2) är uppfyllt.

Alla KKT-villkoren är alltså uppfyllda.

Fall 3: $\mathcal{I}_a(\mathbf{x}) = \{2\}$

Från KKT(4) får vi att $y_1 = 0$. Då säger KKT(1) att

$$\nabla f(\mathbf{x}) + y_2 \nabla g_2(\mathbf{x}) = \begin{bmatrix} 2(x_1 - 3) & 2(x_2 - 2) \end{bmatrix} + y_2 \begin{bmatrix} 1 & 2 \end{bmatrix} = 0,$$

vilket ger att $x_1 = 3 - y_2/2$ och $x_2 = 2 - y_2$.

Antagandet att $g_2(\mathbf{x}) = 0$ ger nu med dessa x_1 och x_2

$$(3 - y_2/2) + 2(2 - y_2) - 6 = 0 \qquad \Rightarrow \qquad y_2 = 2/5$$

Detta y_2 (och y_1) uppfyller KKT(3), och då blir
$x_1 = 3 - 1/5 = 14/5$ och $x_2 = 2 - (2/5) = 8/5$.

Slutligen, $g_1(\mathbf{x}) = 2(14/5) + 8/5 - 6 = 1/5 > 0$ så KKT(2) är *ej* uppfyllt.

Fall 4: $\mathcal{I}_a(\mathbf{x}) = \{1, 2\}$

Antagandet att $g_1(\mathbf{x}) = 2x_1 + x_2 - 6 = 0$ och $g_2(\mathbf{x}) = x_1 + 2x_2 - 6 = 0$ ger att $x_1 = 2$ och $x_2 = 2$.

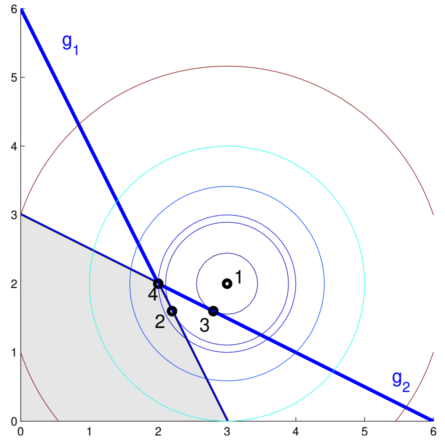
KKT(1) säger att $\nabla f(\mathbf{x}) + y_1 \nabla g_1(\mathbf{x}) + y_2 \nabla g_2(\mathbf{x}) = 0$,

$$\begin{bmatrix} 2(x_1 - 3) & 2(x_2 - 2) \end{bmatrix} + y_1 \begin{bmatrix} 2 & 1 \end{bmatrix} + y_2 \begin{bmatrix} 1 & 2 \end{bmatrix} = 0,$$

med x_1 och x_2 insatta ger det $-2 + 2y_1 + y_2 = 0$ och $y_1 + 2y_2 = 0$, så $y_1 = -2/3$ och $y_2 = 4/3$.

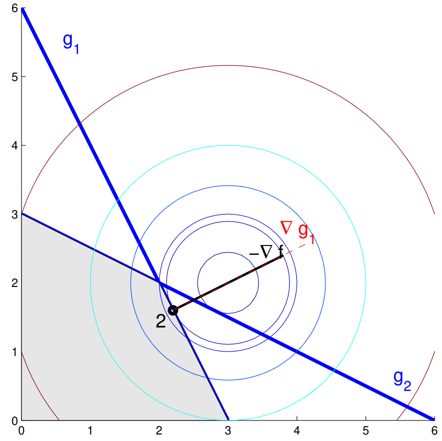
Men detta $\mathbf{y}$ uppfyller inte KKT(3) så detta kan inte vara en lokal minpunkt.

Grafisk illustration



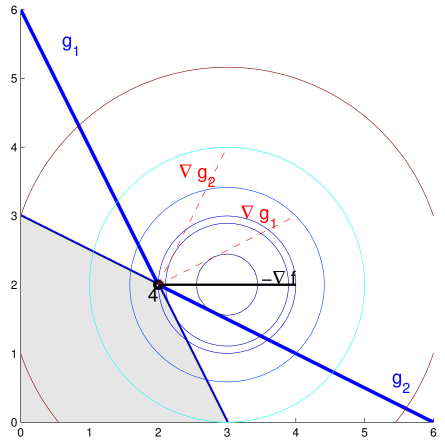
De fyra "lösningarna" ges i figuren. Som vi såg ovan är det bara nummer 2 som uppfyller alla KKT-villkoren.

Fall 2



Vi ser här att minus gradienten till f är en positiv linjärkombination av det enda bindande bivillkoret.

Fall 4



Vi ser här att minus gradienten till f *inte* är en positiv linjärkombination av de bindande bivillkoren.

Läsanvisningar

- Lite blandad optimeringsteori, röda häftet, sidorna 17-19.
- Ickelinjär optimering, grå häftet, sidorna 18-27.