



Föreläsning 10: NLP med olikhetsbivillkor och KKT villkoren

Ickelinjär optimering med olikhetsbivillkor

Vi betraktar optimeringsproblemet

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

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

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

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

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

Sats 1 (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*

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

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

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

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

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

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

Eftersom $g_i(\hat{\mathbf{x}}) \leq 0$ och $\hat{y}_i \geq 0$ följer att $\hat{y}_i g_i(\hat{\mathbf{x}}) = 0$, $i = 1, \dots, m$.

Vi får således följande ekvivalenta villkor

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

$$(3') \quad \hat{y}_i \geq 0, \quad i = 1, \dots, m,$$

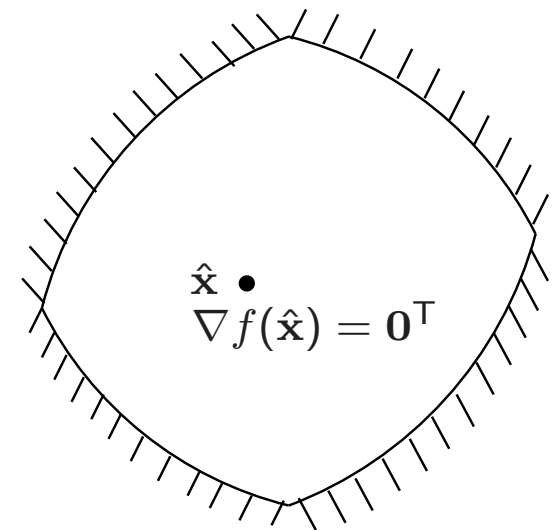
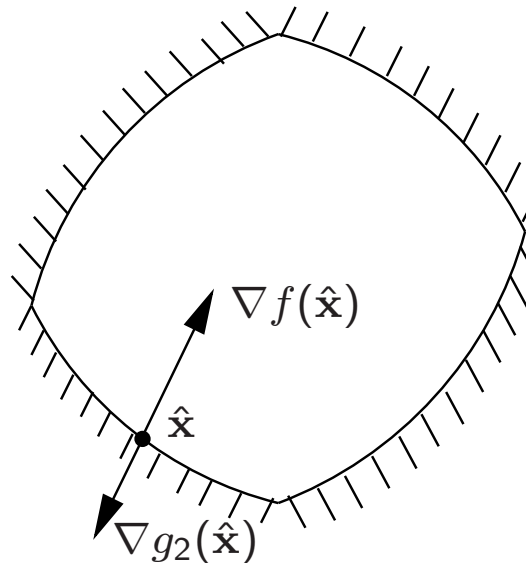
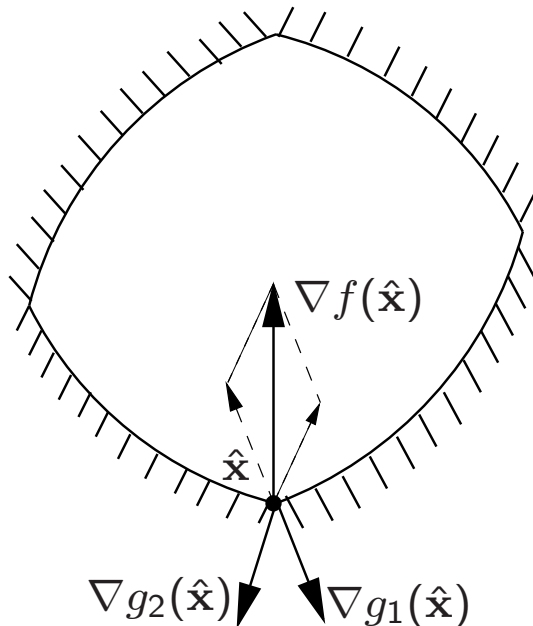
$$(4') \quad \hat{y}_i \cdot g_i(\hat{\mathbf{x}}) = 0, \quad i = 1, \dots, m.$$

Geometrisk tolkning

Komplementaritetsvillkoret (4') implicerar att om $g_i(\hat{\mathbf{x}}) < 0$ så är $y_i = 0$. 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

$$\begin{aligned} &\text{minimera} \quad \frac{1}{2} \mathbf{x}^T \mathbf{H} \mathbf{x} + \mathbf{c}^T \mathbf{x} + c_0 \\ &\text{då} \quad \mathbf{A} \mathbf{x} \geq \mathbf{b}. \end{aligned} \tag{3}$$

Om $\mathbf{H}$ är positivt semidefinit är detta ett konvext optimeringsproblem och vi kan applicera Sats 1.

Sats 2. $\hat{\mathbf{x}}$ är en (globalt) optimal lösning till (3) om och endast om det finns en vektor $\hat{\mathbf{y}} \in \mathbf{R}^m$ så att

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

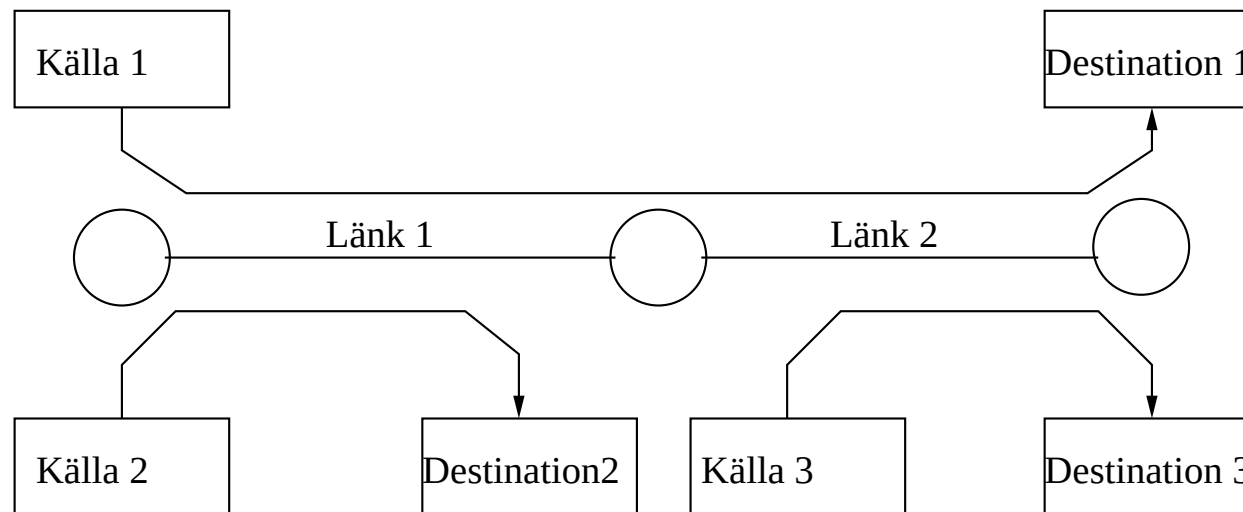
$$(2) \quad \mathbf{A} \hat{\mathbf{x}} \geq \mathbf{b},$$

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

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

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

$$\text{maximera } U_1(x_1) + U_2(x_2) + U_3(x_3)$$

$$\text{då } x_1 + x_2 \leq 2$$

$$x_1 + x_3 \leq 1$$

$$x_1 \geq 0, x_2 \geq 0, x_3 \geq 0$$

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

$$-\text{minimera} \quad -\log(x_1) - \log(x_2) - \log(x_3)$$

$$\text{då} \quad x_1 + x_2 \leq 2$$

$$x_1 + x_3 \leq 1$$

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

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

Optimalitetsvillkoren i Sats 1 blir

$$\begin{aligned} (1) \quad & -\frac{1}{x_1} + y_1 + y_2 = 0 \\ & -\frac{1}{x_2} + y_1 = 0 \\ & -\frac{1}{x_3} + y_2 = 0 \end{aligned}$$

$$\begin{aligned} (2) \quad & x_1 + x_2 - 2 \leq 0 \\ & x_1 + x_3 - 1 \leq 0 \end{aligned}$$

$$\begin{aligned} (3) \quad & y_1 \geq 0 \\ & y_2 \geq 0 \end{aligned}$$

$$\begin{aligned} (4) \quad & y_1(x_1 + x_2 - 2) = 0 \\ & y_2(x_1 + x_3 - 1) = 0 \end{aligned}$$

Ur (1) får vi

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

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

$$\begin{aligned} \frac{1}{y_1 + y_2} + \frac{1}{y_1} &= 2 & \Rightarrow & & y_1 &= \frac{\sqrt{3}}{\sqrt{3} + 1} \\ \frac{1}{y_1 + y_2} + \frac{1}{y_2} &= 1 & & & y_2 &= \sqrt{3} \end{aligned}$$

vilket i sin tur ger de optimala datahastigheterna

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

Allmänna icke-linjära optimeringsproblem med olikhetsbivillkor

$$\begin{aligned} &\text{minimera } f(\mathbf{x}) \\ &\text{då } g_i(\mathbf{x}) \leq 0, \quad i = 1, \dots, m \end{aligned} \tag{4}$$

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

Definition 2. 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 3. En tillåten lösning $\mathbf{x} \in \mathcal{F}$ är en reguljär punkt till (4) om $\nabla g_i(\mathbf{x})$ för $i \in \mathcal{I}_a(\mathbf{x})$ är linjärt oberoende.

Sats 3 (KKT för allmänna problem med olikhetsbivillkor).

Antag att (4) 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

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

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

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

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

Bevis: *Satsen bevisas i Sektion 12 i grå häftet (ganska avancerat). Det är samma bevis som beviset för nödvändighet i Sats 1.*

Ett exempel

minimera $(x_1 - 3)^2 + (x_2 - 2)^2$

då $2x_1 + x_2 - 6 \leq 0$,

$x_1 + 2x_2 - 6 \leq 0$

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

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

Detta y_1 (och y_2) uppfyller KKT(3), och då blir

$$x_1 = 3 - 4/5 = 11/5 \text{ och } x_2 = 2 - 1/2(4/5) = 8/5.$$

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

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

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 $x_2 = 2$.

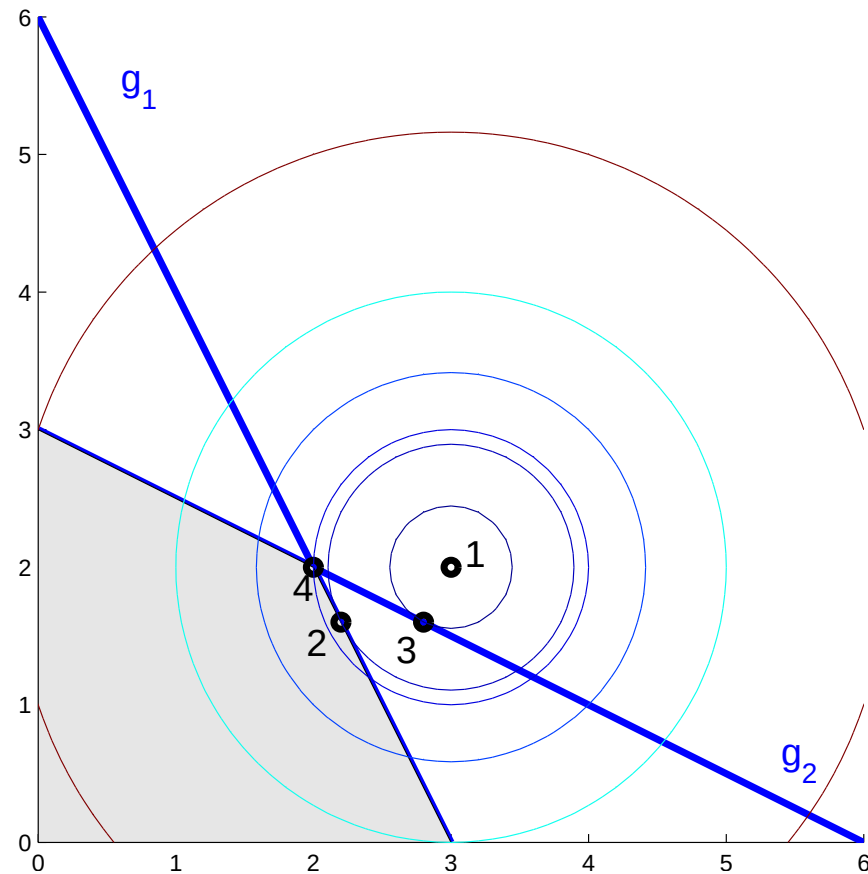
KKT(1) säger att $\nabla f(\mathbf{x}) + y_1 \nabla g_1(\mathbf{x}) + y_2 \nabla g_2(\mathbf{x}) = 0$,

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

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 $y_2 = 4/3$.

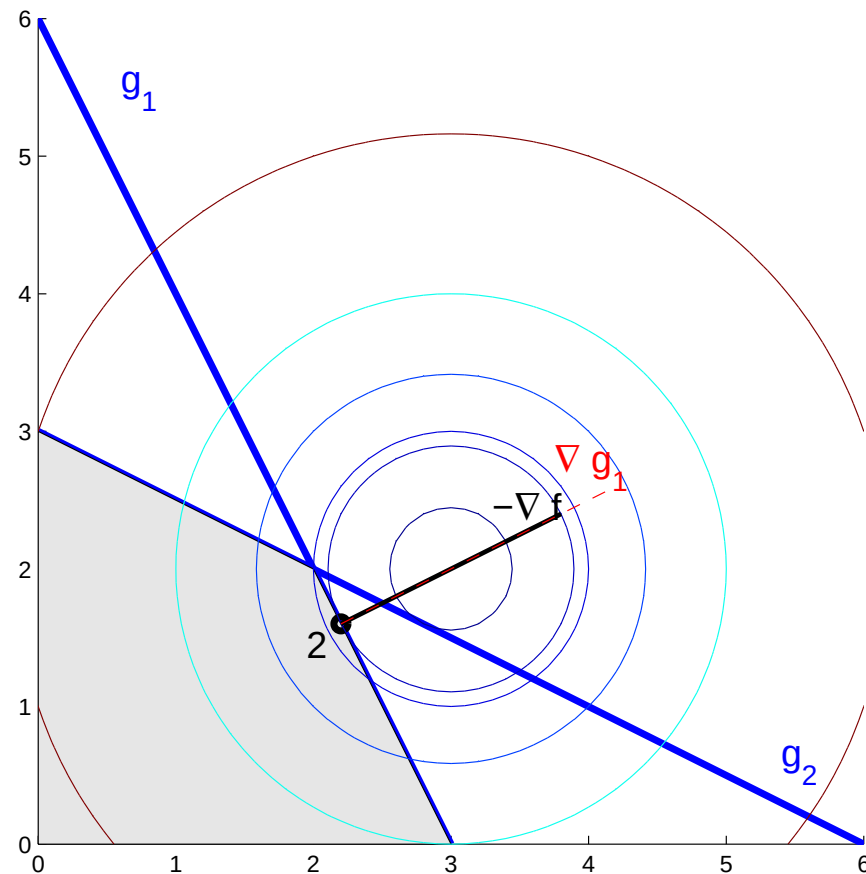
Men detta $\mathbf{y}$ uppfyller inte KKT(3) så detta kan inte vara en lokal minpunkt.

Grafisk illustration



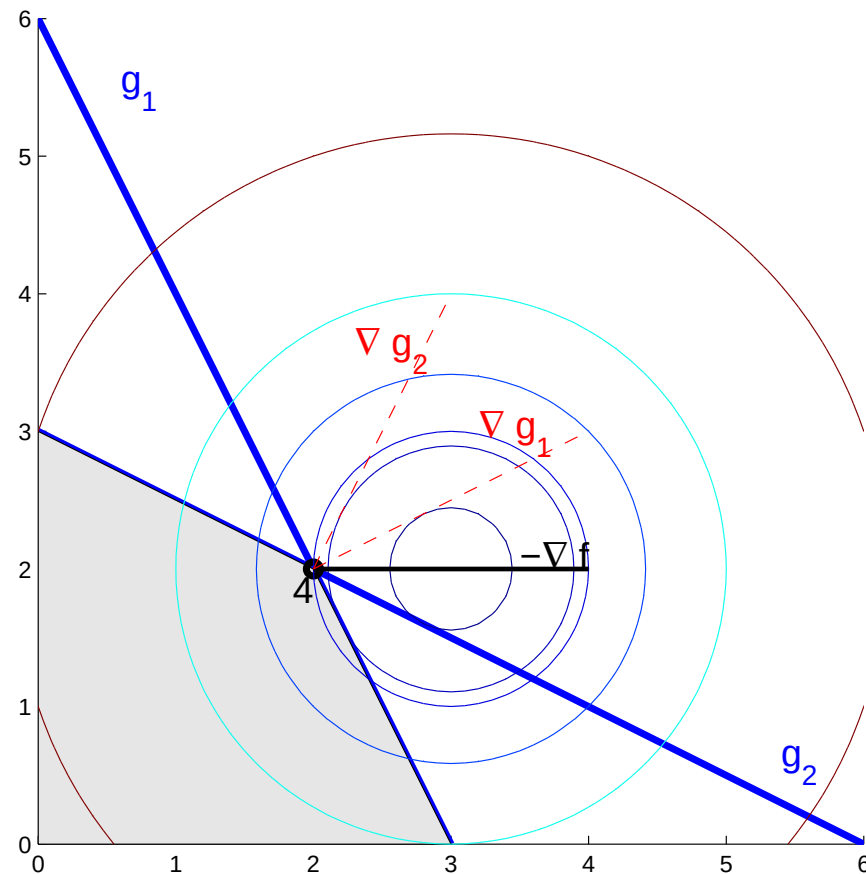
De fyra “lösningarna” ges i figuren. Som vi såg ovan är det bara nummer 2 som uppfyller alla KKT-villkoren.

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

Kapitel 20 och 21 i Optimeringskompendiet

- Ickelinjär optimering, grå häftet, sidorna 21-27.