Föreläsning 9: Konvexitet och optimering

1. Allmänna ickelinjära optimeringsproblem
2. Konvexa mängder
3. Konvexa funktioner
4. Konvexa optimeringsproblem

Allmänna ickelinjära optimeringsproblem

$$\begin{aligned} \text{minimera} \quad & f(\mathbf{x}) \\ \text{då} \quad & \mathbf{x} \in \mathcal{F} \end{aligned} \tag{1}$$

- $\mathcal{F} \subset \mathbf{R}^n$ kallas *tillåtna området*.
- $f : \mathcal{F} \to \mathbf{R}$ kallas *målfunktion*.

Definition 1.

(i) *Punkten* $\mathbf{x} \in \mathbf{R}^n$ *kallas* tillåten *om* $\mathbf{x} \in \mathcal{F}$.

(ii) *Punkten* $\hat{\mathbf{x}} \in \mathcal{F}$ *är en* lokal optimallösning *till (2) om det finns* $\delta > 0$ *så att* $f(\hat{\mathbf{x}}) \leq f(\mathbf{x})$ *för alla* $\mathbf{x} \in \mathcal{F}$ *sådana att* $|\mathbf{x} - \hat{\mathbf{x}}| \leq \delta$.

(iii) *Punkten* $\hat{\mathbf{x}} \in \mathcal{F}$ *är en* global optimallösning *till (2) om* $f(\hat{\mathbf{x}}) \leq f(\mathbf{x})$ *för alla* $\mathbf{x} \in \mathcal{F}$.

Några exempel

Linjär optimering: $f(\mathbf{x}) = \mathbf{c}^{\mathsf{T}}\mathbf{x}$ och $\mathcal{F} = \{\mathbf{x} \in \mathbf{R}^n : \mathbf{A}\mathbf{x} \geq \mathbf{b}\}$, d.v.s.

$$\begin{aligned} \text{minimera} \quad & \mathbf{c}^{\mathsf{T}}\mathbf{x} \\ \text{då} \quad & \mathbf{A}\mathbf{x} \geq \mathbf{b} \end{aligned}$$

Kvadratisk optimering: $f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^{\mathsf{T}}\mathbf{H}\mathbf{x} + \mathbf{c}^{\mathsf{T}}\mathbf{x}$ och $\mathcal{F} = \{\mathbf{x} \in \mathbf{R}^n : \mathbf{A}\mathbf{x} \geq \mathbf{b}\}$, d.v.s.

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

Ickelinjär optimering: $\mathcal{F} = \{\mathbf{x} \in \mathbf{R}^n : g_i(\mathbf{x}) \leq 0,\ i = 1, \ldots, m\}$, d.v.s.

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

Välartade optimeringsproblem (sådana vi kan lösa)

- Linjära optimeringsproblem är välartade eftersom de kan lösas med simplexmetoden.
- Kvadratiska optimeringsproblem är välartade om $\mathbf{H}$ är positivt semidefinit.
- När är allmänna ickelinjära optimeringsproblem välartade?
  - En klass av ickelinjära optimeringsproblem som är välartade är konvexa optimeringsproblem.

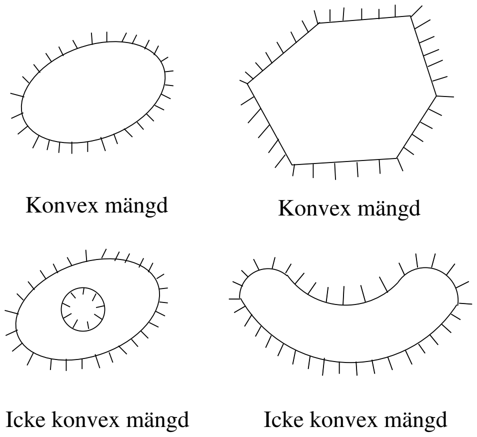
Kommentar 1. *Vi skall se att linjära optimeringsproblem och kvadratiska optimeringsproblem med positivt semidefinit* $\mathbf{H}$ *är specialfall av konvexa optimeringsproblem.*

Konvexa mängder

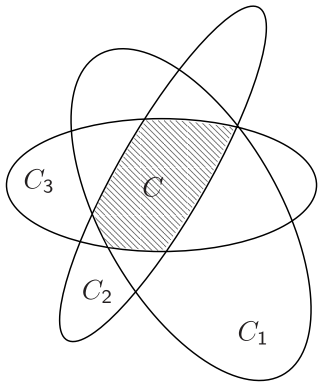
Definition 2. *En mängd $C \subset \mathbf{R}^n$ är konvex om*

$$(1-t)\mathbf{x}_1 + t\mathbf{x}_2 \in C$$

för varje val av $\mathbf{x}_1 \in C$, $\mathbf{x}_2 \in C$ och $t \in (0,1)$.



Sats 1. *Låt $C_1, \dots C_n \subset \mathbf{R}^n$ vara konvexa mängder. Då är även snittet $C = \cap_{k=1}^n C_k$ av mängderna en konvex mängd.*



Exempel 1. *Låt*

$$\mathbf{A} = \begin{bmatrix} \bar{\mathbf{a}}_1^\mathsf{T} \\ \vdots \\ \bar{\mathbf{a}}_m^\mathsf{T} \end{bmatrix} \in \mathbf{R}^{m\times n} \quad \text{och} \quad \mathbf{b} = \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix} \in \mathbf{R}^m$$

Mängden $\mathcal{F} = \{\mathbf{x} \in \mathbf{R}^n : \mathbf{A}\mathbf{x} \geq \mathbf{b}\}$ är konvex. Detta följer eftersom

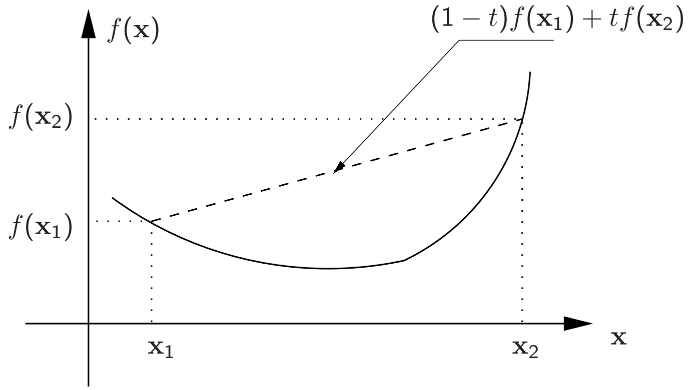
$$\mathcal{F} = \{\mathbf{x} \in \mathbf{R}^n : \bar{\mathbf{a}}_j^\mathsf{T}\mathbf{x} \geq b_j,\ j = 1, \dots, m\} = \cap_{j=1}^n \{\mathbf{x} \in \mathbf{R}^n : \bar{\mathbf{a}}_j^\mathsf{T}\mathbf{x} \geq b_j\}$$

Mängderna $\{\mathbf{x} \in \mathbf{R}^n : \bar{\mathbf{a}}_j^\mathsf{T}\mathbf{x} \geq b_j\}$ är halvplan och därmed uppenbart konvexa.

Konvexa funktioner

Definition 3. *Låt $C \subset \mathbf{R}^n$ vara en konvex mängd. En reellvärd funktion $f : C \to \mathbf{R}$ är konvex om för varje val av $\mathbf{x}_1 \in C$, $\mathbf{x}_2 \in C$ och $t \in (0,1)$*

$$f((1-t)\mathbf{x}_1 + t\mathbf{x}_2) \leq (1-t)f(\mathbf{x}_1) + tf(\mathbf{x}_2)$$



Sats 2. *Om $f_1, \dots, f_m$ är konvexa funktioner på C och $\gamma_1, \dots, \gamma_m$ är icke-negativa reella konstanter, så är funktionen*

$$f(\mathbf{x}) = \gamma_1 f_1(\mathbf{x}) + \ldots + \gamma_m f_m(\mathbf{x})$$

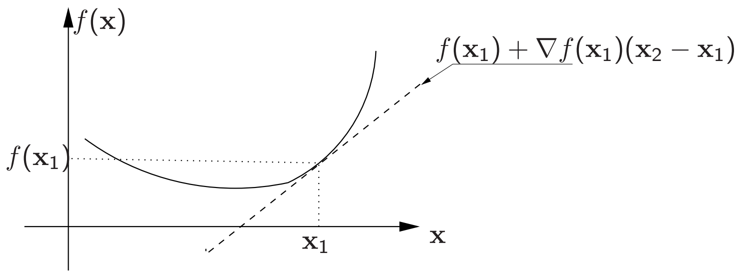
konvex på C.

Bevis: *Låt $\mathbf{x}_1, \mathbf{x}_2 \in C$ och $t \in (0,1)$. Enligt Definition 3 har vi att*

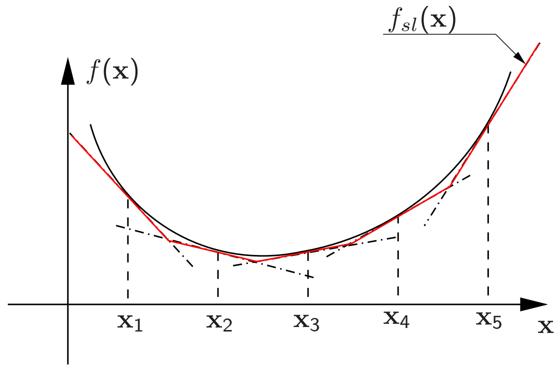
$$\begin{aligned} f((1-t)\mathbf{x}_1 + t\mathbf{x}_2) &= \sum_{k=1}^{m} \gamma_k f_k((1-t)\mathbf{x}_1 + t\mathbf{x}_2) \\ &\le \sum_{k=1}^{m} \gamma_k [(1-t) f_k(\mathbf{x}_1) + t f_k(\mathbf{x}_2)] \\ &= (1-t)\sum_{k=1}^{m} \gamma_k f_k(\mathbf{x}_1) + t \sum_{k=1}^{m} \gamma_k f_k(\mathbf{x}_2) \\ &= (1-t) f(\mathbf{x}_1) + t f(\mathbf{x}_2), \quad \text{vilket visar att } f \text{ är konvex.} \end{aligned}$$

Sats 3. *Låt $C \subset \mathbf{R}^n$ vara en konvex mängd. En kontinuerligt deriverbar funktion $f : C \to \mathbf{R}$ är konvex om och endast om för varje $\mathbf{x}_1 \in C$ och $\mathbf{x}_2 \in C$*

$$f(\mathbf{x}_2) \ge f(\mathbf{x}_1) + \nabla f(\mathbf{x}_1)(\mathbf{x}_2 - \mathbf{x}_1)$$



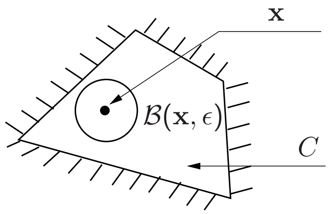
En direkt följd av föregående sats är att en konvex funktion kan underskattas med hjälp av en styckvis linjär konvex funktion.



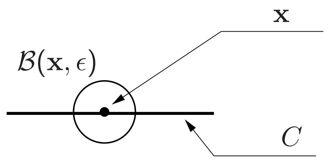
$$f_{sl}(\mathbf{x}) = \max_{\mathbf{x}}\{f(\mathbf{x}_k) + \nabla f(\mathbf{x}_k)(\mathbf{x} - \mathbf{x}_k)\}$$

Sats 4. *Antag att $C \subset \mathbf{R}^n$ är en given konvex mängd med minst en inre punkt. En två gånger kontinuerligt deriverbar funktion $f : C \to \mathbf{R}$ är då konvex om och endast om hessianen $\nabla^2 f(\mathbf{x})$ är positivt semidefinit för varje $\mathbf{x} \in C$.*

En punkt $\mathbf{x} \in C$ är en inre punkt om det finns en sfär med centrum i $\mathbf{x}$, $\mathcal{B}(\mathbf{x}, \epsilon) = \{\mathbf{y} \in \mathbf{R}^n : |\mathbf{x} - \mathbf{y}| < \epsilon\}$, så att $\mathcal{B}(\mathbf{x}, \epsilon) \subset C$.



Konvex mängd med inre punkter



Konvex mängd som saknar inre punkt

Notera att det i allmänhet är viktigt att funktionen f är definierad på en konvex delmängd $C \subset \mathbf{R}^n$.

Exempel 2. *Låt* $C = (0, \infty)$ *vilket är en konvex mängd. Funktionen* $f(x) = -\ln(x)$ *är konvex på* C *eftersom* $\nabla^2 f(\mathbf{x}) = \dfrac{1}{x^2}$ *är positiv för alla* $x \in C = (0, \infty)$. *Funktionen är inte definierad för* $x \leq 0$ *och är därmed inte konvex på* $C = (-\infty, \infty)$.

Exempel 3. *Funktionen* $f(x) = x^3$ *är konvex om den är definierad på* $C = [0, \infty)$, *men den är inte konvex om den är definierad på* $C = (-\infty, \infty)$.

Ibland finns det inga inre punkter och då är följande sats mer användbar.

Sats 5. *Antag att* $C \subset \mathbf{R}^n$ *är en given konvex mängd och att* $f : C \to \mathbf{R}$ *är en två gånger kontinuerligt deriverbar funktion på* C. *Då är* f *konvex på* C *om och endast om följande olikhet är uppfylld för varje val av* $\hat{\mathbf{x}} \in C$ *och* $\mathbf{x} \in C$:

$$(\mathbf{x} - \hat{\mathbf{x}})^{\mathsf{T}} \nabla^2 f(\hat{\mathbf{x}})(\mathbf{x} - \hat{\mathbf{x}}) \geq 0$$

Exempel 4. *Antag att*

$$f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^{\mathsf{T}}\mathbf{H}\mathbf{x} + \mathbf{c}^{\mathsf{T}}\mathbf{x}$$
$$C = \{\mathbf{x} \in \mathbf{R}^n : \mathbf{A}\mathbf{x} = \mathbf{b}\}.$$

Om $\mathbf{x}, \hat{\mathbf{x}} \in C$ *så följer att* $\mathbf{x} - \hat{\mathbf{x}} \in \mathcal{N}(\mathbf{A})$ *eftersom* $\mathbf{A}(\mathbf{x} - \hat{\mathbf{x}}) = \mathbf{b} - \mathbf{b} = \mathbf{0}$. *Om* $\mathbf{Z}$ *är en nollrumsmatris, dvs* $\mathbf{Z} = \begin{bmatrix} \mathbf{z}_1 & \dots & \mathbf{z}_k \end{bmatrix}$ *där kolonnerna utgör en bas till* $\mathcal{N}(\mathbf{A})$ *så följer att* $\mathbf{x} - \hat{\mathbf{x}} = \mathbf{Z}\mathbf{v}$ *för något* $\mathbf{v} \in \mathbf{R}^k$. *Således gäller att*

$$(\mathbf{x} - \hat{\mathbf{x}})^{\mathsf{T}} \nabla^2 f(\hat{\mathbf{x}})(\mathbf{x} - \hat{\mathbf{x}}) = \mathbf{v}^{\mathsf{T}}\mathbf{Z}^{\mathsf{T}}\mathbf{H}\mathbf{Z}\mathbf{v}$$

vilken är positiv för godtyckligt $\mathbf{v} \in \mathbf{R}^k$ *om och endast om den reducerade hessianen* $\mathbf{Z}^{\mathsf{T}}\mathbf{H}\mathbf{Z}$ *är positivt semidefinit. Enligt föregående sats är därför* f *konvex på* C *om och endast om* $\mathbf{Z}^{\mathsf{T}}\mathbf{H}\mathbf{Z}$ *är positivt semidefinit.*

Konvexa optimeringsproblem

Optimeringsproblemet

$$\begin{aligned} \text{minimera} \quad & f(\mathbf{x}) \\ \text{då} \quad & \mathbf{x} \in \mathcal{F} \end{aligned} \tag{2}$$

kallas konvext om

- det tillåtna området $\mathcal{F} \subset \mathbf{R}^n$ är konvext.
- målfunktionen $f : \mathcal{F} \to \mathbf{R}$ är konvex.

Exempel på konvexa optimeringsproblem

Linjär optimering:

$$\begin{aligned} \text{minimera} \quad & \mathbf{c}^\mathsf{T}\mathbf{x} \\ \text{då} \quad & \mathbf{A}\mathbf{x} \geq \mathbf{b} \end{aligned}$$

Vi har redan visat att $\mathcal{F} = \{\mathbf{x} \in \mathbf{R}^n : \mathbf{A}\mathbf{x} \geq \mathbf{b}\}$ är konvex och målfunktionen $f(\mathbf{x}) = \mathbf{c}^\mathsf{T}\mathbf{x}$ är konvex eftersom den uppfyller olikheten i Definition 3 med likhet.

Kvadratisk optimering:

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

Detta problemet är konvext om $\mathbf{H}$ är positivt semidefinit eftersom målfunktionen $f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^\mathsf{T}\mathbf{H}\mathbf{x} + \mathbf{c}^\mathsf{T}\mathbf{x}$ enligt Sats 4 då är konvex.

Kvadratisk optimering under linjära likhetsbivillkor:

$$\begin{aligned} \text{minimera} \quad & \tfrac{1}{2}\mathbf{x}^\mathsf{T}\mathbf{H}\mathbf{x} + \mathbf{c}^\mathsf{T}\mathbf{x} \\ \text{då} \quad & \mathbf{A}\mathbf{x} = \mathbf{b} \end{aligned}$$

Det är enkelt att visa att tillåtna området $\mathcal{F} = \{\mathbf{x} \in \mathbf{R}^n : \mathbf{A}\mathbf{x} = \mathbf{b}\}$ är konvext. Detta problem är därför enligt Exempel 4 konvext om den reducerade hessianen $\mathbf{Z}^\mathsf{T}\mathbf{H}\mathbf{Z}$ är positivt semidefinit.

Ickelinjär optimering:

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

Detta problem är konvext om $g_i : \mathbf{R}^n \to \mathbf{R}$, $i = 1, \ldots, n$ är konvexa funktioner och f är en reellvärd konvex funktion på

$$\mathcal{F} = \{\mathbf{x} \in \mathbf{R}^n : g_i(\mathbf{x}) \leq 0, \;\; i = 1, \ldots, m\}.$$

För att bevisa detta behöver vi visa att $\mathcal{F}$ är konvex. Eftersom

$$\mathcal{F} = \cap_{i=1}^m \{\mathbf{x} \in \mathbf{R}^n : g_i(\mathbf{x}) \leq 0\}$$

räcker det att visa att $\mathcal{F}_i = \{\mathbf{x} \in \mathbf{R}^n : g_i(\mathbf{x}) \leq 0\}$ är konvex.

Om $\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{F}_i$, så gäller att $g_i(\mathbf{x}_1) \leq 0$ och $g_i(\mathbf{x}_2) \leq 0$. Eftersom g_i är konvex följer att

$$g_i((1-t)\mathbf{x}_1 + t\mathbf{x}_2) \leq (1-t)g_i(\mathbf{x}_1) + t g_i(\mathbf{x}_2) \leq 0$$

för $t \in (0, 1)$. Detta visar att $(1-t)\mathbf{x}_1 + t\mathbf{x}_2 \in \mathcal{F}_i$ och $\mathcal{F}_i$ är därmed en konvex mängd.

Följande sats visar en ytterst trevlig egenskap för konvexa optimeringsproblem

Sats 6. *Om $\hat{\mathbf{x}} \in \mathcal{F}$ är en lokal optimallösning till det konvexa optimeringsproblemet (2) så är det också en global optimallösning.*

Bevis: *Antag att $\hat{\mathbf{x}} \in \mathcal{F}$ inte är en global optimallösning. Då finns minst ett $\mathbf{x} \in \mathcal{F}$ så att $f(\mathbf{x}) < f(\hat{\mathbf{x}})$. Eftersom $\mathcal{F}$ är konvex gäller att $\mathbf{x}(t) = (1-t)\hat{\mathbf{x}} + t\mathbf{x} \in \mathcal{F}$ för $t \in (0,1)$. Eftersom f är konvex har vi att*

$$f(\mathbf{x}(t)) \leq (1-t)f(\hat{\mathbf{x}}) + tf(\mathbf{x}) = f(\hat{\mathbf{x}}) + t(f(\mathbf{x}) - f(\hat{\mathbf{x}})) < f(\hat{\mathbf{x}})$$

Men eftersom $\mathbf{x}(t) = \hat{\mathbf{x}} + t(\mathbf{x} - \hat{\mathbf{x}})$ så gäller att godtyckligt nära $\hat{\mathbf{x}}$ (d.v.s. för små t) finns tillåtna punkter vars målfunktionsvärde är mindre än $f(\mathbf{x})$. Detta är omöjligt eftersom $\hat{\mathbf{x}}$ är en lokal optimallösning. Det måste alltså gälla att $f(\hat{\mathbf{x}}) \leq f(\mathbf{x})$ för alla $\mathbf{x} \in \mathcal{F}$.

Följande följdsats är mycket användbar.

Sats 7. *Antag att funktionen f är kontinuerligt deriverbar och konvex på $\mathbf{R}^n$. Då är $\hat{\mathbf{x}} \in \mathbf{R}^n$ en global minpunkt om och endast om $\nabla f(\hat{\mathbf{x}}) = \mathbf{0}$.*

Bevis: *Eftersom $\hat{\mathbf{x}} \in \mathbf{R}^n$ är en global minpunkt så är det också en lokal minpunkt varvid det följer från de nödvändiga villkoren för lokal optimalitet att $\nabla f(\hat{\mathbf{x}}) = \mathbf{0}^\mathsf{T}$.*

Antag omvänt att $\nabla f(\hat{\mathbf{x}}) = \mathbf{0}^\mathsf{T}$. Då gäller enligt Sats 3 att

$$f(\mathbf{x}) \geq f(\hat{\mathbf{x}}) + \underbrace{\nabla f(\hat{\mathbf{x}})}_{=0}(\mathbf{x} - \hat{\mathbf{x}}) = f(\hat{\mathbf{x}})$$

för alla $\mathbf{x} \in \mathbf{R}^n$. Detta visar att $\hat{\mathbf{x}}$ är en global minpunkt till f.

Exempel 5. *Betrakta optimeringsproblemet*

$$\text{minimera } f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^\mathsf{T}\mathbf{H}\mathbf{x} + \mathbf{c}^\mathsf{T}\mathbf{x}$$

där $\mathbf{H}$ är positivt semidefinit. Då är f en konvex funktion och minpunkten uppfyller

$$\nabla f(\hat{\mathbf{x}})^\mathsf{T} = \mathbf{H}\hat{\mathbf{x}} + \mathbf{c} = \mathbf{0}$$

Detta är samma optimalitetsvillkor som vi tidigare härlett med kvadratkomplettering.

Kommentarer

Sats 6 visar att varje lokal optimal lösning till ett konvext optimeringsproblem också är en global optimal lösning. Notera att optimallösningen inte behöver vara unik och att det inte nödvändigtvis existerar någon optimal lösning till ett konvext optimeringsproblem.

Exempel 6.

$$\begin{aligned} \text{minimera} \quad & x_1 + x_2 \\ \text{då} \quad & x_1 \geq 1 \end{aligned}$$

är nedåt obegränsad och saknar därmed optimal lösning.

Nedanstående konvexa optimeringsproblem saknar unik optimal lösning

$$\begin{aligned} \text{minimera} \quad & x_1 + x_2 \\ \text{då} \quad & x_1 + x_2 \geq 1 \end{aligned}$$

Läsanvisningar

- Ickelinjär optimering, grå häftet, sidan 18 och 25 samt halva sidan 26.
- Lite blandad optimeringsteori, röda häftet, sidan 3 samt sidorna 13-16.