KTH Matematik

5B1817 Tillämpad ickelinjär optimering

Föreläsning 2

Optimalitetsvillkor för problem med linjära bivillkor.

Optimalitetsvillkor för ickelinjära programmeringsproblem

Betrakta ett ickelinjärt programmeringsproblem

$$(P) \quad \begin{array}{ll} \min & f(x) \\ \text{då} & x \in F \subseteq \mathbb{R}^n, \end{array} \qquad \text{där } f \in C^2.$$

Definition. *En riktning p är en* tillåten riktning *till F i x^* om det finns $\bar{\alpha} > 0$ så att $x^* + \alpha p \in F$ för $\alpha \in [0, \bar{\alpha}]$.*

Definition. *En riktning p är en* descentriktning *till f i x^* om $\nabla f(x^*)^T p < 0$.*

Definition. *En riktning p är en* negativ krökningsriktning *till f i x^* om $p^T \nabla^2 f(x^*)^T p < 0$.*

Nödvändiga optimalitetsvillkor

$$(P) \quad \begin{array}{ll} \text{min} & f(x) \\ \text{då} & x \in F \subseteq \mathbb{R}^n, \end{array} \qquad \text{där } f \in C^2.$$

Första ordningens nödvändiga optimalitetsvillkor baseras på:

Sats. *Antag att $x^* \in F$. Om x^* är en lokal minpunkt till (P) så finns ingen tillåten riktning p till f i x^* sådan att $\nabla f(x^*)^T p < 0$.*

Andra ordningens nödvändiga optimalitetsvillkor baseras på:

Sats. *Antag att $x^* \in F$. Om x^* är en lokal minpunkt till (P) så finns ingen tillåten riktning p till f i x^* sådan att $\nabla f(x^*)^T p < 0$, och heller ingen tillåten riktning p till f i x^* sådan att $\nabla f(x^*)^T p = 0$ och $p^T \nabla^2 f(x^*) p < 0$.*

Följer av Taylorutveckling av f kring x^*.

Tillräckliga optimalitetsvillkor för problem med linjära bivillkor

$$(P) \quad \begin{array}{ll} \text{min} & f(x) \\ \text{då} & x \in F \subseteq I\!\!R^n, \end{array} \qquad \text{där } f \in C^2.$$

Tillräckliga optimalitetsvillkor är mer komplicerade. Vi kommer idag att studera problem utan bivillkor och problem med linjära bivillkor.

Exempel: $f = (x_2 - 2x_1^2)(x_2 - x_1^2)$, $F = I\!\!R^2$. Här är $x^* = (0\ 0)^T$ *inte* en lokal minpunkt, men f växer i alla riktningar från x^*.

Optimalitetsvillkor för problem utan bivillkor

Betrakta ett ickelinjärt programmeringsproblem utan bivillkor

$$(P) \quad \begin{array}{ll} \min & f(x) \\ \text{då} & x \in I\!R^n, \end{array} \qquad \text{där } f \in C^2.$$

Sats. (Första ordningens nödvändiga optimalitetsvillkor) *Om x^* är en lokal optimalpunkt till (P) så är $\nabla f(x^*) = 0$.*

Sats. (Andra ordningens nödvändiga optimalitetsvillkor) *Om x^* är en lokal optimalpunkt till (P) så är $\nabla f(x^*) = 0$ och $\nabla^2 f(x^*) \succeq 0$.*

Sats. (Andra ordningens tillräckliga optimalitetsvillkor) *Om $\nabla f(x^*) = 0$ och $\nabla^2 f(x^*) \succ 0$ så är x^* en lokal optimalpunkt till (P).*

Bevis. Se Nash och Sofer, sid. 296–298. $\square$

Optimalitetsvillkor för problem utan bivillkor, forts.

Diskussionen om optimalitetsvillkor baseras på Taylorutvecklingen

$$f(x^* + \alpha p) = f(x^*) + \alpha \nabla f(x^*)^T p + \frac{\alpha^2}{2} p^T \nabla^2 f(x^*) p + o(\alpha^2).$$

För en symmetrisk $n \times n$-matris M betecknar $M \succeq 0$ att M är positivt semidefinit. Analogt betecknar $M \succ 0$ att M är positivt definit.

En symmetrisk $n \times n$-matris M sägs vara positivt semidefinit om $p^T M p \geq 0$ för alla $p \in I\!R^n$. Det gäller att $M \succeq 0 \iff \eta_j(M) \geq 0$, $j = 1, \ldots, n$, där $\eta_j(M)$ betecknar Ms $j$te egenvärde.

Analogt sägs en symmetrisk $n \times n$-matris M vara positivt definit om $p^T M p > 0$ för alla $p \in I\!R^n$, $p \neq 0$. Det gäller att $M \succ 0 \iff \eta_j(M) > 0$, $j = 1, \ldots, n$.

Optimalitetsvillkor för problem med linjära likhetsbivillkor

Antag att vi har linjära likhetsbivillkor enligt

$$(P_=) \quad \begin{array}{ll} \min & f(x) \\ \text{då} & Ax = b, \ x \in I\!R^n, \end{array} \qquad \text{där } f \in C^2 \text{ och } A \text{ har full radrang.}$$

Låt $F = \{x \in I\!R^n : Ax = b\}$. Antag att $\bar{x}$ är en känd punkt i F, och låt x vara en godtycklig punkt i F. Då gäller $A(x - \bar{x}) = 0$, dvs $x - \bar{x} \in \text{null}(A)$.

Om Z betecknar en matris vars kolumner bildar bas för null(A), betyder det att $x - \bar{x} = Zv$ för något $v \in I\!R^{n-m}$.

Exempelvis, om $A = (B \ N)$, där B är $m \times m$ och inverterbar, kan vi välja $\bar{x} = \begin{pmatrix} B^{-1}b \\ 0 \end{pmatrix}$ och $Z = \begin{pmatrix} -B^{-1}N \\ I \end{pmatrix}$.

Exempel på Z-matriser

Låt $A = \begin{pmatrix} 1 & 2 & 3 \end{pmatrix}$.

$$B = \begin{pmatrix} 2 \end{pmatrix},\ N = \begin{pmatrix} 1 & 3 \end{pmatrix} \quad \Rightarrow \quad Z = \begin{pmatrix} 1 & 0 \\ -\frac{1}{2} & -\frac{3}{2} \\ 0 & 1 \end{pmatrix}.$$

Låt $A = \begin{pmatrix} 1 & 1 & 2 \\ 0 & 1 & 1 \end{pmatrix}$.

$$B = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix},\ N = \begin{pmatrix} 2 \\ 1 \end{pmatrix} \quad \Rightarrow \quad Z = \begin{pmatrix} -1 \\ -1 \\ 1 \end{pmatrix}.$$

OBS! Z ej unik. I Matlab ger "`null(A)`" en ortogonal Z.

Optimalitetsvillkor för problem med linjära likhetsbivillkor, forts.

Låt $\varphi(v) = f(\bar{x} + Zv)$. Vi kan skriva om problemet enligt

$$(P'_=) \qquad \begin{array}{ll} \text{min} & \varphi(v) \\ \text{då} & v \in I\!R^{n-m}. \end{array}$$

Derivering ger $\nabla\varphi(v) = Z^T\nabla f(\bar{x} + Zv)$, $\nabla^2\varphi(v) = Z^T\nabla^2 f(\bar{x} + Zv)Z$.

Detta är ett problem utan bivillkor, där vi känner optimalitetsvillkoren.

Därmed kan vi tillämpa dessa och identifiera $x^* = \bar{x} + Zv^*$, där v^* associeras med $(P'_=)$.

$Z^T\nabla f(x)$ kallas *reducerade gradienten* till f i x.

$Z^T\nabla^2 f(x)Z$ kallas *reducerade Hessianen* till f i x.

Nödvändiga optimalitetsvillkor för problem med linjära likhetsbivillkor

$$(P_=) \quad \begin{array}{ll} \min & f(x) \\ \text{då} & Ax = b, \ x \in I\!\!R^n, \end{array}$$

där $f \in C^2$ och A har full radrang.

Sats. (Första ordningens nödvändiga optimalitetsvillkor) *Om x^* är en lokal optimalpunkt till $(P_=)$ gäller att*

(i) $Ax^* = b$, *och*

(ii) $Z^T \nabla f(x^*) = 0$.

Sats. (Andra ordningens nödvändiga optimalitetsvillkor) *Om x^* är en lokal optimalpunkt till $(P_=)$ gäller att*

(i) $Ax^* = b$,

(ii) $Z^T \nabla f(x^*) = 0$, *och*

(iii) $Z^T \nabla^2 f(x^*) Z \succeq 0$.

Tillräckliga optimalitetsvillkor för problem med linjära likhetsbivillkor

$$(P_=) \quad \begin{array}{ll} \min & f(x) \\ \text{då} & Ax = b, \; x \in I\!R^n, \end{array} \qquad \text{där } f \in C^2 \text{ och } A \text{ har full radrang.}$$

Sats. (Andra ordningens tillräckliga optimalitetsvillkor) *Om*

(i) $Ax^* = b$,

(ii) $Z^T \nabla f(x^*) = 0$, *och*

(iii) $Z^T \nabla^2 f(x^*) Z \succ 0$,

så är x^* *en lokal optimalpunkt till* $(P_=)$.

Nollrum och bildrum

Påstående. *Låt* $c \in I\!R^n$, $A \in I\!R^{m\times n}$. *Antag att* $\text{rank}(A) = m$ *och låt* Z *vara en matris vars kolumner bildar en bas för* $\text{null}(A)$. *Då existerar* $\lambda \in I\!R^m$ *och* $u \in I\!R^{n-m}$ *så att* $c = A^T\lambda + Zu$.

Bevis. Då A^T och Z har full kolumnrang samt $AZ = 0$ får vi

$$\begin{pmatrix} A \\ Z^T \end{pmatrix} \begin{pmatrix} A^T & Z \end{pmatrix} = \begin{pmatrix} AA^T & 0 \\ 0 & Z^TZ \end{pmatrix} \succ 0.$$

Alltså är $(A^T \;\, Z)$ en ickesingulär $n \times n$-matris. $\square$

Detta medför att $Z^Tc = 0 \iff Z^TZu = 0 \iff u = 0 \iff c = A^T\lambda$.

Följaktligen gäller $Z^T\nabla f(x^*) = 0$ om och endast om $\nabla f(x^*) = A^T\lambda^*$ för något $\lambda^* \in I\!R^m$. Vi kallar λ^* *lagrangemultiplikatorvektor*.

Nödvändiga optimalitetsvillkor för problem med linjära likhetsbivillkor

$$(P_=) \quad \begin{array}{ll} \min & f(x) \\ \text{då} & Ax = b, \ x \in \mathbb{R}^n, \end{array} \qquad \text{där } f \in C^2 \text{ och } A \text{ har full radrang.}$$

Sats. (Första ordningens nödvändiga optimalitetsvillkor) *Om x^* är en lokal optimalpunkt till $(P_=)$ gäller att*

(i) $Ax^* = b$, *och*

(ii) $\nabla f(x^*) = A^T\lambda^*$ *för något* $\lambda^* \in \mathbb{R}^m$.

Sats. (Andra ordningens nödvändiga optimalitetsvillkor) *Om x^* är en lokal optimalpunkt till $(P_=)$ gäller att*

(i) $Ax^* = b$,

(ii) $\nabla f(x^*) = A^T\lambda^*$ *för något* $\lambda^* \in \mathbb{R}^m$, *och*

(iii) $Z^T\nabla^2 f(x^*)Z \succeq 0$.

Tillräckliga optimalitetsvillkor för problem med linjära likhetsbivillkor

$$(P_=) \quad \begin{array}{ll} \min & f(x) \\ \text{då} & Ax = b, \; x \in \mathbb{R}^n, \end{array}$$ där $f \in C^2$ och A har full radrang.

Sats. (Andra ordningens tillräckliga optimalitetsvillkor) *Om*

(i) $Ax^* = b$,

(ii) $\nabla f(x^*) = A^T \lambda^*$ *för något* $\lambda^* \in \mathbb{R}^m$, *och*

(iii) $Z^T \nabla^2 f(x^*) Z \succ 0$,

så är x^* *en lokal optimalpunkt till* $(P_=)$.

Optimalitetsvillkor för problem med linjära likhetsbivillkor, forts.

$$(P_=) \quad \begin{array}{ll} \text{min} & f(x) \\ \text{då} & Ax = b, \ x \in I\!R^n, \end{array}$$ där $f \in C^2$ och A har full radrang.

Om vi definierar lagrangefunktionen $\mathcal{L}(x, \lambda) = f(x) - \lambda^T(Ax - b)$ är första ordningens optimalitetsvillkor ekvivalenta med

$$\begin{pmatrix} \nabla_x \mathcal{L}(x^*, \lambda^*) \\ \nabla_\lambda \mathcal{L}(x^*, \lambda^*) \end{pmatrix} = \begin{pmatrix} \nabla f(x^*) - A^T \lambda^* \\ b - Ax^* \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

Alternativt, kravet är en punkt x^* sådan att $Ax^* = b$ där problemet

$$\begin{array}{ll} \text{min} & \nabla f(x^*)^T p \\ \text{då} & Ap = 0, \ p \in I\!R^n, \end{array}$$

har optimalvärde noll.

Optimalitetsvillkor för problem med linjära olikhetsbivillkor

Antag att vi har linjära olikhetsbivillkor enligt

$$(P_{\geq}) \qquad \begin{array}{ll} \text{min} & f(x) \\ \text{då} & Ax \geq b, \ x \in I\!\!R^n, \end{array} \qquad \text{där } f \in C^2.$$

Betrakta en tillåten punkt x^*. Partitionera $A = \begin{pmatrix} A_A \\ A_I \end{pmatrix}$, $b = \begin{pmatrix} b_A \\ b_I \end{pmatrix}$,

där $A_A x^* = b_A$ och $A_I x^* > b_I$.

Bivillkoren $A_A x \geq b_A$ är *aktiva* i x^*.

Bivillkoren $A_I x \geq b_I$ är *inaktiva* i x^*.

Optimalitetsvillkor för problem med linjära olikhetsbivillkor, forts.

Om x^* är en lokal optimalpunkt till $(P_\geq)$ får det inte finnas någon tillåten descentriktning i x^*. Därmed måste problemen

$$\begin{array}{llcll} \text{min} & \nabla f(x^*)^T p & \qquad & \text{max} & 0^T \lambda_A \\ \text{då} & A_A p \geq 0, & \qquad & \text{då} & A_A^T \lambda_A = \nabla f(x^*), \ \lambda_A \geq 0, \end{array}$$

ha optimalvärde noll. (Det andra problemet är LP-dualen till det första.) Alltså finns $\lambda_A^* \geq 0$ så att $A_A^T \lambda_A^* = \nabla f(x^*)$.

ödvändiga optimalitetsvillkor för problem med linjära olikhetsbivillko

$$(P_{\geq}) \quad \begin{array}{ll} \text{min} & f(x) \\ \text{då} & Ax \geq b, \; x \in I\!\!R^n, \end{array} \qquad \text{där } f \in C^2.$$

Sats. (Första ordningens nödvändiga optimalitetsvillkor) *Om* x^* *är en lokal optimalpunkt till* $(P_{\geq})$ *gäller att*

(i) $Ax^* \geq b$, *och*

(ii) $\nabla f(x^*) = A_A^T \lambda_A^*$ *för något* $\lambda_A^* \geq 0$,

där A_A *svarar mot de aktiva bivillkoren i* x^*.

Första ordningens nödvändiga optimalitetsvillkor kallas ofta *KKT-villkoren*.

ödvändiga optimalitetsvillkor för problem med linjära olikhetsbivillkc

$$(P_{\geq}) \quad \begin{array}{ll} \text{min} & f(x) \\ \text{då} & Ax \geq b, \ x \in I\!\!R^n, \end{array} \qquad \text{där } f \in C^2.$$

Sats. (Andra ordningens nödvändiga optimalitetsvillkor) *Om x^* är en lokal optimalpunkt till $(P_{\geq})$ gäller att*

(i) $Ax^* \geq b$,

(ii) $\nabla f(x^*) = A_A^T \lambda_A^*$ *för något* $\lambda_A^* \geq 0$, *och*

(iii) $Z_A^T \nabla^2 f(x^*) Z_A \succeq 0$,

där A_A svarar mot de aktiva bivillkoren i x^* och Z_A är en matris vars kolumner bildar bas i $\text{null}(A_A)$.

Villkor *(iii)* svarar mot bivillkoren $Ax \geq b$ ersätts med $A_A x = b_A$.

Tillräckliga optimalitetsvillkor för problem med linjära olikhetsbivillkor

$$(P_{\geq}) \quad \begin{array}{ll} \min & f(x) \\ \text{då} & Ax \geq b, \; x \in I\!R^n, \end{array} \qquad \text{där } f \in C^2.$$

Sats. (Andra ordningens tillräckliga optimalitetsvillkor) *Om*

(i) $Ax^* \geq b$,

(ii) $\nabla f(x^*) = A_A^T \lambda_A^*$ *för något* $\lambda_A^* \geq 0$*, och*

(iii) $Z_+^T \nabla^2 f(x^*) Z_+ \succ 0$,

så är x^* *en lokal minpunkt till* $(P_{\geq})$*, där* A_A *svarar mot de aktiva bivillkoren i* x^**,* A_+ *består av de rader ur* A_A *för vilka* λ_A^* *har positiva komponenter, och* Z_+ *är en matris vars kolumner bildar bas i* null(A_+).

Vi bortser här från alla bivillkor för vilka $\lambda_i^* = 0$.

Om $\lambda_A^* > 0$ blir $A_+ = A_A$ och $Z_+ = Z_A$.

ödvändiga optimalitetsvillkor för problem med linjära olikhetsbivillko

$$(P_{\geq}) \qquad \begin{array}{ll} \text{min} & f(x) \\ \text{då} & Ax \geq b, \;\; x \in I\!R^n, \end{array} \qquad \text{där } f \in C^2.$$

Första ordningens nödvändiga optimalitetsvillkor skrivs ofta med en m-dimensionell lagrangemultiplikatorvektor λ^*.

Sats. (Första ordningens nödvändiga optimalitetsvillkor) *Om* x^* *är en lokal optimalpunkt till* $(P_{\geq})$ *uppfyller* x^* *tillsammans med något* $\lambda^* \in I\!R^m$

(i) $Ax^* \geq b$,

(ii) $\nabla f(x^*) = A^T \lambda^*$,

(iii) $\lambda^* \geq 0$, *och*

(iv) $\lambda_i^*(a_i^T x^* - b_i) = 0, \;\; i = 1, \ldots, m$.

Nödvändiga optimalitetsvillkor för problem med linjära bivillkor

$$(P) \quad \begin{array}{ll} \text{min} & f(x) \\ \text{då} & a_i^T x \geq b_i, \ \ i \in \mathcal{I}, \\ & a_i^T x = b_i, \ \ i \in \mathcal{E}, \\ & x \in \mathbb{R}^n, \end{array} \qquad \text{där } f \in C^2.$$

Sats. (Första ordningens nödvändiga optimalitetsvillkor) *Om x^* är en lokal optimalpunkt till (P) uppfyller x^* tillsammans med något $\lambda^* \in \mathbb{R}^m$*

(i) $a_i^T x^* \geq b_i$, $i \in \mathcal{I}$, $a_i^T x^* = b_i$, $i \in \mathcal{E}$,

(ii) $\nabla f(x^*) = A^T \lambda^*$,

(iii) $\lambda_i^* \geq 0$, $i \in \mathcal{I}$, *och*

(iv) $\lambda_i^* (a_i^T x^* - b_i) = 0$, $i \in \mathcal{I}$.

Exempel, nödvändiga optimalitetsvillkor för LP-problem

Som specialfall kan vi titta på ett LP-problem

$$(LP) \quad \begin{array}{ll} \text{min} & c^T x \\ \text{då} & Ax = b, \; x \geq 0, \end{array} \qquad \text{där } A \in \mathbb{R}^{m \times n}.$$

Om vi kallar multiplikatorerna y och s får vi de "vanliga" villkoren.

Sats. (Första ordningens nödvändiga optimalitetsvillkor) *Om x^* är en lokal optimalpunkt till (LP) uppfyller x^* tillsammans med något $y^* \in \mathbb{R}^m$ och $s^* \in \mathbb{R}^n$*

(i) $Ax^* = b$, $x^* \geq 0$,

(ii) $c = A^T y^* + s^*$,

(iii) $s^* \geq 0$, *och*

(iv) $s_j^* x_j^* = 0, \; j = 1, \ldots, n.$