KTH Matematik

5B1817 Tillämpad ickelinjär optimering

Föreläsning 3

Optimalitetsvillkor för problem med ickelinjära bivillkor

Optimalitetsvillkor för ickelinjära programmeringsproblem

Betrakta ett ickelinjärt programmeringsproblem med likhetsbivillkor

$$(P_=) \quad \begin{array}{ll} \text{min} & f(x) \\ \text{då} & g(x) = 0, \end{array} \qquad \text{där } f, g \in C^2,\ g : \mathbb{R}^n \to \mathbb{R}^m.$$

Låt $F = \{x \in \mathbb{R}^n : g(x) = 0\}$.

En *tillåten kurva* till F i x^* är en två gånger kontinuerligt differentierbar funktion $x : \mathbb{R} \to \mathbb{R}^n$ sådan att $g(x(t)) = 0$, $t \in (-\epsilon, \epsilon)$, för något $\epsilon > 0$.

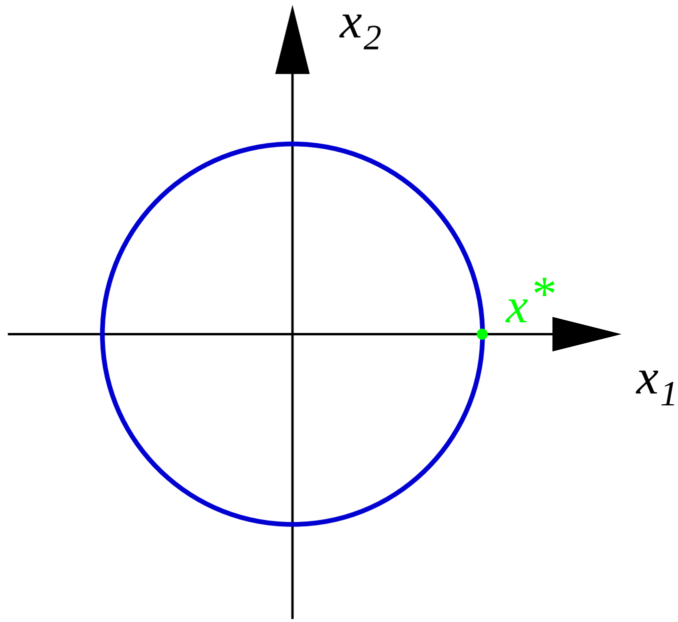
Tangentplanet $T(x^*)$ till F i x^* definieras som

$$T(x^*) = \{x'(0) : x(t) \text{ tillåten kurva till } F \text{ sådan att } x(0) = x^*\}.$$

Då $g_i(x(t)) \equiv 0$ för $t \in (-\epsilon, \epsilon)$ följer att $\nabla g_i(x(t))^T x'(t) \equiv 0$ för $t \in (-\epsilon, \epsilon)$. Speciellt får vi $\nabla g_i(x^*)^T x'(0) = 0$.

Exempel

Låt $g(x) = x_1^2 + x_2^2 - 1$ och $x^* = \begin{pmatrix} 1 & 0 \end{pmatrix}^T$.



Då är $x(t) = \begin{pmatrix} \cos t \\ \sin t \end{pmatrix}$ och $x(t) = \begin{pmatrix} \sqrt{1-t^2} \\ t \end{pmatrix}$ tillåtna kurvor med $x(0) = x^*$.

Optimalitetsvillkor för ickelinjära programmeringsproblem

Låt $A(x) = \begin{pmatrix} \nabla g_1(x)^T \\ \vdots \\ \nabla g_m(x)^T \end{pmatrix}$. Vi får $T(x^*) \subseteq \text{null}(A(x^*))$.

Låt $\varphi(t) = f(x(t))$. Om x^* är en lokal minpunkt till $(P_=)$ måste gälla att $\varphi'(0) = 0$. Men $\varphi'(t) = \nabla f(x(t))^T x'(t)$, vilket ger $\varphi'(0) = \nabla f(x^*)^T x'(0) = 0$.

Om $T(x^*) = \text{null}(A(x^*))$ följer att det till varje p sådant att $A(x^*)p = 0$ finns en tillåten kurva $x(t)$ sådan att $x(0) = x^*$, $x'(0) = p$. Om x^* är en lokal minpunkt måste därför $\nabla f(x^*)^T p = 0$ för alla $p \in \text{null}(A(x^*))$, dvs $\nabla f(x^*) = A(x^*)^T \lambda^*$ för något $\lambda^* \in I\!R^m$.

Speciellt om $g(x) = Ax - b$ får vi $A(x^*) = A$ och $T(x^*) = \text{null}(A)$.

Optimalitetsvillkor för ickelinjära programmeringsproblem

Om $T(x^*) = \text{null}(A(x^*))$ är lineariseringen av bivillkoren "tillräckligt bra" i x^* för att få optimalitetsvillkor analoga till problem med linjära bivillkor. Det ger vissa krav på bivillkoren, "constraint qualification".

Definition. *En punkt $x^* \in F$ är* reguljär *till $(P_=)$ om $A(x^*)$ har full radrang, dvs om $\nabla g_i(x^*)$, $i = 1, \ldots, m$, är linjärt oberoende.*

Antag att $g(x) = Ax - b$. Då blir $g(x^* + pt) = g(x^*) + Apt$.

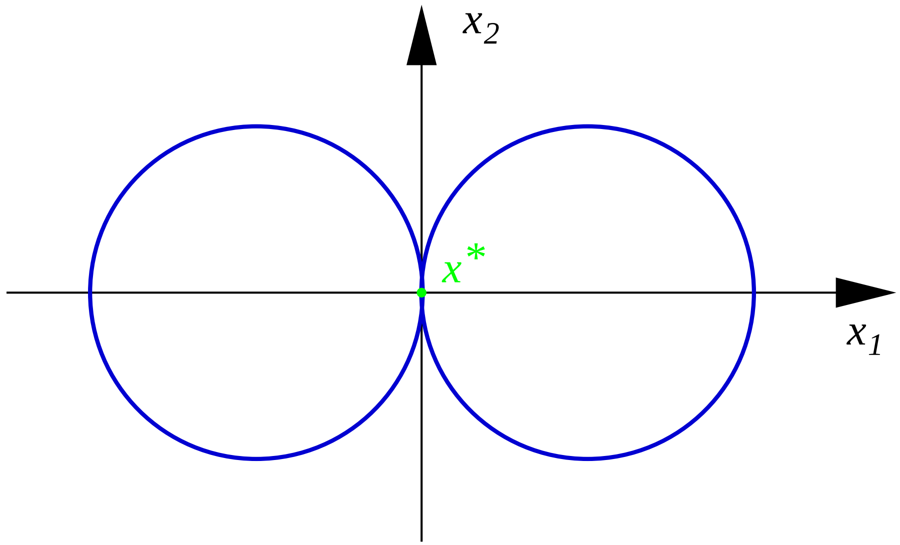
Påstående. *Antag att $x^* \in F$ är reguljär. Låt $p \in I\!\!R^n$. Då finns $\widetilde{x}(t)$ och $\epsilon > 0$ så att $\widetilde{x}(0) = x^*$, $\widetilde{x}'(0) = p$ och $g(\widetilde{x}(t)) = g(x^*) + A(x^*)pt$ för $t \in (-\epsilon, \epsilon)$.*

Bevis. Se Nash och Sofer, Lemma 14.6 (för fallet $A(x^*)p = 0$). $\square$

Corollary. *Om $x^* \in F$ är reguljär, så är $T(x^*) = \text{null}(A(x^*))$.*

Ett exempel där regularitet inte gäller

Låt $g(x) = \begin{pmatrix} (x_1 - 1)^2 + x_2^2 - 1 \\ (x_1 + 1)^2 + x_2^2 - 1 \end{pmatrix}$ med $x^* = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$.



Då är $A(x^*) = \begin{pmatrix} -2 & 0 \\ 2 & 0 \end{pmatrix}$ och $T(x^*) \neq \text{null}(A(x^*))$.

OBS! Inget bivillkor är redundant.

Nödvändiga optimalitetsvillkor för problem med likhetsbivillkor

$$(P_=) \quad \begin{array}{ll} \text{min} & f(x) \\ \text{då} & g(x) = 0, \end{array} \qquad \text{där } f, g \in C^2,\ g : I\!R^n \to I\!R^m.$$

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

(i) $g(x^*) = 0$, *och*

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

Nödvändiga optimalitetsvillkor för problem med likhetsbivillkor

Om x^* är en lokal minpunkt till $(P_=)$ måste gälla att $\varphi''(0) \geq 0$, där $\varphi(t) = f(x(t))$. Men $\varphi''(t) = x'(t)^T \nabla^2 f(x(t)) x'(t) + \nabla f(x(t))^T x''(t)$, vilket ger

$$\varphi''(0) = x'(0)^T \nabla^2 f(x^*) x'(0) + \nabla f(x^*)^T x''(0) \geq 0.$$

Eftersom $g(x(t)) = 0$ får vi analogt

$$x'(0)^T \nabla^2 g_i(x^*) x'(0) + \nabla g_i(x^*)^T x''(0) = 0, \;\; i = 1, \ldots, m,$$

vilket ger

$$\sum_{i=1}^{m} \lambda_i^* (x'(0)^T \nabla^2 g_i(x^*) x'(0) + \nabla g_i(x^*)^T x''(0)) = 0.$$

Nödvändiga optimalitetsvillkor för problem med likhetsbivillkor

Om x^* är reguljär kan vi använda λ^* från (ii), vilket ger

$$
\begin{aligned}
0 \le\ & x'(0)^T \nabla^2 f(x^*) x'(0) + \nabla f(x^*)^T x''(0) \\
& - \sum_{i=1}^{m} \lambda_i^* (x'(0)^T \nabla^2 g_i(x^*) x'(0) + \nabla g_i(x^*)^T x''(0)) \\
& = x'(0)^T \left(\nabla^2 f(x^*) - \textstyle\sum_{i=1}^m \lambda_i^* \nabla^2 g_i(x^*) \right) x'(0) \\
& + \left(\nabla f(x^*) - \textstyle\sum_{i=1}^m \lambda_i^* \nabla g_i(x^*) \right)^T x''(0).
\end{aligned}
$$

Eftersom $\nabla f(x^*) - \sum_{i=1}^m \lambda_i^* \nabla g_i(x^*) = 0$ följer att $x'(0)^T \nabla_{xx}^2 \mathcal{L}(x^*, \lambda^*) x'(0) \ge 0$, där $\mathcal{L}(x, \lambda) = f(x) - \lambda^T g(x)$.

Regularitet ger $p^T \nabla_{xx}^2 \mathcal{L}(x^*, \lambda^*) p \ge 0$ för alla $p \in \text{null}(A(x^*))$, vilket är ekvivalent med $Z(x^*)^T \nabla_{xx}^2 \mathcal{L}(x^*, \lambda^*) Z(x^*) \succeq 0$, där $Z(x^*)$ är en matris vars kolumner bildar bas för $\text{null}(A(x^*))$.

Nödvändiga optimalitetsvillkor för problem med likhetsbivillkor

$$(P_=) \qquad \begin{array}{ll} \min & f(x) \\ \text{då} & g(x) = 0, \end{array} \qquad \text{där } f, g \in C^2,\ g : I\!R^n \to I\!R^m.$$

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

(i) $g(x^*) = 0$,

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

(iii) $Z(x^*)^T \nabla^2_{xx} \mathcal{L}(x^*, \lambda^*) Z(x^*) \succeq 0$.

Tillräckliga optimalitetsvillkor för problem med likhetsbivillkor

$$(P_=) \quad \begin{array}{ll} \min & f(x) \\ \text{då} & g(x) = 0, \end{array} \qquad \text{där } f, g \in C^2,\ g : I\!R^n \to I\!R^m.$$

Sats. (Andra ordningens tillräckliga optimalitetsvillkor) *Om*

(i) $g(x^*) = 0$,

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

(iii) $Z(x^*)^T \nabla^2_{xx} \mathcal{L}(x^*, \lambda^*) Z(x^*) \succ 0$,

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

Tillräckliga optimalitetsvillkor för problem med likhetsbivillkor

Bevis. Antag motsatsen. Då finns $\{x^k\}_{k=1}^{\infty}$ så att $x^k = x^* + \alpha^k p^k$, $\|p^k\| = 1$, $f(x^k) < f(x^*)$, $g(x^k) = 0$, $k = 1, \ldots$, samt $\lim_{k\to\infty} \alpha_k = 0$.

Taylorutveckling ger, för $k \geq 1$,

$$0 > f(x^k) - f(x^*) = \alpha^k \nabla f(x^*)^T p^k + \frac{\alpha^{k2}}{2} p^{kT} \nabla^2 f(x^* + \theta_0^k \alpha^k p^k) p^k,$$

$$0 = g_i(x^k) - g_i(x^*) = \alpha^k \nabla g_i(x^*)^T p^k + \frac{\alpha^{k2}}{2} p^{kT} \nabla^2 g_i(x^* + \theta_i^k \alpha^k p^k) p^k,$$

$i = 1, \ldots, m$, för några $\theta_i^k \in (0, 1)$, $i = 0, \ldots, m$. Ur (ii) får vi

$$0 > p^{kT} \left(\nabla^2 f(x^* + \theta_0^k \alpha^k p^k) - \sum_{i=1}^m \lambda_i^* \nabla^2 g_i(x^* + \theta_i^k \alpha^k p^k) \right) p^k.$$

Vi kan anta att $\lim_{k\to\infty} p^k = p^*$. Då får vi $\|p^*\| = 1$, $A(x^*)p^* = 0$ och $p^{*T} \nabla^2_{xx} \mathcal{L}(x^*, \lambda^*) p^* \leq 0$, vilket motsäger (iii). $\square$

Optimalitetsvillkor för problem med olikhetsbivillkor

Antag att vi har olikhetsbivillkor enligt

$$(P_{\geq}) \qquad \begin{array}{ll} \min & f(x) \\ \text{då} & g(x) \geq 0, \ x \in I\!R^n, \end{array} \qquad \text{där } f, g \in C^2, \ g : I\!R^n \rightarrow I\!R^m.$$

Betrakta en tillåten punkt x^*. Partitionera $g(x^*) = \begin{pmatrix} g_A(x^*) \\ g_I(x^*) \end{pmatrix}$, där

$g_A(x^*) = 0$ och $g_I(x^*) > 0$. Partitionera $A(x^*)$ analogt.

Definition. *En punkt* $x^* \in I\!R^n$ *som är tillåten till* $(P_{\geq})$ *är* reguljär *till* $(P_{\geq})$ *om* $A_A(x^*)$ *har full radrang, dvs om* $\nabla g_i(x^*)$, $i \in \{l : g_l(x^*) = 0\}$, *är linjärt oberoende.*

Nödvändiga optimalitetsvillkor för problem med olikhetsbivillkor

$$(P_{\geq}) \quad \begin{array}{ll} \text{min} & f(x) \\ \text{då} & g(x) \geq 0, \; x \in I\!\!R^n, \end{array} \qquad \text{där } f, g \in C^2, \; g : I\!\!R^n \to I\!\!R^m.$$

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

(i) $g(x^*) \geq 0$, *och*

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

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

Nödvändiga optimalitetsvillkor för problem med olikhetsbivillkor

Bevis. Allt utom teckenkravet $\lambda_A^* \geq 0$ följer från likhetsfallet genom att ersätta $g(x) \geq 0$ med $g_A(x) = 0$.

Regulariteten medför att det finns p så att $A_A(x^*)p = e_i$ och det finns $\widetilde{x}(t)$ så att $\widetilde{x}(0) = x^*$, $\widetilde{x}'(0) = p$ och $g(\widetilde{x}(t)) = e_i t$. Låt $\varphi(t) = f(\widetilde{x}(t))$. Om x^* är en lokal minpunkt måste vi ha $\varphi'(0) \geq 0$, vilket ger $0 \leq \varphi'(0) = \nabla f(x^*)^T p = \lambda_A^{*\,T} A_A(x^*)p = (\lambda_A^*)_i$. $\square$

Nödvändiga optimalitetsvillkor för problem med olikhetsbivillkor

$$(P_{\geq}) \quad \begin{array}{ll} \text{min} & f(x) \\ \text{då} & g(x) \geq 0, \; x \in I\!R^n, \end{array} \qquad \text{där } f, g \in C^2, \; g: I\!R^n \to I\!R^m.$$

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

(i) $g(x^*) \geq 0$,

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

(iii) $Z_A(x^*)^T \nabla^2_{xx} \mathcal{L}(x^*, \lambda^*) Z_A(x^*) \succeq 0$,

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

Villkor *(iii)* svarar mot att bivillkoren $g(x) \geq 0$ ersätts med $g_A(x) = 0$.

Tillräckliga optimalitetsvillkor för problem med olikhetsbivillkor

$$(P_{\geq}) \quad \begin{array}{ll} \text{min} & f(x) \\ \text{då} & g(x) \geq 0, \; x \in I\!R^n, \end{array} \qquad \text{där } f, g \in C^2, \; g : I\!R^n \to I\!R^m.$$

Sats. (Andra ordningens tillräckliga optimalitetsvillkor) *Om*

(i) $g(x^*) \geq 0$,

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

(iii) $Z_+(x^*)^T \nabla^2_{xx} \mathcal{L}(x^*, \lambda^*) Z_+(x^*) \succ 0$,

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

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

Tillräckliga optimalitetsvillkor för problem med olikhetsbivillkor

Bevis. Vi betraktar problemet där $g(x) \geq 0$ ersätts med $g_i(x) \geq 0$, $i \in \{l : \lambda_l^* > 0\}$. Det påverkar inte (i), (ii), eller (iii), men strikt komplementaritet gäller.

Antag motsatsen. Då finns $\{x^k\}_{k=1}^{\infty}$ så att $x^k = x^* + \alpha^k p^k$, $\|p^k\| = 1$, $f(x^k) < f(x^*)$, $g_i(x^k) \geq 0$, $i \in \{l : \lambda_l^* > 0\}$, $k = 1, \ldots,$ samt $\lim_{k\to\infty} \alpha_k = 0$. Vi kan anta att $\lim_{k\to\infty} p^k = p^*$, med $\|p^*\| = 1$.

Antag först att $\nabla g_i(x^*))^T p^* > 0$ för något i sådant att $\lambda_i^* > 0$. Då får vi $\nabla f(x^*)^T p^* = \lambda_+^{*T} A_+(x^*) p^* > 0$. Men $f(x^k) < f(x^*)$ implicerar $\nabla f(x^*)^T p^* \leq 0$, vilket är en motsägelse.

Tillräckliga optimalitetsvillkor för problem med olikhetsbivillkor

Forts. av bevis. Annars gäller $A_+(x^*)p^* = 0$. Taylorutveckling ger, för $k \geq 1$,

$$0 > f(x^k) - f(x^*) = \alpha^k \nabla f(x^*)^T p^k + \frac{\alpha^{k2}}{2} p^{kT} \nabla^2 f(x^* + \theta_0^k \alpha^k p^k) p^k,$$

$$0 \leq g_i(x^k) - g_i(x^*) = \alpha^k \nabla g_i(x^*)^T p^k + \frac{\alpha^{k2}}{2} p^{kT} \nabla^2 g_i(x^* + \theta_i^k \alpha^k p^k) p^k,$$

$i \in \{l : \lambda_l^* > 0\}$, för några $\theta_i^k \in (0, 1)$, $i \in \{l : \lambda_l^* > 0\}$. Ur (ii) får vi

$$0 > p^{kT} \left(\nabla^2 f(x^* + \theta_0^k \alpha^k p^k) - \textstyle\sum_{i:\lambda_i^*>0} \lambda_i^* \nabla^2 g_i(x^* + \theta_i^k \alpha^k p^k) \right) p^k$$

vilket ger $p^{*T} \nabla_{xx}^2 \mathcal{L}(x^*, \lambda^*) p^* \leq 0$, i strid med (iii). $\square$

Första ordningens nödvändiga optimalitetsvillkor

$$(P) \quad \begin{array}{ll} \text{min} & f(x) \\ \text{då} & g_i(x) \geq 0,\ i \in \mathcal{I}, \\ & g_i(x) = 0,\ i \in \mathcal{E}, \\ & x \in I\!R^n, \end{array} \qquad \text{där } f, g \in C^2,\ g : I\!R^n \to I\!R^m.$$

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

(i) $g_i(x^*) \geq 0$, $i \in \mathcal{I}$, $g_i(x^*) = 0$, $i \in \mathcal{E}$,

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

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

(iv) $\lambda_i^* g_i(x^*) = 0$, $i \in \mathcal{I}$.

Konvexitet ger global optimalitet

$$(P) \quad \begin{array}{ll} \text{min} & f(x) \\ \text{då} & g_i(x) \geq 0,\ i \in \mathcal{I}, \quad g_i(x) = 0,\ i \in \mathcal{E}, \quad x \in I\!R^n, \end{array}$$

där $f, g \in C^2$, $g : I\!R^n \longrightarrow I\!R^m$.

Sats. *Antag att g_i, $i \in \mathcal{I}$, är konkava funktioner på $I\!R^n$ och g_i, $i \in \mathcal{E}$, är affina funktioner på $I\!R^n$. Antag att f är en konvex funktion på tillåtna området till (P). Om $x^* \in I\!R^m$ och $\lambda^* \in I\!R^m$ uppfyller*

(i) $g_i(x^*) \geq 0$, $i \in \mathcal{I}$, $g_i(x^*) = 0$, $i \in \mathcal{E}$,

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

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

(iv) $\lambda_i^* g_i(x^*) = 0$, $i \in \mathcal{I}$,

så är x^* en global optimalpunkt till (P).

Konvexitet ger global optimalitet

Bevis. Idé. Tag $y \in F$. De givna förutsättningarna ger relationerna

$$\begin{aligned} f(y) &\geq f(x^*) + \nabla f(x^*)^T (y - x^*) \\ &= f(x^*) + \lambda^{*T} A(x^*)(y - x^*) \\ &= f(x^*) + \textstyle\sum_{i \in I} \lambda_i^* \nabla g_i(x^*)^T (y - x^*) \\ &\geq f(x^*) + \textstyle\sum_{i \in I} \lambda_i^* (g_i(y) - g_i(x^*)) \geq f(x^*). \quad \square \end{aligned}$$

Alternativt kan man visa detta med hjälp av lagrangerelaxering.

Första ordningens nödvändiga optimalitetsvillkor ger alltså tillsammans med konvexitet global optimalitet.