Optimalitetsvillkor för bivillkorsproblem

För en minimerare $x_*$ till ett minimeringsproblem

$$\begin{array}{ll} \min\limits_{x} & f(x) \\ \text{s.t.} & g_i(x) = 0, \quad i \in \mathcal{E} \\ & g_i(x) \geq 0, \quad i \in \mathcal{I} \end{array}$$

gäller att

$$p^T \nabla f(x_*) \geq 0$$

för alla tillåtna riktningar $p$.



Optimalitetsvillkor för linjära likhetsbivillkorsproblem

Om vi endast har linjära likhetsbivillkor, dvs vi har problemet

$$\begin{array}{ll} \min\limits_{x} & f(x) \\ \text{s.t.} & Ax = b, \end{array}$$

där vi antar att $A$ har full rang, går detta att skriva om till

$$\min_{v \in \Re^r} \phi(v) = f(\bar{x} + Zv), \tag{1}$$

där $\bar{x}$ är någon tillåten punkt och $Z \in \Re^{n \times r}$ spänner upp $\mathcal{N}(A)$, vilket är ett problem utan bivillkor. $\phi(v)$ kallas för den *reducerade funktionen*.



Nödvändiga villkor för minimerare

Då problem (1) är ett problem utan bivillkor blir de nödvändiga villkoren för en minimerare:

$$\begin{aligned} \nabla \phi(v) &= Z^T \nabla f(\bar{x} + Zv) \\ &= Z^T \nabla f(x) = 0 \text{ och} \\ \nabla^2 \phi(v) &= Z^T \nabla^2 f(\bar{x} + Zv) Z \\ &= Z^T \nabla^2 f(x) Z \text{ positivt semidefinit,} \end{aligned}$$

där $x = \bar{x} + Zv$. $Z^T \nabla f(x)$ kallas för den *reducerade gradienten* och $Z^T \nabla^2 f(x) Z$ för den *reducerade hessianen*.

Andragradsvillkoret motsvarar

$$v^T Z^T \nabla^2 f(x_*) Z v \geq 0 \quad \forall v,$$

vilket kan skrivas om till

$$p^T \nabla^2 f(x_*) p \geq 0 \quad \forall p \in \mathcal{N}(A),$$

där $p = Zv$, dvs hessianen i $x_*$ måste vara positivt semidefinit på nollrummet till $A$.

Notera att detta inte innebär att hessianen i $x_*$ måste vara positivt semidefinit på hela $\Re^n$.



Tillräckliga villkor för minimerare

Analogt blir de *tillräckliga* villkoren på en minimerare $x_*$:

- $Ax_* = b$,
- $Z^T \nabla f(x_*) = 0$ och
- $Z^T \nabla^2 f(x_*) Z$ positivt definit.

Lagranges multiplikatorer

Låt $x_*$ vara en minimerare och $Z \in \Re^{n\times r}$ en nollrumsmatris för $A$. Om vi delar upp $\nabla f(x_*)$ i sina nollrums- och värderumskomponenter får vi

$$\nabla f(x_*) \;=\; Zv_* + A^T\lambda_*,$$

där $v_* \in \Re^r$ och $\lambda_* \in \Re^m$.

Tillsammans med förstagradvillkoret ger detta

$$\begin{aligned} Z^T\nabla f(x_*) &= Z^T Z v_* + Z^T A^T \lambda_* \\ &= Z^T Z v_* + (\underbrace{AZ}_{=0})^T\lambda_* \\ &= Z^T Z v_* = 0 \\ &\Downarrow \\ Zv_* &= 0. \end{aligned}$$



Lagranges multiplikatorer, forts.

För en minimerare $x_*$ gäller alltså att

$$\nabla f(x_*) = A^T\lambda_*$$

dvs i ett lokalt minimum är gradienten till objektfunktionen en linjärkombination av gradienterna till bivillkoren. Vektorn $\lambda_*$ innehåller koefficienterna och kallas för *Lagranges multiplikatorer*.



Lagranges multiplikatorer, forts.

Lagranges multiplikatorer har betydelse för att uppskatta känsligheten i minvärdet $f(x_*)$. Antag att vi funnit en lösning $x_*$ till

$$\begin{array}{ll} \text{min} & f(x) \\ s.t. & Ax = b. \end{array}$$

Antag nu att bivillkoren blivit "störda" till $Ax = b+\delta$. Om störningen $\delta$ är tillräckligt liten kommer lösningen $\bar{x}$ till det störda problemet att ligga nära $x_*$ och

$$\begin{aligned} f(\bar{x}) &\approx f(x_*) + (\bar{x}-x_*)^T\nabla f(x_*) \\ &= f(x_*) + (\bar{x}-x_*)^T A^T\lambda_* \\ &= f(x_*) + (A\bar{x}-Ax_*)^T\lambda_* \\ &= f(x_*) + (b+\delta-b)^T\lambda_* \\ &= f(x_*) + \delta^T\lambda_* \\ &= f(x_*) + \sum_{i=1}^{m} \delta_i\lambda_{*i} \end{aligned}$$

för $\bar{x}$ nära $x_*$.



Lagranges multiplikatorer, forts.

Om element $i$ i högerledet av bivillkoren modifieras med $\delta_i$ kommer alltså funktionsvärdet att ändras med ungefär $\delta_i\lambda_{*i}$. Därför kallas Lagranges multiplikatorer ibland *skuggpriser* eller *duala variabler*.

Vi kan också sätta ihop bivillkoren och förstagradsvillkoret på en minimerare och få ett problem med $n+m$ ekvationer och $n+m$ variabler i $x$ och $\lambda$:

$$\begin{array}{lcl} \nabla f(x) - A^T\lambda &=& 0 \\ Ax &=& b \end{array}$$

Lagranges funktion

Om vi konstruerar *Lagrange-funktionen* av $x$ och $\lambda$

$$\mathcal{L}(x,\lambda) = f(x) - \sum_{i=1}^{m} \lambda_i (a_i^T x - b_i) = f(x) - \lambda^T (Ax - b),$$

och konstaterar att

$$\nabla_x \mathcal{L}(x,\lambda) = \nabla f(x) - A^T \lambda$$

och

$$\nabla_\lambda \mathcal{L}(x,\lambda) = b - Ax,$$

så kan vi beskriva första ordningens villkor på en minimerare till ett problem med linjära bivillkor som

$$\nabla \mathcal{L}(x_*, \lambda_*) = 0,$$

dvs en lokal minpunkt till bivillkorsproblemet är en stationär punkt till Lagranges funktion.



Optimalitetsvillkor för linjära olikhetsbivillkorsproblem

Om vi har ett problem med linjära olikhetsbivillkor

$$\begin{array}{ll} \min\limits_x & f(x) \\ \text{s.t.} & Ax \geq b, \end{array}$$

där vi antar att $A$ har full rang, så kommer de aktiva bivillkoren vid en punkt $x_*$ att vara avgörande för om $x_*$ är en minimerare. Vårt problem blir då

$$\begin{array}{ll} \min\limits_x & f(x) \\ \text{s.t.} & \widehat{A}x = \widehat{b}, \end{array}$$

där $\widehat{A}$ och $\widehat{b}$ innehåller de aktiva bivillkoren. Om $Z$ är nollrumsmatris till $\widehat{A}$ så blir förstagradsvillkoren

$$Z^T \nabla f(x_*) = 0 \text{ eller } \nabla f(x_*) = \widehat{A}^T \widehat{\lambda}_*,$$

där $\widehat{\lambda}_* \geq 0$ innehåller Lagrangemultiplikatorerna för de aktiva bivillkoren. Andra ordningens nödvändiga villkor innebär att $Z^T \nabla^2 f(x_*) Z$ måste vara positivt semidefinit.



Komplementaritet

Om vi definierar en multiplikator till ett inaktivt bivillkor att vara noll, kan vi beskriva olikhetsbivillkoren med hjälp av likhet:

$$\lambda_{*i}(a_i^T x_* - b_i) = 0, \;\; i = 1, \ldots, m.$$

Detta kallas för *komplementär slackhet* och innebär att bivillkoret är aktivt ($a_i^T x_* - b_i = 0$) och/eller Lagrangemultiplikatorn är noll ($\lambda_{*i} = 0$). Om detta inte kan gälla samtidigt sägs *strikt komplementariet* hålla. Annars måste någon Lagrangemultiplikator vara noll och dess bivillkor sägs vara *degenererat*.



Nödvändiga villkor, linjära olikhetsbivillkor

Sammanfattningsvis måste följande gälla för att en punkt $x_*$ ska vara en minimerare för en funktion $f$ på mängden $\{x : Ax \geq b\}$:

För någon vektor $\lambda_*$ av Lagrangemultiplikatorer:

- $Ax_* \geq b$,
- $\nabla f(x_*) = A^T \lambda_* \Leftrightarrow Z^T \nabla f(x_*) = 0$,
- $\lambda_* \geq 0$,
- $\lambda_*^T (Ax_* - b) = 0$ och
- $Z^T \nabla^2 f(x_*) Z$ positivt semidefinit,

där $Z$ är en nollrumsmatris för matrisen $\widehat{A}$ av de aktiva bivillkoren i $x_*$.

## Tillräckliga villkor, linjära olikhetsbivillkor I

Om vi har strikt komplementaritet går det att utöka till tillräckliga villkor på sedvanligt sätt:

- $Ax_* \geq b$,
- $\nabla f(x_*) = A^T \lambda_*$,
- $\lambda_* \geq 0$,
- strikt komplementaritet gäller och
- $Z^T \nabla^2 f(x_*) Z$ positivt definit.



## Tillräckliga villkor, linjära olikhetsbivillkor I, forts.

Notera att $x_*$ är en strikt lokal minimerare på mängden $\{x : \widehat{A}x = \widehat{b}\}$, dvs $f$ ökar i alla riktningar $p$ sådana att $\widehat{A}p = 0$. Studera en riktning $p$ sådan att $\widehat{A}p \geq 0$, där någon komponent är strikt positiv, dvs $p$ pekar in i den tillåtna mängden. Då $\nabla f(x_*) = A^T \lambda_* = \widehat{A}^T \widehat{\lambda}_*$ blir

$$p^T \nabla f(x_*) = p^T \widehat{A}^T \widehat{\lambda}_* = (\widehat{A}p)^T \widehat{\lambda}_*.$$

Med strikt komplementaritet kommer $(\widehat{A}p)^T \widehat{\lambda}_* > 0$, dvs $p$ är en uppåtriktning och $x_*$ därmed en strikt minimerare.

Utan strikt komplementaritet kan $(\widehat{A}p)^T \widehat{\lambda}_* = 0$ för något $p$, vilket innebär att vi kan inte säga något om $x_*$ med endast förstagradsinformation.

Om vi däremot släpper de degenererade villkoren kan vi formulera tillräckliga villkor baserade på de resterande bivillkoren:



## Tillräckliga villkor, linjära olikhetsbivillkor II

Låt $\widehat{A}_+$ innehålla de rader i $\widehat{A}$ som motsvarar de icke degenererade villkoren i $x_*$. Låt $Z_+$ vara nollrumsmatris till $\widehat{A}_+$. Om $x_*$ då uppfyller

- $Ax_* \geq b$,
- $\nabla f(x_*) = A^T \lambda_*$,
- $\lambda_* \geq 0$,
- $\lambda_*^T (Ax_* - b) = 0$ och
- $Z_+^T \nabla^2 f(x_*) Z_+$ positivt definit,

så är $x_*$ en strikt lokal minimerare till olikhetsbivillkorsproblemet.



## Optimalitetsvillkor för icke-linjära bivillkor

Vi ska studera icke-linjära optimeringsproblem med icke-linjära bivillkor. Ett sådant skrivs

$$\begin{array}{ll} \min\limits_x & f(x) \\ \text{s.t.} & g_i(x) = 0, \;\; i = 1, \dots, m \end{array}$$

för likhetsbivillkor och

$$\begin{array}{ll} \min\limits_x & f(x) \\ \text{s.t.} & g_i(x) \geq 0, \;\; i = 1, \dots, m \end{array}$$

för olikhetsbivillkor.

Vi antar att lösningspunkten $x_*$ är reguljär, dvs att gradienterna till de aktiva bivillkoren i $x_*$ $\{\nabla g_i(x_*) : g_i(x_*) = 0\}$ är linjärt oberoende.

Optimalitetsvillkoren uttrycks i termer av Lagrange-funktionen

$$\mathcal{L}(x, \lambda) = f(x) - \sum_{i=1}^{m} \lambda_i g_i(x) = f(x) - \lambda^T g(x),$$

där $\lambda$ är en vektor med Lagrange-multiplikatorer och $g$ är en vektor av bivillkorsfunktioner $\{g_i\}$.

Nödvändiga villkor för likhetsbivillkor

Låt $x_*$ vara en lokal minimerare för $f$ under bivillkoren $g(x) = 0$ och $Z(x_*)$ vara en nollrumsmatris för jacobianen $\nabla g(x_*)^T$. Om $x_*$ är en reguljär punkt så existerar det en Lagrange-vektor $\lambda_*$ sådan att

- $\nabla_x \mathcal{L}(x_*, \lambda_*) = 0 \Leftrightarrow Z(x_*)^T \nabla f(x_*) = 0$ och
- $Z(x_*)^T \nabla^2_{xx} \mathcal{L}(x_*, \lambda_*) Z(x_*)$ positivt semidefinit.



Tillräckliga villkor för likhetsbivillkor

Låt $x_*$ vara en punkt sådan att $g(x_*) = 0$ och $Z(x_*)$ vara en bas för nollrummet till $\nabla g(x_*)^T$. Om det existerar en vektor $\lambda_*$ sådan att

- $\nabla_x \mathcal{L}(x_*, \lambda_*) = 0$ och
- $Z(x_*)^T \nabla^2_{xx} \mathcal{L}(x_*, \lambda_*) Z(x_*)$ är positivt definit,

så är $x_*$ en strikt lokal minimerare till $f$ på bivillkorsmängden $\{x : g(x) = 0\}$.



Linjära likhetsbivillkor

För linjära bivillkor $g(x) = Ax - b$ blir jacobianen $\nabla g(x)^T = A$.

Första ordningens villkor

$$Z(x_*)^T \nabla f(x_*) = 0 \quad \Leftrightarrow$$

$$\nabla_x \mathcal{L}(x_*, \lambda_*) = \nabla f(x_*) - \nabla g(x_*) \lambda = 0$$

blir då

$$Z^T \nabla f(x_*) = 0 \quad \Leftrightarrow \quad \nabla f(x_*) = A^T \lambda_*.$$

Andra ordningens villkor blir också lika, ty

$$\nabla^2_{xx} \mathcal{L}(x_*, \lambda_*) = \nabla^2 f(x_*).$$



Nödvändiga villkor för olikhetsbivillkor

Låt $x_*$ vara en lokal minimerare för $f$ under bivillkoren $g(x) \geq 0$ och $Z(x_*)$ vara en nollrumsmatris för jacobianen av de aktiva bivillkoren i $x_*$. Om $x_*$ är en reguljär punkt så existerar det en Lagrange-vektor $\lambda_*$ sådan att

- $\nabla_x \mathcal{L}(x_*, \lambda_*) = 0 \Leftrightarrow Z(x_*)^T \nabla f(x_*) = 0$,
- $\lambda_* \geq 0$,
- $\lambda_*^T g(x_*) = 0$ och
- $Z(x_*)^T \nabla^2_{xx} \mathcal{L}(x_*, \lambda_*) Z(x_*)$ positivt semidefinit.

Villkoret $\lambda_*^T g(x_*) = 0$ är den icke-linjära versionen av det komplementära slackhetsvillkoret.

## Tillräckliga villkor för olikhetsbivillkor

Låt $x_*$ vara en punkt sådan att $g(x_*) \geq 0$. Om det existerar en vektor $\lambda_*$ sådan att

- $\nabla_x \mathcal{L}(x_*, \lambda_*) = 0$,
- $\lambda_* \geq 0$,
- $\lambda_*^T g(x_*) = 0$ och
- $Z_+(x_*)^T \nabla^2_{xx} \mathcal{L}(x_*, \lambda_*) Z_+(x_*)$ är positivt definit,

där $Z_+(x_*)$ är en bas för nollrummet av jacobianen till de icke-degenererade bivillkoren i $x_*$, så är $x_*$ en strikt lokal minimerare till $f$ på bivillkorsmängden $\{x : g(x) \geq 0\}$.

De tillräckliga och nödvända villkoren för icke-linjära olikhetsbivillkor kallas ibland för *Karush-Kuhn-Tucker-villkoren* (KKT-villkoren).



## Dualitet

Dualitet innebär att det till ett givet min-problem finns ett annat max-problem som under vissa förutsättningar har samma optimum.

Definiera

$$\mathcal{F}^*(x) = \max_{y \in Y} \mathcal{F}(x, y)$$

och

$$\mathcal{F}_*(y) = \min_{x \in X} \mathcal{F}(x, y).$$

Problemet

$$\min_{x \in X} \mathcal{F}^*(x) = \min_{x \in X} \max_{y \in Y} \mathcal{F}(x, y)$$

kallas för ett min-max-problem och problemet

$$\max_{y \in Y} \mathcal{F}_*(y) = \max_{y \in Y} \min_{x \in X} \mathcal{F}(x, y)$$

för ett max-min-problem. Dessa problem sägs vara varandras *dualer* och vi refererar till min-max-problemet som det *primala* problemet och $\mathcal{F}^*(x)$ för den *primala funktionen*. Max-min-problemet kallar vi det *duala* problemet och $\mathcal{F}_*(y)$ för den *duala funktionen*.



## Dualitet, forts.

För varje $x \in X$ och $y \in Y$ gäller att

$$\mathcal{F}_*(y) = \min_{x \in X} \mathcal{F}(x, y) \leq \mathcal{F}(x, y) \leq \max_{y \in Y} \mathcal{F}(x, y) = \mathcal{F}^*(x)$$

eller

$$\mathcal{F}_*(y) \leq \mathcal{F}^*(x).$$

Detta kallas för *svag dualitet*. En följd av detta är att det primala problemet är begränsat underifrån av $\mathcal{F}_*(y)$.

En punkt $(x_*, y_*)$ uppfyller *sadelpunktsvillkoret* m a p $\mathcal{F}$ om

$$\mathcal{F}(x_*, y) \leq \mathcal{F}(x_*, y_*) \leq \mathcal{F}(x, y_*)$$

för alla $x \in X$ och $y \in Y$. Om det existerar en punkt $(x_*, y_*)$ som uppfyller det så är lösningen till det primala och duala problemet lika, d v s

$$\min_{x \in X} \max_{y \in Y} \mathcal{F}(x, y) = \max_{y \in Y} \min_{x \in X} \mathcal{F}(x, y).$$

Detta kallas för *stark dualitet*.



## Dualitet och Lagranges multiplikatorer

Studera det icke-linjära problemet

$$\begin{array}{ll} \min\limits_{x} & f(x) \\ \text{s.t.} & g_i(x) \geq 0, \quad i = 1, \ldots, m \end{array}$$

och dess motsvarande Lagrange-funktion

$$\mathcal{L}(x, \lambda) = f(x) - \lambda^T g(x),$$

där $x \in \Re^n$, $\lambda \in \Re^m, \lambda \geq 0$.

Dualitet och Lagranges multiplikatorer, forts.

Definiera den primala funktionen

$$\mathcal{L}^*(x) = \max_{\lambda \geq 0} \mathcal{L}(x, \lambda)$$

och studera $\mathcal{L}^*(x)$ för ett fixt $x$:

$$\mathcal{L}^*(x) = \max_{\lambda \geq 0} \left( f(x) - \lambda^T g(x) \right).$$

För en tillåten punkt kommer $g(x) \geq 0$ och $\mathcal{L}^*(x) = f(x)$. För en otillåten punkt kommer $g_i(x) < 0$ för något $i$ och $\mathcal{L}^*(x)$ saknar övre begränsning.

Om vi formulerar det primala problemet som

$$\min_x \mathcal{L}^*(x)$$

så blir det samma som vårt ursprungliga problem med bivillkor.



Dualitet och Lagranges multiplikatorer, forts.

Vi kan nu använda min-max-dualitet och formulera det duala problemet. För varje $\lambda \geq 0$ så definieras den duala funktionen

$$\mathcal{L}_*(\lambda) = \min_x \mathcal{L}(x, \lambda)$$

och det duala max-min-problemet blir

$$\max_{\lambda \geq 0} \mathcal{L}_*(\lambda).$$

Om vi vill kan vi lösa det duala problemet i stället för det primala.