Presentation outline

1. Métodos numéricos para optimização não linear sem restrições
   - Métodos de procura directa
   - Métodos de procura unidireccional (geral)
   - Métodos de procura unidireccional (descida máxima, Newton)
   - Métodos de região de confiança
2. Teoria da optimização não linear com restrições
   - Introdução e qualificações de restrições
   - Condições de primeira ordem
   - Condições de segunda ordem
   - Teoria da dualidade
3. Métodos numéricos para optimização não linear com restrições
   - Método da penalização quadrática
   - Método do Lagrangeano aumentado
   - Programação Sequencial Quadrática (SQP)
   - Métodos de pontos interiores

Consideremos o **problema ou programa de Optimização Não Linear**

$$\boxed{\begin{array}{rrcll} \displaystyle\min_{x\in\mathbb{R}^n} & f(x) & & & \\ \text{sujeito a} & c_i(x) & = & 0, & i\in\mathcal{E} \\ & c_i(x) & \geq & 0, & i\in\mathcal{I} \end{array}} \qquad \text{PNL}$$

onde $f(x)$ é a **função objectivo**, $c_i(x) = 0, i \in \mathcal{E}$, as **restrições de igualdade** e $c_i(x) \geq 0, i \in \mathcal{I}$, as **restrições de desigualdade**.

A **região admissível** Ω é definida pelo conjunto de pontos que satisfazem todas as restrições

$$\Omega = \{x \in \mathbb{R}^n : \ c_i(x) = 0, i \in \mathcal{E} \quad c_i(x) \geq 0, i \in \mathcal{I}\}$$

Definição

Seja $x \in \Omega$.

Uma restrição $i \in \mathcal{I}$ *diz-se activa em* x *se* $c_i(x) = 0$.

O conjunto das restrições activas $\mathcal{A}(x)$ *é constituído pelos índices de* $\mathcal{E}$ *e pelos índices de* $\mathcal{I}$ *das restrições activas em* x, *ou seja,*

$$\mathcal{A}(x) = \mathcal{E} \cup \{i \in \mathcal{I} : \; c_i(x) = 0\}$$

Nota:

- Apenas as restrições activas são consideradas na caracterização da optimalidade local.

Exemplo:

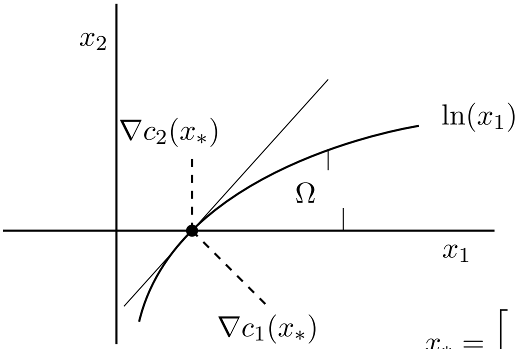
$$\begin{array}{rl} \min\limits_{x \in \mathbb{R}^2} & x_1 + x_2 \\ \text{sujeito a} & \ln(x_1) - x_2 \; \geq \; 0 \\ & x_2 \; \geq \; 0 \end{array}$$

De acordo com a notação anterior, temos

$$c_1(x) = \ln(x_1) - x_2 \quad \text{e} \quad c_2(x) = x_2$$

Assim,

$$\nabla c_1(x) = \begin{bmatrix} 1/x_1 \\ -1 \end{bmatrix} \quad \text{e} \quad \nabla c_2(x) = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$



$$x_* = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

$$\mathcal{A}(x_*) = \{1, 2\}$$

Definição

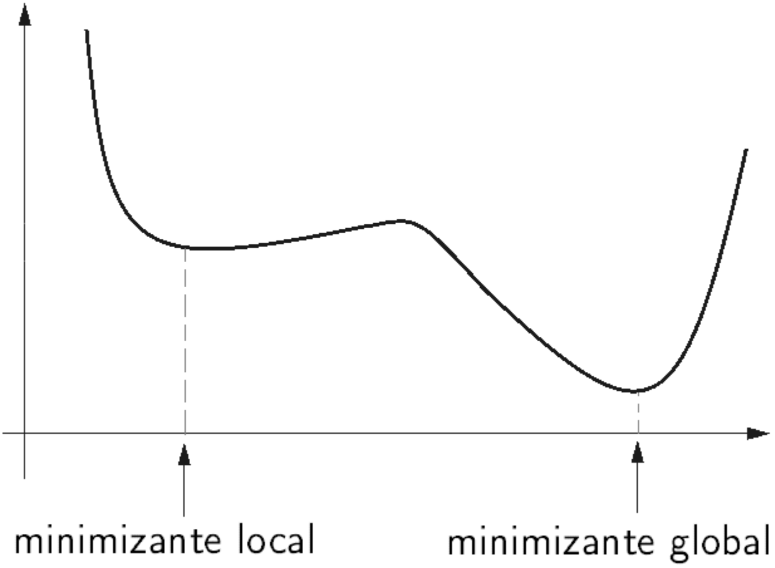
x_* diz-se um minimizante local (estrito) de f em Ω se existir uma vizinhança N de x_* tal que

$$f(x_*) \ \leq (<) \ f(x) \quad \forall x \in (N \cap \Omega) \setminus \{x_*\}$$

Definição

x_* diz-se um minimizante global (estrito) de f em Ω se

$$f(x_*) \ \leq (<) \ f(x) \quad \forall x \in \Omega \setminus \{x_*\}$$


minimizante local
minimizante global

Definição

Seja $x_* \in \Omega$. Diz-se que $\{z_k\}$ é uma sucessão admissível a aproximar-se de x_* se

$$z_k \longrightarrow x_* \quad \text{e} \quad z_k \in \Omega \text{ para } k \text{ suf. grande.}$$

Definição

Seja $x_* \in \Omega$. Um vector $d \in \Omega$ diz-se tangente a Ω em x_* se existir uma sucessão admissível $\{z_k\}$ a aproximar-se de x_* e uma sucessão $\{t_k\}$ em que $t_k > 0$ e $t_k \downarrow 0$ tais que

$$\lim_{k \to +\infty} \frac{z_k - x_*}{t_k} = d$$

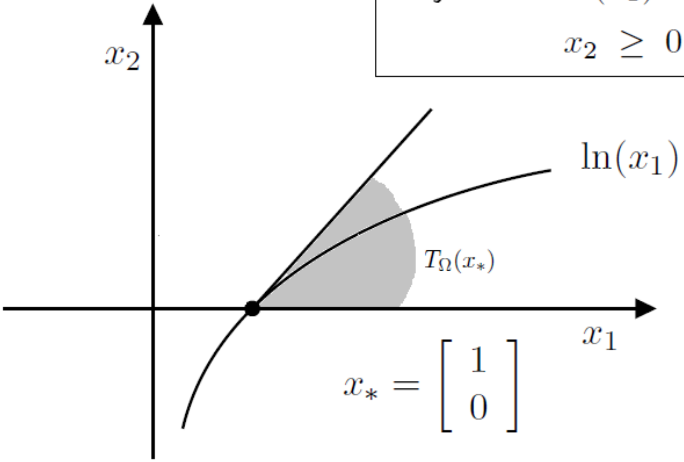
Definição

*Seja $x_* \in \Omega$. O **cone tangente** é o conjunto de todas as direcções tangentes a Ω em x_*, ou seja,*

$$T_\Omega(x_*) = \{d \in \mathbb{R}^n : \; d \text{ é tangente a } \Omega \text{ em } x_*\}$$

Considerando o exemplo anterior, temos

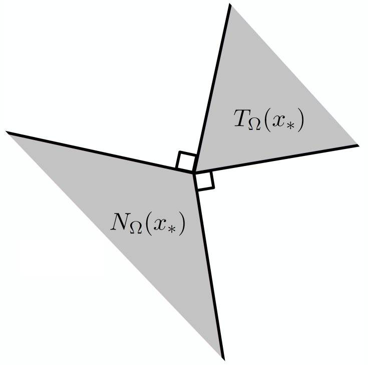
$$
\begin{array}{rl}
\min\limits_{x\in\mathbb{R}^2} & x_1 + x_2 \\
\text{sujeito a} & \ln(x_1) - x_2 \;\geq\; 0 \\
 & x_2 \;\geq\; 0
\end{array}
$$



É possível, desde logo, provar uma condição necessária para optimização local utilizando $T_\Omega(x_*)$.

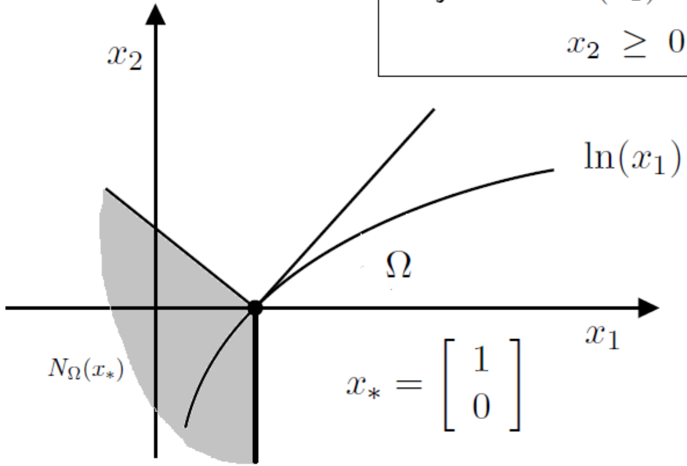
Em particular, utilizaremos o cone normal a Ω no ponto x_* (corresponde ao polar do cone tangente)

$$N_\Omega(x_*) = (T_\Omega(x_*))^\circ = \{v \in \mathbb{R}^n : \ v^\top d \ \leq \ 0, \quad \forall d \in T_\Omega(x_*)\}$$



Considerando o exemplo anterior, temos

$$\begin{array}{rl} \displaystyle\min_{x \in \mathbb{R}^2} & x_1 + x_2 \\ \text{sujeito a} & \ln(x_1) - x_2 \ \geq \ 0 \\ & x_2 \ \geq \ 0 \end{array}$$



Teorema

Seja $x_* \in \Omega$. Seja $f \in C^1(D)$, em que $x_* \in D$ (aberto).

Se x_* for um minimizante local de f em Ω, então

$$-\nabla f(x_*) \in N_\Omega(x_*)$$

Demonstração: Seja $d \in T_\Omega(x_*)$. Por definição de $T_\Omega(x_*)$, existe uma sucessão admissível $\{z_k\}$ a aproximar-se de x_* e uma sucessão $\{t_k\}$ em que $t_k > 0$ e $t_k \downarrow 0$ tais que

$$\lim_{k \to +\infty} \frac{z_k - x_*}{t_k} = d$$

Este limite é equivalente a

$$z_k = x_* + t_k d + \mathrm{o}(t_k)$$

Por sua vez, aplicando uma expansão de Taylor, temos

$$f(z_k) - f(x_*) = (z_k - x_*)^\top \nabla f(x_*) + \mathrm{o}(\|z_k - x_*\|)$$

Como

$$z_k - x_* = t_k d + \mathrm{o}(t_k)$$

então

$$(z_k - x_*)^\top \nabla f(x_*) = t_k d^\top \nabla f(x_*) + \nabla f(x_*)^\top \mathrm{o}(t_k) = t_k d^\top \nabla f(x_*) + \mathrm{o}(t_k)$$

Por outro lado, observe-se que

$$\begin{aligned}\frac{\mathrm{o}(\|z_k - x_*\|)}{t_k} &= \frac{\mathrm{o}(\|z_k - x_*\|)}{t_k}\frac{\|z_k - x_*\|}{\|z_k - x_*\|} \\ &= \underbrace{\frac{\mathrm{o}(\|z_k - x_*\|)}{\|z_k - x_*\|}}_{\to 0}\underbrace{\frac{\|z_k - x_*\|}{t_k}}_{\to\|d\|}\end{aligned}$$

logo,

$$\mathrm{o}(\|z_k - x_*\|) = \mathrm{o}(t_k)$$

Assim, temos

$$f(z_k) - f(x_*) = (z_k - x_*)^\top \nabla f(x_*) + \mathrm{o}(\|z_k - x_*\|) = t_k d^\top \nabla f(x_*) + \mathrm{o}(t_k)$$

Como $f(z_k) \ \geq \ f(x_*)$ para k suf. grande, temos

$$t_k d^\top \nabla f(x_*) + \mathrm{o}(t_k) \ \geq \ 0$$

Dividindo por t_k, vem

$$0 \ \leq \ \nabla f(x_*)^\top d + \frac{\mathrm{o}(t_k)}{t_k}$$

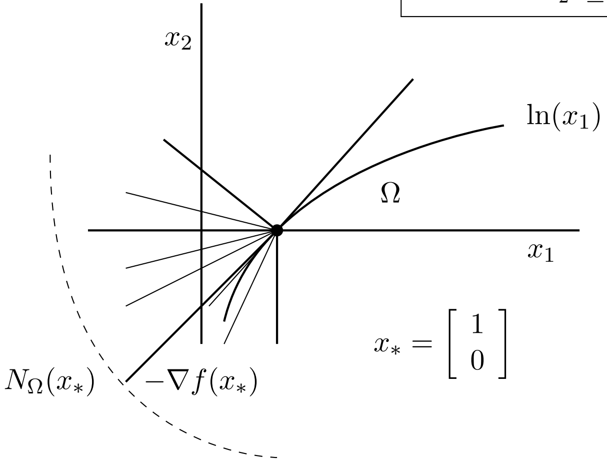
Finalmente, tomando limites, vem

$$-\nabla f(x_*)^\top d \;\leq\; 0$$

■

No âmbito do exemplo anterior, a condição necessária do teorema é ilustrada do seguinte modo

$$\begin{array}{rl} \displaystyle\min_{x\in\mathbb{R}^2} & x_1 + x_2 \\ \text{sujeito a} & \ln(x_1) - x_2 \geq 0 \\ & x_2 \geq 0 \end{array}$$



Exercício:

- O que acontece quando temos $\Omega = \mathbb{R}^n$?

Note-se que quer a definição (geométrica) de $T_\Omega(x_*)$ quer o teorema anterior são válidos para qualquer Ω.

Como representar $T_\Omega(x_*)$ algebricamente e mais concretamente linearmente (caso em que Ω é definido na forma PNL)?

Tal representação é possível sob uma **qualificação de restrições**.

Esta é uma condição suficiente para que uma aproximação linear de Ω em x_* descreva bem as características geométricas essenciais de Ω numa vizinhança de x_* (resumidas por $T_\Omega(x_*)$).

Definição

Dado um ponto $x \in \Omega$ e o conjunto das restrições activas $\mathcal{A}(x)$, temos uma qualificação de restrições LICQ se o conjunto

$$\{\nabla c_i(x),\ i \in \mathcal{A}(x)\}$$

for linearmente independente.

Definição

Dado um ponto $x \in \Omega$ e o conjunto das restrições activas $\mathcal{A}(x)$, o cone de primeira ordem das direcções admissíveis é definido por

$$\mathcal{F}(x) = \left\{ d \in \mathbb{R}^n : \begin{array}{ll} d^\top \nabla c_i(x) = 0, & \forall i \in \mathcal{E} \\ d^\top \nabla c_i(x) \geq 0, & \forall i \in \mathcal{I} \cap \mathcal{A}(x) \end{array} \right\}$$

Sob a qualificação de restrições LICQ

$$\{\nabla c_i(x_*),\ i \in \mathcal{A}(x_*)\} \quad \text{linearmente independente,}$$

tem-se que o cone de primeira ordem das direcções admissíveis

$$\mathcal{F}(x_*) = \left\{ d \in \mathbb{R}^n : \begin{array}{ll} d^\top \nabla c_i(x_*) = 0, & \forall i \in \mathcal{E} \\ d^\top \nabla c_i(x_*) \geq 0, & \forall i \in \mathcal{I} \cap \mathcal{A}(x_*) \end{array} \right\}$$

é igual ao cone tangente $T_\Omega(x_*)$.

Teorema

Seja $x_* \in \Omega$. Sejam $c_i \in C^1(D)$, em que $x_* \in D$ (aberto).

Então temos que:

(i)
$$T_\Omega(x_*) \subseteq \mathcal{F}(x_*)$$

(ii)
$$\text{(sob LICQ)} \quad \mathcal{F}(x_*) \subseteq T_\Omega(x_*)$$

Demonstração: (i) Seja $d \in T_\Omega(x_*)$. Por definição de $T_\Omega(x_*)$, existe uma sucessão admissível $\{z_k\}$ a aproximar-se de x_* e uma sucessão $\{t_k\}$ em que $t_k > 0$ e $t_k \downarrow 0$, tais que

$$\lim_{k \to +\infty} \frac{z_k - x_*}{t_k} = d$$

Este limite é equivalente a

$$z_k = x_* + t_k d + \mathrm{o}(t_k)$$

Consideremos primeiro o caso em que $i \in \mathcal{E}$.

$$c_i(z_k) = \underbrace{c_i(x_*)}_{=0} + \underbrace{(z_k - x_*)^\top \nabla c_i(x_*)}_{t_k d^\top \nabla c_i(x_*) + \mathrm{o}(t_k)} + \underbrace{\mathrm{o}(\|z_k - x_*\|)}_{\mathrm{o}(t_k)}$$

Como, para k suf. grande $c_i(z_k) = 0$, temos

$$0 = \nabla c_i(x_*)^\top d + \frac{\mathrm{o}(t_k)}{t_k}$$

Tomando limites, temos

$$\nabla c_i(x_*)^\top d = 0$$

O caso em que $i \in \mathcal{I}$ prova-se de modo análogo.

(ii) Seja $d \in \mathcal{F}(x_*)$. Vamos construir uma sucessão $\{z_k\}$ a convergir para x_*.

Seja $A = A(x_*) = \left[\nabla c_i(x_*)^\top\right]_{i \in \mathcal{A}(x_*)}$

Seja $Z \in \mathbb{R}^{n \times (n-m)}$ uma matriz cujas colunas são uma base para o espaço nulo $\mathcal{N}(A)$ de A.

Seja $c(z) = [c_i(z)]_{i \in \mathcal{A}(x_*)}$

Considere-se

$$\left[\frac{A}{Z^\top}\right]_{n-m}^{m} \qquad \text{com} \qquad m = |\mathcal{A}(x_*)|$$

Faça-se

$$R(z;t) = \left[\begin{array}{c} c(z) - tAd \\ Z^\top (z - x_* - td) \end{array} \right]$$

Tem-se

$$R(x_*;0) = 0 \quad \text{e} \quad \nabla_z R(x_*;0) = \left[\begin{array}{c} A \\ Z^\top \end{array} \right] \text{ não singular por construção de } Z.$$

Aplicando-se o teorema da função implícita (TFI) com

$$\{t_k\}, \quad t_k > 0 \quad \text{e} \quad t_k \longrightarrow 0$$

garante-se a existência de uma solução z_k para $R(z; t_k) = 0$ com

$$z_k \longrightarrow x_*$$

Verifica-se facilmente que z_k é admissível.

Prova-se, ainda, que $d = \lim_{k \to +\infty} \frac{z_k - x_*}{t_k}$.

■

Exercício:

- Termine explicitamente a demonstração de (ii) do teorema anterior, mostrando que z_k é admissível e $d = \lim_{k \to +\infty} \frac{z_k - x_*}{t_k}$.

  Sugestão: Para provar que

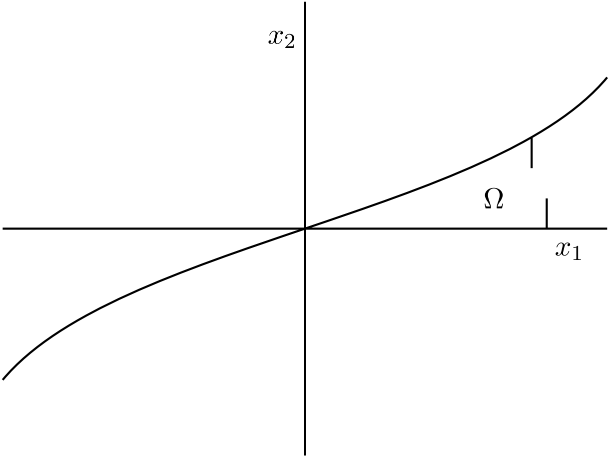
$$d = \lim_{k \to +\infty} \frac{z_k - x_*}{t_k}$$

  comece por multiplicar por $\left[\begin{array}{c} A \\ Z^\top \end{array} \right]^{-1}$ e depois mostre que

$$\lim_{k \to +\infty} \left(\frac{z_k - x_*}{\|z_k - x_*\|} - \frac{t_k}{\|z_k - x_*\|} d \right) = 0$$

Exemplo:

$$\boxed{\Omega = \{x \in \mathbb{R}^2 : \quad x_2 \;\leq\; x_1^3, \quad x_2 \;\geq\; 0\}}$$



Temos

$$\nabla c_1(x) = \left[\begin{array}{c} 3x_1^2 \\ -1 \end{array} \right] \qquad \text{e} \qquad \nabla c_2(x) = \left[\begin{array}{c} 0 \\ 1 \end{array} \right]$$

Seja $x_* = 0$, então

$$\nabla c_1(x_*) = \left[\begin{array}{r} 0 \\ -1 \end{array} \right] \qquad \text{e} \qquad \nabla c_2(x_*) = \left[\begin{array}{c} 0 \\ 1 \end{array} \right]$$

Assim, temos que $\mathcal{A}(x_*) = \{1, 2\}$ e a qualificação de restrições LICQ não é satisfeita.

Temos ainda,

$$\begin{aligned}\mathcal{F}(x_*) &= \left\{ d \in \mathbb{R}^2 : \begin{bmatrix} 0 \\ -1 \end{bmatrix}^\top \begin{bmatrix} d_1 \\ d_2 \end{bmatrix} \geq 0, \quad \begin{bmatrix} 0 \\ 1 \end{bmatrix}^\top \begin{bmatrix} d_1 \\ d_2 \end{bmatrix} \geq 0 \right\} \\ &= \left\{ d \in \mathbb{R}^2 : \; d_2 = 0 \right\}\end{aligned}$$

e, por conseguinte, $\mathcal{F}(x_*) \nsubseteq T_\Omega(x_*)$.

Uma outra qualificação de restrições importante é

$$\boxed{c_i, \quad i \in \mathcal{A}(x_*) \text{ são funções afim}}$$

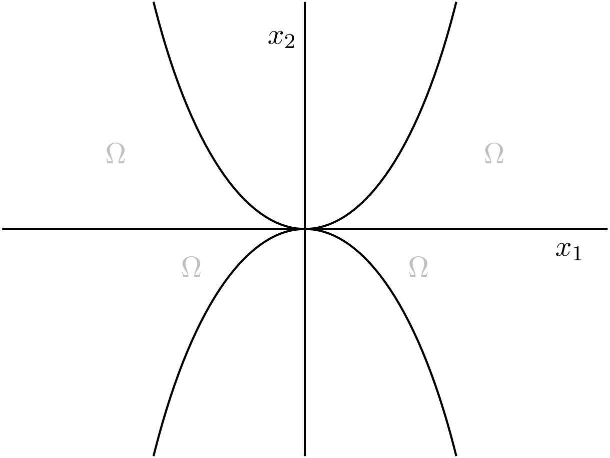
Neste caso prova-se, também, que $\mathcal{F}(x_*) \subseteq T_\Omega(x_*)$.

É importante notar que as restrições de qualificação **não são** condições necessárias para que $\mathcal{F}(x_*) = T_\Omega(x_*)$.

Veja-se o exemplo seguinte.

Exemplo:

$$\boxed{\Omega = \{x \in \mathbb{R}^2 : \; x_2 \; \leq \; x_1^2, \; x_2 \; \geq \; -x_1^2\}}$$



Temos

$$\nabla c_1(x) = \begin{bmatrix} 2x_1 \\ -1 \end{bmatrix} \qquad \text{e} \qquad \nabla c_2(x) = \begin{bmatrix} 2x_1 \\ 1 \end{bmatrix}$$

Seja $x_* = 0$, então

$$\nabla c_1(x_*) = \begin{bmatrix} 0 \\ -1 \end{bmatrix} \qquad \text{e} \qquad \nabla c_2(x_*) = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

Assim,

$$
\begin{aligned}
\mathcal{F}(x_*) &= \left\{ d \in \mathbb{R}^2 : \begin{bmatrix} 0 \\ -1 \end{bmatrix}^\top \begin{bmatrix} d_1 \\ d_2 \end{bmatrix} \geq 0, \quad \begin{bmatrix} 0 \\ 1 \end{bmatrix}^\top \begin{bmatrix} d_1 \\ d_2 \end{bmatrix} \geq 0 \right\} \\
&= \left\{ d \in \mathbb{R}^2 : \ d_2 = 0 \right\}
\end{aligned}
$$

Logo, neste exemplo, tem-se

$$\mathcal{F}(x_*) = T_\Omega(x_*)$$

Exercício:

- Sob a qualificação de restrições

$$c_i, \quad i \in \mathcal{A}(x_*) \text{ são funções afim,}$$

mostre que $\mathcal{F}(x_*) \subseteq T_\Omega(x_*)$.

Presentation outline

1. Métodos numéricos para optimização não linear sem restrições
   - Métodos de procura directa
   - Métodos de procura unidireccional (geral)
   - Métodos de procura unidireccional (descida máxima, Newton)
   - Métodos de região de confiança
2. Teoria da optimização não linear com restrições
   - Introdução e qualificações de restrições
   - Condições de primeira ordem
   - Condições de segunda ordem
   - Teoria da dualidade
3. Métodos numéricos para optimização não linear com restrições
   - Método da penalização quadrática
   - Método do Lagrangeano aumentado
   - Programação Sequencial Quadrática (SQP)
   - Métodos de pontos interiores

Relativamente às condições necessárias de primeira ordem (CN1O), sabemos já que

$$-\nabla f(x_*) \in (T_\Omega(x_*))^\circ \underbrace{=}_{\text{sob LICQ}} (\mathcal{F}(x_*))^\circ$$

Importa agora calcular $(\mathcal{F}(x_*))^\circ$. Tal será conseguido através do conhecido Lema de Farkas.

Lema

Seja $g \in \mathbb{R}^n$ *e* $K = \{By + Cw : \ y \ \geq \ 0\} \subseteq \mathbb{R}^n$. *Então*

(i) *ou*

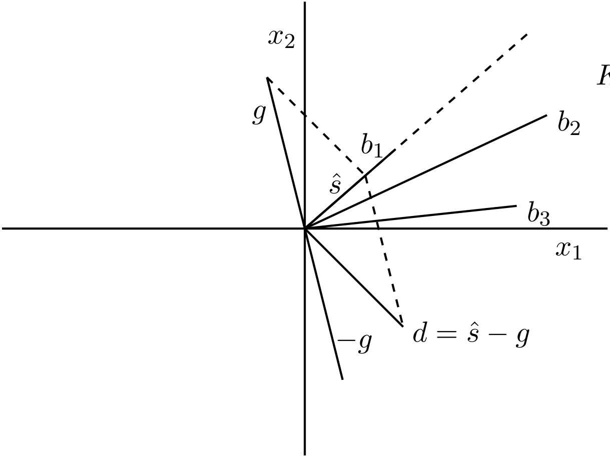
$$g \in K$$

(ii) *ou*

$$\exists d \in \mathbb{R}^n : \quad g^\top d \ < \ 0, \quad B^\top d \ \geq \ 0, \quad C^\top d = 0$$

Notas:

- K é um cone convexo e é fechado.
- O Lema de Farkas é um teorema de alternativa.



$$K = \{By : \; y \; \geq \; 0\}$$

Demonstração: Comecemos por mostrar que (i) e (ii) não podem acontecer simultaneamente.

Se $g \in K$, então $\exists y \geq 0, w: \; g = By + Cw$.

Se $\exists d \in \mathbb{R}^n: \; g^\top d < 0, \quad B^\top d \;\geq\; 0, \quad C^\top d = 0$, então temos

$$0 \;>\; d^\top g \;=\; d^\top By + d^\top Cw \;=\; (\underbrace{B^\top d}_{\geq 0})^\top \underbrace{y}_{\geq 0} + (\underbrace{C^\top d}_{= 0})^\top w \;\geq\; 0\,!$$

Logo (i) e (ii) não podem acontecer simultaneamente.

Suponhamos então que $g \notin K$.

Uma vez que K é fechado e convexo e a função

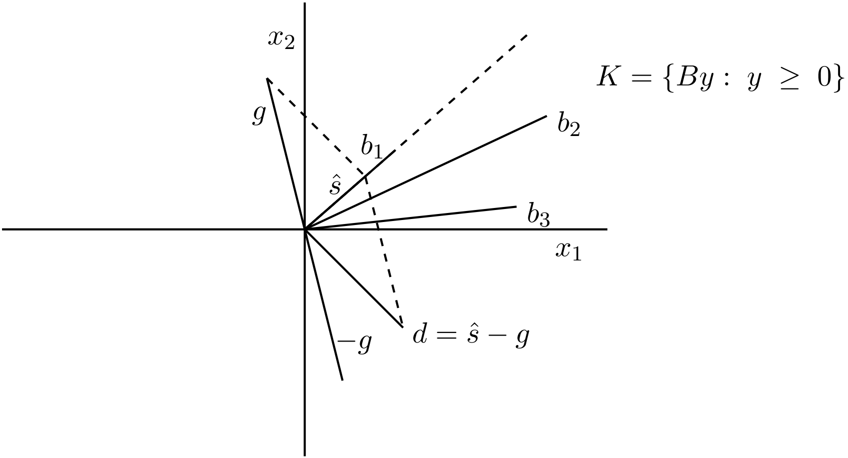
$$f(s) = \|s - g\|^2$$

é contínua e uniformemente convexa, o problema

$$\begin{array}{rl} \displaystyle\min_{s \in \mathbb{R}^n} & \|s - g\|^2 \\ \text{sujeito a} & s \in K \end{array}$$

tem solução e esta é única ($\hat{s}$).

Seja d como na figura



Então, temos (pela figura...)

$$d^\top g = d^\top(\hat{s} - d) = \underbrace{d^\top \hat{s}}_{=0} - \underbrace{\|d\|^2}_{<0} < 0$$

Por outro lado, temos (pela figura...)

$$d^\top s \;\geq\; 0, \; \forall s \in K$$

Logo,

$$0 \;\leq\; d^\top(By + Cw), \quad \forall y \geq 0, w$$

Fazendo

$$\begin{array}{llll} w = 0: & (B^\top d)^\top y \; \geq \; 0, \quad \forall y \geq 0 & \Longrightarrow & B^\top d \; \geq \; 0 \\ y = 0: & (C^\top d)^\top w \; \geq \; 0, \quad \forall w & \Longrightarrow & C^\top d = 0 \end{array}$$

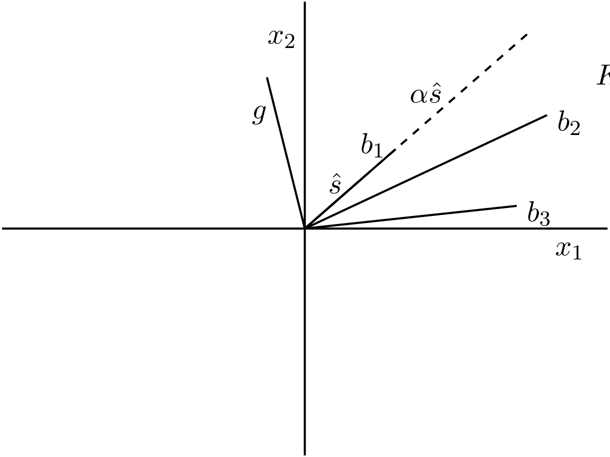
Falta então provar que $d^\top \hat{s} = 0$ e $d^\top s \; \geq \; 0, \, \forall s \in K$.

Provemos que $d^\top \hat{s} = 0$.

Como $\hat{s} \in K$ e uma vez que K é um cone, então $\alpha \hat{s} \in K \quad \forall \alpha > 0$.

A função $h(\alpha) = \|\alpha \hat{s} - g\|^2$ tem como minimizante $\alpha = 1$. Assim,

$$h'(1) = ((2\hat{s}^\top \hat{s})\alpha - 2g^\top \hat{s})|_{\alpha=1} = 0 \Longleftrightarrow \hat{s}^\top \hat{s} - g^\top \hat{s} = 0 \Longleftrightarrow (\underbrace{\hat{s} - g}_{d})^\top \hat{s} = 0$$



$$K = \{By : \; y \; \geq \; 0\}$$

Por último provemos, analiticamente, que $d^\top s \ \geq \ 0, \ \forall s \in K$.

Uma vez que K é convexo e $\hat{s} \in K$, então

$$\theta s + (1-\theta)\hat{s} \in K \quad \forall \theta \in [0,1]$$

Assim, temos

$$\|\theta s + (1-\theta)\hat{s} - g\|^2 = \|\hat{s} + \theta(s-\hat{s}) - g\|^2 \;\geq\; \|\hat{s} - g\|^2 \quad \forall \theta \in [0,1]$$

Fazendo as contas, vem que

$$\theta^2\|s-\hat{s}\|^2 + 2\theta(s-\hat{s})^\top(\hat{s}-g) \;\geq\; 0 \underbrace{\Longrightarrow}_{\theta\to 0^+} (s-\hat{s})^\top d \;\geq\; 0 \Longrightarrow s^\top d \;\geq\; 0$$

■

Sabemos, então, que se x_* for um minimizante local, temos

$$-\nabla f(x_*) \in (T_\Omega(x_*))^\circ \underbrace{=}_{\text{LICQ}} (\mathcal{F}(x_*))^\circ$$

Desta forma

$$d^\top \nabla f(x_*) \;\geq\; 0, \quad \forall d \in \mathcal{F}(x_*)$$

Recorde-se que

$$\mathcal{F}(x_*) = \left\{ d \in \mathbb{R}^n : \begin{array}{ll} d^\top \nabla c_i(x_*) = 0, & \forall i \in \mathcal{E} \\ d^\top \nabla c_i(x_*) \geq 0, & \forall i \in \mathcal{I} \cap \mathcal{A}(x_*) \end{array} \right\}$$

Fazendo

$$B_* = (\nabla c_i(x_*))_{i \in \mathcal{I} \cap \mathcal{A}(x_*)}$$

$$C_* = (\nabla c_i(x_*))_{i \in \mathcal{E}}$$

vem, pelo Lema de Farkas,

$$\nabla f(x_*) \in K_* = \{B_* y + C_* w : \ y \ \geq \ 0\}$$

Surgem, assim, desta forma as condições necessárias de primeira ordem (CN1O) para o PNL, também conhecidas por condições de Karush-Kuhn-Tucker (KKT) de primeira ordem.

Seja

$$L(x, \lambda) = f(x) - \sum_{i \in \mathcal{E} \cup \mathcal{I}} \lambda_i c_i(x)$$

a função Lagrangeana associada ao PNL (λ é o vector dos multiplicadores de Lagrange).

Observe-se que

$$\nabla_x L(x, \lambda) = \nabla f(x) - \sum_{i \in \mathcal{E} \cup \mathcal{I}} \lambda_i \nabla c_i(x)$$

Teorema

Seja $x_* \in \Omega$*.*

Sejam $f, c_i \in C^1(D)$*, em que* $x_* \in D$ *(aberto).*

Suponhamos que a qualificação de restrições LICQ é satisfeita em x_**.*

Se x_* *for um minimizante local, então* $\exists \lambda_*$:

$$\begin{aligned}
\nabla_x L(x_*, \lambda_*) &= 0, & & \text{(dualidade)} \\
c_i(x_*) &= 0, & \forall i \in \mathcal{E} & \quad \text{(admissibilidade)} \\
c_i(x_*) &\geq 0, & \forall i \in \mathcal{I} & \quad \text{(admissibilidade)} \\
\lambda_i^* &\geq 0, & \forall i \in \mathcal{I} & \quad \text{(não negatividade)} \\
\lambda_i^* c_i(x_*) &= 0, & \forall i \in \mathcal{I} & \quad \text{(complementaridade)}
\end{aligned}$$

Demonstração: Vimos que

$$\exists \lambda_i^*, i \in \mathcal{A}(x_*) \quad \text{e} \quad \lambda_i^* \;\geq\; 0, i \in \mathcal{I} \cap \mathcal{A}(x_*)$$

tais que

$$\nabla f(x_*) = \sum_{i \in \mathcal{A}(x_*)} \lambda_i^* \nabla c_i(x_*)$$

Defina-se λ_* do seguinte modo

$$\lambda_i^* \;=\; \begin{cases} \lambda_i, & i \in \mathcal{A}(x_*) \\ \;0, & i \notin \mathcal{A}(x_*) \end{cases}$$

Mostremos, agora, que esta escolha de λ_* satisfaz as CN1O-KKT.

Temos

$$\begin{aligned}
\nabla_x L(x_*, \lambda_*) &= \nabla f(x_*) - \sum_{i \in \mathcal{E} \cup \mathcal{I}} \lambda_i^* \nabla c_i(x_*) \\
&= \underbrace{\nabla f(x_*) - \sum_{i \in \mathcal{A}(x_*)} \lambda_i^* \nabla c_i(x_*)}_{=0} - \sum_{i \notin \mathcal{A}(x_*)} \lambda_i^* \nabla c_i(x_*)
\end{aligned}$$

Tendo em conta a definição de λ_*, vem

$$\nabla_x L(x_*, \lambda_*) = 0$$

dualidade ✓

Uma vez que $x_* \in \Omega$, vem

$$\begin{aligned} c_i(x_*) &= 0, \quad \forall i \in \mathcal{E} \\ c_i(x_*) &\geq 0, \quad \forall i \in \mathcal{I} \end{aligned}$$

admissibilidade ✓

Temos

$$\lambda_i^* \;\geq\; 0,\; i \in \mathcal{A}(x_*) \quad \text{e} \quad \lambda_i^* = 0,\; i \notin \mathcal{A}(x_*)$$

Logo,

$$\lambda_i^* \;\geq\; 0,\; i \in \mathcal{I}$$

não negatividade ✓

Temos,

$$\begin{aligned} c_i(x_*) &= 0, \quad \forall i \in \mathcal{A}(x_*) \\ \lambda_i^* &= 0, \quad \forall i \notin \mathcal{A}(x_*) \end{aligned}$$

Assim,

$$\lambda_i^* c_i(x_*) = 0, \ i \in \mathcal{I}$$

complementaridade ✓

■

Exemplo:

$$\begin{aligned} \min_{x \in \mathbb{R}^2} \quad & f(x) \\ \text{sujeito a} \quad & \ln(x_1) - x_2 \ \geq \ 0 \\ & x_2 \ \geq \ 0 \end{aligned}$$

Temos

$$\nabla c_1(x) = \left[\begin{array}{c} 1/x_1 \\ -1 \end{array} \right] \quad \text{e} \quad \nabla c_2(x) = \left[\begin{array}{c} 0 \\ 1 \end{array} \right]$$

Considerando $f(x) = x_1 + \frac{x_2^2}{3} + x_2$, temos $\nabla f(1,0) = \left[\begin{array}{c} 1 \\ 1 \end{array} \right]$.

Assim, temos

$$\begin{bmatrix} 1 \\ 1 \end{bmatrix} = \lambda_1^* \begin{bmatrix} 1 \\ -1 \end{bmatrix} + \lambda_2^* \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

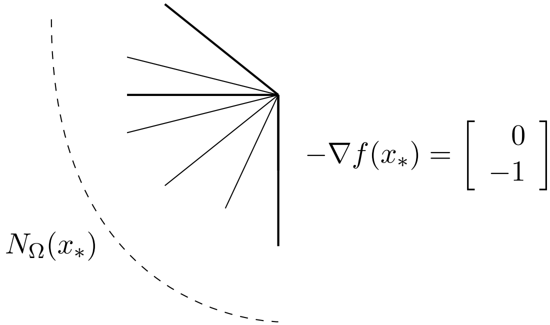
Logo,

$$x_* = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \quad \text{e} \quad \lambda_* = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \quad \text{satisfaz as CN1O-KKT.}$$

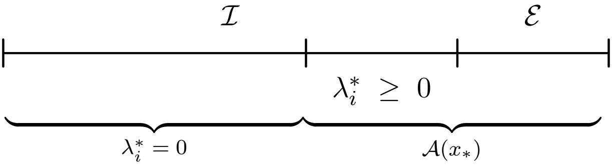
Considerando $f(x) = \dfrac{x_2^3}{3} + x_2$, temos $\nabla f(1,0) = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$

Assim,

$$\begin{bmatrix} 0 \\ 1 \end{bmatrix} = \underbrace{0}_{\lambda_1^*} \begin{bmatrix} 1 \\ -1 \end{bmatrix} + \underbrace{1}_{\lambda_2^*} \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$



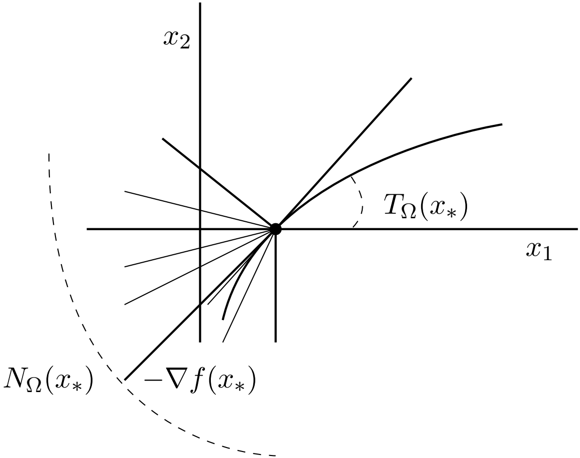
Notas:

- 

- Sob LICQ, λ^* é único.

- Quando $\lambda_i^* > 0, \ \forall i \in \mathcal{I} \cap \mathcal{A}(x_*)$ tem-se complementaridade estrita.

  Nesse caso (e quando $\mathcal{E} = \emptyset$), $-\nabla f(x_*)$ está no *interior relativo* do cone normal $N_\Omega(x_*)$:



- Uma outra qualificação de restrições, muito conhecida, é a MFCQ (Mangasarian-Fromovitz)

$$\exists w \in \mathbb{R}^n : \left\{ \begin{array}{ll} \nabla c_i(x_*)^\top w > 0, & i \in \mathcal{A}(x_*) \cap \mathcal{I} \\ \nabla c_i(x_*)^\top w = 0, & i \in \mathcal{E} \end{array} \right.$$

e $\{\nabla c_i(x_*),\ i \in \mathcal{E}\}$ é linearmente independente.

Pode-se provar que MFCQ é equivalente a

$$\{\lambda_* :\ \lambda_* \text{ satisfaz as CN1O}\} \quad \text{é limitado.}$$

É óbvio, então, que LICQ $\implies$ MFCQ.

Presentation outline

Relativamente às condições necessárias de segunda ordem (CN2O), interessa-nos essencialmente estudar a curvatura da função Lagrangeana ao longo das direcções de indecisão, isto é, ao longo das direcções

$$v \in \mathcal{F}(x_*) \quad \text{para as quais} \quad \nabla f(x_*)^\top v = 0$$

Como

$$\nabla f(x_*)^\top v \overbrace{=}^{\text{CN1O}} \sum_{i \in \mathcal{A}(x_*)} \lambda_i^* \nabla c_i(x_*)^\top v$$

há que considerar $\nabla c_i(x_*)^\top v = 0$, quando $i \in \mathcal{A}(x_*) \cap \mathcal{I}$ e $\lambda_i^* > 0$.

Definição

Considere-se

$$\mathcal{F}(x_*) = \left\{ d \in \mathbb{R}^n : \begin{array}{ll} d^\top \nabla c_i(x_*) = 0, & \forall i \in \mathcal{E} \\ d^\top \nabla c_i(x_*) \geq 0, & \forall i \in \mathcal{I} \cap \mathcal{A}(x_*) \end{array} \right\}$$

Seja λ_* um vector de multiplicadores de Lagrange a satisfazer as KKT-CN1O.

O cone de criticalidade (de segunda ordem) é definido por

$$\mathcal{C}(x_*, \lambda_*) = \{ v \in \mathcal{F}(x_*) : \nabla c_i(x_*)^\top v = 0, \ \forall i \in \mathcal{A}(x_*) \cap \mathcal{I} \text{ com } \lambda_i^* > 0 \}$$

Teorema

Seja $x_* \in \Omega$ *um minimizante local, em que a qualificação de restrições LICQ é satisfeita.*

Sejam $f, c_i \in C^2(D)$*, em que* $x_* \in D$ *(aberto).*

Seja λ_* *o vector dos multiplicadores de Lagrange a satisfazer as KKT-CN1O.*

Então

$$v^\top \nabla^2_{xx} L(x_*, \lambda_*) v \geq 0, \quad \forall v \in \mathcal{C}(x_*, \lambda_*)$$

Demonstração: Seja $v \in \mathcal{C}(x_*, \lambda_*) \subseteq \mathcal{F}(x_*) \underbrace{=}_{\text{LICQ}} T_\Omega(x_*)$

Por definição de $T_\Omega(x_*)$, existe uma sucessão admissível $\{z_k\}$ a aproximar-se de x_* e uma sucessão $\{t_k\}$ em que $t_k > 0$ e $t_k \downarrow 0$, tais que

$$\lim_{k \to +\infty} \frac{z_k - x_*}{t_k} = v$$

Este limite é equivalente a

$$z_k = x_* + t_k v + \mathrm{o}(t_k)$$

Através de um raciocínio, utilizado anteriormente, sabemos que é possível escolher $\{t_k\}$ e $\{z_k\}$ tais que

$$c_i(z_k) = t_k \nabla c_i(x_*)^\top v, \ \forall i \in \mathcal{A}(x_*)$$

Assim e tendo em conta a definição de função Lagrangeana, temos

$$\begin{aligned} L(z_k, \lambda_*) &= f(z_k) - \sum_{i \in \mathcal{A}(x_*)} \lambda_i^* c_i(z_k) \\ &= f(z_k) - \underbrace{t_k \sum_{i \in \mathcal{A}(x_*)} \lambda_i^* \nabla c_i(x_*)^\top v}_{= 0 \text{ (pois } v \in \mathcal{C}(x_*, \lambda_*))} \\ &= f(z_k) \end{aligned}$$

Por outro lado, aplicando uma expansão de Taylor, temos

$$\begin{aligned} L(z_k, \lambda_*) &= L(x_*, \lambda_*) + (z_k - x_*)^\top \nabla_x L(x_*, \lambda_*) \\ &\quad + \tfrac{1}{2}(z_k - x_*)^\top \nabla^2_{xx} L(x_*, \lambda_*)(z_k - x_*) + \mathrm{o}(\|z_k - x_*\|^2) \end{aligned}$$

Pela complementaridade, temos

$$L(x_*, \lambda_*) = f(x_*) - \underbrace{\sum_{i \in \mathcal{A}(x_*)} \lambda_i^* c_i(x_*)}_{=0} = f(x_*)$$

Pela dualidade, temos

$$(z_k - x_*)^\top \nabla_x L(x_*, \lambda_*) = 0$$

Assim, temos

$$\begin{aligned}\underbrace{L(z_k, \lambda_*)}_{= f(z_k)} \quad &= \quad \underbrace{L(x_*, \lambda_*)}_{= f(x_*)} + \underbrace{(z_k - x_*)^\top \nabla_x L(x_*, \lambda_*)}_{= 0} \\ &\qquad + \frac{1}{2}(z_k - x_*)^\top \nabla^2_{xx} L(x_*, \lambda_*)(z_k - x_*) + \mathrm{o}(\|z_k - x_*\|^2)\end{aligned}$$

Consequentemente,

$$f(z_k) = f(x_*) + \frac{1}{2}(z_k - x_*)^\top \nabla^2_{xx} L(x_*, \lambda_*)(z_k - x_*) + \mathrm{o}(\|z_k - x_*\|^2)$$

Uma vez que

$$z_k - x_* = t_k v + \mathrm{o}(t_k)$$

vem

$$\begin{aligned}
&(z_k - x_*)^\top \nabla^2_{xx} L(x_*, \lambda_*)(z_k - x_*) + \mathrm{o}(\|z_k - x_*\|^2) \\
&= \; t_k^2 v^\top \nabla^2_{xx} L(x_*, \lambda_*) v + 2 v^\top \nabla^2_{xx} L(x_*, \lambda_*) \underbrace{t_k \mathrm{o}(t_k)}_{\mathrm{o}(t_k^2)} + \mathrm{o}(t_k^2) + \mathrm{o}(t_k^2) \\
&= \; t_k^2 v^\top \nabla^2_{xx} L(x_*, \lambda_*) v + \mathrm{o}(t_k^2)
\end{aligned}$$

Assim

$$f(z_k) = f(x_*) + \frac{1}{2} t_k^2 v^\top \nabla_{xx}^2 L(x_*, \lambda_*) v + \mathrm{o}(t_k^2)$$

Uma vez que $f(z_k) \ \geq \ f(x_*)$ para k suficientemente grande, então

$$\frac{1}{2} t_k^2 v^\top \nabla_{xx}^2 L(x_*, \lambda_*) v + \mathrm{o}(t_k^2) \ \geq \ 0$$

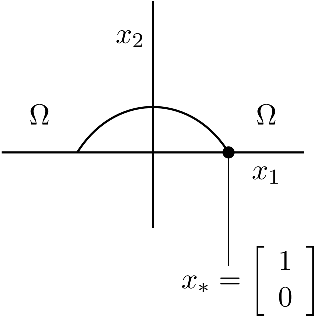
Finalmente, dividindo por t_k^2 e tomando limites, vem

$$\frac{1}{2} v^\top \nabla_{xx}^2 L(x_*, \lambda_*) v \ \geq \ 0 \quad \Longleftrightarrow \quad v^\top \nabla_{xx}^2 L(x_*, \lambda_*) v \ \geq \ 0$$

$\blacksquare$

Exemplo:

$$\begin{array}{rl} \min\limits_{x \in \mathbb{R}^2} & -0.1(x_1-4)^2 + x_2^2 \\ \text{sujeito a} & x_1^2 + x_2^2 - 1 \ \geq \ 0 \\ & x_2 \ \geq \ 0 \end{array}$$



Temos $\mathcal{E} = \emptyset$ e $\mathcal{I} = \{1,2\} = \mathcal{A}(x_*)$.

$$\nabla f(x) = \left[\begin{array}{c} -0.2(x_1-4) \\ 2x_2 \end{array}\right], \quad \nabla c_1(x) = \left[\begin{array}{c} 2x_1 \\ 2x_2 \end{array}\right] \quad \text{e} \quad \nabla c_2(x) = \left[\begin{array}{c} 0 \\ 1 \end{array}\right].$$

$$\nabla_x L(x_*, \lambda_*) = 0 \quad \Longleftrightarrow \quad \left[\begin{array}{c} 0.6 \\ 0 \end{array}\right] - \underbrace{\lambda_1^*}_{=0.3} \left[\begin{array}{c} 2 \\ 0 \end{array}\right] - \underbrace{\lambda_2^*}_{=0} \left[\begin{array}{c} 0 \\ 1 \end{array}\right] = 0$$

Qualificação de restrições LICQ ✓

Dualidade ✓

Complementaridade ✓

Admissibilidade ✓

Não negatividade ✓

CN1O ✓

Temos

$$\begin{aligned} \mathcal{F}(x_*) &= \left\{ d \in \mathbb{R}^2 : \begin{bmatrix} 2 \\ 0 \end{bmatrix}^\top \begin{bmatrix} d_1 \\ d_2 \end{bmatrix} \geq 0, \begin{bmatrix} 0 \\ 1 \end{bmatrix}^\top \begin{bmatrix} d_1 \\ d_2 \end{bmatrix} \geq 0 \right\} \\ &= \left\{ d \in \mathbb{R}^2 : d_1 \geq 0, d_2 \geq 0 \right\} \end{aligned}$$

$$\nabla^2_{xx} L(x_*, \lambda_*) = \begin{bmatrix} -0.2 & 0 \\ 0 & 2 \end{bmatrix} - 0.3 \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} = \begin{bmatrix} -0.8 & 0 \\ 0 & 1.4 \end{bmatrix}$$

$$\begin{aligned} \mathcal{C}(x_*, \lambda_*) &= \left\{ d \in \mathbb{R}^2 : \left[\begin{array}{c} 2 \\ 0 \end{array} \right]^\top \left[\begin{array}{c} d_1 \\ d_2 \end{array} \right] = 0, \left[\begin{array}{c} 0 \\ 1 \end{array} \right]^\top \left[\begin{array}{c} d_1 \\ d_2 \end{array} \right] \geq 0 \right\} \\ &= \left\{ d \in \mathbb{R}^2 : d_1 = 0, \ d_2 \geq 0 \right\} \end{aligned}$$

Assim, temos

$$d^\top \nabla^2_{xx} L(x_*, \lambda_*) d = 1.4 d_2^2 \geq 0, \ \forall d \in \mathcal{C}(x_*, \lambda_*)$$

CN2O ✓

As **condições suficientes de segunda ordem** (CS2O) baseiam-se, também, no cone $\mathcal{C}(x_*, \lambda_*)$.

Teorema

Seja $x_* \in \Omega$*.*

Sejam $f, c_i \in C^2(D)$*, em que* $x_* \in D$ *(aberto).*

Se, para todo o λ_* *vector de multiplicadores de Lagrange a satisfazer as KKT-CN1O,*

$$w^\top \nabla^2_{xx} L(x_*, \lambda_*) w > 0 \quad \forall w \in \mathcal{C}(x_*, \lambda_*), w \neq 0$$

então

x_* *é um minimizante local estrito.*

Demonstração: Vamos provar que $f(z_k) > f(x_*)$ para toda a sucessão admissível $\{z_k\}$ a aproximar-se de x_*.

Se considerarmos todas as sucessões K para as quais

$$\lim_{\substack{k \to +\infty \\ k \in K}} \frac{z_k - x_*}{\|z_k - x_*\|} = d$$

cobrimos toda a sucessão $\{z_k\}$.

No caso em que $d \in \mathcal{C}(x_*, \lambda_*)$

(note que $d \in T_\Omega(x_*) \subseteq \mathcal{F}(x_*)$ e não é preciso qualificação de restrições),

utilizando a condição suficiente e a demonstração do último teorema, temos

$$
\begin{aligned}
f(z_k) \;\geq\; L(z_k, \lambda_*) &= L(x_*, \lambda_*) + \frac{1}{2}(z_k - x_*)^\top \nabla^2_{xx} L(x_*, \lambda_*)(z_k - x_*) \\
&\quad + \mathrm{o}(\|z_k - x_*\|^2) \\
&= f(x_*) + \frac{1}{2}\|z_k - x_*\|^2 \underbrace{d^\top \nabla^2_{xx} L(x_*, \lambda_*) d}_{>0} \\
&\quad + \mathrm{o}(\|z_k - x_*\|^2)
\end{aligned}
$$

Vejamos que

$$\frac{1}{2}\|z_k - x_*\|^2 d^\top \nabla^2_{xx} L(x_*, \lambda_*) d + \mathrm{o}(\|z_k - x_*\|^2) > 0$$

Sabemos que $\forall \varepsilon > 0$,

$$\left| \frac{\mathrm{o}(\|z_k - x_*\|^2)}{\|z_k - x_*\|^2} \right| \;\leq\; \varepsilon \qquad \text{para } k \text{ suf. grande}$$

Assim

$$-\varepsilon \|z_k - x_*\|^2 \;\leq\; \mathrm{o}(\|z_k - x_*\|^2) \;\leq\; \varepsilon \|z_k - x_*\|^2$$

Fazendo $\varepsilon = \frac{1}{4} d^\top \nabla^2_{xx} L(x_*, \lambda_*) d$, vem

$$
\begin{aligned}
-\frac{1}{4} d^\top \nabla^2_{xx} L(x_*, \lambda_*) d \|z_k - x_*\|^2 &\leq \mathrm{o}(\|z_k - x_*\|^2) \\
&\leq \frac{1}{4} d^\top \nabla^2_{xx} L(x_*, \lambda_*) d \|z_k - x_*\|^2
\end{aligned}
$$

Temos, então, o limite inferior

$$
\mathrm{o}(\|z_k - x_*\|^2) \geq -\frac{1}{4} d^\top \nabla^2_{xx} L(x_*, \lambda_*) d \|z_k - x_*\|^2
$$

Desta forma,

$$\begin{aligned}
\tfrac{1}{2}\|z_k - x_*\|^2 d^\top \nabla^2_{xx} L(x_*, \lambda_*) d + \mathrm{o}(\|z_k - x_*\|^2) &> \\
\tfrac{1}{2}\|z_k - x_*\|^2 d^\top \nabla^2_{xx} L(x_*, \lambda_*) d - \tfrac{1}{4}\|z_k - x_*\|^2 d^\top \nabla^2_{xx} L(x_*, \lambda_*) d &> 0
\end{aligned}$$

Assim, $f(z_k) > f(x_*)$ para k suficientemente grande.

No caso em $d \notin \mathcal{C}(x_*, \lambda_*)$ não pode ser aplicada a condição suficiente.

Porém, neste caso, sabemos que

$$\exists j \in \mathcal{A}(x_*) \cap \mathcal{I} : \quad \lambda_j^* > 0, \quad \nabla c_j(x_*)^\top d > 0$$

Para além disso,

$$
\begin{aligned}
L(z_k, \lambda_*) &= f(z_k) - \sum_{i \in \mathcal{A}(x_*)} \lambda_i^* c_i(z_k) \\
&\leq f(z_k) - \lambda_j^* c_j(z_k) \\
&= f(z_k) - \lambda_j^* \Bigg(\overbrace{c_j(x_*)}^{=\,0} + \underbrace{(z_k - x_*)^\top \nabla c_j(x_*) + \mathrm{o}(\|z_k - x_*\|)}_{\|z_k - x_*\| \underbrace{d^\top \nabla c_j(x_*)}_{>\,0} + \mathrm{o}(\|z_k - x_*\|)} \Bigg)
\end{aligned}
$$

Logo, usando $L(z_k, \lambda_*) = f(x_*) + \mathrm{o}(\|z_k - x_*\|)$ (porquê?),

$$f(z_k) \;\geq\; f(x_*) + \|z_k - x_*\| \lambda_j^* d^\top \nabla c_j(x_*) + \mathrm{o}(\|z_k - x_*\|)$$

Pelo mesmo raciocínio utilizado anteriormente, sabemos que

$$\|z_k - x_*\| \lambda_j^* d^\top \nabla c_j(x_*) + \mathrm{o}(\|z_k - x_*\|) > 0$$

e, assim sendo, $f(z_k) > f(x_*)$ para k suficientemente grande.

■

Notas:

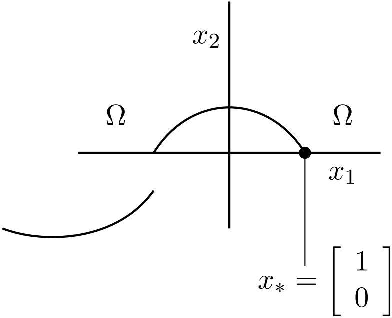
- As CS2O não necessitam de uma qualificação de restrições.

- Acha que a demonstração do último teorema é válida em espaços de dimensão infinita (espaços de Hilbert)? Ou seja, é utilizado, mesmo que implicitamente, algum argumento de compacidade na esfera unitária?

Exemplo:

Retomando o exemplo anterior,

$$\begin{array}{rl} \displaystyle\min_{x \in \mathbb{R}^2} & -0.1(x_1 - 4)^2 + x_2^2 \\ \text{sujeito a} & x_1^2 + x_2^2 - 1 \ \geq \ 0 \\ & x_2 \ \geq \ 0 \end{array}$$



$$d^\top \nabla^2_{xx} L(x_*, \lambda_*) d = 1.4 d_2^2 > 0, \ \forall d \in \mathcal{C}(x_*, \lambda_*), d \neq 0$$

Logo x_* é um minimizante local estrito.

Note-se que não há um minimizante global, pois

$$\lim_{\substack{x \in \Omega \\ x_1 \to +\infty \\ x_2 = 0}} f(x) = -\infty$$

As condições de segunda ordem podem ser apresentadas de uma outra forma, mais fraca mas mais simples de verificar.

Sob complementaridade estrita, temos que

$$\begin{aligned} \mathcal{C}(x_*, \lambda_*) &= \{w \in \mathbb{R}^n : \ \nabla c_i(x_*)^\top w = 0, \ i \in \mathcal{A}(x_*)\} \\ &= \mathcal{N}(A(x_*)), \quad \text{em que} \quad A(x_*) = [\nabla c_i(x_*)^\top]_{i \in \mathcal{A}(x_*)} \end{aligned}$$

ou, de modo equivalente,

$$\mathcal{C}(x_*, \lambda_*) = \{Zu : \ u \in \mathbb{R}^{|\mathcal{A}(x_*)|}\}$$

Deste modo, seja $w \in \mathcal{C}(x_*, \lambda_*)$ e Z uma matriz cujas colunas são uma base para $\mathcal{N}(A(x_*)) = \mathcal{C}(x_*, \lambda_*)$.

Então temos

$$w^\top \nabla^2_{xx} L(x_*, \lambda_*) w = u^\top (Z^\top \nabla^2_{xx} L(x_*, \lambda_*) Z) u$$

onde $Z^\top \nabla^2_{xx} L(x_*, \lambda_*) Z$ é Hessiana projectada (ou reduzida) do Lagrangeano.

Assim, as **CN2O** podem ser apresentadas como no seguinte teorema.

Teorema

Seja $x_* \in \Omega$ *um minimizante local, em que a qualificação de restrições LICQ é satisfeita.*

Sejam $f, c_i \in C^2(D)$, *em que* $x_* \in D$ *(aberto).*

Seja λ_* *o vector dos multiplicadores de Lagrange a satisfazer as KKT-CN1O (com complementaridade estrita).*

Seja Z *uma matriz cujas colunas são uma base para* $\mathcal{N}(A(x_*))$.

Então

$$Z^\top \nabla^2_{xx} L(x_*, \lambda_*) Z \text{ é PSD}$$

Por sua vez, as **CS2O** podem ser apresentadas como no seguinte teorema.

Teorema

Seja $x_* \in \Omega$

Sejam $f, c_i \in C^2(D)$*, em que* $x_* \in D$ *(aberto).*

Seja satisfeita a complementaridade estrita para todo o vector λ_* *dos multiplicadores de Lagrange a satisfazer as KKT-CN1O.*

Seja Z *uma matriz cujas colunas são uma base para* $\mathcal{N}(A(x_*))$*.*

Se, para todo o λ_* *vector de multiplicadores de Lagrange a satisfazer as KKT-CN1O,*

$$Z^\top \nabla^2_{xx} L(x_*, \lambda_*) Z \text{ é PD}$$

então

$$x_* \text{ é um minimizante local estrito.}$$

Nota:

- No caso em que é verificada a qualificação de restrições LICQ, Z pode ser calculada na forma

$$\begin{aligned} A^\top_{(n\times m)} &= Q_{(n\times n)} \left[\begin{array}{c} R \\ 0 \end{array} \right]_{(n\times m)} \\ &= \left[\begin{array}{cc} Q_{1_{(n\times m)}} & Q_{2_{(n\times (n-m))}} \end{array} \right] \left[\begin{array}{c} R \\ 0 \end{array} \right] \end{aligned}$$

onde R é uma matriz não singular e triangular superior e $Q_2 = Z$.

Exercício:

- Mostre que se (x_*, λ_*) satisfaz as CN1O e

$$d^\top \nabla^2_{xx} L(z, \lambda_*) d \ \geq \ 0, \quad \forall z, d \in \mathbb{R}^n$$

então x_* é um minimizante global.

Utilizando este resultado, resolva:

  -
$$\begin{array}{ll} \min\limits_{x \in \mathbb{R}^n} & \|x\|_2 \\ \text{sujeito a} & Ax \ = \ b \quad \text{em que } b \in \mathbb{R}^p, \ A \in \mathbb{R}^{p \times n} \text{ e } car(A) = p. \end{array}$$

  -
$$\begin{array}{ll} \min\limits_{x \in \mathbb{R}^2} & 3x_1^2 + 2x_2^2 \\ \text{sujeito a} & 25 - x_1^2 - x_2 \ \geq \ 0 \\ & -27 + 9x_1 - x_2^2 \ \geq \ 0 \end{array}$$

Presentation outline

1. Métodos numéricos para optimização não linear sem restrições
   - Métodos de procura directa
   - Métodos de procura unidireccional (geral)
   - Métodos de procura unidireccional (descida máxima, Newton)
   - Métodos de região de confiança
2. Teoria da optimização não linear com restrições
   - Introdução e qualificações de restrições
   - Condições de primeira ordem
   - Condições de segunda ordem
   - Teoria da dualidade
3. Métodos numéricos para optimização não linear com restrições
   - Método da penalização quadrática
   - Método do Lagrangeano aumentado
   - Programação Sequencial Quadrática (SQP)
   - Métodos de pontos interiores

Para finalizar a teoria da Optimização ou Programação Não Linear, vamos estudar a **dualidade**.

A dualidade pode ser motivada através do seguinte **jogo**:

$\longmapsto$ O jogador 1 escolhe a estratégia $x \in X$.

$\longmapsto$ O jogador 2 escolhe a estratégia $y \in Y$.

Ambos os jogadores actuam de forma racional mas de modo independente um do outro.

O que um perde é igual ao que o outro ganha (**jogo de soma nula**).

(Suponhamos que o primeiro perde e o segundo ganha.)

$\longmapsto$ O jogador 1 minimiza em X o pior cenário, dado por

$$\max_{y \in Y} F(x,y) \quad \overset{\text{def}}{=} \quad F_p(x) \qquad \text{(função primal)}$$

$\longmapsto$ O jogador 2 maximiza em Y o pior cenário, dado por

$$\min_{x \in X} F(x,y) \quad \overset{\text{def}}{=} \quad F_d(y) \qquad \text{(função dual)}$$

Surgem, assim, os problemas

PRIMAL $\qquad \displaystyle\min_{x \in X} F_p(x)$

DUAL $\qquad \displaystyle\max_{y \in Y} F_d(y)$

Exemplo:

$$X = Y = \{1,2\}, \quad F(i,j) = a_{ij}, \quad \text{em que} \quad A = \begin{bmatrix} 1 & 4 \\ 2 & -3 \end{bmatrix}.$$

PRIMAL $$\min_{i \in \{1,2\}} \max_{j \in \{1,2\}} F(i,j) \;=\; \min_{i \in \{1,2\}} \{4,2\} \;=\; 2 \qquad (i_* = 2)$$

DUAL $$\max_{j \in \{1,2\}} \min_{i \in \{1,2\}} F(i,j) \;=\; \max_{j \in \{1,2\}} \{1,-3\} \;=\; 1 \quad (j_* = 1)$$

Teorema (Dualidade Fraca)

$$F_d(y) \;\leq\; F_p(x),\; \forall x \in X, y \in Y$$

Demonstração:

$$F_d(y) \;=\; \min_{x' \in X}\; F(x', y) \;\leq\; F(x, y) \;\leq\; \max_{y' \in Y}\; F(x, y') \;=\; F_p(x)$$

$\blacksquare$

A dualidade fraca implica:

$$\max_{y \in Y}\; F_d(y) \;\leq\; \min_{x \in X}\; F_p(x)$$

Teorema (Dualidade Forte)

$$\max_{y \in Y} \; F_d(y) \; = \; \min_{x \in X} \; F_p(x)$$

se e só se

$$\exists (x_*, y_*) \in X \times Y \, (\textit{ponto de sela}) :$$

$$F(x_*, y) \; \leq \; F(x_*, y_*) \; \leq \; F(x, y_*), \; \forall (x, y) \in X \times Y$$

Exemplo:

No exemplo anterior não existe ponto de sela. Porém, se considerarmos

$$X = Y = \{1, 2\}, \quad F(i, j) = a_{ij}, \quad \text{em que} \quad A = \begin{bmatrix} 1 & 4 \\ 0 & -3 \end{bmatrix}.$$

O ponto $(i_*, j_*) = (2, 1) \in X \times Y$ é **ponto de sela**.

O zero é o maior na sua linha X.

O zero é o menor na sua coluna Y.

Neste caso, como

$$F(i,j) = a_{ij}, \quad \text{em que} \quad A = \left[\begin{array}{rr} 1 & 4 \\ 0 & -3 \end{array} \right].$$

vem que

PRIMAL $$\min_{i \in \{1,2\}} \max_{j \in \{1,2\}} F(i,j) \;=\; \min_{i \in \{1,2\}} \{4, 0\} \;=\; 0 \qquad (i_* = 2)$$

DUAL $$\max_{j \in \{1,2\}} \min_{i \in \{1,2\}} F(i,j) \;=\; \max_{j \in \{1,2\}} \{0, -3\} \;=\; 0 \quad (j_* = 1)$$

Analisemos, agora, a dualidade (Lagrangeana) em Optimização Não Linear, considerando

$$\begin{array}{rl} \min\limits_{x \in \mathbb{R}^n} & f(x) \\ \text{sujeito a} & x \in X \\ & c_i(x) \geq 0, \ i \in \mathcal{I} \end{array}$$

onde se pressupõe que $X = \mathbb{R}^n$.

Sejam

$$F(x, y) = L(x, \lambda) = f(x) - \sum_{i \in \mathcal{I}} \lambda_i c_i(x)$$

e

$$Y = \{\lambda \in \mathbb{R}^{|\mathcal{I}|} : \ \lambda \ \geq \ 0\}$$

A função primal toma a forma

$$
\begin{aligned}
L_p(x) &= \sup_{\lambda \geq 0} L(x, \lambda) \\
&= f(x) + \sup_{\lambda \geq 0} \sum_{i \in \mathcal{I}} \overbrace{(-\lambda_i)}^{\leq 0} c_i(x) \\
&= \begin{cases} f(x) & \text{se} \quad x \in \Omega \\ +\infty & \text{se} \quad x \notin \Omega \end{cases}
\end{aligned}
$$

Logo, o problema primal

$$
\min_{x \in X} L_p(x) \quad \Longleftrightarrow \quad \text{PNL}
$$

Analisemos, agora, o problema dual.

A função dual é

$$L_d(\lambda) = \inf_{x \in X} L(x, \lambda) = q(\lambda)$$

e o seu domínio é

$$\mathcal{D} = \{\lambda : \ q(\lambda) > -\infty\}$$

Teorema

q é uma função côncava e $\mathcal{D}$ é um conjunto convexo.

Demonstração:

$$L(x, \alpha\bar{\lambda} + (1-\alpha)\bar{\bar{\lambda}}) = \alpha L(x, \bar{\lambda}) + (1-\alpha) L(x, \bar{\bar{\lambda}}), \ \forall \alpha \in [0,1]$$

Logo,

$$q(\alpha\bar{\lambda} + (1-\alpha)\bar{\bar{\lambda}}) \ \geq \ \alpha q(\bar{\lambda}) + (1-\alpha) q(\bar{\bar{\lambda}})$$

Suponhamos, agora, que $\bar{\lambda}, \bar{\bar{\lambda}} \in \mathcal{D}$. A desigualdade anterior implica que

$$q(\alpha\bar{\lambda} + (1-\alpha)\bar{\bar{\lambda}}) > -\infty$$

Logo, $\alpha\bar{\lambda} + (1-\alpha)\bar{\bar{\lambda}} \in \mathcal{D}$.

∎

Nota:

- O problema dual

$$\max_{\lambda \geq 0} \; L_d(\lambda)$$

é sempre convexo.

A dualidade fraca diz-nos que

$$q(\lambda) = L_d(\lambda) \;\leq\; L_p(x), \; \forall \lambda \geq 0, x \in X = \mathbb{R}^n$$

Mas,

$$L_p(x) = \begin{cases} f(x) & \text{se} \quad x \in \Omega \\ +\infty & \text{se} \quad x \notin \Omega \end{cases}$$

Logo a dualidade fraca em Optimização Não Linear é dada por

$$q(\lambda) \;\leq\; f(x), \; \forall \lambda \geq 0, x \in \Omega$$

No caso convexo, é possível relacionar os multiplicadores de Lagrange com a solução do problema dual.

Teorema

Sejam $f, -c_i \in C^1(\mathbb{R}^n)$ funções convexas em $\mathbb{R}^n$.

Então qualquer $\bar{\lambda}$, com $(\bar{x}, \bar{\lambda})$ a satisfazer as CN1O, é uma solução (global!) do problema dual.

Demonstração: Observemos que

(i) $L(\cdot, \bar{\lambda})$ é convexa, porque:

- $L(\cdot, \bar{\lambda}) = f(\cdot) - \sum_{i \in \mathcal{E} \cup \mathcal{I}} \lambda_i c_i(\cdot)$
- $\lambda_i \geq 0, \; i \in \mathcal{I}$
- $f(\cdot)$ e $-c_i(\cdot)$ são convexas

(ii) $L(\cdot, \bar{\lambda}) \in C^1(\mathbb{R}^n)$

Logo,

$$\forall x \in \mathbb{R}^n: \; L(x,\bar{\lambda}) \;\geq\; L(\bar{x},\bar{\lambda}) + \overbrace{\nabla_x L(\bar{x},\bar{\lambda})^\top}^{=0 \text{ (CN1O)}}(x-\bar{x}) = L(\bar{x},\bar{\lambda})$$

Assim, temos

$$q(\bar{\lambda}) = \inf_{x\in\mathbb{R}^n} L(x,\bar{\lambda}) = L(\bar{x},\bar{\lambda}) \overbrace{=}^{\text{CN1O}} f(\bar{x})$$

Mas, pela dualidade fraca,

$$q(\lambda) \;\leq\; f(\bar{x}), \; \forall \lambda \geq 0$$

Desta forma, $\bar{\lambda}$ é uma solução de $\sup_{\lambda \geq 0} q(\lambda)$.

■

Vimos, nesta demonstração, que em programação convexa ocorre dualidade forte:

$$q(\bar{\lambda}) \;=\; f(\bar{x}),$$

o que, aliás, mostra que $\bar{x}$ é solução (global, claro) do primal.

Neste caso diz-se que o *gap* dual é nulo e que $(\bar{x}, \bar{\lambda})$ é um ponto de sela para o par primal–dual (no sentido de minimizar a função primal e maximizar a função dual).

Analisemos a aplicação à **Programação Linear**.

Considere-se o **PL**

$$\begin{array}{rl} \displaystyle\min_{x \in \mathbb{R}^n} & c^\top x \\ \text{sujeito a} & Ax \;\geq\; b \end{array}$$

A função Lagrangeana é dada por

$$L(x, \lambda) = c^\top x - \lambda^\top (Ax - b)$$

A função dual é

$$
\begin{aligned}
q(\lambda) = L_d(\lambda) &= \inf_{x \in \mathbb{R}^n} \{c^\top x - \lambda^\top (Ax - b)\} \\
&= \inf_{x \in \mathbb{R}^n} \{(c - A^\top \lambda)^\top x\} + b^\top \lambda \\
&= \begin{cases} b^\top \lambda & \text{se} \quad c - A^\top \lambda = 0 \\ -\infty & \text{caso contrário} \end{cases}
\end{aligned}
$$

Logo, o problema dual é equivalente a

$$
\begin{aligned}
\max_{\lambda} \quad & b^\top \lambda \\
\text{sujeito a} \quad & A^\top \lambda \, = \, c \\
& \lambda \, \geq \, 0
\end{aligned}
$$

Prova-se, recorrendo ao resultado anterior, que o dual de um programa linear, escrito na **forma standard ou padrão**

$$\begin{array}{rrcll} \min\limits_{x} & c^\top x & & & \\ \text{sujeito a} & Ax & = & b & \qquad \textbf{(P)} \\ & x & \geq & 0 & \end{array}$$

é dado por

$$\begin{array}{rrcll} \max\limits_{\lambda} & b^\top \lambda & & & \\ \text{sujeito a} & A^\top \lambda & \leq & c & \qquad \textbf{(D)} \end{array}$$

Exercício:

- Prove que o dual de um programa linear, escrito na forma standard ou padrão

$$\begin{array}{rrcl} \min\limits_{x} & c^\top x & & \\ \text{sujeito a} & Ax & = & b \\ & x & \geq & 0 \end{array} \qquad \textbf{(P)}$$

  é dado por

$$\begin{array}{rrcl} \max\limits_{\lambda} & b^\top \lambda & & \\ \text{sujeito a} & A^\top \lambda & \leq & c \end{array} \qquad \textbf{(D)}$$

As CN1O-KKT de (P) são

$$
\begin{aligned}
A^\top \lambda + s &= c \\
Ax &= b \\
x &\geq 0 \\
s &\geq 0 \\
x^\top s &= 0
\end{aligned}
$$

Seja (x_*, λ_*, s_*) a satisfazer as CN1O-KKT.

Então, para qualquer x tal que $Ax = b,\ x \geq 0$, temos

$$
\begin{array}{rcccl}
c^\top x & = & \left(A^\top \lambda_* + s_*\right)^\top x & = & b^\top \lambda_* + s_*^\top x \\
\geq \quad b^\top \lambda_* & = & (Ax_*)^\top \lambda_* & = & x_*^\top (c - s_*) \\
 & & & = & c^\top x_*
\end{array}
$$

Logo, estas condições são também **suficientes** para a optimização (global) do primal (P).

Exercício:

- Escreva as CN1O-KKT para (D). Verifique que são também suficientes para a optimização (global) de (D).

Analisemos a aplicação à **Programação Quadrática**.

No caso da programação quadrática estamos interessados em saber qual o dual do **PQ estritamente convexo**

$$\begin{array}{rl} \displaystyle\min_{x\in\mathbb{R}^n} & g^\top x + \frac{1}{2}x^\top Hx \\ \text{sujeito a} & Ax \;\geq\; b \end{array}$$

em que H é uma matriz simétrica e PD. A função dual é dada por

$$\inf_{x\in\mathbb{R}^n} L(x,\lambda) = \inf_{x\in\mathbb{R}^n} \left\{ g^\top x + \frac{1}{2}x^\top Hx - \lambda^\top(Ax-b) \right\}$$

Como $L(\cdot,\lambda)$ é uma quadrática estritamente convexa, o ínfimo/mínimo é atingido em

$$Hx + g - A^\top \lambda = 0 \quad \Longleftrightarrow \quad x = H^{-1}(A^\top \lambda - g)$$

Assim,

$$\begin{aligned} g^\top x + \frac{1}{2} x^\top H x - \lambda^\top (Ax - b) &= \frac{1}{2} x^\top H x + g^\top x - \lambda^\top A x + b^\top \lambda \\ &= \frac{1}{2} x^\top H x + (-A^\top \lambda + g)^\top x + b^\top \lambda \\ &= \frac{1}{2} x^\top H x + (-Hx)^\top x + b^\top \lambda \\ &= -\frac{1}{2} x^\top H x + b^\top \lambda \end{aligned}$$

Substituindo, na última expressão, por

$$x = H^{-1}(A^\top \lambda - g)$$

temos

$$q(\lambda) = L_d(\lambda) = -\frac{1}{2}(A^\top \lambda - g)^\top H^{-1}(A^\top \lambda - g) + b^\top \lambda$$

O dual também é, assim, um PQ estritamente convexo.

No caso, mais geral, em que temos (com H matriz simétrica)

$$\begin{array}{rl} \min\limits_{x\in\mathbb{R}^n} & g^\top x + \frac{1}{2}x^\top H x \\ \text{sujeito a} & a_i^\top x \;=\; b_i, \quad i \in \mathcal{E} \\ & a_i^\top x \;\geq\; b_i, \quad i \in \mathcal{I} \end{array}$$

As CN1O-KKT são dadas por

$$\begin{array}{rcll} Hx + g - \sum\limits_{i\in\mathcal{A}(x)} \lambda_i a_i & = & 0 & \text{(dualidade)} \\ a_i^\top x & = & b_i, \quad i \in \mathcal{A}(x) & \text{(admissib., complem.)} \\ a_i^\top x & > & b_i, \quad i \in \mathcal{I}, i \notin \mathcal{A}(x) & \text{(admissib., complem.)} \\ \lambda_i & \geq & 0, \quad i \in \mathcal{A}(x) \cap \mathcal{I} & \text{(não negatividade)} \end{array}$$

Teorema

Seja $x_* \in \mathbb{R}^n$ a satisfazer as CN1O-KKT, com multiplicadores $\lambda_i^*,\ i \in \mathcal{A}(x_*)$. Seja H PSD.

Então x_* é um minimizante global do PQ (ou seja, estas condições são suficientes para a optimalidade global).

Demonstração: Seja x um qualquer ponto admissível. Temos que

$$f(x) = f(x_*) + \underbrace{(x - x_*)^\top (Hx_* + g)}_{?} + \underbrace{\frac{1}{2}(x - x_*)^\top H(x - x_*)}_{\geq 0}$$

Porém,

$$(x - x_*)^\top (Hx_* + g) = \sum_{i \in \mathcal{E}} \lambda_i^* \underbrace{a_i^\top (x - x_*)}_{= 0} + \sum_{i \in \mathcal{A}(x_*) \cap \mathcal{I}} \overbrace{\lambda_i^*}^{\geq 0} \underbrace{a_i^\top (x - x_*)}_{\geq 0} \ \geq\ 0$$

Logo, $f(x) \geq f(x_*)$. ■

Quando $\mathcal{I} = \emptyset$, ocorre um resultado mais forte.

Consideremos, então

$$\begin{array}{rl} \min\limits_{x \in \mathbb{R}^n} & g^\top x + \frac{1}{2} x^\top H x \\ \text{sujeito a} & Ax \;=\; b \end{array}$$

As CN1O são dadas por

$$\left\{ \begin{array}{rcl} Hx_* + g - A^\top \lambda_* & = & 0 \\ Ax_* & = & b \end{array} \right.$$

Teorema

Seja $x_* \in \mathbb{R}^n$ a satisfazer as CN1O. Seja $Z^\top H Z$ PSD, em que Z é uma matriz cujas colunas são uma base para $\mathcal{N}(A)$.

Então x_* é um minimizante global do PQ (ou seja, estas condições são suficientes para a optimalidade global).

Demonstração: Seja x tal que $Ax = b$ (admissível).

Tem-se

$$\begin{aligned} f(x) &= f(x_*) + (x - x_*)^\top (Hx_* + g) + \tfrac{1}{2}(x - x_*)^\top H (x - x_*) \\ &= f(x_*) + \underbrace{(x - x_*)^\top (A^\top \lambda_*)}_{=0} + \underbrace{\frac{1}{2}(Zu)^\top H (Zu)}_{\geq 0} \end{aligned}$$

Logo, $f(x) \geq f(x_*)$.

■

As CN1O podem ser reescritas (mudando λ_* para $-\lambda_*$), do seguinte modo

$$\left\{ \begin{array}{rcl} Hx_* + g + A^\top \lambda_* & = & 0 \\ Ax_* & = & b \end{array} \right. \Longleftrightarrow \left[\begin{array}{cc} H & A^\top \\ A & 0 \end{array} \right] \left[\begin{array}{c} x_* \\ \lambda_* \end{array} \right] = - \left[\begin{array}{c} g \\ -b \end{array} \right]$$

Teorema

Seja $Z^\top H Z$ *PD, em que* Z *é uma matriz cujas colunas são uma base para* $\mathcal{N}(A)$*.*

Seja car$(A) =$ *número de linhas de* A*.*

A chamada matriz KKT

$$\begin{bmatrix} H & A^\top \\ A & 0 \end{bmatrix}$$

é não singular.

Nota:

- Nas condições do último teorema, os multiplicadores são únicos.

Exercício:

- Prove o último teorema.