IA 810 Otimização de Sistemas de Grande Porte

2-Programação Não Linear

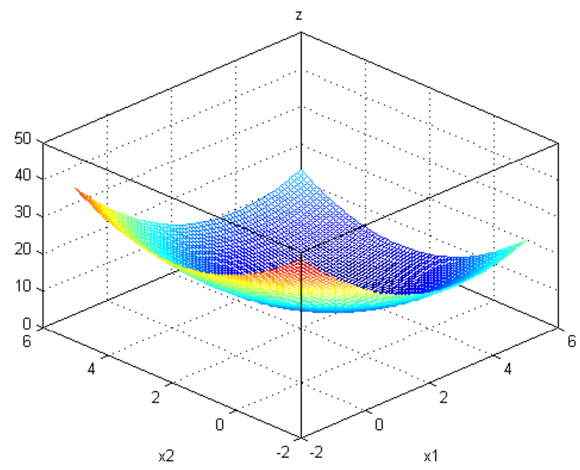
Conteúdo

1. Introdução
2. Convexidade
3. Condições de Karush-Kuhn-Tucker
4. Pontos de sela e condições suficientes
5. Métodos de programação não linear

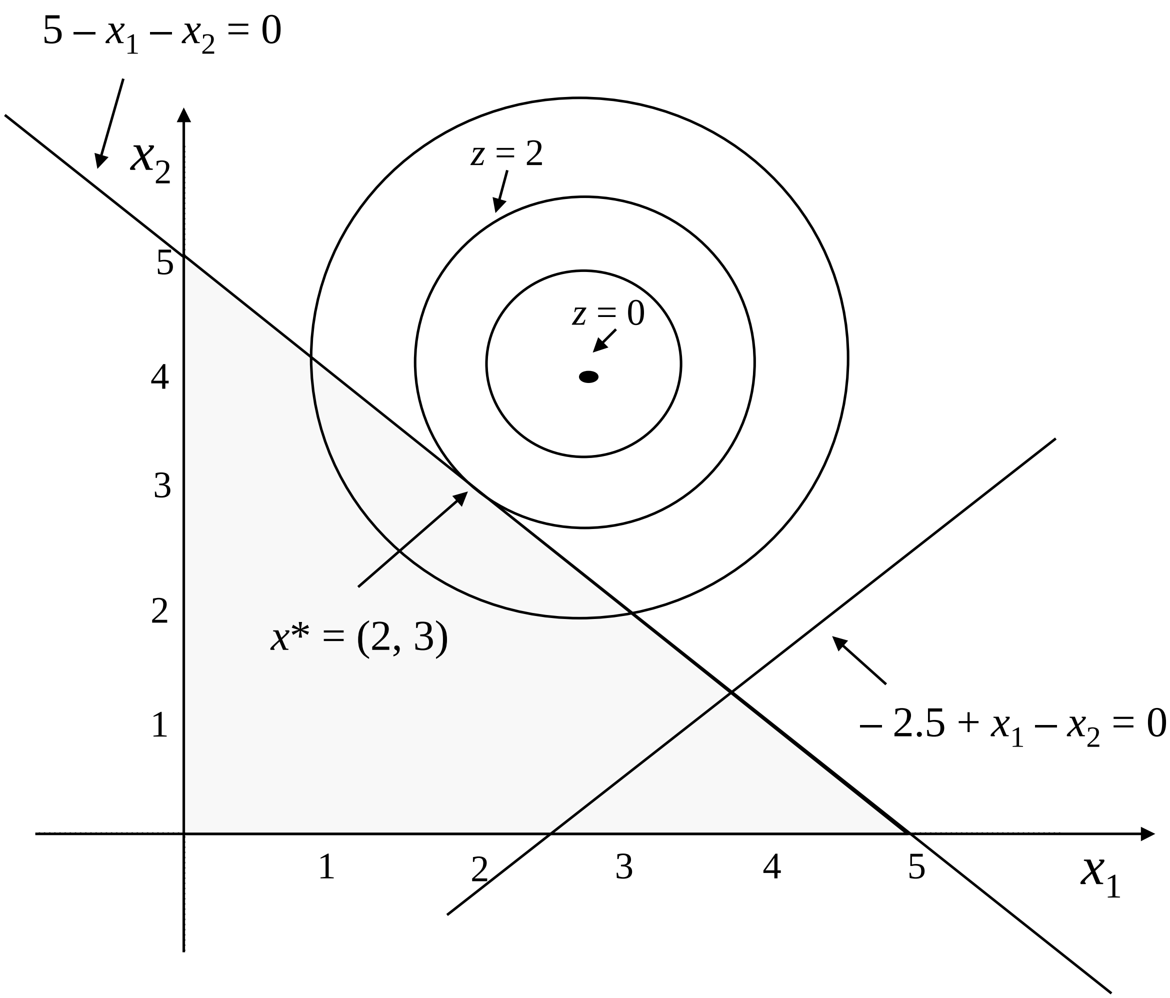
1-Introdução

- Geometria da programação não linear

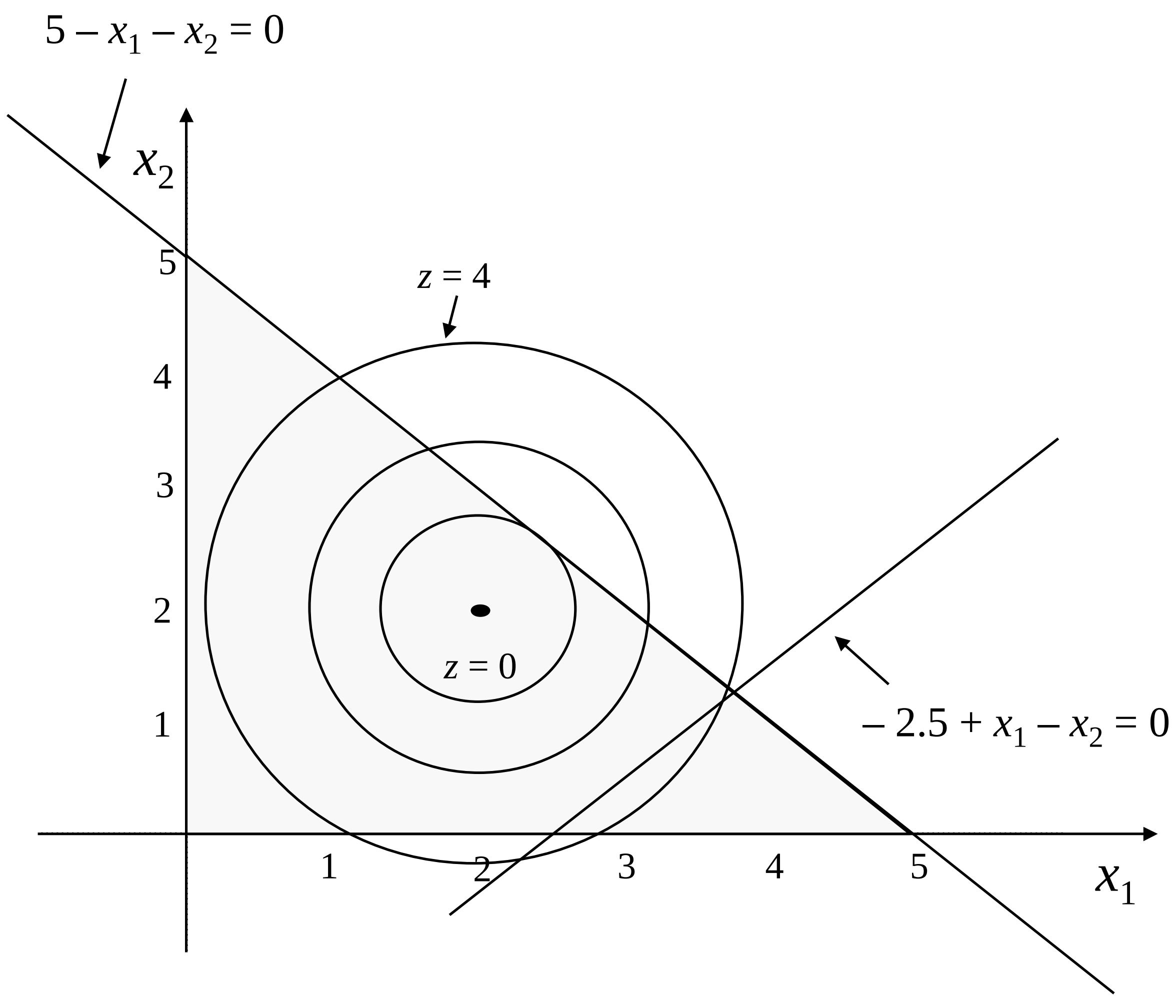
$$
\begin{aligned}
min\ \ z &= (x_1-3)^2+(x_2-4)^2 \\
sa \qquad & 5-x_1-x_2 \geq 0 \\
& -2.5+x_1-x_2 \leq 0 \\
& x_1,\ x_2 \geq 0
\end{aligned}
$$



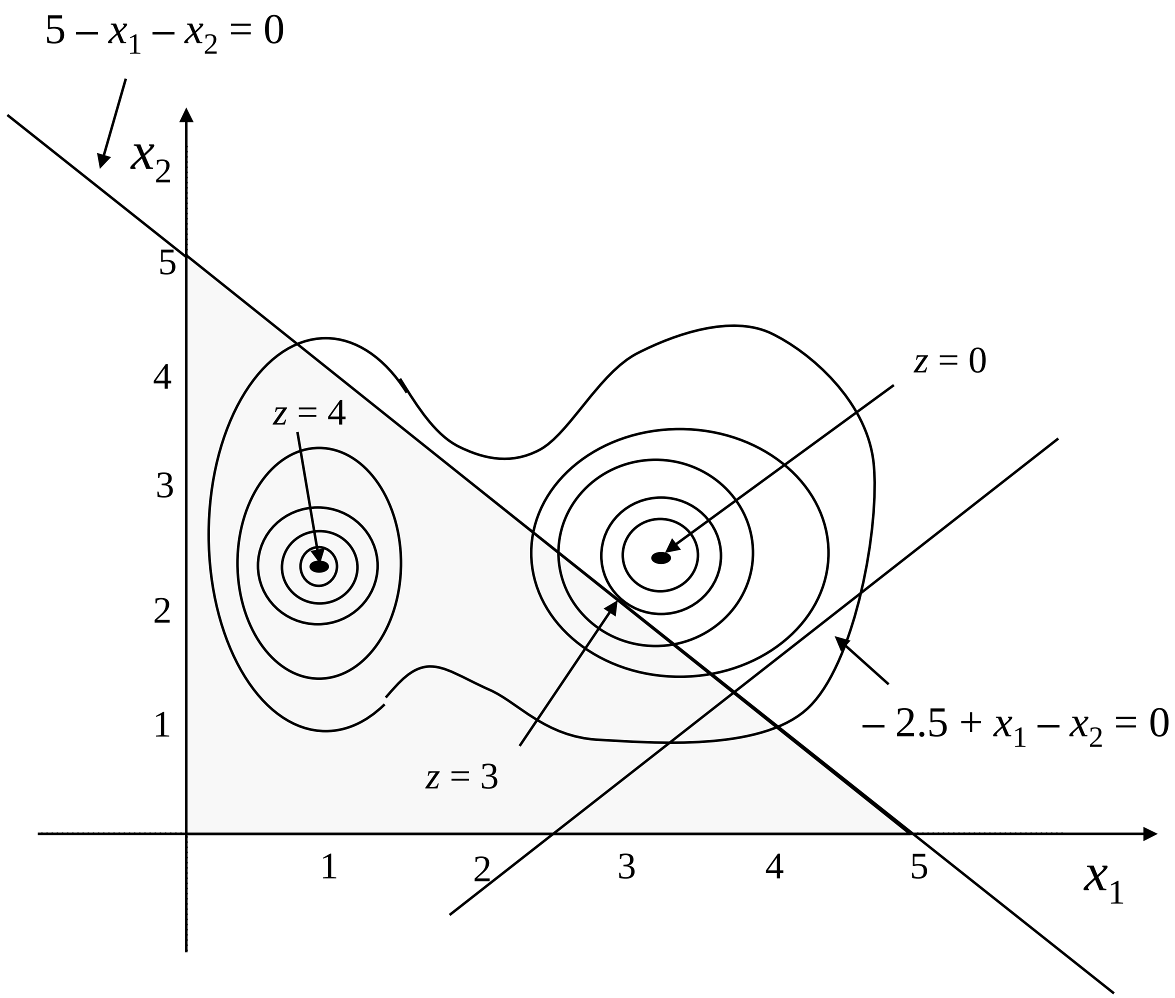
- Solução ótima na fronteira das restrições



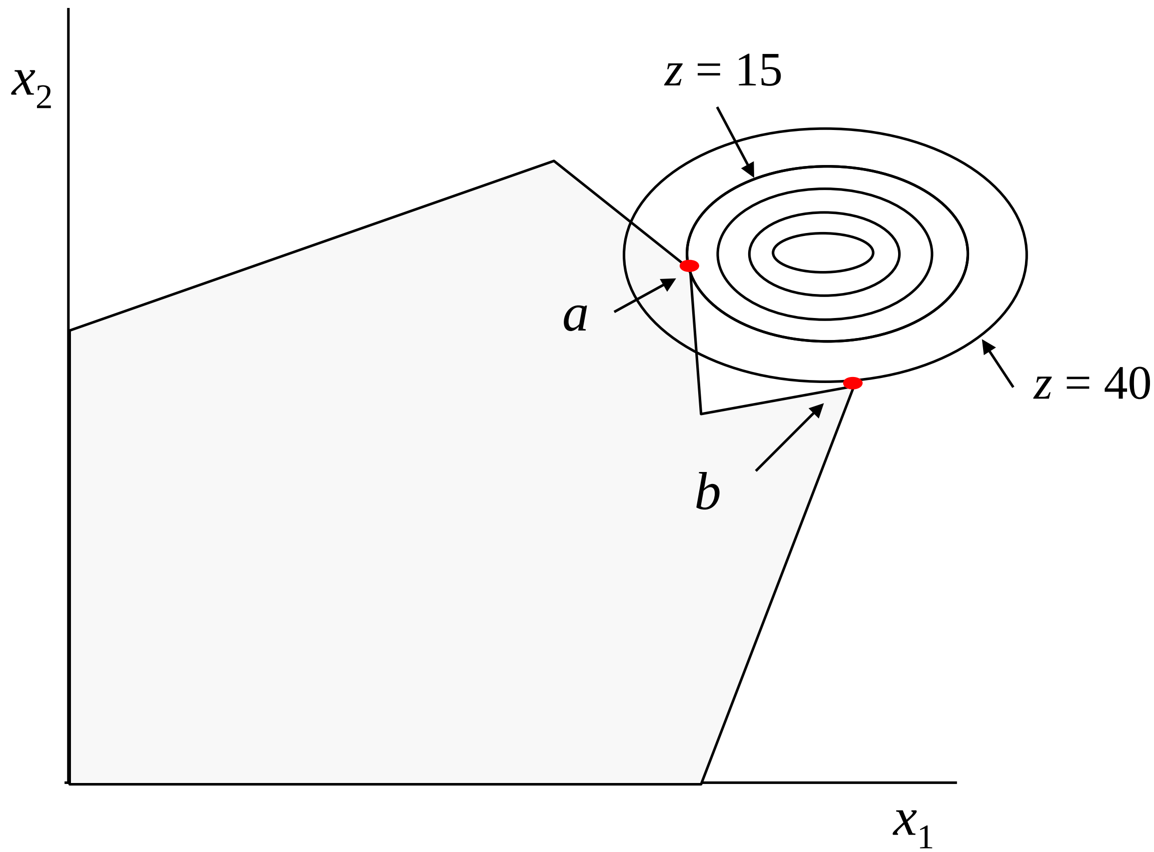
- Solução ótima no interior das restrições



- Ótimo local devido à função objetivo



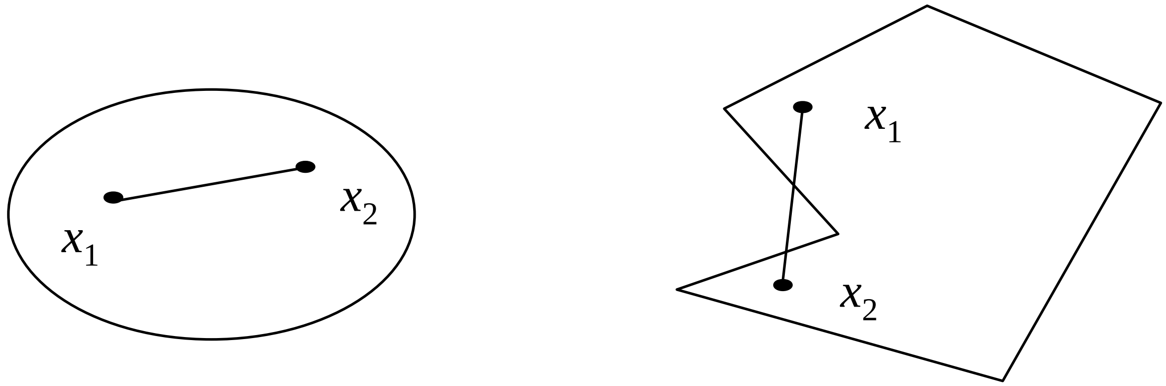
- Ótimo local devido às restrições



2-Convexidade

Definição 1: um conjunto é convexo se, dados dois pontos quaisquer x_1 e x_2 do conjunto, o segmento que os une também pertence ao conjunto.

$$S = \{\ x \mid x = \lambda x_1 + (\ 1 - \lambda)\ x_2,\ 0 \leq \lambda \leq 1\ \}$$

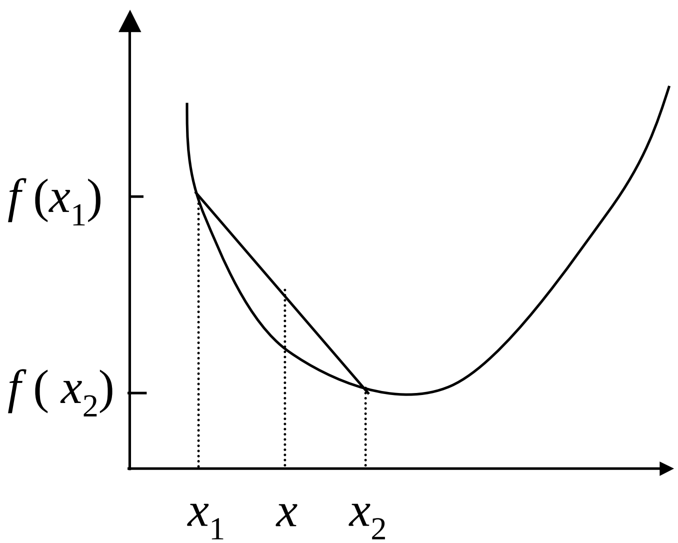


convexo

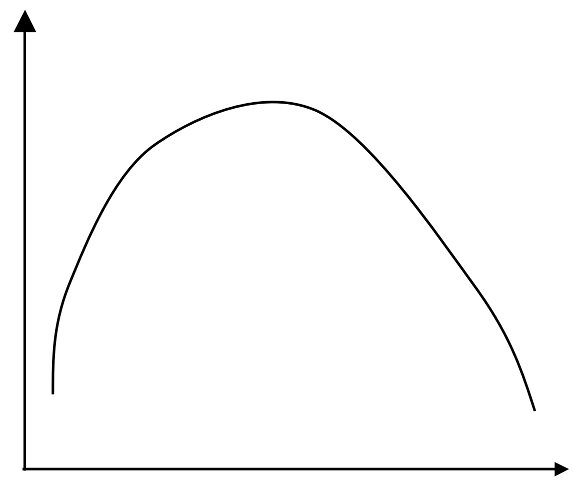
não convexo

Definição 2: A função $f\,(x)$ é convexa se o segmento de reta que une dois pontos quaisquer de seu grafo nunca está abaixo do grafo (côncava se nunca está acima).

$$f\,(\lambda x_1 + (\,1 - \lambda)\,x_2) \leq \lambda f\,(x_1) + (\,1 - \lambda) f\,(\,x_2) \qquad (\geq \text{côncava})$$



convexa



côncava

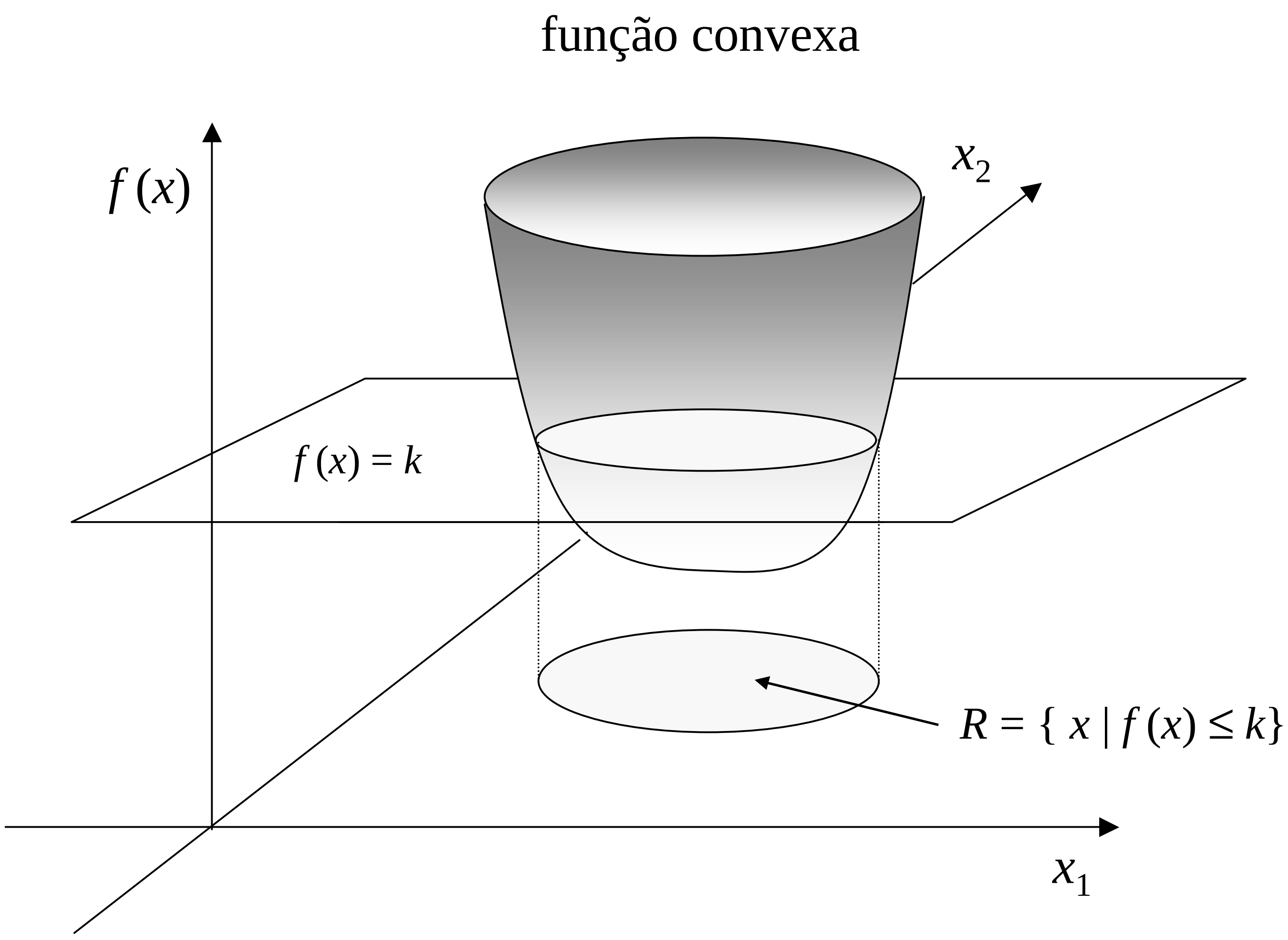
Teoremas básicos da programação não linear

Teorema 1: qualquer mínimo local de um modelo de otimização convexo é um mínimo global.

Teorema 2: Se $f(x)$ é convexa, então o conjunto $R = \{ x \mid f(x) \leq k \}$ é convexo para todo escalar k.

Teorema 3: a intersecção de conjuntos convexos é um conjunto convexo.

Teorema 2

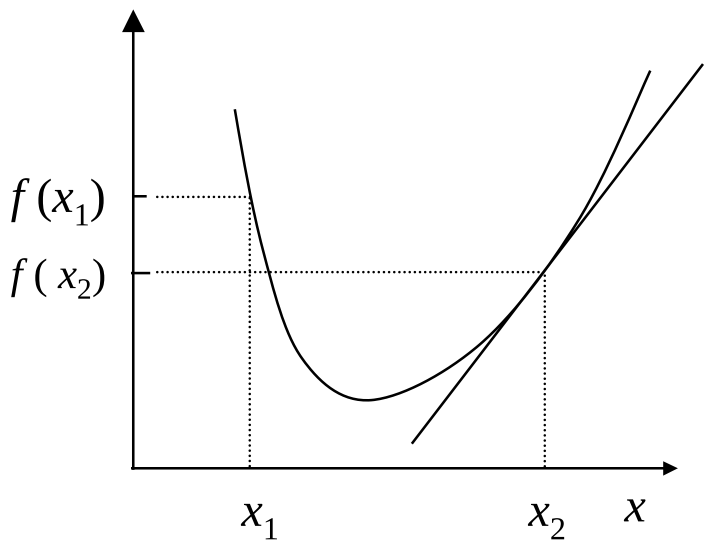


Teorema 4: se $f(x)$ possui a primeira e a segunda derivadas contínuas, então as seguintes afirmações são equivalentes:

(a) f (x) é convexa

(b) $f(x_1) \geq f(x_2) + \nabla f(x_2)(x_1 - x_2)$

(c) a matriz derivada segunda de $f(x)$ é positiva definida para todos x.



(b)

Teorema 5: uma forma quadrática semi positiva definida é convexa.

Teorema 6: uma combinação linear positiva de funções convexas é convexa.

Teorema 7: a função $f(x)$ é convexa se se somente se a função unidimensional $g(\alpha) = f(x + \alpha s)$ é convexa para todo x fixo e s.

- Problemas com restrições de igualdade

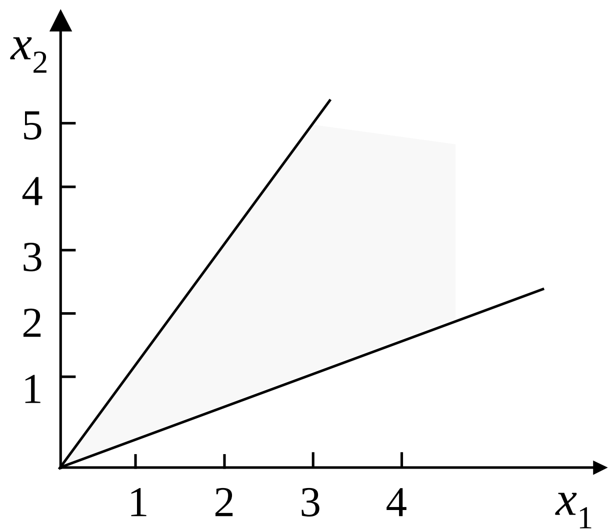
$$\begin{aligned} &min \quad f(x) \\ &sa \quad g_i(x) \leq 0, \quad i = 1, \ldots, m \\ &\qquad h_j(x) = 0, \quad j = 1, \ldots, r < n \end{aligned}$$

$R = \{ x \mid h(x) = 0\}$ não necessariamente convexo

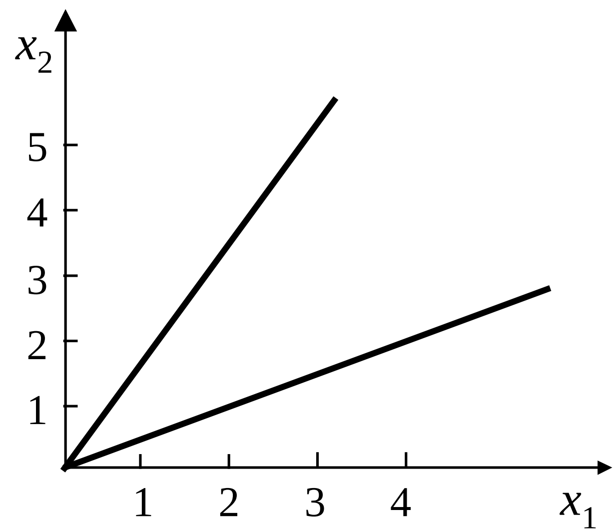
3-Condições Karush-Kuhn-Tucker

Cone: um conjunto R é um cone se $x \in R$, então $\lambda x \in R$ para $\lambda \geq 0$.

$$R = \{ x \mid x \in R \Rightarrow \lambda x \in R , \lambda \geq 0\}$$



cone convexo



cone não convexo

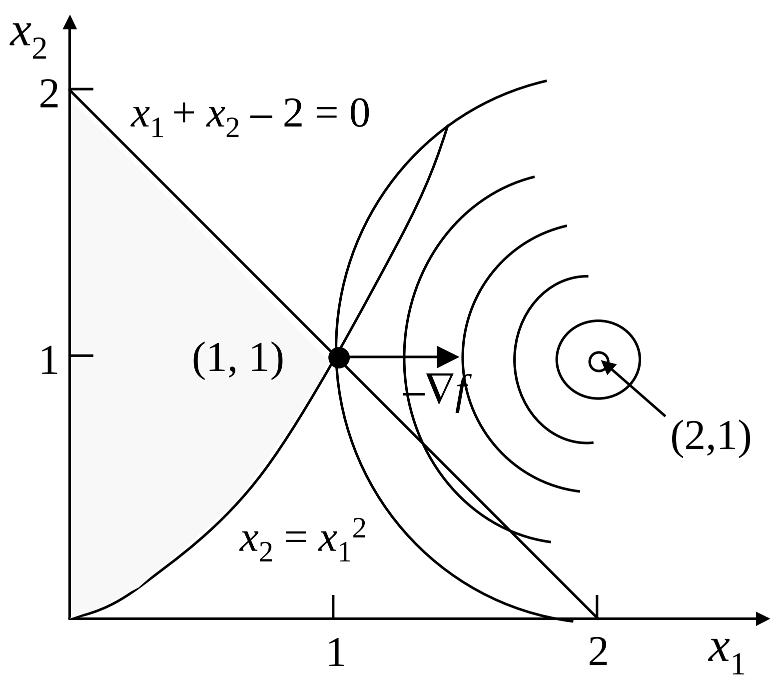
Se x^1, x^2, ..., x^m é um conjunto finito de vetores, então o conjunto R de todas combinações lineares não negativas destes vetores é um cone convexo.

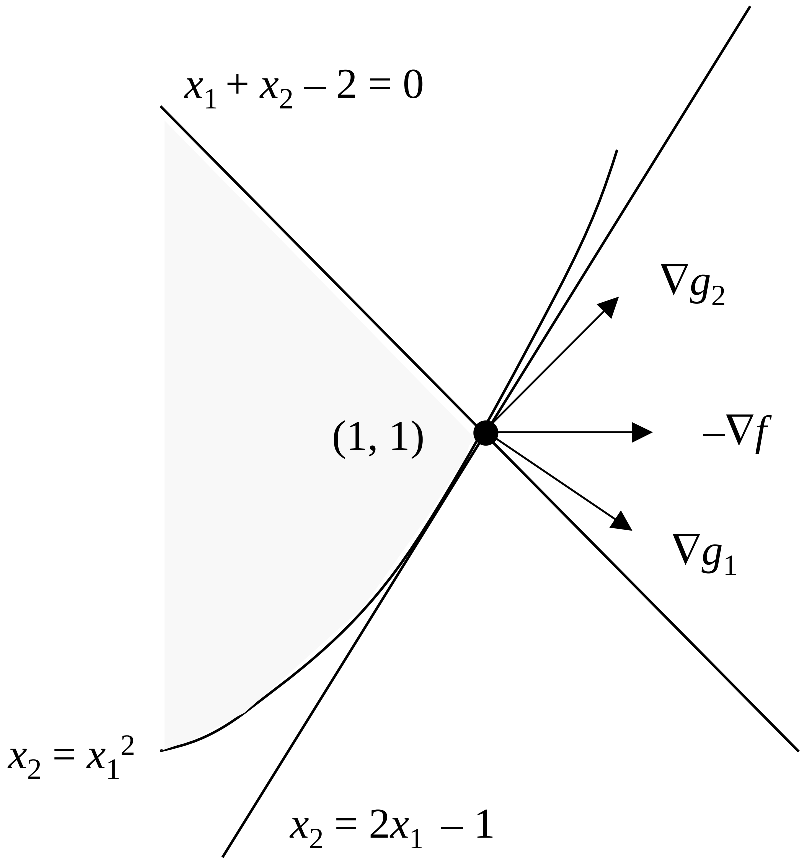
$$R = \{\ x \mid x = \lambda_1 x^1 + \lambda_2 x^2 + \ldots + \lambda_m x^m,\ \lambda_i \geq 0,\ i = 1,\ldots,m\}$$

x^1, x^2, ..., x^m são os vetores geradores do cone

- Teorema de KKT: interpretação geométrica

$$\begin{aligned} &\min \quad f(x_1,x_2) = (x_1-2)^2 + (x_2-1)^2 \\ &sa \quad g_1(x_1,x_2) = x_1^2 - x_2 \le 0 \\ &\qquad g_2(x_1,x_2) = x_1 + x_2 - 2 \le 0 \end{aligned}$$





$$\nabla f(x^\circ) = \sum_{i \in I} u_i^\circ (-\nabla g_i(x^\circ))$$

$$u_i^\circ \geq 0, \; i \in I$$

- Condições de Karush-Kuhn-Tucker

$$\nabla f(x^\circ) = \sum_{i \in I} u_i^\circ (-\nabla g_i(x^\circ)) \qquad (5)$$

$$u_i^\circ \geq 0$$

$$u_i^\circ g_i(x^\circ) = 0$$

$$g_i(x^\circ) \leq 0 \qquad (6)$$

$$i = 1, \ldots, m$$

- Condições de KKT e multiplicadores de Lagrange

$$L(x,u) = f(x) + \sum_{i=1}^{m} u_i g_i(x)$$

(5) e (6) $\Rightarrow$ $L(x, u)$ estacionário em (x^o, u^o), com u^o satisfazendo (6)

- Derivação das condições de Karush-Kuhn-Tucker

$$
\begin{aligned}
& min \quad f(x) \\
& sa \quad g_i(x) \le 0, \quad i = 1,\dots,m
\end{aligned}
\qquad (8)
$$

f, g : diferenciáveis

x^o : mínimo local

$x^o + y$: perturbação no entorno de x^o

$$B^\circ = \{i \mid g_i(x^\circ) = 0\}$$

$x^o + y$ factível (admissível) se $g_i(x^o + y) \le 0$, $i \in B^o$

$$g_i(x^\circ + y) \le 0, i \in B^\circ$$

$$g_i(x^\circ) + \nabla' g_i(x^\circ) y + O(y) \le 0, \quad i \in B^\circ$$ Taylor

$$i \in B^\circ \Rightarrow g_i(x^\circ) = 0$$

$$\nabla' g_i(x^\circ) y \le 0, \quad i \in B^\circ \tag{12}$$

y admissível $\Rightarrow$ satisfaz (12)

satisfazer (12) $\not\Rightarrow$ admissível (a não ser que y esteja qualificada)

- Exemplo

$$g_1(x) = -x_1 \le 0$$

$$g_2(x) = -x_2 \le 0$$

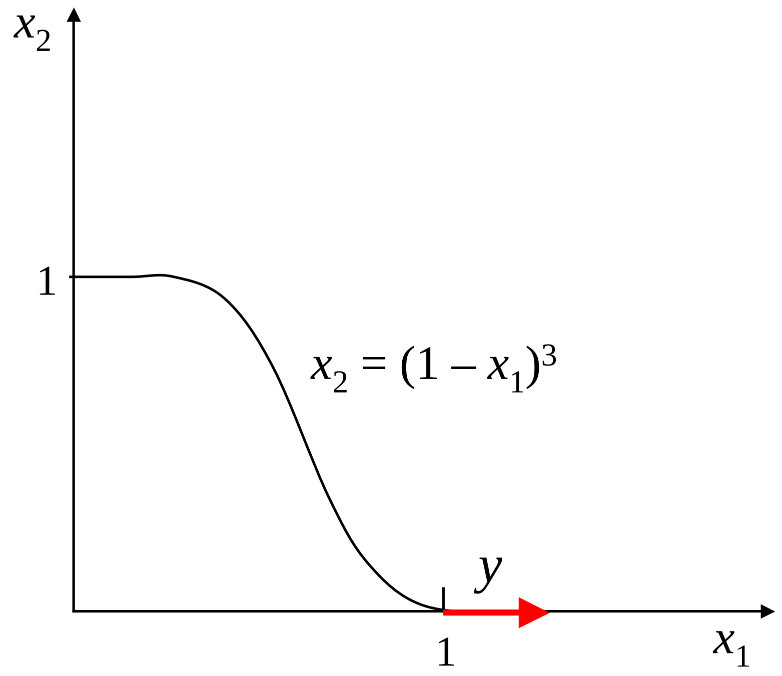
$$g_3(x) = -(1 - x_1)^3 + x_2 \le 0$$

$$x^\circ = (1,0), \quad B^\circ = \{2,3\}$$

$$\nabla' g_2(x^\circ) y = -y_2 \le 0$$

$$\nabla' g_3(x^\circ) y = \quad y_2 \le 0$$

$y = (1,0)$ satisfaz mas não é admissível



y não qualificada

- Mínimo local

  – nenhuma perturbação local factível y

$$\nabla' g_i(x^\circ) y \leq 0, \;\; i \in B^\circ$$

  – diminui função objetivo f

$$\nabla' f(x^\circ) y \geq 0$$

$$Z_2^\circ = \{ y \mid \nabla' g_i(x) y \leq 0, \; i \in B^\circ, \;\; \nabla f'(x^\circ) y < 0 \} = \phi$$

Lema de Farkas

Seja $\{P_0, P_1, .., P_r\}$ um conjunto arbitrário de vetores. Então existem $\beta_i \geq 0$ tal que:

$$P_0 = \sum_{i=1}^{r} \beta_i P_i$$

se e somente se $P_0' y \geq 0$ para todo y que satisfaz $P_i' y \geq 0, i = 1, \ldots, r$

Prova

$\rightarrow$ se $P_0 = \sum_{i=1}^{r} \beta_i P_i$ então

$$P_0' y = \sum_{i=1}^{r} \beta_i P_i' y, \ \beta_i \geq 0$$

$\forall y$ que satisfaz $P_i' y \geq 0$

$\leftarrow$ se $P_0' y \geq 0$ e $P_i' y \geq 0,\ i = 1, \ldots, r$ então o PL

$$\begin{aligned} &\min\ P_0' y \\ &\text{sa}\ \ P_i' y \geq 0,\ i = 1, \ldots, r \end{aligned}$$

tem solução ótima y = 0. Logo o problema dual é factível e tem solução ótima finita. Este dual, com variáveis $\beta_1, \ldots, \beta_r$ tem as restrições (que são factíveis):

$$P_0 = \sum_{i=1}^{r} \beta_i P_i\,,\ \beta_i \geq 0$$

Teorema 8: se f e g são diferenciáveis e x^o é um mínimo local, então existem multiplicadores $u_i^\circ \geq 0$ tal que

$$\nabla f(x^\circ) + \sum_{i=1}^{n} u_i^\circ \nabla g_i(x^\circ) = 0$$

$$u_i^\circ \nabla g_i(x^\circ) = 0, \; i = 1, \ldots, m$$

se e somente se $Z_2^\circ = \phi$

Prova

se escolhermos vetores $\nabla f(x^o)$ e $-\nabla g_i(x^o)$, $i \in B^o$ como $\{P_0,...,P_r\}$ no lema de Farkas, então existem $u_i^\circ \geq 0,\ i \in B^\circ$ tal que

$$\nabla f(x^\circ) = \sum_{i=1}^{n} u_i^\circ(-\nabla g_i(x^\circ)) = 0$$ se e somente se

$$\nabla f'(x^\circ) y \geq 0$$ para todo y que satisfaz $\nabla g_i(x^\circ) y \leq 0,\ i \in B^\circ$

Se definirmos $u_i^\circ = 0$ para $i \notin B^\circ$ então

$$\nabla f(x^\circ) + \sum_{i=1}^{n} u_i^\circ \nabla g_i(x^\circ) = 0$$

$$u_i^\circ \nabla g_i(x^\circ) = 0,\ i = 1, \ldots, m$$

- Classe de problemas onde KKT são necessárias

  – mínimo local quando $Z_2^{\circ} = \phi$

1. todas restrições são lineares

2. todas as restrições são convexas e interior conjunto restrição $\neq \phi$

3. gradientes de todas as restrições ativas são linearmente independentes

4. qualificação de restrição é satisfeita

   se x^o satisfaz $g_i(x) \leq 0$, $i = 1,\ldots,m$, g_i diferenciáveis, então a qualificação de restrição é satisfeita em x^o se todo y tal que $g'_i(x)y \leq 0$, $i \in B^o$ é tangente a um arco diferenciável que emana de x^o e está contido no conjunto restrição.

Teorema de Karush-Kuhn-Tucker

Se

a) todas funções f e g_i são diferenciáveis

b) x^o é um mínimo local para (8)

c) qualificação de restrição é verificada em x^o

então existe um vetor $u^o = (u^o_1, u^o_2, ..., u^o_m) \geq 0$ tal que as condições de KKT

$$\nabla f(x^\circ) + \sum_{i=1}^{n} u_i^\circ \nabla g_i(x^\circ) = 0$$

$$u_i^\circ \nabla g_i(x^\circ) = 0,\; i = 1, \ldots, m$$

são satisfeitas em x^o.

4-Pontos de sela e condições suficientes

– Condições de KKT

• necessárias problemas convexos e não convexos

• supõem função objetivo e restrições diferenciáveis

– Condições baseadas na função Lagrangeana

• verificadas sem necessidade de diferenciabilidade

• suficientes para quase todos modelos de programação matemática

• se o modelo é convexo e satisfaz qualificação de restrição então condições de ponto de sela são necessárias e suficientes

- Problema primal

$$\min \quad f(x) \tag{1}$$

$$\text{sa} \quad g_i(x) \le 0, \quad i = 1, \ldots, m \tag{2}$$

$$x \in S \tag{3}$$

f, g : funções arbitrárias definidas em S

x : vetor n dimensional

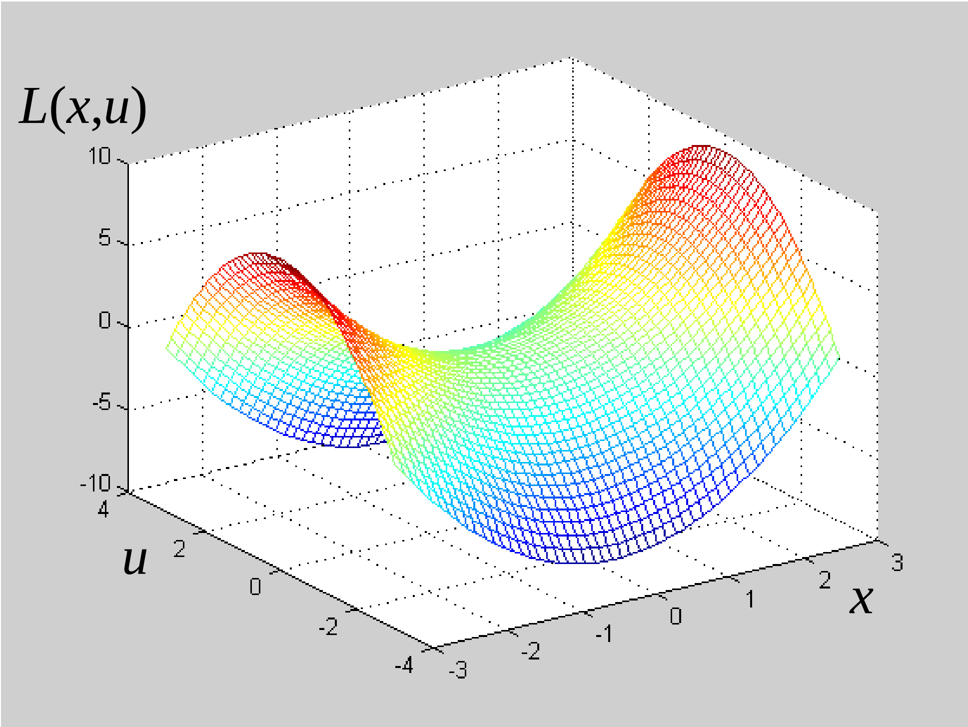
S : subconjunto arbitrário de E^n

- Função Lagrangeana associada

$$L(x,u) = f(x) + \sum_{i=1}^{m} u_i g_i(x), \; u_i \ge 0 \tag{4}$$

Definição 3: o ponto (x^o, u^o) com $u^o \geq 0$ e $x^o \in S$ é um ponto de sela de $L(x,u)$ se

$$L(x^o, u^o) \leq L(x, u^o) \quad \forall x \in S$$
$$L(x^o, u^o) \geq L(x^o, u) \quad \forall u \geq 0$$



Teorema 9: seja $u^o \geq 0$ e $x^o \in S$. Então (x^o, u^o) é um ponto de sela de $L(x, u)$ se e somente se

(a) x° minimiza $L(x, u^\circ)$ em S

(b) $g_i(x^\circ) \leq 0, \quad i = 1, \ldots, m$

(c) $u_i^\circ g_i(x^\circ) = 0, \; i = 1, \ldots, m$

Prova

$\rightarrow \quad L(x^o, u^o) \leq L(x, u^o) \quad \forall x \in S \;\Rightarrow$ (a)

$$L(x^o, u^o) \geq L(x^o, u) \quad \forall u \geq 0$$

$$f(x^\circ) + \sum_{i=1}^{m} u_i^\circ g_i(x^\circ) \geq f(x^\circ) + \sum_{i=1}^{m} u_i g_i(x^\circ), \; u_i \geq 0$$

$$\sum_i (u_i - u_i^\circ) g_i(x^\circ) \leq 0, \; \forall u_i \geq 0 \qquad (11)$$

Se (b) ($g_i(x) \le 0$, $i = 1,...,m$) é violada, então u_i pode ser escolhido suficientemente grande tal que (11) seja violada. Então (b) tem que ser satisfeita.

Se todos $u_i = 0$, então (11) $\Rightarrow \sum_i u_i^\circ g_i(x^\circ) \ge 0$

mas $u_i^\circ \ge 0 \; e \; g_i(x) \le 0 \Rightarrow \sum_i u_i^\circ g_i(x^\circ) \le 0$

logo $u_i^\circ g_i(x^\circ) = 0, \; i = 1, \ldots, m$

$\leftarrow$ (a) x^o minimiza $L(x, u^o)$ e (c) $u^o_i\, g_i(x) = 0$, $i = 1,...,m \Rightarrow L(x^o, u^o) = f(x^o)$

por definição $L(x^\circ, u) = f(x^\circ) + \sum_{i=1}^{m} u_i g_i(x^\circ)$

como $g_i(x) \le 0$, o termo $u_i\, g_i(x)$ é não positivo para todo $u_i \ge 0$ e

$$L(x^\circ, u) \le f(x^\circ) = L(x^\circ, u^\circ), \; \forall u \ge 0$$

Teorema 10: (suficiência do ponto de sela)

Se (x^o, u^o) é um ponto de sela de $L(x, u)$, então x^o é solução do primal

Prova

como (x^o, u^o) é um ponto de sela, então as condições (a), (b) e (c), teorema 9 são satisfeitas. Fazendo $g(x) = (g_1(x),....,g_m(x))'$ a condição (a) torna-se

$$f(x^o) + u^o g(x^o) \leq f(x) + u^o g(x), \quad \forall x \in S \tag{18}$$

Condição (c) $\Rightarrow u^o g(x^o) = 0$ e (18) torna-se

$$f(x^o) \leq f(x) + u^o g(x), \quad \forall x \in S \text{ que satisfaz } g(x) \leq 0 \tag{19}$$

e $\forall x$ primal factível, o termo $u^o g(x)$ é não positivo; isto é

$$f(x^o) \leq f(x), \ \forall x \in S, \ \ g(x) \leq 0 \Rightarrow x^o \text{ é solução do primal}$$

- Comentários sobre suficiência do ponto de sela

  - Teorema 10 se aplica a qualquer modelo de otimização, incluindo
    - S é um conjunto finito
    - f e g não convexas

  - Ponto de sela pode não existir
    - existência garantida somente para modelos convexos
    - mesmo assim minimização de $L(x, u)$ pode ser atrativa

- Exemplo da não existência ponto de sela

$$min\{-x^2 \,|\, 1-2x=0,\ 0\leq x\leq 1\}$$

Fazendo $S=\{0\leq x\leq 1\}$ o problema de minizar o Lagrangeano é

$$\min_{x\in S}\{L(x,u)=-x^2+u(1-2x)\}$$

$$\nexists\ u \mid \ min\, L(x,u) \text{ ocorre em } x^{\circ}=\frac{1}{2}$$

5-Métodos de programação não linear

- Categorias principais
  - métodos das direções factíveis
  - métodos das funções de penalização

Métodos das direções factíveis

– Princípio dos métodos das direções factíveis

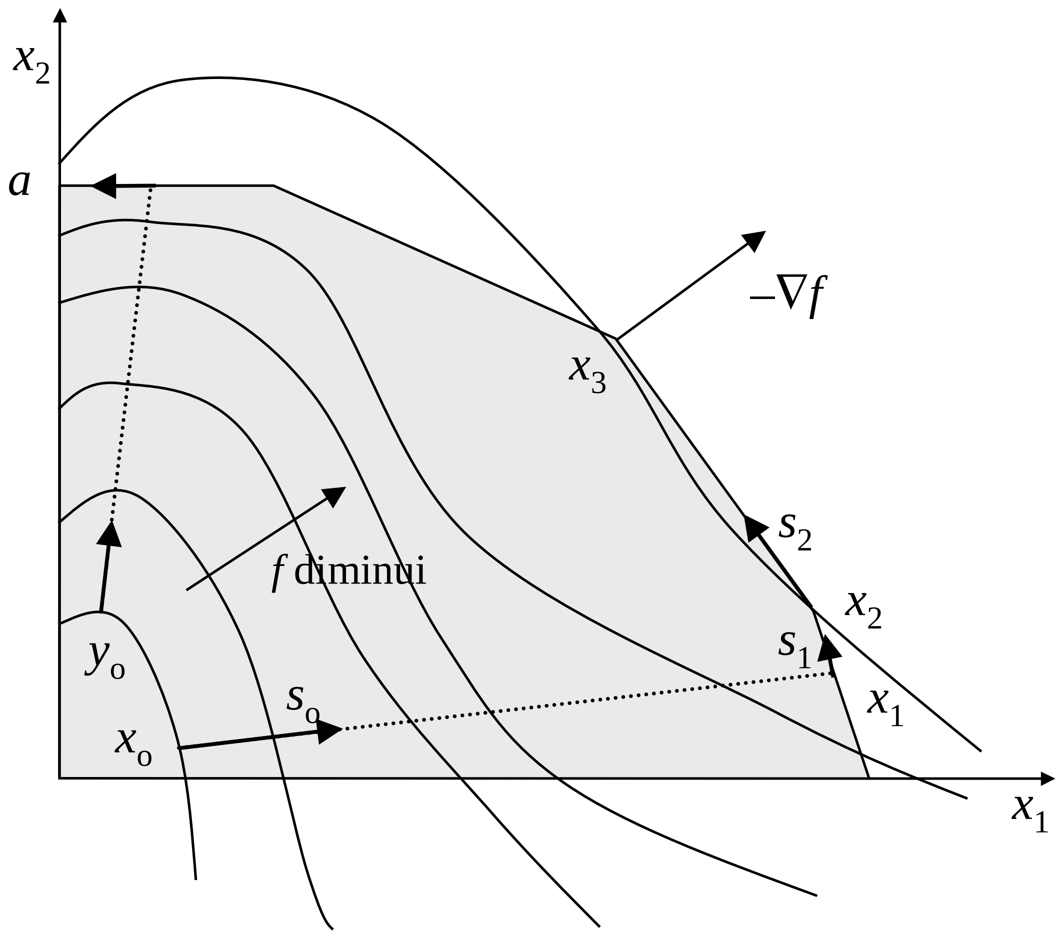
1. escolher uma solução inicial factível

2. determinar uma direção que

   2.1 pequeno movimento nesta direção não viola restrições (factível)

   2.2 valor da função objetivo melhora nesta direção (usável)

- Métodos das direções factíveis



- Algoritmo de Zoutendijk

$$\begin{aligned} &min \quad f(x) \\ &sa \quad g_i(x) \geq 0, \quad i = 1, \dots, m \end{aligned}$$

$R = \{ x \mid g_i(x) \geq 0,\ i = 1,...,m \}$ conjunto restrição

x_o : solução inicial factível

escolher vetor *s* cuja direção seja simultaneamente factível e usável

supor $g_i(x_o) = 0,\ i \in \mathrm{I}$

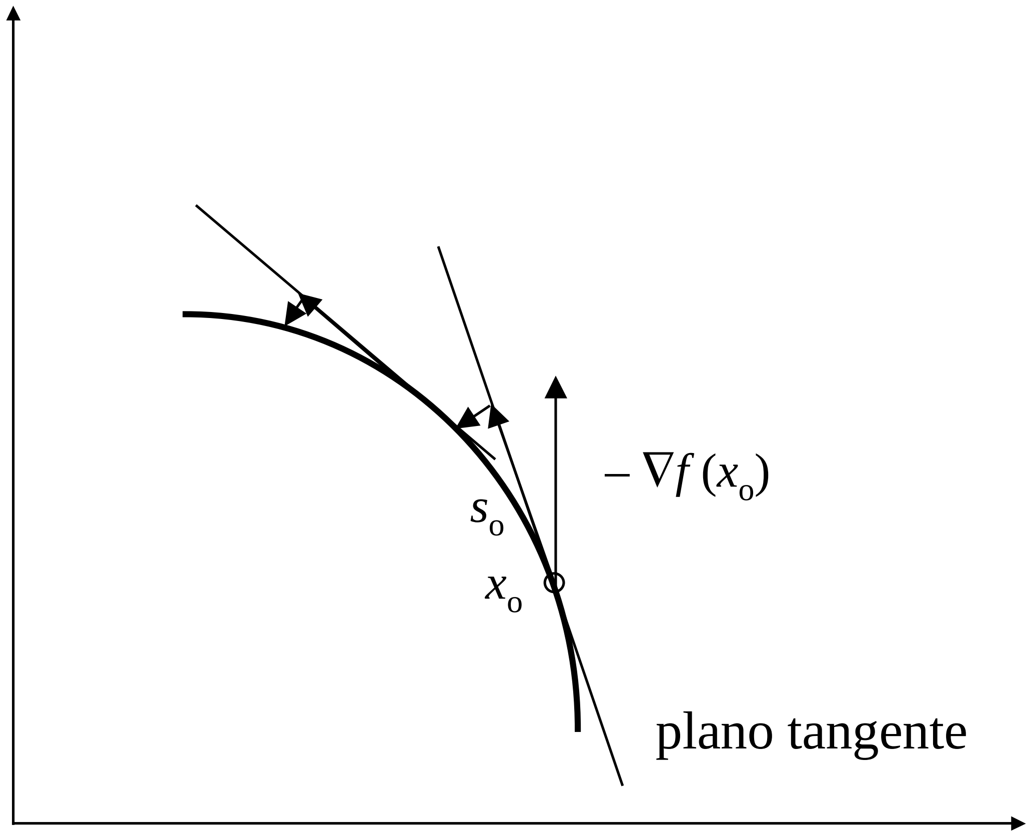
- Direção factível

$$\frac{d}{d\alpha} g_i(x_o + \alpha s)\Big|_{\alpha=0} = \nabla g_i'(x_o)s \geq 0$$

- Direção usável

$$\frac{d}{d\alpha} f(x_o + \alpha s)\Big|_{\alpha=0} = \nabla f_i'(x_o)s < 0$$

- Melhor direção: vetor factível que *min* $\nabla f\,'(x_o)\ s$
- Projeção de – $\nabla f\,(x_o)$ no plano tangente em x_o



- Melhor direção (s, ξ) deve ser tal que
  - diminui função objetivo
  - se afasta da fronteira do conjunto restrição

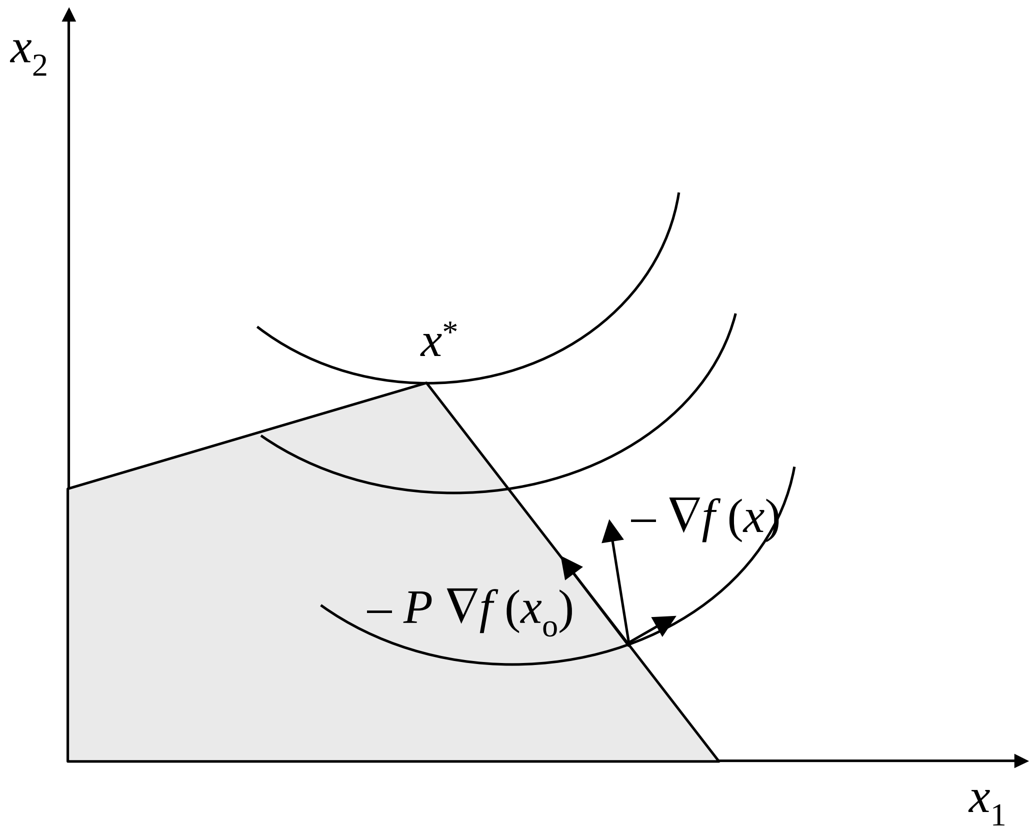
$$
\begin{aligned}
&min \;\; \xi \\
&sa \quad \nabla' g_i(x)s + \theta_i \xi \geq 0, \quad i \in I, \; 0 \leq \theta_i \leq 1 \\
&\qquad \nabla f(x_o)s - \xi \leq 0 \\
&\qquad s's = 1
\end{aligned}
$$

- Problema quase linear
  - pode ser resolvido eficientemente
  - se min $\xi > 0$ então s não factível; pára

- Determinação do passo

$$
\begin{aligned}
&min \quad f(x_k + \alpha s) \\
&sa \quad x_k + \alpha s \in R
\end{aligned}
$$

- Algoritmo do gradiente projetado de Rosen



$$P = I - A(A'A)^{-1}A'$$

Restrições lineares

Resumo algoritmo do gradiente projetado

1. calcular $\nabla f\,(x_k)$

2. associar à x_k as restrições ativas (planos que passam) em x_k

3. projetar $-\nabla f\,(x_k)$ na intersecção dos planos que passam em x_k
   se x_k é um ponto interior, então a projeção é o própio $-\nabla f\,(x_k)$

4. se a projeção não é nula, então minimizar ao longo da projeção,
   sujeitas às restrições; fazer $k = k + 1$ e ir para o passo 1

5. Se a projeção é nula, então $\nabla f\,(x_k)$ pode ser escrito como $\nabla f\,(x_k) = \sum_i u_i a_i$

   5.1 se $u_i \geq 0\ \forall i$, então é solução ótima (satisfaz KKT); senão

   5.2 definir novo conjunto de planos associado à x_k, removendo do conjunto
   corrente de planos aqueles com $u_i < 0$; ir para passo 3

- Métodos das funções de penalização

$$\begin{aligned} &min \quad f(x) \\ &sa \quad g_i(x) \geq 0, \quad i = 1, \ldots, m \end{aligned}$$

$$\phi(x) = \begin{cases} 0 & y \geq 0 \\ \infty & y < 0 \end{cases}$$

$$min \quad f(x) + \sum_{i=1}^{m} \phi(g_i(x))$$

Algoritmo de Fiacco-McCormick

$$P(x,r) = f(x) + r\sum_{i=1}^{m}\frac{1}{g_i(x)} \quad , \qquad r > 0$$

1. escolher $r_1 > 0$ e x_o interior ao conjunto restrição

2. *min* $P(x, r_1)$ iniciando em x_o existe $x(r_1)$ dentro conjunto restrição

3. $r_1 > r_2 > ... > r_k > 0$, cada solução $x(r_1)$ será um ponto interior

4. efeito de r : reduzir influência do terno adicional

- Modelo dual de Wolfe

  - Primal

$$\begin{aligned} &min \quad f(x) \\ &sa \quad g_i(x) \ge 0, \quad i = 1, \dots, m \end{aligned}$$

  - Dual

$$\begin{aligned} &max \quad L(x,u) = f(x) - \sum_{i=1}^{m} u_i g_i(x) \\ &sa \quad \nabla f(x) = \sum_{i=1}^{m} u_i \nabla g_i(x) \\ &\qquad u_i \ge 0,\ i = 1, \dots, m \end{aligned}$$

- Propriedades do modelo dual de Wolfe

1. se primal tem solução x, então $\exists u$ tal que (x, u) resolve o dual

2. valores das funções objetivos primal e dual são iguais

3. para cada x primal factível e (x, u) dual factível $f(x) \geq L(x, u)$

4. $f(x)$ limitante superior e $L(x, u)$ limitante inferior

$$P(x,r) = f(x) + r\sum_{i=1}^{m}\frac{1}{g_i(x)} \quad , \qquad r > 0$$

$$\nabla P(x(r_k), r_k) = \nabla f(x(r_k)) + \sum_{i=1}^{m}\frac{r_k}{g_i^2(x(r_k))}\nabla g_i(x(r_k)) = 0$$

$$u_i(r_k) = \frac{r_k}{g_i^2(x(r_k))} > 0$$

Se

1. se interior conjunto restrição não vazio

2. funções f e g_i são $2\times$continuamente diferenciáveis

3. conjunto ponto no conjunto restrição tal que $f(x) \leq k$ é limitado $\forall k < \infty$

4. $f(x)$ limitada inferiormente para x no conjunto restrição

5. $f(x)$ convexa

6. g_i côncavas

7. $P(x, r)$ estritamente convexa no interior do conjunto restrição $\forall r > 0$

então

$$L(x,u) = f(x(r_k)) - r_k \sum_{i=1}^{m} \frac{1}{g_i^2(x(r_k))} \leq v_\circ \leq f(x(r_k))$$

v_o: P-mínimo

$\{(x(r_k), u(r_k))\}$ converge para a solução dual (x, u)

Observação

Este material refere-se às notas de aula do curso IA 810 Otimização de Sistemas de Grande da Faculdade de Engenharia Elétrica e de Computação da Unicamp. Não substitui o livro texto, as referências recomendadas e nem as aulas expositivas. Este material não pode ser reproduzido sem autorização prévia dos autores. Quando autorizado, seu uso é exclusivo para atividades de ensino e pesquisa em instituições sem fins lucrativos.