Método do Lagrangiano aumentado (LA)

Esquema de penalidade externa.
↳ instável numericamente.
↳ ρ_k cresce muito...

Método de LA pode ser visto como uma melhoria "prática" do esquema de penalidade.
Duas estratégias são adotadas para evitar ρ_k crescer muito:

1) Penalizar $h(x) + a$ e $(g(x) + b)_+$ ao invés de $h(x)$ e $g(x)_+$.
2) Só aumentar ρ_k se a inviabilidade não diminuir.

Obs.: não vamos pedir que x^k seja minimizador global do subproblema, apenas que seja ponto estacionário (anule o ∇ F.O.)

Penalização com deslocamento

Penalidade externa "pura":

$SP(\rho_k)$: $\min f(x) + \frac{\rho_k}{2}\left[\|h(x)\|^2 + \|g(x)_+\|^2\right]$

(subproblema irrestrito)

Resolver $SP(\rho_k)$ significa calcular x^k tq.

$$\nabla f(x^k) + \sum_{i=1}^{m} [\rho_k h_i(x^k)] \nabla h_i(x^k) + \sum_{j=1}^{p} [\rho_k g_j(x^k)_+] \nabla g_j(x^k) = 0.$$

As expressões $\rho_k h_i(x^k)$ e $\rho_k g_j(x^k)_+$ são aproximações dos multiplicadores de Lagrange.

Vamos considerar as estimativas deslocadas $\rho_k h_i(x^k) + \bar{\lambda}_i^k$ e $(\rho_k g_j(x^k) + \bar{\mu}_j^k)_+$, onde $\bar{\lambda}^k$ e $\bar{\mu}^k$ estarão limitados (dentro de um compacto).

Veja que, considerando a função

$$L_\rho(x,\bar{\lambda},\bar{\mu}) = f(x) + \frac{\rho}{2}\left[\left\| h(x) + \frac{\bar{\lambda}}{\rho} \right\|^2 + \left\| \left(g(x) + \frac{\bar{\mu}}{\rho} \right)_+ \right\|^2 \right],$$

derivando obtemos

$$\nabla_x L_\rho(x,\bar{\lambda},\bar{\mu}) = \nabla f(x) + \sum_{i=1}^{m} (\rho h_i(x) + \bar{\lambda}_i) \nabla h_i(x) + \sum_{j=1}^{p} (\rho g_j(x) + \bar{\mu}_j)_+ \nabla g_j(x).$$

Veja que pontos x^k estacionários de $L_{\rho_k}(x,\bar{\lambda}^k,\bar{\mu}^k)$ são associados aos multiplicadores com deslocamento.

O subproblema do método de Lagrangiano aumentado é

$$\min_x L_{\rho_k}(x,\bar{\lambda}^k,\bar{\mu}^k).$$

Obs.: se $\bar{\lambda}^k = 0$ e $\bar{\mu}^k = 0$, então o método se reduz à penalidade externa "pura".

Obs: a função $L_\rho(x, \bar{\lambda}, \bar{\mu})$ é chamada Lagrangiano aumentado, daí vem o nome do método!

Obs: a função Lagrangiano é

$$L(x, \lambda, \mu) = f(x) + \sum_{i=1}^{m} \lambda_i h_i(x) + \sum_{j=1}^{p} \mu_j g_j(x)$$

(Em KKT, temos $\nabla_x L(x, \lambda, \mu) = 0$)

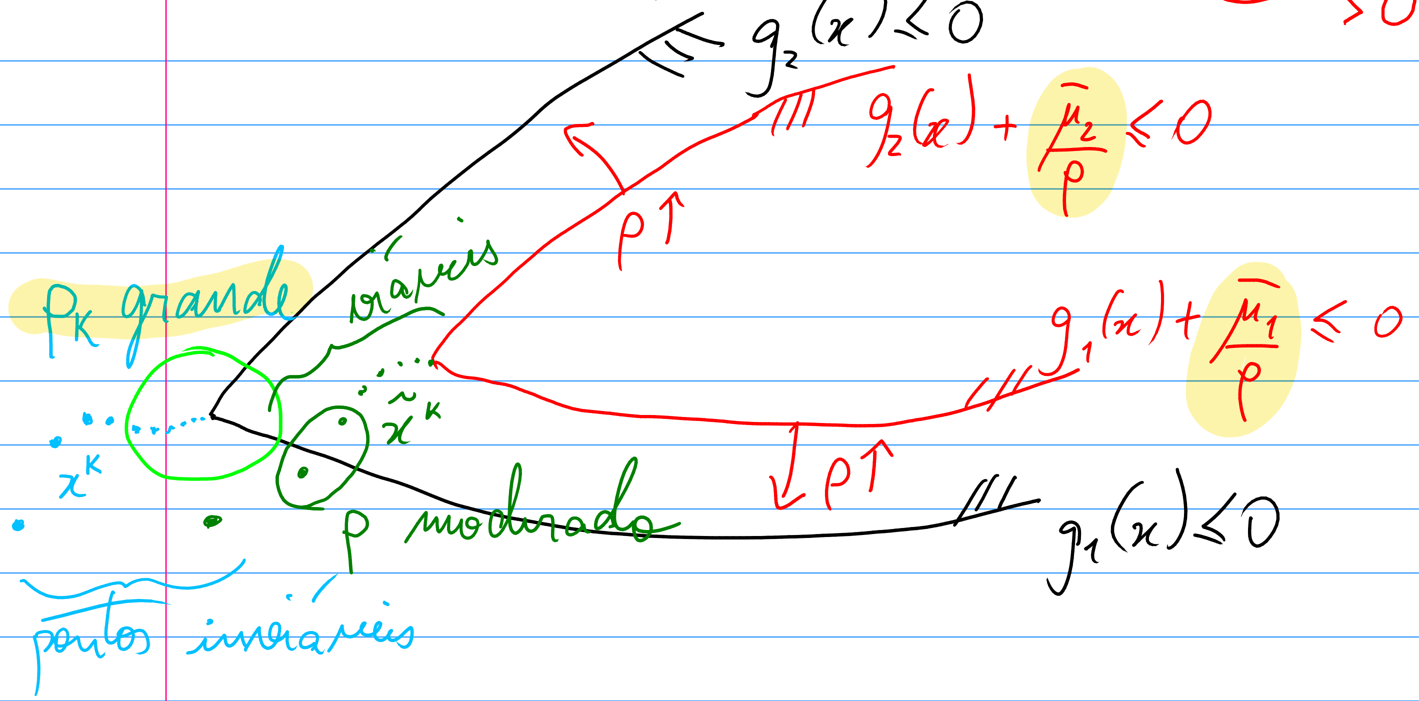
Veja que

$$\nabla_x L(x, \lambda, \mu) = \nabla f(x) + \sum_{i=1}^{m} \lambda_i \nabla h_i(x) + \sum_{j=1}^{p} \mu_j \nabla g_j(x)$$

$$= \nabla_x L_0(x, \lambda, \mu)$$

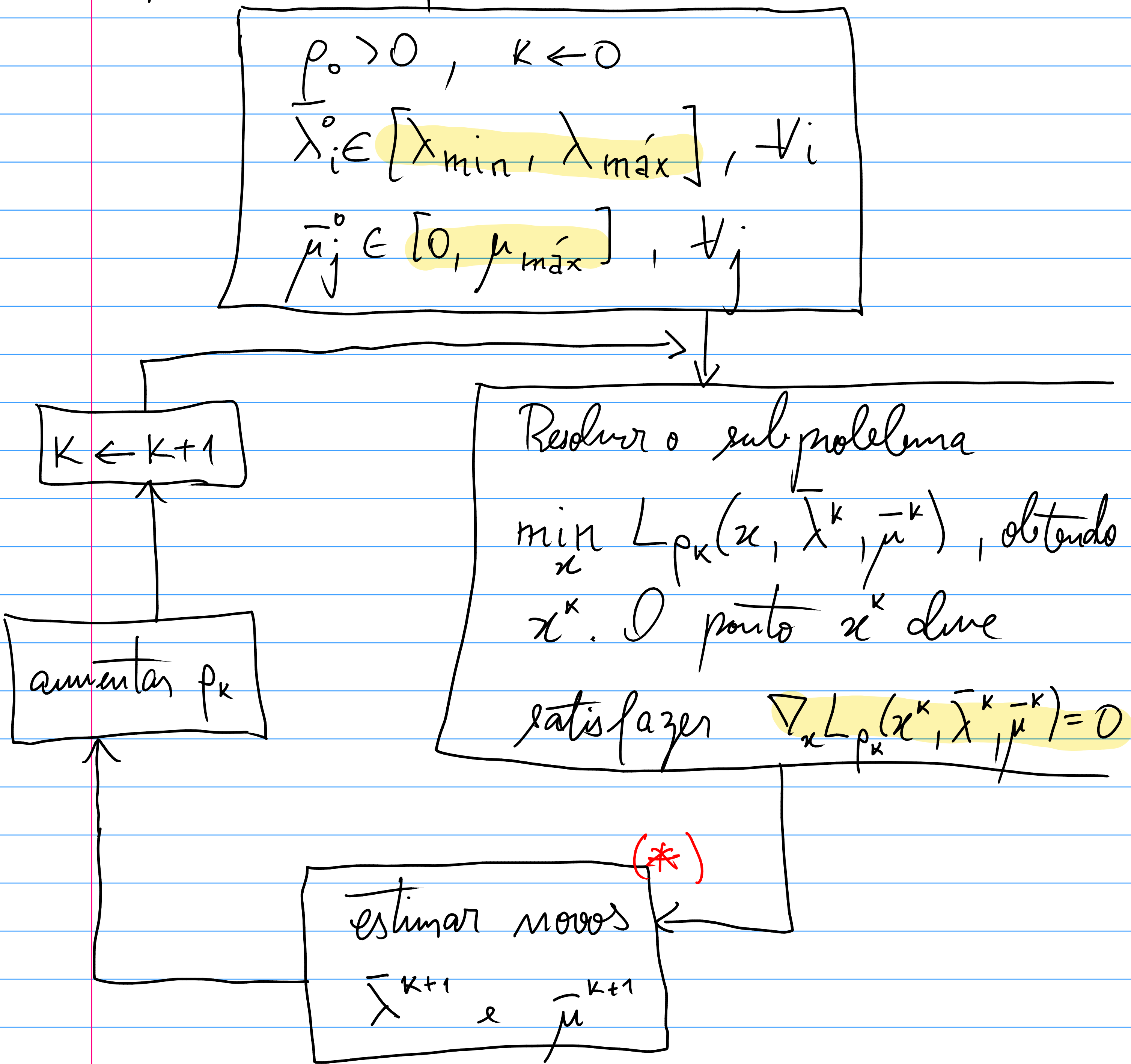
Interpretação geométrica do deslocamento

$$h(x) + \frac{\bar{\lambda}}{\rho} \quad , \quad \underbrace{g(x) + \frac{\bar{\mu}}{\rho}}_{>0}$$



- Penalidade "pura" (sem deslocamento)
- com deslocamento

Esquema de penalidade com deslocamento



Quem são $\bar{\lambda}^{k+1}$ e $\bar{\mu}^{k+1}$ adequados?

Veja que, ao resolver o subproblema, temos

$$\nabla_x L_{\rho_k}(x^k, \bar{\lambda}^k, \bar{\mu}^k) = \nabla f(x^k) + \sum_{i=1}^{m} \left(\rho_k h_i(x^k) + \bar{\lambda}_i^k\right) \nabla h_i(x^k) + \sum_{j=1}^{p} \left(\rho_k g_j(x^k) + \bar{\mu}_j^k\right)_+ \nabla g_j(x^k) = 0$$

Desta expressão, é natural considerar

$$\tilde{\lambda}_i^{k+1} = \rho_k h_i(x^k) + \bar{\lambda}_i^k \quad e \quad \tilde{\mu}_j^{k+1} = (\rho_k g_j(x^k) + \bar{\mu}_j^k)_+ \geq 0$$

Os novos $\bar{\lambda}_i^{k+1}$ e $\bar{\mu}_j^{k+1}$ serão as projeções de $\tilde{\lambda}_i^{k+1}$ e $\tilde{\mu}_j^{k+1}$ nos intervalos

$$[\lambda_{min}, \lambda_{máx}] \quad e \quad [0, \mu_{máx}].$$

Assim,

$$\rightarrow \bar{\lambda}_i^{k+1} = \text{máx}\{\lambda_{min}, \text{min}\{\lambda_{máx}, \rho_k h_i(x^k) + \bar{\lambda}_i^k\}\}.$$

Exercício:

(i) verifique que $\bar{\lambda}_i^{k+1} \in [\lambda_{min}, \lambda_{max}]$, e logo $\{\bar{\lambda}_i^{k+1}\}$ é limitada.

(ii) verifique que, se $\rho_k h_i(x^k) + \bar{\lambda}_i^k \in [\lambda_{min}, \lambda_{max}]$ então $\bar{\lambda}_i^{k+1} = \rho_k h_i(x^k) + \bar{\lambda}_i^k$.

Analogamente,

$$\rightarrow \bar{\mu}_j^{k+1} = \text{máx}\{0, \text{min}\{\mu_{máx}, \rho_k g_j(x^k) + \bar{\mu}_j^k\}\}$$

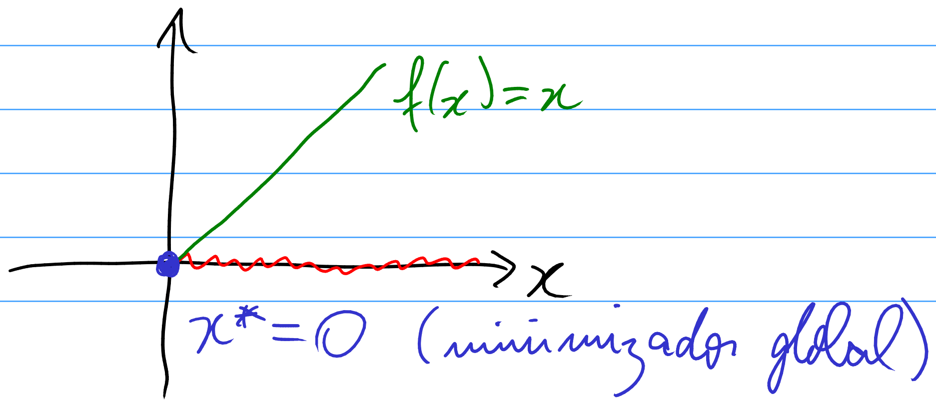
Essas expressões são implementadas no pacote ALGENCAN.

Obs: À medida que $\lambda_{\min}$, $\lambda_{\text{máx}}$ e $\mu_{\text{máx}}$ (parâmetros do método) são escolhidos ≈ 0, o método se assemelha ao esquema de penalidade sem deslocamento. Na prática, escolhemos esses parâmetros grandes. Por exemplo, em ALGENCAN, o padrão é $\lambda_{\min} = -10^{20}$,
$\lambda_{\text{máx}} = \mu_{\text{máx}} = 10^{20}$.

Exemplo: $P: \min\limits_{x} \; x$

$\quad$ s.a. $x \geq 0$. $(-x \leq 0)$

$g(x) \leq 0$



Aplicando o esquema de penalidade externa para:

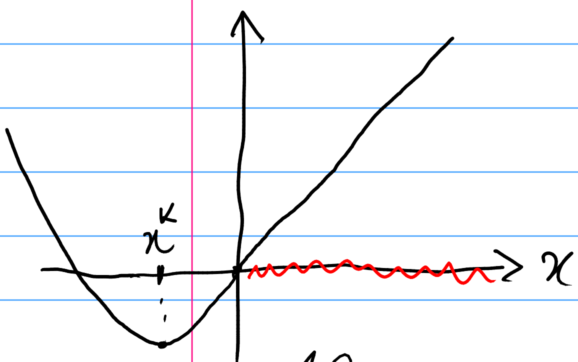
$$SP(\rho_k): \min_{x} \underbrace{x + \frac{\rho_k}{2} \max\{0, -x\}^2}_{f(x) + \frac{\rho_k}{2} \max\{0,\, g(x)\}^2}$$

Suponha que $x \leq 0$. Então

$$\max\{0, -x\} = -x \quad (\geq 0)$$

Assim, neste caso,

$$SP(\rho_k): \min_{x} x + \frac{\rho_k}{2} x^2 .$$



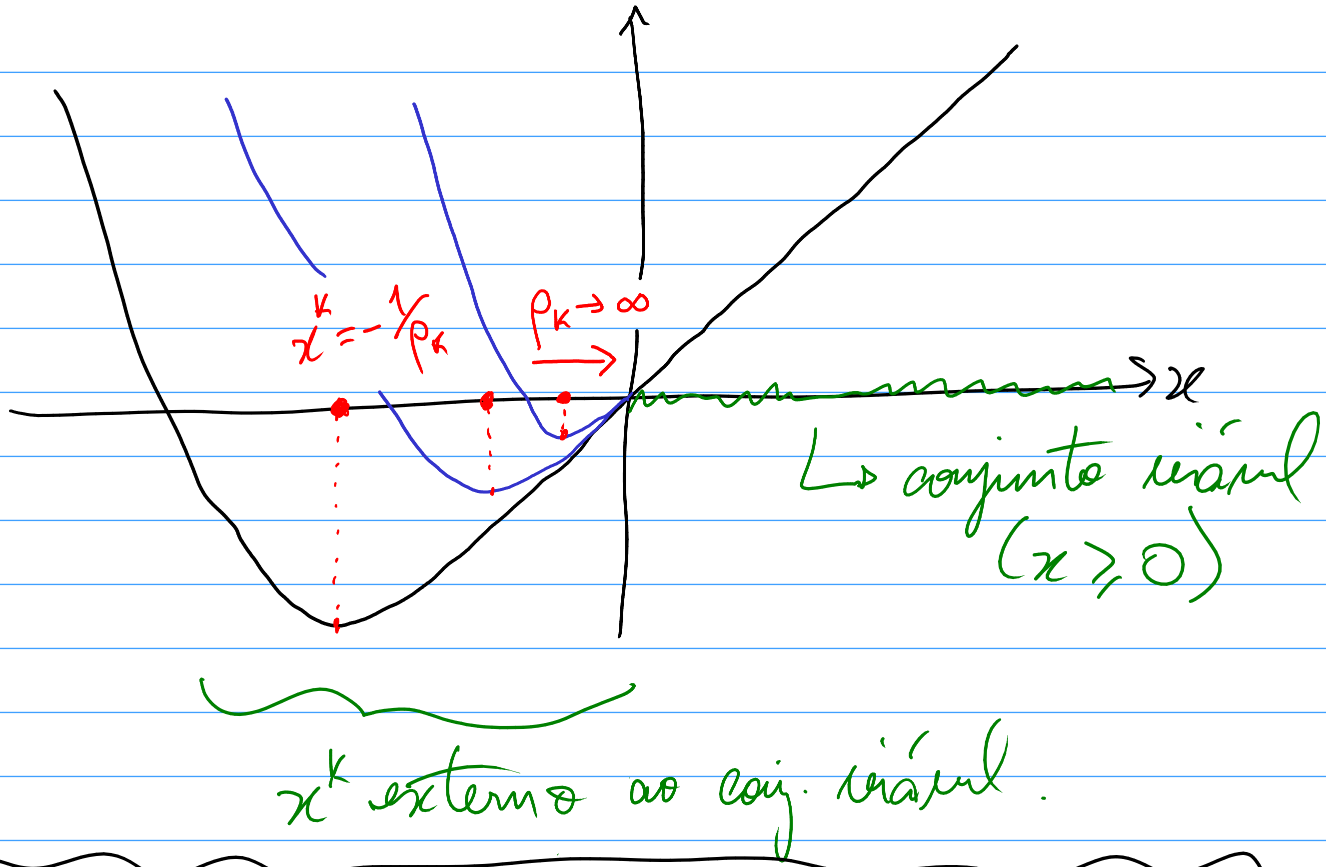
Resolvendo:

$$1 + \rho_k x^k = 0 \iff \boxed{x^k = -1/\rho_k}$$

Veja que x^k de fato é < 0. Assim, x^k é a solução de $SP(\rho_k)$ no caso geral.

Observe que $x^k \to 0$. (solução de P).

Observe também $x^k = -1/\rho_k$ nunca é igual à $x^* = 0$ (ρ_k deve crescer para aproximar x^*).



Aplicando penalidade com deslocamento:

$$SP(\rho_k): \min_x \; x + \frac{\rho_k}{2} \max\left\{0, -x + \frac{\bar{\mu}^k}{\rho_k}\right\}^2$$

Dados iniciais: $\bar{\mu}^0 = 0$, $\rho_0 = 1$

Iteração 0: ($k=0$)

$$SP(\rho_k): \min_x \; x + \frac{1}{2} \max\{0, -x\}^2$$

(esse é o mesmo problema da penalidade pura).

Assim, $x^0 = -\frac{1}{\rho_0} = -1$.

Paramos com x^0? NÃO, pois ñ é KKT.

Iteração 1: $(k=1)$

$$\tilde{\mu}^1 = \rho_0 g(x^0) + \bar{\mu}^0 = 1.(-(-1)) + 0 = 1.$$

Imagine que $\mu_{max} > 1$ (p. ex., 10^{20}).

$$\bar{\mu}^1 = \text{máx}\{0, \min\{10^{20}, \tilde{\mu}^1\}\}$$

$$= \text{máx}\{0, \underbrace{\min\{10^{20}, 1}\}\}$$

$$= \text{máx}\{0, 1\}$$

$$\Rightarrow \boxed{\bar{\mu}^1 = 1}.$$

Atualizamos também $\rho_1 = 10\rho_0 = 10$.

Subproblema $(k=1)$

$$SP(\rho_1): \min_x \; x + \frac{10}{2}\,\text{máx}\{0, \underbrace{-x + 1/10}\}^2$$

Suponha que $-x + 1/10 > 0$. Assim,

$$SP(\rho_1): \min_x \; x + 5\left(-x + 1/10\right)^2$$

Resolvendo:

$$1 + 10\left(-x^1 + 1/10\right)(-1) = 0$$

$$\Rightarrow \cancel{1} + 10x^1 - \cancel{1} = 0 \quad \Rightarrow x^1 = 0.$$

Verificando se paramos, ou seja, se x^1 é KKT:

$$\nabla f(x^1) + \bar{\mu}^1 \nabla g_1(x^1) = 0 \qquad g(x) = -x$$

$$\bar{\mu}^1 \geq 0$$

$$\bar{\mu}^1 g(x^1) = 0$$

$$1 + 1(-1) = 0 \quad \checkmark$$

$$\bar{\mu}^1 = 1 \geq 0 \quad \checkmark$$

$$\bar{\mu}^1 \cdot g(x^1) = 1(-0) = 0 \quad \checkmark$$

Assim, x^1 é ponto KKT do problema original P, e logo o método para com 2 iterações!

Em particular, o uso do deslocamento impediu ρ_k crescer.

Exemplo executado no pacote ALGENCAN

• Com deslocamento ($\mu_{max} = 10^{20}$)

```
out penalt  objective infeas infeas norm   norm   |Grad|  inner Newton
ite         function  ibilty +compl graLag point  infeas  totit forKKT
  0         1.000D+00 0.D+00 0.D+00 1.D+00 1.D+00 0.D+00     0    0    0
  1 1.D+01 -1.000D-01 1.D-01 1.D-01 0.D+00 1.D-01 1.D-01    4C    0    0
  2 1.D+01  0.000D+00 0.D+00 0.D+00 0.D+00 0.D+00 0.D+00    5C    0    0

 Flag of ALGENCAN: Solution was found.
```

← sucesso!

2 iterações → ρ_k pequeno

ρ_k — $f(x^k)$ — $\phi(x^k)$ — $\min\{\mu^k, -g(x^k)\}$ (viabilidade + complementaridade) — $\|\nabla L\|$

• Sem deslocamento ($\mu_{max} = 0$)

↳ = penalidade externa pura (precisão parada $= 10^{-8}$)

```
out penalt  objective infeas infeas norm   norm   |Grad|  inner Newton
ite         function  ibilty +compl graLag point  infeas  totit forKKT
  0         1.000D+00 0.D+00 0.D+00 1.D+00 1.D+00 0.D+00     0    0    0
  1 1.D+01 -1.000D-01 1.D-01 1.D-01 0.D+00 1.D-01 1.D-01    4C    0    0
  2 1.D+01 -1.000D-01 1.D-01 1.D-01 0.D+00 1.D-01 1.D-01    4C    0    0
  3 1.D+02 -1.000D-02 1.D-02 1.D-02 4.D-16 1.D-02 1.D-02    5C    0    0
  4 1.D+02 -1.000D-02 1.D-02 1.D-02 4.D-16 1.D-02 1.D-02    5C    0    0
  5 1.D+03 -1.000D-03 1.D-03 1.D-03 9.D-16 1.D-03 1.D-03    6C    0    0
  6 1.D+03 -1.000D-03 1.D-03 1.D-03 9.D-16 1.D-03 1.D-03    6C    0    0
  7 1.D+04 -1.000D-04 1.D-04 1.D-04 4.D-16 1.D-04 1.D-04    7C    0    0
  8 1.D+04 -1.000D-04 1.D-04 1.D-04 4.D-16 1.D-04 1.D-04    7C    0    0
  9 1.D+05 -1.000D-05 1.D-05 1.D-05 1.D-16 1.D-05 1.D-05    8C    0    0

out penalt  objective infeas infeas norm   norm   |Grad|  inner Newton
ite         function  ibilty +compl graLag point  infeas  totit forKKT
 10 1.D+05 -1.000D-05 1.D-05 1.D-05 1.D-16 1.D-05 1.D-05    8C    0    0
 11 1.D+06 -1.000D-06 1.D-06 1.D-06 1.D-15 1.D-06 1.D-06    9C    0    0
 12 1.D+06 -1.000D-06 1.D-06 1.D-06 1.D-15 1.D-06 1.D-06    9C    0    0
 13 1.D+07 -1.000D-07 1.D-07 1.D-07 0.D+00 1.D-07 1.D-07   10C    0    0
 14 1.D+07 -1.000D-07 1.D-07 1.D-07 0.D+00 1.D-07 1.D-07   10C    0    0
 15 1.D+08 -1.000D-08 1.D-08 1.D-08 3.D-16 1.D-08 1.D-08   11C    0    0

 Flag of ALGENCAN: Solution was found.
```

$\rightarrow x^{15} = -\frac{1}{10^8} < 0$

ρ_k aumenta muito!

Controle de viabilidade / admissibilidade

Queremos evitar crescer ρ... Vimos a estratégia de penalizar restrições com deslocamento. Uma segunda estratégia é aumentar ρ apenas se a inviabilidade não melhora de uma iteração para outra.

$\phi(x) = \|h(x)\|_2^2 + \|g(x)_+\|_2^2 \rightarrow$ poderíamos usar ϕ no teste, mas não é a forma adotada por ALGENCAN. Usamos a norma $\|\cdot\|_\infty$:

$$\|z\|_\infty = \max\{|z_1|, \dots, |z_n|\}$$

- Viabilidade do iterando x^k:

$$\max\{\|h(x^k)\|_\infty, \|g(x^k)_+\|_\infty\}.$$

- Complementaridade: $\mu_j^k g_j(x^k) \approx 0$?

Veja que $\underbrace{\mu_j g_j(x) = 0}_{\text{complementaridade}}$ $+$ $\underbrace{g(x) \leq 0}_{\text{viabilidade } g(x) \leq 0}$ $(\mu \geq 0)$

$$\Leftrightarrow \boxed{V_j = \min\{\mu_j, -g_j(x)\} = 0}$$

(mínimo coord. a coord.)

De fato, suponha que $\mu_j g_j(x) = 0$ e $g(x) \leq 0$. Se fosse $V_j > 0$ então $\mu_j > 0$ e $-g_j(x) > 0$. Assim, $\mu_j g_j(x) < 0$, uma contradição. Se fosse $V_j < 0$ então, em particular, $-g_j(x) < 0 \Rightarrow g_j(x) > 0$, uma contradição.

Reciprocamente, suponha $V_j = 0$. Assim, $(\mu_j = 0$ e $-g_j(x) \geq 0)$ ou $(\mu_j > 0$ e $-g_j(x) = 0)$. No primeiro caso, $\mu_j g_j(x) = 0$ e $-g_j(x) \geq 0 \Rightarrow g_j(x) \leq 0$. No segundo caso $\mu_j g_j(x) = \mu_j(-g_j(x)) = 0$ e $-g_j(x) = 0 \Rightarrow g_j(x) \leq 0$.

Desta forma $V_j = 0$ engloba a complementaridade e a viabilidade das desigualdades $g(x) \leq 0$. Daí, o teste do acréscimo de ρ fica:

(**) Se $K=0$ (primeira iteração) ou

$\max\{\|h(x^K)\|_\infty, \|V^K\|_\infty\}$ viabilidade $g(x)\le 0$ + complement.

viabilid. $h(x)=0$ $\le \sigma \max\{\|h(x^{K-1})\|_\infty, \|V^{K-1}\|_\infty\}$

então $\rho_{K+1}=\rho_K$ (não aumentamos ρ). Caso contrário, $\rho_{K+1}=10\,\rho_K$.

- $\sigma\in[0,1)$: parâmetro.
- $V^K=\min\{\bar{\mu}^K, -g(x^K)\}$.

Método de Lagrangiano aumentado (ALGENCAN)

$\rho_0 > 0$, $\tau \in [0,1)$, $x^0 \in \mathbb{R}^n$,
$\bar{\lambda}_i^0 \in [\lambda_{min}, \lambda_{máx}]$, $\forall i$
$\bar{\mu}_j^0 \in [0, \mu_{máx}]$, $\forall j$; $\{\varepsilon_k\} \downarrow 0$

Parar?

sim → PARE

NÃO

Resolver o subproblema

$$\min_x f(x) + \frac{\rho_k}{2}\left[\left\|h(x) + \frac{\bar{\lambda}^k}{\rho_k}\right\|^2 + \left\|\left(g(x) + \frac{\bar{\mu}}{\rho_k}\right)_+\right\|^2\right]$$

obtendo o iterando x^k. O ponto x^k deve satisfazer $\|\nabla_x L_{\rho_k}(x^k, \bar{\lambda}^k, \bar{\mu}^k)\|_\infty \leq \varepsilon_k$

Atualizar multiplicadores projetados:

$$\bar{\lambda}^{k+1} = \text{máx}\{\lambda_{min}, \min\{\lambda_{máx}, \rho_k h(x^k) + \bar{\lambda}^k\}\}$$

$$\bar{\mu}^{k+1} = \text{máx}\{0, \min\{\mu_{máx}, \rho_k g(x^k) + \bar{\mu}^k\}\}$$

Atualize ρ segundo o critério (**)

$k \leftarrow k+1$

<u>Critério de parada</u>: paramos quando x^k é $\tilde{\imath}$ KKT do problema original

$$P: \min_x f(x) \text{ s.a. } h(x)=0,\ g(x)\leq 0.$$

<u>Parâmetro</u>: $\varepsilon_{opt} > 0$: precisão de parada.

$$\left\| \nabla f(x^k) + \sum_{i=1}^{m} \lambda_i^k \nabla h_i(x^k) + \sum_{j=1}^{p} \mu_j^k \nabla g_j(x^k) \right\| \leq \varepsilon_{opt}$$

$$\mu_j^k \geq 0,\ \forall j.$$

$$\left| \min\{\mu_j^k, -g_j(x^k)\} \right| \leq \varepsilon_{opt},\ \forall j$$

$$\|h(x^k)\|_\infty \leq \varepsilon_{opt},\quad \|g(x^k)_+\|_\infty \leq \varepsilon_{opt}$$

- Padrão: $\varepsilon_{opt} = 10^{-8}$.
- $\lambda^k = \rho_k h(x^k) + \bar{\lambda}^k$, $\quad \mu^k = (\rho_k g(x^k) + \bar{\mu}^k)_+$.

```
Min (x[1] - 2.0) ^ 2.0 + (x[2] - 1.0) ^ 2.0
Subject to
 x[1] + x[2] ≤ 2.0
 (x[1] ^ 2.0 - x[2]) - 0.0 ≤ 0
```

$\max\{\|h(x^k)\|_\infty, \|V^k\|_\infty\}$

out ite	penalt	objective function	infeas ibilty	scaled obj-funct	scaled infeas	infeas +compl	norm graLag	|Grad| infeas	inner totit	Newton forKKT	
0		5.000D+00	0.D+00	1.250D+00	0.D+00	0.D+00	1.D+00	0.D+00	0	0	0
1	1.D+01	9.827D-01	1.D-02	2.457D-01	1.D-02	1.D-02	3.D-09	4.D-02	8C	0	0
2	1.D+01	9.996D-01	4.D-04	2.499D-01	4.D-04	4.D-04	3.D-07	7.D-04	10C	0	0
3	1.D+01	1.000D+00	1.D-05	2.500D-01	1.D-05	1.D-05	1.D-06	1.D-05	11C	0	0
4	1.D+01	1.000D+00	3.D-07	2.500D-01	3.D-07	3.D-07	5.D-10	3.D-07	12C	0	0
5	1.D+01	1.000D+00	1.D-08	2.500D-01	1.D-08	1.D-08	8.D-14	1.D-08	13C	0	0

Flag of ALGENCAN: Solution was found.

ρ não aumenta