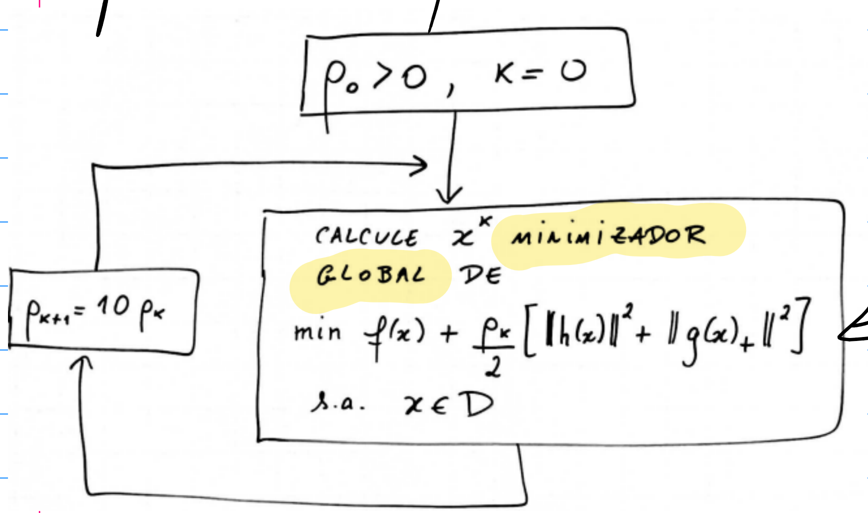
Convergência do esquema de penalidade externa

P: $\min_{x} f(x)$ s.a. $h(x)=0$, $g(x)\le 0$, $x\in D$,
D compacto.

Notação: $g(x)_+ = \max\{g(x), 0\}$.

Esquema de penalidade externa:



$SP(\rho_k)$.

- O esquema gera uma sequência $\{x^k\} \subset D$.
- Como D é compacto, $\{x^k\}$ admite ponto de acumulação x^*.

Vamos mostrar que:

1) P viável $\Rightarrow$ x^* é viável
2) x^* é minimizador global de P.

1) __Teorema__: Seja $\{x^k\}$ a sequência gerada pelo esquema e x^* um ponto de acumulação seu. Se P é viável, então x^* é viável para P.

__Prova__: Como x^* é ponto de acumulação de $\{x^k\}$, existe um conjunto infinito de índices K tal que

$$\lim_{k \in K} x^k = x^*.$$

Suponha por contradição que x^* não é viável. Então existe $z \in D$ viável tal que

$$\phi(x^*) = \|h(x^*)\|^2 + \|g(x^*)_+\|^2 > \|h(z)\|^2 + \|g(z)_+\|^2 = \phi(z) = 0$$

Existe uma constante $c > 0$ tal que

$$\|h(x^k)\|^2 + \|g(x^k)_+\|^2 > \|h(z)\|^2 + \|g(z)_+\|^2 + c,$$

para todo $k \gg 1$ (k suficientemente grande).

Logo,

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

$$> f(x^k) + \frac{\rho_k}{2}\left[\|h(z)\|^2 + \|g(z)_+\|^2\right] + \frac{\rho_k}{2}C$$

$$= f(z) + \frac{\rho_k}{2}\left[\|h(z)\|^2 + \|g(z)_+\|^2\right] + \left(f(x^k) - f(z) + \frac{\rho_k}{2}C\right)$$

Lembre-se que no esquema de penalidade, $\rho_k \to \infty$. Daí, $\lim_{k \in K}\left(f(x^k) - f(z) + \frac{\rho_k C}{2}\right) = \infty$.

Portanto,

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

$$> f(z) + \frac{\rho_k}{2}\left[\|h(z)\|^2 + \|g(z)_+\|^2\right], \quad \forall k \gg 1.$$

Isso contraria o fato de x^k ser minimizador global de $SP(\rho_k)$. Daí, x^* é viável ∎

2) <u>Teorema</u>: Seja $\{x^k\}$ a sequência gerada pelo esquema e x^* um ponto de acumulação seu. Se P é viável, então x^* é minimizador global de P. (o esquema resolve P).

<u>Prova</u>: Pelo teorema anterior, x^* é viável para P. Pela construção do esquema, x^k é min. global de $SP(\rho_k)$, ou seja, para todo k,

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

$$\leq f(z)+\frac{\rho_k}{2}\left[\|h(z)\|^2+\|g(z)_+\|^2\right], \quad \forall z\in D.$$

Em particular, para $z \in D$ viáveis, isto é, $h(z)=0$ e $g(z)_+=0$, temos

$$f(x^k)+\frac{\rho_k}{2}\left[\|h(x^k)\|^2+\|g(x^k)_+\|^2\right] \leq f(z), \quad \forall k.$$

Tomando um conjunto de índices K tal que

$\lim_{k \in K} x^k = x^*$, passamos o limite sobre K na desigualdade acima para obtermos

$$f(x^*) \leq f(z), \quad \forall z \in D \text{ viável para } P.$$

Isto é, x^* é minimizador global de P ∎

Prova das condições KKT usando o esquema de penalidade externa

x^* é KKT se é viável e existirem $\lambda_1, \ldots, \lambda_m$ e $\mu_1, \ldots, \mu_p$ tais que

$$\nabla f(x^*) + \sum_{i=1}^{m} \lambda_i \nabla h_i(x^*) + \sum_{j=1}^{p} \mu_j \nabla g_j(x^*) = 0$$

$$\mu_j \geq 0, \quad \forall j$$

$$\mu_j g_j(x^*) = 0, \quad \forall j$$

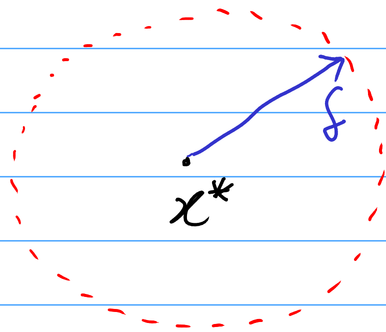
---

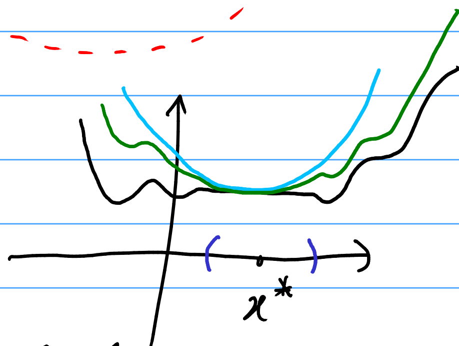
x^* é <u>regular</u> se os gradientes das restrições ativas $\nabla h_i(x^*)$, $i = 1, \ldots, m$ e $\nabla g_j(x^*)$, $j : g_j(x^*) = 0$, são L.I.

Teorema: Se x^* é minimizador local de P e regular então x^* é ponto KKT.

Prova: Como x^* é minimizador local de P, então existe $\delta > 0$ tal que x^* é o único minimizador global de

$$P': \min_x f(x) + \frac{1}{2}\|x - x^*\|^2$$
$$\text{s.a. } h(x) = 0$$
$$g(x) \leq 0$$
$$\|x - x^*\| \leq \delta \quad \Big) D$$





Aplicamos o esquema de penalidade externa ao problema P', penalizando $h(x) = 0$ e $g(x) \leq 0$. Os subproblemas do esquema são

$$SP(\rho_k): \min_x f(x) + \frac{1}{2}\|x^* - x\|^2 + \frac{\rho_k}{2}\left[\|h(x)\|^2 + \|g(x)_+\|^2\right]$$
$$\text{s.a. } \|x - x^*\| \leq \delta.$$

Temos $\rho_k \to \infty$ e $\{x^k\}$ gerada pelo esquema, formada por minimizadores globais de $SP(\rho_k)$. Como $\|x^k - x^*\| \leq \delta$, $\forall k$, a sequência $\{x^k\}$ admite ponto de acumulação. A teoria de penalidade assegura que este ponto de acumulação é min. global de P'. (2º teorema). Como x^* é o único min. global de P', só pode ser $\lim\limits_{k \in K} x^k = x^*$.
Daí, $\|x^k - x^*\| < \delta$, $\forall k \gg 1$, $k \in K$, isto é, a restrição $\|x - x^*\| \leq \delta$ de $SP(\rho_k)$ torna-se inativa nos pontos x^k. Assim, $SP(\rho_k)$ é, ao redor de x^k, essencialmente um problema irrestrito. Daí, x^k anula o gradiente da função objetivo de $SP(\rho_k)$:

$$\nabla f(x^k) + (x^k - x^*) + \sum_{i=1}^{m} [\rho_k h_i(x^k)] \nabla h_i(x^k)$$

$$+ \sum_{j=1}^{p} [\rho_k g_j(x^k)_+] \nabla g_j(x^k) = 0, \; \forall k \gg 1, \; k \in K.$$

Assim,

$$\nabla f(x^k) + \sum_{i=1}^{m} \lambda_i^k \nabla h_i(x^k) + \sum_{j=1}^{p} \mu_j^k \nabla g_j(x^k) = -(x^k - x^*) \qquad (1)$$

onde $\lambda_i^k = \rho_k h_i(x^k)$ e $\mu_j^k = \rho_k g_j(x^k)_+$.

Definimos, para todo k,

$$\gamma_k = \|(\lambda^k, \mu^k)\|_\infty = \max\{|\lambda_1^k|, \dots, |\lambda_m^k|, |\mu_1^k|, \dots, |\mu_p^k|\}.$$

<u>CASO 1:</u> $\{\gamma_k\}$ possui subsequência convergente.

Neste caso, $\{\lambda^k\}$ e $\{\mu^k\}$ admitem pontos de acumulação, digamos λ^* e μ^*.

Assim, passando o limite $K \ni k \to \infty$ em (1), obtemos

$$\nabla f(x^*) + \sum_{i=1}^{m} \lambda_i^* \nabla h_i(x^*) + \sum_{j=1}^{p} \mu_j^* \nabla g_j(x^*) = 0$$

Agora, veja que $\mu_j^k = \rho_k g_j(x^k)_+$ $= \rho_k \max\{g_j(x^k), 0\} \geq 0$, $\forall k, \forall j$. Daí $\mu_j^* \geq 0$, $\forall j$. Mais ainda, se $g_j(x^*) < 0$

então $g_j(x^k) < 0$, $\forall k \gg 1$, $k \in K$.

Daí, $\mu_j^k = \rho_k \max\{g_j(x^k), 0\} = 0$ para estes j. Assim, $\mu_j^* g_j(x^*) = 0$, $\forall j$.

Ou seja, x^* é ponto KKT.

CASO 2: $\{\rho_k\}$ não possui subseq. convergente.

Aqui, $\lim_{k\to\infty} \rho_k = \infty$. Dividindo (1) por ρ_k, obtemos

$$\underbrace{\frac{Df(x^k)}{\rho_k}}_{\to 0} + \underbrace{\sum_{i=1}^{m} \left(\frac{\lambda_i^k}{\rho_k}\right) Dh_i(x^k) + \sum_{j=1}^{p} \left(\frac{\mu_j^k}{\rho_k}\right) Dg_j(x^k)}_{\to 0} = \underbrace{-\frac{(x^k - x^*)}{\rho_k}}_{\to 0}$$

$$\Rightarrow \lim_{k\to\infty} \sum_{i=1}^{m} \underbrace{\left(\frac{\lambda_i^k}{\rho_k}\right)}_{\text{limitada}} Dh_i(x^k) + \sum_{j=1}^{p} \underbrace{\left(\frac{\mu_j^k}{\rho_k}\right)}_{\text{limitada}} Dg_j(x^k) = 0. \qquad (2)$$

Pela definição de ρ_k, as sequências $\left\{\underline{\frac{\lambda_i^k}{\rho_k}}\right\}$ e $\left\{\underline{\frac{\mu_j^k}{\rho_k}}\right\}$ são limitadas em $[-1,1]$.

Assim, elas admitem pontos de acumulação, digamos λ_i^* e μ_j^*, $\forall i,j$.
Novamente pela definição de ρ_k, pelo menos um λ_i^* ou μ_j^* é $\neq 0$.
Assim, (2) fornece

$$\sum_{i=1}^{m} \lambda_i^* \nabla h_i(x^*) + \sum_{j=1}^{p} \mu_j^* \nabla g_j(x^*) = 0, \quad (3)$$

onde $(\lambda^*, \mu^*) \neq 0$. Veja que

$\mu_j^k = \rho_k \max\{g_j(x^k), 0\} = 0$, $\forall k \gg 1$ caso $g_j(x^*) < 0$. Assim, (3) fornece

$$\sum_{i=1}^{m} \lambda_i^* \nabla h_i(x^*) + \sum_{j: g_j(x^*) = 0} \mu_j^* \nabla g_j(x^*) = 0,$$

isto é, os gradientes das restrições ativas em x^* são L.D. Mas isso contraria a hipótese de regularidade em x^*. Assim, não pode ser $\rho_k \to \infty$, e logo recaímos no CASO 1. Portanto, x^* é KKT. ∎