Programação Quadrática Sequencial
(SQP / PQS)

Ribeiro, A. A; Karas, E. W. Otimização contínua. Cengage, 2014

Martínez, J. M.; Santos, S. A. Métodos computacionais de otimização

SQP "básico" aplica-se à problemas com restrições de igualdade apenas

$$\min f(x)$$
$$\text{s.a. } h(x) = 0. \quad \} \; m \text{ restrições}$$

KKT:
$$\nabla f(x) + \sum_{i=1}^{m} \lambda_i \nabla h_i(x) = 0$$
$$h(x) = 0 \qquad h: \mathbb{R}^n \to \mathbb{R}^m$$

Função Lagrangeano (Lagrangeana) $\quad h = (h_1, \ldots, h_m)$

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

$$\begin{bmatrix} | & & | \\ \nabla h_1(x) & \cdots & \nabla h_m \\ | & & | \end{bmatrix}$$

Reescrevemos KKT como

$$\nabla_x L(x,\lambda) = 0, \qquad h(x) = 0.$$

$$\left(\nabla_x L(x,\lambda) = \nabla f(x) + \sum_{i=1}^{m} \lambda_i \nabla h_i(x) = \nabla f(x) + \nabla h(x)\lambda \right)$$

SQP = Newton no sistema KKT

$$F(x,\lambda) = \begin{bmatrix} \nabla_x L(x,\lambda) \\ h(x) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

Direção de Newton $d = (d_x, d_\lambda)$:

$$F'(x,\lambda) = \begin{bmatrix} \nabla^2_{xx} L(x,\lambda) & \nabla h(x) \\ \nabla h(x)^T & 0 \end{bmatrix}$$

$$\begin{bmatrix} \nabla^2_{xx} L(x,\lambda) & \nabla h(x) \\ \nabla h(x)^T & 0 \end{bmatrix} \begin{bmatrix} d_x \\ d_\lambda \end{bmatrix} = - \begin{bmatrix} \nabla_x L(x,\lambda) \\ h(x) \end{bmatrix}$$

$$\Rightarrow \begin{cases} \nabla^2_{xx} L(x,\lambda)\, d_x + \nabla h(x)\, d_\lambda = -\nabla_x L(x,\lambda) \\ \nabla h(x)^T d_x = -h(x) \end{cases}$$

$$\Rightarrow \begin{cases} \left[\nabla^2 f(x) + \sum_{i=1}^{m} \lambda_i \nabla^2 h_i(x)\right] d_x + \sum_{i=1}^{m} (d_\lambda)_i \nabla h_i(x) \\ \qquad = -\nabla f(x) - \sum_{i=1}^{m} \lambda_i \nabla h_i(x) \\ \nabla h(x)^T d_x = -h(x) \end{cases}$$

$$\Rightarrow \begin{cases} \left[\nabla^2 f(x) + \sum_{i=1}^{m} \lambda_i \nabla^2 h_i(x)\right] d_x + \sum_{i=1}^{m} (d_\lambda + \lambda)_i \nabla h_i(x) = -\nabla f(x) \\ \nabla h(x)^T d_x = -h(x) \end{cases}$$

Chamando $d_\lambda^+ = d_\lambda + \lambda$, obtemos

$$\begin{cases} \left[\nabla^2 f(x) + \sum_{i=1}^{m} \lambda_i \nabla^2 h_i(x)\right] d_x + \nabla h(x)\, d_\lambda^+ = -\nabla f(x) \\ \nabla h(x)^T d_x = -h(x) \end{cases}$$

O sistema acima é o sistema KKT do problema quadrático

$$QP: \min_{d_x} \frac{1}{2} d_x^T \left[\nabla^2 f(x) + \sum_{i=1}^{m} \lambda_i \nabla^2 h_i(x)\right] d_x + \nabla f(x)^T d_x$$

$$\text{s.a. } \nabla h(x)^T d_x + h(x) = 0 \quad (*)$$

De fato, KKT de QP é

multiplicador de Lagrange de (*)

$$\begin{cases} \left[\nabla^2 f(x) + \sum_{i=1}^{m} \lambda_i \nabla^2 h_i(x)\right] d_x + \nabla f(x) + \nabla h(x)\, d_\lambda^+ = 0 \\ \nabla h(x)^T d_x + h(x) = 0 \end{cases}$$

Observação:

$\lambda_1 \nabla h_1 + \lambda_2 \nabla h_2 + \cdots + \lambda_m \nabla h_m$

$$F(x,\lambda) = \begin{bmatrix} \nabla_x L(x,\lambda) \\ h(x) \end{bmatrix} = \begin{bmatrix} \nabla f(x) + \sum \lambda_i \nabla h_i(x) \\ h(x) \end{bmatrix}$$

$$F'(x,\lambda) = \begin{array}{c} \\ \\ \end{array}\begin{bmatrix} \vdots & \nabla h_1(x) & \nabla h_2(x) & & \nabla h_m(x) \\ \vdots & 0 & 0 & \cdots & 0 \\ & \vdots & \vdots & & \vdots \\ & 0 & 0 & & 0 \end{bmatrix} \quad \nabla h(x)$$

$$\begin{array}{ccccc} \uparrow & \uparrow & \uparrow & \cdots & \uparrow \\ \partial x & \partial \lambda_1 & \partial \lambda_2 & & \partial \lambda_m \end{array}$$

Observação: a minimização em QP é somente em d_x. A componente d_λ^+ do sistema Newtoniano é o multiplicador de Lagrange associado à restrição linear $\nabla h(x)^T d_x + h(x) = 0$. Ou seja, ao resolver QP, obtemos a solução d_x e o multiplicador d_λ^+.

Problema com QP: Se $\hat{H} = \nabla^2 f + \sum \lambda_i \nabla^2 h_i$ não for definida positiva, QP pode ser não convexo.

IDEIA: Trocar $\hat{H}$ por outra matriz H simétrica e definida positiva. Neste caso, resolvemos

$$QP: \min_{d_x} \frac{1}{2} d_x^T H d_x + \nabla f(x)^T d_x$$

$$\text{s.a. } \nabla h(x)^T d_x + h(x) = 0$$

} linearização de $h(x) = 0$ no ponto x.

Algumas opções para H:

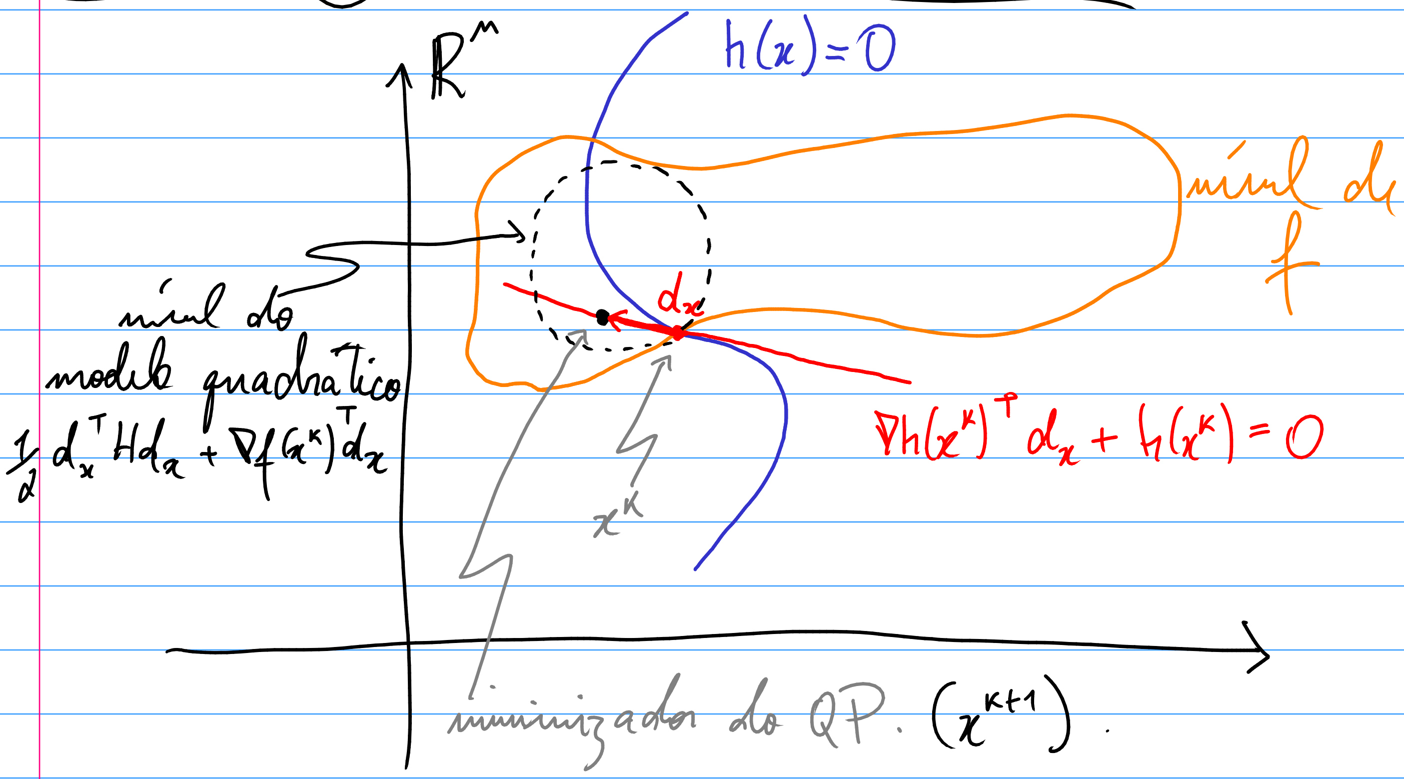
1) H como atualização quase-Newton (BFGS, DFP, ...)

2) $H = \alpha I$ (problemas de grande porte)

3) $H = \hat{H} + \sigma I$, para $\sigma > 0$ grande.

(σ pode ser computado calculando o menor autovalor de $\hat{H}$ ou utilizando discos de

Gershgorin + heurísticas).



SQP básico

(x^0, λ^0) qualquer , $k=0$

$\downarrow$

calcule H_k

$\downarrow$

resolva o QP: $\min_{d_x} \frac{1}{2} d_x^T H_k d_x + \nabla f(x^k)^T d_x$

s.a. $\nabla h(x^k)^T d_x + h(x^k) = 0$

obtendo a solução d_x^k com multip. $d_\lambda^{+,k}$

$\downarrow$

$x^{k+1} = x^k + d_x^k$, $\lambda^{k+1} = \lambda^k + d_\lambda^{+,k}$

$k \leftarrow k+1$

Observações:

1) Este método só tem convergência local, isto é, se iniciamos (x_0, λ_0) próximo à solução (do sistema KKT do problema original) então o método converge à solução.

2) Há formas de obter convergência global, isto é, iniciando (x_0, λ_0) q.q.

↳ busca linear: $x^{k+1} = x^k + \alpha d_x^k$, $\alpha \in (0,1]$.
(o mesmo p/ λ)

↳ usar regiões de confiança no QP.

3) Há pacotes computacionais que implementam SQP.

- SNOPT: versão demonstração com AMPL.
- WORHP: completo para academia
  ↳ há interface para Julia.