Método (primal dual) seguidor de caminhos 1

P: $\min_{x} c^t x$ s.a. $Ax=b$, $x \geq 0$

D: $\max_{y,z} b^t y$ s.a. $A^t y + z = c$, $z \geq 0$

Otimalidade:

$$\begin{cases} Ax - b = 0 & (-r_p = 0) \\ A^t y + z - c = 0 & (-r_d = 0) \\ XZe = 0 & (-r_c = 0) \\ (x \geq 0, \ z \geq 0) & \end{cases}$$

Direção afim escala: (Newton)

$$(ADA^t) d_y = r_p + AD(r_d - X^{-1} r_c)$$

$$d_x = D(A^t d_y - r_d + X^{-1} r_c)$$

$$d_z = X^{-1}(r_c - Z d_x).$$

O método primal dual afim escala não é bom na prática pois seus iterandos x^k, z^k levam à complementaridades $x_i^k z_i^k$, $i=1,\dots,n$ muito diferentes entre si:

Newton:

$$\begin{bmatrix} A & 0 & 0 \\ 0 & A^t & I \\ Z & 0 & X \end{bmatrix} \begin{bmatrix} d_x \\ d_y \\ d_z \end{bmatrix} = \begin{bmatrix} r_p \\ r_d \\ r_c \end{bmatrix}$$

(lembre-se que $x_i^k z_i^k > 0$).

$$\rightarrow r_{c_i} = -x_i z_i \neq -x_j z_j = r_{c_j}$$

<u>Ideia</u>: forçar todos $x_i z_i$, $i=1,\ldots,n$, serem iguais à um mesmo valor $\mu > 0$:

Otimalidade aproximada (KKT perturbado):

$$\begin{cases} Ax - b = 0 & (x \geq 0) \\ A^t y + z - c = 0 & (z \geq 0) \\ XZe = \mu e \iff x_i z_i = \mu, \ \forall i \end{cases}$$

Quando $\mu \to 0^+$, recuperamos a otimalidade verdadeira.

No método seguidor de caminhos, calculamos a direção de Newton relativa à otimalidade perturbada e levamos $\mu^k \to 0$:

$$F(x,y,z) = \begin{bmatrix} Ax - b \\ A^t y + z - c \\ XZe - \mu e \end{bmatrix} = \begin{bmatrix} -r_p \\ -r_d \\ -r_c \end{bmatrix}$$

$$\left(r_c = -XZe + \underline{\underline{\mu e}} \right).$$

$$F'(x,y,z) = \begin{bmatrix} A & 0 & 0 \\ 0 & A^t & I \\ Z & 0 & X \end{bmatrix} \quad (= \text{afim escala} \dots)$$

Resolvendo $F'd = -F = r$, obtemos, de maneira análoga ao método afim escala,

$$(ADA^t)\, d_y = r_p + AD(r_d - X^{-1} r_c)$$
$$d_x = D(A^t d_y - r_d + X^{-1} r_c)$$
$$d_z = X^{-1}(r_c - Z d_x)$$

(idêntico ao caso anterior, exceto que agora

$r_c = -XZe + \mu e$ carrega o parâmetro de centralidade $\mu > 0$).

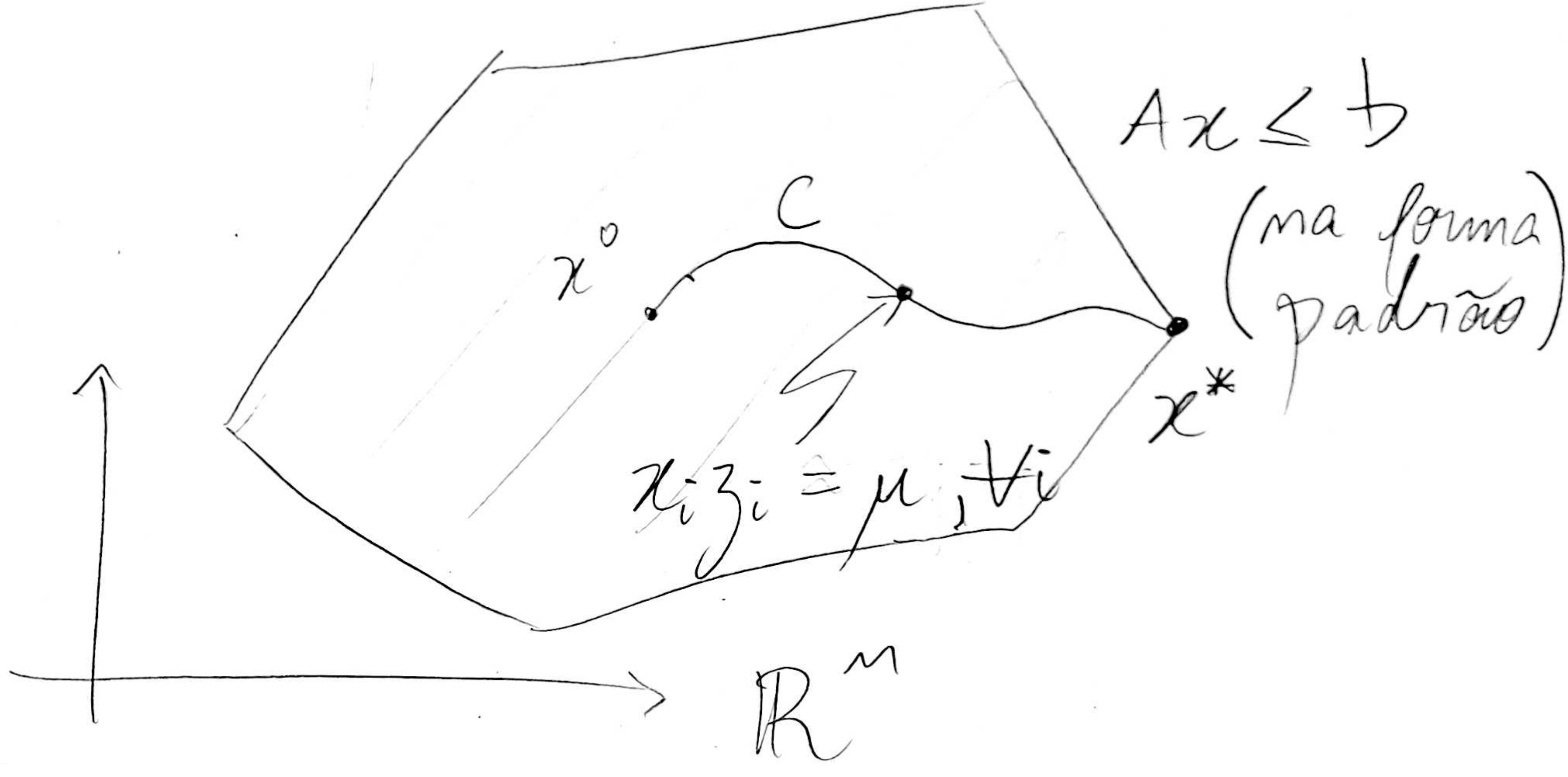
- Os resíduos r_p e r_d são como anteriormente
- Os tamanhos de passo α_p e α_d são como anteriormente.

Por que "seguidor de caminhos"?

↳ porque agora $\{(x^k, y^k, z^k)\}$ segue o caminho central

$C = \{(x_\mu, y_\mu, z_\mu)$; solução da otimalidade perturbada para $\mu > 0\}$

$(\mu \downarrow 0)$.



Como levar $\mu^k \to 0$?

↳ tomar μ^k considerando a média aritmética dos produtos $x_i z_i$, $i = 1, \dots, n$:

$$\mu^k = \sigma^k \frac{(x^k)^t z^k}{n}, \quad \sigma^k \in (0,1)$$

Veja que $(x^k)^t z^k \to 0 \Rightarrow \mu^k \to 0$.

Escolhas práticas para σ^k:

1) $\sigma^k \equiv 1/\sqrt{n}$ (constante)

2) $$\sigma^k = \begin{cases} \dfrac{(x^k)^t z^k}{\sqrt{n}}, & \text{se } (x^k)^t z^k < 1 \\ \dfrac{1}{\sqrt{n}}, & \text{caso contrário.} \end{cases}$$

9

Esta escolha vem da prática e visa acelerar o método caso todos os produtos $x_i^k z_i^k$ forem pequenos ($(x^k)^t z^k < 1 \rightarrow$ próximo à solução). Veja que $\dfrac{(x^k)^t z^k}{\sqrt{n}} < \dfrac{1}{\sqrt{n}}$ se $(x^k)^t z^k < 1$.

Método seguidor de caminhos.

Dado (x^0, y^0, z^0) com $x^0 > 0$ e $z^0 > 0$.
(não necessariamente primal dual viável)

para $k = 0, \dots, maxit$

$$\boxed{\mu^k = \sigma^k \frac{(x^k)^t z^k}{n}}$$

$r_p^k = b - Ax^k$, $r_d^k = c - A^t y^k - z^k$, $\boxed{r_c^k = \mu^k e - X_k Z_k e}$

$D_k = Z_k^{-1} X_k$

$d_y^k = (A D_k A^t)^{-1} \left[r_p^k + A D_k (r_d^k - X_k^{-1} r_c^k) \right]$

$d_x^k = D_k (A^t d_y^k - r_d^k + X_k^{-1} r_c^k)$

$$d_z^k = X_k^{-1}(r_c^k - Z_k d_x^k)$$

$$\alpha_p^k = \tau \min_i \{ -x_i^k / d_{x_i}^k \ ; \ d_{x_i}^k < 0 \}$$

$$\alpha_d^k = \tau \min_i \{ -z_i^k / d_{z_i}^k \ ; \ d_{z_i}^k < 0 \}$$

$$x^{k+1} = x^k + \alpha_p^k d_x^k$$

$$y^{k+1} = y^k + \alpha_d^k d_y^k$$

$$z^{k+1} = z^k + \alpha_d^k d_z^k$$

até "convergir"

Comentários.

1) o custo por iteração é o mesmo dos métodos anteriores (= afim escala a menos de μ e r_c)

2) este método balanceia $x_i^k z_i^k$ ao redor de μ^k, ao contrário do método afim escala.

3) Como anteriormente, na prática fazemos $\alpha_p^k \leftarrow \min\{1, \alpha_p^k\}$ e $\alpha_d^k \leftarrow \min\{1, \alpha_d^k\}$.

4) Inicialização de (x^0, y^0, z^0) e critério de parada são os mesmos do método afim escala.

5) todas as questões sobre o sistema de Newton são as mesmas (Cholesky, AMD...)

---

6) Parâmetros de referência: $\tau = 0,99995$ e opção (2) para σ^k.

---

7) Este método é polinomial (com σ^k variando com k). Ele é bom numericamente.