

# Método preditor-corretor (de Mehrotra). (1

$$Ax - b = 0, \quad A^t y + z - c = 0, \quad XZe = \mu e$$

- primal dual afim escala: direção de Newton na otimalidade verdadeira ( $\mu = 0$ )
- seguidor de caminhos: Newton na otimalidade perturbada ( $\mu > 0$ ), com  $\mu^k \rightarrow 0^+$ .
  - ↳ segue trajetória central ( $x_i z_i \approx \mu, \forall i$ )
  - ↳ + estável.

## Ideia do método preditor-corretor:

2

- 1) calcule a direção afim escala (direção preditora —  $\mu = 0$ );
- 2) corrija a direção afim escala somando a ela uma direção corretora obtida com  $\mu > 0$ , mas considerando os resíduos que seriam obtidos após o passo afim escala.

Direção preditora (afim escala)  $\hat{d}$ : (3)

$$A\hat{d}_x = r_p^k, \quad A^t\hat{d}_y + \hat{d}_z^k = r_d^k, \quad Z_k\hat{d}_x + X_k\hat{d}_z^k = \hat{r}_c^k$$

onde  $\hat{r}_c^k = -X_k Z_k e$ .

Sendo otimistas, vamos supor que os iterandos do método afim escala sejam interiores com passos  $\hat{\alpha}_p^k = \hat{\alpha}_d^k = 1$  (passo de Newton "puro").

Pergunta: quais os resíduos no novo ponto?

14

$$\begin{aligned} \bullet \quad r_p^{k+1} &= b - A \hat{x}^{k+1} = b - A(\hat{x}^k + \hat{d}_x^k) \\ &= (b - A \hat{x}^k) - A \hat{d}_x^k = (b - A \hat{x}^k) - \hat{r}_p^k = 0 \end{aligned}$$

$$\begin{aligned} \bullet \quad \hat{r}_d^{k+1} &= A^t \hat{y}^{k+1} + \hat{z}^{k+1} - c = A^t(\hat{y}^k + \hat{d}_y^k) + \hat{z}^k + \hat{d}_z^k - c \\ &= \underbrace{(A^t \hat{y}^k + \hat{z}^k - c)}_{-\hat{r}_d^k} + \underbrace{(A^t \hat{d}_y^k + \hat{d}_z^k)}_{\hat{r}_d^k} = 0 \end{aligned}$$

$$\begin{aligned} \bullet \quad \hat{r}_{c_i}^{k+1} &= -(\hat{x}_i^k + \hat{d}_{x_i}^k)(\hat{z}_i^k + \hat{d}_{z_i}^k) = -\cancel{\hat{x}_i^k \hat{z}_i^k} - \hat{d}_{x_i}^k \hat{d}_{z_i}^k \\ &\quad - \cancel{(\hat{z}_i^k \hat{d}_{x_i}^k + \hat{x}_i^k \hat{d}_{z_i}^k)} = -\hat{d}_{x_i}^k \hat{d}_{z_i}^k \end{aligned}$$

Matricialmente, escrevemos

15

$$\hat{\pi}_c^{k+1} = -\hat{D}_k^x \hat{D}_k^z e, \text{ onde } \hat{D}_k^x = \text{diag}(\hat{d}_x^k) \\ \text{e } \hat{D}_k^z = \text{diag}(\hat{d}_z^k).$$

Direção corretora  $\tilde{d}^k$ : estabilidade perturbada  
da com os resíduos afim escala com passo 1:

$$A \tilde{d}_x^k = 0, \quad A^t \tilde{d}_y^k + \tilde{d}_z^k = 0, \quad Z_k \tilde{d}_x^k + X_k \tilde{d}_z^k = \tilde{\pi}_c^k,$$

$$\text{onde } \tilde{\pi}_c^k = \hat{\pi}_c^k + \mu^k e = -\hat{D}_k^x \hat{D}_k^z e + \mu^k e.$$

Direção preditora-corretora:  $d^k = \hat{d}^k + \tilde{d}^k$  16

A direção  $d^k$  é calculada tomando os sistemas relativos a  $\hat{d}$  e  $\tilde{d}$ :

$$\begin{cases} A d_x^k = r_p^k \\ A^t d_y^k + d_z^k = r_d^k \\ Z_k d_x^k + X_k d_z^k = \hat{r}_s^k - \tilde{r}_c^k \end{cases}$$

onde  $r_s^k = \hat{r}_c^k + \tilde{r}_c^k = -X_k Z_k e - \hat{D}_k^x \hat{D}_k^z + \mu^k e$ .

- Devemos resolver 2 sistemas, um para  $\hat{d}$  e outro para  $d$ .

Fazendo as contas, recaímos no mesmo sistema com matriz  $AD_k A^t \dots$

---

- os tamanhos de passo devem ser calculados para cada direção:  $\hat{\alpha}_p$ ,  $\hat{\alpha}_d$  e  $\alpha_p$ ,  $\alpha_d$  de modo a manter interioridade de do passo afim escala e do passo composto.

- calculada a direção afim escala  $\hat{d}^k$ , (8  
devemos calcular o parâmetro de centralidade  $\mu^k$  para a direção composta.

Da mesma forma que fizemos no método seguidor de caminhos, fazemos

$$\mu^k = \sigma^k \frac{p^k}{n},$$

onde  $p^k = (x^k)^t z^k$ . Mas agora, a direção corretora usa os resíduos no ponto.

$$(x^k + \hat{d}_x^k, y^k + \hat{d}_y^k, z^k + \hat{d}_z^k).$$



agregamos a informação após o passo  $19$   
afim escala em  $\sigma^k$ , tomando

$$\sigma^k = \left( \frac{\hat{y}^k}{y^k} \right)^p,$$

onde  $\hat{y}^k = (x^k + \hat{\alpha}_p^k \hat{d}_x^k)^t (z^k + \hat{\alpha}_d^k \hat{d}_z^k)$  e

$p = 2$  ou  $3$ . Veja que usamos os passos  
 $\hat{\alpha}_p^k$  e  $\hat{\alpha}_d^k$  para garantir  $\hat{y}^k > 0$ .

- Inicialização e parada são os mesmos dos métodos afim escala e seguidor de caminhos.

# Método preditor-corretor

10

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

para  $k=0, \dots, \maxit$

$$r_p^k = b - Ax^k, \quad r_d^k = c - A^t y^k - z^k, \quad r_c^k = -X_k z^k e$$

$$D_k = Z_k^{-1} X_k$$

$$\hat{d}_y^k = (AD_k A^t)^{-1} [r_p^k + AD_k (r_d^k - X_k^{-1} r_c^k)]$$

$$\hat{d}_x^k = D_k (A^t \hat{d}_y^k - r_d^k + X_k^{-1} r_c^k)$$

$$\hat{d}_z^k = X_k^{-1} (r_c^k - Z_k \hat{d}_x^k)$$

direção  
preditora

$$\hat{\alpha}_p^k = \min \{ 1, \min_i \} - \frac{x_i^k}{\hat{d}_{x_i}^k}; \hat{d}_{x_i}^k < 0 \{ \{ \underline{11}$$

$$\hat{\alpha}_d^k = \min \{ 1, \min_i \} - \frac{z_i^k}{\hat{d}_{z_i}^k}; \hat{d}_{z_i}^k < 0 \{ \{$$

$$\hat{y}^k = (x^k + \hat{\alpha}_p^k \hat{d}_x^k)^t (z^k + \hat{\alpha}_d^k \hat{d}_z^k)$$

$$y^k = (x^k)^t z^k$$

$$\mu^k = \left( \frac{\hat{y}^k}{y} \right)^p \frac{y^k}{n}$$

$$n_s^k = n_c^k - \hat{D}_x^k \hat{D}_z^k e + \mu^k e.$$

direção composta  
(pred. + correção)

12

$$d_y^k = (AD_k A^t)^{-1} [r_p^k + AD_k (r_d^k - X_k^{-1} r_s^k)]$$

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

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

$$\alpha_p^k = \min\{1, \min_i \frac{x_i^k}{d_{x_i}^k} ; d_{x_i}^k < 0\}$$

$$\alpha_d^k = \min\{1, \min_i \frac{z_i^k}{d_{z_i}^k} ; d_{z_i}^k < 0\}$$

$$x^{k+1} = x^k + \alpha_p^k d_x^k, \quad y^{k+1} = y^k + \alpha_d^k d_y^k, \quad z^{k+1} = z^k + \alpha_d^k d_z^k$$

até "convergir".

## Comentários finais:

13

1) A iteração deste método é mais cara, pois devemos resolver 2 sistemas: um para  $\hat{d}^k$  e outro para  $d^k$ . No entanto, note que ambos têm mesma matriz  $AD_k A^t$ . Logo, realizamos apenas uma única fatoração Cholesky.

2) Este método costuma fazer menos iterações que os outros, o que compensa o custo.

3) Este método é polinomial e possui 14  
convergência quadrática se considerarmos  
a resolução dos dois sistemas numa única  
iteração.

---

4) Este método é a base das implementa-  
ções modernas e eficientes (CPLEX etc)

---

5) As implementações mais eficientes possuem [15]

(i) AMD em  $AA^t$  (fizemos)

(ii) pré-processamento (eliminação de restrições e variáveis redundantes)

(iii) pré-condicionadores para resolução dos sistemas com matriz  $ADA^t$ .

(iv) Tratamento de colunas densas em  $A$  (fatoração de  $ADA^t$  refinada)

(v) múltiplas correções:  $\hat{d} + \tilde{d}^1 + \dots + \tilde{d}^q$ .