Método primal dual afim escala

Ideia: dar um passo de Newton para resolver o sistema de otimalidade

$$Ax - b = 0$$

$$A^t y + z - c = 0$$

$$XZe = 0$$

$$\cancel{x \geq 0}$$

$$\cancel{z \geq 0}$$

Vamos descartar $x \geq 0$ e $z \geq 0$. A interioridade será tratada ajustando o tamanho dos passos.

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

Ponto corrente: w^k

Newton: $w^{k+1} = w^k + d^k$ onde $F'(w^k)\, d^k = -F(w^k)$,

$w = (x, y, z)$

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

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

Assim $F'd = -F$

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

$$\Rightarrow \begin{cases} A d_x = r_p & (1) \\ A^t d_y + d_z = r_d & (2) \\ Z d_x + X d_x = r_c & (3) \end{cases}$$

Supor que $x > 0$ e $z > 0$.

De (3) vem $\boxed{d_z = X^{-1}(r_c - Z d_x)}$

De (2),

$$A^t d_y + d_z = r_d \Rightarrow A^t d_y = X^{-1} Z d_x + r_d - X^{-1} r_c$$

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

$$\Rightarrow (Z^{-1}X) A^t d_y = d_x + (Z^{-1}X)(r_d - X^{-1} r_c) \quad (*)$$

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

$$\underset{(1)}{\Rightarrow} \boxed{(ADA^t)\, d_y = r_p + AD(r_d - X^{-1} r_c)}$$

onde $\boxed{D = Z^{-1} X}$.

Finalmente, de (*) segue que

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

Em resumo (1), (2), (3) fica (supondo $x, z > 0$)

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

onde $D = Z^{-1} X$

Queremos que

$x^{k+1} = x^k + \alpha_p^k d_x^k > 0$:

$$\alpha_p^k \leq \tau \min_i \left\{ -\frac{x_i^k}{d_{x_i}^k} \; ; \; d_{x_i}^k < 0 \right\}.$$

$z^{k+1} = z^k + \alpha_d^k d_z^k > 0$:

$$\alpha_d^k \leq \tau \min_i \left\{ -\frac{z_i^k}{d_{z_i}^k} \; ; \; d_{z_i}^k < 0 \right\},$$

onde $\tau \in (0,1)$. α_p e α_d: "passos primal / dual."

Método primal dual afim escala

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

para $k = 0, \ldots, maxit$

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

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

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

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

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

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

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

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

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

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

até "convergir"

Observação: a convergência teórica foi estabelecida supondo um único passo:

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

onde

$$\alpha^k = \min\{\alpha_p^k, \alpha_d^k\}.$$

Porém, isso é um pouco pior numericamente.

Observações:

1) o custo por iteração é similar aos métodos primal e dual.

2) este método não acumula erros de arredondamento como os anteriores

3) É fácil inicializar pois não requer viabilidade de (x^0, y^0, z^0).

4) Na prática, fazemos $\alpha_p^k \leftarrow \min\{1, \alpha_p\}$ e $\alpha_d^k \leftarrow \min\{1, \alpha_d\}$ (não ultrapassa o passo de Newton)

Critério de parada

Ideia: já que trabalhamos com variáveis primal (x) e duais (y, z) simultaneamente, parar com as condições de otimalidade satisfeitas aproximadamente:

- $\dfrac{\|b - Ax^k\|}{1 + \|b\|} = \dfrac{\|r_p^k\|}{1 + \|b\|} \leq \varepsilon$ (viab. primal)

- $\dfrac{\|c - A^t y^k - z^k\|}{1 + \|c\|} = \dfrac{\|r_d^k\|}{1 + \|c\|} \leq \varepsilon$ (viab. dual)

- $$\frac{|c^t x^k - b^t y^k|}{1 + |c^t x^k| + |b^t y^k|} \leq \varepsilon$$ (brecha de dualidade).

Usamos a brecha de dualidade ao invés da complementaridade $X_k Z_k e \approx 0$ por questões de estabilidade numérica.

Note que o método já fornece $x^k \geq 0$ e $z^k \geq 0$.

Ponto inicial

Em tese, quaisquer $x^0 > 0$, y^0 e $z^0 > 0$ servem... Porém, na prática adota-se a seguinte heurística:

1) Calcula-se $\tilde{x} = A^t(AA^t)^{-1}b$. Note que $A\tilde{x} = b$, mas não necessariamente $\tilde{x} > 0$.

2) $x_j^0 = \max\{\tilde{x}_j, \varepsilon_1\}$ onde

$$\varepsilon_1 = \max\left\{-\min_j \tilde{x}_j,\ 100,\ \frac{\|b\|_1}{100\,\|A\|_1}\right\}.$$

3) $y^\circ = 0$

4) $$z^\circ_j = \begin{cases} c_j + \varepsilon_3 & , \ c_j \geq 0 \\ -c_j & , \ c_j \leq -\varepsilon_3 \\ \varepsilon_3 & , \ -\varepsilon_3 < c_j < 0 \end{cases}$$

"Questão computacional":

O passo caro é o cálculo de d_y^k. Devemos resolver $$(AD_kA^t)d_y = r_p^k + AD_k(r_d^k - X_k^{-1}r_c^k).$$

A matriz $m \times m$ $AD_k A^t$ é definida positiva pois $D_k = Z_k^{-1} X_k$ e $z^k > 0$, $x^k > 0$.

$\longrightarrow$ Usamos Cholesky!

- $AD_k A^t = G_k G_k^t$
- Resolva o sistema triangular inferior

$$G_k s = r_p^k + AD_k (r_d^k - X_k^{-1} r_c^k)$$

- Resolva o sistema triangular superior

$$G_k^t d_y = s.$$

<u>Problema com Cholesky</u>: mesmo que AD_xA^t seja esparsa, pode ocorrer de G ser cheia... Isso ocorre quando os elementos não nulos de AD_xA^t estão longe da diagonal.

<u>Solução</u>: reordenar linhas/colunas de AD_xA^t de modo a concentrar seus elementos não nulos ao redor da diagonal: $P^t(AD_xA^t)P$, P matriz de permutação.

Como encontrar uma boa P ?

- encontrar <u>a melhor</u> P é NP-difícil (esqueça!)
- <u>Heurísticas</u>. Uma delas é a "Approximate Minimum Degree" (AMD).

AMD é rápida e dá bons resultados. O comando "cholesky" do Julia implementa a AMD.

Cálculo de d_y^k usando uma permutação P:

1) aplique AMD à matriz AD_kA^t obtendo P_k

2) decomponha $P_k^t(AD_kA^t)P_k = G_kG_k^t$.

3) resolva $G_k s = P_k^t[r_p^k + AD_k(r_d^k - X_k^{-1}r_c^k)]$

4) resolva $G_k^t \tilde{d}_y = s$

5) calcule $d_y = P\tilde{d}_y$.

→ Verifique que este d_y é válido.

Comentários finais

- O método primal dual afim escala não é tão bom na prática. Seu defeito é que os produtos $x_i^k z_i^k$, $i=1,\dots,n$, ficam muito diferentes entre si.

- Solução: seguir um "caminho central" (método primal dual seguidor de caminhos). Este método pode ser visto como o pontos interiores com barreira logarítmica da programação não linear.

- Porém, a direção afim escala é boa, pois é a direção de Newton no sistema de otimalidade original.

- O método preditor-corretor de Mehrotra combina as duas direções: primeiro toma a direção afim escala; depois a "corrige" com a tal direção que aproxima o ponto da trajetória central → funciona bem e é estável!