

Solução de Problemas de Programação Linear

Melissa de Moraes Carvalho - RA118165

Disciplina MS777 - Projeto Supervisionado 1

Orientador: Professor Doutor Aurelio Ribeiro Leite de Oliveira

1 Resumo

Este projeto teve por objetivo fornecer uma base teórica em programação linear, com o intuito de subsidiar o trabalho com novos algoritmos que solucionam problemas lineares, como o algoritmo de ajustamento ótimo para p coordenadas.

Este algoritmo consiste numa generalização do algoritmo de ajustamento pelo par ótimo.

O algoritmo de ajustamento pelo par ótimo foi baseado no algoritmo de Von Neumann e desenvolvido por João Gonçalves, Robert Storer e Jacek Gondzio.

Os conceitos desenvolvidos no projeto incluem o estudo do método simplex como forma de se resolver os problemas primais de programação linear, desenvolvimento da teoria de dualidade lagrangiana, estudo dos métodos de pontos interiores para determinação da solução ótima primal-dual e estudo da resolução e condicionamento do sistema conhecido como “sistema aumentado”.

2 Introdução

Definição 1. Um problema de programação linear tem por objetivo a otimização de uma função linear que está sujeita a restrições lineares.

Um problema deste tipo está na forma padrão, se tem o seguinte formato:

$$\text{Determinar } x^* \text{ que minimiza } f(x) = c^T x = \sum_{i=1}^n c_i x_i,$$

sujeita às condições:

$$\begin{array}{rcl} a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n & = & b_1 \\ a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n & = & b_2 \\ \vdots & & \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \dots + a_{mn}x_n & = & b_m \end{array}$$

E também,

$$x_1 \geq 0, x_2 \geq 0, \dots, x_n \geq 0.$$

Terminologia

1. A função linear f a ser otimizada é chamada de *função objetivo*.
2. A matriz $A = [a_1 a_2 \dots a_n]_{m \times n}$ é chamada de *matriz de restrições*.
3. Um vetor $x \in \mathbb{R}^n$ tal que $x_i \geq 0, \forall i \in \{1, 2, \dots, n\}$ e que satisfaça o conjunto de restrições $Ax = b$ é uma solução *factível* do problema linear.

Todo problema de programação linear pode ser colocado na forma padrão por meio de operações elementares. Estas operações incluem:

(i) *Trabalhar-se com o oposto da função objetivo.* Esta operação transforma um problema de maximização em um de minimização, e reciprocamente, um de minimização em um de maximização. Basta, para isso, observar que $\max(f(x)) = \min(-f(x))$ e $\min f(x) = \max(-f(x))$.

(ii) *Determinação de variáveis não negativas auxiliares.* Se um problema apresenta uma variável livre x_j , a qual não tem como restrição ser não negativa, ela passa a ser escrita como $x_j = u_j - v_j$, sendo $u_j \geq 0$ e $v_j \geq 0$ para que assim, todas as variáveis tratadas sejam condicionadas a serem não negativas.

(iii) *Transformação de desigualdades em igualdades.* Dada uma restrição do tipo $a_j^T x \leq b$ sendo a_j a j -ésima linha de A , a desigualdade é convertida em uma igualdade criando-se a variável auxiliar z_j tal que $a_j^T x + z_j = b$ e adicionando-se ao conjunto de restrições a condição que $z_j \geq 0$. Analogamente, para uma desigualdade do tipo, $a_j^T x \geq b$, faz-se $a_j^T x - z_j = b$.

Feita esta observação, os conceitos de programação linear podem ser desenvolvidos a partir de um problema na forma padrão, sem se perder em generalidade, já que problemas mais gerais podem ser facilmente colocados no formato padrão.

3 O Método Simplex

Seja $A \in \mathbb{R}^{m \times n}$ a matriz de restrições de um problema de programação linear padrão.

Considere uma partição de A da seguinte forma: $A = [B \ N]$, em que $B \in \mathbb{R}^{m \times m}$ é invertível e $N \in \mathbb{R}^{m \times n-m}$.

Com esta partição,

$$Ax = b \Leftrightarrow Bx_B + Nx_N = b,$$

sendo que $x_B = (x_1, x_2, \dots, x_m)$ e $x_N = (x_{m+1}, \dots, x_n)$.

A solução $\hat{x}_N = 0$ e $\hat{x}_B = B^{-1}b$ é chamada de solução básica do problema linear. Se $\hat{x}_B \geq 0$, esta solução básica é uma solução básica factível, pois satisfaz todas as restrições do problema.

É possível ser demonstrado que:

1. Toda solução básica factível é um vértice da região factível.
2. Se um problema de PL padrão possui uma solução ótima, ele possui um vértice ótimo.

Logo, os vértices da região factível são candidatos à solução ótima do problema tratado, sendo intuitivo buscar a solução que o otimize dentro de um subconjunto dos vértices.

Esta estratégia é exatamente a utilizada no método conhecido como Método Simplex, o qual foi desenvolvido por George Dantzig e publicado em 1947.

A partir de uma solução básica factível (que é um vértice da região factível), o algoritmo testa se o vértice corrente é o ótimo. Se for, o algoritmo termina. Caso contrário, uma nova partição é feita na matriz A , trocando-se uma variável não básica ($x_i \in x_N$) com uma básica ($x_j \in x_B$).

Esta nova partição gera uma nova solução básica que é, novamente, submetida ao teste de otimalidade, e assim por diante.

Na prática, o método simplex apresenta convergência em tempo polinomial, quando aplicado a problemas reais.

4 O problema Lagrangiano

Seja P_1 um problema de programação linear padrão, cujo vetor b de restrições sofre uma pequena perturbação u , de modo que $Ax = b - u$. Este novo problema poderia ser tratado como qualquer outro problema de programação linear padrão.

Porém, ao invés disso, vamos definir um novo problema chamado de *problema lagrangiano*, o qual constituirá a base para o desenvolvimento da teoria de dualidade lagrangiana.

O problema lagrangiano visa a minimizar tanto a função objetivo do problema original P_1 quanto os custos gerados pela perturbação realizada.

Seja y_i o custo gerado pela perturbação do i -ésimo recurso, presente no problema P_1 , em u_i unidades.

Então, $\sum_{i=1}^m y_i u_i$ é o custo adicional a ser minimizado.

Com isso, o problema lagrangiano tem por objetivo:

$$\begin{aligned} \text{minimizar } L(x, y) &= c^T x + y^T u, \text{ sendo que} \\ x &\geq 0 \text{ e } u = b - Ax. \end{aligned}$$

A função objetivo do problema lagrangiano, $L(x, y)$, costuma ser chamada de *função lagrangiana*.

A função lagrangiana também pode ser expressa da seguinte forma:

$$\begin{aligned} L(x, y) &= c^T x + y^T u = c^T x + y^T (b - Ax) = \\ &= (c^T - y^T A)x + y^T b. \end{aligned}$$

Se denotarmos $A = [a_1 a_2 \dots a_n]_{m \times n}$ e $c = (c_1, c_2, \dots, c_n)$, temos que:

$$L(x, y) = (c_1 - y^T a_1)x_1 + (c_2 - y^T a_2)x_2 + \dots + (c_n - y^T a_n)x_n + y^T b.$$

Definição 2. A função $g : \mathbb{R}^m \mapsto \mathbb{R}$, que a cada $y \in \mathbb{R}^m$ associa o mínimo da função Lagrangiana restrita ao conjunto $\{x \in \mathbb{R}^n \mid x \geq 0\}$ é definida como função dual, isto é:

$$g(y) := \min_{x \geq 0} L(x, y).$$

De maneira mais explícita:

$$\begin{aligned} g(y) &= \min_{x \geq 0} L(x, u) = \min_{x \geq 0} \{(c_1 - y^T a_1)x_1 + \dots + (c_n - y^T a_n)x_n + y^T b\} = \\ &= \min_{x_1 \geq 0} \{c_1 - y^T a_1\}x_1 + \dots + \min_{x_n \geq 0} \{c_n - y^T a_n\}x_n + y^T b. \end{aligned}$$

Observe que se existe algum índice i para o qual $c_i - y^T a_i < 0$, conclui-se que

$$\min_{x_i \geq 0} \{(c_i - y^T a_i)x_i\} = -\infty$$

Logo, $g(y) = \sum_{i=1}^n \min_{x_i \geq 0} \{(c_i - y^T a_i)x_i\} + y^T b = -\infty$

Portanto, se tivermos por objetivo garantir que a função dual g tenha solução finita, temos como restrição que para todo $i \in \{1, 2, \dots, n\}$, $c_i - y^T a_i \geq 0$.

Vetorialmente, basta ser garantido que $(c_1, c_2, \dots, c_n) \geq (y^T a_1, y^T a_2, \dots, y^T a_n)$, ou seja:

$$A^T y \leq c.$$

Como consequência desta condição, $\min_{x_i \geq 0} \{(c_i - y^T a_i)x_i\} = 0$. Logo, a função dual passa a ser,

$$g(y) = y^T b.$$

Tais conceitos constituem a base da teoria de dualidade lagrangiana presente na próxima sessão.

5 Dualidade Lagrangiana

Considere o seguinte problema de programação linear na forma padrão:

$$\begin{aligned} & \text{minimizar } f(x) = c^T x \\ & \text{sujeita a } Ax = b, x \geq 0 \end{aligned}$$

Este problema passará a ser denotado como problema primal. O problema dual, a ele associado, é definido como:

$$\begin{aligned} & \text{maximizar } g(y) = y^T b \\ & \text{sujeita a } A^T y \leq c \end{aligned}$$

Todo vetor y que satisfaça as restrições duais é chamado de *solução dual factível*.

Note que a condição $A^T y \leq c$ é equivalente a afirmar que existe um vetor $s \in \mathbb{R}^n, s \geq 0$, tal que $A^T y + s = c$.

O vetor s é chamado de vetor das *folgas complementares* do problema dual.

Teorema 1. (Teorema da Dualidade Fraca) *Se $x \in \mathbb{R}^n$ é solução factível do problema primal e $y \in \mathbb{R}^m$ é solução factível do problema dual, então $g(y) \leq f(x)$.*

Demonstração. Sejam $R = \{x \in \mathbb{R}^n | Ax = b \text{ e } x \geq 0\}$ e $R' = \{x \in \mathbb{R}^n | x \geq 0\}$.

Pela definição dos conjuntos, temos que: $v \in R \Rightarrow v \in R'$, concluindo-se que: $R \subset R'$.

Portanto, o mínimo da função lagrangiana em R' deve ser menor ou igual ao seu mínimo tomado em R . Assim,

$$\begin{aligned} g(y) &= \min_{\{x \geq 0\}} L(x, y) = \min_{\{x \geq 0 = R'\}} c^T x + y^T (b - Ax) \leq \\ &\leq \min_{\{R\}} c^T x + y^T (b - b) = \min_{\{x \geq 0, Ax = b\}} c^T x = \min f(x) \leq f(x) \end{aligned}$$

□

Teorema 2. *Se \bar{x} e \bar{y} são soluções factíveis dos problemas primal e dual, respectivamente, e $f(\bar{x}) = g(\bar{y})$ então, \bar{x} e \bar{y} são as soluções ótimas dos seus problemas correspondentes.*

Demonstração. Suponha que \bar{x} não seja solução ótima primal. Então, existe \hat{x} tal que $c^T \hat{x} < c^T \bar{x} = \bar{y}^T b$, o que contradiz o teorema 1. Da mesma forma, supondo que \bar{y} não é solução ótima dual, existe um \hat{y} tal que $c^T \bar{x} = \bar{y}^T b < \hat{y}^T b$, o que novamente, contradiz o teorema 1.

Portanto, \bar{x} e \bar{y} são soluções ótimas, primal e dual, respectivamente.

□

Teorema 3. (Teorema da Dualidade Forte) *Se o problema primal tem uma solução ótima \bar{x} , então o problema dual também tem uma solução ótima \bar{y} tal que $f(\bar{x}) = g(\bar{y})$.*

Teorema 4. (Teorema das folgas complementares) *Sejam \bar{x} e \bar{y} soluções factíveis dos problemas primal e dual, respectivamente, ou seja, $A\bar{x} = b$ e $A^T \bar{y} \leq c$. Então, \bar{x} e \bar{y} são soluções ótimas, se e somente se,*

$$s_i \bar{x}_i = 0, \quad \forall i \in \{1, 2, \dots, n\} \tag{1}$$

Demonstração. (\Leftarrow) Supondo (1) válido, tem-se que:

$$\begin{aligned} 0 &= s_1 \bar{x}_1 + s_2 \bar{x}_2 + \dots + s_n \bar{x}_n = (s_1, s_2, \dots, s_n) \bar{x} = \\ &= (c_1 - \bar{y}^T a_1, c_2 - \bar{y}^T a_2, \dots, c_n - \bar{y}^T a_n) \bar{x} = \\ &= (c^T - \bar{y}^T A) \bar{x} \end{aligned}$$

Logo,

$$0 = (c^T - \bar{y}^T A) \bar{x} \implies f(\bar{x}) = c^T \bar{x} = \bar{y}^T A \bar{x} = \bar{y}^T b = g(\bar{y})$$

Daí, pelo teorema (2), conclui-se que \bar{x} e \bar{y} são soluções ótimas.

(\implies) Suponha que \bar{x} e \bar{y} sejam soluções ótimas primal e dual, respectivamente. Então, pelo teorema da dualidade forte, temos que $f(\bar{x}) = g(\bar{y})$. Daí,

$$(c^T - \bar{y}^T A)\bar{x} = 0 \implies \sum_{i=1}^n s_i \bar{x}_i = 0$$

Sabemos que $x_i \geq 0$ e $s_i \geq 0$. Logo, a igualdade anterior só é válida se,

$$s_i \bar{x}_i = 0, \forall i \in \{1, 2, \dots, n\}$$

□

O teorema anteriormente demonstrado é de grande importância na busca da solução ótima primal-dual.

As hipóteses do teorema das folgas complementares, podem ser resumidas no seguinte sistema:

$$Ax = b \tag{2}$$

$$A^T y + s = c \tag{3}$$

$$s_i x_i = 0, \quad i \in \{1, 2, \dots, n\} \tag{4}$$

$$(x, s) \geq 0$$

Neste sistema, as equações $Ax = b$ e $A^T y + s = c$, que garantem a factibilidade de x e y , são lineares e as equações $s_i x_i = 0$ que garantem a otimalidade são não lineares.

Portanto, a solução ótima dos problemas primal e dual é obtida por meio da resolução de um sistema não linear. Os métodos analisados para resolvê-lo foram os métodos de pontos interiores.

6 Métodos de Pontos Interiores

Os métodos de pontos interiores têm por objetivo determinar uma solução ótima primal-dual (x, y, s) do sistema não linear definido pelas equações de (2) a (4), com a condição adicional de que $(x, s) > 0$.

Esta propriedade faz com que a solução ótima seja buscada no interior do ortante positivo e é, justamente, esta característica que origina o nome do método.

A solução do sistema não linear é aproximada por meio do método de Newton com algumas variações. Estas variações são feitas para se garantir que a restrição $(x, s) > 0$ não seja violada em cada iteração e para se permitir que a solução ótima seja buscada através de direções mais eficientes do que a direção padrão do método de Newton.

O sistema não linear a ser resolvido define uma função não linear $F : \mathbb{R}^{2n+m} \mapsto \mathbb{R}^{2n+m}$ sobre a qual será aplicado o Método de Newton:

$$F(x, y, s) = \begin{bmatrix} A^T y + s - c \\ Ax - b \\ X S e \end{bmatrix} = \begin{bmatrix} r_c \\ r_b \\ r_{xs} \end{bmatrix}$$

em que, $X = \text{diag}(x_1, x_2, \dots, x_n)$ e $S = \text{diag}(s_1, s_2, \dots, s_n)$.

Aplicando-se o método de Newton à função F , deve-se resolver, em cada iteração, o seguinte sistema:

$$\begin{bmatrix} 0 & A^T & I \\ A & 0 & 0 \\ S & 0 & X \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \\ \Delta s \end{bmatrix} = \begin{bmatrix} -r_c \\ -r_b \\ -r_{xs} \end{bmatrix} \tag{5}$$

sendo que a matriz dos coeficientes é a jacobiana da função F .

Uma das variações do método clássico de Newton é a de que o vetor direção de Newton $(\Delta x, \Delta y, \Delta s)$ é multiplicado por uma constante $\alpha \in (0, 1)$. Devido a isso, a k -ésima aproximação para a solução ótima é dada por:

$$(x, y, s)_k = (x, y, s)_{k-1} + \alpha(\Delta x, \Delta y, \Delta s)$$

O parâmetro α é inserido, pois um passo completo na direção canônica de newton ($\alpha = 1$), geralmente, viola a condição de que $(x, s) > 0$.

O parâmetro α é obtido garantindo-se que:

$$x_i + \alpha\Delta x_i > 0, \quad \forall i \in \{1, 2, \dots, n\} \text{ e} \quad (6)$$

$$s_i + \alpha\Delta s_i > 0, \quad \forall i \in \{1, 2, \dots, n\} \quad (7)$$

Ou seja, precisamos garantir que $x_i > -\alpha\Delta x_i$ e $s_i > -\alpha\Delta s_i$.

Como α , x_i e s_i são ambos positivos, podemos dividir as desigualdades anteriores por αx_i e αs_i , respectivamente, sem que o sinal da desigualdade seja alterado. Assim, para todo $i \in \{1, 2, \dots, n\}$, é necessário que:

$$\frac{1}{\alpha} > \frac{-\Delta x_i}{x_i} \quad \text{e} \quad \frac{1}{\alpha} > \frac{-\Delta s_i}{s_i}$$

Logo, basta escolhermos α como:

$$\frac{1}{\alpha} = \max_{i,j} \left\{ \frac{-\Delta x_i}{x_i}, \frac{-\Delta s_j}{s_j} \right\}$$

Mas, como a igualdade anterior nem sempre garante a desigualdade estrita que é exigida em (6) e (7), introduz-se à definição de α um parâmetro r próximo de 1, e, portanto:

$$\alpha = r \left(\max_{i,j} \left\{ \frac{-\Delta x_i}{x_i}, \frac{-\Delta s_j}{s_j} \right\} \right)^{-1} \wedge 1$$

Em geral, para que a não negatividade de x e s seja garantida, α acaba sendo muito pequeno, o que faz com que um algoritmo com apenas esta modificação, não avance muito de uma iteração para a outra.

Com isso, tem-se como motivação alterar-se a direção do vetor $(\Delta x, \Delta y, \Delta s)$ mais para o interior do ortante positivo, para que na próxima iteração o passo feito consiga ser mais longo, sem se violar a condição sobre x e s .

Um bom caminho para orientar o vetor do passo é o caminho conhecido como “caminho central” definido a seguir.

O caminho central

O caminho central C é um caminho parametrizado por um escalar $\tau > 0$, sendo que cada ponto $(x_\tau, y_\tau, s_\tau) \in C$ é solução do seguinte sistema:

$$\begin{aligned} cA^T y + s - c &= 0 \\ Ax - b &= 0 \\ x_i s_i &= \tau \\ (x, s) &> 0 \end{aligned} \quad (8)$$

Em síntese, os pontos de C são soluções estritamente factíveis do problema primal-dual.

Observe que caso $\tau \rightarrow 0$, o sistema (8), que define o caminho central, se aproxima do sistema que determina a solução ótima. Assim, se $C = \{(x_\tau, y_\tau, s_\tau) | \tau > 0\}$ converge a um ponto, quando $\tau \rightarrow 0$, este ponto é a solução ótima.

Os passos tomados na direção do caminho central geralmente são maiores do que os passos na direção padrão de Newton, pois como $x_\tau s_\tau = \tau > 0$, os pontos pertencentes a C localizam-se mais

no interior do ortante positivo conseguindo-se o objetivo de ter maior avanço de uma iteração para a outra.

Devido a isto, muitos algoritmos primais duais visam a determinar passos nesta direção, ao invés da direção padrão de Newton.

Dado $(x_\tau, y_\tau, s_\tau) \in C$, a média (μ) dos produtos $x_i s_i$ vale exatamente τ , pois:

$$\mu = \frac{1}{n} \sum_{i=1}^n x_i s_i = \frac{n\tau}{n} = \tau$$

Logo, esta média é um dos indicativos do quanto a direção tomada está próxima do caminho central.

Além da centralidade, outro objetivo que deve ser buscado é a convergência a zero do parâmetro τ . Por isso, é necessário que o parâmetro τ seja ponderado por um fator que limite seu crescimento.

Feitas estas duas considerações, define-se a variável τ como:

$$\tau = \mu \cdot \sigma,$$

em que $\mu = \frac{1}{n} \sum_{i=1}^n x_i s_i$ e $\sigma \in (0, 1)$.

Algoritmo Primal Dual

O algoritmo primal dual consiste em se aplicar o método de Newton à seguinte função não linear, definida a partir do sistema (8) :

$$F(x, y, s) = \begin{bmatrix} A^T y + s - c \\ Ax - b \\ X S e - \mu \sigma e \end{bmatrix} \quad (9)$$

em que, $X = \text{diag}(x_1, x_2, \dots, x_n)$ e $S = \text{diag}(s_1, s_2, \dots, s_n)$.

O algoritmo é iniciado com um ponto factível $p_0 = (x^0, y^0, s^0)$, isto é $A^T y_0 + s_0 - c = 0$ e $Ax_0 - b = 0$. Daí, em cada iteração deve-se resolver o seguinte sistema:

$$\begin{bmatrix} 0 & A^T & I \\ A & 0 & 0 \\ S^k & 0 & X^k \end{bmatrix} \begin{bmatrix} \Delta x^k \\ \Delta y^k \\ \Delta s^k \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ -X^k S^k e + \sigma_k \mu_k e \end{bmatrix} \quad (10)$$

sendo que a matriz dos coeficientes é a jacobiana da função F .

Então, $(x, y, s)_{k+1} = (x, y, s)_k + \alpha(\Delta x^k, \Delta y^k, \Delta s^k)$, sendo α escolhido de modo a se garantir $(x, s) > 0$.

Outros algoritmos primais-duais consistem de modificações deste. A seguir apresentam-se alguns outros métodos semelhantes a este.

Métodos seguidores de Caminhos

Um algoritmo seguidor de caminhos restringe suas iterações a uma vizinhança do caminho central C . Estas vizinhanças são determinadas a partir de alguma norma.

As normas mais utilizadas são:

$$(i) N_2(\theta) = \{(x, y, s) \in F^0 \mid \|X S e - \mu e\|_2 \leq \theta \mu\}, \theta \in (0, 1)$$

(ii) $N_{-\infty}(\gamma) = \{(x, y, s) \in F^0 \mid x_i s_i \geq \gamma \mu, i \in \{1, 2, \dots, n\}\}, \gamma \in (0, 1)$, sendo F^0 o conjunto dos pontos estritamente factíveis.

Típicos valores utilizados são $\theta = 0,5$ e $\gamma = 10^{-3}$.

Observe que se γ é muito próximo de zero, a vizinhança $N_{-\infty}(\gamma)$ aproxima-se do conjunto factível, tornando-se pouco restritiva.

Em geral, a vizinhança $N_2(\theta)$ é mais restritiva, no sentido de que muitos pontos factíveis não pertencem a esta vizinhança.

Um método seguidor de caminhos que gera bons resultados é o método preditor-corretor descrito a seguir.

Método Preditor-Corretor

O método preditor corretor é um método seguidor de caminhos que utiliza duas vizinhanças do tipo N_2 , sendo uma delas interna à outra. As duas vizinhanças são utilizadas de forma alternada da seguinte forma: Em iterações pares utiliza-se a vizinhança interna e em iterações ímpares utiliza-se a vizinhança externa.

A partir da vizinhança interna é realizado um passo de newton com a direção afim escala até se atingir a borda da vizinhança externa. Ao se adotar a direção afim escala, o sistema resolvido nas iterações pares é o sistema não linear originado do teorema das folgas complementares, concluindo-se que o objetivo deste passo, chamado de passo preditor é o de favorecer a otimalidade do problema.

Na próxima iteração (ímpar) a partir da vizinhança externa, realiza-se um passo chamado de passo “corretor” com $(\sigma = 1)$, em direção, novamente, à vizinhança interna. A correção feita é no sentido de que com esta iteração visa-se a garantir maior centralidade, para que na próxima iteração o passo seja maior.

Assim, o algoritmo alterna entre os dois principais objetivos na busca da solução ótima: garantir-se as condições de otimalidade e as de centralidade.

O algoritmo em linguagem de alto nível está descrito a seguir.

Algoritmo Preditor Corretor

Dado $(x^0, y^0, s^0) \in N_2(0, 25)$;

Para $k = 0, 1, 2, \dots$

Se k é par: (passo preditor)

Resolva o sistema (10) com $\sigma_k = 0$ para obter $(\Delta x^k, \Delta y^k, \Delta s^k)$;

Escolha α_k o maior valor no intervalo $[0, 1]$ tal que,

$$(x^k(\alpha), y^k(\alpha), s^k(\alpha)) \in N_2(0, 5)$$

$$\text{Faça } (x^{k+1}, y^{k+1}, s^{k+1}) = (x^k(\alpha_k), y^k(\alpha_k), s^k(\alpha_k));$$

Se k é ímpar: (passo corretor)

Resolva o sistema (10) com $\sigma_k = 1$ para obter $(\Delta x^k, \Delta y^k, \Delta s^k)$;

$$\text{Faça } (x^{k+1}, y^{k+1}, s^{k+1}) = (x^k, y^k, s^k) + (\Delta x^k, \Delta y^k, \Delta s^k);$$

Fim do laço.

A seguir temos um método que se assemelha ao preditor-corretor, mas que apresenta algumas outras vantagens, o que o torna o mais eficiente dentre os analisados.

Método de Mehrotra Preditor Corretor

O algoritmo de Mehrotra se assemelha ao algoritmo preditor corretor no sentido de determinar, em cada iteração, um vetor de passo que é combinação de um vetor na direção afim escala (i), de um vetor na direção do caminho central (ii) e outro que visa a corrigir a não linearidade da função trabalhada (iii).

Uma das vantagens desse algoritmo é que o peso do passo na direção primal é diferente do peso na direção dual, pois, computacionalmente, mostra-se ser mais vantajoso fazer esta separação.

Os pesos α_{primal}^{aff} e α_{dual}^{aff} são escolhidos da seguinte forma:

$$\alpha_{primal}^{aff} = \max\{\alpha \in [0, 1] | x + \alpha \Delta x^{aff} \geq 0\}$$

$$\alpha_{dual}^{aff} = \max\{\alpha \in [0, 1] | s + \alpha \Delta s^{aff} \geq 0\}$$

e tanto α_{primal}^{aff} , quanto α_{dual}^{aff} são calculados como foi feito no algoritmo primal-dual.

Outra vantagem é que neste método a constante σ não é fixa ao longo da execução de todo o algoritmo, mas sim escolhida em cada iteração de forma a obter-se melhores resultados. Sua escolha é feita a partir de uma comparação entre a média μ dos produtos $x_i^k s_i^k$ com a média μ_{aff} dos produtos $x_i^{aff} s_i^{aff}$ do ponto obtido com a direção afim escala. De maneira mais explícita,

$$\mu_{aff} = \frac{(x + \alpha_{primal}^{aff} \Delta x^{aff})(s + \alpha_{dual}^{aff} \Delta s^{aff})}{n}$$

Se μ_{aff} é próximo de μ , significa que a direção afim escala manteve os produtos $x_i s_i$ aproximadamente constantes, ou seja, a direção afim escala move próxima ao caminho central, sendo preferível tomar σ próximo a 1 para que mantendo esta centralidade, seja possível na próxima iteração dar passos mais longos. Por outro lado, se $\mu_{aff} \ll \mu$, significa que a direção afim escala foi capaz de contribuir para uma grande redução dos produtos $x_i s_i$, favorecendo a convergência para a solução ótima, podendo também ser mantida através da escolha de σ próximo de zero.

Mehrotra sugeriu uma expressão fixa para σ a qual mostrou-se muito eficiente depois de muitos testes computacionais:

$$\sigma = \left(\frac{\mu_{aff}}{\mu} \right)^3$$

O vetor do tipo (i), como já se sabe, é encontrado resolvendo-se um sistema como em (5).

Já, o vetor do tipo (ii), que visa a corrigir a centralidade, poderia ser obtido resolvendo-se um sistema como em (10), o qual determina um vetor na direção do caminho central.

Porém, veremos que os vetores do tipo (ii) e (iii) não serão determinados separadamente, mas sim, será encontrado um único vetor que visa a obter os dois objetivos.

Como o método de Newton faz uma aproximação linear da função F , que é não linear nas n entradas definidas pelos produtos $x_i s_i$, após realizar-se o passo na direção afim escala, o produto $(x + \Delta x^{aff})(s + \Delta s^{aff})$ não é zero como prevê o modelo linear, sendo necessária uma correção, a qual é obtida resolvendo-se o seguinte sistema:

$$\begin{bmatrix} 0 & A^T & I \\ A & 0 & 0 \\ S^k & 0 & X^k \end{bmatrix} \begin{bmatrix} \Delta x^{cor} \\ \Delta y^{cor} \\ \Delta s^{cor} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ -\Delta X^{aff} \Delta Y^{aff} e \end{bmatrix} \quad (11)$$

Veja que este sistema determina um passo a mais na direção afim escala dentro da mesma iteração.

Ademais, observe que a matriz dos coeficientes em (11) que determina o vetor corretor do tipo (iii) é igual à matriz dos coeficientes que define o vetor na direção do caminho central (ii). Devido a isto, e ao fato de que tais direções são linearmente independentes, o algoritmo de Mehrotra calcula um único vetor a partir do seguinte sistema:

$$\begin{bmatrix} 0 & A^T & I \\ A & 0 & 0 \\ S^k & 0 & X^k \end{bmatrix} \begin{bmatrix} \Delta x^{cor} \\ \Delta y^{cor} \\ \Delta s^{cor} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \sigma \mu e - \Delta X^{aff} \Delta Y^{aff} e \end{bmatrix} \quad (12)$$

Por fim, o vetor do passo é dado pela soma entre o vetor na direção afim escala e o vetor corretor. A seguir tem-se o algoritmo explicitamente em linguagem de alto nível:

Algoritmo de Mehrotra Preditor Corretor

Dado (x_0, y_0, s_0) com $(x_0, s_0) > 0$;

Para $k = 0, 1, 2, \dots$

Resolva (5) para obter $(\Delta x^{aff}, \Delta y^{aff}, \Delta s^{aff})$

Calcule

$$\alpha_{primal}^{aff} = \max\{\alpha \in [0, 1] | x^k + \alpha \Delta x^{aff} \geq 0\}$$

$$\alpha_{dual}^{aff} = \max\{\alpha \in [0, 1] | s^k + \alpha \Delta s^{aff} \geq 0\}$$

Faça $\sigma = \left(\frac{\mu_{aff}}{\mu}\right)^3$

Resolva (12) para obter $(\Delta x^{cor}, \Delta y^{cor}, \Delta s^{cor})$

Determine o vetor do passo fazendo,

$$(\Delta x^k, \Delta y^k, \Delta s^k) = (\Delta x^{aff}, \Delta y^{aff}, \Delta s^{aff}) + (\Delta x^{cor}, \Delta y^{cor}, \Delta s^{cor})$$

$$\alpha_{max}^{primal} = \max\{\alpha \in [0, 1] | x^k + \alpha \Delta x^k \geq 0\}$$

$$\alpha_{max}^{dual} = \max\{\alpha \in [0, 1] | s^k + \alpha \Delta s^k \geq 0\}$$

$$\alpha_k^{primal} = \min(0.99\alpha_{max}^{primal}, 1) \text{ e } \alpha_k^{dual} = \min(0.99\alpha_{max}^{dual}, 1)$$

Faça:

$$x^{k+1} = x_k + \alpha_k^{primal} \Delta x^k \text{ e}$$

$$(y^{k+1}, s^{k+1}) = (y^k, s^k) + \alpha_k^{dual} (\Delta y^k, \Delta s^k)$$

Fim do laço

Determinação do Vetor Direção

Observando os métodos descritos, percebe-se que, independentemente das alterações feitas, todos os sistemas trabalhados tem a seguinte forma:

$$\begin{bmatrix} 0 & A^T & I \\ A & 0 & 0 \\ S & 0 & X \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \\ \Delta s \end{bmatrix} = \begin{bmatrix} r_1 \\ r_2 \\ r_3 \end{bmatrix}$$

Vamos, então, reorganizar as equações visando a obter a melhor forma de resolvê-lo. As equações obtidas a partir deste sistema são:

$$A^T \Delta y + \Delta s = r_1 \tag{13}$$

$$A \Delta x = r_2 \tag{14}$$

$$S \Delta X + X \Delta S = r_3 \tag{15}$$

em que $A \in \mathbb{R}^{m \times n}$, $\Delta x \in \mathbb{R}^n$, $\Delta s \in \mathbb{R}^n$ e $\Delta y \in \mathbb{R}^m$.

Na equação (15) temos que $\Delta s = X^{-1}(r_3 - S \Delta x)$. Substituindo na equação (13):

$$\begin{aligned} X^{-1}(r_3 - S \Delta x) + A^T \Delta y &= r_1 \Rightarrow \\ \Rightarrow -X^{-1} S \Delta x &= r_1 - A^T \Delta y - X^{-1} r_3 \end{aligned} \tag{16}$$

Esta última equação, juntamente com a equação (14) definem um sistema conhecido como Sistema Aumentado como mostrado a seguir:

$$\begin{bmatrix} -X^{-1} S & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix} = \begin{bmatrix} r_1 - X^{-1} r_3 \\ r_2 \end{bmatrix}$$

Observe que isolando Δx na equação (16) e multiplicando-a por A , tem-se que:

$$r_2 = A \Delta x = -A X^{-1} S \Delta x + A X^{-1} r_3 + A X^{-1} X A^T \Delta y$$

E, isolando Δy na equação anterior:

$$A X^{-1} X A^T \Delta y = r_2 + A X^{-1} S \Delta x - A X^{-1} r_3 \tag{17}$$

A matriz $A X^{-1} X A^T$ é conhecida como *complemento de Schur*.

Esta matriz por ser simétrica e definida positiva, admite fatoração de Cholesky, isto é, existe uma matriz quadrada G triangular inferior, tal que $A X^{-1} X A^T = G G^T$.

Assim, em problemas pequenos, a estratégia utilizada para se obter o vetor direção é a de se aplicar a fatoração de Cholesky ao complemento de Schur, fazendo com que o vetor Δy seja, facilmente, determinado através da resolução de dois sistemas triangulares inferiores.

No algoritmo de Mehrotra, o passo na direção afim escala é determinado a partir de um sistema em que $r_1 = c - A^T y - s$, $r_2 = -Ax + b$ e $r_3 = -XSe$.

A implementação dos dois sistemas triangulares obtidos após a fatoração de Cholesky, fica bem simples, pois as matrizes S e X são diagonais. Veja que,

$$\begin{aligned} r_2 + AS^{-1}Xr_1 - AS^{-1}r_3 &= (b - Ax) + AS^{-1}Xr_1 - AS^{-1}(-XSe) \\ &= b - Ax + AS^{-1}Xr_1 + Ax = b + AS^{-1}Xr_1 \end{aligned}$$

Além disso, se definirmos $r_1 = (r_{11}, r_{12}, \dots, r_{1n})$ e $AS^{-1}Xr_1 := AW$, conclui-se que:

$$W(i) = \frac{x(i)r_{1i}}{s(i)}, \quad i \in \{1, 2, \dots, n\}$$

Definindo: $G^T \Delta y := U$, concluímos que $GU = b - AW$, que são os dois sistemas triangulares a serem resolvidos.

Denotando A_i como a i -ésima coluna de A , o sistema triangular inferior $GU = b - AW$ tem a seguinte solução:

$$U(1) = \frac{b(1) + A_1 W}{G(1, 1)}$$

E, para $i = 2, 3, \dots, m$:

$$U(i) = \frac{b(i) + A_i W - \sum_{j=1}^{i-1} G(i, j)U(j)}{G(i, i)}$$

Obtido o vetor U , o sistema triangular superior $G^T \Delta y = U$ é resolvido da seguinte forma:

$$\Delta y^{aff}(m) = \frac{U(m)}{G(m, m)}$$

E, para $i = m - 1, m - 2, \dots, 1$:

$$\Delta y^{aff}(i) = \frac{U(i) - \sum_{j=i+1}^n G^T(i, j)\Delta y(j)}{G(i, i)}$$

Obtido Δy^{aff} , se definirmos $K := A^T \Delta y^{aff}$, temos que,

$$\Delta x^{aff}(i) = -W(i) - x(i) + \frac{k(i)x(i)}{s(i)}, \quad i \in \{1, 2, \dots, n\}$$

E, por fim,

$$\Delta s^{aff}(i) = -s(i) - \frac{s(i)Dx(i)}{x(i)}, \quad i \in \{1, 2, \dots, n\}$$

Para a obtenção do vetor corretor, temos que: $r_1 = r_2 = 0$ e $r_3 = -\Delta X^{aff} \Delta S^{aff} e$

Fazendo os mesmos cálculos, que os feitos para o passo na direção afim escala, conclui-se que o sistema que gera o vetor Δy^{cor} é dado por:

$$AS^{-1}XA^T \Delta y^{cor} = AS^{-1} \Delta X^{aff} \Delta S^{aff} e - AS^{-1} \sigma \mu e$$

Como a matriz dos coeficientes deste sistema é a mesma que a matriz dos coeficientes em (17), a fatoração de Cholesky só precisa ser calculada uma vez.

O vetor Δy^{cor} novamente é determinado pela solução de dois sistemas triangulares inferiores, e as implementações tanto de Δx^{cor} , quanto de Δs^{cor} são simples por serem determinadas em termos de matrizes diagonais.

$$\begin{aligned}\Delta x^{cor} &= -S^{-1}\Delta x^{aff}\Delta s^{aff}e + S^{-1}\sigma\mu e + S^{-1}XA^T\Delta y^{cor}e \\ \Delta s^{cor} &= X^{-1}(\sigma\mu e - \Delta x^{aff}\Delta s^{aff}e - S\Delta x^{cor})\end{aligned}$$

Apresentamos, enfim, um exemplo que visa a determinar uma solução ótima primal-dual via o Algoritmo de Mehrotra.

Exemplo: Considere o seguinte problema de programação linear:

$$\text{Minimizar } f(x) = c^T x = -6x_1 + 4x_2 - 4x_3 + 4x_4$$

Com as seguintes restrições:

$$\begin{bmatrix} 2 & 0 & 0 & -4 \\ -1 & 1 & -1 & 0 \\ 0 & 1 & -1 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} -4 \\ -1 \\ 0 \end{bmatrix}$$

Seu problema dual correspondente é:

$$\text{Maximizar } g(y) = b^T y = -4y_1 - y_2$$

Restrita a:

$$\begin{bmatrix} 2 & -1 & 0 \\ 0 & 1 & 1 \\ 0 & -1 & -1 \\ -4 & 0 & 0 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} -6 \\ 4 \\ -4 \\ 4 \end{bmatrix}$$

Implementando-se o algoritmo de Mehrotra e resolvendo-se os sistemas por meio da estratégia de se trabalhar com o complemento de Schur, obteve-se os seguintes resultados:

Iteração	X^T	Y^T	S^T
1	(1, 1.0134, 1.0134, 1.500)	(-0.9474, 4.5368, -0.5568)	(0.5216, 0.0100, 0.0100, 0.1504)
2	(1, 1.0134, 1.0134, 1.500)	(-0.9995, 4.0054, -0.0056)	(0.0052, 0.0001, 0.0001, 0.0015)
3	(1, 1.0134, 1.0134, 1.500)	(-1, 4.0001, -0.0001)	(0.5216, 0.0100, 0.0100, 0.1504) 10^{-4}
4	(1, 1.0134, 1.0134, 1.500)	(-1, 4, 0)	(0.5216, 0.0100, 0.0100, 0.1504) 10^{-6}
⋮	⋮	⋮	⋮
k	(1, 1.0134, 1.0134, 1.500)	(-1, 4, 0)	(0.5216, 0.0100, 0.0100, 0.1504) $10^{-2(k-1)}$

Tabela 1: **Iterações do Algoritmo de Mehrotra**

Observe que a partir da quarta iteração os valores de X e Y tornaram-se fixos e o vetor S foi decaindo com o fator de 10^{-2} . As soluções $X^* = (1, 1.0134, 1.0134, 1.500)^T$ e $Y^* = (-1, 4, 0)^T$ dos problemas primal e dual, respectivamente, são soluções factíveis por satisfazerem as restrições do problema dado.

Além disso, como

$$f(X^*) = g(Y^*) = 0,$$

pelo teorema (2), conclui-se que X^* e Y^* são as soluções ótimas.

Em problemas de grande porte, ao invés de se trabalhar com o complemento de Schur, trabalha-se com o sistema aumentado, determinando-se Δx e Δy iterativamente via método dos gradientes conjugados.

À medida que o método iterativo avança, a matriz $D = X^{-1}S$ fica muito mal condicionada, sendo necessário pré-condicionar o sistema aumentado.

O trabalho em aberto nesta área é o de se determinar bons pré-condicionadores ao longo da execução do algoritmo iterativo. Já existem bons pré-condicionadores tanto para o início, quanto para

o final da execução do método iterativo, mas para as iterações intermediárias, ainda não se conseguiu encontrar este bom pré-condicionador.

A técnica de pré-condicionamento, bem como alguns pré-condicionadores existentes para o sistema aumentado são apresentados na próxima sessão.

7 Pré-condicionadores

7.1 Introdução

Definição 3. *Seja $\|\cdot\|_v$ uma norma vetorial. Então, a norma matricial $\|\cdot\|_M$ induzida por $\|\cdot\|_v$ é definida por:*

$$\|A\|_M = \max_{X \neq \theta} \frac{\|AX\|}{\|X\|}.$$

Definição 4. *O número de condição de uma matriz $A \in \mathbb{R}^n$ é definido como*

$$K(A) = \|A\| \|A^{-1}\|,$$

sendo $\|\cdot\|$ qualquer norma matricial.

Observe que para qualquer norma induzida: $\|I\|_M = \max_{X \neq \theta} \frac{\|IX\|}{\|X\|} = 1$ e com isso,

$$K(A) = \|A\| \|A^{-1}\| \geq \|AA^{-1}\| = \|I\| = 1,$$

ou seja, para este tipo de norma, o número de condição de uma matriz é limitado inferiormente de 1.

Como as normas em \mathbb{R}^n são equivalentes, o número de condição de uma matriz é limitado inferiormente por 1 também em outras normas.

Uma matriz com um número de condição muito maior do que 1 é dita mal condicionada. Caso contrário, ela é dita bem condicionada.

Matrizes mal condicionadas ao sofrerem pequenas perturbações em seus coeficientes, podem gerar grande erro numérico na solução de sistemas lineares a elas associados.

Além disso, métodos iterativos envolvendo tais matrizes podem ter convergência muito lenta, ou nem chegam a convergir.

Com isso, ao invés de se resolver um problema associado a uma matriz mal condicionada, busque-se determinar uma transformação que gere uma matriz com uma maior estabilidade numérica. Esta técnica é conhecida como pré-condicionamento de matrizes.

Definição 5. *Uma matriz A é dita pré-condicionada se existem matrizes M e N de mesma ordem de A , tal que $M^{-1}AN^{-1}$ tem um melhor condicionamento do que A .*

Um bom pré-condicionador, além de melhorar o número de condição da matriz, deve contribuir com a redução do tempo e do custo computacional total associados ao problema tratado.

Esta redução pode ser conseguida se as matrizes M e N forem simples de serem determinadas e de serem manipuladas, como as matrizes "bloco diagonais".

Note que, sem estas restrições, poderíamos escolher $M = A$ e $N = I$ de modo que

$$K(M^{-1}AN^{-1}) = K(I) = 1 \leq K(A).$$

Porém, muitas vezes isto é inviável, já que determinar a inversa de uma matriz gera um alto custo computacional.

Com todas estas condições, vê-se que pré-condicionadores ideais são difíceis de serem encontrados.

Os pré-condicionadores também podem ser construídos de modo que a matriz pré-condicionada tenha outras propriedades mais vantajosas em relação à matriz original. Como por exemplo, pode-se desejar que a matriz pré-condicionada seja simétrica definida positiva, propriedade muito importante, pois permite a utilização do método dos gradientes conjugados.

Este objetivo não é conseguido com pré-condicionadores conhecidos como pré-condicionadores simétricos, que são pré-condicionadores tais que $N^T = M$, ou então, $N = (M^{-1})^T$.

A afirmação anterior é justificada pelo fato de que, pela lei da inércia de Sylvestre, a matriz original e a matriz pré-condicionada têm autovalores com os mesmos sinais.

7.2 Exemplos de Pré-condicionadores

Pré-condicionador obtido da fatoração incompleta de Cholesky

Seja A uma matriz simétrica definida positiva e seja G seu fator de Cholesky, tal que $A = GG^T$. A fatoração de Cholesky incompleta de A consiste em se determinar uma matriz L cuja diferença em relação ao fator de Cholesky G é a de que todas as entradas nulas de A terão suas respectivas entradas em L também nulas, isto é:

$$\begin{aligned} L_{i,j} &= G_{i,j}, & \text{se } A_{i,j} \neq 0 \\ L_{i,j} &= 0, & \text{caso contrário} \end{aligned}$$

As entradas não nulas de L , cujas entradas de mesmo índice em A são nulas, são chamadas de *fill-ins*.

Reduzir estes *fill ins* é desejável porque quanto mais elementos não nulos em L , maiores são os custos computacionais gerados, por exemplo, com o armazenamento e com a resolução dos sistemas associados à matriz $A = LL^T$.

A quantidade de *fill ins* depende da permutação a que a matriz A é submetida. Uma das técnicas para reduzir estes *fill-ins* é a técnica da "ordenação de mínimo grau".

Se A é uma matriz simétrica positiva definida, o pré-condicionador originado pela fatoração de Cholesky incompleta é definido da seguinte forma:

$$M = N^T = L,$$

ou seja, a matriz pré-condicionada é $L^{-1}AL^{-T}$.

Pré-condicionador SSOR

Seja A uma matriz simétrica decomposta da seguinte forma:

$$A = L + D + L^T,$$

sendo que $D = \text{diag}\{a_{11}, a_{22}, \dots, a_{nn}\}$ e $L_{ij} = A_{ij}$ para $i > j$ e $L_{ij} = 0$ caso contrário. O pré-condicionador SSOR é definido da seguinte forma:

$$\begin{aligned} N &= I \text{ e} \\ M &= (L + D)D^{-1}(L^T + D) = (LD^{-1} + I)(L^T + D) = \\ &= LD^{-1}L^T + L + L^T + D = A + LD^{-1}L^T. \end{aligned}$$

Este pré-condicionador é geralmente utilizado com matrizes definidas positivas, cujos elementos da diagonal principal são todos estritamente positivos.

Como problemas de grande porte trabalham com o sistema aumentado, vamos analisar uma classe de pré-condicionadores já existentes para tal sistema.

7.3 Classe de pré-condicionadores simétricos para o Sistema Aumentado

Um dos pré-condicionadores, que visa a reduzir o mal condicionamento da matriz $D = X^{-1}S$ à medida que o método dos gradientes conjugados avança é definido da seguinte forma:

$$M = N^T = \begin{pmatrix} H' & 0 \\ 0 & I \end{pmatrix}, \quad (18)$$

sendo que H' é uma matriz diagonal de dimensão n assim definida:

$$\begin{aligned} h'_{ii} &= \delta_{ii}^{-\frac{1}{2}} & \text{para as } n-m \text{ maiores entradas da matriz } D \text{ e} \\ h'_{ii} &= 1 & \text{caso contrário.} \end{aligned}$$

Existe uma grande classe de pré-condicionadores simétricos que possuem a seguinte forma:

$$\begin{pmatrix} H & 0 \\ F & G \end{pmatrix} \begin{pmatrix} -D & A^T \\ A & E \end{pmatrix} \begin{pmatrix} H^T & F^T \\ 0 & G^T \end{pmatrix} = \begin{pmatrix} -HDH^T & B^T \\ B & C \end{pmatrix},$$

em que $B = -FDH^T + GAH^T$ e $C = -FDF^T + FA^TG^T + GAF^T + GEG^T$.

O pré-condicionador anteriormente descrito pertence à esta classe. Para ver isto, basta tomar-se $H \equiv H'$, $F \equiv 0$ e $G \equiv I$.

Justificativas para a utilização do Sistema Aumentado

Ao analisar-se o modo simples como o problema primal-dual de pequeno porte foi tratado, por meio da fatoração de Cholesky do complemento de Schur, vê-se necessário justificar o porquê de não se utilizar tal técnica para problemas de grande porte.

Uma das razões é que em problemas de grande porte o complemento de Schur geralmente é esparsa e ao aplicar-se a fatoração de Cholesky, esta esparsidade é mudada, fazendo com que o complemento de Schur torne-se mais denso, implicando em cálculos mais custosos.

Além disso, como o tempo computacional exigido para se pré-condicionar o sistema aumentado é elevado, não é vantajoso gerar ainda mais custo para reduzir o sistema aumentado ao sistema positivo definido do complemento de Schur.

Por fim, métodos empíricos mostram, que ao se optar por resolver o sistema definido pelo complemento de Schur, muitas propriedades mais facilmente perceptíveis no sistema aumentado são perdidas, deixando de se entender o problema por completo.

Conclusão

Através da realização deste projeto foi possível adquirir um bom embasamento teórico na área de programação linear.

A metodologia utilizada foi a de um estudo dirigido que incluiu desde os conceitos mais básicos de programação linear, até atingir-se a parte teórica na qual se baseia o problema de se determinar um bom pré-condicionador para o sistema aumentado à medida que o método dos gradientes conjugados avança.

Com este problema em aberto, espera-se que, combinando o embasamento teórico adquirido com métodos empíricos, seja possível, um dia, resolvê-lo eficientemente.

Referências Bibliográficas

1. A.R.L. Oliveira, A New Class of Preconditioners for Large-Scale Linear Systems from Interior Point Methods for Linear Programming, Rice University, 1997
2. R. J. Vanderbei, Linear Programming - Foundations and Extensions, Springer.
3. S.J. Wright, Primal-Dual Interior-Point Methods, Siam, 1997.