**O PROBLEMA DE APROXIMAÇÃO LINEAR
NO L_1 E EXTENSÕES**

Márcia Aparecida Zanolli Meira e Silva

Orientador: Prof. Dr. Marcos Nereu Arenales

Dissertação apresentada ao Instituto de Ciências Matemáticas de São Carlos, da Universidade de São Paulo, para obtenção do Título de "Mestre em Ciências - Área: Ciências de Computação e Matemática Computacional".

São Carlos - SP

Dez/1994

**Aos meus pais, Fausto e Otília,
e ao meu marido, Júlio,
com carinho.**

AGRADECIMENTOS

Ao Marcos, pela orientação e amizade.

À Denise Pizarro Vieira Sato, pela amizade e incentivo dispensados nos momentos mais importantes.

Às amigas, Ana Cristina dos Santos, Andréa Carla Gonçalves Vianna e Edméa Cássia Baptista, pela ajuda e amizade.

À minha família, pelo apoio, compreensão e incentivo.

Em especial ao meu marido, pela paciência e compreensão nos momentos difíceis, e pelo apoio e incentivo constantes, sem os quais esse trabalho não seria concluído.

À FAPESP e a UNESP, pela parcial ajuda financeira.

RESUMO

Este trabalho apresenta uma especialização do Método Primal Simplex para resolver o Problema de Aproximação Linear no L_1 e o Problema de Regressão Quantil, os quais são casos particulares de Problema de Programação Linear por Partes.

No Problema de Regressão Quantil a função objetivo linear por partes depende de um parâmetro θ e, com pequenas adaptações da pós otimização clássica da Programação Linear, pode-se determinar o intervalo para θ onde a solução do problema fica invariante. Assim, este trabalho apresenta também uma maneira simples para realizar esta análise pós otimização.

Além disso, este trabalho apresenta alguns resultados computacionais, utilizando-se de exemplos da literatura.

ABSTRACT

This work presents a specialized of the Primal Simplex Method in order to solve the Least Absolute Approximation Problem as well as the Regression Quantile Problem, which are particular instances of Piecewise Linear Programming Problem.

The piecewise linear objective function of the Regression Quantile Problem is defined using a parameter θ and slight modification in the classic post-optimality analyses of Linear Programming can be obtained in order to determine the interval for θ where the solution remains invariant. This work presents a simple manner to get this interval.

Furthermore, this work illustrates some computational performance of the implemented method using examples from the literature.

ÍNDICE

Capítulo 1 Introdução	01
Capítulo 2 Propriedades de Factibilidade	09
2.1 Preliminares	09
2.1.1 Direções Factíveis	10
2.1.2 Solução Geral do Sistema $X\beta + \varepsilon = y$	12
2.1.3 Direções Simplex	16
2.1.4 Solução Básica do Sistema $X\beta + \varepsilon = y$	19
Capítulo 3 Propriedades de Otimalidade	20
3.1 Aproximação Linear no L_1	20
3.1.1 Direções de Descida	20
3.1.2 Solução Ótima	23
3.2 Regressão Quantil	26
3.2.1 Direções de Descida	26
3.2.2 Solução Ótima	28
Capítulo 4 Método Primal Simplex	31
4.1 Método Primal Simplex Aplicado ao Problema de Aproximação Linear no L_1	32
4.1.1 Iniciação	32
4.1.2 Teste de Otimalidade	32
4.1.3 Escolha do Ponto que Deixará de ser Interpolado	33
4.1.4 Determinação do Tamanho do Passo	36
4.1.5 Escolha do Ponto a ser Interpolado	42

4.1.6 Caso Degenerado	43
4.1.7 Algoritmo	44
4.2 Método Primal Simplex Aplicado ao Problema de Regressão Quantil	46
4.2.1 Iniciação	46
4.2.2 Teste de Otimalidade	46
4.2.3 Escolha do Ponto que Deixará de ser Interpolado	47
4.2.4 Determinação do Tamanho do Passo	47
4.2.5 Escolha do Ponto a ser Interpolado	51
4.2.6 Algoritmo	52
 Capítulo 5 Análise Pós Otimização	 54
 Capítulo 6 Resultados Computacionais	 61
6.1 Exemplo 1	61
6.2 Exemplo 2	65
6.3 Exemplo de Pós Otimização	66
 Conclusões	 68
 Bibliografia	 69

CAPÍTULO 1

INTRODUÇÃO

É muito comum em Programação Linear problemas com estruturas particulares. A simples aplicação de um método geral de Programação Linear normalmente é muito dispendiosa. Nestes casos a exploração da estrutura particular leva a grandes economias, tanto de memória como de tempo de processamento, pois evita-se armazenar e operar com zeros.

Neste trabalho, estudamos o Problema de Aproximação Linear no L_1 também chamado de Problema de Minimização dos Desvios Absoluto Mínimo, e o Problema de Regressão Quantil, explorando o fato da função ser linear por partes.

Como motivação, suponha que temos m pares de observações (x_i, y_i) , $i = 1, 2, \dots, m$. Podemos traçar esses pontos e tentar ajustar a eles uma função, dentre uma família de funções pré estabelecidas (por exemplo, funções lineares, quadráticas, etc.), tal que os pontos estejam "o mais próximo possível" da função (veja figura abaixo).

Mesmo que as variáveis x e y tivessem uma relação exata, os pontos traçados não satisfariam esta relação, devido aos erros de medidas.

Geralmente o tipo de função a ser ajustada é sugerida por evidências empíricas ou argumentos teóricos. Sugerimos um modelo, e então, devemos analisá-lo para ver se está adequado. Este é um problema de estatística e não será objeto deste estudo. Nos limitaremos ao problema de determinar o melhor ajuste, uma vez definida a família de funções.

Suponhamos que a relação entre as duas variáveis x e y seja linear, isto é,

$$y = \beta_1 + \beta_0 X + \varepsilon \quad (1.1)$$

tal relação é conhecida como uma "regressão linear simples" de y em x , onde ε é o erro de medida (veja figura abaixo).

Para os valores observados temos:

$$y_i = \beta_1 + \beta_0 X_i + \varepsilon_i \quad i = 1, \dots, m. \quad (1.2)$$

Queremos determinar β_0 e β_1 tal que

$$\sum_{i=1}^m |y_i - \beta_1 - \beta_0 X_i| = \sum_{i=1}^m |\epsilon_i| = \|\epsilon\|_1 \quad (1.3)$$

seja minimizado. A expressão (1.3) é chamada de "soma dos desvios absoluto".

Observe que, se o critério usado em (1.3) fosse a soma dos quadrados dos desvios, teríamos o clássico problema de "Mínimos Quadrados".

Em notação matricial, podemos escrever (1.2) como:

$$y = X\beta + \epsilon \quad (1.4)$$

onde

$$\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}, \quad X = \begin{bmatrix} X_1 & 1 \\ \vdots & \vdots \\ X_m & 1 \end{bmatrix} \quad \text{e} \quad \epsilon = \begin{pmatrix} \epsilon_1 \\ \vdots \\ \epsilon_m \end{pmatrix}.$$

Assim, o problema de ajustar uma reta usando o critério (1.3) consiste em determinar β tal que:

$$\begin{aligned} \min \|\epsilon\|_1 &= \min \left(\sum_{i=1}^m |\epsilon_i| \right) \\ \text{s.a. } X\beta + \epsilon &= y. \end{aligned} \quad (1.5)$$

De uma maneira geral, o Problema de Aproximação Linear no L_1 é descrito por (1.5) onde:

- $X \in \mathcal{R}^{m \times n}$, denota uma matriz obtida por um conjunto de m medidas de observações em n variáveis independentes;
- $\beta \in \mathcal{R}^n$, denota o vetor de parâmetros do modelo;
- $y \in \mathcal{R}^m$, denota as medidas na variável dependente;
- $\epsilon \in \mathcal{R}^m$, denota os erros aleatórios (desvios ou resíduos).

Assumimos por hipótese que $n < m$ e $\text{posto}(\mathbf{X}) = n$.

De acordo com Eisenhart (1961), (veja [9]), o Problema da Aproximação Linear no L_1 foi sugerido por Boscovitch na metade do século XVIII antes da introdução do Método dos Mínimos Quadrados, porém não foi muito usado devido ao desconhecimento de algoritmos eficientes para problemas com vários parâmetros.

Mais de 100 anos depois, Edgeworth (1887) propôs o ajuste de curvas e de modelos mais complicados pela minimização irrestrita de (1.5). Entretanto, os cálculos são mais complexos que a solução das equações lineares que resultam do Método dos Mínimos Quadrados.

Conforme Bloomfield e Steiger [9], um novo método foi introduzido por Rhodes (1930) e discutido por Singleton (1940). O desenvolvimento de Programação Linear e a observação de Harris (1950), de que o Problema de Aproximação Linear no L_1 pode ser transformado em um Problema de Programação Linear, um grande avanço. Esta linha foi seguida por Wagner [13] e muitos outros, incluindo Barrodale e Roberts [5] Narula e Wellington (1977), Abdelmalek [2] e Robers e Ben-Israel (citado também em [4]).

Embora com motivações diferentes, estes métodos são equivalentes [4].

Em 1959, Wagner [13] propôs a transformação: Defina $\beta = \beta^+ - \beta^-$ e $\varepsilon = \varepsilon^+ - \varepsilon^-$ com β^+ , β^- , ε^+ e ε^- todos não negativos, então o problema (1.5) é equivalente ao seguinte problema padrão de Programação Linear:

$$\min \sum_{i=1}^m (\varepsilon_i^+ + \varepsilon_i^-)$$

$$\text{s.a. } \mathbf{X}\beta^+ - \mathbf{X}\beta^- + \varepsilon^+ - \varepsilon^- = \mathbf{y}$$

$$\begin{aligned} \beta^+ &\geq 0, \beta^- \geq 0, \\ \varepsilon^+ &\geq 0, \varepsilon^- \geq 0. \end{aligned} \tag{1.6}$$

O artigo pioneiro de Barrodale e Roberts é baseado no quadro simplex explorando as particularidades do problema.

Bloomfield e Steiger desenvolve o método simplex, porém sem o uso do quadro, explorando também as particularidades do problema. Esta abordagem, com ligeiras modificações será adotada neste trabalho. Já Abdelmalek e Robers e Ben-Israel aplicam o método dual simplex ao problema dual.

Outros autores têm usado o método simplex revisado, podendo ainda acrescentar a decomposição L.U..

Embora o desempenho computacional possa diferir, tais métodos geram sempre a mesma sequência de pontos, exceção feita ao método de Bloomfield e Steiger que tem uma regra diferente para escolha da variável a entrar na base (novo ponto a ser interpolado), e possuem como essência o método simplex.

Ainda de acordo com Bloomfield e Steiger, Bowditch em uma tradução inglesa de um trabalho de Laplace observou que o Problema da Aproximação Linear no L_1 é menos sensível a erros extremos que o ajuste por Mínimos Quadrados. Observações semelhantes foram feitas por Edgeworth e por Rhodes (1930). Assim, o conceito simplista de estimar desvios absoluto mínimo, associado ao desenvolvimento de técnicas com custo computacional competitivo torna o problema consideravelmente importante.

Entretanto, a aplicação direta do Método Simplex ao problema (1.6) não leva a um algoritmo eficiente. Barrodale e Young [7] exploraram algumas particularidades do problema (1.6): i) uma solução factível básica está sempre disponível, sendo desnecessário a fase I; ii) as colunas de β_i^+ e β_i^- , (ϵ_i^+ e ϵ_i^-) diferem apenas do sinal, sendo desnecessário o armazenamento e atualização de todas as colunas do quadro Simplex, levando a uma grande economia de memória. Barrodale e Roberts [6] recomendaram que nas n primeiras iterações as variáveis β_i^+ ou β_i^- fossem escolhidas para constituir a base (isto leva a interpolação de n pontos, que é propriedade de uma solução ótima). Apesar destas observações representarem economia de memória, o número de iterações do Método Simplex poderia ser reduzido ainda mais.

Finalmente em 1973, Barrodale e Roberts [5] conseguiram reduzir o número de iterações pelas seguintes observações: i) quando uma variável básica β_i^+ atinge o nível zero, pode-se mudar a variável básica β_i^+ por β_i^- (o mesmo vale para uma variável básica β_i^-) e não há "pivotamento" - isto decorre do fato de β_i ser irrestrita de sinal, portanto não limita o crescimento de qualquer variável não básica; ii) quando uma variável básica ε_i^+ (ε_i^-) atinge o nível zero, pode-se trocar ε_i^+ por ε_i^- (ε_i^- por ε_i^+), porém o custo relativo da variável não básica se altera, isto decorre do fato de ε_i ser irrestrita de sinal, porém com valor crítico na origem, isto é, há uma mudança no "custo" de ε_i quando muda de sinal.

Neste trabalho estudamos também o Problema de Regressão Quantil.

Regressões Quantis são introduzidas em Koenker & Bassett [10] como quantidades de interesse no desenvolvimento de procedimentos de estimativas robustas.

As Regressões Quantis são definidas como soluções $\beta(\theta) \in \mathcal{R}^n$ do problema:

$$\min \left(\sum_{i/\varepsilon_i \leq 0} (1-\theta)(\varepsilon_i) + \sum_{i/\varepsilon_i \geq 0} \theta(\varepsilon_i) \right)$$

$$\text{s.a. } X\beta + \varepsilon = y$$

também escrito como

$$\min \left(\sum_{i \in I_0} \rho_\theta(\varepsilon_i) + \sum_{i \in I_1} \sigma_i(\varepsilon_i) \right)$$

$$\text{s.a. } X\beta + \varepsilon = y.$$

com

$$I_0 = \{i / \hat{\varepsilon}_i = 0\}$$

$$I_1 = \{i / \hat{\varepsilon}_i \neq 0\}$$

$$\rho_\theta \in [\theta - 1, \theta]$$

$$\sigma_i = \begin{cases} \theta & \text{se } i \in I^+ = \{i / \hat{\varepsilon}_i > 0\} \\ (\theta - 1) & \text{se } i \in I^- = \{i / \hat{\varepsilon}_i < 0\} \end{cases}$$

ou ainda

$$\begin{aligned} \min_{\beta} f_\theta(\beta, \varepsilon) &= \min \left(\sum_{i=1}^m f_\theta^i(\varepsilon_i) \right) \\ \text{s.a. } X\beta + \varepsilon &= y \end{aligned} \quad (1.7)$$

onde

$$X \in \mathcal{R}^{m \times n}, \text{ posto}(X) = n$$

$$\theta \in (0, 1)$$

f_θ^i é uma função linear por partes ilustrada no gráfico abaixo:

Observe que para $\theta = 1/2$ temos o Problema de Aproximação Linear no L_1 , pois $f_{1/2}(\beta, \epsilon) = \sum_{i=1}^m f_{1/2}^i(\epsilon_i) = \frac{1}{2} \sum_{i=1}^m |\epsilon_i|$ e o fator $1/2$ não interfere na minimização.

O objetivo principal deste trabalho é o estudo do Problema de Aproximação Linear no L_1 e do Problema de Regressão Quantil, com o intuito de determinar a solução numérica desses problemas, utilizando para isso uma especialização do Método Primal Simplex da Programação Linear.

Portanto, desenvolvemos todos os passos da especialização do Método Primal Simplex aplicados aos problemas (1.5) e (1.7).

Assim, no Capítulo 2 são desenvolvidas as propriedades de factibilidade para os problemas em estudo. Tais propriedades são as mesmas para ambos os problemas, uma vez que têm o mesmo conjunto de restrições.

No Capítulo 3, mostramos as propriedades de otimalidade para (1.5) e (1.7).

Baseado na teoria apresentada nos Capítulos 2 e 3 desenvolvemos o Método Primal Simplex para os problemas em estudo. Este desenvolvimento é feito no Capítulo 4, onde também apresentamos seus respectivos algoritmos.

De acordo com Osborne [11], as soluções quantis podem ser calculadas combinando Programação Linear com técnicas de pós otimização. Assim, no Capítulo 5, fazemos um estudo detalhado da análise pós otimização.

Finalmente, no Capítulo 6, apresentamos alguns resultados computacionais obtidos pelos métodos aqui apresentados. Desenvolvemos também um exemplo de pós otimização utilizando a teoria apresentada no Capítulo 5.

CAPÍTULO 2

PROPRIEDADES DE FACTIBILIDADE

Neste Capítulo, detalhamos propriedades de factibilidade, do Problema de Aproximação Linear no L_1 bem como do Problema de Regressão Quantil, salientando direcções factíveis, direcções simplex e soluções básicas.

2.1 PRELIMINARES

Introduzimos aqui propriedades de factibilidade que serão utilizadas no desenvolvimento dos métodos do Capítulo 4.

Considere o problema de Aproximação Linear no L_1

$$\begin{aligned} \min \quad & \|\varepsilon\|_1 \\ \text{s.a.} \quad & X\beta + \varepsilon = y \end{aligned} \tag{2.1.a}$$

e o problema de Regressão Quantil

$$\begin{aligned} \min \quad & \left(\sum_{i \in I_0} \rho_{\theta}(\varepsilon_i) + \sum_{i \in I_1} \sigma_i(\varepsilon_i) \right) \\ \text{s.a.} \quad & X\beta + \varepsilon = y. \end{aligned} \tag{2.1.b}$$

onde

$$I_0 = \{i / \hat{\varepsilon}_i = 0\}$$

$$I_1 = \{i / \hat{\varepsilon}_i \neq 0\}$$

$$\rho_\theta \in [\theta - 1, \theta]$$

$$\sigma_i = \begin{cases} \theta & \text{se } i \in I^+ = \{i / \hat{\varepsilon}_i > 0\} \\ (\theta - 1) & \text{se } i \in I^- = \{i / \hat{\varepsilon}_i < 0\} \end{cases}$$

$$\theta \in (0, 1).$$

Consideremos aqui apenas a restrição dos problemas:

$$X\beta + \varepsilon = y \quad (2.2)$$

com $X \in \mathcal{R}^{m \times n}$, $\text{posto}(X) = n$ e $n < m$.

2.1.1 DIREÇÕES FACTÍVEIS

Seja $(\hat{\beta}, \hat{\varepsilon})$ uma solução factível, isto é, $(\hat{\beta}, \hat{\varepsilon})$ satisfaz (2.2).

Uma estratégia usual de métodos numéricos consiste em perturbar $(\hat{\beta}, \hat{\varepsilon})$ e obter uma nova solução também factível.

Essa nova solução é, então, obtida da seguinte forma:

$$\begin{pmatrix} \beta \\ \varepsilon \end{pmatrix} = \begin{pmatrix} \hat{\beta} \\ \hat{\varepsilon} \end{pmatrix} + \delta \begin{pmatrix} d_\beta \\ d_\varepsilon \end{pmatrix}, \quad \delta > 0 \quad (2.3)$$

com $d_\beta \in \mathcal{R}^n$ e $d_\varepsilon \in \mathcal{R}^m$.

Dizemos que $(\hat{\beta}, \hat{\varepsilon})$ é perturbada na "direção $(d_{\beta}, d_{\varepsilon})$ " e δ é o "passo" adotado nesta direção.

Uma direção $(d_{\beta}, d_{\varepsilon})$ é chamada "direção factível" em $(\hat{\beta}, \hat{\varepsilon})$, se existe $\bar{\delta} > 0$ tal que

$$\begin{pmatrix} \beta \\ \varepsilon \end{pmatrix} = \begin{pmatrix} \hat{\beta} \\ \hat{\varepsilon} \end{pmatrix} + \delta \begin{pmatrix} d_{\beta} \\ d_{\varepsilon} \end{pmatrix}, \quad \forall \delta \in [0, \bar{\delta}]$$

seja factível, ou seja, a restrição (2.2) deve ser verificada.

Desta forma, para que $(d_{\beta}, d_{\varepsilon})$ seja direção factível devemos ter

$$\begin{aligned} X\beta + \varepsilon &= X(\hat{\beta} + \delta d_{\beta}) + (\hat{\varepsilon} + \delta d_{\varepsilon}) \\ &= (X\hat{\beta} + \hat{\varepsilon}) + \delta(Xd_{\beta} + d_{\varepsilon}) \\ &= y. \end{aligned}$$

Como $X\hat{\beta} + \hat{\varepsilon} = y$ e $\delta > 0$ temos,

$$Xd_{\beta} + d_{\varepsilon} = 0$$

$$(X, I) \begin{pmatrix} d_{\beta} \\ d_{\varepsilon} \end{pmatrix} = 0$$

$$(X, I)d = 0.$$

Assim, uma direção factível $d^t = (d_{\beta}, d_{\varepsilon})^t$ deve pertencer ao núcleo de (X, I) que é denotado por $N(X, I)$. A recíproca é facilmente verificada.

Assim, temos provado o seguinte:

Teorema 2.1: Seja $(\hat{\beta}, \hat{\varepsilon})$ uma solução factível. Uma direcção $d^t = (d_\beta, d_\varepsilon)^t$ é factível em $(\hat{\beta}, \hat{\varepsilon})$ se e somente se $d \in N(X, I)$.

2.1.2 SOLUÇÃO GERAL DO SISTEMA $X\beta + \varepsilon = y$

Nesta secção estudaremos as soluções gerais para o sistema

$$X\beta + \varepsilon = y.$$

Teorema 2.2: Seja $(\hat{\beta}, \hat{\varepsilon})$ uma solução do sistema $X\beta + \varepsilon = y$. Então, (β, ε) com $\beta \in \mathcal{R}^n$ e $\varepsilon \in \mathcal{R}^m$ satisfaz $X\beta + \varepsilon = y$ se e somente se

$$\begin{pmatrix} \beta \\ \varepsilon \end{pmatrix} = \begin{pmatrix} \hat{\beta} \\ \hat{\varepsilon} \end{pmatrix} + \delta \begin{pmatrix} d_\beta \\ d_\varepsilon \end{pmatrix}$$

onde $d \in N(X, I)$, com $d = \begin{pmatrix} d_\beta \\ d_\varepsilon \end{pmatrix}$ e $\delta > 0$.

Prova:

($\Rightarrow$) Por hipótese temos que $X\hat{\beta} + \hat{\varepsilon} = y$ e $X\beta + \varepsilon = y$.

Seja $d = \begin{pmatrix} \beta \\ \varepsilon \end{pmatrix} - \begin{pmatrix} \hat{\beta} \\ \hat{\varepsilon} \end{pmatrix}$. Logo,

$$X\beta + \varepsilon = X\hat{\beta} + \hat{\varepsilon}$$

$$X(\beta - \hat{\beta}) + (\varepsilon - \hat{\varepsilon}) = 0$$

$$Xd_\beta + d_\varepsilon = 0.$$

Portanto, $d \in N(X, I)$.

($\Leftarrow$) Temos agora que $d \in N(X, I)$, isto é, $Xd_p + d_\varepsilon = 0$, e

$$\begin{pmatrix} \beta \\ \varepsilon \end{pmatrix} = \begin{pmatrix} \hat{\beta} \\ \hat{\varepsilon} \end{pmatrix} + \delta \begin{pmatrix} d_p \\ d_\varepsilon \end{pmatrix}$$

e queremos mostrar que $X\beta + \varepsilon = y$.

$$\begin{aligned} X\beta + \varepsilon &= X(\hat{\beta} + \delta d_p) + (\hat{\varepsilon} + \delta d_\varepsilon) \\ &= (X\hat{\beta} + \hat{\varepsilon}) + \delta (Xd_p + d_\varepsilon) \\ &= X\hat{\beta} + \hat{\varepsilon} \\ &= y. \end{aligned}$$

Portanto, $X\beta + \varepsilon = y$. ■

Podemos explicitar todas as soluções do sistema $X\beta + \varepsilon = y$ como descreveremos abaixo.

Temos por hipótese que $\text{posto}(X) = n$. Assim, podemos selecionar n linhas linearmente independentes da matriz X , obtendo uma matriz $B \in \mathcal{R}^{n \times n}$ (note que B é inversível) e seja $N \in \mathcal{R}^{(m-n) \times n}$ a matriz formada pelas linhas restantes de X . Podemos escrever X como

$$X = \begin{pmatrix} B \\ N \end{pmatrix}. \quad (2.4)$$

A partição acima na matriz X é chamada "partição básica", a matriz B de "matriz básica" e a matriz N de "matriz não básica".

Considerando a partição básica (2.4) podemos reescrever (2.2) como

$$\begin{pmatrix} B \\ N \end{pmatrix} \beta + \begin{pmatrix} \varepsilon_B \\ \varepsilon_N \end{pmatrix} = \begin{pmatrix} y_B \\ y_N \end{pmatrix} \quad (2.5)$$

onde $\varepsilon_B \in \mathcal{R}^n$ e $y_B \in \mathcal{R}^n$ são formados, respectivamente, pelas componentes de ε e y que possuem os mesmos índices das linhas de B . Analogamente para $\varepsilon_N \in \mathcal{R}^{m-n}$ e $y_N \in \mathcal{R}^{m-n}$.

Então, (2.5) pode ser escrito na forma abaixo:

$$\begin{aligned} B\beta + \varepsilon_B &= y_B \\ N\beta + \varepsilon_N &= y_N \end{aligned} \quad (2.6)$$

Assim, de (2.6) obtemos:

$$\begin{cases} \beta = B^{-1}y_B - B^{-1}\varepsilon_B \\ \varepsilon_N = y_N - N\beta \end{cases} \quad (2.7)$$

Então, se atribuirmos valores a $\varepsilon_B = \hat{\varepsilon}_B$ em (2.7), teremos o valor das demais variáveis $\beta = \hat{\beta}$ e $\varepsilon_N = \hat{\varepsilon}_N$; isto é, basta atribuirmos valores a ε_B para obtermos uma solução do sistema (2.2). As variáveis ε_B são chamadas "variáveis livres" (em Programação Linear são chamadas de variáveis não básicas. Note que para este problema, as variáveis "não básicas" ε_B estão associadas à base B).

Temos, assim,

$$\begin{cases} \hat{\beta} = B^{-1}y_B - B^{-1}\hat{\varepsilon}_B \\ \hat{\varepsilon}_N = y_N - N\hat{\beta} \end{cases} \quad (2.8)$$

tal que, $(\hat{\beta}, \hat{\epsilon})$, com $\hat{\epsilon}^t = (\hat{\epsilon}_B, \hat{\epsilon}_N)$, satisfaz o sistema $X\beta + \epsilon = y$.

Se perturbarmos $\hat{\epsilon}_B$ por um vetor d_{ϵ_B} (d_{ϵ_B} é qualquer vetor de $\mathcal{R}^n$),

$$\epsilon_B = \hat{\epsilon}_B + d_{\epsilon_B} \quad (2.9)$$

podemos obter uma nova solução para o sistema $X\beta + \epsilon = y$, bastando substituir (2.9) em (2.7):

$$\begin{aligned} \beta &= B^{-1}y_B - B^{-1}\epsilon_B \\ &= B^{-1}y_B - B^{-1}(\hat{\epsilon}_B + d_{\epsilon_B}) \\ &= \hat{\beta} - B^{-1}d_{\epsilon_B} \\ &= \hat{\beta} + d_{\beta} \end{aligned}$$

$$\text{onde } d_{\beta} = -B^{-1}d_{\epsilon_B} \quad \text{e} \quad (2.10)$$

$$\begin{aligned} \epsilon_N &= y_N - N\beta \\ &= y_N - N(\hat{\beta} - B^{-1}d_{\epsilon_B}) \\ &= \hat{\epsilon}_N + NB^{-1}d_{\epsilon_B} \\ &= \hat{\epsilon}_N + d_{\epsilon_N} \end{aligned}$$

$$\text{com } d_{\epsilon_N} = \hat{N}d_{\epsilon_B} \text{ onde } \hat{N} = NB^{-1}. \quad (2.11)$$

Assim, a nova solução (β, ϵ) é

$$\begin{pmatrix} \beta \\ \varepsilon \end{pmatrix} = \begin{pmatrix} \hat{\beta} \\ \hat{\varepsilon} \end{pmatrix} + \begin{pmatrix} d_\beta \\ d_\varepsilon \end{pmatrix},$$

com $\varepsilon = \begin{pmatrix} \varepsilon_B \\ \varepsilon_N \end{pmatrix}$, $\hat{\varepsilon} = \begin{pmatrix} \hat{\varepsilon}_B \\ \hat{\varepsilon}_N \end{pmatrix}$, $d_\varepsilon = \begin{pmatrix} d_{\varepsilon_B} \\ d_{\varepsilon_N} \end{pmatrix}$, com d_{ε_B} escolhido e d_{ε_N} dado por (2.11).

Por construção, sabemos que a solução (β, ε) obtida acima satisfaz o sistema dado e, pelo Teorema 2.2, necessariamente devemos ter $d \in N(X, I)$, que pode ser facilmente verificado por:

$$d_\beta = -B^{-1}d_{\varepsilon_B} \Leftrightarrow Bd_\beta + d_{\varepsilon_B} = 0$$

$$d_{\varepsilon_N} = \hat{N}d_{\varepsilon_B} \Leftrightarrow d_{\varepsilon_N} = -Nd_\beta \Leftrightarrow Nd_\beta + d_{\varepsilon_N} = 0.$$

Assim,

$$\begin{pmatrix} B \\ N \end{pmatrix} d_\beta + \begin{pmatrix} d_{\varepsilon_B} \\ d_{\varepsilon_N} \end{pmatrix} = 0.$$

Portanto, $Xd_\beta + d_\varepsilon = 0$ e $d \in N(X, I)$.

2.1.3 DIREÇÕES SIMPLEX

Podemos também determinar n vetores de $N(X, I)$ linearmente independentes escolhendo

$$d_{\varepsilon_B}^j = e_j \quad j = 1, 2, \dots, n \quad (2.12)$$

onde e_j é um vetor do $\mathbb{R}^n$ com 1 na componente j e zero nas demais.

Temos assim, de (2.10) e (2.11) os vetores de $N(X, I)$:

$$\begin{cases} d_{\beta}^j = -(B^{-1})^j \\ d_{\varepsilon}^j = \begin{pmatrix} d_{\varepsilon_B}^j \\ d_{\varepsilon_N}^j \end{pmatrix} = \begin{pmatrix} e^j \\ \hat{N}^j \end{pmatrix} \end{cases}$$

ou

$$d^j = \begin{pmatrix} d_{\beta}^j \\ d_{\varepsilon}^j \end{pmatrix} \quad j = 1, 2, \dots, n \quad (2.13)$$

com d_{β}^j e d_{ε}^j dados acima, onde $(B^{-1})^j$ é a j -ésima coluna de B^{-1} e $\hat{N}^j = N(B^{-1})^j$.

Uma nova solução obtida por uma perturbação na direção d^j e passo δ

$$\begin{pmatrix} \beta \\ \varepsilon \end{pmatrix} = \begin{pmatrix} \hat{\beta} \\ \hat{\varepsilon} \end{pmatrix} + \delta \begin{pmatrix} d_{\beta}^j \\ d_{\varepsilon}^j \end{pmatrix}, \quad \delta > 0$$

corresponde a estratégia de alterar apenas a j -ésima componente do vetor das variáveis livres ε_B

$$\begin{cases} \varepsilon_i = \hat{\varepsilon}_i & i \in B \\ \varepsilon_j = \hat{\varepsilon}_j + \delta & j \in B, j \neq i \end{cases}$$

onde ε_i é a i -ésima componente de ε_B .

Tal estratégia é chamada "**estratégia simplex**" e corresponde a adotar a direção d^j definida em (2.13). As direções d^j , $j = 1, \dots, n$, são chamadas "**direções simplex**". Note que, $-d^j$ é também uma direção simplex e corresponde a estratégia:

$$\varepsilon_j = \hat{\varepsilon}_j - \delta.$$

Teorema 2.3: Se (d_β, d_ε) for uma direcção factível, então

$$d_\varepsilon = \sum_{i=1}^n d_{\varepsilon_i} d_\varepsilon^i$$

com $i \in B$.

Prova: Note inicialmente que:

$$\begin{aligned} [d_\varepsilon^1 \cdots d_\varepsilon^n] &= \left[\frac{I}{NB^{-1}} \right] \\ &= \left[\frac{B}{N} \right] B^{-1} \\ &= XB^{-1}. \end{aligned}$$

Por hipótese d é uma direcção factível. Então, pelo Teorema 2.1 temos que $d \in N(X, I)$, ou seja

$$Xd_\beta + d_\varepsilon = 0.$$

Segue que,

$$\begin{aligned} d_\varepsilon &= -Xd_\beta \\ &= XB^{-1}d_{\varepsilon_B} \end{aligned}$$

pois de acordo com (2.10) temos que $d_\beta = -B^{-1}d_{\varepsilon_B}$.

Assim,

$$\begin{aligned} d_\varepsilon &= [d_\varepsilon^1 \cdots d_\varepsilon^n] d_{\varepsilon_B} \\ &= \sum_{i=1}^n d_{\varepsilon_i} d_\varepsilon^i. \end{aligned}$$

■

2.1.4 SOLUÇÃO BÁSICA DO SISTEMA $X\beta + \varepsilon = y$

Vamos agora determinar as soluções básicas do sistema

$$X\beta + \varepsilon = y.$$

Definição 2.1: Seja X^i a i -ésima linha de X . Se $\varepsilon_i = 0$ (por conseguinte $X^i\beta = y_i$), então y_i é chamado "ponto de interpolação".

Uma solução (β, ε) de (2.1) com a particularidade de $\varepsilon_B = 0$ é muito importante para este trabalho. Será mostrado que os problemas (2.1) sempre admitem uma solução ótima com esta propriedade, isto é, interpola n pontos.

Definição 2.2: Se $\hat{\varepsilon}_B = 0$ na solução geral (2.8), dizemos que a solução $(\hat{\beta}, \hat{\varepsilon})$ é uma "solução básica".

Assim, uma solução básica $(\hat{\beta}, \hat{\varepsilon})$ de $X\beta + \varepsilon = y$, tem a seguinte forma:

$$\begin{cases} \hat{\beta} = B^{-1}y_B \\ \hat{\varepsilon}_N = y_N - N\hat{\beta} \\ \hat{\varepsilon}_B = 0 \end{cases} \quad (2.14)$$

Note que, para uma solução básica temos pelo menos n pontos de interpolação, onde y_i com $i \in B$ e $i = 1, \dots, n$ são os pontos de interpolação.

Se $\hat{\varepsilon}_{N_i} \neq 0$, $i = 1, \dots, m-n$, a solução básica é chamada "não degenerada" (isto é, a solução básica interpola exatamente n pontos). Caso contrário é "degenerada".

CAPÍTULO 3

PROPRIEDADES DE OTIMALIDADE

Neste Capítulo, estudaremos propriedades do Problema de Aproximação Linear no L_1 e do Problema de Regressão Quantil, com o objetivo de caracterizar soluções ótimas.

3.1 APROXIMAÇÃO LINEAR NO L_1

Seja o Problema de Aproximação Linear no L_1 :

$$\begin{aligned} \min \quad & \|\varepsilon\|_1 \\ \text{s.a.} \quad & X\beta + \varepsilon = y. \end{aligned} \tag{3.1}$$

Nesta seção estudaremos em que condições obtém-se o ótimo do problema acima.

3.1.1 DIREÇÕES DE DESCIDA

Vimos no Capítulo 2 o conceito de direções factíveis, as quais preservam a factibilidade de novas soluções obtidas através da perturbação de uma solução factível.

Em problemas de minimização, procuramos perturbar uma solução factível em determinadas direções, de tal forma que as novas soluções assim obtidas sejam factíveis e produzam valores para a função objetivo menores que os anteriores.

Definição 3.1: Seja $(\hat{\beta}, \hat{\varepsilon})$ uma solução do problema (3.1). Dizemos que uma direção factível $d = \begin{pmatrix} d_\beta \\ d_\varepsilon \end{pmatrix}$ é "direção de descida" em $(\hat{\beta}, \hat{\varepsilon})$ se a solução

$$\begin{pmatrix} \beta \\ \varepsilon \end{pmatrix} = \begin{pmatrix} \hat{\beta} \\ \hat{\varepsilon} \end{pmatrix} + \delta \begin{pmatrix} d_\beta \\ d_\varepsilon \end{pmatrix},$$

com $\delta > 0$ suficientemente pequeno, satisfizer

$$\|\varepsilon\|_1 < \|\hat{\varepsilon}\|_1.$$

Considerando d uma direção factível e $(\hat{\beta}, \hat{\varepsilon})$ solução de (3.1), temos que $\exists \delta > 0$ tal que

$$\begin{pmatrix} \beta \\ \varepsilon \end{pmatrix} = \begin{pmatrix} \hat{\beta} \\ \hat{\varepsilon} \end{pmatrix} + \delta \begin{pmatrix} d_\beta \\ d_\varepsilon \end{pmatrix},$$

seja uma solução factível.

Definindo

$$\text{sgn}(\hat{\varepsilon})^t = (\text{senal}(\hat{\varepsilon}_1), \dots, \text{senal}(\hat{\varepsilon}_m))$$

então, para $\delta > 0$, suficientemente pequeno, tal que

$$\text{sgn}(\hat{\varepsilon}_i) = \text{sgn}(\hat{\varepsilon}_i + \delta d_{\varepsilon_i})$$

com $i \in I_1 = \{i/\hat{\varepsilon}_i \neq 0\}$, temos

$$\begin{aligned}\|\varepsilon\|_1 &= \|\hat{\varepsilon} + \delta d_s\|_1 \\ &= \sum_{i \in I_0} |\delta d_{s_i}| + \sum_{i \in I_1} \text{sgn}(\hat{\varepsilon}_i + \delta d_{s_i})(\hat{\varepsilon}_i + \delta d_{s_i}) \\ \|\varepsilon\|_1 &= \|\hat{\varepsilon}\|_1 + \delta \left[\sum_{i \in I_0} |d_{s_i}| + \sum_{i \in I_1} \text{sgn}(\hat{\varepsilon}_i) d_{s_i} \right] \quad (3.2)\end{aligned}$$

com $I_0 = \{i/\hat{\varepsilon}_i = 0\}$.

Teorema 3.1: Uma direção factível d em $(\hat{\beta}, \hat{\varepsilon})$ é de descida se e somente se

$$\left[\sum_{i \in I_0} |d_{s_i}| + \sum_{i \in I_1} \text{sgn}(\hat{\varepsilon}_i) d_{s_i} \right] < 0 \quad (3.3)$$

com $I_0 = \{i/\hat{\varepsilon}_i = 0\}$ e $I_1 = \{i/\hat{\varepsilon}_i \neq 0\}$.

Prova:

($\Leftarrow$) Suponha que

$$\left[\sum_{i \in I_0} |d_{s_i}| + \sum_{i \in I_1} \text{sgn}(\hat{\varepsilon}_i) d_{s_i} \right] < 0.$$

Assim, de (3.2) temos:

$$\|\varepsilon\|_1 < \|\hat{\varepsilon}\|_1.$$

Portanto, d é uma direção de descida.

($\Rightarrow$) Suponha agora que d seja uma direção de descida, isto é, existe uma solução factível

$$\begin{pmatrix} \beta \\ \varepsilon \end{pmatrix} = \begin{pmatrix} \hat{\beta} \\ \hat{\varepsilon} \end{pmatrix} + \delta \begin{pmatrix} d_\beta \\ d_\varepsilon \end{pmatrix}, \quad \delta > 0$$

tal que

$$\|\varepsilon\|_1 < \|\hat{\varepsilon}\|_1.$$

Para $\delta > 0$, suficientemente pequeno, (3.2) é válido e portanto

$$\left[\sum_{i \in I_0} |d_{\varepsilon_i}| + \sum_{i \in I_1} \text{sgn}(\hat{\varepsilon}_i) d_{\varepsilon_i} \right] < 0. \quad \blacksquare$$

3.1.2 SOLUÇÃO ÓTIMA

Nesta seção estudaremos as condições para que uma solução básica seja ótima.

Definição 3.2: A solução (β^*, ε^*) para o problema (3.1) é uma "solução ótima" se

$$\|\varepsilon^*\|_1 \leq \|\varepsilon\|_1$$

com $\varepsilon = y - X\beta$, para $\forall \beta \in \mathcal{R}^n$.

Considerando uma solução básica $(\hat{\beta}, \hat{\varepsilon})$ não degenerada, isto é, $\hat{\varepsilon}_i \neq 0$ para $i \in N$, podemos escrever (3.3) usando o Teorema 2.3 como:

$$\left[\sum_{i=1}^n |d_{\varepsilon_i}| + \sum_{i=1}^n \text{sgn}(\hat{\varepsilon}_N)^t d_{\varepsilon_i} \hat{N}^i \right] < 0$$

$$\sum_{i=1}^n d_{\varepsilon_i} \left[\text{sgn}(d_{\varepsilon_i})^t + \text{sgn}(\hat{\varepsilon}_N)^t \hat{N}^i \right] < 0. \quad (3.4)$$

Consideremos agora uma direção simplex d^i ($d_{\varepsilon_j} = 0$, $j \neq i$). Temos duas possibilidades $d_{\varepsilon_i} > 0$ ou $d_{\varepsilon_i} < 0$.

a) $d_{\varepsilon_i} > 0$

Neste caso, temos:

$$d_{\varepsilon_i} (1 + \text{sgn}(\hat{\varepsilon}_N)^t \hat{N}^i) < 0.$$

Logo,

$$\text{sgn}(\hat{\varepsilon}_N)^t \hat{N}^i < -1.$$

b) $d_{\varepsilon_i} < 0$

$$d_{\varepsilon_i} (-1 + \text{sgn}(\hat{\varepsilon}_N)^t \hat{N}^i) < 0.$$

Assim,

$$\text{sgn}(\hat{\varepsilon}_N)^t \hat{N}^i > 1.$$

Seja $\hat{\lambda}_i = -\text{sgn}(\hat{\varepsilon}_N)^t \hat{N}^i$, a condição (3.3) para uma direção simplex numa solução básica não degenerada é:

$$\hat{\lambda}_i > 1 \text{ se } d_{\varepsilon_i} > 0$$

$$\hat{\lambda}_i < -1 \text{ se } d_{\varepsilon_i} < 0,$$

isto é, se $\hat{\lambda}_i > 1$ ou $\hat{\lambda}_i < -1$, então d^i ou $-d^i$ é direção de descida, e portanto a solução não é ótima.

A inexistência de direções simplex de descida caracteriza uma solução ótima, conforme o teorema a seguir.

Teorema 3.2: Considere uma partição básica. Se

$$-1 \leq \hat{\lambda}_i \leq 1 \quad \forall i \in B,$$

onde $\hat{\lambda}_i = -\text{sgn}(\hat{\varepsilon}_N)^t \hat{N}^i$, então a solução básica associada é ótima.

Prova: De (3.2) e (3.4) temos:

$$\begin{aligned} \|\varepsilon\|_1 &= \|\hat{\varepsilon}\|_1 + \delta \sum_{i=1}^n d_{\varepsilon_i} \left[\text{sgn}(d_{\varepsilon_i}) + \text{sgn}(\hat{\varepsilon}_N)^t \hat{N}^i \right] \\ &= \|\hat{\varepsilon}\|_1 + \delta \sum_{i=1}^n d_{\varepsilon_i} \left[\text{sgn}(d_{\varepsilon_i}) - \hat{\lambda}_i \right]. \end{aligned} \quad (3.5)$$

Como cada parcela de (3.4) pertence ao caso (a) ou ao caso (b) discutidos anteriormente, e por hipótese $-1 \leq \hat{\lambda}_i \leq 1$, temos:

$$\text{Se } d_{\varepsilon_i} > 0 \text{ então } d_{\varepsilon_i} (1 - \hat{\lambda}_i) \geq 0.$$

$$\text{Se } d_{\varepsilon_i} < 0 \text{ então } d_{\varepsilon_i} (-1 - \hat{\lambda}_i) \geq 0.$$

Assim, todas as parcelas de (3.4) são maiores ou iguais a zero. Logo, como $\delta > 0$, de (3.5) temos:

$$\|\varepsilon\|_1 \geq \|\hat{\varepsilon}\|_1.$$

Portanto, $(\hat{\beta}, \hat{\varepsilon})$ é uma solução básica ótima. ■

3.2 REGRESSÃO QUANTIL

Nesta seção temos por objetivo estudar propriedades de otimalidade para o Problema de Regressão Quantil:

$$\begin{aligned} \min & \left(\sum_{i \in I_0} \rho_\theta(\varepsilon_i) + \sum_{i \in I_1} \sigma_i(\varepsilon_i) \right) \\ \text{s.a. } & X\beta + \varepsilon = y. \end{aligned} \quad (3.6)$$

3.2.1 DIREÇÕES DE DESCIDA

Utilizando a definição de direção factível vista no Capítulo 2 seção 2.1.1, temos:

Definição 3.3: Seja $(\hat{\beta}, \hat{\varepsilon})$ uma solução do problema (3.6). Dizemos que uma direção factível $d = \begin{pmatrix} d_\beta \\ d_\varepsilon \end{pmatrix}$ é "direção de descida" em $(\hat{\beta}, \hat{\varepsilon})$ se a solução

$$\begin{pmatrix} \beta \\ \varepsilon \end{pmatrix} = \begin{pmatrix} \hat{\beta} \\ \hat{\varepsilon} \end{pmatrix} + \delta \begin{pmatrix} d_\beta \\ d_\varepsilon \end{pmatrix},$$

com $\delta > 0$ suficientemente pequeno, satisfizer

$$\left(\sum_{i \in I_0} \rho_\theta(\varepsilon_i) + \sum_{i \in I_1} \sigma_i(\varepsilon_i) \right) < \left(\sum_{i \in I_0} \rho_\theta(\hat{\varepsilon}_i) + \sum_{i \in I_1} \sigma_i(\hat{\varepsilon}_i) \right)$$

ou seja

$$f_{\theta}(\beta, \varepsilon) < f_{\theta}(\hat{\beta}, \hat{\varepsilon}).$$

Observe que esta definição é análoga a Definição 3.1 dada para o Problema de Aproximação Linear no L_1 .

Supondo d uma direção factível e $(\hat{\beta}, \hat{\varepsilon})$ solução de (3.6), temos que

$$\begin{pmatrix} \beta \\ \varepsilon \end{pmatrix} = \begin{pmatrix} \hat{\beta} \\ \hat{\varepsilon} \end{pmatrix} + \delta \begin{pmatrix} d_{\beta} \\ d_{\varepsilon} \end{pmatrix}$$

é uma solução factível.

Desta forma

$$\begin{aligned} f_{\theta}(\beta, \varepsilon) &= \sum_{i \in I_0} \rho_{\theta}(\varepsilon_i) + \sum_{i \in I_1} \sigma_i(\varepsilon_i) \\ &= \sum_{i \in I_0} \rho_{\theta}(\hat{\varepsilon}_i + \delta d_{\varepsilon_i}) + \sum_{i \in I_1} \sigma_i(\hat{\varepsilon}_i + \delta d_{\varepsilon_i}) \\ &= \left[\sum_{i \in I_0} \rho_{\theta}(\hat{\varepsilon}_i) + \sum_{i \in I_1} \sigma_i(\hat{\varepsilon}_i) \right] + \\ &\quad + \delta \left[\sum_{i \in I_0} \rho_{\theta}(d_{\varepsilon_i}) + \sum_{i \in I_1} \sigma_i(d_{\varepsilon_i}) \right] \\ f_{\theta}(\beta, \varepsilon) &= f_{\theta}(\hat{\beta}, \hat{\varepsilon}) + \delta \left[\sum_{i \in I_0} \rho_{\theta}(d_{\varepsilon_i}) + \sum_{i \in I_1} \sigma_i(d_{\varepsilon_i}) \right]. \quad (3.7) \end{aligned}$$

Assim, podemos enunciar o teorema correspondente ao Teorema 3.1:

Teorema 3.3: Uma direção factível d em $(\hat{\beta}, \hat{\varepsilon})$ é de descida se e somente se

$$\left[\sum_{i \in I_0} \rho_{\theta}(d_{\varepsilon_i}) + \sum_{i \in I_1} \sigma_i(d_{\varepsilon_i}) \right] < 0 \quad (3.8)$$

com

$$I_0 = \{i / \hat{\varepsilon}_i = 0\}$$

$$I_1 = \{i / \hat{\varepsilon}_i \neq 0\}$$

$$\rho_{\theta} \in [\theta - 1, \theta]$$

$$\sigma_i = \begin{cases} \theta & \text{se } i \in I^+ = \{i / \hat{\varepsilon}_i > 0\} \\ (\theta - 1) & \text{se } i \in I^- = \{i / \hat{\varepsilon}_i < 0\} \end{cases}$$

$$\theta \in (0, 1).$$

Prova: Análoga a prova do Teorema 3.1.

3.2.2 SOLUÇÃO ÓTIMA

Para que uma solução básica seja ótima, é necessário que determinadas condições sejam satisfeitas. Tais condições são estudadas nessa seção.

Utilizando o Teorema 2.3 da seção 2.1.3, e considerando a solução $(\hat{\beta}, \hat{\varepsilon})$ não degenerada, podemos escrever (3.8) como:

$$\left[\sum_{i=1}^n \rho_{\theta}(d_{\varepsilon_i}) + \sum_{i=1}^n \sigma^t(d_{\varepsilon_i} \hat{N}^i) \right] < 0$$

$$\sum_{i=1}^n d_{\varepsilon_i} (\rho_\theta + \sigma^t \hat{N}^i) < 0 \quad (3.9)$$

com $\sigma^t = (\sigma_1 \cdots \sigma_{m-n})$.

Considerando uma direção simplex e conforme o sinal de d_{ε_i} , esta será direção de descida se:

Caso 1: $d_{\varepsilon_i} > 0$

Temos

$$d_{\varepsilon_i} (\theta + \sigma^t \hat{N}^i) < 0.$$

Logo,

$$\sigma^t \hat{N}^i < -\theta.$$

Caso 2: $d_{\varepsilon_i} < 0$

Assim,

$$d_{\varepsilon_i} ((\theta - 1) + \sigma^t \hat{N}^i) < 0.$$

Portanto,

$$\sigma^t \hat{N}^i > -(\theta - 1).$$

Resumindo, seja $\hat{\lambda}_i = -\sigma^t \hat{N}^i$ com $i \in B$, a direção simplex d^i ou $-d^i$ será de descida se $\hat{\lambda}_i > \theta$ ou $\hat{\lambda}_i < (\theta - 1)$, respectivamente.

Observe que σ corresponde a $\text{sgn}(\hat{\varepsilon}_N)$ utilizado no cálculo de $\hat{\lambda}_i$ para o Problema no L_1 .

Desta forma podemos enunciar o teorema que nos dá a condição de otimalidade para o problema (3.6):

Teorema 3.4: Considere uma partição básica. Se

$$(\theta - 1) \leq \hat{\lambda}_i \leq \theta, \quad \forall i \in B$$

com $\hat{\lambda}_i = -\sigma^t \hat{N}^i$ então a solução básica é ótima.

Prova: Análoga a prova do Teorema 3.2.

Concluimos assim, que se uma solução básica factível não for ótima, existe uma direcção de descida disponível tal que uma nova solução básica factível melhor pode ser determinada através de uma perturbação em $(\hat{\beta}, \hat{\varepsilon})$.

CAPÍTULO 4

MÉTODO PRIMAL SIMPLEX

Neste Capítulo aplicaremos o Método Primal Simplex de Programação Linear ao Problema de Aproximação Linear no L_1

$$\begin{aligned} \min \quad & \|\varepsilon\|_1 \\ \text{s.a.} \quad & X\beta + \varepsilon = y \end{aligned} \tag{4.1}$$

e ao Problema de Regressão Quantil

$$\begin{aligned} \min \quad & \left(\sum_{i \in I_0} \rho_\theta(\varepsilon_i) + \sum_{i \in I_1} \sigma_i(\varepsilon_i) \right) \\ \text{s.a.} \quad & X\beta + \varepsilon = y. \end{aligned} \tag{4.2}$$

Um Método Primal tem a seguinte estrutura:

- i) Escolha uma solução primal factível;
 - ii) Repita
 - determine uma direcção factível de descida;
 - determine o passo nesta direcção;
 - obtenha a nova solução;
- Até que "uma condição de otimalidade" seja satisfeita.

A estratégia Simplex (que define o Método Primal Simplex) consiste em alterar uma solução básica factível perturbando apenas uma variável ε_i (a i -ésima componente de ε_B), mantendo as demais nulas. Isto corresponde a escolher o i -ésimo ponto para deixar de ser interpolado. Este argumento foi usado por Bloomfield e Steiger [9] para construir seu método.

4.1 MÉTODO PRIMAL SIMPLEX APLICADO AO PROBLEMA DE APROXIMAÇÃO LINEAR NO L_1

Desenvolvemos nessa seção o Método Simplex para o problema (4.1).

4.1.1 INICIAÇÃO

Considere uma partição básica nas linhas de X :

$$X = \begin{pmatrix} B \\ N \end{pmatrix},$$

com $B \in \mathcal{R}^{n \times m}$, $\text{posto}(X) = n$ e considere também uma solução básica factível primal, não degenerada, associada a essa partição básica:

$$\begin{cases} \hat{\beta} = B^{-1}y_B \\ \hat{\varepsilon}_N = y_N - N\hat{\beta} \\ \hat{\varepsilon}_B = 0 \end{cases} \quad (4.3)$$

4.1.2 TESTE DE OTIMALIDADE

Seja $\hat{\lambda}_i = -\text{sgn}(\hat{\varepsilon}_N)^t \hat{N}^i$, $i \in B$. Pelo Teorema 3.2 do Capítulo 3, seção 3.1.2, temos que, se $-1 \leq \hat{\lambda}_i \leq 1$, $\forall i \in B$, então a solução básica $(\hat{\beta}, \hat{\varepsilon})$ é ótima para o problema (4.1).

4.1.3 ESCOLHA DO PONTO QUE DEIXARÁ DE SER INTERPOLADO

Se existe $i \in B$ tal que $\hat{\lambda}_i < -1$ ou $\hat{\lambda}_i > 1$, então a solução atual não é ótima, e podemos determinar uma nova solução básica factível melhor perturbando $\hat{\varepsilon}_B$ da seguinte forma:

$$\varepsilon_B = \hat{\varepsilon}_B + \delta d_{\varepsilon_B}^i$$

isto é (veja a equação (2.12), onde $d_{\varepsilon_B}^i = (0 \dots 1 \dots 0)^t$ foi escolhido como uma solução particular, produzindo uma direção simplex):

$$\begin{cases} \varepsilon_i = \delta \geq 0 & i \in B \\ \varepsilon_j = 0 & j \in B, j \neq i \end{cases}$$

que corresponde a alterar apenas uma variável livre (variável não básica em P.L.), mantendo as demais nulas, ou seja, o i -ésimo ponto não será mais interpolado. Um novo ponto de interpolação deverá ser determinado.

As demais variáveis são alteradas por:

$$\begin{cases} \varepsilon_N = \hat{\varepsilon}_N + \delta \hat{N}^i \\ \beta = \hat{\beta} - \delta (B^{-1})^i \end{cases}$$

a fim de manter a restrição $X\beta + \varepsilon = y$ satisfeita (veja a equação (2.8)).

Vamos então mostrar que essa nova solução decresce a função objetivo.

Caso a) $\hat{\lambda}_i > 1$

Para $\delta > 0$ temos:

$$\begin{cases} \beta = \hat{\beta} - \delta(B^{-1})^i \\ \varepsilon_N = \hat{\varepsilon}_N + \delta \hat{N}^i \\ \varepsilon_i = \delta \quad i \in B \\ \varepsilon_j = 0 \quad j \in B, j \neq i \end{cases} \quad (4.4)$$

Assim,

$$\begin{aligned} \|\varepsilon\|_1 &= \|\varepsilon_B\|_1 + \|\varepsilon_N\|_1 \\ &= \delta + \|\hat{\varepsilon}_N + \delta \hat{N}^i\|_1. \end{aligned} \quad (4.5)$$

Supondo que o problema seja não degenerado, isto é, $\hat{\varepsilon}_k \neq 0, \forall k \in N$, e para $\delta > 0$ suficientemente pequeno temos que:

$$\text{sgn}(\hat{\varepsilon}_N) = \text{sgn}(\hat{\varepsilon}_N + \delta \hat{N}^i).$$

Como $\|\varepsilon\|_1 = \text{sgn}(\varepsilon)^t \varepsilon$ temos:

$$\|\varepsilon\|_1 = \delta + \|\hat{\varepsilon}_N\|_1 + \delta \text{sgn}(\hat{\varepsilon}_N)^t \hat{N}^i. \quad (4.6)$$

Considerando $\hat{\lambda}_i = -\text{sgn}(\hat{\varepsilon}_N)^t \hat{N}^i$, $\|\hat{\varepsilon}\|_1 = \|\hat{\varepsilon}_N\|_1$ (pois $\hat{\varepsilon}_B = 0$), de (4.6) temos:

$$\begin{aligned} \|\varepsilon\|_1 &= \|\hat{\varepsilon}\|_1 + \delta(1 - \hat{\lambda}_i) \\ &= \|\hat{\varepsilon}\|_1 + \delta \Delta_0 \end{aligned} \quad (4.7)$$

com $\Delta_0 = 1 - \hat{\lambda}_i$.

Note que (4.7) só vale se $\text{sgn}(\hat{\varepsilon}_N) = \text{sgn}(\hat{\varepsilon}_N + \delta \hat{N}^i)$.

Como por hipótese $\hat{\lambda}_i > 1$ e $\delta > 0$ temos

$$\|\varepsilon\|_1 < \|\hat{\varepsilon}\|_1.$$

Logo, a nova solução (4.4) diminui o valor da função objetivo e, portanto, a solução (4.3) com $\hat{\lambda}_i > 1$ para algum $i \in B$ não é solução ótima.

Caso b) $\hat{\lambda}_i < -1$

Para $\delta > 0$ temos:

$$\begin{cases} \varepsilon_i = -\delta & i \in B \\ \varepsilon_j = 0 & j \in B, j \neq i \\ \varepsilon_N = \hat{\varepsilon}_N - \delta \hat{N}^i \end{cases} \quad (4.8)$$

Assim,

$$\|\varepsilon\|_1 = \delta + \|\hat{\varepsilon}_N - \delta \hat{N}^i\|_1.$$

Com as mesmas suposições feitas para $\hat{\lambda}_i > 1$ temos:

$$\begin{aligned} \|\varepsilon\|_1 &= \delta + \|\hat{\varepsilon}\|_1 - \delta \text{sgn}(\hat{\varepsilon}_N)^t \hat{N}^i \\ &= \|\hat{\varepsilon}\|_1 + \delta(1 + \hat{\lambda}_i) \\ &= \|\hat{\varepsilon}\|_1 + \delta \Delta_0 \end{aligned}$$

com $\Delta_0 = 1 + \hat{\lambda}_i$.

Portanto,

$$\|\varepsilon\|_1 < \|\hat{\varepsilon}\|_1$$

ou seja, a solução (4.3) com $\hat{\lambda}_i < -1$ para algum $i \in B$ não é solução ótima.

4.1.4 DETERMINAÇÃO DO TAMANHO DO PASSO

Analisaremos o caso em que $\hat{\lambda}_i > 1$.

De (4.7) temos que quanto maior o valor de δ , menor será o valor da função objetivo. Porém, o valor de δ deve ser limitado, pois seu aumento pode fazer com que os resíduos ε_N mudem de sinal, e então é necessário recalcular $\hat{\lambda}_i$.

Considere os valores de δ que promovem uma mudança de sinal em ε_k , $\forall k \in N$:

$$\delta_k = -\frac{\hat{\varepsilon}_k}{\hat{N}_k^i} \geq 0.$$

Escrevendo estes valores em ordem crescente temos:

$$0 < \delta_{k_1} \leq \delta_{k_2} \leq \dots \leq \delta_{k_r}$$

supondo r razões.

Para $0 < \delta \leq \delta_{k_1}$ não há mudança de sinal e segundo (4.7) a função objetivo decresce.

Para $\delta_{k_1} < \delta \leq \delta_{k_2}$, fazemos $\delta = \delta_{k_1} + \alpha$ com $0 < \alpha \leq \delta_{k_2} - \delta_{k_1}$ e definimos:

$$\epsilon^{(1)} = \hat{\epsilon} + \delta_{k_1} \begin{pmatrix} e_i \\ \hat{N}^i \end{pmatrix}.$$

Então,

$$\epsilon = \hat{\epsilon} + \delta \begin{pmatrix} e_i \\ \hat{N}^i \end{pmatrix} = \epsilon^{(1)} + \alpha \begin{pmatrix} e_i \\ \hat{N}^i \end{pmatrix} \quad (4.9)$$

$$\text{sgn}(\epsilon) = \text{sgn}(\epsilon^{(1)}) + \begin{pmatrix} 0 \\ -2\text{sgn}(\hat{\epsilon}_{k_1})e_{k_1} \end{pmatrix}$$

onde $\hat{\epsilon}_{k_1}$ é a k_1 -ésima componente de $\hat{\epsilon}_N$.

Assim, $\text{sgn}(\epsilon_{k_1}^{(1)}) = +1$ ou -1 . Se em um determinado momento $\epsilon_{k_1}^{(1)}$ muda de sinal temos:

i) se $\text{sgn}(\epsilon_{k_1}^{(1)}) = 1$ então $\text{sgn}(\epsilon_{k_1}^{(1)})$ passa a ser -1

ii) se $\text{sgn}(\epsilon_{k_1}^{(1)}) = -1$ então seu novo sinal passa a ser $+1$

ou seja, quando há troca de sinal o novo sinal de $\epsilon_{k_1}^{(1)}$ passa a ser $\text{sgn}(\epsilon_{k_1}^{(1)}) - 2\text{sgn}(\epsilon_{k_1}^{(1)})$.

Daí dizermos que

$$\text{sgn}(\epsilon) = \text{sgn}(\epsilon^{(1)}) + \begin{pmatrix} 0 \\ -2\text{sgn}(\hat{\epsilon}_{k_1})e_{k_1} \end{pmatrix}. \quad (4.10)$$

Por definição, se $\delta = \delta_{k_j}$ temos $\epsilon_{k_j}^{(1)} = \hat{\epsilon}_{k_j} + \delta_{k_j} \hat{N}_{k_j}^i$.

Logo, $\epsilon_{k_j}^{(1)} = 0$.

Assim, utilizando (4.9) e (4.10) temos:

$$\begin{aligned}\|\varepsilon\|_1 &= \text{sgn}(\varepsilon)^t \varepsilon \\ &= \|\varepsilon^{(1)}\|_1 + \\ &\quad + \alpha \left[1 + \text{sgn}(\varepsilon_N^{(1)})^t \hat{N}^i - 2 \text{sgn}(\hat{\varepsilon}_{k_1}) \hat{N}_{k_1}^i \right].\end{aligned}$$

Como $\text{sgn}(\hat{\varepsilon}_{k_1}) = -\text{sgn}(\hat{N}_{k_1}^i)$ e $\text{sgn}(\varepsilon_N^{(1)}) = \text{sgn}(\hat{\varepsilon}_N)$ temos:

$$\begin{aligned}\|\varepsilon\|_1 &= \|\varepsilon^{(1)}\|_1 + \alpha \left[1 + \text{sgn}(\hat{\varepsilon}_N)^t \hat{N}^i + 2|\hat{N}_{k_1}^i| \right] \\ &= \|\varepsilon^{(1)}\|_1 + \alpha \left[1 - \hat{\lambda}_i + 2|\hat{N}_{k_1}^i| \right] \\ &= \|\varepsilon^{(1)}\|_1 + \alpha \Delta_1\end{aligned}$$

com $\Delta_0 = 1 - \hat{\lambda}_i$ e $\Delta_1 = \Delta_0 + 2|\hat{N}_{k_1}^i|$.

Se $\Delta_1 \geq 0$ então $\|\varepsilon\|_1 \geq \|\varepsilon^{(1)}\|_1$ para $\delta > \delta_{k_1}$. Devemos então escolher $\delta = \delta_{k_1}$, e obtemos $\varepsilon_i = \delta_{k_1}$ e $\varepsilon_{k_1} = 0$, com $i \in B$ e $k_1 \in N$.

Logo, ε_i é o ponto que deixará de ser interpolado e ε_{k_1} é o novo ponto de interpolação.

Devemos agora redefinir a partição básica, ou seja, devemos trocar a i -ésima linha de B pela k_1 -ésima linha de N .

Em seguida repetimos o processo para a nova partição básica até que o critério de otimalidade seja satisfeito.

Se $\Delta_1 < 0$ então $\|\varepsilon\|_1 < \|\varepsilon^{(1)}\|_1$. Devemos então tomar $\delta_{k_2} < \delta \leq \delta_{k_3}$ e refazer a análise acima para verificar se $\|\varepsilon\|_1 \geq \|\varepsilon^{(2)}\|_1$ para $\delta > \delta_{k_2}$.

Generalizando temos:

$$\varepsilon^{(j)} = \hat{\varepsilon} + \delta_{k_j} \begin{pmatrix} e_i \\ \hat{N}^i \end{pmatrix} \quad (4.11)$$

$$\Delta_j = \Delta_0 + 2 \sum_{s=1}^j |\hat{N}_{k_s}^i| \quad (4.12)$$

com $j = 1, \dots, r$ e $i \in B$.

Podemos escrever (4.11) e (4.12) como:

$$\begin{aligned} \varepsilon^{(j)} &= \varepsilon^{(j-1)} + (\delta_{k_j} - \delta_{k_{j-1}}) \begin{pmatrix} e_i \\ \hat{N}^i \end{pmatrix} \\ \Delta_j &= \Delta_{j-1} + 2 |\hat{N}_{k_j}^i| \quad j = 1, \dots, r. \end{aligned} \quad (4.13)$$

Assim,

$$\varepsilon = \hat{\varepsilon} + \delta \begin{pmatrix} e_i \\ \hat{N}^i \end{pmatrix}$$

onde $\delta_{k_j} < \delta \leq \delta_{k_{j+1}}$ pode ser escrito como

$$\varepsilon = \varepsilon^{(j)} + \alpha \begin{pmatrix} e_i \\ \hat{N}^i \end{pmatrix} \quad (4.14)$$

com $0 < \alpha \leq \delta_{k_{j+1}} - \delta_{k_j}$.

Podemos então escrever o seguinte teorema:

Teorema 4.1: Sejam ε e Δ_j dados, respectivamente, por (4.14) e (4.13). Então,

$$\|\varepsilon\|_1 = \|\varepsilon^{(j)}\|_1 + \alpha \Delta_j$$

com $j = 0, 1, \dots, r$, $\varepsilon^{(0)} = \hat{\varepsilon}$, $\delta_{k_0} = 0$, $0 < \alpha \leq \delta_{k_{j+1}} - \delta_{k_j}$ e $\Delta_0 = 1 - \hat{\lambda}_i$.

Prova: Para j qualquer considere:

$$\text{i) } \text{sgn}(\varepsilon) = \text{sgn}(\varepsilon^{(j)}) + \begin{pmatrix} 0 \\ -2\text{sgn}(\hat{\varepsilon}_{k_j})e_{k_j} \end{pmatrix}$$

$$\text{ii) } \text{sgn}(\varepsilon^{(j)}) = \text{sgn}(\hat{\varepsilon}) + \sum_{s=1}^{j-1} \begin{pmatrix} 0 \\ -2\text{sgn}(\hat{\varepsilon}_{k_s})e_{k_s} \end{pmatrix}$$

iii) a k_j -ésima componente de $\varepsilon_N^{(j)}$ é nula por definição de δ_{k_j} .

A demonstração para $j = 0$ foi feita anteriormente.

Suponhamos que vale para $j = r-1$, isto é,

$$\|\varepsilon\|_1 = \|\varepsilon^{(r-1)}\|_1 + \alpha \Delta_{r-1}.$$

Mostraremos que o teorema vale para $j = r$.

De (4.13) e (4.14) temos:

$$\Delta_r = \Delta_{r-1} + 2|\hat{N}_{k_r}^i| \quad (4.15)$$

$$\varepsilon = \varepsilon^{(r)} + \alpha \begin{pmatrix} e_i \\ \hat{N}^i \end{pmatrix}. \quad (4.16)$$

Temos ainda:

$$\text{sgn}(\boldsymbol{\varepsilon}) = \text{sgn}(\boldsymbol{\varepsilon}^{(r)}) + \begin{pmatrix} 0 \\ -2\text{sgn}(\hat{\boldsymbol{\varepsilon}}_{k_r})\mathbf{e}_{k_r} \end{pmatrix} \quad (4.17)$$

$$\text{sgn}(\boldsymbol{\varepsilon}^{(r)}) = \text{sgn}(\hat{\boldsymbol{\varepsilon}}) + \sum_{s=1}^{r-1} \begin{pmatrix} 0 \\ -2\text{sgn}(\hat{\boldsymbol{\varepsilon}}_{k_s})\mathbf{e}_{k_s} \end{pmatrix}. \quad (4.18)$$

Assim, como $\|\boldsymbol{\varepsilon}\|_1 = \text{sgn}(\boldsymbol{\varepsilon})^t \boldsymbol{\varepsilon}$, de (4.16) e (4.17) temos:

$$\begin{aligned} \|\boldsymbol{\varepsilon}\|_1 &= \|\boldsymbol{\varepsilon}^{(r)}\|_1 + \alpha \text{sgn}(\boldsymbol{\varepsilon}^{(r)})^t \begin{pmatrix} \mathbf{e}_i \\ \hat{N}^i \end{pmatrix} + \\ &+ \begin{pmatrix} 0 \\ -2\text{sgn}(\hat{\boldsymbol{\varepsilon}}_{k_r})\mathbf{e}_{k_r} \end{pmatrix}^t \left(\boldsymbol{\varepsilon}^{(r)} + \alpha \begin{pmatrix} \mathbf{e}_i \\ \hat{N}^i \end{pmatrix} \right). \end{aligned} \quad (4.19)$$

De (4.18) e (4.19) temos

$$\begin{aligned} \|\boldsymbol{\varepsilon}\|_1 &= \|\boldsymbol{\varepsilon}^{(r)}\|_1 + \alpha \left[\text{sgn}(\hat{\boldsymbol{\varepsilon}}_i) + \text{sgn}(\hat{\boldsymbol{\varepsilon}}_N)^t \hat{N}^i - \right. \\ &\quad \left. - 2 \sum_{s=1}^{r-1} \text{sgn}(\hat{\boldsymbol{\varepsilon}}_{k_s}) \hat{N}_{k_s}^i - 2\text{sgn}(\hat{\boldsymbol{\varepsilon}}_{k_r}) \hat{N}_{k_r}^i \right] \\ &= \|\boldsymbol{\varepsilon}^{(r)}\|_1 + \alpha \left[1 - \hat{\lambda}_i + 2 \sum_{s=1}^{r-1} |\hat{N}_{k_s}^i| + 2|\hat{N}_{k_r}^i| \right] \end{aligned}$$

pois, $\text{sgn}(\hat{\boldsymbol{\varepsilon}}_{k_t}) = -\text{sgn}(\hat{N}_{k_t}^i)$ para qualquer t .

Portanto,

$$\|\boldsymbol{\varepsilon}\|_1 = \|\boldsymbol{\varepsilon}^{(r)}\|_1 + \alpha \left[\Delta_0 + 2 \sum_{s=1}^{r-1} |\hat{N}_{k_s}^i| + 2|\hat{N}_{k_r}^i| \right].$$

De (4.12) e (4.15) obtemos:

$$\|\varepsilon\|_1 = \|\varepsilon^{(r)}\|_1 + \alpha \Delta_r.$$

Portanto, o teorema vale para $j = 0, 1, \dots, r$. ■

Pelo Teorema 4.1, a função objetivo na direção simplex é linear por partes e a inclinação no trecho $(\delta_{k_j}, \delta_{k_{j+1}})$ é dada por Δ_j . A figura abaixo ilustra o comportamento da função objetivo.

4.1.5 ESCOLHA DO PONTO A SER INTERPOLADO

Seja j tal que

$$\Delta_{j-1} < 0 \text{ e } \Delta_j \geq 0.$$

Fazemos $\varepsilon_i = \delta_{k_j}$ e assim $\varepsilon_{k_j} = 0$.

Desta forma, o resíduo ε_{k_j} atinge o valor zero e, portanto, o k_j -ésimo ponto em N é interpolado. Trocamos a i -ésima linha de B pela k_j -ésima linha de N , obtendo assim uma nova partição.

Assim, devemos repetir o processo para esta nova partição, até a condição de otimalidade ser satisfeita.

Toda a análise foi feita para o Caso a onde $\hat{\lambda}_i > 1$.

A análise para o Caso b, $\hat{\lambda}_i < -1$, é análoga e as fórmulas diferem apenas nos cálculos de:

$$\delta_k = \frac{\hat{\varepsilon}_k}{\hat{N}_k^i} \geq 0$$

$$\Delta_0 = 1 + \hat{\lambda}_i < 0.$$

4.1.6 CASO DEGENERADO

Quando uma ou mais componentes de $\hat{\varepsilon}_k$ com $k \in N$ se anulam, a solução básica é chamada "solução básica degenerada".

Nas seções anteriores consideramos sempre a solução sendo não degenerada.

Na seção 4.1.4, considerando o problema não degenerado e $\hat{\lambda}_i > 1$ para algum $i \in B$, vimos que os valores de δ que causam mudança de sinal em ε_k , para $\forall k \in N$, eram dados por:

$$\delta_k = -\frac{\hat{\varepsilon}_k}{\hat{N}_k^i} \geq 0.$$

No caso degenerado, apesar de $\hat{\varepsilon}_k = 0$, adotamos $\text{sgn}(\hat{\varepsilon}_k)$ igual a 1 ou -1. A razão δ_k será considerada apenas se $\text{sgn}(\hat{N}_k^i) = -\text{sgn}(\hat{\varepsilon}_k)$.

Assim, o problema de degeneração é resolvido e o método prossegue normalmente.

4.1.7 ALGORITMO

Determine uma partição básica inicial conforme a seção 4.1.1, isto é:

$$\mathbf{X} = \begin{pmatrix} \mathbf{B} \\ \mathbf{N} \end{pmatrix},$$

com $\mathbf{B} \in \mathfrak{R}^{n \times n}$.

Passo 1: Iniciação

- Escolha uma solução básica factível não degenerada:

$$\begin{cases} \hat{\beta} = \mathbf{B}^{-1} \mathbf{y}_B \\ \hat{\varepsilon}_N = \mathbf{y}_N - \mathbf{N} \hat{\beta} \\ \hat{\varepsilon}_B = 0 \end{cases}$$

Passo 2: Teste de Otimalidade

- Calcular

$$\hat{\lambda}^t = -\text{sgn}(\hat{\varepsilon}_N)^t \mathbf{N} \mathbf{B}^{-1}$$

- Se $-1 \leq \hat{\lambda}_i \leq 1$ para $\forall i \in B$

Então

- $(\hat{\beta}, \hat{\varepsilon})$ é ótima. Pare.

Senão

- o i -ésimo ponto que deixará de ser interpolado é calculado tal que:

$$\min \{1 - \hat{\lambda}_i, \hat{\lambda}_i + 1\}$$

Passo 3: Determinação do Tamanho do Passo

- Considere $\Delta_0 = 1 - \hat{\lambda}_i$ (quando $\Delta_0 = 1 + \hat{\lambda}_i$ apenas 3.1 se altera)

3.1 Calcule os pontos críticos $k \in N$

$$\delta_k = -\frac{\hat{\varepsilon}_k}{\hat{N}_k^i} \geq 0$$

(Se $\hat{\varepsilon}_k = 0$, $\delta_k = 0$ é considerado apenas se $\text{sgn}(\hat{N}_k^i) = -\text{sgn}(\hat{\varepsilon}_k)$)

3.2 Ordenação dos δ_k (r pontos)

$$0 < \delta_{k_1} \leq \delta_{k_2} \leq \dots \leq \delta_{k_r}$$

3.3 Passo ótimo

$$j = 0$$

Enquanto $(\Delta_{j-1} + 2|\hat{N}_{k_j}^i|) < 0$ faça

$$j = j + 1$$

$$\Delta_j = \Delta_{j-1} + 2|\hat{N}_{k_j}^i|$$

(k_j é o novo ponto a ser interpolado)

Passo 4: Atualização

- Atualização da partição básica: **B** e **N** são alteradas por

$$x_{B_i} \leftrightarrow x_{N_{k_j}}$$

- Volte para o *Passo 2*.

4.2 MÉTODO PRIMAL SIMPLEX APLICADO AO PROBLEMA DE REGRESSÃO QUANTIL

Desenvolvemos aqui o Método Simplex para o problema (4.2).

4.2.1 INICIAÇÃO

Vimos que a restrição para os problemas (4.1) e (4.2) é a mesma. Desta forma, a partição e a solução básica inicial de (4.2) é a mesma de (4.1), isto é:

$$X = \begin{pmatrix} B \\ N \end{pmatrix},$$

com $B \in \mathcal{R}^{n \times n}$, $\text{posto}(X) = n$ e

$$\begin{cases} \hat{\beta} = B^{-1}y_B \\ \hat{\varepsilon}_N = y_N - N\hat{\beta} \\ \hat{\varepsilon}_B = 0 \end{cases} \quad (4.20)$$

4.2.2 TESTE DE OTIMALIDADE

Denotando $\hat{\lambda}_i = -\sigma^t \hat{N}^i$

onde

$$\hat{N}^i = N(B^{-1})^i$$

$$\sigma^t = (\sigma_1 \cdots \sigma_{m-n}) \text{ com } \sigma_i = \begin{cases} \theta & \text{se } i \in I^+ = \{i / \text{sgn}(\hat{\varepsilon}_i) = 1\} \\ \theta - 1 & \text{se } i \in I^- = \{i / \text{sgn}(\hat{\varepsilon}_i) = -1\} \end{cases}$$

e utilizando o Teorema 3.4 do Capítulo 3 seção 3.2.2, temos que, se $(\theta - 1) \leq \hat{\lambda}_i \leq \theta$, $\forall i \in B$, então a solução básica $(\hat{\beta}, \hat{\varepsilon})$ é ótima para o problema (4.2).

4.2.3 ESCOLHA DO PONTO QUE DEIXARÁ DE SER INTERPOLADO

Se a condição de otimalidade não for satisfeita para algum $i \in B$ então o i -ésimo ponto não será mais interpolado e devemos determinar uma nova solução básica.

Assim, essa nova solução é dada por

$$\begin{cases} \beta = \hat{\beta} - \delta(B^{-1})^i \\ \varepsilon_N = \hat{\varepsilon}_N + \delta \hat{N}^i \\ \varepsilon_i = \delta & i \in B \\ \varepsilon_j = 0 & j \in B, j \neq i \end{cases} \quad (4.21)$$

tal que $X\beta + \varepsilon = y$. Note que (4.21) é igual a nova solução para o problema (4.1). (Veja a equação (4.4).)

Analogamente ao que foi feito na seção 4.1.3, mostramos que essa nova solução decresce a função objetivo do problema (4.2).

4.2.4 DETERMINAÇÃO DO TAMANHO DO PASSO

Esta seção visa determinar os valores de δ que causam mudança de sinal em ε_k , $\forall k \in N$, para que possamos determinar uma nova solução factível que decresce a função objetivo.

Devido ao fato das restrições dos problemas (4.1) e (4.2) serem as mesmas, os valores de δ que promovem uma mudança de sinal em ε_k , $\forall k \in N$, são obtidos da mesma forma, ou seja:

$$\delta_k = -\frac{\hat{\varepsilon}_k}{\hat{N}_k^i} \geq 0 \quad (4.22)$$

para o caso onde $\hat{\lambda}_i > \theta$.

Colocando estes valores em ordem crescente, como na seção 4.1.4, vemos que para $0 < \delta \leq \delta_{k_1}$ não há mudança de sinal e a função objetivo decresce.

Para $\delta_{k_1} < \delta \leq \delta_{k_2}$, fazemos $\delta = \delta_{k_1} + \alpha$ com $0 < \alpha \leq \delta_{k_2} - \delta_{k_1}$ e definimos:

$$\varepsilon^{(1)} = \hat{\varepsilon} + \delta_{k_1} \begin{pmatrix} e_i \\ \hat{N}^i \end{pmatrix}.$$

Então,

$$\varepsilon = \hat{\varepsilon} + \delta \begin{pmatrix} e_i \\ \hat{N}^i \end{pmatrix} = \varepsilon^{(1)} + \alpha \begin{pmatrix} e_i \\ \hat{N}^i \end{pmatrix} \quad (4.23)$$

$$\sigma = \sigma^{(1)} - \text{sgn}(\hat{\varepsilon}_{k_1}) e_{k_1} \quad (4.24)$$

onde $\hat{\varepsilon}_{k_1}$ é a k_1 -ésima componente de $\hat{\varepsilon}_N$.

Se $\varepsilon_{k_1}^{(1)}$ muda de sinal temos:

i) se $\sigma_{k_1} = \theta$ então σ_{k_1} passa a ser $(\theta - 1)$

ii) se $\sigma_{k_1} = (\theta - 1)$ então seu valor passa a ser θ

ou seja, quando há troca de sinal temos:

$$\sigma_{k_1} = \sigma_{k_1}^{(1)} - \text{sgn}(\varepsilon_{k_1}^{(1)}).$$

Por definição, se $\delta = \delta_{k_j}$ temos $\varepsilon_{k_j}^{(1)} = \hat{\varepsilon}_{k_j} + \delta_{k_j} \hat{N}_{k_j}^i$. Logo, $\varepsilon_{k_j}^{(1)} = 0$.

Utilizando (4.23) e (4.24) temos:

$$\begin{aligned} f_{\theta}(\beta, \varepsilon) &= \sum_{i \in B} \rho_{\theta}(\varepsilon_i) + \sum_{i \in N} \sigma_i(\varepsilon_i) \\ &= \rho_{\theta}(\varepsilon_B) + \sigma^t(\varepsilon_N) \\ &= \rho_{\theta}(\varepsilon_B^{(1)} + \alpha \varepsilon_i) + \\ &\quad + (\sigma^{(1)} - \text{sgn}(\hat{\varepsilon}_{k_1}) e_{k_1})^t (\varepsilon_N^{(1)} + \alpha \hat{N}^i) \\ &= f_{\theta}(\beta, \varepsilon^{(1)}) + \alpha [\theta + \sigma^{(1)} \hat{N}^i - \text{sgn}(\hat{\varepsilon}_{k_1}) \hat{N}_{k_1}^i] - \\ &\quad - \text{sgn}(\hat{\varepsilon}_{k_1}) (\varepsilon_{k_1}^{(1)}). \end{aligned}$$

De (4.22) temos que $\text{sgn}(\hat{\varepsilon}_{k_1}) = -\text{sgn}(\hat{N}_{k_1}^i)$, então

$$\begin{aligned} f_{\theta}(\beta, \varepsilon) &= f_{\theta}(\beta, \varepsilon^{(1)}) + \alpha [\theta - \hat{\lambda}_i + |\hat{N}_{k_1}^i|] \\ &= f_{\theta}(\beta, \varepsilon^{(1)}) + \alpha \Delta_1 \end{aligned}$$

com $\Delta_0 = \theta - \hat{\lambda}_i$ e $\Delta_1 = \Delta_0 + |\hat{N}_{k_1}^i|$.

Se $\Delta_1 \geq 0$ então $f_{\theta}(\beta, \varepsilon) \geq f_{\theta}(\beta, \varepsilon^{(1)})$ para $\delta > \delta_{k_1}$. Assim, escolhemos $\delta = \delta_{k_1}$, e obtemos $\varepsilon_i = \delta_{k_1}$ e $\varepsilon_{k_1} = 0$, com $i \in B$ e $k_1 \in N$.

Portanto, ε_i é o ponto que deixará de ser interpolado e ε_{k_1} é o novo ponto de interpolação.

Logo, devemos atualizar a partição básica e repetir o processo até a condição de otimalidade ser satisfeita.

Se $\Delta_1 < 0$ então $f_\theta(\beta, \varepsilon) < f_\theta(\beta, \varepsilon^{(1)})$. Devemos então tomar $\delta_{k_2} < \delta \leq \delta_{k_3}$ e refazer a análise acima para verificar se $f_\theta(\beta, \varepsilon) \geq f_\theta(\beta, \varepsilon^{(2)})$ para $\delta > \delta_{k_2}$.

Generalizando como na seção 4.1.4, temos as mesmas equações, ou seja:

$$\varepsilon^{(j)} = \hat{\varepsilon} + \delta_{k_j} \begin{pmatrix} e_i \\ \hat{N}^i \end{pmatrix} \quad (4.25)$$

$$\Delta_j = \Delta_0 + \sum_{s=1}^j |\hat{N}_{k_s}^i| \quad (4.26)$$

com $j = 1, \dots, r$.

Podemos escrever (4.25) e (4.26) como:

$$\varepsilon^{(j)} = \varepsilon^{(j-1)} + (\delta_{k_j} - \delta_{k_{j-1}}) \begin{pmatrix} e_i \\ \hat{N}^i \end{pmatrix} \quad (4.27)$$

$$\Delta_j = \Delta_{j-1} + |\hat{N}_{k_j}^i| \quad j = 1, \dots, r. \quad (4.28)$$

Assim,

$$\varepsilon = \hat{\varepsilon} + \delta \begin{pmatrix} e_i \\ \hat{N}^i \end{pmatrix}$$

onde $\delta_{k_j} < \delta \leq \delta_{k_{j+1}}$ pode ser escrito como

$$\varepsilon = \varepsilon^{(j)} + \alpha \begin{pmatrix} e_i \\ \hat{N}^i \end{pmatrix} \quad (4.29)$$

com $0 < \alpha \leq \delta_{k_{j+1}} - \delta_{k_j}$.

Logo, o Teorema 4.1 apresentado na seção 4.1.4, continua valendo se na sua demonstração considerarmos:

$$\text{i) } \sigma = \sigma^{(j)} - \text{sgn}(\hat{\varepsilon}_{k_j}) e_{k_j}$$

$$\text{ii) } \sigma^{(j)} = \sigma + \sum_{s=1}^{j-1} [-\text{sgn}(\hat{\varepsilon}_{k_s}) e_{k_s}]$$

$$\text{iii) } \varepsilon_{k_j}^{(j)} = 0 \text{ por definição de } \delta_{k_j}.$$

Portanto, temos o seguinte teorema:

Teorema 4.2: Sejam ε e Δ_j dados, respectivamente, por (4.29) e (4.28). Então,

$$f_\theta(\beta, \varepsilon) = f_\theta(\beta, \varepsilon^{(j)}) + \alpha \Delta_j$$

com $j = 0, 1, \dots, r$, $\varepsilon^{(0)} = \hat{\varepsilon}$, $\delta_{k_0} = 0$, $0 < \alpha \leq \delta_{k_{j+1}} - \delta_{k_j}$ e $\Delta_0 = \theta - \hat{\lambda}_i$.

4.2.5 ESCOLHA DO PONTO A SER INTERPOLADO

O novo ponto de interpolação é obtido através do mesmo processo realizado para o problema (4.1), isto é, se

$$\Delta_{j-1} < 0 \text{ e } \Delta_j \geq 0$$

então fazemos $\varepsilon_i = \delta_{k_j}$ e conseqüentemente $\varepsilon_{k_j} = 0$, e assim, o k_j -ésimo ponto em N é interpolado.

Assim, devemos atualizar a partição básica e repetir o método até a condição de otimalidade ser satisfeita.

Para o caso onde $\hat{\lambda}_i < (\theta - 1)$ o raciocínio é análogo, diferindo apenas nos cálculos de:

$$\delta_k = \frac{\hat{\varepsilon}_k}{\hat{N}_k^i} \geq 0$$

$$\Delta_0 = \hat{\lambda}_i - (\theta - 1) < 0.$$

Em caso de degeneração, o procedimento é análogo à seção 4.1.6.

4.2.6 ALGORITMO

Como podemos perceber, o algoritmo para o Problema de Regressão Quantil, (4.2), é semelhante ao do Problema de Minimização no L_1 com as seguintes diferenças:

Teste de Otimalidade

- Calcular

$$\hat{\lambda}^t = -\sigma^t N B^{-1} \text{ onde}$$

$$\sigma_i = \begin{cases} \theta & \text{se } i \in I^+ = \{i / \text{sgn}(\hat{\varepsilon}_i) = 1\} \\ \theta - 1 & \text{se } i \in I^- = \{i / \text{sgn}(\hat{\varepsilon}_i) = -1\} \end{cases}$$

- O teste de otimalidade será:

$$(\theta - 1) \leq \hat{\lambda}_i \leq \theta \quad \forall i \in B$$

Determinação do Tamanho do Passo

Passo Ótimo

Suponha $\Delta_0 = \theta - \hat{\lambda}_i$ (se $\Delta_0 = \hat{\lambda}_i - (\theta - 1)$,
o procedimento é análogo)

$$j = 0$$

Enquanto $(\Delta_{j-1} + |\hat{N}_{k_j}^i|) < 0$ faça

$$j = j + 1$$

$$\Delta_j = \Delta_{j-1} + |\hat{N}_{k_j}^i|$$

(k_j é o novo ponto a ser interpolado)

CAPÍTULO 5

ANÁLISE PÓS OTIMIZAÇÃO

Neste Capítulo fazemos um estudo de análise pós otimização para o Problema de Regressão Quantil.

Resolvido o Problema de Regressão Quantil para um θ_0 particular, obtemos uma solução básica (β_0, ε_0) e λ_0 correspondente.

Assim, uma aplicação padrão de técnicas de pós otimização da Programação Linear permite que soluções sejam desenvolvidas como funções de perturbações em θ_0 . Para pequenas perturbações em θ_0 , a solução básica fica inalterada (a menos de degeneração).

Desta forma, colocamos a seguinte questão:

"Ao perturbarmos θ_0 , isto é, $\theta = \theta_0 + \mu$, para quais valores de μ a base anterior continua ótima?"

Por hipótese temos que $\lambda_0^t = -\sigma_0^t \hat{N}$ onde $\sigma_0^t = (\theta_0, \dots, \theta_0, \theta_0 - 1, \dots, \theta_0 - 1)$, supondo que os resíduos positivos - componentes de ε_0 - sejam os primeiros, por conveniência de notação.

Para resolvermos o problema em questão devemos considerar dois casos.

Supondo que a base ótima em θ_0 continue ótima, os novos valores para σ e λ , (após a perturbação) são dados por:

$$\begin{aligned}\sigma^t &= (\theta_0 + \mu, \dots, \theta_0 + \mu, (\theta_0 + \mu) - 1, \dots, (\theta_0 + \mu) - 1) \\ &= \sigma_0^t + \mu e^t\end{aligned}\quad (5.1)$$

onde $e^t = (1 \dots 1)$.

Assim,

$$\begin{aligned}\lambda(\mu) &= -\sigma^t \hat{N} \\ &= -(\sigma_0^t + \mu e^t) \hat{N} \\ &= -\sigma_0^t \hat{N} - \mu e^t \hat{N} \\ &= \lambda_0^t - \mu e^t \hat{N}.\end{aligned}\quad (5.2)$$

Para que, com esses valores de $\theta = \theta_0 + \mu$ a solução básica seja ótima, devemos ter o critério de otimalidade satisfeito.

Logo devemos ter:

$$(\theta_0 + \mu) - 1 \leq (\lambda(\mu))_i \leq (\theta_0 + \mu) \quad i = 1, \dots, n$$

ou

$$\theta_0 + \mu - 1 \leq (\lambda_0)_i - \mu e^t \hat{N}^i \leq \theta_0 + \mu.$$

Analisemos as duas situações:

Situação 1: Para $(-e^t \hat{N}^i) > 1$, $i = 1, \dots, n$ devemos ter:

a) Se $\mu > 0$

$$(\lambda_0)_i - \mu(e^t \hat{N}^i) \leq \theta_0 + \mu$$

$$\mu(1 + e^t \hat{N}^i) \geq (\lambda_0)_i - \theta_0$$

$$\mu \leq \frac{(\lambda_0)_i - \theta_0}{1 + e^t \hat{N}^i} = \bar{\mu}_i.$$

b) Se $\mu < 0$

$$\theta_0 + \mu - 1 \leq (\lambda_0)_i - \mu(e^t \hat{N}^i)$$

$$\mu(1 + e^t \hat{N}^i) \leq (\lambda_0)_i - (\theta_0 - 1)$$

$$\mu \geq \frac{(\lambda_0)_i - (\theta_0 - 1)}{1 + e^t \hat{N}^i} = \underline{\mu}_i.$$

Assim, devemos escolher μ tal que

$$\underline{\mu} \leq \mu \leq \bar{\mu}$$

onde

$$\bar{\mu} = \min_{i=1, \dots, n} \{ \bar{\mu}_i, \text{ com } 1 + e^t \hat{N}^i < 0 \}$$

e

$$\underline{\mu} = \max_{i=1, \dots, n} \{ \underline{\mu}_i, \text{ com } 1 + e^t \hat{N}^i < 0 \}.$$

Situação 2: Para $(-e^t \hat{N}^i) < 1$, $i = 1, \dots, n$ devemos ter:

a) Se $\mu > 0$

$$\theta_0 + \mu - 1 \leq (\lambda_0)_i - \mu(e^t \hat{N}^i)$$

$$\mu \leq \frac{(\lambda_0)_i - (\theta_0 - 1)}{1 + e^t \hat{N}^i} = \bar{\mu}_i.$$

b) Se $\mu < 0$

$$(\lambda_0)_i - \mu(e^t \hat{N}^i) \leq \theta_0 + \mu$$

$$\mu(1 + e^t \hat{N}^i) \geq (\lambda_0)_i - \theta_0$$

$$\mu \geq \frac{(\lambda_0)_i - \theta_0}{1 + e^t \hat{N}^i} = \underline{\mu}_i$$

ou seja, μ deve ser escolhido tal que

$$\underline{\mu} \leq \mu \leq \bar{\mu}$$

onde

$$\bar{\mu} = \min_{i=1, \dots, n} \{ \bar{\mu}_i, \text{ com } 1 + e^t \hat{N}^i > 0 \}$$

e

$$\underline{\mu} = \max_{i=1, \dots, n} \{ \underline{\mu}_i, \text{ com } 1 + e^t \hat{N}^i > 0 \}.$$

Note que, quando $(-e^t \hat{N}^i) = 1$, para qualquer valor de μ a solução é ótima, isto é, a solução ótima independe de θ_0 (um exemplo dessa situação é ilustrado em [11]).

Geometricamente temos:

Fig. 5.1: $-e^t \hat{N}^i > 1$

Fig. 5.2: $-e^t \hat{N}^i < 1$

Fig. 5.3: $-e^t \hat{N}^i = 1$

Portanto, a perturbação feita em θ_0 fornecerá uma solução ótima desde que $\mu \in (\underline{\mu}, \bar{\mu})$.

No capítulo seguinte apresentamos um exemplo.

CAPÍTULO 6

RESULTADOS COMPUTACIONAIS

Como o algoritmo da seção 4.1.7 corresponde a $\theta = 0.5$ na seção 4.2.7, optamos por implementar o algoritmo descrito na seção 4.2.6 que foi executado para vários valores de θ , inclusive $\theta = 0.5$.

A implementação do algoritmo foi feita em um microcomputador PC 486 DX50 com 4 MB de memória e sistema operacional D.O.S. v.6.2, na linguagem Turbo Pascal v.7.0.

A seguir são apresentados dois exemplos extraídos de [11], onde são resolvidos através de um método baseado no algoritmo do gradiente reduzido.

6.1 EXEMPLO 1

Os dados do problema são apresentados na Tabela 6.1. Este exemplo é discutido por Bassett & Koenker [8], que identificaram uma degeneração correspondendo a $I_0 = \{6, 7, 13, 14, 16, 17, 18, 19\}$ que utiliza $0.13 < \theta < 0.275$.

O método usado por estes autores diferem da nossa abordagem. Eles sugerem que o problema seja resolvido por programação paramétrica.

Como os resultados foram publicados muito resumidamente, não foi possível fazer uma comparação mais efetiva, uma vez que não foi nosso propósito implementar o método por eles apresentado.

O resultado obtido para vários valores de θ é mostrado na Tabela 6.2.

Tabela 6.1

Exemplo 1

i	X				y
1	1	80	27	89	42
2	1	80	27	88	37
3	1	75	25	90	37
4	1	62	24	87	28
5	1	62	22	87	18
6	1	62	23	87	18
7	1	62	24	93	19
8	1	62	24	93	20
9	1	58	23	87	15
10	1	58	18	80	14
11	1	58	18	89	14
12	1	58	17	88	13
13	1	58	18	82	11
14	1	58	19	93	12
15	1	50	18	89	8
16	1	50	18	86	7
17	1	50	19	72	8
18	1	50	19	79	8
19	1	50	20	80	9
20	1	56	20	82	15
21	1	70	20	91	15

Tabela 6.2

Resultados para o Exemplo 1

θ	NUM ITER	SOLUÇÃO (QUANTILE)				PTOS. INTERP.
0.120	12	-29.014	0.315	1.224	-0.028	09 21 17 19
0.124	10	-29.014	0.315	1.224	-0.028	19 21 09 17
0.126	9	-36.078	0.352	1.750	-0.094	21 17 07 19
0.130	9	-36.078	0.352	1.750	-0.094	21 17 07 19
0.140	11	-36.000	0.500	1.000	0.000	17 07 19 18
0.149	9	-36.000	0.500	1.000	0.000	17 07 13 16
0.150	9	-36.000	0.500	1.000	0.000	17 07 13 16
0.161	9	-36.000	0.500	1.000	0.000	17 07 13 16
0.185	8	-36.000	0.500	1.000	0.000	13 19 07 06
0.195	7	-36.000	0.500	1.000	0.000	13 19 07 06
0.200	7	-36.000	0.500	1.000	0.000	13 19 07 06
0.223	8	-36.000	0.500	1.000	0.000	13 19 07 06
0.238	6	-36.000	0.500	1.000	0.000	13 19 07 17
0.254	8	-36.000	0.500	1.000	0.000	06 17 14 18
0.256	7	-36.000	0.500	1.000	0.000	06 17 14 07
0.259	7	-36.000	0.500	1.000	0.000	06 17 14 07
0.267	7	-36.000	0.500	1.000	0.000	06 17 14 07
0.275	7	-36.000	0.500	1.000	0.000	06 17 14 07
0.300	7	-37.897	0.757	0.794	-0.098	17 05 14 07
0.331	7	-37.897	0.757	0.794	-0.098	17 05 14 07
0.355	7	-38.527	0.843	0.726	-0.131	07 02 14 17
0.375	7	-32.638	0.825	0.740	-0.186	07 02 14 18
0.382	7	-32.638	0.825	0.740	-0.186	07 02 14 18
0.392	8	-33.246	0.849	0.627	-0.166	07 02 10 18
0.400	8	-33.246	0.849	0.627	-0.166	07 02 10 18
0.409	8	-33.246	0.849	0.627	-0.166	07 02 10 18
0.450	6	-39.652	0.830	0.581	-0.062	18 02 10 08
0.490	7	-39.690	0.832	0.574	-0.061	18 02 08 16
0.500	6	-39.690	0.832	0.574	-0.061	18 02 08 16
0.565	10	-39.731	0.834	0.566	-0.060	18 02 12 08
0.582	8	-39.731	0.834	0.566	-0.060	12 02 18 08

Tabela 6.2 (continuação)

θ	NUM ITER	SOLUÇÃO (QUANTILE)				PTOS. INTERP.
0.592	8	-39.731	0.834	0.566	-0.060	12 02 18 08
0.600	7	-39.681	0.836	0.566	-0.062	18 08 12 16
0.604	8	-39.681	0.836	0.566	-0.062	18 08 12 16
0.610	9	-54.032	0.873	0.979	-0.003	18 01 12 16
0.620	7	-54.065	0.873	0.980	-0.002	18 01 12 10
0.641	7	-54.065	0.873	0.980	-0.002	18 01 12 10
0.651	7	-54.065	0.873	0.980	-0.002	18 01 12 10
0.680	8	-54.070	0.871	0.984	-0.002	10 01 19 12
0.690	7	-54.190	0.871	0.983	0.000	19 01 11 10
0.732	5	-54.190	0.871	0.983	0.000	10 01 19 11
0.762	7	-54.190	0.871	0.983	0.000	10 01 19 11
0.765	6	-54.091	0.864	1.000	0.000	10 01 12 11
0.768	6	-54.091	0.864	1.000	0.000	10 01 12 11
0.770	5	-54.338	0.777	1.189	0.024	10 01 15 12
0.774	5	-56.683	0.786	1.254	0.032	10 15 20 12
0.776	5	-56.683	0.786	1.254	0.032	10 15 20 12
0.778	6	-58.548	0.808	1.274	0.034	10 03 20 12
0.800	6	-58.548	0.808	1.274	0.034	10 03 20 12
0.814	5	-58.548	0.808	1.274	0.034	10 03 20 12
0.824	6	-59.375	0.806	1.256	0.050	10 03 15 20
0.834	5	-58.543	0.793	1.305	0.038	10 03 15 12
0.903	5	-58.543	0.793	1.305	0.038	10 03 15 12
0.913	6	-58.543	0.793	1.305	0.038	15 03 12 10
0.950	7	-58.462	0.525	1.858	0.107	03 10 04 12

A execução para um valor de θ consome, em média, 6 centésimos de segundos.

6.2 EXEMPLO 2

Este exemplo é obtido pela aproximação de e^x por um polinômio de grau 3, cujos dados, arredondados para três casas decimais, são listados na Tabela 6.3.

Os resultados dos cálculos dos quantis são dados na Tabela 6.4.

Tabela 6.3

Exemplo 2

i	X				y
1	1	0.000	0.000	0.000	1.000
2	1	0.100	0.010	0.001	1.105
3	1	0.200	0.040	0.008	1.221
4	1	0.400	0.160	0.064	1.492
5	1	0.500	0.250	0.125	1.649
6	1	0.600	0.360	0.216	1.822
7	1	0.800	0.640	0.512	2.226
8	1	0.900	0.810	0.729	2.460
9	1	1.000	1.000	1.000	2.718

Tabela 6.4
Resultados para o Exemplo 2

θ	NUM ITER	SOLUÇÃO (QUANTILE)				PTOS. INTERP.
0.200	4	0.999	1.009	0.441	0.268	02 08 07 03
0.333	5	0.999	1.009	0.441	0.268	02 08 07 03
0.400	5	0.998	1.026	0.407	0.287	02 08 06 07
0.444	5	0.998	1.026	0.407	0.287	02 08 06 07
0.464	8	0.998	1.029	0.399	0.292	02 09 06 07
0.484	7	0.998	1.029	0.399	0.292	07 02 09 06
0.490	6	0.998	1.028	0.400	0.292	07 04 09 06
0.495	5	1.000	1.018	0.417	0.283	01 09 06 04
0.500	5	1.000	1.018	0.417	0.283	01 09 06 04
0.600	4	1.000	1.018	0.417	0.283	01 09 06 04
0.797	3	1.000	1.011	0.440	0.267	01 09 05 04
0.900	3	1.000	1.011	0.440	0.267	01 09 05 04

A execução para um valor de θ consome, em média, 3 centésimos de segundos.

6.3 EXEMPLO DE PÓS OTIMIZAÇÃO

Nesta seção apresentamos resultados de algumas perturbações feitas em θ_0 , tal que a solução ótima fica inalterada.

Esses resultados foram obtidos através da implementação da análise de pós otimização feita no Capítulo 5.

Os dados apresentados na Tabela 6.5 são referentes a análise de pós otimização realizada para o Exemplo 1.

Na Tabela 6.6 apresentamos os resultados obtidos para o Exemplo 2.

Tabela 6.5

Análise Pós Otimização para o Exemplo 1

θ_0	μ	$\theta = \theta_0 + \mu$
0.120	0.004	0.124
0.126	0.004	0.130
0.256	0.019	0.275
0.300	0.031	0.331
0.392	0.017	0.409
0.600	0.004	0.604
0.620	0.031	0.651
0.732	0.030	0.762
0.765	0.003	0.768
0.778	0.036	0.814
0.834	0.079	0.913

Tabela 6.6

Análise Pós Otimização para o Exemplo 2

θ_0	μ	$\theta = \theta_0 + \mu$
0.200	0.133	0.333
0.400	0.044	0.444
0.464	0.020	0.484
0.495	0.301	0.796

CONCLUSÕES

Neste trabalho o Problema de Aproximação Linear no L_1 e o Problema de Regressão Quantil foram resolvidos considerando-os como problemas de Programação Linear por Partes.

Através da especialização do Método Primal Simplex, aqui desenvolvido, conseguimos mostrar, de forma clara, que o Problema de Regressão Quantil consiste em uma extensão do Problema de Aproximação Linear no L_1 , pois o Problema Quantil é resolvido para qualquer $\theta \in (0,1)$ e o Problema no L_1 é obtido fazendo $\theta = 0.5$. Na literatura existente este fato é confirmado, porém não é mostrado com clareza.

Podemos verificar que a aplicação do nosso método leva-nos a soluções extremamente satisfatórias se comparadas às soluções obtidas por outros métodos da literatura. Devemos lembrar que essa comparação foi feita analisando apenas os valores numéricos e os pontos de interpolação, devido ao fato de não termos na literatura disponível dados suficientes para realizarmos comparações mais eficientes.

Além disso, apresentamos uma maneira simples de se realizar a análise de pós otimização uma vez determinada uma solução ótima. Tal forma não consta na literatura. Osborne [11] faz esta análise utilizando uma abordagem diferente, através de um algoritmo de homotopia linear por partes. Tal algoritmo possui duas partes: uma identifica o domínio de θ para o qual a solução atual é ótima, e a outra troca o plano quantil definindo a solução. Este segundo passo envolve a atualização dos dados.

É necessário ressaltar que os algoritmos foram desenvolvidos especificamente para os problemas tratados neste trabalho, o que é muito importante, já que os pacotes computacionais de Programação Linear não exploram a estrutura do problema, apresentando assim um custo computacional elevado.

BIBLIOGRAFIA

- [1] Abdelmalek, N.N., *On the Discrete Linear L_1 Approximation and L_1 Solutions of Overdetermined Linear Equations*. J. Approximation Theory, vol. 11, (1974), pp. 38-53.
- [2] Abdelmalek, N.N., *An Efficient Method for the Discrete Linear L_1 Approximation Problem*. Mathematics of Computation, vol. 29, n. 131, (1975), pp. 844-850.
- [3] Arenales, M.N., *Programação Linear: Novos Métodos e Alguns Problemas Particulares*. Tese de Doutorado, Unicamp - Fec Departamento de Engenharia Elétrica, (1984), 144p.
- [4] Arenales, M.N., *On Techniques for Curve Fitting Based on Minimum Absolute Deviations*. Atas da II Escola de Modelos de Regressão - Universidade Federal do Rio de Janeiro, (1991), pp. 06 -11.
- [5] Barrodale, I. & Roberts, F.D.K., *An Improved Algorithm for Discrete L_1 Linear Approximation*. SIAM J. Numer. Anal., vol. 10, n.5, (1973), pp. 839-848.
- [6] Barrodale, I. & Roberts, F.D.K., *Applications of Mathematical Programming to L_p Approximations*. New York, Academic Press, (1970), pp. 447-464.
- [7] Barrodale, I. & Young, A., *Algorithms for Best L_1 and L_∞ Linear Approximations on a Discrete Set*. Numer. Math., vol. 8, (1966), pp. 295-306.

- [8] Bassett, G.Jr. & Koenker, R., *An Empirical Quantile Function for Linear Models with iid Errors*. J. Amer. Statist. Assoc., vol. 77, n.378, (1982), pp. 407-415.
- [9] Bloomfield, P. & Steiger, W., *Least Absolute Deviations Curve - Fitting*. SIAM J. Sci. Stat. Comput., vol. 1, n.2, (1980), pp. 290-301.
- [10] Koenker, R. & Bassett, G.Jr., *Regression Quantiles*. Econometrica, vol. 46, n.1, (1978), pp. 33-50.
- [11] Osborne, M.R., *An Effective Method for Computing Regression Quantiles*. IMA J. Numer. Anal., vol. 12, (1992), pp. 151-166.
- [12] Talarico, R.M.P.M., *Regressão Linear Múltipla Através de Minimização da Soma dos Desvios Absolutos*. Tese de Mestrado, USP - ICMSC, (1984), 111p.
- [13] Wagner, H.M., *Linear Programming Techniques for Regression Analysis*. Amer. Statist. Assoc. J., vol.54, (1959), pp. 206-212.