

# Regressão pela Norma $L_p$ via Métodos de Pontos Interiores Primais Duais

Aurelio R. L. Oliveira, [aurelio@ime.unicamp.br](mailto:aurelio@ime.unicamp.br),  
Daniela R. Cantane, [dcantane@ime.unicamp.br](mailto:dcantane@ime.unicamp.br)

IMECC, UNICAMP,  
Praça Sérgio Buarque de Holanda, 651,  
13081-970, Campinas, Brasil

## Abstract

Os métodos de pontos interiores primaldual e preditor-corretor são desenvolvidos para o problema de regressão pela norma  $L_p$  e a estrutura matricial resultante é explorada objetivando uma implementação eficiente. Uma implementação em Matlab dos métodos desenvolvidos é comparada com implementação de um método de convergência global já existente. Testes computacionais indicam que os métodos de pontos interiores embora necessitem de mais iterações para convergir, obtém menor tempo computacional que o método existente pois possuem uma busca linear rápida.

**Palavras-chave:** métodos de pontos interiores, programação não linear, problema de regressão  $L_p$ .

# 1. Introdução

Desde o surgimento dos métodos de pontos interiores para otimização linear, códigos computacionais baseados nessas idéias vem se apresentando como alternativas eficientes para solução de problemas de grande porte [1].

O objetivo deste trabalho consiste no desenvolvimento dos métodos de pontos interiores para o problema de regressão pela norma  $L_p$  e na implementação eficiente do método desenvolvido.

O método IRLS *iteratively reweighted least-squares* [7] foi por muito tempo a única alternativa prática para a resolução deste problema. Mais recentemente, foi proposto um novo método que apresenta características similares aos métodos de pontos interiores [5]. Este método apresentou resultados computacionais superiores ao IRLS.

O método desenvolvido em [2] é um método de relaxação por coluna. A iteração básica é composta de  $n$  passos. No  $j$ -ésimo passo, somente  $x_j$  é modificado na tentativa de reduzir o valor da função objetivo, enquanto todas as outras variáveis são mantidas fixas. Resultados computacionais não são apresentados.

O método desenvolvido em [5] têm uma importante desvantagem: a busca linear é computacionalmente cara ao contrário dos métodos de pontos interiores aplicados a este problema, que obtém resultados computacionais superiores, repetindo o desempenho obtido na minimização pelas normas  $L_1$  e  $L_\infty$  em [8,9], respectivamente.

## 2. O Problema de Regressão pela Norma $L_p$

O problema de regressão  $\min_{x \in \mathbb{R}^n} \|A^t x - b\|_p^p$  onde  $A \in \mathbb{R}^{n \times m}$ ,  $b \in \mathbb{R}^m$  e  $m > n$ , tem inúmeras aplicações em diversas áreas de ciências e engenharias. As normas mais utilizadas são as normas 1, 2 e  $\infty$ . A norma-2 é muito popular, entre outros motivos, por permitir uma solução direta. Por sua vez a norma-1 permite diminuir o efeito de pontos discrepantes enquanto que a norma- $\infty$  garante proteção contra o pior caso. Os dois últimos problemas podem ser formulados por programação linear e os métodos de pontos interiores aplicados a estes problemas permitem a exploração da estrutura matricial do problema de forma bastante eficiente [8,9].

O objetivo deste trabalho consiste na aplicação de métodos de pontos interiores ao problema de regressão  $L_p$  com  $1 < p < 2$ :

$$\begin{aligned} &\text{minimize } \|r\|_p^p && (1) \\ &\text{sujeito a } Ax + r = b \end{aligned}$$

Definindo  $r = u - v$ , podemos reescrever o problema (1) da seguinte forma:

$$\begin{aligned} &\text{minimize } \sum_{i=1}^n (u_i + v_i)^p && (2) \\ &\text{sujeito a } Ax + u - v = b, \quad (u, v) \geq 0, \end{aligned}$$

pois sempre existe uma solução ótima tal que  $u = 0$  ou  $v = 0$ .

Com este modelo não é necessário trabalhar com valores absolutos e, portanto, temos uma função cuja primeira derivada é definida para qualquer ponto. Além disso, quando  $p = 1$  temos exatamente o modelo de regressão  $L_1$  resultando em um problema de otimização linear.

O método desenvolvido em [5] para o problema de regressão pela norma  $L_p$ , referido como GNCS, é um método de Newton globalizado com convergência superlinear, sob certas condições, para o problema da norma  $L_p$ . A cada iteração este método necessita resolver um sistema linear cuja matriz tem a forma  $A'DA$  onde D é uma matriz diagonal, e envolve uma busca linear da ordem de  $n^2$  no pior caso.

### 3. Métodos de Pontos Interiores

A função objetivo do problema (2) será denotada por  $\phi(u, v) = \sum_{i=1}^n (u_i + v_i)^p$ , o gradiente  $\nabla\phi(u, v)$  é dado por

$$g = \begin{bmatrix} g_u \\ g_v \end{bmatrix} \text{ e } \nabla^2\phi = \begin{bmatrix} \nabla g_u \\ \nabla g_v \end{bmatrix}, \text{ onde}$$

$$g_{u_i} = g_{v_i} = p(u_i + v_i)^{p-1}$$

e

$$\nabla g_{u_{ij}} = \nabla g_{v_{ij}} = \begin{cases} p(p-1) & \text{se } i = j, \\ 0, & \text{se } i \neq j \end{cases}$$

é uma matriz diagonal denotada por G.

Utilizando o método barreira logarítmica [4], temos:

$$\min \sum_{i=1}^n (u_i + v_i)^p - \mu \sum_{i=1}^n \ln(u_i) - \mu \sum_{i=1}^n \ln(v_i)$$

sa  $Ax + u - v - b = 0,$

onde  $\mu > 0$  é o parâmetro barreira ( $\mu \rightarrow 0$ ).

A Lagrangiana é dada por  $L = \sum_{i=1}^n (u_i + v_i)^p - \mu \sum_{i=1}^n \ln(u_i) - \mu \sum_{i=1}^n \ln(v_i) + y'(Ax + u - v - b)$ , onde  $y$  é o multiplicador de Lagrange.

Aplicando as condições de otimalidade obtemos:

$$\nabla_{(x,y,u,v)} L = \begin{bmatrix} A^t y \\ Ax + u - v - b \\ g - \mathbf{m}U^{-1} + y \\ g - \mathbf{m}V^{-1} - y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \quad (3)$$

onde U e V são matrizes diagonais cujos elementos diagonais são u e v respectivamente. Para obter um problema estritamente primal-dual definimos  $z_u = \mu U^{-1}e$  e  $z_v = \mu V^{-1}e$  [4]. Utilizando o Método de Newton, chegamos ao seguinte sistema linear:

$$A^t dy = A^t y \equiv r_1 \quad (4)$$

$$Adx + du - dv = -Ax - u + v + b \equiv r_2 \quad (5)$$

$$dy + Gdu + Gdv - dz_u = -g + z_u - y \equiv r_3 \quad (6)$$

$$-dy + Gdu + Gdv - dz_v = -g + z_v + y \equiv r_4 \quad (7)$$

$$Z_u du + Udz_u = -UZ_u e + \mu e \equiv r_5 \quad (8)$$

$$Z_v dv + V dz_v = -V Z_v e + \mu e \equiv r_6. \quad (9)$$

Através de eliminação de variáveis o sistema (4-9) se reduz a

$$\begin{aligned} dx &= (A^t DA)^{-1} r, & du &= D_u^{-1} (r_3 - dy - G dv + U^{-1} r_5), \\ dv &= (D_v - G^2 D_u^{-1})^{-1} (dy + \hat{r}), & dy &= D(\hat{r} - A dx), \\ dz_u &= U^{-1} (r_5 - Z_u du), & dz_v &= V^{-1} (r_6 - Z_v dv), \end{aligned}$$

onde  $r = r_1 + A^t D \hat{r}$ ,  $D^{-1} = D_u^{-1} + (D_v - G^2 D_u^{-1})^{-1} (I + D_u^{-1} G)^2$ ,

$D_u = (G + U^{-1} Z_u)$ ,  $D_v = (G + V^{-1} Z_v)$ ,  $\hat{r} = r_4 + D_u^{-1} G (r_3 + U^{-1} r_5) + V^{-1} r_6$  e  $\bar{r} = r_2 - D_u^{-1} r_3 + (D_v - G^2 D_u^{-1})^{-1} (I + D_u^{-1} G) \hat{r}$ .

O passo  $\alpha$  é calculado mantendo  $u$  e  $v$  estritamente positivas:

$$\alpha = \min \left\{ \tau \min_{du_i < 0} \left( -\frac{u_i}{du_i} \right), \tau \min_{dv_i < 0} \left( -\frac{v_i}{dv_i} \right), 1 \right\}, \text{ onde } \tau = .99999\%.$$

Conhecendo as direções e os passos, todas as variáveis do problema podem ser atualizadas por:

$$\begin{aligned} x^{k+1} &= x^k + \alpha dx, & u^{k+1} &= u^k + \alpha du, & v^{k+1} &= v^k + \alpha dv, \\ y^{k+1} &= y^k + \alpha dy, & z_u^{k+1} &= z_u^k + \alpha dz_u, & z_v^{k+1} &= z_v^k + \alpha dz_v. \end{aligned}$$

A atualização do parâmetro barreira é dada por  $\mu = \gamma \phi$ , onde  $\phi = \frac{1}{m\sqrt{m}}$  e  $\gamma = u^t z_u + v^t z_v$ .

### 3.1 Método Preditor-Corretor

No método preditor-corretor, primeiramente tomamos a direção afim em que o parâmetro barreira  $\mu = 0$ . Estas direções são encontradas resolvendo um sistema linear semelhante ao obtido para o método primal-dual. Agora, podemos impor um valor para o parâmetro barreira  $\mu$  e resolver um segundo sistema linear com as restrições referentes à complementaridade incluindo  $\mu$  e a correção não linear [6].

O critério de convergência é baseado nas condições de otimalidade:

$$N = \frac{\|\nabla L\|}{(1 + \|x\| + \|u\| + \|v\| + \|y\| + \|z_u\| + \|z_v\|)(2n)} \leq \epsilon. \text{ Estabelecemos } \epsilon = 10^{-6}.$$

O ponto inicial é calculado baseado nas idéias desenvolvidas em [3],

$$\begin{aligned} x^0 &= (A^t A)^{-1} A^t b, \quad r^0 = b - A x^0, & y^0 &= \frac{\kappa r^0}{\|r^0\|_\infty} \\ u_i^0 &= \begin{cases} \frac{\lambda + 1}{2} r_i^0 & \text{se } r_i^0 > 0, \\ -\frac{\lambda - 1}{2} r_i^0 & \text{caso contrário,} \end{cases} & v_i^0 &= \begin{cases} \frac{\lambda - 1}{2} r_i^0 & \text{se } r_i^0 > 0, \\ -\frac{\lambda + 1}{2} r_i^0 & \text{caso contrário,} \end{cases} \end{aligned}$$

Em [3]  $u$  e  $v$  não são definidos. Foi proposto em [9] esta escolha para eles de modo que satisfaça a relação  $u^0 + v^0 = |\lambda r^0|$ . Assim, se  $\lambda$  é  $O(1)$ , ambos  $u^0$  e  $v^0$  são da mesma ordem de  $r^0$ . Estabelecemos  $\kappa = 0,97\%$ .

## 4. Resultados Computacionais e Conclusões

O sistema linear obtido em nosso trabalho é simétrico e definido positivo e assim pode ser resolvido pela fatoração de Cholesky. Obtemos, então, um sistema muito menor que o sistema original (4-9) e da mesma dimensão dos sistemas resolvidos pelo método GNCS [5].

No método preditor-corretor a solução do segundo sistema linear pode ser obtida utilizando a fatoração de Cholesky utilizada na resolução do primeiro sistema linear.

O método de pontos interiores primal-dual e preditor-corretor, referidos como MPIPDP e MPIPDC e o método GNCS foram implementados em Matlab 6.0, com o sistema operacional Solaris em uma estação Sun Blade 100.

Utilizamos um problema real de grande porte para comparar o desempenho dos métodos. O conjunto de dados compreende valores dos juros diários *Prime Rate* ao longo de 40 anos, finais-de-semanas e feriados não considerados, totalizando 10958 valores observados. Para evitar problemas de estabilidade numérica dos métodos, os dados (observados) foram normalizados no intervalo [0,1]. A Tabela 1 mostra os resultados obtidos pelos métodos de pontos interiores e GNCS com  $p$  variando de 1,1 a 2 e ajuste dos dados por uma reta.

p	MPIPDP		MPIPDC		GNCS	
	Iter	Tempo	Iter	Tempo	Iter	Tempo
1,1	19	1,13e4	25	2,10e4	5	1,37e4
1,2	15	1,14e4	15	1,41e4	5	1,38e4
1,3	13	1,11e4	12	8,98e3	4	8,11e3
1,4	11	9,13e3	10	7,78e3	4	7,87e3
1,5	11	1,23e4	8	7,02e3	4	8,56e3
1,6	10	8,76e3	7	5,10e3	3	6,91e3
1,7	10	6,99e3	7	6,67e3	3	6,63e3
1,8	9	7,63e3	7	6,38e3	3	6,71e3
1,9	6	6,16e3	7	5,23e3	2	5,32e3
2,0	4	3,93e3	4	3,09e3	2	3,37e3

**Tabela 1:** Número de Iterações e Tempo Computacional em segundos.

Podemos verificar que o método GNCS sistematicamente obtém menor número de iterações enquanto que os métodos de pontos interiores são superiores em relação ao tempo computacional sendo o método primal-dual melhor para  $p$  próximo de 1 e o método preditor-corretor com melhor desempenho para  $p$  próximo de 2. O melhor desempenho do método primal-dual em relação ao preditor-corretor para os menores  $p$  se explica pelo número de iterações que é maior para o preditor-corretor. Vale observar que os valores de função objetivo obtidos estão próximos para os três métodos.

O melhor desempenho dos métodos de pontos interiores se deve ao fato da busca linear destes métodos ser rápida, da ordem de  $n$  operações de ponto flutuante em comparação com a busca linear do método GNCS que é da ordem de  $n^2$ . Assim, o ganho em iterações não compensa o esforço computacional utilizado no cálculo do tamanho do passo mesmo em casos onde o número de iterações obtido pelo GNCS é menos da metade que o número de iterações do método de pontos interiores.

É importante observar que o método GNCS não converge após mais de cem iterações quando utilizamos o critério de parada apresentado em [5].

Testes computacionais realizados com as funções utilizadas em [3,5] e polinômios de grau superior confirmam as conclusões deste experimento.

## Agrade cimentos

Este trabalho tem apoio parcial da FAPESP (Fundação de Amparo à Pesquisa do Estado de São Paulo) e CNPq (Conselho Nacional Brasileiro de Desenvolvimento Científico e Tecnológico).

## Referências

- [1] I. Adler, M. G. C. Resende, G. Veiga, and N. Karmarkar *An Implementation of Karmarkar's Algorithm for Linear Programming*, Mathematical Programming, 44 (1989), pp. 297-335.
- [2] A. Dax and B. Berkowitz, *Column relaxation methods for least norm problems*, SIAM J. Sci. Stat. Comput., 11 (1990), 975-989.
- [3] T. F. Coleman and Y. Li, *A globally and quadratically convergent affine scaling method for linear  $l_1$  problems*, Math. Programming, 56 (1992), pp. 189-222.
- [4] A. S. El-Bakry, R. A. Tapia, T. Tsuchiya and Y. Zhang *On the Formulation and the Theory of the Newton Interior-Point Method for Nonlinear Programming*, J. of Optimization Theory and Applications, 89 (1996), pp. 507-541.
- [5] Y. Li, *A globally convergent method for  $l_1$  problems*, SIAM J. Optimization, 3 (1993), pp. 609-629.
- [6] S. Mehrotra, *On the implementation of a primal-dual interior point method*, SIAM Journal on Optimization, 2 (1992), 575-601.
- [7] G. Merle and H. Späth, *Computational experience with discrete  $l_1$  approximation*, Computing, 12 (1974), pp. 315-321.
- [8] A. R. L. Oliveira and C. Lyra, *Interior point methods for the polynomial  $l_\infty$  fitting problems*, Internacional Transactions in Operational Research, 11 (2004), pp. 309-322.
- [9] A. R. L. Oliveira, M. A. Nascimento and C. Lyra, *Efficient implementation and benchmark of interior point methods for the polynomial  $l_1$  fitting problems*, Statistics & Data Analysis, 35 (2000), pp. 119-135.