Proceeding Series of the Brazilian Society of Computational and Applied Mathematics

Aproximação de Dados Através do Processo de Mínimos Quadrados

Ítalo F. Nunes da Paz[1]

Centro de Tecnologia, Departamento de Engenharia Mecânica, UFRN, Natal, RN

Fabiana T. Santana[2]

Escola de Ciências e Tecnologia, UFRN, Natal, RN

Resumo. Neste trabalho foi feito um estudo detalhado de aproximações por Mínimos Quadrados no espaço coluna de uma matriz A. Em particular, foi feito o desenvolvimento matemático para se obter a melhor função que se ajusta aos pontos (x_i, y_i), obtidos experimentalmente, utilizando o produto interno euclidiano em R^m. Tal função é denominada aproximação de mínimos quadrados. Por fim, foi aplicado os conceitos estudados na obtenção da estimativa da população brasileira para o ano de 2020.

Palavras-chave. Mínimos Quadrados, Ajuste de Curvas, Produto Interno.

1 Introdução

Muitas vezes, os sistemas lineares oriundos de um modelo matemático podem não apresentar solução devido às aproximações feitas durante a modelagem do problema. Nessas situações, a Álgebra Linear fornece a melhor solução para o problema através dos mínimos quadrados, [2,3]. Neste trabalho, dado um conjunto de pontos $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$ a curva que mais se aproxima é obtida minimizando o erro de aproximação. Isto é, a quantidade $|| \vec{e} || = ||A \vec{x} - \vec{b} ||$ pode ser vista como uma medida de quão próximo o vetor solução $\vec{x}$ está da solução do sistema. A regressão por mínimos quadrados tem importância especial e aparece em várias aplicações físicas obtidas experimentalmete onde se possuem um conjunto de pontos, como por exemplo em determinação de constante de molas, determinação de velocidade, dentre outros [2].

2 Obtenção de Soluções por Mínimos Quadrados e Aplicação

No caso de mínimos quadrados para um conjunto de dados, a projeção do vetor $\vec{b}$ sobre o espaço coluna da matriz dos coeficientes A corresponde a encontrar uma solução

[1]italonpaz@gmail.com
[2]fabianatsantana@gmail.com

para $A\vec{x} = \vec{b}$ com erro mínimo. A abordagem utilizada foi alterar o problema de modo que não tenhamos que exigir que a equação matricial $A\vec{x} = \vec{b}$ seja satisfeita, em vez disso, procura-se um vetor solução $\vec{x}$ substituindo $\vec{b}$ por um vetor que esteja tão próximo de $A\vec{x}$ quanto possível. Tal vetor é a projeção ortogonal de $\vec{b}$ no espaço coluna de A, [1].

Pelo Teorema da Melhor Aproximação, se W for o espaço coluna de A com produto interno euclidiano de R^m, então a $proj_W\vec{b} \in W$ satisfaz $||\vec{b} - proj_W\vec{b}|| < ||\vec{b} - \vec{w}||$ para qualquer vetor $\vec{w} \in W$ distinto de $proj_W\vec{b}$. Este teorema, cuja demonstração pode ser vista em [2], garante que a projeção de $\vec{b}$ em W é a função que melhor se aproxima de $\vec{b} \notin W$ em W. Por outro lado, como $\vec{b} - A\vec{x} = \vec{b} - proj_W\vec{b}$ é ortogonal à W e W é o espaço coluna de A, segue que $\vec{b} - A\vec{x}$ está no espaço nulo de A^T, pois o espaço nulo de A^T e o espaço coluna são complementos ortogonais. Desse modo, uma solução de mínimos quadrados de $A\vec{x} = \vec{b}$ deve satisfazer $A^T(\vec{b} - A\vec{x}) = \vec{0}$ ou $A^TA\vec{x} = A^T\vec{b}$, que é chamado de sistema normal associado à $A\vec{x} = \vec{b}$. Assim, o problema de encontrar uma solução de míminos quadrados foi reduzido a encontrar uma solução exata do sistema normal associado.

O estudo feito foi aplicado em obter uma estimativa da população brasileira no ano de 2020. Para isso, com base nos dados do Instituto Brasileiro de Geografia e Estatística, obtivemos o conjunto de dados (x_i, y_i), onde os dados x_i representam o ano de cada pesquisa realizada e os dados y_i representa a quantidade de milhões de habitantes na respectiva data da pesquisa.

Ao aproximarmos o problema por polinômios de grau elevado, como o de grau 6, $p_6(x) = 37762892 - 107152.51x + 112.44565x^2 - 0.0507168x^3 + 0.0000059x^4 + 2.28210^{-9}x^5 - 5.55210^{-13}x^6$, percebemos que os coeficientes de x^4 já começa a ter relevância pequena em comparação ao de x^3. No caso do polinômio de grau 4, $p_4(x) = -49025107.0 + 100561.53x - 77.330493x^2 + 0.0264214x^3 - 0.0000034x^4$, o coeficiente de x^3 torna-se pequeno, mas não tanto quando comparado com os outros polinômios de graus maiores. Logo, a curva que apresenta uma aproximação satisfatoria, relevância dos coeficientes e comportamento gráfico satisfatório, aos dados apresentados é o polinômio de grau 4, tendo uma população estimada de 203.07 milhões de habitantes.

Agradecimentos

Este trabalho foi financiado pela PROPESQ-UFRN.

Referências

[1] H. Anton e R. C. Busby, Álgebra linear contemporânea, Porto Alegre: Bookman, (2006).

[2] H. Anton e C. Rorres, Álgebra linear com aplicações, Porto Alegre: Bookman, (2001).

[3] S. C. Chapra e R. P. Canale, Métodos numéricos para engenharia, São Paulo: McGraw-Hill, (2008).

