Proceeding Series of the Brazilian Society of Computational and Applied Mathematics

Decomposição em Valores Singulares e Técnicas de Compressão de Dados

Giovanna Castello de Andrade[1]
Estudante de Graduação - Instituto de Matemática, Estatística e Computação Científica, UNICAMP, Campinas, SP
Sandra Augusta Santos[2]
Departamento de Matemática Aplicada, Instituto de Matemática, Estatística e Computação Científica, UNICAMP, Campinas, SP

A decomposição em valores singulares, conhecida como SVD, do inglês *Singular Value Decomposition*, é uma ferramenta importante da análise matricial. Sua determinação eficiente em problemas de grande porte é um tema relevante e atual de investigação, sobretudo considerando que o volume de dados a serem armazenados, processados ou analisados cresce dia a dia. De acordo com [1], a SVD "ordena" as informações contidas na matriz, de forma que seu "conteúdo dominante" torna-se visível. Esta propriedade torna a decomposição extremamente útil em áreas como a de compressão de dados. Em específico, destacamos a compressão de imagens, que busca reduzir a quantidade de dados necessários para representar uma imagem digital.

Seja $A \in \mathbb{R}^{m\times n}$ uma matriz de posto r contendo dados do problema em questão, sua decomposição em valores singulares reduzida pode ser escrita como $A = U\Sigma V^t$, em que $U \in \mathbb{R}^{m\times n}$ é ortogonal, $V \in \mathbb{R}^{n\times n}$ é ortogonal, e $\Sigma \in \mathbb{R}^{n\times n}$ é diagonal, com entradas $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$ chamadas de valores singulares; ou através da soma

$$A = \sum_{i=1}^{n} \sigma_i u_i v_i^t, \tag{1}$$

em que cada um dos termos é uma matriz de posto um.

Uma estratégia possível para a compressão de dados é a análise de componentes principais, conhecida como PCA, do inglês *Principal Components Analysis*. Esta técnica se apoia na equação (1) e na possibilidade de truncar tal soma após k termos, obtendo, pelas propriedades da SVD, a melhor aproximação possível de posto k para a matriz A.

Para a determinação numérica eficiente da SVD de uma matriz, G. Golub e W. Kahan propõem em [2] a prévia bidiagonalização da matriz através de transformações ortogonais. Em [6], o autor apresenta três formas diferentes de se realizar este processo, todas através de refletores de Householder: Bidiagonalização de Golub-Kahan, Bidiagonalização

[1]g168633@unicamp.br
[2]sandra@ime.unicamp.br

de Lawson-Hanson-Chan e Método dos Três Passos. No entanto, considerando problemas de grande porte, o uso de refletores representa um alto custo computacional e, portanto, torna-se vantajoso o método da Bidiagonalização de Lanczos, descrito em [4] pelos criadores do *software* `PROPACK`, e que computa as entradas da matriz de forma direta a partir de relações de recorrências. Com esta etapa concluída, o problema de computar os valores e vetores singulares da matriz pode ser reduzido ao de encontrar os autovalores e autovetores de uma matriz simétrica a ela associada.

A fundamentação teórica deste projeto contemplou o estudo do método de Lanczos para determinação de autovalores e autovetores extremos de matrizes simétricas, com base em [3]. A partir da realização de testes experimentais em `MATLAB`, foi também observado e estudado o aparecimento dos "autovalores fantasmas", descrito em [5], e a possibilidade do uso da reortogonalização para minimizar seus efeitos. O método da Bidiagonalização de Lanczos é obtido de forma análoga ao Método de Lanczos e, portanto, os conhecimentos adquiridos na etapa inicial serviram como base para seu estudo, tanto em uma abordagem teórica, quanto computacional. Do ponto de vista prático, investigamos como a quantidade de componentes principais em matrizes associadas a imagens pode interferir na qualidade da imagem compactada obtida.

Agradecimentos

Agradecemos à FAPESP pelo apoio financeiro através do processo 2018/07318-0.

Referências

[1] L. Eldén. *Matrix Methods in Data Mining and Pattern Recognition.* Society for Industrial and Applied Mathematics, Philadelphia, 2007.

[2] G. Golub, W. Kahan. *Calculating the Singular Values and Pseudo-Inverse of a Matrix* Journal of the Society for Industrial and Applied Mathematics: Series B, Numerical Analysis, Volume 2, Issue 2, 205-224, 1965.

[3] G. Golub, C. F. Van Loan. *Matrix Computations*, 4th ed. The Johns Hopkins University Press, Baltimore 2013.

[4] R. M. Larsen *Computing the SVD for large and sparse matrices*, Stanford University, 2000. Disponível em: http://sun.stanford.edu/ rmunk/PROPACK/.

[5] C.C. Paige *The Computation of Eigenvalues and Eigenvectors of Very Large Sparse Matrices*, PhD thesis, University of London, 1971.

[6] L. N. Trefethen, D. Bau, III. *Numerical Linear Algebra*, Philadelphia: Society for Industrial and Applied Mathematics, 1997.