# Scientific Computing in Data Analysis

Prof. Efstratios Gallopoulos[1]

[1] *University of Patras*
*Department of Computer Engineering & Informatics*
*Email: stratis@ceid.upatras.gr*

**Timetable:** 10 hours. The first lecture will be on Tuesday, June 7, 2011, Torre Archimede, Room 2BC/30. Class meets Tuesday to Friday:
June 7, 2011, 15:00-17:00
June 8, 2011, 10:30-12:30, 15:00-17:00
June 9, 2011, 15:00-17:00
June 10, 2011, 10:30-12:30

**Course requirements:** Basic undergraduate knowledge of: Linear Algebra, Structure of Computers and Computations (Computer Architecture, Algorithms and Data Structures). Some knowledge of MATLAB.

**Examination and grading:**

**Aim:** This course is designed to introduce Scientific Computing, with emphasis on matrix methods at a graduate level, for students with interests in Data Analysis.

**Course contents:**
The course is divided into three parts with a choice of topics as described below:

1. What is Scientific Computing. Models: Discrete, arithmetic and computational. The interaction of computer architecture, compilers and algorithm design. Problem solving environments. Measuring performance and benchmarks. Computational kernels. The role of matrix computations in scientific computing. Floating-point arithmetic and error analysis. Fundamental problems in matrix computations. Structure in matrix computations. Sparse matrix technology.
2. The linear algebra in Information Retrieval (IR). Dimensionality reduction, clustering, retrieval. and text classification. Matrix factorizations and low rank matrix approximation: The Wedderburn-Guttman framework, SVD and SDD. The LSI model. Elements of positive matrix theory. Nonnegative rank factorization and approximation. Clustering and representatives in matrix approximation methods. The MATLAB TMG toolbox and applications. Tensor methods for data representation in IR and HOSVD.
3. Methods in Web IR: Ranking the Web.Google's PageRank.
4. Approximation methods for very large matrices and the CUR decomposition. Application in uncertainty quantification.

**References**

[1] Lars Elden, *Matrix Methods in Data Mining and Pattern Recognition*, SIAM, 2007.

[2] A. Langville and C. Meyer, *Google's PageRank and beyond*, Princeton University Press, 2006.

[3] C. Ueberhuber, *Numerical Computing*, vol. 1, Springer, 2008.

[4] Online course notes and papers.