SSM
Scuola Superiore Meridionale

Course title: Numerical Methods for Data Mining

Duration: 24

PhD Program: MERC

Name and Contact details of unit organizer(s):

Dr. Dimitrios Patsatzis, Prof. Constantinos Siettos
Affiliation(s): Scuola Superiore Meridionale, Dipartimento di Matematica e Applicazioni "Renato Caccioppoli"
Email: d.patsatzis@ssmeridionale.it, constantinos.siettos@unina.it

Course Description:

The objective of the course is to provide a solid theoretical and computational background on various mathematical methods and numerical analysis methods for the analysis of the dynamics of complex nonlinear large-scale models with a wide range of applications (socioeconomic, neuroscience, epidemiology, finance). Starting from linear algebra and topological spaces, emphasis is placed on least square optimization and regularization, as well as the problem of data embedding, that is finding latent spaces (manifolds) to accurately represent the properties of the data and their similarities. Building on manifold learning techniques, the course also provides an introduction for modelling and analysing large-scale problems with the use of Equation free framework and Matrix-free methods. Pseudocodes and algorithms are also provided, in order for the students to build their own libraries of algorithmic tools for the analysis of large-scale, complex dynamical systems.

Syllabus:

A. Matrix algebra for the numerical analysis of complex systems.
- The fundamental theorem of linear algebra. Least squares estimation. The problem of regularization.
B. The problem of data embedding
- SVD, PCA, ICA, MDS, Kernel PCA, ISOMAP, Diffusion Maps
C. Modelling
- The General Linear Model and the Generalized Linear Model, Gaussian Process
- The Dynamic Mode Decomposition method and the Koopman operator theory
D. Analysis of Complex Systems
- The Equation-Free multiscale framework for the analysis of multiscale complex systems
- Matrix-free methods for large-scale problems. Krylov sub-spaces. GMRES and Arnoldi methods.

Assessment:

This will be a hands-one course. Based on the provided pseudocodes and algorithms, the students will develop their own library of numerical algorithms for modelling and solving large-scale complex problems. The assessment and cores will be based on a set of assignments, to be delivered as short reports throughout the bimester, and a short final oral exam.

Suggested reading and online resources:

1. Wang, J, Geometric Structure of High-Dimensional Data and Dimensionality Reduction, Springer-Verlag, 2012
2. Kelley, CT, Iterative Methods for Linear and Nonlinear Equations, SIAM, 1995. https://archive.siam.org/books/textbooks/fr16_book.pdf
3. Karlin, S, An introduction to Stochastic Modelling, Academic Press, 1998.
4. Dobson, A, Barnett, A, Introduction to Generalized Linear Models, Chapman and Hall/CRC2008
5. Notes and papers given to the students.