Department of Mathematical Sciences, University of Essex

Research Group: Data Science

PhD scholarship

https://www.jobs.ac.uk/job/BTR972/department-of-mathematical-sciences-doctoral-scholarships-mathematics-mphil-phd
(closes 28 August 2019)

To apply for the **Data Science scholarship** an applicant needs to have an offer as MPhil/PhD student by one of the supervisors of following PhD projects 2019-2022 of the Department of Mathematical Sciences. First you will need to apply for your PhD through the online portal https://www1.essex.ac.uk/pgapply/login.aspx as normal. Afterwards you follow the guidance under https://www.essex.ac.uk/departments/mathematical-sciences/scholarships-and-funding

PhD Projects 2019-2020

Supervisor: Joe Bailey, Lecturer in Environmetrics
Topic: to be announced
Email: jbailef@essex.ac.uk

Supervisor: Dr Yanchun Bao, Lecturer in Data Science and Statistics
Topic: to be announced
Email: ybaoa@essex.ac.uk *(Dr Bao joins the department 1 September 2019)*

Supervisor: Dr Hongsheng Dai, Reader in Statistics
Topic: Bayesian Fusion
Email: hdaia@essex.ac.uk

Supervisor: Dr Stella Hadjiantoni, Lecturer in Data Science and Statistics
Topic: Numerical methods for the estimation of high-dimensional linear regression models
Email: tba *(Dr Hadjiantoni joins the department 1 September 2019)*

Supervisor: Dr Andrew Harrison, Senior Lecturer in Data Science
Topic: Improving the analysis of Hi-C experiments on genome folding (to be co-supervised by Dr Radu Zabet, School of Life Sciences)
Email: harry@essex.ac.uk

Supervisor: Dr Haslifah Hashim, Lecturer in Actuarial Science
Topic: to be announced
Email: hhashim@essex.ac.uk

Supervisor: Dr Junlei Hu, Lecturer in Actuarial Science
Topic: to be announced
Email: j.hu@essex.ac.uk

Supervisor: Professor Berthold Lausen, Professor of Data Science
Topic: to be announced
Email: blausen@essex.ac.uk

Supervisor: Dr Fanlin Meng, Lecturer in Data Science
Topic: to be announced
Email: fanlin.meng@essex.ac.uk

Supervisor: Professor Abdel Salhi, Professor of Operational Research
Topic: Timetabling optimization
Email: as@essex.ac.uk

Supervisor: Dr Spyros Vrontos, Senior Lecturer in Actuarial Science
Topic: to be announced
Email: svrontos@essex.ac.uk

Supervisor: Dr Jackie Wong, Lecturer in Actuarial Science
Topic: to be announced
Email: jw19203@essex.ac.uk

Supervisor: Dr Xinan Yang, Senior Lecturer in Operational Research
Topic: to be announced
Email: xyangk@essex.ac.uk

Bayesian Fusion

Supervisor: Dr Hongsheng Dai, Reader in Statistics
Email: hdaia@essex.ac.uk

Spatial statistics are involved in many different research disciplines, such as knife crime studies in social sciences, geographic pollution studies in environmental studies and studies of the placement of galaxies in astronomy. One big challenge in spatial statistics is the ultra-high dimensionality of the data. Existing statistical methods in spatial analysis are still in early development and can only deal with spatial data with relatively lower dimension. Typical methods include: Generalised Linear Mixture models (GLMM) and Gaussian (Markov) random field models. However, all these existing methods cannot deal with high-dimensional data (1000-by-1000 spatial matrix or larger). Even for much lower dimensional (500-by-500), approximated posterior or likelihood functions have be used, however, it is usually difficult to justify the accuracy of the approximation.

This project will develop more efficient and practical algorithms, via Hermite polynomial expansions. It will also develop a method via the combination of Hermite polynomials and Bayesian fusion.

Future outcomes: 1. Two four-star REF papers. 2. Collaboration with ISER and Essex Police for landscape decision analysis and (knife) crime study; also possible funding from Google.

Numerical methods for the estimation of high-dimensional linear regression models

Supervisor: Dr Stella Hadjiantoni, Lecturer in Data Science and Statistics
Email: tba *(Dr Hadjiantoni joins the department 1 September 2019)*

Recursive estimation is an essential procedure when the model needs to be re-estimated after modifying a big data set by adding or deleting examples (observations) or features (variables), in regression diagnostics and in forecasting exercises, to name but a few. Since large-scale models and big data sets are intractable to implement and their use implies a high computational cost, it is computationally infeasible to sequentially estimate afresh such models when only a small number of examples or features has been altered. Instead, recursive methods which utilise efficiently previous computations should be employed.

This research project will investigate new numerical methods to tackle computationally intensive problems arising in data science, statistics and computer science. The aim is to develop theoretically sound numerical methods for the efficient and robust estimation of high-dimensional or large-scale multivariate linear models.

A starting point for this project would be to continue current research work on the recursive estimation of multivariate linear models when the original dataset is amended with the inclusion of new observations (examples) and/or the exclusion of old ones. The recursive estimation of the vector autoregressive and seemingly unrelated regressions models when the number of features is larger than the number of examples will be examined. Within the context of high-dimensionality, the project can then extend current research to develop new advanced numerical methods and algorithms for regularization, model selection, forecasting, nowcasting and data reduction. Parallelization of existing algorithms will also be studied and the development of new methods based on parallel algorithms will be investigated.

As matrices are used for the representation and statistical analysis of data, it is expected that the data matrix structures will be studied and their properties will be investigated. The new numerical methods will focus on using tools from numerical linear algebra. A possible application of the new methods is the estimation of high-dimensional network models.

Hadjiantoni, S. and Kontoghiorghes, E.J. (2017). Estimating large-scale general linear and seemingly unrelated regressions models after deleting observations. Statistics and Computing, 27(2):349- 361.

Hadjiantoni, S. and Kontoghiorghes, E.J. (2018). A recursive three-stage least squares method for large-scale systems of simultaneous equations. Linear Algebra and its Applications, 536:210-227.

Improving the analysis of Hi-C experiments on genome folding

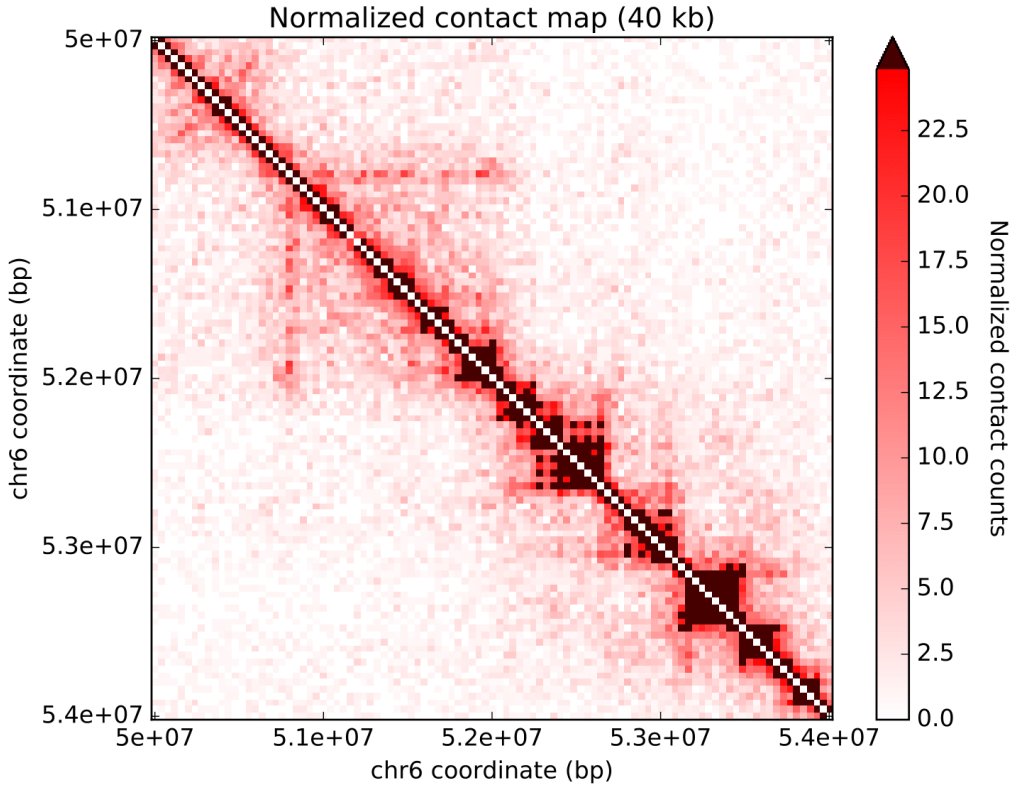
Supervisors
Dr Andrew Harrison (Department of Mathematical Sciences), harry@essex.ac.uk
Dr Radu Zabet (School of Life Sciences), nzabet@essex.ac.uk

Project

DNA in the nuclei of cells is tightly packed. Understanding how this packing is modulated in different cell types, across different organisms and between healthy and disease conditions will provide insights into how genes are regulated across the genome.

Hi-C is an experimental technique for measuring which regions of the genome are in close spatial proximity. The data from Hi-C experiments is large and requires the development of statistical and computational techniques to fully extract all the signals of interest in the data.

The focus of this project is to develop methods for identifying horizontal/vertical lines seen in Hi-C genomic interaction matrices. An example of such a line can be seen starting at genomic position 5.08e+07 and running to 5.2e+07. To identify such lines will require careful treatment of the background signals. The ultimate aim of the research will be to understand the biological significance of the lines and to identify whether lines seen in related experiments have a common cause.



Timetabling optimization

Supervisor: Professor Abdel Salhi, Professor of Operational Research
Email: as@essex.ac.uk

University timetabling is an intractable problem. Unfortunately, it has to be solved frequently, particularly in education institutions. The problem is to schedule teaching, learning and assessment slots, allocate to them relevant staff and appropriate premises, all under reasonable timetabling constraints. Some of these can be treated as soft constraints but many others have to be hard. Note that it can be challenging to tell which constraint should fall in which category, and the impact of adding/removing constraints on the resulting timetables.
At the moment the timetabling office uses a commercial package to find a baseline solution, which takes days. It is updated manually to satisfy special requirements. This is not efficient, and may lead to timetables which are substantially far from the optimum. Last minute requirements is also dealt with which reduces efficiency.
The aims of the project are:
1. To build a general model for the timetabling problem of Essex University;
2. To solve it using existing or new heuristics to generate good timetables which are quantifiably better than current ones;
3. To generate timetables quickly, i.e. in a shorter time than current practice;
4. To handle last minute requests in a more systematic way;
5. To categorise constraints.