**PhD Project Proposal**

School of Electronics, Electrical Engineering and Computer Science

|  |
| --- |
| **Proposed Project Title: A Random Matrix Inference Framework for Big-Data Analytics** |
| **Principal Supervisor: Dr. David Morales Second Supervisor: Dr. Michalis Matthaiou** |
| **Project Description:** A fundamental problem in statistical signal processing is to understand the interactions and relationships between variables (signals) from large-dimensional datasets. This is very important in a myriad of applications in fields such as financial engineering, wireless communications, or bioinformatics. For example:   * In the financial context, certain assets (stocks) are influenced by others, e.g., we would expect that stocks in the automobile sector are coupled (to some degree) with those in oil or energy industries, and this is critical for investors who want to maximize the return or minimize the risk of their investment. This is an important problem in financial engineering, known as *portfolio optimization*, i.e., the allocation of wealth to a selection of assets (stocks) in order to minimize the risk or maximize the return of the investment. * In wireless communications, to deliver fast data rates and reliable links, optimal multi-antenna filters require certain statistical knowledge of the coupling between signals. With emerging wireless networks becoming increasingly large, this needs to be estimated from large-dimensional datasets. * In bioinformatics, an important problem is to infer the links (connections) between genes or proteins from high-dimensional sequence data. Having accurate knowledge of these interactions can help e.g., to design effective vaccines against viruses like HIV, or to understand the underlying mechanisms behind cancer.   Key to resolve these problems is to accurately estimate covariance matrices with become larger than ever before in modern applications (like the ones mentioned above). The traditional estimator, known as the sample covariance matrix (SCM) is known to yield a very poor estimate when the number of variables (e.g., stocks, wireless signals, genes) is very large, and the number of available training samples is limited.  In this project, we will investigate high-dimensional solutions for covariance estimation, building upon state-of-the-art solutions and aiming at developing solutions with improved performance. These will then be applied to some of the relevant problems above, e.g., to design optimal financial portfolios.  This project is best suited to a student with an interest in random matrix theory and its cross-disciplinary applications. Advanced concepts of statistics and probability will be required, and the student should be confident in the use of MATLAB (or similar) simulation platforms to test and validate the theoretical findings.  The research will typically involve collaborations with renowned and internationally-leading researchers from Stanford University, Centrale Supelec, and the Hong Kong University of Science and Technology; it is expected that there will be opportunities for the student to engage in these collaborative activities and to visit some of these collaborators. |
| **Contact details**  Supervisor Name: David Morales  Tel: +44 (0)28 9097 1736 Email: d.morales@qub.ac.uk  QUB Address: Room 01/25, ECIT Institute  NI Science Park, Queens Rd,  Belfast BT3 9DT |