IITB-Monash Research Academy

An Indian-Australian research partnership

Indian Institute of Technology Bombay | MONASH University

| | | |
|---|---|---|
| **Project Title:** | Data Science Algorithms for Robust Debiased High-dimensional Sparse Regression | |
| **Project Number** | **IMURA1238** | |
| **Monash Main Supervisor** (Name, Email Id, Phone) | James Saunderson; james.Saunderson@monash.edu | *Full name, Email* |
| **Monash Co-supervisor(s)** (Name, Email Id, Phone) | | |
| **Monash Head of Dept/Centre** (Name,Email) | Prof Yan Wong, yan.wong@monash.edu | *Full name, email* |
| **Monash Department:** | Electrical and Computer Systems Engineering | |
| **Monash ADGR** (Name,Email) | A/Prof Timothy Scott, timothy.scott@monash.edu | *Full name, email* |
| **IITB Main Supervisor** (Name, Email Id, Phone) | Radhendushka Srivastava; radhe@math.iitb.ac.in | *Full name, Email* |
| **IITB Co-supervisor(s)** (Name, Email Id, Phone) | Ajit Rajwade; ajitvr@cse.iitb.ac.in | *Full name, Email* |
| **IITB Head of Dept** (Name, Email, Phone) | Prof. S. Krishnan | *Full name, email* |
| **IITB Department**: | Department of Mathematics | |

## Research Clusters:

## Research Themes:

| **Highlight which of the Academy's CLUSTERS this project will address?** *(Please nominate JUST **one.** For more information, see www.iitbmonash.org)* | | **Highlight which of the Academy's Theme(s) this project will address?** *(Feel free to nominate more than one. For more information, see www.iitbmonash.org)* | |
|---|---|---|---|
| 1 | *Material Science/Engineering (including Nano, Metallurgy)* | ***1*** | *Artificial Intelligence and Advanced Computational Modelling* |
| 2 | *Energy, Green Chem, Chemistry, Catalysis, Reaction Eng* | 2 | *Circular Economy* |
| 3 | *Math, CFD, Modelling, Manufacturing* | 3 | *Clean Energy* |
| 4 | *CSE, IT, Optimisation, Data, Sensors, Systems, Signal Processing, Control* | 4 | *Health Sciences* |
| 5 | *Earth Sciences and Civil Engineering (Geo, Water, Climate)* | 5 | *Smart Materials* |
| 6 | *Bio, Stem Cells, Bio Chem, Pharma, Food* | 6 | *Sustainable Societies* |
| 7 | *Semi-Conductors, Optics, Photonics, Networks, Telecomm, Power Eng* | 7 | *Infrastructure* |
| 8 | *HSS, Design, Management* | | |

## The research problem

Sparse regression is an area of significant interest in statistics, machine learning, data science and signal processing. Here we are given a linear system of equations where the number of knowns is less than the number of unknowns. Typically, there is no unique solution to such a problem. However a unique solution exists if the vector of unknowns is sparse and if the matrix of coefficients (called the sensing matrix) obeys the property that its nullspace contains no sparse vector apart from the zero vector. There are efficient techniques such as the LASSO which determine a solution to such a system of equations. The LASSO solutions have rigorous theoretical bounds for the squared error of the estimate [3]. However in order to produce any statistical confidence intervals for the **individual** elements of the unknown vector, one needs to include an additional procedure called LASSO Debiasing [1,2], which has been mostly developed for Gaussian noise in the known values. This PhD project seeks to examine novel properties of the debiased LASSO estimator for different noise models, using techniques from the theory of bootstrapping in statistics. The merit of this approach is to be able to generalize the theory to a variety of realistic noise models - for example, Poisson noise in imaging and lognormal noise in RT-PCR (real time polymerase chain reaction) testing. The project will also aim to examine different applications of LASSO debiasing such as in medical image reconstruction from noisy and undersampled measurements.

[1] Javanmard and Montanari, Confidence Intervals and Hypothesis Testing for High-Dimensional Regression, JMLR 2014
[2] Shuvayan Banerjee, Radhendushka Srivastava, James Saunderson and Ajit Rajwade, Robust Non-adaptive Group Testing under Errors in Group Membership Specifications, https://arxiv.org/pdf/2409.05345
[3] Hastie, Tibshirani, Wainwright, Statistical Learning with Sparsity: the LASSO and Generalizations, CRC press, https://dl.acm.org/doi/10.5555/2834535
[4] Hoppe et al, Imaging with Confidence: Uncertainty Quantification for High-Dimensional Undersampled MR Images, ECCV 2024
[5] Shuvayan Banerjee, Sudhansh Peddabomma, Radhendushka Srivastava, James Saunderson and Ajit Rajwade, "Identification and Correction of Permutation Errors in Compressed Sensing Based Group Testing", ICASSP 2025

## Project aims

Algorithm development
Theoretical/statistical analysis of the proposed algorithms

## What is expected of the student when at IITB and when at Monash?

The student is expected to work on algorithmic development as well as theoretical/statistical analysis. The student is also expected to perform numerical simulations to support the theory.

## Expected outcomes

Research papers in prestigious peer-reviewed journals and conferences
Algorithms and programs

## How will the project address the Goals of the above Themes?

The project seeks to discover new theory in the area of sparse regression with interesting applications. The project lies at the intersection of computer science, data science, optimization and statistics. These are well in tune with the theme of AI and Advanced Computational Modelling.

## Potential RPCs from IITB and Monash

Nikhil Karamchandani, Department of EE, IITB
Debraj Das, Department of Mathematics, IITB
Emanuele Viterbo, Department of Electrical and Computer Systems Engineering, Monash
Mehrtash Harandi, Department of Electrical and Computer Systems Engineering, Monash

## Capabilities and Degrees Required

Strong mathematical skills, good grasp of concepts in probability, statistical and linear algebra
Good numerical programming skills (C/C++ or MATLAB or

## Potential Collaborators

Faculty members from the departments of computer science, electrical engineering, or mathematics/statistics from IITB and Monash.

K**eywords** relating to this project to make it easier for the students to apply.

sparse regression, debiasing, compressed sensing, statistics, signal processing, data science, AI and ML