The Second Sydney Workshop on Mathematics of Data Science
10–12, December 2025, University of Sydney, Sydney, Australia

# Program

**Venue for all sessions: F19.03.315.Eastern Avenue Auditorium and Theatre**
A campus map for the location is available here `https://maps.sydney.edu.au/?room=F19.03.315`
Campus Wi-Fi: "eduroam" or "UniSydney-Guest" (Username : yying Password: 864301) )

## Day 1, December 10, 2025, Wednesday

- 08:00–08:50, Registration
- 08:50–09:00, Welcome Speech, by Dingxuan Zhou

**Session I, Chair: Ding-Xuan Zhou**

- 09:00–09:30, **Andreas Christmann**, On qualitative robustness for multi-source data (See abstract, p. 5)
- 09:30–10:00, **Xiaoming Huo**, A new type of Uniform Concentration Inequalities and its application with Kernel-Based Two-Sample Statistics (See abstract, p. 8)
- 10:00–10:30, **Bharath Sriperumbudur**, (De)-regularized maximum mean discrepancy gradient flow (See abstract, p. 13)
- 10:30–11:00, **Coffee Break and Group Photo**

**Session II, Chair: Lixin Shen**

- 11:00–11:30, **Fred Roosta**, Uncertainty Quantification with the Empirical Neural Tangent Kernel (See abstract, p. 11)
- 11:30–12:00, **Georg Gottwald**, State-of-the-art Learning of Chaotic Dynamical Systems with Random Feature Maps (See abstract, p. 6)
- 12:00–12:30, **Song Li**, Theoretical Analysis of Phase Retrieval with Adversarial Sparse Outlier (See abstract, p. 10)
- 12:30–13:30, **Lunch Break (Complimentary lunch boxes)**

**Session III, Chair: Yunwen Lei**

- 13:30–14:00, **Dmytro Matspura**, Centrality of shortest paths: algorithms and complexity results (See abstract, p. 10)
- 14:00–14:30, **Nam Ho-Nguyen**, Maximum Margins Revisited: A New Efficient Online Algorithm (See abstract, p. 8)
- 14:30–15:00, **Matthew Tam**, Convergence of non-stationary algorithms in nonsmooth minimisation (See abstract, p. 13)
- 15:00–15:30, **Coffee Break**

**Session IV, Chair: Matthew Tam**

- 15:30–16:00, **Yunwen Lei**, Non-vacuous Generalization Bounds for Overparameterized Shallow Neural Networks (See abstract, p. 9)
- 16:00–16:30, **Minh Ha Quang**, Geometric mean and geometric Jensen-Shannon divergence between Gaussian measures and Gaussian processes (See abstract, p. 11)

## Day 2, December 11, 2025, Thursday

**Session I, Chair: Georg Gottwald**

- 09:00–09:30, **Weihong Guo**, MS-MGNet: Assembling a Learnable Mumford–Shah Type Model with Multigrid Technique for Image Segmentation (See abstract, p. 7)
- 09:30–10:00, **Lixin Shen**, Explicit Characterization of the $\ell_p$ Proximity Operator for $0 < p < 1$ (See abstract, p. 12)
- 10:00–10:30, **Guohui Song**, Hyper-parameters estimation in sparse Bayesian learning models (See abstract, p. 12)
- 10:30–11:00, **Coffee Break**

**Session II, Chair: Lei Shi**

- 11:00–11:30, **Sakshi Arya**, Interpretable Decision-Making via Single-Index Bandits: Theory, Inference, and Practice (See abstract, p. 4)
- 11:30–12:00, **Anuj Abhishek**, Learning Operators for coefficient-determination in PDE-based inverse problems (See abstract, p. 4)
- 12:00–12:30, **Andi Han**, Feature Learning in the Presence of Label Noise: A Theoretical Perspective (See abstract, p. 8)
- 12:30–13:30, **Lunch Break (Complimentary lunch boxes)**

**Session III, Chair: Xin Guo**

- 13:30–14:00, **Edwin Bonilla**, ProDAG: Projected Variational Inference for Directed Acyclic Graphs (See abstract, p. 4)
- 14:00–14:30, **Cheng Soon Ong**, Some ways to use a binary classifier (See abstract, p. 11)
- 14:30–15:00, **Sevvandi Kandanaarachchi**, Graphon Mixtures (See abstract, p. 9)
- 15:00–15:30, **Coffee Break**

**Session IV, Chair: Jun Fan**

- 15:30–16:00, **Xiang Zhou**, Probability Flow and Generative Model (See abstract, p. 14)
- 16:00–16:30, **Xin Guo**, Learning Green's functions from data (See abstract, p. 7)

## Day 3, December 12, 2025, Friday

**Session I, Chair: Yiming Ying**

- 09:00–09:30, **Clara Grazian**, Deep generalised regression on orthogonal components (DGROC) model (See abstract, p. 6)
- 09:30–10:00,**Hien Nguyen**, When does the variational measure converge given a convergent measure sequence? (See abstract, p. 10)
- 10:00–10:30, **Piotr Koniusz**, Vector task arithmetic for machine unlearning (See abstract, p. 9)
- 10:30–11:00, **Coffee Break**

**Session II, Chair: Andi Han**

- 11:00–11:30, **Susan Wei**, TabMGP: Martingale Posterior with TabPFN (See abstract, p. 13)
- 11:30–12:00, **Tiangang Cui**, An ultra fast sequential experimental design for high-dimensional inverse problems (See abstract, p. 5)
- 12:00–12:30, **Xiaosheng Zhuang**, Spherical Framelets from Spherical Designs (See abstract, p. 14)
- 12:30–13:30, **Lunch Break (Complimentary lunch boxes)**

**Session III, Chair: Yiming Ying**

- 13:30–14:00, **Jun Fan**, Functional data analysis via neural networks (See abstract, p. 5)
- 14:00–14:30, **Lei Shi**, Learning Theory of Classification with Deep Neural Networks (See abstract, p. 12)
- 14:30–15:00, **Zhengchu Guo**, Stochastic Gradient Descent for Two-layer Neural Networks (See abstract, p. 7)

## Titles and Abstracts of the Talks

### Learning Operators for coefficient-determination in PDE-based inverse problems

**Anuj Abhishek**
Case Western Reserve University, United States
axa1828@case.edu

Operator learning methods are emerging as powerful tools for inverse problems governed by PDEs. In this talk, I will discuss recent advances showing how operator learning can accelerate Bayesian sampling and improve reconstructions in challenging imaging problems. The central idea is to replace expensive model evaluations with trained neural surrogates, thereby making previously intractable computations feasible. I will illustrate the approach with examples from PDE-based imaging modalities, highlighting both computational gains and accuracy. More broadly, the results show how operator learning can bridge modern machine learning and classical inverse problems to achieve new levels of efficiency.

### Interpretable Decision-Making via Single-Index Bandits: Theory, Inference, and Practice

**Sakshi Arya**
Case Western Reserve University, United States
sxa1351@case.edu

We present a class of single-index bandit algorithms that blend the interpretability of generalized linear models with the flexibility of non-parametric modeling for contextual decision-making. In the multi-armed bandit with covariates (MABC) setting, we exploit low-dimensional structure to achieve both regret guarantees and statistical inference, enabling confidence intervals for index parameters. Our methods are theoretically grounded and practically motivated, with empirical results on synthetic and real-world data. This work highlights the value of structured yet interpretable approaches to online learning, particularly in high-stakes domains like agriculture and healthcare.

### ProDAG: Projected Variational Inference for Directed Acyclic Graphs

**Edwin Bonilla**
Data61, CSIRO, Australia
Edwin.Bonilla@data61.csiro.au

Directed acyclic graph (DAG) learning is a central task in structure discovery and causal inference. Despite remarkable advances in recent years, it remains both statistically and computationally challenging to learn even a single DAG from data, let alone quantify uncertainty. We tackle this challenge by developing ProDAG, a Bayesian variational inference framework built on novel, provably valid distributions that live directly on the space of sparse DAGs. These distributions, used for both the prior and the variational posterior, arise from a projection that maps a continuous distribution onto the space of sparse, acyclic adjacency matrices. Although this projection is combinatorial in nature, it can be solved efficiently through recent continuous reformulations of acyclicity constraints. Empirically, ProDAG outperforms state-of-the-art methods in both accuracy and uncertainty quantification.

## On qualitative robustness for multi-source data

**Andreas Christmann**
University of Bayreuth, Germany
andreas.christmann@uni-bayreuth.de

Statistical methods and machine learning algorithms should have some robustness or stability concerning small model violations, because the occurrence of "1-10from mixture distributions are also common in practice. One important notion of statistical robustness is qualitative robustness,which is a kind of equicontinuity of a sequence of distributions. A generalization of the notion of qualitative robustness will be given for the case that the data is coming from different sources, e.g. from different hospitals in a multi-center clinical study. It will be shown that this notion of qualitative robustness is interesting for big data in machine learning, too.

## An ultra fast sequential experimental design for high-dimensional inverse problems

**Tiangang Cui**
University of Sydney , Australia
tiangang.cui@sydney.edu.au

Optimal experimental design for inverse problems—a task of identifying the most informative observation functionals—is notoriously difficult. This difficulty stems primarily from the evaluation and optimisation of design criteria. The challenge is further exacerbated by several factors: the high computational cost of forward model evaluations in PDE-constrained settings; the growing number of such evaluations required as the dimensionality increases; and, most critically, the demands of the sequential setting, where each new design must be computed conditional on the updated solution of the inverse problem after assimilating the accumulated data.

We propose to address these difficulties using functional inequalities, which provide (i) an ultra fast surrogate for evaluating design criteria via a closed-form solution, and (ii) a quantitative measure of the effective dimensionality of the inverse problem, thereby bypassing the curse of dimensionality. To ensure the validity of the assumptions underlying these functional-inequality-based design criteria, we couple them with measure transport. This results in an integrated framework that not only overcomes the core challenges of optimal design but also makes the methodology naturally applicable to the otherwise intractable sequential design setting.

## Functional data analysis via neural networks

**Jun Fan**
Hong Kong Baptist University, China
junfan@hkbu.edu.hk

Neural networks excel at learning from finite-dimensional data, but how do they handle infinite-dimensional functional inputs? This talk addresses this question by investigating functional neural networks, a powerful architecture designed to approximate nonlinear smooth functionals. We provide a theoretical analysis of their convergence rates for both approximation and generalization errors within an empirical risk minimization framework. Our findings not only deepen the theoretical understanding of these networks but also establish a foundation for their practical use in functional data analysis.

## State-of-the-art Learning of Chaotic Dynamical Systems with Random Feature Maps

**Georg Gottwald**
University of Sydney, Australia
georg.gottwald@sydney.edu.au

Reservoir computers, in particular, echo state networks have long been established as the leading machine leaning architecture for the task of learning and forecasting chaotic dynamical systems from data. Typically the performance of such models are sensitively dependent on several hyperparameters and a lot of effort has been made in the literature to optimize them with minimal computation. In this talk we consider the a far simpler architecture known as the random feature maps. We show that a judicious data-aware sampling of the internal parameters well as introducing skip connections, localisation and a deep architecture of sequentially ordered random feature map units can be used to achieve state-of-the-art results for learning chaotic systems even in high dimensions. This approach reduces the number of hyperparameters to be optimized to just one. Apart from short-term forecasting, these models are also able to capture long-term behavior of the underlying dynamical system in the test cases. This is joint work with Pinak Mandal.

## Deep generalised regression on orthogonal components (DGROC) model

**Clara Grazian**
Universty of Sydney, Australia
clara.grazian@sydney.edu.au

This work is carried out within an ante-hoc interpretable neural-network framework, in which neural networks are integrated with a statistical model to construct an interpretable model architecture. In this chapter, we proposed a Deep Generalised Regression on Orthogonal Components (DGROC) model based on the GFLSR model structure. The novel model structure and training algorithm are designed to integrate with existing neural network architectures. The mathematical definition of the DGROC model is provided. This specialised neural network structure extracts orthogonal components from the original data and performs regression based on all prior components. Consequently, the model gains interpretability from the orthogonal component framework while leveraging the neural network's capacity to handle large, modern datasets and nonlinear patterns. Furthermore, the model extends naturally to classification tasks, as shown in the real-world application section.

## MS-MGNet: Assembling a Learnable Mumford–Shah Type Model with Multigrid Technique for Image Segmentation

**Weihong Guo**
Case Western Reserve University, United States
weihong.guo@case.edu

The classical Mumford–Shah (MS) model has been successful in image segmentation tasks, providing segmentation results with smooth boundaries of objects. However, the MS model, which operates at the pixel level of the images, faces challenges when dealing with images with low contrast or unclear edges. In this paper, we begin by using a feature extractor to capture high-dimensional deep features that contain more comprehensive semantic information than pixel-level data alone. Inspired by the MS model, we develop a variational model that incorporates threshold dynamics (TD) regularization for segmenting each feature. We obtain the final segmentation result for the original image by assembling segmentation results of all the features. This process results in MS-MGNet, a lightweight trainable segmentation network with a similar architecture to many encoder–decoder networks. The intermediate layers of MS-MGNet are designed by unrolling the numerical scheme based on the multigrid method for solving the variational model. We provide interpretability for the encoder–decoder architecture by elucidating the roles of each layer and offering explanations of the underlying mathematical models. By incorporating the TD regularizer, we integrate spatial priors from the variational models into the network architecture, resulting in better segmentation results with smoother edges and a certain robustness to noise. Compared to some relevant methods, experimental results on the selected data sets with low contrast or unclear edges show that the proposed method can achieve better segmentation performance with fewer parameters, even when trained on smaller data sets.

## Learning Green's functions from data

**Xin Guo**
University of Queensland, Australia
xin.guo@uq.edu.au

We studied the problem of learning the Green's functions of partial differential equations from data, through reproducing kernel methods. With the help of a novel kernel design, we derived an algorithm of time complexity $O(m^3 + m^2N)$ only, where $N$ was the size of training sample, and $m$ was the number of grid points. Minimax lower bound and upper bound of learning rates were derived. Numerical examples on elliptic equations demonstrated accurate approximation of Green's functions.

## Stochastic Gradient Descent for Two-layer Neural Networks

**Zhengchu Guo**
Zhejiang University, China
guozc@zju.edu.cn

In this talk, we will analyze the convergence rates of stochastic gradient descent (SGD) for overparameterized two-layer neural networks. By combining the Neural Tangent Kernel (NTK) approximation with convergence analysis in the associated Reproducing Kernel Hilbert Space (RKHS), we provide a rigorous theoretical framework for understanding the optimization dynamics of SGD in this setting. We establish sharp convergence rates for the last iterate of SGD in overparameterized two-layer networks, significantly relaxing prior assumptions on network width. Specifically, we reduce the required number of neurons from an exponential dependence to a polynomial dependence on the sample size (the number of iterations). This advancement not only enhances the flexibility of neural network design and scaling but also deepens the theoretical understanding of SGD-trained neural networks.

## Feature Learning in the Presence of Label Noise: A Theoretical Perspective

**Andi Han**
University of Sydney, Australia
andi.han@sydney.edu.au

Deep learning models excel at extracting meaningful patterns from data while also risking the memorization of spurious noise. This talk presents a unified perspective on the role of label noise in deep learning. We begin with a signal-plus-noise framework, analyzing how a two-layer convolutional network behaves under partial label corruption. We show a two-stage behaviour in training dynamics: an initial stage where the model prioritizes clean samples to learn robust, generalizable features, followed by a later phase where it overfits noisy labels, degrading test accuracy. We then study a setting where label noise is deliberately injected into the gradient descent updates in a low signal-to-noise regime. This approach suppresses noise memorization and yields superior test performance relative to vanilla gradient descent. Finally, we examine label-noise gradient descent in overparameterized linear models, demonstrating how it transitions the solution from a lazy kernel regime to a richer feature-learning regime. These insights illuminate the role of label noise in shaping model behaviour and offer guidance for robust training strategies.

## Maximum Margins Revisited: A New Efficient Online Algorithm

**Nam Ho-Nguyen**
University of Sydney, Australia
nam.ho-nguyen@sydney.edu.au

Online classification is of high interest in optimization, statistical learning and data science. Online algorithms such as the celebrated perceptron are well-studied, and can guarantee finitely many mistakes on a separable stream of data. In this paper, we revisit the naïve algorithm of computing an offline maximum margin classifier on all points seen so far. Although this naïve approach is inefficient since problem sizes grow as we see more data, it has been shown to make very few mistakes in practice. Complementing this empirical evidence, we conduct a careful analysis of its optimality conditions, and derive mistake bounds and margin guarantees which hold for any strictly convex and differentiable norm, greatly generalizing existing results. We then develop an efficient version of the online maximum margin algorithm, which reduces the per-iteration complexity to a constant number of vector products, comparable with the perceptron and other existing methods. Our new efficient algorithm has the exact same bounds as the naïve algorithm. We show that there are settings where our mistake bound is two orders of magnitude better than perceptron, due to its dependence on much smaller problem parameters. In addition, our algorithms can naturally accommodate a bias term, while existing methods require data transformations that rely on possibly unknown problem parameters.

## A new type of Uniform Concentration Inequalities and its application with Kernel-Based Two-Sample Statistics

**Xiaoming Huo**
Georgia Institute of Technology, United States
huo@isye.gatech.edu

I will present a new uniform concentration inequality for kernel-based two-sample statistics. These statistics include the Maximum Mean Discrepancy (MMD), the Hilbert-Schmidt Independence Criterion (HSIC), the Energy Distance (ED), and the Distance Covariance (dCov), which are widely used. We further demonstrate that the new inequality yields the first-known or better upper bounds for estimation errors in optimization problems that involve kernel-based statistics in the objective function. Consequently, finite-sample and asymptotic performance guarantees are established. Exemplary applications include dCov-based dimension reduction, dCov-based independent component analysis, MMD-based fairness-constrained inference, MMD-based generative model search, and MMD-based generative adversarial networks.

## Graphon Mixtures

**Sevvandi Kandanaarachchi**
Data61, CSIRO, Australia
Sevvandi.Kandanaarachchi@data61.csiro.au

Graphons are graph limits. They are symmetric, measurable functions traditionally defined on a unit square. Intuitively, a graphon can be obtained by scaling the adjacency matrix to the unit square and taking its limit. However, as a result of the Aldous-Hoover theorem, traditional graphons can only represent dense graphs, because sparse graphs converge to the zero graphon. Notwithstanding this challenge, several approached have been proposed to model sparse graphs, often relying on sophisticated mathematical machinery. In this talk we present a simple construction to generate sparse graphs using line graphs. A line graph L(G) maps edges of G to vertices of L(G) and connects two vertices if the corresponding edges in G share a vertex. We show that a subset of sparse graphs has dense line graphs, allowing them to be modelled via the graphon of their line graphs. Furthermore, we propose a mixture that combines a dense graph sequence generated from a standard graphon W, with a sparse graph sequence generated from a line graph graphon U. This (U,W) mixture can generate both sparse and dense graphs depending on the mixture properties. In addition, sparse graphs generated by the (U,W) mixture matches the structure of many real-world graphs including social networks featuring large hubs alongside tightly knit communities.

## Vector task arithmetic for machine unlearning

**Piotr Koniusz**
Data61, CSIRO, Australia
piotr.koniusz@data61.csiro.au

In this talk I will briefly introduce the notion of task vector arithmetic which facilitates fast adaptation of pre-trained networks to new datasets by simple additive arithmetic on network parameter space. Equivalently, subtraction of parameters between pre-trained network and a network fine-tuned on a given "forget" dataset can serve as simple unlearning mechanism. To this end, many authors fine-tune several networks and aggregate task vectors before subtracting them from the original set of parameters. I will introduce our approach which formulates unlearning as an ensemble of functions whose parameters are i.i.d. drawn from the task simplex which represents an unlearning set of parameters. I will show this simple approach can interpolate between function-level ensemble unlearning and parameter-level ensemble unlearning, and it enjoys some interesting properties such a natural formulation as a variance-bias trade-off which can be further balanced for the best performance.

## Non-vacuous Generalization Bounds for Overparameterized Shallow Neural Networks

**Yunwen Lei**
University of Hong Kong , China
leiyw@hku.hk

Overparameterized neural networks often exhibit a benign overfitting phenomenon, achieving excellent generalization performance despite having more parameters than training samples. Traditional generalization analyses typically yield vacuous bounds due to overparameterization. In this talk, we establish non-vacuous generalization bounds by controlling the Rademacher complexity of overparameterized shallow neural networks (SNNs), supported by empirical studies involving highly overparameterized SNNs. Our complexity bounds are fully dependent on the distance from the initialization point and are expressed in terms of the path-norm of the networks.

## Theoretical Analysis of Phase Retrieval with Adversarial Sparse Outlier

**Song Li**
Zhejiang University, China
songli@zju.edu.cn

The problem of solving linear mathematical models with sparse noise is an important issue in the field of compressive sensing. Regarding the solution to this problem, E. Candes and T. Tao have raised an open question about the optimal sparsity of sparse noise. Afterwards,this problem was solved by internationally renowned applied mathematician C. Dwork and others using statistical methods.Based on similar ideas,we will discuss the relevant issues in two other representative mathematical models(low-rank matrix recovery and phase retrieval).Through creatively establish probability density functions for some random variables,we also provide the feature characterization of the optimal sparsity of sparse noise for the two types of models separately.Our work essentially relies on these new probability density function.

## Centrality of shortest paths: algorithms and complexity results

**Dmytro Matspura**
University of Sydney, Australia
dmytro.matsypura@sydney.edu.au

The centrality of a node is often used to measure its importance within a network's structure. Certain centrality measures can be extended to evaluate the importance of a group of nodes that share a common property or form a specific structure, such as a path. In this talk, we focus on identifying the most central shortest path. We demonstrate that the computational complexity of this problem depends on the centrality measure used and, in the case of degree centrality, whether the network is weighted or not. We develop a polynomial algorithm for the most degree-central shortest path problem, with the worst-case running time of $O(|E||V|^2\Delta(G))$, where $|V|$ represents the number of vertices in the network, $|E|$ is the number of edges, and $\Delta(G)$ denotes the maximum degree of the graph. Additionally, we show that this problem is NP-hard on a weighted graph. Furthermore, we demonstrate that the problem of finding the most betweenness-central shortest path can be solved in polynomial time, while finding the most closeness-central shortest path is NP-hard, irrespective of whether the graph is weighted or not. We also develop an algorithm for identifying the most betweenness-central shortest path with a running time of $O(|E|^2|V|^2)$ for both weighted and unweighted graphs. Finally, we present a numerical study of our algorithms on synthetic and real-world networks, comparing our results to those found in the existing literature.

## When does the variational measure converge given a convergent measure sequence?

**Hien Nguyen**
La Trobe University, Australia
H.Nguyen5@latrobe.edu.au

In Bayesian statistics, a sequence of measures $(\mu_n)_n$ is called consistent for a point $\theta_0$ if it weakly converges to the delta measure at $\theta_0$. An approximation approach is often taken whereby, for each $\mu_n$, one approximates $\mu_n$ by a $\nu_n$ that is a minimiser of the Kullback–Leibler (KL) divergence over some approximation family. An elementary question that arises in such a setting is whether the sequence $(\nu_n)_n$ converges to the same delta measure as the original sequence $(\mu_n)_n$. We provide necessary and sufficient conditions for this to hold on Polish spaces, assuming that the minimal KL divergence is eventually finite, and we give some sufficient conditions guaranteeing this eventual finiteness.

## Some ways to use a binary classifier

**Cheng Soon Ong**
Data61, CSIRO, Australia
Cheng-Soon.Ong@data61.csiro.au

We revisit binary classification, more precisely class probability estimation, so that we can discuss downstream applications of a trained binary classifier. First we show that binary classification can directly solve density ratio estimation, and we link the loss used for training the classifier and the density ratio estimator. Then we consider the problem of Bayesian optimisation, and discuss how it can be seen as estimating a density ratio (and hence binary classification). Finally, leaning on recent advances in generative sequence models, we consider the problem of active generation where the search space of Bayesian optimisation is large and sparse.

## Geometric mean and geometric Jensen-Shannon divergence between Gaussian measures and Gaussian processes

**Minh Ha Quang**
RIKEN Center, Japan
minh.haquang@riken.jp

The Jensen-Shannon divergence, a symmetrized version of the Kullback-Leibler (KL) divergence, is an important object in probablity theory and information theory. Despite much research, as of current writing, a closed form formula for the divergence between two Gaussian measures remains unknown. In this talk, we first present a generalized version of the Jensen-Shannon divergence via the concept of abstract means, as introduced by Nielsen (2019). We then present the Geometric Jensen-Shannon divergence, based on the notion of geometric mean of probability measures, in the setting of Gaussian measures on an infinite-dimensional Hilbert space. On the set of all Gaussian measures equivalent to a fixed one, we present closed form expressions for the geometric mean and geometric divergence that directly generalize the corresponding finite-dimensional versions. Using the notion of infinite-dimensional Log-Determinant divergences between positive definite unitized trace class operators, we then define a Regularized Geometric Jensen-Shannon divergence that is valid for any pair of Gaussian measures and that recovers the exact Geometric Jensen-Shannon divergence between two equivalent Gaussian measures when the regularization parameter tends to zero. For the setting of Gaussian processes with squared integrable paths, consistent finite-dimensional approximations are obtained via RKHS methodology. The mathematical formulations are illustrated by numerical experiments with Gaussian processes.

## Uncertainty Quantification with the Empirical Neural Tangent Kernel

**Fred Roosta**
University of Queensland, Australia
fred.roosta@uq.edu.au

While neural networks have demonstrated impressive performance across various tasks, accurately quantifying uncertainty in their predictions is essential to ensure their trustworthiness and enable widespread adoption in critical systems. Several Bayesian uncertainty quantification (UQ) methods exist that are either cheap or reliable, but not both. We propose a post-hoc, sampling-based UQ method for over-parameterized networks at the end of training. Our approach constructs efficient and meaningful deep ensembles by employing a (stochastic) gradient-descent sampling process on appropriately linearized networks. We demonstrate that our method effectively approximates the posterior of a Gaussian process using the empirical Neural Tangent Kernel. Through a series of numerical experiments, we show that our method not only outperforms competing approaches in computational efficiency (often reducing costs by multiple factors) but also maintains state-of-the-art performance across a variety of UQ metrics for both regression and classification task.

## Explicit Characterization of the $\ell_p$ Proximity Operator for $0 < p < 1$

**Lixin Shen**
Syracuse University, United States
lshen03@syr.edu

The nonconvex $\ell_p$ quasi-norm with $0 < p < 1$ is a powerful surrogate for sparsity but complicates the evaluation of proximal maps that underpin many modern algorithms. In this talk, we give an explicit characterization of the scalar proximal operator of $|\cdot|^p$ for all $0 < p < 1$, including the structure and admissible ranges of global minimizers, as well as conditions ensuring strict, isolated solutions. By applying the Lagrange–Bürmann inversion formula to the stationarity equation, we derive a uniformly convergent series for the larger positive root, yielding an exact and numerically stable formula for the $\ell_p$ proximal map above the classical threshold. We further provide a Mellin–Barnes integral representation and identify the series as a Fox–Wright function, which determines its radius of convergence. Specializations recover the known closed forms for $p = \frac{1}{2}$ and $p = \frac{2}{3}$, and we obtain compact hypergeometric expressions for additional rational cases (e.g., $p = \frac{1}{3}$). These results unify scattered formulas into a single framework and enable high-accuracy evaluation of $\ell_p$ proximity operators across the full range $0 < p < 1$.

## Learning Theory of Classification with Deep Neural Networks

**Lei Shi**
Fudan university, China
leishi@fudan.edu.cn

Deep neural networks have achieved remarkable success in various binary classification tasks. Despite their practical effectiveness, our theoretical understanding of their generalization in binary classification remains limited. In this talk, I will present our recent progress on classification using deep neural networks. This talk is based on joint work with Dr. Zihan Zhang and Prof. Ding-Xuan Zhou.

## Hyper-parameters estimation in sparse Bayesian learning models

**Guohui Song**
Old Dominion University, United States
gsong@odu.edu

Hyper-parameters play a crucial role in shaping the behavior of Bayesian models and can have a significant impact on their performance and generalization ability. Bayesian models can have a high number of hyper-parameters, particularly in complex models like hierarchical Bayesian models or deep probabilistic models. Managing and tuning a large number of hyperparameters can be computationally intensive and require careful attention. We will propose an efficient numerical algorithm of hyper-parameters estimation in sparse Bayesian learning models and present its comparisons with other popular methods of hyper-parameters estimation.

## (De)-regularized maximum mean discrepancy gradient flow

**Bharath Sriperumbudur**
Pennsylvania State University, United States
bks18@psu.edu

We introduce a (de)-regularization of the Maximum Mean Discrepancy (DrMMD) and its Wasserstein gradient flow. Existing gradient flows that transport samples from the source distribution to the target distribution with only target samples either lack tractable numerical implementation (f-divergence flows) or require strong assumptions, and modifications such as noise injection, to ensure convergence (Maximum Mean Discrepancy flows). In contrast, DrMMD flow can simultaneously (i) guarantee near-global convergence for a broad class of targets in both continuous and discrete time, and (ii) be implemented in closed form using only samples. The former is achieved by leveraging the connection between the DrMMD and the $\chi^2$-divergence, while the latter comes by treating DrMMD as MMD with a de-regularized kernel. Our numerical scheme uses an adaptive de-regularization schedule throughout the flow to optimally trade off between discretization errors and deviations from the $\chi^2$ regime. The potential application of the DrMMD flow is demonstrated across several numerical experiments, including a large-scale setting of training student/teacher networks.

## Convergence of non-stationary algorithms in nonsmooth minimisation

**Matthew Tam**
University of Melbourne, Australia
matthew.tam@unimelb.edu.au

We study minimisation problems whose objective involves the sum of three functions: two nonsmooth and one smooth. Algorithms which can exploit this structure are typically related to "Douglas–Rachford method" and, consequently, they are only well-understood when the stepsize fixed and the smooth function has Lipschitz gradient. In this talk, we develop an approach to developing "non-stationary" algorithms including the Douglas–Rachford method with non-constant stepsize. The approach relies on a simple but so-far unused observation about proximity operators of convex functions.

## TabMGP: Martingale Posterior with TabPFN

**Susan Wei**
University of Monash, Australia
susan.wei@monash.edu

Bayesian inference provides principled uncertainty quantification but is often limited by challenges of prior elicitation, likelihood misspecification, and computational burden. The martingale posterior (MGP, Fong et al., 2023) offers an alternative, replacing prior-likelihood elicitation with a predictive rule - namely, a sequence of one-step-ahead predictive distributions - for forward data generation. The utility of MGPs depends on the choice of predictive rule, yet the literature has offered few compelling examples. Foundation transformers are well-suited here, as their autoregressive generation mirrors this forward simulation and their general-purpose design enables rich predictive modeling. We introduce TabMGP, an MGP built on TabPFN, a transformer foundation model that is currently state-of-the-art for tabular data. TabMGP produces credible sets with near-nominal coverage and often outperforms both existing MGP constructions and standard Bayes.

## Probability Flow and Generative Model

**Xiang Zhou**
City University of Hong Kong , China
xiang.zhou@cityu.edu.hk

The generative model, as well as optimal transport and Markov bridge, is to seek a geometric path in the space of probability measures under various settings in practice. We may refer that is driven by a Probability Flow to highlight that the same path can be achieved by different transport dynamics (stochastic or deterministic) and different directions for the arrows of time, which opens a vast opportunity for different algorithms across a large number of applications. From geometric curve to time direction, the irreversible dynamics also have a deep connection with the Second Law of Thermodynamics in stochastic thermodynamics.This talk will demonstrate how this viewpoint, with the numerical tools from generative models (score-based diffusion models, normalising flow, etc.), contributes to challenging problems in various applications, including hyperparameter tuning in score-based diffusion, sampling distribution of SDEs, computing the entropy production rate in non-equilibrium active matter, and addressing challenges even with the presence of non-Gaussian stochastic systems. The relevant preprints include arXiv 2306.02063, 2409.01340,2412.19520, 2504.06628, 2405.19256. The main collaborator is Dr Yuanfei HUANG.

## Spherical Framelets from Spherical Designs

**Xiaosheng Zhuang**
City University of Hong Kong, China
xzhuang7@cityu.edu.hk

In this talk, we discuss the structures of the variational characterization of the spherical t-design, its gradient, and its Hessian in terms of fast spherical harmonic transforms. Moreover, we propose solving the minimization problem of the spherical t-design using the trust-region method to provide spherical t-designs with large values of t. Based on the obtained spherical t-designs, we develop (semi-discrete) spherical tight framelets as well as their truncated systems and their fast spherical framelet transforms for practical spherical signal/image processing. Thanks to the large spherical t-designs and localization property of our spherical framelets, we are able to provide signal/image denoising using local thresholding techniques based on a fine-tuned spherical cap restriction. Many numerical experiments are conducted to demonstrate the efficiency and effectiveness of our spherical framelets and spherical designs, including Wendland function approximation, ETOPO data processing, and spherical image denoising.

# List of Participants

**Anuj Abhishek** (See abstract, p. 4)
Case Western Reserve University, United States, axa1828@case.edu

**Sakshi Arya** (See abstract, p. 4)
Case Western Reserve University, United States, sxa1351@case.edu

**Edwin Bonilla** (See abstract, p. 4)
Data61, CSIRO, Australia, Edwin.Bonilla@data61.csiro.au

**Andreas Christmann** (See abstract, p. 5)
University of Bayreuth, Germany, andreas.christmann@uni-bayreuth.de

**Tiangang Cui** (See abstract, p. 5)
University of Sydney , Australia, tiangang.cui@sydney.edu.au

**Jun Fan** (See abstract, p. 5)
Hong Kong Baptist University, China, junfan@hkbu.edu.hk

**Georg Gottwald** (See abstract, p. 6)
University of Sydney, Australia, georg.gottwald@sydney.edu.au

**Clara Grazian** (See abstract, p. 6)
Universty of Sydney, Australia, clara.grazian@sydney.edu.au

**Weihong Guo** (See abstract, p. 7)
Case Western Reserve University, United States, weihong.guo@case.edu

**Xin Guo** (See abstract, p. 7)
University of Queensland, Australia, xin.guo@uq.edu.au

**Zhengchu Guo** (See abstract, p. 7)
Zhejiang University, China, guozc@zju.edu.cn

**Andi Han** (See abstract, p. 8)
University of Sydney, Australia, andi.han@sydney.edu.au

**Nam Ho-Nguyen** (See abstract, p. 8)
University of Sydney, Australia, nam.ho-nguyen@sydney.edu.au

**Xiaoming Huo** (See abstract, p. 8)
Georgia Institute of Technology, United States, huo@isye.gatech.edu

**Sevvandi Kandanaarachchi** (See abstract, p. 9)
Data61, CSIRO, Australia, Sevvandi.Kandanaarachchi@data61.csiro.au

**Piotr Koniusz** (See abstract, p. 9)
Data61, CSIRO, Australia, piotr.koniusz@data61.csiro.au

**Yunwen Lei** (See abstract, p. 9)
University of Hong Kong , China, leiyw@hku.hk

**Song Li** (See abstract, p. 10)
Zhejiang University, China, songli@zju.edu.cn

**Dmytro Matspura** (See abstract, p. 10)
University of Sydney, Australia, dmytro.matsypura@sydney.edu.au

**Hien Nguyen** (See abstract, p. 10)
La Trobe University, Australia, H.Nguyen5@latrobe.edu.au

**Cheng Soon Ong** (See abstract, p. 11)
Data61, CSIRO, Australia, Cheng-Soon.Ong@data61.csiro.au

**Minh Ha Quang** (See abstract, p. 11)
RIKEN Center, Japan, minh.haquang@riken.jp

**Fred Roosta** (See abstract, p. 11)
University of Queensland, Australia, fred.roosta@uq.edu.au

**Lixin Shen** (See abstract, p. 12)
Syracuse University, United States, lshen03@syr.edu

**Lei Shi** (See abstract, p. 12)
Fudan university, China, leishi@fudan.edu.cn

**Guohui Song** (See abstract, p. 12)
Old Dominion University, United States, gsong@odu.edu

**Bharath Sriperumbudur** (See abstract, p. 13)
Pennsylvania State University, United States, bks18@psu.edu

**Matthew Tam** (See abstract, p. 13)
University of Melbourne, Australia, matthew.tam@unimelb.edu.au

**Susan Wei** (See abstract, p. 13)
University of Monash, Australia, susan.wei@monash.edu

**Yiming Ying**
University of Sydney, Australia, yiming.ying@sydney.edu.au

**DingXuan Zhou**
University of Sydney, Australia, dingxuan.zhou@sydney.edu.au

**Xiang Zhou** (See abstract, p. 14)
City University of Hong Kong , China, xiang.zhou@cityu.edu.hk

**Xiaosheng Zhuang** (See abstract, p. 14)
City University of Hong Kong, China, xzhuang7@cityu.edu.hk