15th Symposium of SIAM Student Chapter at NUS

Date: March 3-4, 2026

Organizer: SIAM Student Chapter at NUS

- Homepage: https://siamnus.github.io/website/

Organizing Committee:

- Prof. Bao Weizhu (matbaowz@nus.edu.sg, Faculty Advisor)
- Mr. Hou Di (dihou@u.nus.edu, President)
- Mr. Cheng Jingpu (chengjingpu@u.nus.edu, Vice President)
- Ms. Liu Caoyi (e1353374@u.nus.edu, Secretary)

Venue:

- IMS Executive Seminar Room
- Bock S17, Level 3
- 10 Lower Kent Ridge Rd Singapore 119076

Scan for Full Schedule

Plenary Speaker:

- Prof. Ying Chen, *NUS, Department of Mathematics*
- Prof. Shaoning Han, *NUS, Department of Mathematics*

Student Speaker

- Mr. Laziz Abdullaev, *NUS, Department of Mathematics*
- Mr. Tran Hoang Anh, *NUS, Department of Mathematics*
- Ms. Mengyi Chen, *NUS, Department of Mathematics*
- Mr. Dongmin Wang, *NUS, Department of Mathematics*
- Mr. Penghao Yu, *NUS, Department of Mathematics*
- Ms. Tianjing Zhang, *NUS, Department of Mathematics*
- Mr. Xiangming Gu, *NUS, School of Computing*
- Mr. Renyuan Li, *NUS, Industrial Systems Engineering and Management*
- Mr. Haozhe Tan, *NUS, Industrial Systems Engineering and Management*
- Mr. Anh Duc Nguyen, *NUS, Institute of Operations Research and Analytics*
- Ms. Zilan Cheng, *NTU, School of Physical and Mathematical Sciences*
- Mr. Xixian Wang, *NTU, School of Physical and Mathematical Sciences*
- Mr. Yu Xiong, *NTU, School of Physical and Mathematical Sciences*
- Mr. Deheng Yuan, *NTU, School of Physical and Mathematical Sciences*

Supported by

SIAM   Department of Mathematics, NUS

Program:

15th Symposium of SIAM Student Chapter at NUS		
March 3, 2026 (IMS Executive Seminar Room)		
Time	Title	Name
8:55–9:00 AM	Opening Remarks	
Chair: Di Hou		
9:00–9:45 AM	**Plenary Talk:** Algorithmic Intelligence in the Quantum-AI Era	Ying Chen
9:45–10:15 AM	Demystifying Attention Sink in LLMs and its Applications to Architecture Design	Xiangming Gu
10:15–10:45 AM	Scalable Learning of Macroscopic Stochastic Dynamics	Mengyi Chen
10:45–11:00 AM	Coffee Break	
Chair: Caoyi Liu		
11:00–11:30 AM	Attention Head Count and Transformer Approximation	Penghao Yu
11:30–12:00 PM	Concept Heterogeneity-aware Representation Steering	Laziz Abdullaev
12:00–1:30 PM	Lunch Break	
Chair: Jingpu Cheng		
1:30–2:15 PM	**Plenary Talk:** A Geometric Perspective on Polynomially Solvable Convex Maximization	Shaoning Han
2:15–2:45 PM	Wasserstein Bounds for generative diffusion models with Gaussian tail targets	Xixian Wang
2:45–3:15 PM	PODNO: Proper Orthogonal Decomposition Neural Operators	Zilan Cheng

3:15–3:45 PM	Zero-Shot Blind-Spot Image Denoising via Cross-Scale Non-Local Pixel Refilling	Tianjing Zhang
3:45–4:00 PM	Coffee Break	
	Chair: Caoyi Liu	
4:00–4:30 PM	Layer-wise Quantization and Optimistic Dual Averaging	Anh Duc Nguyen
4:30–5:00 PM	Distributed Statistical Learning under Communication Constraints	Deheng Yuan

March 4, 2026 (IMS Executive Seminar Room)		
Time	Title	Name
	Chair: Shiqi Wu	
9:00–9:30 AM	A priori and a posteriori error analyses of a pressure-robust virtual element method for the two-dimensional Brinkman problem	Yu Xiong
9:30–10:00 AM	An energy-stable parametric finite element method for the Willmore flow in three dimensions	Dongmin Wang
10:00–10:30 AM	Coffee Break	
	Chair: Junyu Chen	
10:30–11:00 AM	On the convergence rates of moment-SOS hierarchies approximation of truncated moment sequences	Tran Hoang Anh
11:00–11:30 AM	Spectral Unification of Recursive Matrix Bounds and Lyapunov Methods in Stochastic Optimization	Renyuan Li
11:30–12:00 AM	Fast Presolving Framework For Sparsity Constrained Convex Quadratic Programming: Screening-Based Cut Generation and Selection	Haozhe Tan

Abstracts

Algorithmic Intelligence in the Quantum-AI Era

Prof. Ying Chen

Department of Mathematics, NUS

As we enter an era dominated by artificial intelligence and quantum computing, the role of applied mathematics is more vital than ever — not only in making sense of data, but in shaping algorithms that drive intelligent decision-making under uncertainty. This talk presents an applied mathematics lens on algorithmic intelligence, grounded in recent work on hybrid AI–quantum models, digital twins, and optimization in complex systems such as finance, healthcare, and logistics. It will share insights involving AI-driven forecasting, reinforcement learning, and quantum circuit learning — highlighting how algorithmic intelligence thinking ensures robustness, interpretability, and transferability in modern learning systems.

A Geometric Perspective on Polynomially Solvable Convex Maximization

Prof. Shaoning Han

Department of Mathematics, NUS

Convex maximization arises in many applications but is generally NP-hard, even for fixed-rank objectives. This paper introduces a set of broadly applicable conditions that certify when such problems are polynomially solvable. Our main condition is a new property of the feasible set, which we term comonotonicity. Under comonotonicity and mild additional assumptions, we develop a geometric framework that generates polynomially many candidate solutions, at least one of which is optimal, yielding a polynomial-time algorithm. We further introduce a lifting technique to improve the complexity bound under standard comonotonicity. Our framework unifies existing tractability results and delivers the first polynomial-time guarantees for several applications.

A priori and a posteriori error analyses of a pressure-robust virtual element method for the two-dimensional Brinkman problem

Mr. Yu Xiong

School of Physical and Mathematical Sciences, NTU

This talk investigates both a priori and a posteriori error estimates for a pressure-robust and divergence-free virtual element method to approximate the incompressible Brinkman problem on polygonal meshes. The exactly divergence-free property of virtual space preserves the mass-conservation of the system. By extending the lowest-order Raviart–Thomas element to polygonal meshes, we construct a divergence-preserving reconstructor for the discretization of the right-hand side. A rigorous a priori error analysis is developed, showing that the velocity error is independent of both the continuous pressure and the viscosity. Taking advantage of the virtual element method's ability to handle more general polygonal meshes, we design an adaptive mesh refinement approach and construct a residual-type a posteriori error indicator. This indicator is proven to provide global upper and local lower bounds for the discretization error. Finally, some numerical experiments demonstrate the robustness, accuracy, reliability and efficiency of the method.

An energy-stable parametric finite element method for the Willmore flow in three dimensions

Mr. Dongmin Wang

Department of Mathematics, NUS

In this talk, I will introduce an energy-stable parametric finite element method (PFEM) for simulating Willmore flow in three dimensions. The Willmore flow describes the geometric evolution of surfaces driven by curvature and arises in various applications in materials science and biophysics. As the L^2–gradient flow of the Willmore energy, the Willmore flow is intrinsically energy dissipative. Designing numerical schemes that rigorously preserve this dissipation structure at the discrete level, however, remains a fundamental and nontrivial challenge. By incorporating a reformulated normal velocity equation together with the temporal

evolution equation of the mean curvature, we propose a new variational formulation that ensures energy stability of the PFEM at the fully discrete level. Moreover, this framework can be extended to a broad class of curvature-dependent geometric flows, including mean curvature flow and Gauss curvature flow.

PODNO: Proper Orthogonal Decomposition Neural Operators

Ms. Zilan Cheng

School of Physical and Mathematical Sciences, NTU

We introduce Proper Orthogonal Decomposition Neural Operators (PODNO) for solving partial differential equations (PDEs) dominated by high-frequency components. Building on the structure of Fourier Neural Operators (FNO), PODNO replaces the Fourier transform with (inverse) orthonormal transforms derived from the Proper Orthogonal Decomposition (POD) method to construct the integral kernel. Due to the optimality of POD basis, the PODNO has potential to outperform FNO in both accuracy and computational efficiency for high-frequency problems. From analysis point of view, we established the universality of a generalization of PODNO, termed as Generalized Spectral Operators (GSO). In addition, we evaluate PODNO's performance numerically on dispersive equations such as the Nonlinear Schr\"odinger (NLS) equation and the Kadomtsev–Petviashvili (KP) equation.

Concept Heterogeneity-aware Representation Steering

Mr. Laziz Abdullaev

Department of Mathematics, NUS

Representation steering offers a lightweight mechanism for controlling the behavior of large language models (LLMs) by intervening on internal activations at inference time. Most existing methods rely on a single global steering direction, typically obtained via difference-in-means over contrastive datasets. This approach implicitly assumes that the target concept is homogeneously represented across the embedding space. In practice, however, LLM

representations can be highly non-homogeneous, exhibiting clustered, context-dependent structure, which renders global steering directions brittle. In this work, we view representation steering through the lens of optimal transport (OT), noting that standard difference-in-means steering implicitly corresponds to the OT map between two unimodal Gaussian distributions with identical covariance, yielding a global translation. To relax this restrictive assumption, we theoretically model source and target representations as Gaussian mixture models and formulate steering as a discrete OT problem between semantic latent clusters. From the resulting transport plan, we derive an explicit, input-dependent steering map via barycentric projection, producing a smooth, kernel-weighted combination of cluster-level shifts. We term this method Concept Heterogeneity-aware Representation Steering (CHaRS). Through numerous experimental settings, we show that CHaRS yields more effective behavioral control than global steering.

Demystifying Attention Sink in LLMs and its Applications to Architecture Design

Mr. Xiangming Gu

School of Computing, NUS

Currently, the Large Language Model (LLM) widely adopts the Decoder-only Transformer architecture, in which there is a phenomenon known as attention sink. This refers to the first token, even if it lacks semantic meaning, being assigned an excessively high attention score. This phenomenon is closely related to the long-text capabilities of large language models, efficient computation, kv cache optimization, and model quantization. This talk will explain why this phenomenon exists in LLMs and how it influences the model architecture design of LLMs. Additionally, the talk will interpret why OpenAI's open-source model, GPT-OSS, incorporates a learnable sink and why Qwen3-Next employs a gating mechanism to address the attention sink phenomenon.

On the convergence rates of moment-SOS hierarchies approximation of truncated moment sequences

Mr. Tran Hoang Anh

Department of Mathematics, NUS

The moment-SOS hierarchy, which is based on Putinar-type (quadratic module) and Schmudgen-type (preordering) sum-of-squares positivity certificates, is a widely applicable framework to address polynomial optimization problems over basic semi-algebraic sets. Recent works show that the convergence rate of this hierarchy over certain simple sets, namely, the unit ball, hypercube, and standard simplex, is of the order $O(1/r^2)$, where r denotes the level of the moment-SOS hierarchy. This paper aims to provide a comprehensive understanding of the convergence rate of the Schm¨udgen-type moment-SOS hierarchy by estimating the Hausdorff distance between the set of truncated pseudo-moment sequences and the set of truncated moment sequences specified by projection-based method derived from Tchakaloff's theorem. Our results provide a connection between the convergence rate of the Schmudgen-type moment-SOS hierarchy and the Lojasiewicz exponent L of the domain under the compactness assumption, where we establish the convergence rate of $O(1/r^L)$. Consequently, we obtain the convergence rate of $O(1/r)$ for polytopes and sets satisfying the constraint qualification condition, $O(1/\sqrt{r})$ for domains that either satisfy the Polyak-Lojasiewicz condition or are defined by locally strongly convex polynomials. We also use our method to reprove the convergence rate of $O(1/r^2)$ for general polynomials over a sphere.

Distributed Statistical Learning under Communication Constraints

Mr. Deheng Yuan

School of Physical and Mathematical Sciences, NTU

In classical statistical learning problems, it is often assumed that all training data is available at a central location. However, in many practical cases such as federated learning, data is collected by distributed sensors and needs to be communicated to a learner via capacity-constrained channels. Analysing the theoretical limits for this model is a recent interest. Our work focused

on a specific discrete distribution estimation problem under communication constraints. We obtained the minimax optimal rates of the problem for a wider range than previous works, letting each terminal hold $n>1$ samples and imposing general ℓ^p losses simultaneously. Interestingly, we found that there exists an elbow effect of the optimal rates at $p=2$. Novel adaptive refinement protocols were designed to achieve the optimal rate, and their optimality was shown by deriving compatible minimax lower bounds.

Attention Head Count and Transformer Approximation

Mr. Penghao Yu

Department of Mathematics, NUS

In this talk, we show how the number of attention heads controls the approximation efficiency of (single-layer) transformers as sequence-to-vector maps. Our analysis is built around a structured family of targets, called generalized D-retrieval tasks, which is dense in $C(K)$ for a compact sequence domain K. Within this framework, we establish upper and lower bounds on the parameter complexity needed for ε-approximation. In particular, having sufficiently many heads enables efficient approximation, whereas with too few heads the required number of parameters must grow as $\Omega(1/\varepsilon^{cT})$ for some constant c and sequence length T. We also analyze the single-head regime and show that an embedding dimension of order $O(T)$ permits full memorization of the input, shifting approximation to the feed-forward block.

Scalable Learning of Macroscopic Stochastic Dynamics

Ms. Mengyi Chen

Department of Mathematics, NUS

Macroscopic dynamical descriptions of complex physical systems are crucial for understanding and controlling material behavior. With the growing availability of data and compute, machine learning has become a promising alternative to first-principles methods to build accurate

macroscopic models from microscopic trajectory simulations. However, for spatially extended systems, direct simulations of sufficiently large microscopic systems that inform macroscopic behavior is prohibitive. In this talk, I will introduce a framework that learns the macroscopic dynamics of large microscopic systems using only small-system simulations. Our framework employs a partial evolution scheme to generate training data pairs by evolving large-system snapshots within local patches. We subsequently identify the closure variables associated with the macroscopic observables and learn the macroscopic dynamics using a custom loss. I will also briefly introduce our ongoing work to apply this framework to a realistic high-entropy alloy system.

Layer-wise Quantization and Optimistic Dual Averaging

Mr. Nguyen Anh Duc

Institute of Operations Research and Analytics, NUS

Modern deep neural networks exhibit heterogeneity across numerous layers of various types such as residuals, multi-head attention, etc., due to varying structures (dimensions, activation functions, etc.), distinct representation characteristics, which impact predictions. We develop a general layer-wise quantization framework with tight variance and code-length bounds, adapting to the heterogeneities over the course of training. We then apply a new layer-wise quantization technique within distributed variational inequalities (VIs), proposing a novel Quantized Optimistic Dual Averaging (QODA) algorithm with adaptive learning rates, which achieves competitive convergence rates for monotone VIs. We empirically show that QODA achieves up to a 150% speedup over the baselines in end-to-end training time for training Wasserstein GAN on 12+ GPUs.

Spectral Unification of Recursive Matrix Bounds and Lyapunov Methods in Stochastic Optimization

Mr. Renyuan Li

Industrial Systems Engineering and Management, NUS

This report presents a unified theoretical framework for analyzing the stability and convergence of stochastic optimization algorithms governed by recursive matrix inequalities. We rigorously compare two distinct bounding methodologies: the forward-propagating Recursive Inequality approach (spectral decomposition) and the backward-projecting Lyapunov approach (scalar potential functions). We demonstrate that for time-invariant, positive systems, these approaches are asymptotically equivalent, with convergence rates governed by the Perron-Frobenius spectral radius. Beyond this regime, we characterize their structural similarities and differences under reducibility, indefinite dynamics, and introduce lifting techniques to restore applicability. The practical utility of this framework is highlighted through rigorous derivations for SPSTORM.

Wasserstein Bounds for generative diffusion models with Gaussian tail targets

Mr. Xixian Wang

School of Physical and Mathematical Sciences, NTU

Diffusion models have emerged as state-of-the-art generative models, yet their theoretical understanding remains incomplete, particularly regarding convergence guarantees in high dimensions.

In this talk, I will discuss convergence bounds for score-based diffusion models under a Gaussian tail assumption on the target distribution. The key challenge lies in establishing uniform Lipschitz and growth estimates for the score function, which are necessary for well-posedness and numerical convergence of the backward SDE.

By exploiting a heat kernel representation and a suitable Gaussian decomposition of the density, we derive dimension-free gradient and Hessian estimates for the score. These bounds lead to optimal Wasserstein convergence guarantees and a sharp sampling complexity of order $O(\sqrt{d})$.

Fast Presolving Framework For Sparsity Constrained Convex Quadratic Programming: Screening-Based Cut Generation and Selection

Mr. Haozhe Tan

Industrial Systems Engineering and Management, NUS

Screening is widely utilized for Mixed-Integer Programming (MIP) presolving. It aims to certify a \emph{priori} whether one or multiple specific binary variables can be fixed to optimal values based on solutions from convex relaxations. This paper studies the challenge of solving Sparsity-constrained (strongly) Convex Quadratic Programming (SCQP) and proposes the Screening-based Cut Presolving Framework (SCPF). SCPF contains two parts: a Screening-based Cut Generation (SCG) rule and a Screening-based Cut Selection (SCS) method. We show that the SCG provides superior screening ability compared to existing screening methods, and achieves a finer balance between screening effectiveness and computational overhead. We then provide theoretical guarantees for the SCS method to ensure the selection of generated cuts with high screening ability. Extensive numerical experiments validate the theoretical findings and demonstrate that the proposed framework significantly outperforms state-of-the-art screening methods. Notably, our SCPF achieves a $1.7\times$ to $3.0\times$ acceleration in total running time, especially in challenging phases, across high-dimensional synthetic datasets, complex real-world instances, and simulation libraries from sparse identification of nonlinear dynamics.

Zero-Shot Blind-Spot Image Denoising via Cross-Scale Non-Local Pixel Refilling

Ms. Tianjing Zhang

Department of Mathematics, NUS

Blind-spot denoising (BSD) enables zero-shot image denoising by training directly on a single noisy image. However, real-world noise often exhibits strong local correlation, which limits the effectiveness of standard BSD approaches. In this talk, I will present a theoretical analysis revealing a trade-off between noise decorrelation and pixel-value correlation in BSD, and introduce a cross-scale non-local pixel refilling strategy to address this challenge. Experimental results demonstrate consistent improvements over existing zero-shot methods on real-world benchmarks.