Bayesian Learning of Very High-Dimensional Physical Process Models

The Workshop Program

David Gunawan, Ajay Jasra, Leah South, Matthew T. Moores, Sumeetpal S. Singh, and Susan Wei

June 14, 2025

Dates: 23rd June to 4th July 2025.

Organising Committee :

- David Gunawan (University of Wollongong, Australia)
- Ajay Jasra (The Chinese University of Hong Kong, China)
- Leah South (Queensland University of Technology, Australia)
- Matthew T. Moores (University of Wollongong, Australia)
- Sumeetpal S. Singh (University of Wollongong, Australia)
- Susan Wei (Monash University, Australia)

DG is a Senior Lecturer, MTM is a Lecturer and SSS holds the Tibra Chair in Mathematical Sciences, all based at the School of Mathematics and Applied Statistics. AJ is a Professor at the School of Data Science. LS is a Senior Lecturer and DECRA fellow at the School of Mathematical Sciences. SW is an Associate Professor at the Department of Econometrics and Business Statistics. E-mail:

dgunawan@uow.edu.au, ajayjasra@cuhk.edu.cn, l1.south@qut.edu.au, mmoores@uow.edu.au, sumeetpals@uow.edu.au susan.wei@monash.edu

1 Schedule

Time	Monday	Tuesday	Wednesday	Thursday	Friday
09:30 - 10:30	Xin Tong	Maurizio Filippone	Marina Riabiz	Axel Finke	Joshua Bon
10:30 - 11:00	Morning Tea	Morning Tea	Morning Tea	Morning Tea	Morning Tea
11:00 - 12:00	Shuigen Liu	Dootika Vats	Patricia Ning	Susan Wei	Matti Vihola
12:00 - 14:00	Lunch	Lunch	Lunch	Lunch	Lunch
14:00 - 15:00	Collaboration	Collaboration	Collaboration	Collaboration	Collaboration
15:00 - 15:30	Tea Break	Tea Break	Tea Break	Tea Break	Tea Break
15:30 - 17:00	Collaboration	Collaboration	Collaboration	Collaboration	Collaboration

Table 1: Schedule for week 1.

Time	Monday	Tuesday	Wednesday	Thursday	Friday
09:30 - 10:30	Daniel Paulin	Clara Grazian	Sam Power	Pierre Del Moral	Sahani Pathiraja
10:30 - 11:00	Morning Tea	Morning Tea	Morning Tea	Morning Tea	Morning Tea
11:00 - 12:00	Yingzhen Li	Florian Maire	David Gunawan	Joona Karjalainen	Sumeetpal Singh
12:00 - 14:00	Lunch	Lunch	Lunch	Lunch	Lunch
14:00 - 15:00	Adrien Corenflos	Maria De Iorio	Collab/Talk	Collab/Talk	
15:00 - 15:30	Tea Break	Tea Break	Tea Break	Tea Break	
15:30 - 17:00	Collaboration	Collaboration	Collaboration	Collaboration	

Table 2: Schedule for week 2.

2 List of Abstracts

1. Name: Joshua Bon (Weeks 1 and 2)

   Title: Variance reduction in sequential Monte Carlo with knots

   Abstract: Sequential Monte Carlo (SMC) methods, or particle filters, are widely used for approximating intractable integrals, particularly in Bayesian inference and state-space models. We introduce a new variance reduction technique, the knot operation, which improves the efficiency of SMC algorithms by incorporating potential function information earlier in the computation of these particle approximations. We show that knots guarantee asymptotic variance reduction in all cases but those that are trivial, where there is no change. Moreover, an SMC algorithm with knots can be implemented with no additional computation cost for a large class of models. This class includes models where the fully adapted auxiliary particle filter (FA-APF) is applicable, and our study of the FA-APF using knots show when and how it must be altered to guarantee this variance reduction. Additionally, we demonstrate that knots induce partial orderings of Feynman-Kac models based on asymptotic variance, offering a new approach to algorithm design. Connections between knots and existing SMC algorithms are also established.

2. Name: Matti Vihola (Weeks 1 and 2)

   Title: Mixing time of the conditional backward sampling particle filter

   Abstract: The conditional backward sampling particle filter (CBPF) is a powerful Markov chain Monte Carlo sampler for general state space hidden Markov model smoothing. It was proposed as an improvement over the conditional particle filter, which is known to have an $O(T^2)$ computational time complexity under a general 'strong' mixing assumption, where T is the time horizon. We provide the first proof that the CBPF admits an $O(T log T)$ time complexity under strong mixing, complementing strong empirical evidence of the superiority of the CBPF in practice. In particular, the CBPF's mixing time is upper bounded by $O(log T)$, for any sufficiently large number of particles N that depends only on the mixing assumptions and not T. We show that an $O(log T)$ mixing time is optimal. The proof involves the analysis of a novel coupling of two CBPFs, which involves a maximal coupling of two particle systems at each time instant. The coupling is implementable, and thus can also be used to construct unbiased, finite variance, estimates of functionals which have arbitrary dependence on the latent state's path, with a total expected cost of $O(T log T)$.

3. Name: Xin Tong (Week 1)

   Title: Diffusion models for high dimensional distributions

   Abstract: Diffusion model is a popular tool to generate new data samples. However, rigorous understanding of diffusion model is still lacking. One issue is how to train these models for high dimensional problems as score function estimation is subject to the curse of dimension. Another issue is how to avoid the memorization effect, where the diffusion model is bound to generate an exact copy from the training data. We will provide solutions to the first issue by focusing on high dimensional distributions with sparse dependence. We will leverage the sparse dependence to provide a local estimation of the score functions. As for the second issue, we will modify the diffusion model in the final stage and generate new samples close to the same manifold where the training data is originated.

4. Name: Patricia Ning (Weeks 1 and 2, prefer week 1)

   Title: Scalable Bayesian Inference for Large Language Model Analysis

   Abstract: Understanding and controlling large language models (LLMs) remains a fundamental challenge in AI research. Bayesian inference provides a principled approach to modeling uncertainty and structure in LLMs, enabling improved interpretability and adaptability. In this talk, I will explore scalable Bayesian inference techniques, including sequential Monte Carlo and probabilistic programming, to enhance LLM analysis. I will discuss how these methods help uncover latent structures, improve in-context learning, and enforce constraints during text generation. By bridging theoretical insights with practical applications, I will demonstrate how Bayesian methods can provide better control, robustness, and efficiency in LLMs. Finally, I will highlight open challenges and future research directions in probabilistic modeling for large-scale deep learning systems.

5. Name: Marina Riabiz (Week 1 only)

   Abstract: Tempering enables the transformation of complicated probability distributions into simpler ones, underpinning a range of numerical methods in the Bayesian context.

   Such methods construct a sequence $\{p_t\}_{0<t<1}$, with p_0 typically representing the prior and p_1 representing the posterior of interest, and proceed to numerically approximate terms in this sequence, starting with p_0 and proceeding through intermediate distributions until an approximation to p_1 is obtained. Our contribution reveals that approximation of terms up to p_1 is not essential, as knowledge of the intermediate distributions enables posterior quantities of interest to be extrapolated.

   Specifically, we establish weak sufficient conditions under which maps of the form $t \mapsto$

$E_{X \sim p_t}[f(X)]$ are not merely smooth, but analytic, implying that knowledge of the tempered expectations corresponding to t in any open set fully determines the posterior expectation of interest.

Harnessing this result, we propose novel computational methodology for extrapolation of tempered posteriors based on the waste-free sequential Monte Carlo method of Dau and Chopin (2022) and the importance tempering method of Gramacy et al. (2010), and illustrate its effectiveness on a range of examples.

6. Name: Florian Maire (Weeks 1 and 2)

   Title: The spectrum of the Optimal Self-Regenerative and Independent Metropolis Markov chains with applications to MCMC

   Abstract: A standard route to establish quantitative asymptotic properties of MCMC algorithms used in Bayesian statistics is to study their so-called spectral gap. For example, for reversible chains, bounds on the spectral gap translates into bounds on the convergence (e.g. in total variation) and on the asymptotic variance of MCMC estimators. But more specific information on the algorithm can be inferred by knowing more about the spectrum, than only its upper end.

   Our work provides a detailed analysis of the spectrum for two elementary MCMC algorithms on general state-space, namely the Self-Regenerative and Independent Metropolis frameworks. This includes characterizing the whole spectrum of each corresponding operator and their generalized eigenspaces.

   Applications of these results allow to get a finer understanding of the asymptotic behaviour of the MCMC algorithm. In some instances, they can also provide guidelines on how to tune these algorithms if one is particularly interested in a class of test functions sharing some similarities. Equivalently, they may help choosing between several competing algorithms, for a given task at hand.

7. Name: Yingzhen Li (Week 2 only)

   Title: Variational Uncertainty Decomposition for In-Context Learning

Abstract: As large language models (LLMs) gain popularity in conducting prediction tasks in-context, understanding the sources of uncertainty in in-context learning becomes essential to ensuring reliability. The recent hypothesis of in-context learning performing predictive Bayesian inference opens the avenue for Bayesian uncertainty estimation, particularly for decomposing uncertainty into epistemic uncertainty due to lack of in-context data and aleatoric uncertainty inherent in the in-context prediction task. However, the decomposition idea remains under-explored due to the intractability of the latent parameter posterior from the underlying Bayesian model. In this work, we introduce a variational uncertainty decomposition framework for in-context learning without explicitly sampling from the latent parameter posterior, by optimising auxiliary inputs as probes to obtain an upper bound to the aleatoric uncertainty of an LLM's in-context learning procedure. Through experiments on synthetic and real-world tasks, we show quantitatively and qualitatively that the decomposed uncertainties obtained from our method exhibit desirable properties of epistemic and aleatoric uncertainty.

8. Name: Clara Grazian (Week 2 only)

   Title: Model Probability Maps of Kinetic Heterogeneity in Dynamic Total Body PET using Approximate Bayesian Computation

   Abstract: The quantification of [18F]Fluorodeoxyglucose (FDG) metabolism in oncology via positron emission tomography (PET) is essential for accurate diagnosis and treatment planning. Traditional semi-quantitative methods like the Standard Uptake Value (SUV) and Patlak plots, although widely used, often fall short in addressing the complexities of tissue heterogeneity. We introduce a voxelwise Approximate Bayesian Computation (ABC) for total body PET (TB-PET) imaging heterogeneity study and kinetic parameter estimation, demonstrating high accuracy for both estimation and model selection tasks. The study successfully computed model probability maps and kinetic parameters across an extensive real dataset of over 44 million voxels, achieving a significant computational speed-up.

9. Name: Dootika Vats (Week 1 only)

   Title: MCMC Importance Sampling via Moreau-Yosida Envelopes

Abstract: Markov chain Monte Carlo (MCMC) is the workhorse computational algorithm employed for inference in Bayesian statistics. Gradient-based MCMC algorithms are known to yield faster converging Markov chains. In modern parsimonious models, the use of non-differentiable priors is fairly standard, yielding non-differentiable posteriors. Without differentiability, gradient-based MCMC algorithms cannot be employed effectively. Recently proposed proximal MCMC approaches, however, can partially remedy this limitation. These approaches employ the Moreau-Yosida (MY) envelope to smooth the nondifferentiable prior enabling sampling from an approximation to the target posterior. In this work, we leverage properties of the MY envelope to construct an importance sampling paradigm to correct for this approximation error. We establish asymptotic normality of the importance sampling estimators with an explicit expression for the asymptotic variance which we use to derive a practical metric of sampling efficiency. Numerical studies show that the proposed scheme can yield lower variance estimators compared to existing proximal MCMC alternatives.

10. Name: Sam Power (Weeks 1 and 2)

Title: A State-Space Perspective on Modelling and Inference for Online Skill Rating

Abstract: In the quantitative analysis of competitive sports, a fundamental task is to estimate the skills of the different agents ('players') involved in a given competition based on the outcome of pairwise comparisons ('matches') between said players, often in an online setting. In this talk, I will discuss recent work in which we advocate for adoption of the state-space modelling paradigm in solving this problem. This perspective facilitates the decoupling of modeling from inference, enabling a more focused approach to development and critique of model assumptions, while also fostering the development of general-purpose inference tools. I will first describe some illustrative model classes which arise in this framework, before turning to a careful discussion of inference and computation strategies for these models. A key challenge throughout is to develop methodology which scales gracefully to problems with a large number of players and a high frequency of matches. I then conclude by describing some real-data applications of our approach, demonstrating how this framework facilitates a practical workflow across different sports. This is joint work with Samuel Duffield (Normal

Computing) and Lorenzo Rimella (Università degli Studi di Torino).

11. Name: Susan Wei (Weeks 1 and 2)

    Title: What Is Prediction For? Turning Pre-Trained Transformers into Bayesian Engines

    Abstract: The Bayesian predictive approach replaces a likelihood–prior specification with a sequence of one-step ahead predictive densities. In special cases the sequence coincides—exactly or asymptotically—with the predictions of an underlying Bayesian model; more generally it defines its own distribution, as in the martingale-posterior (MGP) framework. Existing MGP engines—chiefly the bivariate-copula recursive update—hew closely to theoretical requirements but invite scepticism about their scalability to modern data settings. Prioritising computational ease, we ask whether a black-box foundation model can act as the MGP engine. We plug TabPFN, a pretrained Transformer for tabular data, directly into the MGP framework, yielding turn-key uncertainty quantification. A single forward pass per new datapoint updates the predictive density, and the MGP construction then delivers a posterior for a target functional—no priors, tuning, or training required. Preliminary experiments are encouraging: large off-the-shelf Transformers can already function as sensible predictive rules. These findings merit deeper theoretical investigation.

12. Name: David Gunawan (Week 2)

    Title: Bayesian inference for complex spatial statistical models

    Abstract: I will talk about my recent work in spatial statistics.

    The first is Bayesian copula based spatial random effects model.

    "In this article, we develop a new class of copula-based hierarchical spatial-statistical models for large, noisy, incomplete, and non-Gaussian spatial data. Our approach includes novel constructions of Gaussian and t copulas that accommodate a spatial random effects structure, enabling low-rank representations and computationally efficient Bayesian inference for large, non-Gaussian spatial datasets. These spatial copulas are inserted into the process model of a hierarchical spatial-statistical model, which allows the models to handle measurement errors and missing data. Through extensive simulation studies, we show that our Bayesian approach delivers accurate and fast inference for parameter estimation and spatial prediction. Moreover, our copula-based hierarchical spatial-statistical models consistently outperform several benchmark models, including fixed rank kriging, across a range of non-Gaussian spatial processes. The new class of copula-based spatial random effects models is used to map atmospheric methane in the Bowen Basin, Queensland, Australia."

The second is about variational Bayes method for estimating non-Gaussian simultaneous autoregressive models with missing data.

"Standard simultaneous autoregressive (SAR) models are usually assumed to have normally distributed errors, an assumption that is often violated in real-world datasets, which are frequently found to exhibit non-normal, skewed, and heavy-tailed characteristics. New SAR models are proposed to capture these non-Gaussian features. In this project, the spatial error model (SEM), a widely used SAR-type model, is considered. Three novel SEMs are introduced that extend the standard Gaussian SEM by incorporating Student's t-distributed errors after a one-to-one transformation is applied to the response variable. Variational Bayes (VB) estimation methods are developed for these models, and the framework is further extended to handle missing response data. Standard variational Bayes (VB) methods perform well with complete datasets; however, handling missing data requires a Hybrid VB (HVB) approach, which integrates a Markov chain Monte Carlo sampler to generate missing values from their full conditional posterior distributions. The proposed VB methods are evaluated using both simulated and real-world datasets, demonstrating their robustness and effectiveness in dealing with non-normal errors and missing data in spatial models. Although the method is demonstrated using SAR models, the proposed model specifications and estimation approaches are widely applicable to various types of models for handling non-Gaussian data with missing values."

13. Name: Sumeetpal Singh (Weeks 1 and 2)

    Title: Bayesian learning of the optimal action-value function in a Markov decision process Abstract: The Markov Decision Process (MDP) is a popular framework for sequential decision-making problems, and uncertainty quantification is an essential component of it to learn optimal decision-making strategies. In particular, a Bayesian framework is used to maintain beliefs about the optimal decisions and the unknown ingredients of the model, which are also to be learned from the data, such as the rewards and state dynamics. However, many existing Bayesian approaches for learning the optimal decision-making strategy are based on unrealistic modelling assumptions and utilise approximate inference techniques. This raises doubts whether the benefits of Bayesian uncertainty quantification are fully realised or can be relied upon.

We focus on infinite-horizon and undiscounted MDPs, with finite state and action spaces, and a terminal state. We provide a full Bayesian framework, from modelling to inference to decision-making. For modelling, we introduce a likelihood function with minimal assumptions for learning the optimal action-value function based on Bellman's optimality equations, analyse its properties, and clarify connections to existing works. For deterministic rewards, the likelihood is degenerate and we introduce artificial observation noise to relax it, in a controlled manner, to facilitate more efficient Monte Carlo-based inference. For inference, we propose an adaptive sequential Monte Carlo algorithm to both sample from and adjust the sequence of relaxed posterior distributions. For decision-making, we choose actions using samples from the posterior distribution over the optimal strategies. While commonly done, we provide new insight that clearly shows that it is a generalisation of Thompson sampling from multi-arm bandit problems. Finally, we evaluate our framework on the Deep Sea benchmark problem and demonstrate the exploration benefits of posterior sampling in MDPs.

14. Name: Axel Finke (Weeks 1 and 2)

    Title: Particle-MALA and Particle-mGRAD: Gradient-based MCMC methods for high-dimensional state-space models

    Abstract: State-of-the-art methods for Bayesian inference in state-space models are (a) conditional sequential Monte Carlo (CSMC) algorithms; (b) sophisticated "classical" gradient-based MCMC algorithms like MALA, the preconditioned Crank—Nicholson Langevin (PCNL) algorithm, or the mGRAD algorithm recently introduced in Titsias & Papaspiliopoulos

(JRSSB, 2018):

The former (i.e. CSMC algorithms) propose N particles at each time step to exploit the model's decorrelation-over-time property and scale favourably with the time horizon, T, but break down when the dimension of the latent states, D, increases. The latter (i.e. "classical" MCMC algorithms) leverage gradient-/prior-informed local proposals to scale favourably with D but exhibit sub-optimal scaling with T due to a lack of model-structure exploitation.

We introduce methods combining the strengths of both approaches:

(a) The first, Particle-MALA, spreads N particles around the current state using gradient information, extending MALA to $T > 1$ time steps and $N > 1$ proposals.

(b) The second, Particle-mGRAD, additionally incorporates (conditionally) Gaussian prior dynamics into the proposal, extending mGRAD to $T, N > 1$. Particle-mGRAD provably resolves the tuning problem of choosing between CSMC (superior for informative dynamics) and Particle-MALA (superior for uninformative dynamics).

We similarly extend other "classical" MCMC approaches like auxiliary MALA (aMALA), auxiliary gradient (aGRAD), and PCNL.

15. Name: Sahani Pathiraja (Weeks 1 and 2)

Title: On connections between sequential Bayesian inference and evolutionary dynamics

Abstract: It has long been posited that there is a connection between the dynamical equations describing birth-death & evolutionary processes in biology (so-called replicator-mutator dynamics) and sequential Bayesian learning methods. This talk describes new research in which this precise connection is rigorously established in the continuous time setting. Here we focus on a class of interacting particle methods for solving the sequential Bayesian inference problem which are characterised by a McKean-Vlasov Stochastic differential equation. Of particular importance is a piecewise smooth approximation of the observation path from which the discrete time filtering equations are shown to converge to a Stratonovich interpretation of the Kushner equation. This smooth formulation will then be used to draw precise connections between nonlinear filtering and replicator-mutator dynamics. Additionally, gradient flow formulations will be investigated as well as a particular form of replicator-mutator dynamics which is shown to be beneficial for filtering with misspecified models. It is hoped this work will spur further research into exchanges between sequential learning and evolutionary biology and to inspire new algorithms in filtering and sampling.

16. Name: Pierre Del Moral (Weeks 1 and 2)

    Title: On the stability of Schrödinger bridges and Sinkhorn semigroups

    Abstract: Entropic optimal transport techniques, including Schrödinger bridges and Sinkhorn algorithm have become state-of-the-art tools in generative modeling and machine learning. For finite spaces, Sinkhorn iterations reduces to the simple iterative proportional fitting procedure. However, in more general settings these iterations are based on nonlinear conditional/conjugate transformations and exact finite-dimensional solutions cannot be computed. In the first part of the talk we present a finite-dimensional recursive formulation of Sinkhorn iterations and closed form expressions of Schrödinger bridge maps for general Gaussian multivariate models in terms of a class of discrete time algebraic Riccati equations (a.k.a. DARE) arising in filtering and optimal control theory. There are very few articles on the exponential convergence of Sinkhorn iterates on non-compact spaces and unbounded cost functions that apply to Gaussian models. In the second part of the lecture, we present some methodologies to analyze the stability of Sinkhorn semigroups and Schrödinger bridges for general marginals and reference distributions, including Lyapunov approaches and transportation cost inequalities. The strength of this approach is that it is applicable to a large class of models arising in machine learning and artificial intelligence algorithms. We illustrate the impact of our results in the context of regularized entropic transport, proximal samplers and diffusion generative models as well as diffusion flow matching models.

17. Name: Maria de Iorio (Week 2 only)

    Latent Structure Learning in Multi-View Data

    Mixture models are essential tools for analysing heterogeneous populations. From a Bayesian nonparametric perspective, we introduce a novel class of priors, the Normalised Independent Point Process, and develop both marginal and conditional algorithms for finite mixture models with a random number of components. By employing an auxiliary variable MCMC approach, we efficiently address the challenges posed by intractable posterior distributions, creating a flexible and extensible framework for Bayesian modelling.

In this talk, we present key methodological advances for the analysis of multiview data, where complex dependencies arise across multiple data domains. To address these challenges, we introduce a probabilistic framework for conditional partial exchangeability, specifically designed for multiview and longitudinal settings. This framework enables the construction of flexible random partitions that vary across features, allowing for nuanced modelling of dependencies within and between data domains.

Moreover, the framework can be extended to link subject-level partitions across datasets via an underlying shared latent structure, promoting information sharing and facilitating robust, interpretable inference. By jointly modelling within-subject dependencies and marginal relationships across views, the approach enhances both the flexibility and interpretability of the resulting clustering structures.

We also describe scalable Bayesian solutions, such as Bayesian Distance Clustering, which strike a balance between computational efficiency and probabilistic rigour. These methods are particularly effective in handling large, high-dimensional datasets, offering accurate predictive performance while maintaining scalability.

We validate these methods through extensive simulations and applications to real-world datasets, demonstrating their effectiveness across a variety of contexts. These advancements provide a robust and versatile toolkit for addressing modern data analysis challenges, particularly in the context of multiview and longitudinal data.

Bayesian Structural Learning: Applications in Biological Systems

Graphical models provide a powerful framework for understanding the conditional independence structure in multivariate data, offering valuable insights into complex dependencies among variables. In this work, we propose a novel Bayesian approach for inferring multiple graphical models, leveraging information across heterogeneous groups. Moreover, we introduce a formal statistical method for the differential analysis of molecular associations through network representations. Differential networks reveal novel insights by characterising differences in biological states. We demonstrate how the comprehensive analysis of molecular association patterns and their changes under varying physiological conditions can elucidate the biological basis of disease-specific phenotype variations.

Our methodology extends beyond the inference of individual edges in latent graphs to incorporate more complex structures, addressing the growing interest in sub-graph patterns such as communities. Such structures are crucial for improved information retrieval and interpretability. To address these challenges, we propose two complementary strategies.

Firstly, we exploit advances in random graph theory by incorporating Bayesian nonparametric stochastic blockmodels as priors on the graph. This approach enables the propagation of uncertainty in graph estimation to large-scale structure learning.

Secondly, we propose a hierarchical graphical model, referred to as graph of graphs, that combines Bayesian nonparametrics with Voronoi tessellation. In this approach, nodes are grouped into "supernodes" using a size-biased tessellation prior, allowing dependencies within supernodes to be captured effectively. Relationships among these supernodes are then represented by a Gaussian graphical model, facilitating inference on "superedges" and large-scale structural relationships.

The methodological framework is underpinned by tailored Markov chain Monte Carlo algorithms, designed to support parallel computations and make our approach feasible for high-dimensional datasets. We validate our methods through extensive simulations and applications to real-world datasets, including metabolomics and transcriptomics, demonstrating their effectiveness in uncovering meaningful biological insights.

18. Name: Adrien Corenflos (Weeks 1 and 2)

    Title: A random-walk in the land of denoising diffusions

    Abstract: Conditional sampling within denoising diffusion models has been the main driver of mass-market applications, but to this day all direct conditioning techniques either (i) are biased in some way, and do not provide consistent posterior distributions p(x—y) p(x)p(y — x) for a given likelihood model p(y — x) and a denoising prior p(x); (ii) are highly specialized and limit the type of queries that can be made (e.g., only image in-painting, or Gaussian deblurring; (iii) exhibit a large variance which dominates the removal of bias. A characteristic of most of these prior works, however, is that they rely on global approximations to the posterior distribution, which are hardly practical in such high-dimensional systems.

In this work, we describe a local alternative: we define an analogue to pre-conditioned Crank-Nicholson MCMC for distributions defined as the posterior of denoising diffusion model. Our algorithm takes the form of a particle Gibbs sampler for a Feynman–Kac model targetting the full denoising diffusion model. We discuss its properties and illustrate the resulting performance of our method on posterior sampling tasks for denoising diffusion models.

19. Name: Daniel Paulin (Week 2)

    Title: Unbiased MCMC for high dimensional Bayesian inference

    Abstract: We present an unbiased method for Bayesian posterior means based on kinetic Langevin dynamics that combines advanced splitting methods with enhanced gradient approximations. Our approach avoids Metropolis correction by coupling Markov chains at different discretization levels in a multilevel Monte Carlo approach. Theoretical analysis demonstrates that our proposed estimator is unbiased, attains finite variance, and satisfies a central limit theorem. It can achieve accuracy $O(eps)$ with $O(d^1/4/sqrt(eps))$ expected gradient evaluations, without assuming warm start. We exhibit similar bounds using both approximate and stochastic gradients, and our method's computational cost is shown to scale independently of the size of the dataset. The proposed method is tested using a multinomial regression problem on the MNIST dataset and a Poisson regression model for soccer scores. Experiments indicate that the number of gradient evaluations per effective sample is independent of dimension, even when using inexact gradients. For product distributions, we give dimension-independent variance bounds. Our results demonstrate that in large-scale applications, the unbiased algorithm we present can be 2-3 orders of magnitude more efficient than the gold-standard randomized Hamiltonian Monte Carlo. This is joint work with Ben

Leimkuhler, Neil Chada and Peter Whalley.

20. Name: Shuigen Liu (Week 1)

    Title: Stein's method for marginals on large graphical models

    Abstract: Many spatial models exhibit locality structures that effectively reduce their intrinsic dimensionality, enabling efficient approximation and sampling of high-dimensional distributions. However, existing approximation techniques mainly focus on joint distributions, and do not guarantee accuracy for low-dimensional marginals. By leveraging the locality structures, we establish a dimension independent uniform error bound for the marginals of approximate distributions. Inspired by the Stein's method, we introduce a novel δ-locality condition that quantifies the locality in distributions, and link it to the structural assumptions such as the sparse graphical models. The theoretical guarantee motivates the localization of existing sampling methods, as we illustrate through the localized likelihood-informed subspace method and localized score matching. We show that by leveraging the locality structure, these methods greatly reduce the sample complexity and computational cost via localized and parallel implementations.

21. Name: Joona Karjalainen (Weeks 1 and 2)

    Title: On forgetting properties of particle filter algorithms

    Abstract: We discuss the forgetting properties of the particle filter (a sequential Monte Carlo algorithm) when its state — the collection of particles — is regarded as a Markov chain. Under a strong mixing assumption on the underlying Feynman–Kac model, we find that the particle filter forgets its state (in total variation sense) in O(log N) time, where N is the number of particles and time refers to the number of particle filter algorithm steps, each comprising a selection (or resampling) and mutation (or prediction) operation. An example shows that the rate is optimal. We discuss implications of our findings e.g. to coupling particle filters and propagation of chaos.

22. Name: Maurizio Filippone (Week 1, 22 - 27th June 2025)

    Title: GANs are Secretly Bayesian

Abstract: Generative Adversarial Networks (GANs) are popular models achieving impressive performance in various generative modeling tasks. In this work, we aim at explaining the undeniable success of GANs by interpreting them as probabilistic generative models. In this view, GANs transform a distribution over latent variables Z into a distribution over inputs X through a function parameterized by a neural network, which is usually referred to as the generator. This probabilistic interpretation enables us to cast the GAN adversarial-style optimization as a proxy for marginal likelihood optimization. More specifically, it is possible to show that marginal likelihood maximization with respect to model parameters is equivalent to the minimization of the Kullback-Leibler (KL) divergence between the true data generating distribution and the one modeled by the GAN. By replacing the KL divergence with other divergences and integral probability metrics we obtain popular variants of GANs such as f-GANs, Wasserstein-GANs, and Maximum Mean Discrepancy (MMD)-GANs. This connection has profound implications because of the desirable properties associated with marginal likelihood optimization, such as (i) lack of overfitting, which explains the success of GANs, and (ii) allowing for model selection. The aim of this talk, is to stimulate a discussion around the use of this framework for general high-dimensional inference problems.