3RD WORKSHOP ON

LONDON / 2026

PROGRAMME / 16 NOVEMBER 2026

Gradient Flows

for Sampling, Inference, and Learning

23–25 November 2026

Imperial College London

170 Queen's Gate / South Kensington

Programme & *book of abstracts*

ORGANISED BY

O. Deniz Akyildiz / Louis Sharrock / Grigorios A. Pavliotis

WITH SUPPORT FROM

Prob_AI

G RESEARCH

Imperial × QRT Labs

ROYAL STATISTICAL SOCIETY

DATA | EVIDENCE | DECISIONS

Statistical Computing Section

23–25 NOVEMBER / LONDON

Timetable & Book of Abstracts

The 3rd Workshop on Gradient Flows for Sampling, Inference, and Learning brings together researchers in statistics, machine learning, and applied mathematics. This booklet presents the workshop timetable and the complete collection of invited talk and poster abstracts.

VENUE

170 Queen's Gate
Imperial College London
South Kensington, London
Talks: 170QG Council Room
Posters: 170QG Drawing Room

DATES & TIMES

23–25 November 2026
Monday to Wednesday
All times are local to London.

ORGANISERS

O. Deniz Akyildiz	Imperial College London
Louis Sharrock	University College London
Grigorios A. Pavliotis	Imperial College London

CONTACT

deniz.akyildiz@imperial.ac.uk / gradientflows.com

INSIDE THIS PROGRAMME

Monday programme 3
Tuesday programme and poster session 4
Wednesday programme 5
Talk abstracts 6
Poster directory and book of abstracts 22

Programme / 16 November 2026
15 invited talks and 17 poster abstracts.

DAY 01 / 23 NOVEMBER 2026

Monday.

170QG Council Room — Workshop programme

Morning chair: O. Deniz Akyildiz

09:45–10:00	**Opening remarks** O. Deniz Akyildiz
10:00–10:50 TALK 01	**Sebastian Reich** University of Potsdam Interacting particle formulations for stochastic optimal control
10:50–11:40 TALK 02	**Giacomo Borghi** TU Munich Convergence of replicator–mutator particle systems via a bounded Wasserstein distance
11:40–12:10	**Coffee break**
12:10–13:00 TALK 03	**Claudia Schillings** Freie Universität Berlin Ensemble Kalman Inversion/Optimisation: Choosing and Escaping the Subspace
13:00–14:00	**Lunch break**

Afternoon chair: Louis Sharrock

14:00–14:50 TALK 04	**Andrew Duncan** Imperial College London Soft Specialists: Mean-Field Langevin Dynamics and the Emergence of Diversity in LLM Post-Training
14:50–15:40 TALK 05	**Pierre Monmarché** Université Gustave Eiffel Convergence rates for Wasserstein gradient flows of free energies and associated particle systems
15:40–16:10	**Coffee break**
16:10–17:00 TALK 06	**Anna Korba** ENSAE Paris / CREST Stein Transport for Generative Data Augmentation
17:00	**Closing notes**

DAY 02 / 24 NOVEMBER 2026

Tuesday.

170QG Council Room — Workshop programme

Morning chair: Louis Sharrock

09:45–10:00	**Opening remarks** Louis Sharrock
10:00–10:50 TALK 07	**Francesca Romana Crucinio** University of Turin Wasserstein–Fisher–Rao gradient flows for sampling: algorithms, convergence, and numerical analysis
10:50–11:40 TALK 08	**Andrea Agazzi** University of Bern A multiscale analysis of information propagation dynamics in deep transformer models
11:40–12:10	**Coffee break**
12:10–13:00 TALK 09	**Olivier Zahm** Inria / Université Grenoble Alpes Gradient-free Riemannian Langevin sampler
13:00–14:00	**Lunch break**

Afternoon chair: Grigorios A. Pavliotis

14:00–14:50 TALK 10	**Marie-Therese Wolfram** University of Warwick On Boltzmann mean field games
14:50–15:40 TALK 11	**Urbain Vaes** Inria Paris / CERMICS Rare events and the narrow escape problem
15:40–16:00	**Coffee break / poster setup**
16:00–18:00	**Poster session** 170QG Drawing Room / Directory: page 22

DAY 03 / 25 NOVEMBER 2026

Wednesday.

170QG Council Room — Workshop programme

Morning chair: Grigorios A. Pavliotis

09:00–09:10	**Opening remarks** Grigorios A. Pavliotis
09:10–10:00 TALK 12	**André Schlichting** University of Ulm A fourth-order Langevin dynamics? Diffusive transport, the DLSS equation and structure-preserving discretizations
10:00–10:50 TALK 13	**Anna Shalova** University of Amsterdam Gradient flow of a random quadratic functional: an RDS perspective
10:50–11:20	**Coffee break with snacks** 30 minutes
11:20–12:10 TALK 14	**Eric Vanden-Eijnden** New York University The Inner Workings of Generative Flows and Maps
12:10–13:00 TALK 15	**Hong Duong** University of Birmingham GENERIC formulation and asymptotic limits for kinetic models
13:00–13:10	**Workshop closing** O. Deniz Akyildiz

POSTER AWARDS / G-RESEARCH

Poster awards will be announced at the workshop closing.

BOOK OF ABSTRACTS

Talk *abstracts.*

23–25 November 2026

170QG Council Room

TALK	TITLE / SPEAKER	P.
T01	Interacting particle formulations for stochastic optimal control Sebastian Reich	7
T02	Convergence of replicator–mutator particle systems via a bounded Wasserstein distance Giacomo Borghi	8
T03	Ensemble Kalman Inversion/Optimisation: Choosing and Escaping the Subspace Claudia Schillings	9
T04	Soft Specialists: Mean-Field Langevin Dynamics and the Emergence of Diversity in LLM Post-Training Andrew Duncan	10
T05	Convergence rates for Wasserstein gradient flows of free energies and associated particle systems Pierre Monmarché	11
T06	Stein Transport for Generative Data Augmentation Anna Korba	12
T07	Wasserstein–Fisher–Rao gradient flows for sampling: algorithms, convergence, and numerical analysis Francesca Romana Crucinio	13
T08	A multiscale analysis of information propagation dynamics in deep transformer models Andrea Agazzi	14
T09	Gradient-free Riemannian Langevin sampler Olivier Zahm	15
T10	On Boltzmann mean field games Marie-Therese Wolfram	16
T11	Rare events and the narrow escape problem Urbain Vaes	17
T12	A fourth-order Langevin dynamics? Diffusive transport, the DLSS equation and structure-preserving discretizations André Schlichting	18
T13	Gradient flow of a random quadratic functional: an RDS perspective Anna Shalova	19
T14	The Inner Workings of Generative Flows and Maps Eric Vanden-Eijnden	20
T15	GENERIC formulation and asymptotic limits for kinetic models Hong Duong	21

TALK ABSTRACT / T01

Monday 23 November 10:00–10:50 170QG Council Room

Interacting particle formulations for stochastic optimal control

Sebastian Reich

University of Potsdam

The work is motived by optimal control problems for partially observed Markov decision processes (POMDP). It builds upon two strands of work: Interacting particle formulations for stochastic filtering in continuous time, such as the ensemble Kalman-Bucy filter (EnKBF), and a novel mean field formulation of the Pontryagin minimum principle (PMP) for (fully observed) Markov decision processes (MDP). The first part of the talk will motive the mathematical setting in the context of the digital twin paradigm. The second part, develops the novel mean field PMP formulation for MDP including numerical demonstrations. The final part outlines how to combine particle based filtering algorithms, such as the EnKBF, with mean field PMP in order to tackle optimal control problems for POMDP.

TALK ABSTRACT / T02

Monday 23 November 10:50–11:40 170QG Council Room

Convergence of replicator–mutator particle systems via a bounded Wasserstein distance

Giacomo Borghi

TU Munich

Interacting particle systems with selection and mutation arise in many contexts: biological modelling, evolutionary optimization, particle filtering, and the approximation of Wasserstein–Fisher–Rao gradient flows. A quantitative link between the microscopic dynamics and their continuum description is important for identifying appropriate parameter regimes and scalings, particularly for numerical algorithms. In this talk, I will consider the replicator–mutator PDE, which combines diffusion with a source term that depends on an external potential. I will show how to derive quantitative error estimates for the distance between the PDE solution and the empirical measure of an unweighted particle system. The proof uses a Wasserstein-1 distance with a bounded ground cost, which allows us to treat unbounded potentials on unbounded state spaces. Interestingly, a dependence on the space dimension already appears in the coupling estimate, where mean-field-type systems typically exhibit a dimension-independent Monte Carlo rate instead. Based on the work:

G. Borghi, “Convergence of replicator–mutator particle systems via a bounded Wasserstein distance” (2026, soon on arXiv)

TALK ABSTRACT / T03

Monday 23 November 12:10–13:00 170QG Council Room

Ensemble Kalman Inversion/Optimisation: Choosing and Escaping the Subspace

Claudia Schillings

Freie Universität Berlin

Ensemble Kalman inversion (EKI) is a derivative-free method for inverse problems. In its continuous-time limit it is a gradient flow preconditioned by the ensemble covariance, so its dynamics are confined to the span of the initial ensemble. This subspace property is both a strength and a limitation: it reduces the dimension in which the method works, but the quality of the solution depends on how well the subspace captures the problem. In this talk we take both views. First, we embrace the subspace and ask how to choose it well. Second, we ask how to escape it.

In the first part, we present a subspace optimisation approach in which the subspace is built adaptively, steering the dynamics towards the directions that matter most for the problem. In the second part, we introduce a randomised Ensemble Kalman Optimisation scheme that adds noise to the ensemble update. The noise turns the degenerate preconditioned flow into a stochastic dynamics that can leave the initial subspace. We also discuss a projected variant that controls the ensemble spread with a moving-ball constraint in observation space, which can be read as a constrained gradient flow. We outline theoretical results on how both schemes behave and illustrate their performance on representative examples.

The first part is joint work with Ruben Harris (FU Berlin). The second part is joint work with Jonas Latz (University of Manchester), Alix Leroy (University of Edinburgh) and Simon Weissmann (University of Mannheim).

TALK ABSTRACT / T04

Monday 23 November 14:00–14:50 170QG Council Room

Soft Specialists: Mean-Field Langevin Dynamics and the Emergence of Diversity in LLM Post-Training

Andrew Duncan

Imperial College London

Post-training a large language model usually returns a single set of weights, even when the data are heterogeneous or contradictory. We study a variational objective over probability measures on parameter space in which each observation is scored by the α-power mean of its likelihood, interpolating between generalised variational Bayes and predictively oriented posteriors. The resulting free energy is linearly convex and entropy-regularised, so its Wasserstein gradient flow is a mean-field Langevin dynamics whose interaction acts through likelihood-based responsibilities, and whose unique stationary state satisfies a self-consistent fixed-point equation. The associated particle system gives a practical algorithm: an ensemble of LoRA adapters trained with soft routing of examples between members, which I will illustrate on supervised fine-tuning and preference optimisation. While convexity rules out phase transitions at positive temperature, I will argue that sharp transitions nonetheless appear in the limit where regularisation is small relative to the data, and that they govern when the ensemble specialises through a nucleation process. I will discuss different scaling regimes, the metastability of the non-convex particle system, and some connections to phase transitions and coarsening in McKean–Vlasov equations.

TALK ABSTRACT / T05

Monday 23 November 14:50–15:40 170QG Council Room

Convergence rates for Wasserstein gradient flows of free energies and associated particle systems

Pierre Monmarché

Université Gustave Eiffel

For gradient flows, exponential convergence is characterised by a Polyak–Łojasiewicz (PL) inequality. For the Wasserstein gradient flows of entropic-regularised energies, the PL inequality for the corresponding system of mean-field interacting particles is a log-Sobolev inequality (LSI). A fast convergence thus corresponds to a log-Sobolev constant independent from the number of particles. We will review some recent results on this topic : equivalence between the mean-field PL and uniform LSI, degenerate situations at phase transitions, local convergence, Nesterov acceleration, time-discretisation, and sampling of local minimisers.

TALK ABSTRACT / T06

Monday 23 November 16:10–17:00 170QG Council Room

Stein Transport for Generative Data Augmentation

Anna Korba

ENSAE Paris / CREST

Generating additional samples from limited observations is important when data acquisition or Monte Carlo simulation is costly. We propose a nonparametric approach to generative modeling based on Stein operators and probability transport that learns from target samples alone. Using an Ornstein–Uhlenbeck probability path, we derive a tractable conditional regression objective for learning the transport field, without evaluating the target density or its gradient. Restricting the field to a reproducing kernel Hilbert space yields closed-form kernel ridge regression solutions at each time step, with deterministic, stochastic, and coordinatewise variants. We establish convergence rates for the excess conditional Stein risk and show that the regression targets are noise-free given the augmented inputs, identifying a structural distinction from denoising score matching. Experiments on Gaussian mixtures and Bayesian posterior emulation show effective generation in small-data regimes and greater robustness to hyperparameter choices than kernel score matching. On tabular benchmarks, we obtain competitive downstream predictive performance and explore the use of generated samples for data augmentation. This is a joint work with Clémentine Chazal, Linfeng Wang, Nikolas Nusken.

TALK ABSTRACT / T07

Tuesday 24 November 10:00–10:50 170QG Council Room

Wasserstein–Fisher–Rao gradient flows for sampling: algorithms, convergence, and numerical analysis

Francesca Romana Crucinio

University of Turin

Sampling from complex probability distributions can be formulated as an optimisation problem over probability measures, where the objective is to minimise the Kullback–Leibler (KL) divergence to a target distribution. Wasserstein–Fisher–Rao (WFR) gradient flows provide a natural framework for this task by combining Wasserstein transport, which moves probability mass through space, with Fisher–Rao dynamics, which allow changes in mass through birth–death mechanisms.

In this talk, I will present recent developments on WFR gradient flows from three complementary perspectives: convergence theory, numerical approximation, and sampling algorithms.

First, I will turn to the convergence theory of WFR flows targeting distributions known up to a normalisation constant. For strongly log-concave targets satisfying suitable curvature conditions, I will show that strong log-concavity is preserved along the flow. This property yields explicit non-asymptotic convergence rates for the symmetrised KL divergence, without requiring a warm start. Moreover, the rates decompose additively into Wasserstein and Fisher–Rao contributions.

Second, I will discuss operator splitting schemes for approximating WFR dynamics by sequentially applying Wasserstein and Fisher–Rao flows. I will present variational characterisations of the resulting time-discrete dynamics and examine the role of operator ordering. In particular, I will show that, with an appropriate step size and ordering, a splitting scheme can converge to the target faster than the exact WFR flow in model time, and identify conditions under which each splitting order is advantageous.

Finally, I will discuss the connections between KL-divergence minimisation flows and Monte Carlo methods, highlighting how the WFR geometry can be exploited to construct practical sampling algorithms based on importance sampling and sequential Monte Carlo. Building on these connections, I will introduce a novel algorithm for approximating WFR flows and present numerical comparisons with established Monte Carlo methods.

Overall, the talk will illustrate how the interplay between transport and mass variation in WFR geometry leads to new theoretical insights, numerical methods, and practical sampling algorithms.

TALK ABSTRACT / T08

Tuesday 24 November 10:50–11:40 170QG Council Room

A multiscale analysis of information propagation dynamics in deep transformer models

Andrea Agazzi

University of Bern

In this talk, we study the evolution of tokens across the depth of encoder-only transformer models at inference time, modeling them as a system of interacting particles in the infinite-depth limit. Motivated by techniques for extending the context length of large language models, we focus on the moderate interaction regime, where the number of tokens is large and the inverse temperature parameter scales accordingly. In this setting, the dynamics exhibit a multiscale structure. Using PDE analysis, we identify different phases depending on the choice of parameters.

TALK ABSTRACT / T09

Tuesday 24 November 12:10–13:00 170QG Council Room

Gradient-free Riemannian Langevin sampler

Olivier Zahm

Inria / Université Grenoble Alpes

In this talk, we address the problem of efficiently sampling multimodal probability distributions, where standard Markov Chain Monte Carlo methods often suffer from poor mixing and mode trapping. To mitigate these issues, we propose GRiLS (Gradient-free Riemannian Langevin Sampler), a novel preconditioning framework that enhances exploration without requiring gradient evaluations. Our approach introduces a Riemannian metric which reshapes the local geometry in order to facilitate transitions across modes. The resulting gradient-free MCMC algorithm is particularly suitable for complex, computationally expensive targets where derivatives are unavailable or impractical.

TALK ABSTRACT / T10

Tuesday 24 November 14:00–14:50 170QG Council Room

On Boltzmann mean field games

Marie-Therese Wolfram

University of Warwick

In this talk I will focus on a generalisation of a Boltzmann mean field game (BMFG) for knowledge growth, originally introduced by the economists Lucas and Moll. In BMFG the evolution of the agent density with respect to their knowledge level is described by a Boltzmann equation. Agents increase their knowledge through binary interactions with others; their increase is modulated by the interaction and learning rate.

I will discuss different aspects and challenges of this model – first the existence of special solutions, which relate to overall exponential growth. Second the connection to variational mean-field games and gradient flows in suitable scaling limits.

TALK ABSTRACT / T11

Tuesday 24 November 14:50–15:40 170QG Council Room

Rare events and the narrow escape problem

Urbain Vaes

Inria Paris / CERMICS

Many stochastic dynamics, such as Langevin dynamics in molecular simulation, are metastable: they remain trapped for very long times in some region of the state space before transitioning to another one. When the trapping is due to energetic barriers, such rare events are well described by Eyring-Kramers type formulas, but metastability can also have an entropic origin. The narrow escape problem is a prototypical model of this situation: a Brownian particle evolves in a domain with reflecting boundary, except on small windows through which it can escape.

In this talk, I will present recent work on this problem based on the quasi-stationary distribution approach to metastability. Starting from the quasi-stationary distribution in the domain, we derive asymptotic expansions, with error bounds, of the mean exit time and of the law of the exit point in the limit where the sizes of the windows tend to zero, in general domains of arbitrary dimension. The analysis relies on the study of the principal eigenvalue and eigenfunction of the Laplacian with mixed Dirichlet-Neumann boundary conditions. Theoretical predictions are illustrated by numerical simulations based on dedicated Monte Carlo techniques.

Joint work with Louis Carillo, Tony Lelièvre, Thomas Normand and Gabriel Stoltz.

TALK ABSTRACT / T12

Wednesday 25 November 09:10–10:00 170QG Council Room

A fourth-order Langevin dynamics? Diffusive transport, the DLSS equation and structure-preserving discretizations

André Schlichting

University of Ulm

Langevin dynamics is the Wasserstein gradient flow of the relative entropy with respect to a target measure. We ask whether a natural fourth-order analogue exists. Our starting point is diffusive transport, which replaces the continuity equation in the Benamou–Brenier formula by a second-order one and relaxes martingale transport. In this geometry, the gradient flow of the entropy is the fourth-order Derrida–Lebowitz–Speer–Spohn (DLSS) equation.

A pure diffusive transport flow conserves the mean and cannot sample on its own. Combining it with the Wasserstein geometry leads to a fourth-order Langevin equation. For Gaussian targets, this is the confined DLSS equation. It has the same spectral gap as the Ornstein–Uhlenbeck process, but higher Hermite modes decay quadratically faster.

We present a structure-preserving discretization with an exact discrete Gaussian equilibrium and a decay rate uniform in the mesh size. It has a particle interpretation as a chemical reaction network of single-particle and pair jumps.

Joint work with Daniel Matthes, Eva-Maria Rott and Giuseppe Savaré, and with Alexander Mielke and Artur Stephan.

TALK ABSTRACT / T13

Wednesday 25 November 10:00–10:50 170QG Council Room

Gradient flow of a random quadratic functional: an RDS perspective

Anna Shalova

University of Amsterdam

We introduce the Random Quadratic Form (RQF): a stochastic differential equation which formally corresponds to the gradient flow of a random quadratic functional on a sphere, and show how the properties of the deterministic gradient flow translate into the random setting. In particular, we prove that the random attractor of the RQF inherits the structure of the minimisers of its deterministic counterpart. We then consider the RQF model perturbed by a random independent forcing and recover the resulting multi-scale synchronising behaviour which illustrates the ‘metastable synchronisation by noise’ phenomenon.

The RQF model is motivated by the study of the role of linear layers in transformers and demonstrates the *synchronization by common noise* phenomena arising in the simplified models of transformers. In particular, we provide an alternative (independent of self-attention) explanation of the clustering behaviour in deep transformers and show that tokens cluster even in the absence of the self-attention mechanism.

TALK ABSTRACT / T14

Wednesday 25 November 11:20–12:10 170QG Council Room

The Inner Workings of Generative Flows and Maps

Eric Vanden-Eijnden

New York University

Generative models based on dynamical transport have recently led to significant advances in unsupervised learning. At mathematical level, these models are primarily designed around the construction of a map between two probability distributions that transform samples from the first into samples from the second. While these methods were first introduced in the context of image generation, they have found a wide range of applications. In this talk, I will discuss the mathematical underpinning of generative models based on flows and their associated maps, and show how a better understanding of their inner workings can help improve their design. These results indicate how to structure the transport to best reach complex target distributions while maintaining computational efficiency, both at learning and sampling stages.

TALK ABSTRACT / T15

Wednesday 25 November 12:10–13:00 170QG Council Room

GENERIC formulation and asymptotic limits for kinetic models

Hong Duong

University of Birmingham

In non-equilibrium thermodynamics, GENERIC (General Equation for Non-Equilibrium Reversible-Irreversible Coupling), which was introduced by Grmela and Oettinger in 1997 in the context of complex fluids, describes a class of thermodynamic evolution equations for systems with both reversible and irreversible dynamics. It is formulated in terms of an energy functional, an entropy functional, and two geometric structures for the Hamiltonian and dissipative dynamics, such that it automatically fulfils the first and second laws of thermodynamics by ensuring conservation of energy and monotonic increase of entropy. Over the last three decades, the GENERIC framework has proved to be a versatile modelling tool for a variety of complex systems in physics and engineering, ranging from anisotropic inelastic solids to viscoplastic solids and thermoelastic dissipative materials. In recent years, there has been a considerable progress on establishing a mathematically rigorous foundation for the GENERIC framework.

In this talk, I will show rigorously the GENERIC formulation and using it to derive asymptotic limits for various kinetic models, including the (fuzzy) Boltzmann and Landau equations, and the kinetic wave equations. This talk is based on joint works with Zihui He and Boris Golubkov (Bielefield University).

BOOK OF ABSTRACTS

Poster *directory.*

Tuesday 24 November / 16:00–18:00

170QG Drawing Room

ID	TITLE / PRESENTER(S)	P.
P01	Efficient Function Approximation Under Heteroskedastic Noise Yifu Zhang	23
P02	On the implicit regularization of Langevin dynamics with projected noise Adrien Vacher	24
P03	Convergence of adaptive self-normalized importance sampling Nicola Branchini	25
P04	Consistency Regularised Gradient Flows for Inverse Problems Tim Y. J. Wang / Alessio Spagnoletti	26
P05	A Unifying View of Variational Generative Wasserstein Flows Paul Caucheteux	27
P06	The Fenchel Game of Underdamped Langevin: Insights into Accelerated Convergence Alexandra Borkowski	28
P07	Annealed Underdamped Langevin Sampling Andreas Habring	29
P08	Fractional Posteriors under Model Misspecification in Generalised Linear Models Emily Ehrhardt	30
P09	Sobolev Regularized MMD Gradient Flow Chenyang Tian	31
P10	L^2 over Wasserstein: Statistical analysis for optimal transport Rohan Shenoy / Pengcheng Ye	32
P11	Robust MMD-Regularised Particle Filter as a Wasserstein Gradient Flow Shang Ling Yap	33
P12	Sharp Asymptotics for Regularised Optimal Transport Carlos Alberto Cardoso Correia Perelló	34
P13	Optimal Stabilization for Global Optimization Yuyang Huang	35
P14	Coresets for preserving attention and learning dynamics Paula Cordero Encinar	36
P15	An analysis of Bures-Wasserstein variational inference with Gaussian Mixtures Marguerite Petit Talamon	37
P16	Feedback Control of Wasserstein Gradient Flows: Local Stabilisation, Convexification and Particle Implementation Lucas M. Moschen	38
P17	Self-Compression: A gradient-based approach towards training LLMs for binary inference. Ben Boys	39

POSTER ABSTRACT / P01

Tuesday 24 November 16:00–18:00 170QG Drawing Room

Efficient Function Approximation Under Heteroskedastic Noise

Yuji Nakatsukasa — Mathematical Institute, University of Oxford
Yifu Zhang — Department of Mathematics, Imperial College London

Approximating a function on $[-1, 1]$ from $N + 1$ noisy samples is a classical problem in numerical analysis. When the noise is heteroskedastic, with variance $\sigma(x)^2$ depending on the sampling location, existing methods do not provide an $O(N \log N)$ algorithm tailored to this setting. We introduce *HeteroChebtrunc*, an adaptation of NoisyChebtrunc designed for heteroskedastic observations. Using high-dimensional probability, we show that, with high probability, HeteroChebtrunc attains a sharper uniform-norm error bound than NoisyChebtrunc under heteroskedastic noise. Its computational cost is $O(N + \widehat{N} \log \widehat{N})$, where $\widehat{N} \ll N$ is a user-chosen parameter. In the course of the analysis, we also obtain a non-asymptotic high-probability relative-error bound for the sample variance estimator of subgaussian random variables. Numerical experiments illustrate the improved uniform approximation accuracy of the proposed method.

Presenter: Yifu Zhang

POSTER ABSTRACT / P02

Tuesday 24 November 16:00–18:00 170QG Drawing Room

On the implicit regularization of Langevin dynamics with projected noise

Govind Menon — Brown University, USA
Austin J. Stromme and Adrien Vacher — ENSAE, France

Keywords: Implicit Bias, Langevin Dynamics, Isometric Group Actions, Geometry

We study Langevin dynamics with noise projected onto the directions orthogonal to an isometric group action. This mathematical model is introduced to shed new light on the effects of symmetry on stochastic gradient descent for over-parametrized models. Our main result identifies a novel form of implicit regularization: when the initial and target density are both invariant under the group action, Langevin dynamics with projected noise is equivalent in law to Langevin dynamics with isotropic diffusion but with an additional drift term proportional to the negative log volume of the group orbit. We prove this result by constructing a coupling of the two processes via a third process on the group itself, and identify the additional drift as the mean curvature of the orbits.

References

[1] G. Menon, A.J. Stromme, A. Vacher, *On the implicit regularization of Langevin dynamics with projected noise*, Conference On Learning Theory, 2026.

Presenter: Adrien Vacher

POSTER ABSTRACT / P03

Tuesday 24 November 16:00–18:00 170QG Drawing Room

Convergence of adaptive self-normalized importance sampling

Nicola Branchini — University of Warwick

Estimating expectations under complex posterior distributions is a central task in machine learning and statistics. In certain challenging problems like Bayesian cross-validation and sensitivity analysis, even i.i.d sampling from the posterior distribution can produce estimators with high or infinite variance. Then using self-normalized importance sampling (SNIS) with an appropriate proposal, a distribution which differs from the posterior, is required. Yet, there is a lack of approaches which learn a proposal tailored to SNIS.

We revisit the SNIS-optimal proposal of Hesterberg (1988), long overlooked because it is doubly intractable: the very expectation being estimated enters its definition, and the resulting density cannot be evaluated. We propose a generic adaptive framework that approximates it by interleaving MCMC sampling from an approximation of the optimal proposal with running SNIS estimates, connecting stochastic approximation and adaptive importance sampling. We prove a martingale central limit theorem for the resulting estimator, show that learning the proposal incurs little asymptotic cost, and establish a non-asymptotic chi-squared contraction rate between the learned and optimal proposals. Experiments in Bayesian cross-validation confirm the theory and show substantial variance reduction over standard alternatives.

Presenter: Nicola Branchini

POSTER ABSTRACT / P04

Tuesday 24 November 16:00–18:00 170QG Drawing Room

Consistency Regularised Gradient Flows for Inverse Problems

Alessio Spagnoletti[1,*], Tim Y. J. Wang[2,*], Marcelo Pereyra[3,†], O. Deniz Akyildiz[2,†]
[1]Laboratoire MAP5, UMR 8145, Université Paris Cité, CNRS
[2]Department of Mathematics, Imperial College London
[3]Heriot-Watt University, MACS & Maxwell Institute
[*]Joint first authors. [†]Equal contribution.

Vision-Language Latent Diffusion Models (LDMs) (Rombach et al., 2022) provide powerful generative priors for inverse problems. However, existing LDM-based inverse solvers typically require a large number of neural function evaluations (NFEs) and backpropagation through large pretrained components, leading to substantial computational cost and, in some cases, degraded reconstruction quality. We propose a unified Euclidean-Wasserstein-2 gradient-flow framework that jointly performs posterior sampling and prompt optimization in the latent space through a single flow that aligns the prior and posterior with the observed data. Combined with few-step latent text-to-image models, this formulation enables low-NFE inference without backpropagation through autoencoders. Experiments across several canonical imaging inverse problems show that our method achieves state-of-the-art performance with significantly reduced computational cost.

Presenters: Tim Y. J. Wang / Alessio Spagnoletti

POSTER ABSTRACT / P05

Tuesday 24 November 16:00–18:00 170QG Drawing Room

A Unifying View of Variational Generative Wasserstein Flows

Paul Caucheteux[1], Clément Bonet[2], Anna Korba[1]
[1]ENSAE, CREST [2]CMAP, Ecole Polytechnique

Many modern generative models can be viewed as minimizing divergences between probability distributions, yet they rely on different algorithmic and geometric principles. Wasserstein gradient flows provide a continuous-time formulation for optimizing over distributions, and can be approximated through their implicit discretization via the Jordan–Kinderlehrer–Otto (JKO) scheme. In this work, we present a unified theoretical framework for generative modeling based on Wasserstein gradient flows, which we refer to as Generative Wasserstein Flows (GWF). We show that a broad class of existing methods can be derived as instances of parametric JKO schemes for f-divergence objectives, and we establish equivalences between several recently proposed algorithms. We extend this framework beyond f-divergences to Integral Probability Metrics and squared Maximum Mean Discrepancy, deriving new JKO-based generative algorithms and clarifying their connections with GANs. We study empirically the impact of JKO regularization for a wide range of objectives. Finally, we analyze parametric Wasserstein flows, where the dynamics are restricted to distributions induced by parametrized maps.

arXiv:2605.31369

Presenter: Paul Caucheteux

POSTER ABSTRACT / P06

Tuesday 24 November 16:00–18:00 170QG Drawing Room

The Fenchel Game of Underdamped Langevin: Insights into Accelerated Convergence

Alexandra Borkowski* and Nik Nüsken — King's College London
*Presenting author.

Following the work of Wang, Abernethy, and Levy (2023), a two-player online optimisation game can be used to describe accelerated convex optimisation algorithms (and many others). Based on the prevailing parallels between optimisation and sampling, the question arises as to whether an analogous perspective may yield accelerated sampling dynamics.

In this context, we construct a two-player game for sampling from a probability measure $\pi(dx) \propto e^{-V(x)}\,dx$, where V is convex or strongly convex, and show that Underdamped Langevin dynamics arise from a combination of distinct strategies for the two players. Moreover, this interpretation provides a framework for analysing the convergence of Underdamped Langevin dynamics in KL divergence. In particular, for appropriate choices of the damping parameters, we show convergence rates that are accelerated compared with those of Overdamped Langevin dynamics.

Preprint: arXiv:2609.07812. **Reference:** Wang, Abernethy, and Levy (2023), *No-Regret Dynamics in the Fenchel Game: A Unified Framework for Algorithmic Convex Optimization.*

Presenter: Alexandra Borkowski

POSTER ABSTRACT / P07

Tuesday 24 November 16:00–18:00 170QG Drawing Room

Annealed Underdamped Langevin Sampling

Laurenz Nagler, Alexander Falk, Andreas Habring

Approaches for accelerating SDE-based sampling broadly fall into two categories: using dynamics with faster ergodic convergence or better discretization properties, such as underdamped Langevin; or non-equilibrium sampling, where the target is annealed or diffused to improve its conditioning. We combine the two and propose Annealed Underdamped Langevin Sampling (ANULD), which runs underdamped Langevin dynamics with a time-varying target interpolating from a simple reference (e.g. a standard Gaussian) to the complex target of interest. We bound the sampling error in terms of the action of the annealing path, i.e., its kinetic energy in Wasserstein space, linking sampling complexity to the geometry of the path. Because the noise in the underdamped system acts only on the velocity, standard Girsanov arguments do not apply, and we rely on a new proof strategy based on shifted Girsanov arguments. In the convex case obtain a complexity of $O(\epsilon^{-2})$ to reach accuracy ϵ in TV, compared to $O(\epsilon^{-6})$ for annealed overdamped Langevin (Guo et al., 2025; Cordero-Encinar et al., 2025). We demonstrate the method's efficacy in experiments for posterior sampling for toy problems and Bayesian MRI.

Presenter: Andreas Habring

POSTER ABSTRACT / P08

Tuesday 24 November 16:00–18:00 170QG Drawing Room

Fractional Posteriors under Model Misspecification in Generalised Linear Models

Emily C. Ehrhardt, Randolf Altmeyer, Kolyan Ray — Imperial College London

Polynomial time sampling guarantees ensure that posterior inference remains computationally tractable in high-dimensional problems. Nickl & Wang (2022) and Altmeyer (2025) showed that for the unadjusted Langevin algorithm (ULA) the usual global log-concavity assumption of the posterior can be relaxed to a local curvature assumption combined with posterior contraction. We extend this framework to the situation when the likelihood is misspecified, so the assumed statistical model does not accurately represent the underlying data-generating process. A common approach to Bayesian inference under misspecification is the fractional posterior, which is obtained by raising the likelihood to a power $\alpha \in (0, 1]$. We focus on misspecified generalized linear models (GLMs) and prove that the fractional posterior contracts around the Kullback-Leibler projection θ^* at rate $\sqrt{p \log(\alpha n)/(\alpha n)}$ in Euclidean distance. Notably, our proof includes $\alpha = 1$ contrary to previously obtained contraction results. Given a suitable initialiser, we obtain polynomial time sampling guarantees for ULA applied to the fractional posterior of misspecified GLMs.

References

Altmeyer, R. (2025), 'Polynomial time guarantees for sampling based posterior inference in high-dimensional generalised linear models'. arXiv:2208.13296 [math.ST].
URL: https://doi.org/10.48550/arXiv.2208.13296

Nickl, R. & Wang, S. (2022), 'On polynomial-time computation of high-dimensional posterior measures by Langevin-type algorithms', *Journal of the European Mathematical Society* **26**(3), 1031–1112.
URL: https://doi.org/10.4171/jems/1304

Presenter: Emily Ehrhardt

POSTER ABSTRACT / P09

Tuesday 24 November 16:00–18:00 170QG Drawing Room

Sobolev Regularized MMD Gradient Flow

Chenyang Tian[1,3,†], Bharath K. Sriperumbudur[2], Arthur Gretton[3,4], Zonghao Chen[3,†]
[1]Tsinghua University [2]Pennsylvania State University
[3]University College London [4]Google Deepmind
[†]Equal contribution.

We analyze Sobolev descent [MSR19], a regularized variant of maximum mean discrepancy (MMD) gradient flow based on a gradient penalty on the witness function. Such regularization mitigates the non-convexity of the MMD objective and yields provable global convergence guarantees in MMD in both continuous and discrete time. A surprising appeal is that our convergence analysis does not rely on isoperimetric assumptions on the target distribution. Instead, it is based on a regularity condition on the difference between kernel mean embeddings. We extend Sobolev descent such that it is applicable in both sampling (from an unnormalized target distribution)—using Stein kernels—and generative modeling settings, unlike previous works, where a gradient flow is suitable for only generative modeling or sampling but not both. We empirically validate the method on a broad range of tasks across both settings.

References

[MSR19] Youssef Mroueh, Tom Sercu, and Anant Raj. Sobolev descent. In K. Chaudhuri and M. Sugiyama, editors, *The 22nd International Conference on Artificial Intelligence and Statistics*, pages 2976–2985. PMLR, PMLR, 2019.

Presenter: Chenyang Tian

POSTER ABSTRACT / P10

Tuesday 24 November 16:00–18:00 170QG Drawing Room

L^2 over Wasserstein: Statistical analysis for optimal transport

Rohan Shenoy and Pengcheng Ye

Optimal transport provides an inherently geometric and highly structured framework for studying spaces of probability measures, supplying a rich theoretical toolkit for contemporary statistics, machine learning, and generative modelling. In applications, however, the measures of interest are almost never known precisely, calling for a theory of optimal transport that accounts for statistical uncertainty. We construct such a framework, lifting the classical theory to the setting of random probability measures. We introduce the space of Wasserstein random variables, establishing that it inherits the formal Riemannian structure of the Wasserstein space by characterising distances and geodesic geometry. The structure induces random flows with Wasserstein gradient flow sample paths, making it the natural extension of the Wasserstein space which allows for random gradient flow dynamics. We ensemble statistical convergence results of the optimal transport machinery using the random empirical measure within the framework. Moreover, in the setting of Bayesian analysis, we refine Schwartz's consistency theorem to the Wasserstein topology and deduce posterior convergence of the same machinery. We demonstrate that the growing theory of random token sampling for transformer models using self-attention flow paths can be embedded into the framework. The results provide a unified treatment of random optimal transport and its consequences for principled inference and generative modelling under the statistical uncertainty of random sampling.

Presenters: Rohan Shenoy / Pengcheng Ye

POSTER ABSTRACT / P11

Tuesday 24 November 16:00–18:00 170QG Drawing Room

Robust MMD-Regularised Particle Filter as a Wasserstein Gradient Flow

Shang Ling Yap, Randolf Altmeyer, Ömer Deniz Akyildiz
Department of Mathematics, Imperial College London

Standard filters presuppose a correct forecast model and degrade confidently when it is wrong. Motivated by generalised variational inference, we recast the filtering update at each assimilation step as an optimisation problem over $\mathcal{P}_2(\mathbb{R}^{d_X})$ — a data-fit term plus a maximum mean discrepancy penalty to the forecast ensemble — and solve it by its Wasserstein gradient flow, a deterministic McKean–Vlasov dynamics whose Euler discretisation is a weight-free, resampling-free update requiring no densities. Under standard regularity assumptions, we establish well-posedness and monotone decrease of the objective along the flow, and show that the discrete scheme tracks the continuous flow to first order in the step size. Under partial observation the flow is structurally degenerate, as the unobserved coordinates receive no information from the data; we repair this by descending in a covariance-preconditioned metric frozen at the forecast, and show that all guarantees transfer. On Lorenz-63 the filter outperforms ensemble Kalman and particle filters under misspecification; on Lorenz-96 it remains competitive with a localised and inflated ensemble Kalman filter, with low variability across seeds in both cases.

Presenter: Shang Ling Yap

POSTER ABSTRACT / P12

Tuesday 24 November 16:00–18:00 170QG Drawing Room

Sharp Asymptotics for Regularised Optimal Transport

Carlos Alberto Cardoso Correia Perelló — Columbia University

We establish sharp asymptotic expansions for entropically regularised optimal transport and L^p-regularised optimal transport, providing explicit limiting constants under substantially weaker assumptions than previously known. The proofs are based on a new and flexible framework that unifies these settings, and is designed to apply more broadly to general regularised optimal transport problems. This is joint work with Alberto González Sanz and Marcel Nutz.

References

[CGN26] Carlos Cardoso-Perelló, Alberto González-Sanz, and Marcel Nutz. “Sharp Asymptotics for Regularized Optimal Transport”. In: (2026). arXiv: 2607.18191 [math.OC].

[EN24] Stephan Eckstein and Marcel Nutz. “Convergence Rates for Regularized Optimal Transport via Quantization”. In: *Mathematics of Operations Research* 49 (2024), pp. 1223–1240.

[GGM24] Álvaro Garriz-Molina, Alberto González-Sanz, and Gilles Mordant. “Infinitesimal Behavior of Quadratically Regularized Optimal Transport and Its Relation with the Porous Medium Equation”. In: (2024). arXiv: 2407.21528 [math.OC].

[GLN22] Alberto González-Sanz, Jean-Michel Loubes, and Jonathan Niles-Weed. “Weak Limits of Entropy Regularized Optimal Transport: Potentials, Plans and Divergences”. In: (2022). arXiv: 2207.07427 [math.OC].

[Pal24] Soumik Pal. “On the Difference Between Entropic Cost and the Optimal Transport Cost”. In: *The Annals of Applied Probability* 34.1B (2024), pp. 1003–1028.

Presenter: Carlos Alberto Cardoso Correia Perelló

POSTER ABSTRACT / P13

Tuesday 24 November 16:00–18:00 170QG Drawing Room

Optimal Stabilization for Global Optimization

Yuyang Huang — Mathematical Institute, University of Oxford

We reformulate non-convex global minimization as a discounted infinite-horizon optimal control problem, whose associated Hamilton–Jacobi–Bellman (HJB) equation yields a value function that captures the full landscape of the objective. The optimal feedback is driven by the gradient of this value function, so trajectories follow a gradient flow of the value function rather than of original objective function and are steered towards the global minimizers. We establish explicit convergence rates for the variational behavior of the value function, for the objective value, and for pathwise convergence of trajectories to the minimizer set. These results do not require convexity, differentiability or Lojasiewicz-type conditions on the objective.

Presenter: Yuyang Huang

POSTER ABSTRACT / P14

Tuesday 24 November 16:00–18:00 170QG Drawing Room

Coresets for preserving attention and learning dynamics

Paula Cordero Encinar, Carles Domingo-Enrich and Lester Mackey

Approximating attention can reduce the computational cost of transformers (Carrell et al., 2025; Gong et al., 2026; Han et al., 2024), but accurate forward predictions alone do not guarantee accurate training gradients (Liang et al., 2024). We develop kernel-based coreset methods that simultaneously approximate attention outputs and gradients. Our construction yields guarantees for both unmasked and causal attention and extends to multilayer transformers under suitable regularity assumptions. From a gradient-flow perspective, our gradient bounds quantify the perturbation of the vector field driving parameter learning. This connects attention compression to the broader problem of efficiently approximating interacting systems while preserving their dynamics, with maximum mean discrepancy providing a common framework for controlling approximation errors.

References

Carrell, A. M., Gong, A., Shetty, A., Dwivedi, R., and Mackey, L. (2025). Low-rank thinning. In *Forty-second International Conference on Machine Learning*.

Gong, A., Carrell, A. M., Dwivedi, R., and Mackey, L. (2026). Express language modeling. arXiv preprint arXiv:2606.10944.

Han, I., Jayaram, R., Karbasi, A., Mirrokni, V., Woodruff, D., and Zandieh, A. (2024). Hyperattention: Long-context attention in near-linear time. In *The Twelfth International Conference on Learning Representations*.

Liang, Y., Sha, Z., Shi, Z., Song, Z., and Zhou, Y. (2024). Multi-layer transformers gradient can be approximated in almost linear time. arXiv preprint arXiv:2408.13233.

Presenter: Paula Cordero Encinar

POSTER ABSTRACT / P15

Tuesday 24 November 16:00–18:00 170QG Drawing Room

An analysis of Bures-Wasserstein variational inference with Gaussian Mixtures

Marguerite Petit Talamon — ENSAE

Gaussian mixtures provide a flexible variational family, but their optimization remains challenging: even for a convex and smooth target potential, the Kullback–Leibler (KL) objective is neither geodesically convex nor globally smooth in the component-wise Bures–Wasserstein geometry. We establish non-asymptotic optimization guarantees for Bures–Wasserstein gradient schemes over finite Gaussian mixtures with uniform weights. Our main result is a descent lemma for the deterministic scheme under a smooth target potential and suitable covariance and step-size conditions. We prove this by controlling the curvature of the KL objective along each update, despite its lack of global smoothness. This yields an $O(1/k)$ bound on the average squared component-wise gradient norm. Under an additional log-Sobolev inequality, we derive finite-time KL bounds with an explicit residual term arising from the Bures–Wasserstein geometry. We also analyze the associated continuous-time dynamics. This is a joint work with Anna Korba.

Presenter: Marguerite Petit Talamon

POSTER ABSTRACT / P16

Tuesday 24 November 16:00–18:00 170QG Drawing Room

Feedback Control of Wasserstein Gradient Flows: Local Stabilisation, Convexification and Particle Implementation

Dante Kalise, Lucas M. Moschen*, and Grigorios A. Pavliotis
Department of Mathematics, Imperial College London, London (UK)
*Poster Presenter.

We consider particle systems on the torus driven by a confining potential, a pairwise interaction and additive Brownian noise. Their mean-field limit is a McKean–Vlasov equation, which can be seen as the Wasserstein gradient flow of a free energy. This PDE can have multiple equilibria, both stable and unstable, and near a stable equilibrium, the convergence rate is set by the spectrum of the Wasserstein Hessian of the free energy. Under suitable assumptions on the Hessian, we prove that only finitely many modes decay slowly. At the particle level, these slow modes are reflected in metastability and slow sampling. We design a feedback control at the mean-field level that stabilises the nonlinear dynamics, including around unstable equilibria, while also convexifying the free energy locally. Using the Hessian at the target, we build a finite-rank control that acts only on the eigendirections whose decay rates fall below a chosen threshold, with gains computed from a Riccati equation. The controlled mean-field dynamics is still a gradient flow, of a modified free energy, and it is locally displacement convex around the target. Since the feedback law depends on the measure only through finitely many moments, it can be applied directly to the particle system through its empirical measure. We show that the particles converge at close to the accelerated mean-field rate until they reach the finite-N fluctuation level, and then stay near the target for times exponentially large in N. When the target is a Gibbs measure, the feedback can be built from the unnormalised density alone. Numerical tests on interacting models and on multimodal and banana-shaped Gibbs targets show faster convergence than the uncontrolled dynamics.

Presenter: Lucas M. Moschen

POSTER ABSTRACT / P17

Tuesday 24 November 16:00–18:00 170QG Drawing Room

Self-Compression: A gradient-based approach towards training LLMs for binary inference.

Ben Boys — Imagination Technologies

A major cost of LLM inference is moving weights, activations, and KV-cache data between memory and the inference engine in the decode phase. Quantization reduces this cost by lowering numerical precision, but aggressive quantization typically degrades accuracy and requires costly precision-allocation schemes. We introduce Self-Compression: a simple and general framework that jointly learns model parameters and their precision by augmenting the training objective with a communication-cost regularizer. This regularizer encourages the removal of redundant parameters while reducing the representation cost of the remaining weights and KV-cache data. As a result, Self-Compression achieves state-of-the-art accuracy at 4-bit precision, and offers variable bitdepth and self-pruning as further advantages. We apply Self-Compression to LLM pretraining and demonstrate stable optimization with standard optimizers down to 1-bit weights and activations.

Presenter: Ben Boys