Third workshop on Monte Carlo methods in Warsaw

University of Warsaw, 13-15.12.2023

Book of abstracts

	Wednesday 13.12	Thursday 14.12	Friday 15.12
09:30-10:20		Matti Vihola	Giorgos Vasdekis
10:20-11:10		Łukasz Szpruch	Abdul Haji-Ali
11:10-11:40		Coffee break	Coffee break
11:40-12:30		Tomasz Szarek	Alex Beskos
12:30-14:10	Lunch break/Welcome	Lunch break	Lunch break
14:10-15:00	Joris Bierkens	Katharina Schuh	Federica Milinanni
15:00-15:50	Dawid Czapla	Pierre Jacob	Krzysztof Łatuszyński
15:50-16:20	Coffee break	Coffee break	Coffee break
16:20-17:10	Elena Akhmatskhaya	Christophe Andrieu	
17:30-19:30		Poster Session	
19:30-	Dinner		

Localization:
Faculty of Mathematics, Informatics and Mechanics
Banacha 2a, Warsaw
Room 2180

Adaptive Irreversible Hamiltonian/Hybrid Monte Carlo Methods with Applications

Elena Akhmatskaya
BCAM - Basque Center for Applied Mathematics, Spain

Hamiltonian (Hybrid) Monte Carlo (HMC) method, initially proposed for Quantum Field theory problems, has recently become a popular tool for solving complex and otherwise intractable problems of statistical inference. The HMC replaces the random walk, inherent to standard Markov Chain Monte Carlo (MCMC), by Hamiltonian trajectories. This helps to produce less correlated samples and achieve faster convergence to the target distribution. The HMC is, therefore, more efficient than MCMC when applied to indirect observation models, generalized linear and nonlinear mixed models, and neural networks. Applications of HMC range from modelling of biomolecules, materials and chemical processes to Bayesian parameter estimation. One serious disadvantage of the HMC is, however, its reversibility, which diminishes the mixing rate of the sampler. The other bottleneck is strong dependence of the HMC's performance on the choice of parameters associated with an integration method for underlying Hamiltonian equations. To date, this choice remains largely heuristic. In this talk we discuss the advantages, over conventional HMC, of two its irreversible variants – the Generalized Hamiltonian/Hybrid Monte Carlo (GHMC) and the Modified Hamiltonian Monte Carlo (MHMC). We also present an adaptive integration approach that provides an intelligent (problem domain-, model-, sampler- and step-size specific) choice of the most appropriate (in terms of the best conservation of energy for harmonic forces/Gaussian targets) multi-stage splitting integrator for Hamiltonian Monte Carlo methods. The approach automatically avoids the values of the simulation parameters, likely to cause undesired extreme scenarios, such as resonance artefacts, low accuracy or especially poor sampling. The sampling efficiency of GHMC and MHMC combined with adaptive splitting integrators is demonstrated in dynamic modelling of morphology development in multiphase latex particle.

Weak Poincaré Inequalities for Markov chains: theory and applications

Christophe Andrieu
University of Bristol, UK

We investigate the application of Weak Poincaré Inequalities (WPI) to Markov chains to study their rates of convergence and to derive complexity bounds. At a theoretical level we investigate the necessity of the existence of WPIs to ensure L^2-convergence, in particular by establishing equivalence with the Resolvent Uniform Positivity-Improving (RUPI) condition and providing a counterexample. From a more practical perspective, we extend the celebrated Cheeger's inequalities to the subgeometric setting, and further apply these techniques to study random-walk Metropolis algorithms for heavy-tailed target distributions and pseudo-marginal algorithms.
Joint work with Anthony Lee, Sam Power, Andi Q. Wang

The G-Wishart Weighted Proposal Algorithm: Efficient Posterior Computation for Gaussian Graphical Models

Alexandros Beskos
University College London, UK

Gaussian graphical models can capture complex dependency structures among variables. For such models, Bayesian inference is attractive as it provides principled ways to incorporate prior information and to quantify uncertainty through the posterior distribution. However, posterior computation under the conjugate G-Wishart prior distribution on the precision matrix is expensive for general non-decomposable graphs. We therefore propose a new Markov chain Monte Carlo (MCMC) method named the G-Wishart weighted proposal algorithm (WWA). WWA's distinctive features include delayed acceptance MCMC, Gibbs updates for the precision matrix and an informed proposal distribution on the graph space that enables embarrassingly parallel computations. Compared to existing approaches, WWA reduces the frequency of the relatively expensive sampling from the G-Wishart distribution. This results in faster MCMC convergence, improved MCMC mixing and reduced computing time. Numerical studies on simulated and real data show that WWA provides a more efficient tool for posterior inference than competing state-of-the-art MCMC algorithms.

Distributed Bayesian Computation via Metropolis-Hastings and Piecewise Deterministic Monte Carlo

Joris Bierkens
Delft University of Technology, Netherlands

A challenge in Bayesian computation is distributing the computation over multiple machines. Reasons for doing so may be to aim for improved computational efficiency, distributed memory usage, or there may be privacy issues which do not allow to communicate data between different machines.
In (Bierkens, Duncan, 2022) we introduced a federated approach employing Piecewise Deterministic Monte Carlo methods. Recent work (to appear a forthcoming publication) shows that also the classical Metropolis-Hastings algorithm may be designed to allow for distributed computing. We will discuss how the distributed approach affects the computational efficiency of these approaches. To make either of these approaches efficient, a control variates approach may be employed. The resulting algorithms efficiently combine local likelihood computations without the requirement to transfer data between machines, thus allowing for strong privacy guarantees without introducing bias in the computation.

On the Exponential Ergodicity in the Bounded-Lipschitz Distance of some Piecewise Deterministic Markov Processes on Polish Spaces

Dawid Czapla
University of Silesia, Poland

Let (Y, ρ_Y) be a Polish metric space endowed with the Borel σ-field $\mathcal{B}(Y)$. Consider a finite collection $\{S_i : i \in I\}$ of arbitrary semiflows acting from $\mathbb{R}_+ \times Y$ to Y, a matrix $\{\pi_{ij}\}_{i,j\in I}$ of continuous maps from Y to $[0,1]$ satisfying

$$\sum_{j\in I} \pi_{ij}(y) = 1 \quad \text{for any} \quad y \in Y,\ i \in I,$$

and a bounded continuous function $\lambda : Y \to (0,\infty)$ with positive infimum. Moreover, let $J : Y \times \mathcal{B}(Y) \to [0,1]$ be an arbitrary stochastic Markov kernel.

The object of our study will be a Markov process $\Psi := \{(Y(t), \xi(t))\}_{t\in\mathbb{R}_+}$ evolving on the space $X := Y \times I$ in such a way that

$$Y(t) = S_{\xi_{n-1}}(t - \tau_{n-1}, Y_{n-1}),\ \ \xi(t) = \xi_{n-1} \ \text{ for }\ t \in [\tau_{n-1}, \tau_n),\ n \in \mathbb{N}.$$

Here $\{\tau_n\}_{n\in\mathbb{N}_0}$ is an a.s. infinitely increasing sequence of positive random variables (representing the jump times of Ψ) with mutually independent increments $\Delta\tau_n := \tau_n - \tau_{n-1}$, $n \in \mathbb{N}$, such that, for all $t \in \mathbb{R}_+$, $(y,i) \in X$, and $n \in \mathbb{N}$,

$$\mathbb{P}(\Delta\tau_n \le t \,|\, \Phi_{n-1} = (y,i)) = 1 - \exp\left(-\Lambda(y,i,t)\right), \quad \text{where} \quad \Lambda(y,i,t) := \int_0^t \lambda(S_i(s,y))\, ds,$$

while $\Phi := \{(Y_n, \xi_n)\}_{n\in\mathbb{N}_0}$ is a given time-homogeneous Markov chain describing the post-jump locations of Ψ, i.e., $\Psi(\tau_n) = \Phi_n$ for $n \in \mathbb{N}_0$. The transition law of Φ is defined using S_i, π_{ij}, λ and J, so that, for any $y \in Y$, $B \in \mathcal{B}(Y)$, $i, j \in I$ and $n \in \mathbb{N}$, one has

$$\mathbb{P}(Y_n \in B \mid \Phi_{n-1};\ Y(\tau_n-) = y) = J(y,B),$$
$$\mathbb{P}(\xi_n = j \mid \Phi_{n-1};\ \xi_{n-1} = i,\ Y_n = y) = \pi_{ij}(y)$$
$$\mathbb{P}(\tau_n \le t \,|\, \Phi_{n-1} = (y,i),\ \tau_{n-1} = s) = 1_{[s,\infty)}(t)\big(1 - \exp\left(-\Lambda(y,i,t-s)\right)\big)$$

Within the lecture, we will present the results regarding:

(1) A one-to-one correspondence between the stationary distributions of the process Ψ and those of the chain Φ (given in an explicit form).

(2) A criterion on the exponential ergodicity of the process Ψ in the bounded Lipschitz distance, based on certain conditions involving the semiflows S_i and the kernel J.

The bounded Lipschitz distance d_{BL} is defined on the set $\mathcal{M}_1(X)$ of Borel probability measures on X by

$$d_{BL}(\mu,\nu) := \sup\left\{ \left| \int_X f\, d(\mu - \nu) \right| :\ f : X \to [0,1],\ \mathrm{Lip}\, f \le 1 \right\} \quad \text{for} \quad \mu,\nu \in \mathcal{M}_1(X),$$

where X is equipped with the metric ρ_X of the form

$$\rho_X(x_1,x_2) := \rho_Y(y_1,y_2) + c\,\mathbf{d}(i,j) \quad \text{for} \quad x_1 := (y_1,i_1),\ x_2 = (y_2,i_2) \in X,$$

with $\mathbf{d}$ standing for the discrete metric on I and some positive constant c.

By saying that Ψ is exponentially ergodic in d_{BL} we mean that it has a unique stationary distribution μ_*, and its Markov transition semigroup on $\mathcal{M}_1(X)$, say $\{P_t\}_{t\ge0}$, enjoys the following property: there exist constants $C < \infty$ and $\gamma \in (0,1)$ such that, for any $\mu \in \mathcal{M}_1(X)$ and $t \ge 0$,

$$\|\mu P_t - \mu_*\|_{BL} \le Ce^{-\gamma t}\left(\int_X V\, d\mu + 1\right)^{1/2},$$

with $V(x) := \rho_Y(y_*, y)$ for $x = (y,i) \in X$, where y_* is an arbitrary point of Y.

(3) An application of the result mentioned in (2) to a special case of the proces Ψ, where J is the transition law of the Markov chain arising from a random iterated function system with place-dependent probabilities. It is noteworthy that, in this setting, the model under consideration can serve as a mathematical framework for gene expression analysis.

The result concerning (1) can be found in [1]. The criterion indicated in (2) is an extension of the results of [3] to the case where the jump intensity λ and the probabilities π_{ij} depend on the state of the process. Finally, the application of this criterion mentioned in (3) is based on [4], where the exponential ergodicity has been established for the discrete-time chain Φ.

If time allows, we shall additionally present a version of the central limit theorem for additive functionals of the process Ψ (in the case of constant λ and π_{ij}), proved in [2] by utilizing the results of [3].

References

[1] D. Czapla, *On the Existence and Uniqueness of Stationary Distributions for some Class of Piecewise Deterministic Markov Processes with State-Dependent Jump Intensity*, preprint, ArXiv: 2303.11576.

[2] D. Czapla, K. Horbacz, H. Wojewódka-Ściażko, *The Central Limit Theorem for Markov Processes that are Exponentially Ergodic in the Bounded-Lipschitz Norm*, Qualitative Theory of Dynamical Systems 23(1), 2024, art. no. 7.

[3] D. Czapla, K. Horbacz, H. Wojewódka-Ściażko, *Exponential Ergodicity in the Bounded-Lipschitz Distance for some Piecewise-Deterministic Markov processes with Random Switching Between Flows*, Nonlinear Analysis 215, 2022, art. no. 112678.

[4] D. Czapla, J. Kubieniec, *Exponential Ergodicity of some Markov Dynamical Dystems with Application to a Poisson Driven Stochastic Differential Equation*, Dynamical Systems 34(1), 2019, pp. 130–156.

An Antithetic Multilevel Monte Carlo-Milstein Scheme for Stochastic Partial Differential Equations

Abdul-Lateef Haji-Ali
Heriot-Watt University, UK

I will present a novel multilevel Monte Carlo approach for estimating quantities of interest for stochastic partial differential equations (SPDEs). Drawing inspiration from (Giles and Szpruch, 2014), the approach extends the antithetic Milstein scheme for finite-dimensional stochastic differential equations to Hilbert space-valued SPDEs. The resulting method combines the truncated Milstein scheme with appropriate spatial discretizations and noise approximations on all scales to obtain a fully discrete scheme. The method has the advantages of both Euler and Milstein discretizations, as it is easy to implement and does not involve intractable Levy area terms. Moreover, it leads to the same variance decay in an MLMC algorithm as the standard Milstein method, resulting in a significantly lower computational complexity than a corresponding MLMC Euler scheme.

A Coupling-Based Estimator of the Asymptotic Variance in the Central Limit Theorem for Markov Chain Ergodic Averages

Pierre Jacob
ESSEC Business School, France

Coupled Markov chains that meet exactly after a random number of iterations can generate unbiased estimators of point-wise evaluations of the solutions to the Poisson equation. This presentation describes how to do this and connects it to recently proposed unbiased estimators of Markov chain equilibrium expectations. The proposed estimators can also be used to construct unbiased estimators of the asymptotic variance of Markov chain ergodic averages, providing alternatives to batch means and spectral variance estimators with distinct properties, both favorable and unfavorable.

Air Markov Chain Monte Carlo

Krzystof Latuszyński
University of Warwick, UK

We introduce a class of Adapted Increasingly Rarely Markov Chain Monte Carlo (AirMCMC) algorithms where the underlying Markov kernel is allowed to be changed based on the whole available chain output but only at specific time points separated by an increasing number of iterations. The main motivation is the ease of analysis of such algorithms as well as the reduced computational cost of adapting. Under the assumption of either simultaneous or (weaker) local simultaneous geometric drift condition, or simultaneous polynomial drift we prove the L^2-convergence, Weak and Strong Laws of Large Numbers (WLLN, SLLN), Central Limit Theorem (CLT), and discuss how our approach extends the existing results. We argue that many of the known Adaptive MCMC algorithms may be transformed into the corresponding Air versions, and provide an empirical evidence that performance of the Air version stays virtually the same. This is joint work with Cyril Chimisov and Gareth Roberts.

Large deviation principle for Metropolis-Hastings based Markov chains

Federica Milinanni
KTH Royal Institute of Technology, Sweden

Good performance measures for the convergence of Markov chain Monte Carlo (MCMC) methods are essential. For instance, they can be used to compare different algorithms, or to tune parameters within a given method. Common tools that are used for analysing convergence properties of MCMC algorithms are, e.g., mixing times, spectral gap and functional inequalities (e.g. Poincaré, log-Sobolev). A further, rather novel, approach consists in the use of large deviations theory to study the convergence of empirical measures of MCMC chains. At the heart of large deviations theory is the large deviation principle, which allows us to describe the rate of convergence of the empirical measures through a so-called rate function.

In this talk I will consider Markov chains generated via MCMC methods of Metropolis-Hastings type for sampling from a target distribution on a Polish space. I will state a large deviation principle for the corresponding empirical measure, show examples of algorithms from this class for which the theorem applies, and illustrate how the result can be used to tune algorithm parameters.

Nonlinear Hamiltonian Monte Carlo and its Particle Approximation

Katharina Schuh
TU Wien, Austria

In the talk, a nonlinear (in the sense of McKean) generalization of Hamiltonian Monte Carlo (HMC) termed nonlinear HMC (nHMC) is presented to sample from nonlinear probability measures of mean-field type. Provided the underlying confinement potential is L-gradient Lipschitz and asymptotically K-strongly convex, we prove contraction in L^1-Wasserstein distance. Further, we show that uniformly in time nHMC can be approximated by unadjusted HMC for its corresponding mean-field particle system in L^1-Wasserstein distance. In particular, we analyse the number of gradient evaluations needed for unadjusted HMC with randomized time integration to achieve an ϵ-accuracy of a d-dimensional nonlinear probability measure in L^1-Wasserstein distance. The talk is based on joint work with Nawaf Bou-Rabee (arXiv:2308.11491).

Unique ergodicity and limit theorems for random maps on the circle and interval.

Tomasz Szarek
Gdańsk Technical University and PAS Sopot, Poland

The talk will be devoted to ergodic properties of Markov operators corresponding to random iterated function systems acting on simple one-dimensional manifolds: the circle and interval. We formulate criteria for unique ergodicity. We show also that such systems satisfy the Central Limit Theorem and the Law of the Iterated Logarithm.

Exploration-exploitation trade-off for continuous-time episodic reinforcement learning

Łukasz Szpruch
University of Edinburgh, UK

We develop a probabilistic framework for analysing model-based reinforcement learning in the episodic setting. We then apply it to study finite-time horizon stochastic control problems with linear dynamics but unknown coefficients and convex, but possibly irregular, objective function. Using probabilistic representations, we study the regularity of the associated cost functions and establish precise estimates for the performance gap between applying optimal feedback control derived from estimated and true model parameters. We identify conditions under which this performance gap is quadratic. Next, we propose a phase-based learning algorithm for which we show how to optimise exploration-exploitation trade-off and achieve sublinear regrets in high probability and expectation. Next, we study exploration-exploitation learning using noisy policies. Again we achieve sub-linear regrets for a class of entropy regularised stochastic control problems.

Skew-symmetric sampling schemes and locally balancing algorithms

Giorgios Vasdekis
University College London, UK

Locally balancing algorithms are a new class of MCMC algorithms, recently introduced in (Livingstone and Zanella, 2022). One of these algorithms, the Barker algorithm, has been shown to be robust to heteroskedasticity of the posterior target and the step size of the algorithm. At the same time, the algorithm seems to preserve high dimensional properties of state of the art MCMC, making it an interesting alternative to the existing literature. In the first part of the talk we will review the main results on these locally balancing algorithms. It turns out that in order to sample from the Barker algorithm, one can use ideas of sampling from skew-symmetric distributions. We will transfer these ideas in the context of discretising and simulating from diffusion processes and we will suggest a new class of unadjusted MCMC algorithms, which are robust with respect to the step size.
This is joint work with S. Livingstone, N. Nusken and R. Zhang.

On the forgetting of particle filters

Matti Vihola
University of Jyväskylä, Finland

We study the forgetting properties of the particle filter when its state — the collection of particles — is regarded as a Markov chain. Under a strong mixing assumption on the particle filter's underlying Feynman–Kac model, we find that the particle filter is exponentially mixing, and forgets its initial state in $O(logN)$ '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. We present an example which suggests that this rate is optimal. We also study the conditional particle filter (CPF) and extend our forgetting result to this context. We establish a similar conclusion, namely, CPF is exponentially mixing and forgets its initial state in $O(logN)$ time. To support this analysis, we establish new time-uniform L^p error estimates for CPF, which can be of independent interest.
Based on joint work with Joona Karjalainen, Sumeetpal S. Singh and Anthony Lee: https://arxiv.org/abs/2309.08517

Poster Session

Adaptive Stereographic Particle Sampler

Cameron Bell
University of Warwick, UK

Stereographic MCMC methods were first introduced in Yang, Latuszynski, and Roberts (2022) and were shown to perform very well when targeting high dimensional or heavy tailed densities. The Stereographic Projection Sampler (SPS) and Stereographic Bouncy Particle Sampler (SBPS) both rely on projecting Euclidean space onto the unit hypersphere.
However, these methods are highly dependent on the choice of parameters used in the transformation. These are equivalent to preconditioning Euclidean space to centre and scale the target distribution. In this poster, we present an Adaptive MCMC method which allows us to update the parameters of the SBPS whilst preserving asymptotic convergence properties.

Discretization of the Gradient Flow

Matej Benko
Brno University of Technology, Czechia

We consider the gradient flow along the functional consisting of the sum of potential, interaction, and Boltzmann entropy functional. It is equivalent to McKean-Vlasov stochastic differential equation and McKean-Vlasov-Fokker-Planck partial differential equation under sufficient assumptions.
The discretization of the gradient flow has an advantage over the discretization of space at large dimensions in computational complexity, which includes applications in machine learning and finance. The interaction term describing the forces which move the particle in the direction of a higher concentration of particles makes it impossible to approximate gradient flow as independent particles.
Therefore, we consider the propagation of chaos technique, which provides a linearisation of the McKean-Vlasov SDE to the Itô's SDE system, which converges to the McKean-Vlasov system with the number of independent particles. We show that this works for the interaction energy potential function with a bounded value below. On the approximating system of SDEs, we incorporate proximal step to compute the drift (consist from potential and linearised interaction term) and consider gradient flow along entropy as adding up a Gaussian random variable.

Parameter Estimation with Increased Precision for Elliptic and Hypo-elliptic Diffusions

Yuga Iguchi
University College London, UK

This work aims at making a comprehensive contribution in the general area of parametric inference for discretely observed diffusion processes. Established approaches for likelihood-based estimation invoke a time-discretisation scheme for the approximation of the intractable transition dynamics of the Stochastic Differential Equation (SDE) model over finite time periods. The scheme is applied for a step-size that is either user-selected or determined by the data. Recent research has highlighted the critical effect of the choice of numerical scheme on the behaviour of derived parameter estimates in the setting of hypo-elliptic SDEs. In brief, in our work, first, we develop two weak second-order sampling schemes (to cover both hypo-elliptic and elliptic SDEs) and produce a small time expansion for the density of the schemes to form a proxy for the true intractable SDE transition density. Then, we establish a collection of analytic results for likelihood-based parameter estimates obtained via the formed proxies, thus providing a theoretical framework that showcases advantages from the use of the developed methodology for SDE calibration. We present numerical results from carrying out classical or Bayesian inference, for both elliptic and hypo-elliptic SDEs.

Quantifying the effectiveness of linear preconditioning in Markov chain Monte Carlo

Max Hird
University College London, UK

Linear transformation of the state variable (linear preconditioning) is a common technique that often drastically improves the practical performance of a Markov chain Monte Carlo algorithm. Despite this, however, the benefits of linear preconditioning are not well-studied theoretically, and rigorous guidelines for choosing preconditioners are not always readily available. Mixing time bounds for various samplers have been produced in recent works for the class of strongly log-concave and Lipschitz target distributions and depend strongly on a quantity known as the condition number. We study linear preconditioning for this class of distributions, and under appropriate assumptions we provide bounds on the condition number after using a given linear preconditioner. We provide bounds on the spectral gap of RWM that are tight in their dependence on the condition number under the same assumptions. Finally we offer a review and analysis of popular preconditioners. Of particular note, we identify a surprising case in which preconditioning with the diagonal of the target covariance can actually make the condition number *increase* relative to doing no preconditioning at all

On the convergence of dynamic implementations of Hamiltonian Monte Carlo and No U-Turn Samplers

Miika Kailas
University of Jyväskylä, Finland

There is substantial empirical evidence about the success of dynamic implementations of Hamiltonian Monte Carlo (HMC), such as the No U-Turn Sampler (NUTS), in many challenging inference problems but theoretical results about their behavior are scarce. The aim of this paper is to fill this gap. We consider a general class of MCMC algorithms we call dynamic HMC. First, we show that this general framework encompasses NUTS as a particular case, implying the invariance of the target distribution as a by-product. Second and most importantly, we establish conditions under which NUTS is irreducible and aperiodic and as a corollary ergodic. Under conditions similar to the ones existing for HMC, we also show that NUTS is geometrically ergodic. Finally, we improve existing convergence results for HMC showing that this method is ergodic without any boundedness condition on the stepsize and the number of leapfrog steps in the case where the target is a perturbation of a Gaussian distribution.

Adaptive multi-stage integration schemes for Hybrid/Hamiltonian Monte Carlo

Lorenzo Nagar
BCAM - Basque Center for Applied Mathematics, Spain

Hybrid and Hamiltonian Monte Carlo, commonly denoted as HMC, are recognized as powerful sampling tools for molecular simulation and Bayesian inference applications. Their effectiveness stems from the combination of a deterministic proposal generated through Hamiltonian dynamics with a stochastic Monte Carlo, which allows to sample complex and high-dimensional target distributions more accurately than conventional MCMC.

The numerical integration of underlying Hamiltonian equations plays a fundamental role in the overall performance of the method. Recently, multi-stage

splitting integrators demonstrated to be a promising alternative to the Verlet method, traditionally used in HMC. Nevertheless, their stability intervals are shorter than Verlet's, and their accuracy depends on the choice of an integration step size. The poster presents an Adaptive Integration Approach for applications in computational statistics, which addresses those issues. More precisely, the method detects optimal, problem-specific integration schemes (in terms of the best conservation of energy for harmonic forces/Gaussian targets) within the families of 2- and 3-stage splitting integrators.

Numerical experiments on well-known statistical models show that the adaptive schemes reach the best possible performance - in terms of sampling, acceptance rate and convergence - within the family of 2-, 3-stage splitting schemes.

Skew–Symmetric Approximations of Posterior Distributions

Francesco Pozza
Università Bocconi, Italy

Modern Bayesian statistics relies heavily on deterministic approximations of intractable posterior distributions to facilitate inference in difficult contexts. From a practical perspective, a common drawback of many of these approximations is that they are typically Gaussian or, more generally, symmetric about a given location parameter. As a consequence, such solutions may miss important properties of the posterior, such as, trivially, asymmetry. To alleviate this problem, we propose a simple skewness-inducing correction that can improve any off-the-shelf symmetric posterior approximation. In addition to provide asymmetric approximations that are provably never less accurate than their symmetric counterparts with respect to any alpha divergence, the proposed method is also computationally attractive. Indeed, Monte Carlo estimates of functionals of interest for statistical inference can be obtained at a cost usually comparable to that of simulating from the original symmetric density. Real-world data examples confirm that the proposed method can yield remarkable accuracy improvements over state-of-the-art symmetric deterministic approximations.

Metropolis Adjusted Langevin Trajectories: a robust alternative to Hamiltonian Monte Carlo

Lionel Riou-Durand
INSA Rouen, France

Hamiltonian Monte Carlo (HMC) is a widely used sampler, known for its efficiency on high dimensional distributions. Yet HMC remains quite sensitive to the choice of integration time. Randomizing the length of Hamiltonian trajectories (RHMC) has been suggested to smooth the Auto-Correlation Functions (ACF), ensuring robustness of tuning. We present the Langevin diffusion as an alternative to control these ACFs by inducing randomness in Hamiltonian trajectories through a continuous refreshment of the velocities. We connect and compare the two processes in terms of quantitative mixing rates for the 2-Wasserstein and L2 distances. The Langevin diffusion is presented as a limit of RHMC achieving the fastest mixing rate for strongly log-concave targets. We introduce a robust alternative to HMC built upon these dynamics, named Metropolis Adjusted Langevin Trajectories (MALT). Studying the scaling limit of MALT, we obtain optimal tuning guidelines similar to HMC, and recover the same scaling with respect to the dimension without additional assumptions. We illustrate numerically the efficiency of MALT compared to HMC and RHMC.

A Data Augmentation Framework for Sequential Monte Carlo Methods for Inference in Diffusions Observed with Noise

Christopher Stanton
University College London, UK

Applying Sequential Monte Carlo methods to inference in diffusions is challenging, due to the intractability of transition densities. The filtering problem can be overcome through data augmentation, under which we show that guided proposals, either based on likelihoods for conditioned or unconditioned diffusions, can readily be applied. However, this framework does not naturally extend to particle smoothing, due to the dependence between path-valued particles resulting in degenerate updates. We introduce a reparameterisation of the augmented space, that simultaneously breaks this dependency, and when the inference problem is extended to parameter estimation, enables non-degenerate updates of the diffusion parameters in particle Gibbs. We aim to emphasise the practical benefits of the methodology on joint smoothing in diffusions using Particle Gibbs with backward sampling, for which performance for a fixed number of particles is independent of the time horizon.