[Filippo Ascolani (Friday 14:45-15:30)]

[Joris Bierkens (Friday 11:00-11:45)]

[Gabriel Ducrocq (Thursday 14:00-14:45)]

[Alain Durmus (Thursday 9:45-10:30)]

[Gersende Fort (Thursday 11:45-12:30)]

[Samuel Livingston (Friday 14:00-14:45)]

[Krzysztof Łatuszyński (Saturday 11:30-12:15)]

[Błażej Miasojedow (Wednesday 14:00-14:45)]

[Omiros Papaspiliopoulos (Friday 11:45 -12:30)]

[Ardjen Pengel (Saturday 10:45-11:30)]

[Umut Simsekli (Wednesday 14:45-15:30)]

[Timothee Stumpf-Fetizon (Thursday 14:45-15:30)]

[Jun Yang (Thursday 11:00-11:45)]

[Giacomo Zanella (Friday 9:45-10:30)]

Filippo Ascolani (Friday 14:45-15:30)

Complexity of Gibbs sampler through Bayesian asymptotics

Gibbs samplers are popular algorithms to approximate posterior distributions arising from Bayesian hierarchical models. Despite their popularity and good empirical performances, however, there are still relatively few quantitative theoretical results on their scalability or lack thereof, e.g. much less than for gradient-based sampling methods.

We introduce a novel technique to analyse the asymptotic behaviour of mixing times of Gibbs Samplers, based on tools of Bayesian asymptotics. Our methodology applies to high-dimensional regimes where both number of datapoints and parameters increase, under random data-generating assumptions. This allows us to provide a fairly general framework to study the complexity of Gibbs samplers fitting Bayesian hierarchical models.

The methodology is illustrated on two-level hierarchical models with generic likelihood. For this class, we are able to provide dimension-free convergence results for Gibbs Samplers under mild conditions.

Joris Bierkens (Friday 11:00-11:45)

Some developments in PDMC methods

Piecewise Determinstic Monte Carlo (PDMC) is a relatively new and versatile method for simulating complex probability distributions. Key examples are the Bouncy Particle Sampler, Randomized Hamiltonian Monte Carlo and the Zig-Zag Sampler.

In this talk I will give an overview of some developments of the last years in this area. In particular I will discuss adaptive and approximate PDMC methods, application of the Zig-Zag sampler to variable selection, and, if time permits, spectral theory for the one-dimensional Zig-Zag process.

This concerns joint work with Andrea Bertazzi, Paul Dobson, Sebastiano Grazzi, Frank van der Meulen, Moritz Schauer, and Sjoerd Verduyn Lunel.

Gabriel Ducrocq (Thursday 14:00-14:45)

Fast compression of MCMC output

We propose cube thinning, a novel method for compressing the output of an MCMC (Markov chain Monte Carlo) algorithm when control variates are available. It allows resampling of the initial MCMC sample (according to weights derived from control variates), while imposing equality constraints on the averages of these control variates, using the cube method (an approach that originates from survey sampling). The main advantage of cube thinning is that its complexity does not depend on the size of the compressed sample. This compares favourably to previous methods, such as Stein thinning, the complexity of which is quadratic in that quantity.

Alain Durmus (Thursday 9:45-10:30)

Non-Equilibrium Sampling

joint work with Achille Thin, Yazid Janati, Sylvain Le Corff, Charles Ollion, Arnaud Doucet, Eric Moulines, Christian Robert

Sampling from a complex distribution π and approximating its intractable normalizing constant Z are challenging problems. In this talk, a novel family of importance samplers (IS) and Markov chain Monte Carlo (MCMC) samplers is derived. Given an invertible map T, these schemes combine (with weights) elements from the forward and backward orbits through points sampled from a proposal distribution ρ. The map T does not leave the target π invariant, hence the name NEO, standing for Non-Equilibrium Orbits. NEO-IS provides unbiased estimators of the normalizing constant and self-normalized IS estimators of expectations under π while NEO-MCMC combines multiple NEO-IS estimates of the normalizing constant and an iterated sampling-importance resampling mechanism to sample from π.

Joint work with Achille Thin, Yazid Janati, Sylvain Le Corff, Charles Ollion, Arnaud Doucet, Eric Moulines, Christian Robert

Gersende Fort (Thursday 11:45-12:30)

Federated Expectation Maximization with heterogeneity mitigation and variance reduction,

The Expectation Maximization (EM) algorithm is the default algorithm for inference in latent variable models. As in any other field of machine learning, applications of latent variable models to very large datasets make the use of advanced parallel and distributed architectures mandatory. In this talk, we will introduce FedEM, which is the first extension of the EM algorithm to the federated learning context. FedEM is a new communication efficient method, which handles partial participation of local devices, and is robust to heterogeneous distributions of the datasets. To alleviate the communication bottleneck, FedEM compresses appropriately defined complete data sufficient statistics. We also develop and analyze an extension of FedEM to further incorporate a variance reduction scheme. Numerical results will be presented to support our theoretical findings. Finally, we will comment the finite-time complexity bounds we obtained for these federated EM algorithms. Based on the paper: https://hal.archives-ouvertes.fr/hal-03333516v3

Authors : Aymeric DIEULEVEUT (CMAP, Ecole Polytechnique), Gersende FORT (IMT, CNRS), Eric MOULINES (CMAP, Ecole Polytechnique), Geneviève ROBIN (LAMME, CNRS).

Samuel Livingston (Friday 14:00-14:45)

Adaptive random neighbourhood samplers for Bayesian variable selection

We introduce a framework for efficient Markov Chain Monte Carlo (MCMC) algorithms targeting discrete-valued high-dimensional distributions, such as posterior distributions in Bayesian variable selection (BVS) problems. We show that many recently introduced algorithms, such as the locally informed sampler and the Adaptively Scaled Individual adaptation sampler (ASI), can be viewed as particular cases within the framework. We then describe a novel algorithm, the Adaptive Random Neighbourhood Informed sampler (ARNI), by combining ideas from both of these existing approaches. We show using several examples of both real and simulated datasets that a computationally efficient point-wise implementation (PARNI) leads to relatively more reliable inferences on a range of variable selection problems, particularly in the very large p setting. This is joint work with Xitong Liang and Jim Griffin.

Krzysztof Łatuszyński (Saturday 11:30-12:15)

A framework for adaptive MCMC for multimodal targets

I will present a class of adaptive auxiliary variable MCMC algorithms as a framework for designing algorithms for multimodal targets. The framework facilitates local within mode adaptation and adaptive between mode jumps. Limiting theorems will be presented for this class of algorithms. This is joint work with Emilia Pompe and Chris Holmes.

Błażej Miasojedow (Wednesday 14:00-14:45)

Particle MCMC with Poisson resampling.

We introduce a new version of particle filter in which the number of “children” of a particle at a given time has a Poisson distribution. As a result, the number of particles is random and varies with time. An advantage of this scheme is that descendants of different particles can evolve independently. It makes easy to parallelize computations. Moreover, particle filter with Poisson resampling is readily adapted to the case when a hidden process is a continuous time, piecewise deterministic semi-Markov process. We show that the basic techniques of particle MCMC, namely particle independent Metropolis-Hastings, particle Gibbs sampler and its version with ancestor sampling, work under our Poisson resampling scheme. Our version of particle Gibbs sampler is uniformly ergodic under the same assumptions as its standard counterpart. We present simulation results which indicate that our algorithms can compete with the existing methods.

Omiros Papaspiliopoulos (Friday 11:45 -12:30)

Functional umbrella sampling

I will overview past and current work on estimating marginal likelihood surfaces. It is joint work with Tim Stumpf-Fetizon (Warwick) and Jonathan Weare (Courant Institute)

Ardjen Pengel (Saturday 10:45-11:30)

Strong Invariance Principles for Ergodic Markov Processes

Strong invariance principles describe the error term of a Brownian approximation of the partial sums of a stochastic process. While these strong approximation results have many applications, the results for continuous-time settings have been limited. In this paper, we obtain strong invariance principles for a broad class of ergodic Markov processes. The main results rely on ergodicity requirements and an application of Nummelin splitting for continuous-time processes. Strong invariance principles provide a unified framework for analysing commonly used estimators of the asymptotic variance in settings with a dependence structure. We demonstrate how this can be used to analyse the batch means method for simulation output of Piecewise Deterministic Monte Carlo samplers. We also derive a fluctuation result for additive functionals of ergodic diffusions using our strong approximation results.

Umut Simsekli (Wednesday 14:45-15:30)

Fractional Underdamped Langevin Dynamics: Retargeting SGD with Momentum under Heavy-Tailed Gradient Noise

Stochastic gradient descent with momentum (SGDm) is one of the most popular optimization algorithms in deep learning. While there is a rich theory of SGDm for convex problems, the theory is considerably less developed in the context of deep learning where the problem is non-convex and the gradient noise might exhibit a heavy-tailed behavior, as empirically observed in recent studies. In this talk, we will consider a continuous-time variant of SGDm, known as the underdamped Langevin dynamics (ULD), and investigate its asymptotic properties under heavy-tailed perturbations. Supported by recent studies from statistical physics, I will argue both theoretically and empirically that the heavy-tails of such perturbations can result in a bias even when the step-size is small, in the sense that the optima of stationary distribution of the dynamics might not match the optima of the cost function to be optimized. As a remedy, we will develop a framework, which we call fractional ULD (FULD), and prove that FULD targets the so-called Gibbs distribution, whose optima exactly match the optima of the original cost. I will illustrate that the Euler discretization of FULD has noteworthy algorithmic similarities with natural gradient methods and gradient clipping, bringing a new perspective on understanding their role in deep learning. I will illustrate the developed theory with experiments conducted on a synthetic model and neural networks. The talk will be based on the following paper:

Simsekli, U., Zhu, L., Teh, Y. W., & Gurbuzbalaban, M. (2020). Fractional underdamped langevin dynamics: Retargeting sgd with momentum under heavy-tailed gradient noise. In International Conference on Machine Learning (ICML). https://arxiv.org/pdf/2002.05685.pdf

Timothee Stumpf-Fetizon (Thursday 14:45-15:30)

Exact Bayesian Inference for Markov Switching Diffusions

We address the problem of Bayesian inference for discretely observed regime switching diffusions. Switching diffusion models extend ordinary diffusions by allowing for discrete shifts in instantaneous drift and volatility. These shifts are driven by an underlying, usually latent Markov jump process. As for ordinary diffusion models, the transition density implied by a Markov switching SDE is intractable, which complicates likelihood-based inference. We design an MCMC algorithm that targets the joint posterior of diffusion parameters and the latent regime process. The algorithm is /exact/ in the sense that the MCMC central limit theorem applies, which isn't the case for methods that discretely approximate the diffusion. In the process, we improve on existing exact methods for ordinary diffusions

Jun Yang (Thursday 11:00-11:45)

Stereographic Markov Chain Monte Carlo

High dimensional distributions, especially those with heavy tails, are notoriously difficult for off the shelf MCMC samplers: the combination of unbounded state spaces, diminishing gradient information and local moves, results in empirically observed "stickiness" and theoretical mixing properties - lack of geometric ergodicity. In this paper we introduce a new class of MCMC samplers that map the original high dimensional problem in Euclidean space onto a sphere and remedy these notorious mixing problems. In particular, we develop Random Walk Metropolis type algorithms as well as versions of Bouncy Particle Sampler that are uniformly ergodic for a large class of light and heavy tailed distributions and also empirically exhibit rapid convergence.

Giacomo Zanella (Friday 9:45-10:30)

Robust leave-one-out cross-validation for high-dimensional Bayesian models.

Leave-one-out cross-validation (LOO-CV) is a popular method for estimating out-of-sample predictive accuracy. However, computing LOO-CV criteria can be computationally expensive due to the need to fit the model multiple times. In the Bayesian context, importance sampling provides a possible solution but classical approaches can easily produce estimators whose variance is infinite, making them potentially unreliable. Here we propose and analyze a novel mixture estimator to compute Bayesian LOO-CV criteria. Our method retains the simplicity and computational convenience of classical approaches, while guaranteeing finite variance of the resulting estimators. Both theoretical and numerical results are provided to illustrate the improved robustness and efficiency. The computational benefits are particularly significant in high-dimensional problems, allowing to perform Bayesian LOO-CV for a broader range of models.