Hausdorff School
"MCMC: Recent developments and new connections"

September 14 to 25, 2020

organized by
Nawaf Bou-Rabee, Andreas Eberle

Abstracts

Francis Bach (INRIA and ENS Paris)

Optimization for machine learning

Abstract: In these lectures, I will present recent results relating optimization and machine learning, with a particular focus on exponentially convergent stochastic algorithms for convex problems and global convergence of gradient descent for specific non-convex problems.

Arnaud Doucet (Oxford)

Differentiable Particle Filtering

Abstract: Particle Filters (PF) are a powerful class of methods for performing state inference in state-space models and for computing likelihood estimates for fixed parameters. Resampling is a key ingredient of PF, necessary to obtain low variance estimates. However, resampling operations result in the simulated likelihood function being non-differentiable with respect to parameters, even if the true likelihood is itself differentiable. These resampling operations also yield high variance gradient estimates of the Evidence Lower Bound (ELBO) when performing variational inference. By leveraging Optimal Transport (OT) ideas, we introduce differentiable PF, providing a differentiable simulated likelihood function. This allows one to perform parameter estimation via maximization of the simulated likelihood using gradient techniques and to compute low variance gradient estimates for variational inference. We provide consistency results and demonstrate the performance of differentiable PF on various examples.

Arnaud Doucet (Oxford)

Controlled Sequential Monte Carlo

Abstract: Sequential Monte Carlo methods, also known as particle methods, are a popular set of techniques for approximating high-dimensional probability distributions and their normalizing constants. These methods have found numerous applications in statistics and related fields; e.g. for inference in non-linear non-Gaussian state space models, and in complex static models. Like many

Monte Carlo sampling schemes, they rely on proposal distributions which crucially impact their performance. We introduce here a class of controlled sequential Monte Carlo algorithms, where the proposal distributions are determined by approximating the solution to an associated optimal control problem using an iterative scheme. This method builds upon a number of existing algorithms in econometrics, physics, and statistics for inference in state space models, and generalizes these methods so as to accommodate complex static models. We demonstrate significant gains over state-of-the-art methods at a fixed computational complexity on a variety of applications.

Arnaud Doucet (Oxford)

Unbiased Markov chain Monte Carlo

Abstract: MCMC methods provide consistent estimators of integrals as the number of iterations goes to infinity but typically exhibit a bias after any fixed number of iterations. We will review two approaches that have been recently proposed by Glynn and Rhee (2014) and Jacob, O'Leary and Atchade (2019) to remove the bias using random truncations and coupling ideas. We will illustrate how these methods can be applied to complex algorithms. We will present conditions under which the resulting unbiased estimators can be computed in expected finite time and exhibit finite variance. Finally, we will illustrate the benefits and limitations of such ideas.

Tony Lelièvre (CERMICS and École des Ponts Paris))

Sampling problems in computational statistical physics

Abstract: Computational statistical physics is typically a domain where efficient sampling methods are crucial. The objective is indeed to obtain macroscopic properties of materials starting from a microscopic description at the molecular level, using ensemble averages (thermodynamic properties) or averages over paths (dynamical properties). Applications are numerous in very different scientific fields such as molecular biology, chemistry or materials science. The objective of these lectures will be, starting from some prototypical sampling problems raised in statistical physics, to introduce general purpose algorithms which are useful to sample multimodal distributions, distributions supported on manifolds and metastable trajectories. More precisely, the first lecture will be devoted to free energy adaptive biasing techniques, and their analysis using entropy techniques or techniques useful to prove the convergence of stochastic algorithms. In the second lecture, we will present Hamiltonian Monte Carlo methods to sample measures on submanifolds of R^n. Finally, the third lecture will be devoted to a discussion of the link between metastable dynamics and jump Markov processes, using the notion of quasi-stationary distribution.

Jesus María Sanz Serna (Madrid)

Numerical integrators for the Hamiltonian Monte Carlo method

Abstract: When using the Hamiltonian Monte Carlo (HMC) method and its variants, most computational effort goes into the numerical simulation of the Hamiltonian dynamics. In practice, that simulation is (almost) always performed by means of the Leapfrog/Verlet/Störmer integrator. In my talks I will first explain why the demands that HMC imposes on the integrator have a number of peculiar features not present in other uses of numerical methods for ordinary differential equations. I will then explain the reasons why Leapfrog is so successful in some settings. Finally I will argue that, nevertheless, Leapfrog is not in general the optimal integrator and discuss alternative algorithms that afford an enhanced performance.

Aaron Smith (Ottawa)

Methods for Bounding MCMC Error: Recent Advances and Comparisons

Abstract: One of the main questions in the theory of Markov chains is to bound or estimate the rate at which a chain, or its ergodic averages, converges to a stationary measure. In the context of MCMC or computer science, this rate determines the efficiency of an algorithm; in statistical physics this might determine the qualitative behaviour of a system being modeled, while in other areas of probability a rapidly-mixing chain might be used to prove efficient concentration inequalities. Although the different communities studying Markov chain convergence are aware of each other, methods have not always traveled quickly between them. One major difficulty has been the different systems being studied - in particular, many probabilists and statistical physicists work on finite state spaces, while statisticians tend to be interested in continuous state spaces.

The primary goal of this lecture series is to discuss some recent advances in estimating convergence of Markov chains that may be relevant to statisticians interested in MCMC. We will start with an overview of important parts of the classical theory of Markov chain convergence. This will include the mainstays of MCMC analysis (drift-and-minorization and its frequent companion, contraction) and some methods that are much more popular in other communities (conductance and canonical path approaches).

From there, we will explore the relevance of various spectral methods to MCMC analysis. This will include some recent work on basic path bounds, the extension of the spectral profile method to continuous state spaces, and the sharpness of the spectral profile. We will also explore some recent breakthroughs involving the conductance method. This will include some work on both new methods for proving conductance inequalities and using profile-based methods to obtain sharp bounds. Throughout, there will be an emphasis on illustrative examples and and comparison of these methods to classical methods.

Finally, we will devote some time to survey (de-)coupling methods that have been developed in the statistics community. We will emphasize the distinction between coupling methods for analyzing the convergence of chains and those for analyzing the convergence of ergodic averages.

Joris Bierkens (TU Delft)

Spectral theory and asymptotic variance of piecewise deterministic samplers

Abstract: Markov Chain Monte Carlo (MCMC) is an essential computational tool in quantitative fields such as Bayesian statistics, physics and machine learning. In recent years piecewise deterministic Markov processes (PDMPs) have emerged as a promising alternative to classical MCMC algorithms. In this talk PDMP based algorithms such as the Zig-Zag Sampler and the Bouncy Particle Sampler will be introduced and recent progress in our understanding of the underlying processes will be presented. My aim is to zoom in on two types of convergence properties, as described by (i) the asymptotic variance and (ii) the spectral properties of the process. As it turns out, these can sometimes provide apparently conflicting information.

Alain Durmus (ENS Paris Saclay)

Quantitative convergence of Unadjusted Langevin Monte Carlo and application to stochastic approximation

Abstract: Stochastic approximation methods play a central role in maximum likelihood estimation problems involving intractable likelihood functions, such as marginal likelihoods arising in problems with missing or incomplete data, and in parametric empirical Bayesian estimation. Combined with Markov chain Monte Carlo algorithms, these stochastic optimisation methods have been successfully applied to a wide range of problems in science and industry. However, this strategy scales poorly to large problems because of methodological and theoretical difficulties related to using high-dimensional Markov chain Monte Carlo algorithms within a stochastic approximation scheme. This paper proposes to address these difficulties by using unadjusted Langevin algorithms to construct the stochastic approximation. This leads to a highly efficient stochastic optimisation methodology with favourable convergence properties that can be quantified explicitly and easily checked. The proposed methodology is demonstrated with three experiments, including a challenging application to high-dimensional statistical audio analysis and a sparse Bayesian logistic regression with random effects problem.

Ioannis Kontoyannis (Cambridge)

Variable-dimension MCMC samplers for variable-memory Markov models

Abstract: A new Bayesian modelling framework was recently developed for discrete time series, based on the class of higher-order, variable-memory Markov chain models. In particular, an exact inference algorithm was introduced, which identifies the a posteriori most likely models and computes their exact posterior probabilities. But it is computationally infeasible beyond the top 5 or 10 most likely models. To facilitate further, effective exploration of the posterior distribution on both models a parameters, we describe a new family of variable-dimension Markov chain Monte Carlo samplers. Their performance is illustrated both on simulated and on real-world data sets, for model selection, Markov order estimation, and parameter estimation. The proposed samplers are found to perform at least as well as – and usually better than – the most commonly used and the state-of-the-art approaches, in applications with data from finance, genetics, neuroscience, and animal communication.

Jianfeng Lu (Duke)

Quantitative convergence analysis of hypocoercive sampling dynamics

---

Abstract: In this talk, we will discuss some recent advances on quantitative analysis of convergence of hypocoercive sampling dynamics, including underdamped Langevin dynamics, randomized Hamiltonian Monte Carlo, zigzag process and bouncy particle sampler. The analysis is based on a variational framework for hypocoercivity which combines a Poincare-type inequality in time-augmented state space and an energy estimate. Based on joint works with Yu Cao (NYU) and Lihan Wang (Duke).

Manon Michel (Université de Clermont-Auvergne)

Using symmetries as an efficiency compass in MCMC

Abstract: This talk will aim at giving some historical perspectives into MCMC developments in computational physics, with the overall goal to give insights on what could be tomorrow's more efficient MCMC schemes. The key concepts will rely on trading the reversibility symmetry for distribution ones and on extracting local information.

Eric Moulines (École Polytechnique)

Variance reduction for MCMC algorithms

Abstract: New methodologies are presented for the construction of control variates to reduce the variance of additive functionals of Markov Chain Monte Carlo (MCMC) samplers. I will present three approaches that we have recently developed.
The first approach defines control variates through the minimization of the asymptotic variance of the Langevin diffusion over a family of functions, which can be seen as a quadratic risk minimization procedure. The use of these control variates is theoretically justified. We show that the asymptotic variances of some well-known MCMC algorithms, including the Random Walk Metropolis and the (Metropolis) Unadjusted/Adjusted Langevin Algorithm, are close to the asymptotic variance of the Langevin diffusion.
The second approach relies on a novel discrete time martingale representation for Markov chains. Our approach is fully non-asymptotic and does not require any type of ergodicity or special product structure of the underlying density. By rigorously analyzing the convergence properties of the proposed algorithm, we show that it's complexity is indeed asymptotically smaller than one of the original MCMC algorithm.
The third approach combines the use of control functions with the minimisation of an empirical asymptotic variance estimate. We analyse finite sample properties of the proposed method and derive convergence rates of the excess asymptotic variance to zero.
We present empirical results carried out on a number of real-world benchmarks showing that our variance reduction methods achieve significant improvement as compared to state-of-the-art methods at the expense of a moderate increase of computational overhead.
(Joint work with Alain Durmus, Nicolas Brosse, Sean Meyn, Denis Belomestny, Alexei Naumov, Leonid Iosispoi, Serguei Samsonov)

Michela Ottobre (Heriot-Watt University Edinburgh)

Uniform in time approximations of stochastic dynamics

Abstract: Complicated models, for which a detailed analysis is too far out of reach, are routinely approximated via a variety of procedures, for example by use of numerical schemes. When using a numerical scheme we make an error which is small over small time-intervals but it typically compounds over longer time-horizons. Hence, in general, the approximation error grows in time so that the results

of our simulations are less reliable when the simulation is run for longer. However this is not necessarily the case and one may be able to find dynamics and corresponding approximation procedures for which the error remains bounded, uniformly in time. We will discuss some criteria and general procedures to understand when this is possible. We will do this both for approximations generated via numerical schemes, but also for more general approximation procedures, i.e. averaging.

Daniel Rudolf (Universität Göttingen)

Spectral gap of slice sampling

Abstract: We provide results on Wasserstein contraction of simple slice sampling for approximate sampling with respect to distributions with log-concave and rotational invariant Lebesgue densities. This yields, in particular, an explicit quantitative lower bound of the spectral gap of simple slice sampling. Moreover, this lower bound carries over to more general target distributions depending only on the volume of the (super-)level sets of their unnormalized density. This allows us to deduce convergence results of hybrid slice sampling approaches.