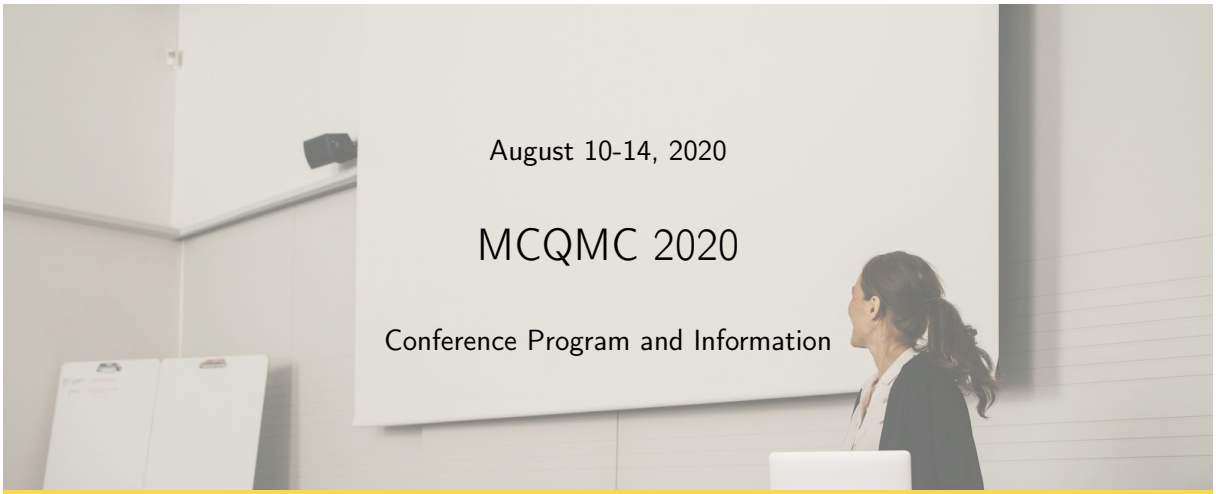


August 10-14, 2020

MCQMC 2020

Conference Program and Information

14th International Conference in Monte Carlo & Quasi-Monte Carlo Methods in Scientific Computing

Revision September 9, 2020

Contents

Welcome to MCQMC 2020

Welcome to the *14th Conference on Monte Carlo and Quasi-Monte Carlo Methods in Scientific Computing*, MCQMC 2020. The MCQMC conference series is the major event for researchers investigating and developing Monte Carlo and quasi-Monte Carlo methods, held in alternate years with the International Conference on Monte Carlo Methods and Applications (MCM).

The Local Organizing Committee had been looking forward to welcoming you all to Oxford. The plan was for the conference to be held in the Mathematical Institute, and the conference dinner would have been nearby in the Dining Hall of Keble College. On Wednesday afternoon we would have had a tour of Oxford, and we might even have arranged for some punting on the river if the weather was good.

Unfortunately of course, the Coronavirus pandemic has completely ruined all of these plans. Meeting in person is not possible; indeed the UK has been one of the worst affected countries with all of the Local Organizing Committee significantly impacted in their jobs. We are therefore very grateful to the MCQMC Steering Committee, led by Alex Keller, who have stepped in to assist greatly in putting on this online version of MCQMC 2020. We also appreciate the assistance of the International Centre for Mathematical Sciences (ICMS) in hosting the plenary talks and tutorials.

We are very pleased that so many contributors have confirmed that they are happy to continue in this new conference format. We have nine one-hour invited plenary talks and two 90 mins tutorials, all from leading researchers; these will take place live via Zoom, with the talks and Q&A being recorded for those in incompatible timezones. Minisymposia and special sessions are pre-recorded so that they may be watched from within any time zone.

We hope that despite the different form of this year's conference you will nevertheless enjoy a stimulating week of talks at MCQMC 2020, and we can look forward to seeing each other in person at MCM 2021, and MCQMC 2022.

The organizers,

Mike Giles, Arnaud Doucet, and Alex Keller

The MCQMC Conference Series

The MCQMC conference series is a biennial meeting focused on Monte Carlo (MC) and quasi-Monte Carlo (QMC) methods in scientific computing. The conference attracts between 150 and 200 participants. Its aim is to provide a forum where leading researchers and users can exchange information on the latest theoretical developments and important applications of these methods.

In a nutshell, MC methods study complex systems by simulations fed by computer-generated pseudorandom numbers. QMC methods replace these random numbers by more evenly distributed (carefully selected) numbers to improve their effectiveness. A large variety of special techniques are developed and used to make these methods more effective in terms of speed and accuracy. The conference focuses primarily on the mathematical study of these techniques, their implementation and adaptation for concrete applications, and their empirical assessment.

The conference was initiated by Harald Niederreiter, who co-chaired the first seven conferences. In 2006 Harald Niederreiter announced his wish to step down from the organizational role, and a Steering Committee was formed to ensure and oversee the continuation of the conference series. The locations of the 13 first conferences are set out below.

If you are interested in hosting a future MCQMC at your institution, then please contact any member of the steering committee.

Year	**Location**
1994	Las Vegas, NV USA
1996	Salzburg, Austria
1998	Claremont, CA USA
2000	Hong Kong
2002	Singapore
2004	Juan-Les-Pins, France
2006	Ulm, Germany
2008	Montréal, Canada
2010	Warsaw, Poland
2012	Sydney, Australia
2014	KU Leuven, Belgium
2016	Stanford, CA USA
2018	Rennes, France

Committees

University of Oxford Organizing Committee

Mike Giles	UK, Oxford
Arnaud Doucet	UK, Oxford
Abdul-Lateef Haji-Ali	UK, Heriot-Watt
Jeremy Heng	Singapore, ESSEC
Daniel Paulin	UK, Edinburgh
Alex Shestopaloff	UK, Alan Turing Institute
Lukasz Szpruch	UK, Edinburgh
Aretha Teckentrup	UK, Edinburgh

Berlin Organizing Committee

Alexander Keller	Germany, NVIDIA
Nikolaus Binder	Germany, NVIDIA
Thomas Müller	Germany, NVIDIA
Merlin Nimier-David	Switzerland, EPFL

Steering Committee

Alexander Keller (Chair)	Germany, NVIDIA
Josef Dick	Australia, University of New South Wales
Fred J. Hickernell	USA, Illinois Institute of Technology
Aicke Hinrichs	Austria, JKU Linz
Pierre L'Ecuyer	Canada, University Montréal
Art Owen	USA, Stanford University
Friedrich Pillichshammer	Austria, JKU Linz

Scientific Committee

Christophe Andrieu	UK, Bristol
Andrea Barth	Germany, Stuttgart
Zdravko Botev	Australia, University of New South Wales
Hector Cancela	Uruguay, University of the Republic
Frédéric Cérou	France, Inria
Nicolas Chopin	France, ENSAE
Ronald Cools	Belgium, KU Leuven
Josef Dick	Australia, University of New South Wales
Arnaud Doucet	UK, Oxford
Stefan Geiss	Finland, Jyväskylä
Mike Giles	UK, Oxford
Mark Girolami	UK, University of Warwick
Paul Glasserman	USA, Columbia
Michael Gnewuch	Germany, Osnabrück
Emmanuel Gobbet	France, Ecole Polytechnique
Takashi Goda	Japan, The University of Tokyo
Arnaud Guyader	France, University Pierre et Marie Curie

Scientific Committee (continued)

Stefan Heinrich	Germany, University Kaiserslautern
Fred J. Hickernell	USA, IIT
Aicke Hinrichs	Austria, JKU Linz
Wenzel Jakob	Switzerland, EPFL
Bert Kappen	Netherlands, Radboud
Alexander Keller	Germany, NVIDIA
Dirk Kroese	Australia, University Queensland
Frances Kuo	Australia, University of New South Wales
Gerhard Larcher	Austria, JKU Linz
Christian Lécot	France, University Savoie
Pierre L'Ecuyer	Canada, University Montréal
Christiane Lemieux	Canada, Waterloo
Gunther Leobacher	Austria, Graz
Faming Liang	USA, University Florida, Gainesville
Makoto Matsumoto	Japan, Hiroshima
Eric Moulines	France, Ecole Polytechnique
Thomas Müller-Gronbach	Germany, Passau
Andreas Neuenkirch	Germany, Mannheim
Harald Niederreiter	Austria, Academy of Sciences
Erich Novak	Germany, FSU, Jena
Dirk Nuyens	Belgium, KU Leuven
Art Owen	USA, Stanford University
Gilles Pagès	France, UPMC
Gareth Peters	UK, University College London
Friedrich Pillichshammer	Austria, JKU Linz
Mike Pitt	UK, KCL
Sebastian Reich	Germany, Potsdam
Klaus Ritter	Germany, Kaiserslautern
Gerardo Rubino	(France, Inria
Wolfgang Schmid	Austria, Salzburg
Ian Sloan	Australia, University of New South Wales
Lukasz Szpruch	UK, Edinburgh
Aretha Teckentrup	UK, Edinburgh
Raul Tempone	Germany, Aachen
Bruno Tuffin	France, INRIA
Grzegorz Wasilkowski	USA, University Kentucky
Henryk Woźniakowski	USA, Columbia University

Sponsors

We are very grateful for the support from the following sponsors

- Mathematical Institute, University of Oxford
- Department of Statistics, University of Oxford
- Engineering and Physical Sciences Research Council (EPSRC)
- International Centre for Mathematical Sciences (ICMS)

Practical Information

The 14th International Conference in Monte Carlo & Quasi-Monte Carlo in Scientific Computing will be held online due to the pandemic. We are planning to have the tutorials and plenary talks to be presented live. If you missed one of these presentations, you may watch them at https://media.ed.ac.uk/playlist/dedicated/51612401/1_0z0wec2z/1_ydrnqw0f. In order to attend the first live stream in the BST timezone, please visit the MCQMC2020 event web page of the International Center of Mathematical Sciences and use the link there to register. Some conference photos taken during the live sessions on Zoom will be posted on https://twitter.com/ICMS_Edinburgh.

The ICMS is conducting a conference survey. Your valuable feedback is very much appreciated!

Minisymposia and special sessions will comprise of recorded presentations released on Youtube. We may have the opportunity to have some live session there as well. Recordings of the presentations will be available on the MCQMC2020 YouTube channel.

As things are still in flow, please check our two companion web pages MCQMC 2020 and MCQMC 2020 Online for the latest updates and schedule. The book of abstracts will continuously be updated once the recordings are coming in.

Instructions for Speakers

Plenary talks are 50 minutes plus 10 minutes for questions and discussion. All other talks are 25 minutes, plus 5 minutes for questions and discussion.

Please make sure that you do not exceed your time. Given the short time allowed to each speaker, it is generally not possible to give the full details of your work. Focus on the essential of your message. You should concentrate on providing a clear explanation of your main results and their significance.

Presenting Live

Please check our two companion web pages MCQMC 2020 and MCQMC 2020 Online for the latest updates and schedule.

Recording Presentations

Presentations in minisymposia and special sessions will have to be recorded and submitted to a file server prior to their publication on the corresponding MCQMC2020 Youtube channel. The details are still being worked out.

Here are some recording hints for the most common operating systems and programs:

- On Windows, pressing the Windows-Key in combination with 'G' will start a screen recording. You may use your favorite presentation software and talk to record your contribution.
- On macOS, you may use Quicktime to record a screen capture of your presentation with audio. As an option, you may check out the presentation tool http://iihm.imag.fr/blanch/software/osx-presentation/ at your own risk.
- If you want to go professional, the Open Broadcaster Software https://obsproject.com allows for anything from simple screen capture to professional production. The software is free and available for Windows, macOS 10.13+, and Linux.
- There will be many more options like recording your presentation using Webex or Zoom, if available to you.
- Finally, both Powerpoint and Keynote offer an option to record your presentation with audio as a video.

Beyond that, please check the web for advice.

Uploading Presentations

Presentations in minisymposia and special sessions will have to be recorded and submitted to the organizers prior to the conference. With your permission, we will publish your recorded presentations on the MCQMC2020 YouTube channel for other participants to watch.

Instructions for the uploading procedure will be sent by email.

Instructions for Session Chairs

Session chairs have the responsibility to make sure the speakers adhere tightly to the schedule. Session chairs will notify the speaker once 5, 3, and 1 minutes, respectively, are left of the scheduled time.

Session chairs should also contact their session speakers ahead of time to verify their presence and inform the organizers of any potential no-shows. Technical modalities shall be checked safely ahead of time.

Conference Proceedings

There is a long standing tradition that a selection of strictly refereed scientific articles are published after the conference. The proceedings will be entitled Monte Carlo and Quasi-Monte Carlo Methods 2020, edited by Mike Giles, Arnaud Doucet.

Every speaker may submit an article on the topic of the presented talk. Submissions must be formatted using the LaTeX type setting system and a document style provided on the conference web page. While a submission via an https://arxiv.org link is preferred, for the reviewing process, files in the PDF format will be accepted as well.

The deadline for submission will be Dec. 18, 2020. Detailed instructions will be made available through the conference website.

Conference Schedule

Sessions are color-coded:

- Tutorials and plenary talks are in **gold**,
- minisymposia are in **green**,
- special sessions are in **blue**, and
- all other sessions are in **red**.

Clickable links are in magenta. In the schedule, these include the links to the presentations on Youtube, while the page number takes you to the presentation abstract. The abstracts link back to the schedule.

Note that both tutorials and plenary talks will be run in British Summer Time (BST), UTC +1. Check https://time.is/United_Kingdom.

Tutorial and plenary presentations will be presented live and recorded for viewing afterwards, check the playlist at The University of Edinburgh. Minisymposia and special session presentations will be provided as videos. Please check our two companion web pages MCQMC 2020 and MCQMC 2020 Online for the latest updates and schedule.

Registration

In order to attend the first live stream in the BST timezone, please visit the MCQMC2020 event web page of the International Center of Mathematical Sciences and use the link there to register.

Monday, August 10

14:00 – 15:30 BST	**Tutorial**
	Aretha Teckentrup
	Markov Chain Monte Carlo Methods
	Links: Recording, Slides
	Chair: Mike Giles p. 34
15:45–17:15 BST	**Tutorial**
	Fred J. Hickernell
	Quasi-Monte Carlo Software
	Links: Recording, Slides, Google Colaboratory Notebook, Blog
	Chair: Mike Giles p. 35

Tuesday, August 11

14:00 – 14:15 BST	**Welcome**
	Mike Giles
14:15 – 15:15 BST	**Plenary Talk**
	Mario Ullrich
	Random vs. Optimal Information for L_2-Approximation
	Links: Recording, Slides
	Chair: Aicke Hinrichs p. 38
15:15 – 16:00 BST	**Breakout Sessions**
16:00 – 17:00 BST	**Plenary Talk**
	Jing Dong
	Can Algorithms Collaborate? The Replica Exchange Method
	Links: Recording, Slides
	Chair: Pierre L'Ecuyer p. 39
17:00 – 17:30 BST	**Breakout Sessions**

Wednesday, August 12

14:00 – 15:00 BST	**Plenary Talk**
	Mark Jerrum
	A complexity-theoretic perspective on MCMC
	Links: Recording, Slides
	Chair: Matti Vihola p. 40
15:15 – 16:15 BST	**Plenary Talk**
	David Pfau
	Quantum Monte Carlo and Solving the Many-Electron Schrödinger Equation with Deep Neural Networks
	Links: Recording, Slides
	Chair: Geoff Nicholls p. 41
16:30 – 17:30 BST	**Plenary Talk**
	Ives Atchadé
	Approximate Spectral Gap for MCMC Mixing Times in High Dimensions
	Links: Recording, Slides, Article
	Chair: Christophe Andrieu p. 42

Thursday, August 13

14:00 – 15:00 BST	**Plenary Talk**
	Peter Kritzer
	Exponential Tractability
	Links: Recording, Slides
	Chair: Fred Hickernell p. 43
15:00 – 15:45 BST	**Breakout Sessions**
15:45 – 16:45 BST	**Plenary Talk**
	Thomas Müller
	Variance Reduction using Neural Networks
	Links: Recording, Slides
	Chair: Alex Keller p. 44
16:45 – 17:00 BST	**MCQMC 2022 announcement**
	Alex Keller and Peter Kritzer: Slides

Friday, August 14

14:00 – 15:00 BST	**Plenary Talk**
	Claudia Schillings
	Optimization approaches for Bayesian inverse problems: Preconditioning integration methods in the small noise or large data limit
	Links: Recording, Slides
	Chair: Aretha Teckentrup p. 45
15:00 – 15:30 BST	**Breakout Sessions**
15:30 – 16:30 BST	**Plenary Talk**
	Pierre L'Ecuyer
	Quasi-Monte Carlo for Density Estimation
	Links: Recording, Slides
	Chair: Art Owen p. 46
16:30 – 16:45 BST	**Goodbyes**
	Mike Giles

Minisymposium

Stein's Method in Computational Statistics

Chair: Chris Oates p. 48

Chris Oates

Recasting Sampling as Optimization via Stein's Method

Youtube p. 48

Leah South

Monte Carlo Variance Reduction Using Stein Operators

Youtube p. 49

François-Xavier Briol

Learning to Reduce Variance Using Stochastic Gradient Descent

Youtube p. 50

Andrew Duncan

On the Geometry of Stein Variational Gradient Descent

Youtube p. 50

Minisymposium

Hierarchical Methods for Variance Reduction

Chair: Chiheb Ben Hammouda p. 51

Chiheb Ben Hammouda

Importance Sampling for a Robust and Efficient Multilevel Monte Carlo Estimator for Stochastic Reaction Networks

Youtube p. 52

Nadhir Ben Rached

Dynamic splitting method for rare events simulation

Youtube p. 52

Raúl Tempone

Combining numerical smoothing with multilevel Monte Carlo for efficient option pricing and density estimation

Youtube p. 53

Minisymposium

Combinations of Importance Sampling and MCMC

Chair: Daniel Rudolf and Björn Sprungk p. 54

Víctor Elvira

Anti-tempered layered adaptive importance sampling

Youtube p. 54

Ilja Klebanov

Markov Chain Importance Sampling

Youtube p. 55

Björn Sprungk

On a Metropolis–Hastings importance sampling estimator

Youtube p. 55

Matti Vihola

Conditional particle filters with diffuse initial distributions

Youtube p. 56

Minisymposium

Deep Thought - Analysis and Application of Hierarchical Models

Chair: Jonas Latz p. 57

Aretha Teckentrup

Convergence of Gaussian process regression with estimated hyper-parameters

Youtube p. 57

Subhadip Mukherjee

Unsupervised Deep Learning Approaches for Inverse Problems

Youtube p. 58

Lassi Roininen

Hierarchical and non-Gaussian Models for Bayesian Inversion

Youtube p. 59

Jonas Latz

Fast and even faster sampling of parameterised Gaussian random fields

Youtube p. 59

Tapio Helin

Hyperparameter Estimation in Bayesian Inverse Problems

Youtube p. 60

Laura Scarabosio

Model-based multilevel Monte Carlo methods for local quantities of interest in random heterogeneous media

Youtube p. 60

Minisymposium

Probabilistic Numerical and Kernel-Based Methods (Part 1 of 2)

Chair: Toni Karvonen p. 61

Jonathan Cockayne

Bayesian Probabilistic Numerical Methods in Integration

Youtube p. 61

Matthew Fisher

Adaptive Algorithms in Bayesian Quadrature

Youtube p. 62

Luc Pronzato

Design of Computer Experiments based on Bayesian Quadrature

Youtube p. 62

Takuo Matsubara

Quadrature of Bayesian Neural Networks

Youtube p. 63

Minisymposium

Probabilistic Numerical and Kernel-Based Methods (Part 2 of 2)

Chair: Toni Karvonen p. 64

Toni Karvonen

Kernel Methods, Gaussian Processes and Uncertainty Quantification for Bayesian Quadrature

Youtube p. 64

Motonobu Kanagawa

On the Positivity of Bayesian Quadrature Weights

Youtube p. 65

Ayoub Belhadji

Variations around kernel quadrature with DPPs

Youtube p. 65

Ken'ichiro Tanaka

Generation of Point Sets by Global Optimization for Kernel-Based Numerical Integration

Youtube p. 66

Minisymposium

Piecewise Deterministic Markov Chain Monte Carlo Methods and Hypocoercivity

Chair: Daniel Paulin p. 67

Minisymposium

Scalable Markov Chain Monte Carlo Algorithms (Part 2 of 2)

Chair: Deborshee Sen p. 76

Gareth Roberts

Bayesian Fusion

Youtube p. 76

Inass Sekkat

Removing the Mini-Batching Error in Bayesian Inference using Adaptive Langevin Dynamics

Youtube p. 77

Deborshee Sen

Efficient sub-sampling for stochastic gradient MCMC for hidden Markov models

Youtube p. 77

Minisymposium

Variance-Reduced Estimators for Expected Information Gains in Bayesian Optimal Experimental Design

Chair: Joakim Beck p. 78

Joakim Beck

Multilevel Double-loop Monte Carlo to Simulation-based Bayesian Optimal Experimental Design

Youtube p. 79

Tomohiko Hironaka

Two Applications of Multilevel Monte Carlo Methods to Bayesian Experimental Designs

Youtube p. 79

André Gustavo Carlon

Stochastic Optimization for Bayesian Design of Experiments

Youtube p. 80

Minisymposium

Stochastic Computation and Complexity

Chair: Larisa Yaroslavtseva p. 87

Monika Eisenmann

Backward Euler–Maruyama method for SDEs with multi-valued drift coefficient

Youtube p. 88

Máté Gerencsér

Approximation of SDEs – a stochastic sewing approach

Youtube p. 88

Emmanuel Gobet

Sampling scheme for intractable copula function, application to the computation of tail events in factor copula model

Youtube p. 89

Stefan Heinrich

Monte-Carlo Algorithms with Restricted Access to Randomness

- p. 89

Andreas Neuenkirch

Weak convergence rates of semi-exact discretization schemes for the Heston model

Youtube p. 90

Paweł Przybyłowicz

Stability and Convergence of Randomized Runge-Kutta Method Under Inexact Information

- p. 90

Daniel Rudolf

Spectral gap of slice sampling

Youtube p. 91

Sotirios Sabanis

A fully data-driven approach to minimizing CVaR for portfolio of assets via SGLD with discontinuous updating

Youtube p. 92

Dai Taguchi

Multi-dimensional Avikainen's estimate

Youtube p. 92

Yue Wu

Semi-implicit Taylor schemes for stiff rough differential equations

Youtube p. 93

Larisa Yaroslavtseva

On sharp lower error bounds for strong approximation of SDEs with discontinuous drift coefficient

Youtube p. 93

Special Session

MCQMC and Machine Learning

Chair: First Last p. 96

Art Owen

Mean dimension of ridge functions

Youtube p. 96

Avinash Prasad

Quasi-Monte Carlo for Multivariate Distributions via Generative Neural Networks

Youtube p. 97

Gurprit Singh

How to Train Your Samples?

Youtube p. 98

Alexander Keller

Quasi-Monte Carlo Methods and Neural Networks

Youtube p. 99

Special Session

Sampling

Chair: First Last p. 100

Special Session

Multi-Level Monte Carlo

Chair: First Last p. 108

Emil Løvbak

Multilevel Monte Carlo with improved correlation for kinetic equations in the diffusive scaling

Youtube p. 108

Giray Ökten

Multilevel Monte Carlo for LIBOR Market Model

Youtube p. 109

Sundar Ganesh

Efficient Multi-Level Monte Carlo Estimators for Risk-Averse Engineering Design

Youtube p. 110

Philippe Blondeel

p-refined Multilevel Quasi-Monte Carlo for Galerkin Finite Element Methods with applications in Geotechnical Engineering

Youtube p. 111

Special Session

Markov Chain Monte Carlo

Chair: First Last p. 112

Luiz Max Carvalho

Bayesian Estimation of Time-Trees

Youtube p. 112

Florian Puchhammer

Hamiltonian Monte Carlo Applied to Inverse Problems in Imaging the Earth's Subsurface

Youtube p. 114

Robin J. Ryder

Component-wise approximate Bayesian computation via Gibbs-like steps

Youtube p. 115

Thomas Catanach

Multifidelity Sequential Tempered Markov Chain Monte Carlo for Bayesian Inference

Youtube p. 116

Tobias Siems

A Note on the Metropolis-Hastings Acceptance Probabilities for Mixture Spaces

Youtube p. 117

Alexander Dorndorf

A Bayesian Robust Nonlinear Multivariate Time Series Model with Autoregressive and t-Distributed Errors – A Case Study for GNSS Data

Youtube p. 118

Shin Harase

Implementing short-period Tausworthe generators for Markov chain quasi-Monte Carlo

Youtube p. 119

Matthew Graham

Manifold MCMC Methods for Efficient Inference in a Wide Class of Diffusion Models

Youtube p. 120

Samuel Livingstone

The Barker Proposal: Robust, Gradient-based MCMC

Youtube p. 121

Special Session

Differential Equations

Chair: First Last p. 122

Adrian Ebert

Construction of QMC Finite Element Methods for Elliptic PDEs with Random Coefficients by a Reduced CBC Construction Algorithm

Youtube p. 122

Chi-Ok Hwang

Infinite Parallel Plates Algorithm

Youtube p. 123

Flavius Guiaș

Stochastic Runge–Kutta methods based on Markov jump processes and applications to deterministic systems of differential equations

Youtube p. 124

Special Session

Variance Reduction

Chair: First Last p. 125

Jamie Fox

Polynomial Chaos as a Control Variate Method

Youtube p. 125

Emanouil Atanassov

On the Use of Global Sensitivity Analysis for Optimisation of Numerical Schemes for Evaluation of Option Prices and Greeks

Youtube p. 126

Nabil Kahalé

Randomized Dimension Reduction for Monte Carlo Simulations

Youtube p. 127

Ad Ridder

Rare-Event Analysis and Simulation of Queues with Time-Varying Rates

Youtube p. 128

Nikolaus Binder

Solving Integral Equations in real-time

Youtube p. 129

Special Session

Optimization and Optimality

Chair: First Last p. 130

Kamélia Daudel

Infinite-Dimensional Alpha-Divergence Minimisation for Variational Inference

Youtube p. 130

Stefka Fidanova

Monte Carlo Optimization Including Weights

Youtube p. 131

Miroslav Vořechovský

Monte Carlo integration using designs obtained via Periodic Maximin and ϕ_p (Phi) Criteria

Youtube p. 132

Ryan White

On the Evolution of Minimal-Volume, Sufficient-Probability Sets for Stochastic Paths

Youtube p. 133

Special Session

Methods for Engineering and Finance

Chair: First Last p. 134

Chao Zheng

Higher-order weak schemes for the Heston stochastic volatility model by extrapolation

Youtube p. 134

Soumya Roy

Statistical Analysis of Progressive Type-I Interval Censored Data under Competing Risks

Youtube p. 135

Natalya Tracheva

A new combined kernel-projection statistical estimator with applications to the study of polarized radiation intensity

Youtube p. 136

Edmund Ryan

Efficient Monte Carlo methods for forecasting returns from large-scale consumer debt portfolios

Youtube p. 137

Special Session

Discrepancy and Quasi-Monte Carlo Methods

Chair: First Last p. 138

Mario Neumüller

On the asymptotic behaviour of the Sudler product of sines

- p. 138

Robert Nasdala

Multivariate approximation based on transformed rank-1 lattices

Youtube p. 139

Marcin Wnuk

Reproducing Kernel Banach Spaces and QMC Integration

Youtube p. 140

Michael Gnewuch

Optimal cubature rules on Haar wavelet spaces and on spaces with fractional smoothness

Youtube p. 141

Marcello Longo

A-posteriori QMC error estimation

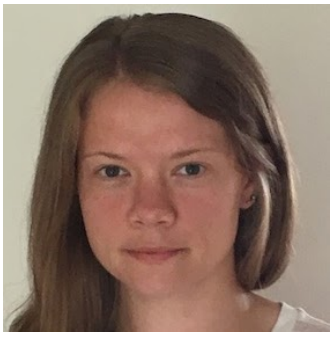
Youtube p. 142

Abstracts of Tutorials

Monday 14:00 – 15:30 BST **Live on Zoom**

Markov Chain Monte Carlo Methods

Chair: Mike Giles



Aretha Teckentrup

Aretha Teckentrup
School of Mathematics, University of Edinburgh, Scotland,
a.teckentrup@ed.ac.uk

Sampling methods are frequently used to compute statistics of quantities of interest in applications, and are invaluable tools in modern simulation tasks.

This tutorial will give an algorithmic introduction to Markov chain Monte Carlo methods, a class of sampling methods frequently used in Bayesian statistics and computational physics, in which the probability distribution of interest is usually not known in closed form.

We focus on the Metropolis-Hastings framework and discuss well known methods including random walk (RW), pre-conditioned Crank-Nicholson (pCN), and Metropolis-adjusted Langevin (MALA). Some theoretical properties of the algorithms will be discussed to highlight their advantages or disadvantages, but the focus will be on the algorithmic foundations.

Monday 15:45–17:15 BST **Live on Zoom**

Quasi-Monte Carlo Software

Chair: Mike Giles

Fred J. Hickernell

Sou-Cheng Choi
Kamakura Corporation, USA and Department of Applied Mathematics, Illinois Institute of Technology, USA, schoi32@iit.edu

Fred J. Hickernell
Department of Applied Mathematics, Illinois Institute of Technology, USA, hickernell@iit.edu

R. Jagadeeswaran
Department of Applied Mathematics, Illinois Institute of Technology, USA, jrathin1@hawk.iit.edu

Michael J. McCourt
SigOpt, USA, mccourt@sigopt.com

Aleksei Sorokin
Department of Applied Mathematics, Illinois Institute of Technology, USA, asorokin@hawk.iit.edu

Quasi-Monte Carlo (QMC) methods achieve substantial efficiency gains by replacing independent and identically distributed (IID) random points by low discrepancy (LD) points. LD point generators and QMC algorithms are active research areas. Practitioners are attracted to QMC by the promise of efficiency gains.

This tutorial highlights several readily available QMC software libraries in various languages. We describe the components of a QMC calculation: the LD point generators, problem specification, methods for speeding up the computation, and stopping criteria. We argue that excellent QMC software requires the collaboration of a community—not only the efforts of individual research groups.

During this tutorial we provide hands-on experience with QMCPy [2], a Python 3 library that draws on the work of several experts [1, 3, 4, 5, 6, 7]. We do this through the Google Colaboratory notebook at https://tinyurl.com/QMCPyTutorial. QMCPy grew out of discussions held at MCQMC 2018. Minimal experience with QMC or Python is assumed.

[1] S.-C. T. Choi, Y. Ding, F. J. Hickernell, L. Jiang, Ll. A. Jiménez Rugama, D. Li, R. Jagadeeswaran, X. Tong, K. Zhang, Y. Zhang, and X. Zhou, GAIL: Guaranteed Automatic Integration Library (Version 2.3.1) [MATLAB Software], http://gailgithub.github.io/GAIL_Dev/, 2020.

[2] S.-C. T. Choi, F. J. Hickernell, R. Jagadeeswaran, M. J. McCourt, and A. Sorokin, QMCPy: A quasi-Monte Carlo Python Library, https://qmcsoftware.github.io/QMCSoftware/, 2020.

[3] M. B. Giles, Multilevel Monte Carlo software, http://people.maths.ox.ac.uk/~gilesm/mlmc/, 2020.

[4] M. Hofert and C. Lemieux, QRNG: (Randomized) Quasi-Random Number Generators. R package version 0.0-7, https://CRAN.R-project.org/package=qrng, 2019.

[5] P. L'Ecuyer and D. Munger, Lattice Builder, http://simul.iro.umontreal.ca/latbuilder/, 2016.

[6] D. Nuyens, MPS: Magic Point Shop, https://people.cs.kuleuven.be/~dirk.nuyens/qmc-generators/, 2017.

[7] A. B. Owen, A randomized Halton algorithm in R, https://arxiv.org/abs/1706.02808, 2017.

Abstracts of Plenary Talks

Tuesday 14:15 – 15:15 BST **Live on Zoom**

Random vs. Optimal Information for L_2-Approximation

Chair: Aicke Hinrichs

Mario Ullrich

Mario Ullrich
Institut für Analysis, Johannes Kepler University Linz, Austria,
mario.ullrich@jku.at

In this talk we discuss the general problem of how good certain admissible information of a function f is compared to optimal information for L_2-approximation in a Hilbert space H. This means that, as usual, we allow a class $\Lambda \subset H'$ (=the dual of H) of admissible information, like function values of Fourier coefficients, and consider the worst-case error of the best algorithm based on information from this class. We denote by $e_n(H;\Lambda)$ the minimal achievable worst case error (in the unit ball of H) based on n pieces of information from Λ. In the case $\Lambda = H'$, i.e., we allow all linear information, we have that $e_n(H;H')$ equal the approximation numbers $a_n = a_n(H)$, and we would like to know which classes of information Λ lead to $e_n(H;\Lambda) \asymp a_n$.

In [1, 2] we approached this problem from a different side, by considering information that is "randomly chosen" in H' (w.r.t. a suitable Gaussian). In particular, we proved that $a_n \lesssim n^{-\alpha}$ for some $\alpha > 1/2$ implies that random information is with overwhelming probability as good as the optimal information. This shows, to some extent, that optimal information is not special at all in these cases.
This motivates the study in [3, 4, 5] where we proved, somewhat surprisingly, that a similar statement holds if we restrict ourselves to much smaller classes of information Λ^*, like function values or coefficients w.r.t. an arbitrary fixed ONB of L_2. In particular, we obtain for these classes that $e_{n\log(n)}(H;\Lambda^*) \lesssim a_n$, if $a_n \lesssim n^{-\alpha}$ for some $\alpha > 1/2$. As this holds again with high probability, we obtain, again to some extent, that optimal sampling points for L_2-approximation are not something special. Moreover, our general bounds are tight enough to improve upon existing bounds for some special Hilbert spaces.

[1] A. Hinrichs, D. Krieg, E. Novak, J. Prochno and M. Ullrich. On the power of random information. In: F.J. Hickernell, P. Kritzer (eds.), *Multivariate Algorithms and Information-Based Complexity*, pp. 43–64, DeGruyter, Berlin/Boston, 2020.

[2] A. Hinrichs, D. Krieg, E. Novak, J. Prochno and M. Ullrich. Random sections of ellipsoids and the power of random information. Submitted.

[3] D. Krieg and M. Ullrich. Function values are enough for L_2-approximation. Subm.

[4] M. Ullrich. On the worst-case error of least squares algorithms for L_2-approximation with high probability. *J. Complexity* (to appear).

[5] A. Hinrich, D. Krieg and M. Ullrich. On Gelfand width w.r.t. an arbitrary orthonormal basis. In preparation.

Tuesday 16:00 – 17:00 BST **Live on Zoom**

Can Algorithms Collaborate? The Replica Exchange Method

Chair: Pierre L'Ecuyer

Jing Dong

Jing Dong
Graduate School of Business, Columbia University, USA,
jing.dong@gsb.columbia.edu

Xin T. Tong
Department of Mathematics, National University of Singapore, Singapore,
mattxin@nus.edu.sg

The division of labor is the secret of any efficient enterprise. By collaborating with individuals with different skillsets, we can focus on tasks within our own expertise and produce better outcomes than working independently.

In this talk, we investigate whether the same principle can be applied when designing an algorithm. We introduce a simple collaboration mechanism called replica exchange, which was first developed in molecular dynamics. We demonstrate two applications of the method: one is to solve nonconvex optimization problems, the other is to sample from mixture-type target distributions. For the nonconvex optimization problem, we use replica exchange to facilitate the collaboration between gradient descent and Langevin dynamics.

We show that this new algorithm converges to the global minimum linearly with high probability, assuming the objective function is strongly convex in a neighborhood of the unique global minimum. By replacing gradients with stochastic gradients, and adding a proper threshold to the exchange mechanism, our algorithm can also be used in online settings. For the sampling problem where the target distribution is a mixture of multiple log-concave densities concentrated around isolated modes, we show that replica exchange Langevin diffusions with properly chosen temperature and exchange intensity can achieve constant or better convergence rates. We further quantify the benefit of replica exchange for multiple Langevin diffusions sampling at different temperatures.

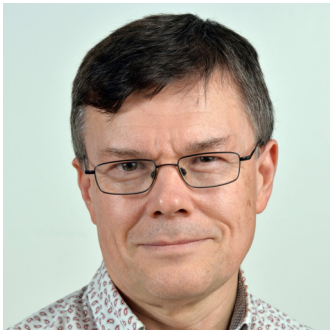
[1] Jing Dong and Xin T. Tong, Replica Exchange for Non-Convex Optimization, available on arxiv: 2001.08356, 2020

[2] Jing Dong and Xin T. Tong, Spectral Gap of Replica Exchange Langevin Diffusion on Mixture Distributions, available on arxiv:2006.16193, 2020

Wednesday 14:00 – 15:00 BST **Live on Zoom**

A complexity-theoretic perspective on MCMC

Chair: Matti Vihola



Mark Jerrum

Mark Jerrum
Queen Mary University of London, UK, m.jerrum@qmul.ac.uk

Computational complexity attempts to quantify the computational resources required to achieve given computational goals. The complexity classes P, NP and NP-complete that arise in the classification of decision problems are by now quite familiar outside of theoretical computer science. However, MCMC is connected with sampling and (approximate) counting problems as opposed to decision problems. For example, in the context of models in statistical physics we are interested in sampling a configuration from a Gibbs distribution, or in estimating a partition function. Again, in Bayesian analysis we would like to sample from from a posterior distribution. This leads us into the study of the computational complexity of sampling and (approximate) counting problems, a less familiar area. I will briskly survey MCMC from a computational complexity perspective, covering the following.

- What is tractable: the design and analysis of MCMC algorithms, including the analytical techniques used to derive a priori bounds on mixing times of Markov chains.
- What is intractable: the evidence for certain sampling and counting problems being intrinsically computationally hard.
- What is currently unknown: problems that are inaccessible to our current methods.

This is a rapidly developing area at the moment, and whatever I say under the third heading will probably be out of date in a few months.

Wednesday 15:15 – 16:15 BST **Live on Zoom**

Quantum Monte Carlo and Solving the Many-Electron Schrödinger Equation with Deep Neural Networks

Chair: Geoff Nicholls

David Pfau

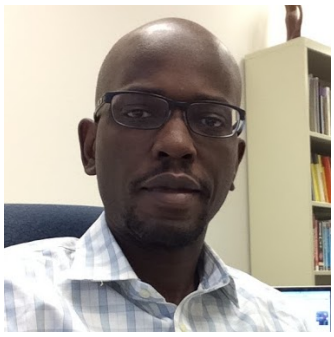
David Pfau
Google Deep Mind, pfau@google.com

Monte Carlo methods originated in computational physics, and only later were adopted by the statistics and machine learning communities. One physics application, Quantum Monte Carlo, uses Monte Carlo methods to solve the complex high-dimensional integrals for the energy of quantum systems, enabling direct computational solution of the otherwise intractable Schrödinger equation for many-particle systems. In this talk, I will present work on how ideas from the machine learning community can give back to computational physics, in particular deep neural networks and approximate natural gradient descent. I will present a novel deep neural network architecture, the Fermionic Neural Network (Fermi Net), which can be used as an expressive class of approximate solutions to the Schrödinger equation for many-electron systems. We optimize the Fermi Net by Kronecker-Factorized Approximate Curvature (KFAC), which makes it possible to scale algorithms from the Quantum Monte Carlo community to much more complex approximate wave functions. We show that the Fermi Net is able to achieve much higher accuracy than other Quantum Monte Carlo methods on challenging systems. I will also discuss some counterintuitive results on how different MCMC methods interact with optimization of the Fermi Net, suggesting directions for future work on the problem of simultaneously sampling and optimizing a probability distribution.

Wednesday 16:30 – 17:30 BST **Live on Zoom**

Approximate Spectral Gap for MCMC Mixing Times in High Dimensions

Chair: Christophe Andrieu



Ives Atchade

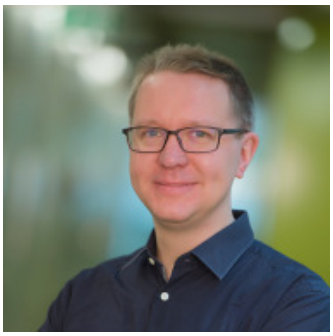
Ives Atchadé
Boston University, atchade@bu.edu

Understanding the type of problems for which fast Markov Chain Monte Carlo (MCMC) sampling is possible is a question of fundamental interest. The study of the size of the spectral gap is a widely used approach. However this technique may be inappropriate when dealing with distributions with small isolated local modes. This talk introduces a concept of approximate spectral gap. The approximate spectral gap discounts the ill effects of small local modes, but still describes well the overall mixing behavior of the Markov chain under certain conditions. We use the idea to analyze a class of MCMC algorithms for mixtures of densities. Applications to high-dimensional Bayesian variable selection problems will also be presented.

Thursday 14:00 – 15:00 BST **Live on Zoom**

Exponential Tractability

Chair: Fred Hickernell



Peter Kritzer

Fred J. Hickernell
Illinois Institute of Technology, USA, hickernell@iit.edu

Friedrich Pillichshammer
JKU Linz, Austria, friedrich.pillichshammer@jku.at

Henryk Woźniakowski
University of Warsaw, Poland, h.wozniakowski@mimuw.edu.pl

Peter Kritzer
RICAM, Austrian Academy of Sciences, Austria,
peter.kritzer@oeaw.ac.at

We consider approximations of compact linear multivariate operators defined over Hilbert spaces. This talk focuses on studying the information complexity of the problem, which is the minimal amount of information needed to have an approximation with an error of at most $\varepsilon > 0$. The notion of tractability is used to describe how the information complexity depends on the dimension d of the problem and the error threshold ε. Much progress has been made in recent years in studying tractability of multivariate problems in different settings.

In this talk, we give an overview and compare necessary and sufficient conditions on various kinds of tractability, and in particular we present findings regarding exponential tractability. These conditions are mainly given in terms of sums of certain functions depending on the singular values of the multivariate problem.

Thursday 15:45 – 16:45 BST **Live on Zoom**

Variance Reduction using Neural Networks

Chair: Alex Keller

Thomas Muller

Thomas Müller
NVIDIA, tmueller@nvidia.com

Neural networks have proven themselves as powerful high-dimensional function approximators, but they offer little in terms of error and convergence guarantees. This is in contrast to Monte Carlo methods, which admit known statistical behavior, such as consistency or unbiasedness.

We show how neural networks can be used for variance reduction—either for importance sampling [1] or as control variates [2]—such that their approximation power is leveraged without compromising the desirable properties of Monte Carlo methods. Furthermore, we discuss gradient-based neural optimization strategies that guarantee convergence of the neural network's parameters to a locally variance-optimal configuration. The optimization is fueled from the noisy Monte Carlo samples themselves, allowing it to happen online during the Monte Carlo estimation, as opposed to requiring a pre-computation.

We will demonstrate results in the high-dimensional setting of light transport simulation and discuss applications in other domains, such as reinforcement learning, particle scattering, and Bayesian inference.

The presented work was conducted in collaboration with Markus Gross, Brian McWilliams, Fabrice Rousselle, Jan Novák, and Alex Keller.

[1] Thomas Müller, Brian Mcwilliams, Fabrice Rousselle, Markus Gross, and Jan Novák. Neural importance sampling. *ACM Transactions on Graphics*, 38(5):145:1–145:19, October 2019.

[2] Thomas Müller, Fabrice Rousselle, Jan Novák, and Alexander Keller. Neural control variates. *arXiv:2006.01524*, June 2020.

Friday 14:00 – 15:00 BST **Live on Zoom**

Optimization approaches for Bayesian inverse problems: Preconditioning integration methods in the small noise or large data limit

Chair: Aretha Teckentrup

Claudia Schillings

Claudia Schillings
Institute of Mathematics, University of Mannheim, Germany,
c.schillings@uni-mannheim.de

Björn Sprungk
Faculty of Mathematics and Computer Science, Technische Universität Bergakademie Freiberg, Germany, bjoern.sprungk@math.tu-freiberg.de

Philipp Wacker
Department of Mathematics, Friedrich-Alexander-Universität Erlangen-Nürnberg, Germany, phkwacker@gmail.com

The Bayesian approach to inverse problems provides a rigorous framework for the incorporation and quantification of uncertainties in measurements, parameters and models. We are interested in designing numerical methods which are robust w.r.t. the size of the observational noise, i.e., methods which behave well in case of concentrated posterior measures. The concentration of the posterior is a highly desirable situation in practice, since it relates to informative or large data. However, it can pose a computational challenge for numerical methods based on the prior measure. We propose to use the Laplace approximation of the posterior as the reference measure for the numerical integration and analyze the efficiency of Monte Carlo methods based on it.

[1] C. Schillings, B. Sprungk and P. Wacker. On the convergence of the Laplace approximation and noise-level-robustness of Laplace-based Monte Carlo methods for Bayesian inverse problems *Numerische Mathematik*, 2020.

Friday 15:30 – 16:30 BST **Live on Zoom**

Quasi-Monte Carlo for Density Estimation

Chair: Art Owen

Pierre L'Ecuyer

Pierre L'Ecuyer
DIRO, Université de Montréal, Canada, lecuyer@iro.umontreal.ca

Estimating the density of a continuous random variable X has been studied extensively in statistics, in the setting where n independent observations of X are given a priori and one wishes to estimate the density from that.

Popular methods include histograms and kernel density estimators. In this talk, we are interested instead in the situation where the observations are generated by Monte Carlo simulation from a model. In that case, it is possible to take advantage of variance reduction methods such as stratification, conditional Monte Carlo, and randomized quasi-Monte Carlo (RQMC), and obtain a more accurate density estimator than with standard Monte Carlo for a given computing budget. We examine various ways of doing this.

One approach is to combine directly RQMC with a kernel density estimator [1]. Another one is to adapt simulation-based derivative estimation methods such as smoothed perturbation analysis or the likelihood ratio method [2, 3, 5] to obtain an unbiased derivative estimator, and then use RQMC points with this unbiased density estimator [4]. We provide both theoretical results and numerical illustrations showing an improved convergence rate of the mean square integration error.

This talk is based on joint work with Amal Ben Abdellah from Université de Montréal, Florian Puchhammer from the Basque Center for Applied Mathematics, and Art B. Owen from Stanford University.

[1] A. Ben Abdellah, P. L'Ecuyer, A. Owen, and F. Puchhammer. Density estimation by randomized quasi-Monte Carlo. Manuscript, http://arxiv.org/abs/1807.06133, 2019.

[2] M. Fu and J.-Q. Hu. *Conditional Monte Carlo: Gradient Estimation and Optimization Applications*. Kluwer Academic, Boston, 1997.

[3] P. L'Ecuyer. A unified view of the IPA, SF, and LR gradient estimation techniques. *Management Science*, 36(11):1364–1383, 1990.

[4] P. L'Ecuyer, F. Puchhammer, and A. Ben Abdellah. Monte Carlo and quasi-Monte Carlo density estimation via conditioning. Manuscript, http://arxiv.org/abs/1906.04607, 2019.

[5] L. Lei, Y. Peng, M. C. Fu, and J.-Q. Hu. Applications of generalized likelihood ratio method to distribution sensitivities and steady-state simulation. *Discrete Event Dynamic Systems*, 28(1):109–125, 2018.

Abstracts of Minisymposia

Minisymposium

Stein's Method in Computational Statistics

Organizer(s): Chris Oates, Leah South, François-Xavier Briol, Andrew Duncan

Chair: Chris Oates

Stein's method is a technique from probability theory used to bound the distance between probability measures using differential operators adapted to the distributions of interest. In recent years it was realised that Stein's method enables computable discrepancy measures to be designed for use with un-normalised distributions in the Bayesian statistical context. Recent applications of Stein discrepancies include measuring the performance of Markov chain Monte Carlo algorithms, the development of optimisation-based sampling methods, construction of control variates for Monte Carlo variance reduction, parameter inference for generative models and goodness-of-fit testing. This minisymposium will present a tour of recent research in this active field.

Online **Youtube**

Recasting Sampling as Optimization via Stein's Method

Speaker: Chris Oates

Chris Oates
School of Mathematics, Statistics & Physics, Newcastle University, UK, chris.oates@ncl.ac.uk

There is a recent trend in computational statistics to move away from sampling methods and towards optimization methods for posterior approximation. These include discrepancy minimization, gradient flows and control functionals - all of which have the potential to deliver faster convergence than a Monte Carlo method. In this talk we will provide a basic introduction to some of these algorithms, such as [1, 2, 3], and then we will attempt to unify these emergent research themes in the context of Stein's method.

[1] W.Y. Chen, L. Mackey, J. Gorham, F.-X. Briol, C.J. Oates. Stein Points. In *Proceedings of the 35th International Conference on Machine Learning*, 2018.

[2] W.Y. Chen, A. Barp, F.-X. Briol, J. Gorham, M. Girolami, L. Mackey, C.J. Oates. Stein Point Markov Chain Monte Carlo. In *Proceedings of the 36th International Conference on Machine Learning*, 2019.

[3] M. Riabiz, W.Y. Chen, J. Cockayne, P. Swietach, S.A. Niederer, L. Mackey, C.J. Oates. Optimal Thinning of MCMC Output. *arXiv preprint*, 2005.03952, 2020.

Online **Youtube**

Monte Carlo Variance Reduction Using Stein Operators

Speaker: Leah South

Leah South
School of Mathematical Sciences, Queensland University of Technology, Australia, l1.south@qut.edu.au

Toni Karvonen
Alan Turing Institute, UK, tkarvonen@turing.ac.uk

Chris Drovandi
School of Mathematical Sciences, Queensland University of Technology, Australia, c.drovandi@qut.edu.au

Antonietta Mira
Faculty of Economics, Università della Svizzera italiana, Switzerland, antonietta.mira@usi.ch

Chris Nemeth
Department of Mathematics and Statistics, Lancaster University, UK, c.nemeth@lancaster.ac.uk

Mark Girolami
Department of Engineering, University of Cambridge, UK, mag92@eng.cam.ac.uk

Chris Oates
School of Mathematics, Statistics & Physics, Newcastle University, UK, chris.oates@ncl.ac.uk

This talk will focus on two new methods for estimating posterior expectations when the derivatives of the log posterior are available. The proposed methods are in a class of estimators that use Stein operators to generate control variates or control functionals. The first method applies regularisation to improve the performance of popular Stein-based control variates for high-dimensional Monte Carlo integration. The second method, referred to as semi-exact control functionals (SECF), is based on control functionals and Sard's approach to numerical integration. The use of Sard's approach ensures that our control functionals are exact on all polynomials up to a fixed degree in the Bernstein-von-Mises limit. Several Bayesian inference examples will be used to illustrate the potential for reduction in mean square error. If time permits, I will also briefly describe some benefits and challenges of Stein-based control variates in the unbiased Markov chain Monte Carlo setting.

Online **Youtube**

Learning to Reduce Variance Using Stochastic Gradient Descent

Speaker: François-Xavier Briol

Shijing Si
Department of Electrical and Computer Engineering, Duke University, US, shijing.si@duke.edu

Chris Oates
School of Mathematics, Statistics & Physics, Newcastle University, UK, chris.oates@ncl.ac.uk

Andrew Duncan
Department of Mathematics, Imperial College London, UK, a.duncan@imperial.ac.uk

Lawrence Carin
Department of Electrical and Computer Engineering, Duke University, US, lcarin@duke.edu

François-Xavier Briol
Department of Statistical Science, University College London, UK, f.briol@ucl.ac.uk

Control variates are a popular method for variance reduction of Monte Carlo or MCMC estimators of intractable integrals. Rather than considering improved approximations of the probability distribution, the focus is on finding an alternative function with the same integral value, but for which the variance is significantly reduced. In this talk, we will demonstrate how to automatically construct control variates through

the use of Stein's method, by creating classes of zero-mean functions based on polynomials, kernels and neural networks. We will then introduce a framework for selecting the best control variate in this class through the use of stochastic optimisation. In comparison with existing methods, our construction leads to orders of magnitude speed-ups and also allows for the use of more expressive classes of functions.

Online **Youtube**

On the Geometry of Stein Variational Gradient Descent

Speaker: Andrew Duncan

Andrew Duncan
Department of Mathematics, Imperial College London, UK, a.duncan@imperial.ac.uk

Nikolas Nüsken
Institute of Mathematics, University of Potsdam, Germany, nuesken@uni-potsdam.de

Lukasz Szpruch
School of Mathematics, University of Edinburgh, UK, l.szpruch@ed.ac.uk

Bayesian inference problems require sampling or approximating high-dimensional probability distributions. The focus of this talk is on the recently introduced Stein variational gradient descent methodology, a class of algorithms that rely on iterated steepest descent steps with respect to a reproducing kernel Hilbert space norm. This construction leads to interacting particle systems, the mean-field limit of which is a gradient flow on the space of probability distributions equipped with a certain geometrical structure. We leverage this viewpoint to shed some light on the convergence properties of the algorithm, in particular addressing the problem of choosing a suitable positive definite kernel function. Our analysis leads us to considering certain singular kernels with adjusted tails.

Minisymposium

Hierarchical Methods for Variance Reduction

Organizer(s): Chiheb Ben Hammouda, Nadhir Ben Rached, Raúl Tempone

Chair: Chiheb Ben Hammouda

Variance reduction methods play a crucial role in improving the complexity of the Monte Carlo (MC) methods. In this mini-symposium, we focus on advanced topics related to hierarchical-type variance reduction techniques, in particular, improved multilevel Monte Carlo (MLMC) for non-regular diffusion problems and continuous-time Markov chains, and splitting and subset approximations for rare event simulations. We are also interested in the connection of MLMC methods with importance sampling and regularization techniques.

Online **Youtube**

Importance Sampling for a Robust and Efficient Multilevel Monte Carlo Estimator for Stochastic Reaction Networks

Speaker: Chiheb Ben Hammouda

Chiheb Ben Hammouda
King Abdullah University of Science and Technology (KAUST), chiheb.benhammouda@kaust.edu.sa

Nadhir Ben Rached
RWTH Aachen University, Aachen, Germany, benrached@uq.rwth-aachen.de

Raúl Tempone
RWTH Aachen University & King Abdullah University of Science and Technology (KAUST), tempone@uq.rwth-aachen.de

The multilevel Monte Carlo (MLMC) method for continuous-time Markov chains, first introduced by Anderson and Higham (SIAM Multiscal Model. Simul. 10(1), 2012), is a highly efficient simulation technique that can be used to estimate various statistical quantities for stochastic reaction networks (SRNs), in particular for stochastic biological systems. Unfortunately, the robustness and performance of the multilevel method can be affected by the high kurtosis, a phenomenon observed at the deep levels of MLMC, which leads to inaccurate estimates of the sample variance. In this work, we address cases where the high-kurtosis phenomenon is due to *catastrophic coupling* (characteristic of pure jump processes where coupled consecutive paths are identical in most of the simulations, while differences only appear in a tiny proportion) and introduce a pathwise-dependent importance sampling (IS) technique that improves the robustness and efficiency of the multilevel method. Our theoretical results, along with the conducted numerical experiments, demonstrate that our proposed method significantly reduces the kurtosis of the deep levels of MLMC, and also improves the strong convergence rate from $\beta = 1$ for the standard case (without IS), to $\beta = 1 + \delta$, where $0 < \delta < 1$ is a user-selected parameter in our IS algorithm. Due to the complexity theorem of MLMC, and given a pre-selected tolerance, TOL, this results in an improvement of the complexity from $\mathcal{O}\left(\text{TOL}^{-2}\log(\text{TOL})^2\right)$ in the standard case to $\mathcal{O}\left(\text{TOL}^{-2}\right)$, which is the optimal complexity of the MLMC estimator. We achieve all these improvements with

a negligible additional cost since our IS algorithm is only applied a few times across each simulated path.

Online **Youtube**

Dynamic splitting method for rare events simulation

Speaker: Nadhir Ben Rached

Nadhir Ben Rached
RWTH Aachen University, Aachen, Germany, benrached@uq.rwth-aachen.de

Daniel MacKinlay
University of New South Wales, Sydney, Australia, d.mackinlay@unsw.edu.au

Zdravko Botev
University of New South Wales, Sydney, Australia., botev@unsw.edu.au

Raúl Tempone
RWTH Aachen University & King Abdullah University of Science and Technology,
tempone@uq.rwth-aachen.de

Mohamed-Slim Alouini
King Abdullah University of Science and Technology, Thuwal, Saudi Arabia, slim.alouini@kaust.edu.sa

We propose a unified rare-event estimator based on the multilevel splitting algorithm. In its original form, the splitting algorithm cannot be applied to time-independent problems, because splitting requires an underlying continuous-time Markov process whose trajectories can be split. We embed the time-independent problem within a continuous-time Markov process, so that the target static distribution corresponds to the distribution of the Markov process at a given time instant. To illustrate the large scope of applicability of the proposed approach, we apply it to the problem of estimating the cumulative distribution function (CDF) of sums of random variables (RVs), the CDF of partial sums of ordered RVs, the CDF of ratios of RVs, and the CDF of weighted sums of Poisson RVs. We investigate the computational efficiency of the proposed estimator via a number of simulation studies and find that it compares favorably with existing estimators.

Online **Youtube**

Combining numerical smoothing with multilevel Monte Carlo for efficient option pricing and density estimation

Speaker: Raúl Tempone

Christian Bayer
Weierstrass Institute for Applied Analysis and Stochastics (WIAS), Berlin, Germany,

Chiheb Ben Hammouda
King Abdullah University of Science and Technology (KAUST), chiheb.benhammouda@kaust.edu.sa

Raúl Tempone
RWTH Aachen University & King Abdullah University of Science and Technology (KAUST),
tempone@uq.rwth-aachen.de

When approximating the expectation of a functional of a certain stochastic process, the robustness and performance of multilevel Monte Carlo (MLMC) method, may be highly deteriorated by the low regularity of the integrand with respect to the input parameters. To overcome this issue, a smoothing procedure is needed to uncover the available regularity and improve the performance of the MLMC estimator. In this work, we consider cases where we cannot perform an analytic smoothing. Thus, we introduce a novel numerical smoothing technique based on root-finding combined with a one dimensional integration with respect to a single well-chosen variable. Our study is motivated by option pricing problems and our main focus is on dynamics where a discretization of the asset price is needed. Through our analysis and numerical experiments, we demonstrate how numerical smoothing significantly reduces the kurtosis at the deep levels of MLMC, and also improves the strong convergence rate, when using Euler scheme. Due to the complexity theorem of MLMC, and given a pre-selected tolerance, TOL, this results in an improvement of the complexity from $\mathcal{O}\left(\mathrm{TOL}^{-2.5}\right)$ in the standard case to $\mathcal{O}\left(\mathrm{TOL}^{-2}\log(\mathrm{TOL})^2\right)$. Moreover, we show how our numerical smoothing combined with MLMC enables us also to estimate density functions, which standard MLMC (without smoothing) fails to achieve.

Minisymposium

Combinations of Importance Sampling and MCMC

Organizer(s): Daniel Rudolf, Björn Sprungk

Chair: Daniel Rudolf and Björn Sprungk

Markov chain Monte Carlo (MCMC) and importance sampling have both become well-established and broadly used tools for approximate integration in computational statistics and scientific computation. Despite the rather simple underlying ideas of importance sampling and the Metropolis–Hastings (MH) algorithm, respectively, there have been still many new advances in recent years concerning algorithmic as well as theoretical aspects of both methods. One of these advances is the beneficial combination of both sampling approaches. For instance, importance sampling techniques can be employed to derive an estimator based on the proposed states generated in the MH algorithm, or vice versa, use the states of the Markov chain as importance distributions. Furthermore, MCMC simulations can guide the adaption of the proposal densities in adaptive (multiple) importance sampling. The goal of this session is to discuss these recent and related developments from a theoretical as well as a practical point of view.

Online **Youtube**

Anti-tempered layered adaptive importance sampling

Speaker: Víctor Elvira

Víctor Elvira
School of Mathematics, University of Edinburgh, United Kingdom, victor.elvira@ed.ac.uk

In this work, we first introduce an adaptive importance sampler which mixes together the benefits of the importance sampling (IS) and Markov chain Monte Carlo (MCMC) approaches. The method is called *layered adaptive importance sampling* (LAIS). Different parallel MCMC chains (upper layer) provide the location parameters of the proposal probability density functions (pdfs) used in an IS method (lower layer). Then, we consider a variation of LAIS, called *anti-tempered* LAIS (AT-LAIS) where the MCMC algorithms sample from an *anti-tempered* version of the posterior distribution. We also provide an exhaustive theoretical support explaining why, in the presented technique, even an anti-tempering strategy (reducing the scaling of the posterior) can be beneficial. Numerical results also confirm the advantages of the proposed scheme.

[1] L. Martino, V. Elvira, D. Luengo, and J. Corander. Layered adaptive importance sampling. *Statistics and Computing*, 27(3): 599–623, 2017.

[2] L. Martino, V. Elvira, and D. Luengo. Anti-tempered layered adaptive importance sampling. In *2017 22nd International Conference on Digital Signal Processing (DSP)*, pages 1–5. IEEE, 2017.

Online **Youtube**

Markov Chain Importance Sampling

Speaker: Ilja Klebanov

Ingmar Schuster

Zalando Research, Berlin, Germany, ingmar.schuster@zalando.de

Ilja Klebanov
Zuse Institute Berlin, Germany, klebanov@zib.de

Markov chain (MC) algorithms are ubiquitous in machine learning and statistics and many other disciplines. Typically, these algorithms can be formulated as acceptance rejection methods. In this work we present a novel estimator applicable to these methods, dubbed Markov chain importance sampling (MCIS), which efficiently makes use of rejected proposals. For the unadjusted Langevin algorithm, it provides a novel way of correcting the discretization error. Our estimator satisfies a central limit theorem and improves on error per CPU cycle, often to a large extent. As a by-product it enables estimating the normalizing constant, an important quantity in Bayesian machine learning and statistics.

Online Youtube

On a Metropolis–Hastings importance sampling estimator

Speaker: Björn Sprungk

Daniel Rudolf
Institute for Mathmatical Stochastics, University of Göttingen, Germany,
daniel.rudolf@uni-goettingen.de

Björn Sprungk
Faculty of Mathematics and Computer Science, TU Bergakademie Freiberg, Germany,
bjoern.sprungk@math.tu-freiberg.de

A classical approach for approximating expectations of functions w.r.t. partially known distributions is to compute the average of function values along a trajectory of a Metropolis–Hastings (MH) Markov chain. A key part in the MH algorithm is a suitable acceptance/rejection of a proposed state, which ensures the correct stationary distribution of the resulting Markov chain. However, the rejection of proposals causes highly correlated samples. In particular, when a state is rejected it is not taken any further into account. In contrast to that we consider a MH importance sampling estimator which explicitly incorporates all proposed states generated by the MH algorithm. The estimator satisfies a strong law of large numbers as well as a central limit theorem, and, in addition to that, we provide an explicit mean squared error bound. Remarkably, the asymptotic variance of the MH importance sampling estimator does not involve any correlation term in contrast to its classical counterpart. Moreover, although the analyzed estimator uses the same amount of information as the classical MH estimator, it can outperform the latter in scenarios of moderate dimensions as indicated by numerical experiments.

[1] Daniel Rudolf and Björn Sprungk. On a Metropolis–Hastings importance sampling estimator. *Electron. J. Statist.*, 14 (1): 857–889, 2020.

Online Youtube

Conditional particle filters with diffuse initial distributions

Speaker: Matti Vihola

Santeri Karppinen
Department of Mathematics and Statistics, University of Jyväskylä, Finland, santeri.j.karppinen@jyu.fi

Matti Vihola
Department of Mathematics and Statistics, University of Jyväskylä, Finland, matti.s.vihola@jyu.fi

Conditional particle filters (CPFs) [1] and their backward sampling variants [4, 3] are powerful MCMC algorithms for general nonlinear and/or non-Gaussian hidden Markov model smoothing. However, CPFs can be inefficient or difficult to apply with diffuse (mildly informative or non-informative) initial distributions, which are common in statistical applications.

We discuss a simple but generally applicable auxiliary variable method [2], which can be used together with the CPF in order to perform efficient inference with diffuse initial distributions. The method only requires simulatable Markov transitions that are reversible with respect to the initial distribution, which can be improper. We focus in particular on random-walk type transitions which are reversible with respect to a uniform initial distribution (on some domain), and their on-line adaptation based on the estimated covariance and an acceptance rate heuristic.

The experimental findings demonstrate that our method works reliably with little user specification, and can be substantially better mixing than a direct particle Gibbs algorithm that treats initial states as (additional) parameters.

[1] C. Andrieu, A. Doucet and R. Holenstein. Particle Markov chain Monte Carlo methods *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 72(3):269–342, 2010.

[2] S. Karppinen and M. Vihola. Conditional particle filters with diffuse initial distributions *arXiv:2006.14877*, 2020. https://arxiv.org/abs/2006.14877

[3] F. Lindsten, M. I. Jordan and T. B. Schön. Particle Gibbs with ancestor sampling *J. Mach. Learn. Res.*, 15(1):2145–2184, 2014.

[4] N. Whiteley. Discussion on "Particle Markov chain Monte Carlo methods" *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 72(3):306–307, 2010.

Minisymposium

Deep Thought - Analysis and Application of Hierarchical Models

Organizer(s): Jonas Latz, Aretha Teckentrup

Chair: Jonas Latz

Deep learning can be viewed as the training of complex and flexible mathematical models, which have been constructed by composing simpler models. Examples range from prior measures in Bayesian inference with uncertain hyper-parameters, to structures as complicated as deep Gaussian processes consisting of hundreds of layers.

Hierarchical models pose many theoretical and practical questions, some of which will be addressed in this mini symposium: modeling and estimation of hierarchical statistical models, computational challenges and strategies, statistical learning of machine learning models.

Online **Youtube**

Convergence of Gaussian process regression with estimated hyper-parameters

Speaker: Aretha Teckentrup

Aretha Teckentrup
School of Mathematics, University of Edinburgh, Scotland, a.teckentrup@ed.ac.uk

We consider hierarchical Gaussian process regression, where hyper-parameters appearing in the mean and covariance structure of the Gaussian process emulator are a-priori unknown, and are learnt from the data, along with the posterior mean and covariance. We work in the framework of empirical Bayes, where a point estimate of the hyper-parameters is computed, using the data, and then used within the standard Gaussian process prior to posterior update. Using results from scattered data approximation, we provide a convergence analysis of the method used to learn an unknown, deterministic function.

Online **Youtube**

Unsupervised Deep Learning Approaches for Inverse Problems

Speaker: Subhadip Mukherjee

Subhadip Mukherjee
Dept. of Applied Mathematics and Theoretical Physics, University of Cambridge, UK, sm2467@cam.ac.uk

Sören Dittmer
Center for Industrial Mathematics, University of Bremen, Germany, sdittmer@math.uni-bremen.de

Zakhar Shumaylov
Dept. of Applied Mathematics and Theoretical Physics, University of Cambridge, UK, zs334@cam.ac.uk

Ozan Öktem
Department of Mathematics, KTH – Royal Institute of Technology, Sweden, ozan@kth.se

Carola-Bibiane Schönlieb
Dept. of Applied Mathematics and Theoretical Physics, University of Cambridge, UK, cbs31@cam.ac.uk

Data-driven methods for inverse problems have recently received considerable research attention due to their excellent numerical performance. Training deep learning models in such methods typically requires measurement–ground-truth pairs, which are difficult to obtain in large numbers, especially in medical imaging applications. To circumvent this problem, we develop unsupervised learning strategies that do not rely on paired data and yet achieve competitive performance. Our approaches are primarily inspired by the recently proposed unsupervised learning framework referred to as *adversarial regularization* (AR) [1]. Firstly, we develop a convex variant of AR with the two-fold objective of (i) establishing analytical convergence guarantees for the corresponding variational reconstruction problem and (ii) devising efficient and provable algorithms for computing the solution [2]. The resulting adversarial convex regularization (ACR) approach retains the practical advantage of being data-adaptive, while offering convergence guarantees due to convexity. Secondly, we address the shortcoming of computationally expensive iterative reconstruction in AR by employing an unrolled primal-dual network [3], which is trained adversarially in conjunction with a regularizer. The learned primal-dual network leads to a couple of orders of magnitude reduction in the reconstruction time as compared to iterative gradient-based methods. For performance evaluation, we consider the task of image reconstruction in computed tomography (CT) and show that the proposed approaches outperform model-based techniques, and are competitive with state-of-the-art supervised data-driven methods.

[1] S. Lunz, O. Öktem, and C.-B. Schönlieb. Adversarial Regularizers in Inverse Problems. *Advances in Neural Info. Processing Systems*, pp. 8507–8516, 2018.

[2] S. Mukherjee, S. Dittmer, Z. Shumaylov, O. Öktem, and C.-B. Schönlieb. Adversarially Learned Convex Regularizers for Inverse Problems. *preprint available on request*.

[3] J. Adler and O. Öktem. Learned Primal-Dual Reconstruction. *IEEE Transactions on Medical Imaging*, vol. 37, issue 6, pp. 1322–1332, 2018.

Online **Youtube**

Hierarchical and non-Gaussian Models for Bayesian Inversion

Speaker: Lassi Roininen

Lassi Roininen
School of Engineering Science, LUT University, Finland, lassi.roininen@lut.fi

We consider the construction of shallow-layer deep Gaussian process models for Bayesian inversion and spatial statistics. Our models are based on presenting Matérn priors with Markov property via their equivalent stochastic partial differential equation presentation in a stacked form [1]. For discretisation, we use finite difference, finite element and series expansion methods. For computing estimators we use Metropolis-within-Gibbs, elliptical slice sampling [2] and preconditioned Crank–Nicolson algorithm which is modified to work with multi-layered Gaussian fields [3]. We show via numerical experiments in interpolation, signal deconvolution and computerised X-ray tomography problems that the proposed method can offer both smoothing and edge preservation at the same time.

[1] L. Roininen, M. Girolami, S. Lasanen, and M. Markkanen, Hyperpriors for Matérn Fields with Applications in Bayesian Inversion. *Inverse Problems and Imaging*, 13(1): 1–29, 2019.

[2] K. Monterrubio-Gómez, L. Roininen, S. Wade, T. Damoulas, and M. Girolami, Posterior Inference for Sparse Hierarchical Non-stationary Models. *Computational Statistics & Data Analysis*, 148: 106954, 2020.

[3] M. Emzir, S. Lasanen, Z. Purisha, L. Roininen, and S. Särkkä, Non-stationary Multi-layered Gaussian Priors for Bayesian Inversion. *arXiv preprint*, arXiv:2006.15634, 2020.

Online **Youtube**

Fast and even faster sampling of parameterised Gaussian random fields

Speaker: Jonas Latz

Jonas Latz
Department of Applied Mathematics and Theoretical Physics, University of Cambridge, UK,
jl2160@cam.ac.uk

Gaussian random fields are popular models for spatially varying uncertainties, arising, e.g., in geotechnical engineering, hydrology, or image processing. A Gaussian random field is fully characterised by its mean and covariance operator. In more complex models these can also be partially unknown. In this case we need to handle a family of Gaussian random fields indexed with hyperparameters. Sampling for a fixed configuration of hyperparameters is already very expensive due to the nonlocal nature of many classical covariance operators. Sampling from multiple configurations increases the total computational cost severely. In this talk we employ parameterised Karhunen-Loève expansions and adaptive cross approximations for sampling. To reduce the cost we construct a reduced basis surrogate built from snapshots of Karhunen-Loève eigenvectors in the first case. In the second case, we propose a parameterised version of the adaptive cross scheme.

In numerical experiments we consider Matérn-type covariance operators with unknown correlation length and standard deviation. Here, we study the approximation accuracy of reduced basis and cross approximation. As an application we consider Bayesian inversion with an elliptic partial differential equation where the logarithm of the diffusion coefficient is a parameterised Gaussian random field. Indeed, we employ Markov chain Monte Carlo on the reduced space to generate samples from the posterior measure.

[1] D. Kressner, J. Latz, S. Massei, E. Ullmann. Certified and fast computations with shallow covariance kernels. *arXiv:2001.09187*, 2020.

[2] J. Latz, M. Eisenberger, E. Ullmann. Fast sampling of parameterised Gaussian random fields. *Comput. Methods in Appl. Mech. Engrg.*, 348: 978–1012, 2019.

Online **Youtube**

Hyperparameter Estimation in Bayesian Inverse Problems

Speaker: Tapio Helin

Matt Dunlop
Courant Institute of Mathematical Sciences, New York University, US, matt.dunlop@nyu.edu

Tapio Helin
School of Engineering Science, LUT University, Finland, tapio.helin@lut.fi

Andrew Stuart
Department of Computing and Mathematical Sciences, Caltech, US, astuart@caltech.edu

Bayesian inverse problems are often high dimensional and it can be computationally tempting to calibrate hierarchical parameters of the prior by maximum a posteriori (MAP) estimates. Whilst these are relatively cheap to compute, a key drawback is their lack of parametrization invariance. This is an especially significant issue for hierarchical priors in the non-parametric setting. In this talk we discuss the effect of the choice of parameterization when the prior distribution is conditionally Gaussian. Specifically we consider the centered parametrization, the natural parametrization in which the unknown state is solved for directly, and the non-centered parameterization, which arises by considering dimension-robust sampling algorithms.).

Online **Youtube**

Model-based multilevel Monte Carlo methods for local quantities of interest in random heterogeneous media

Speaker: Laura Scarabosio

Laura Scarabosio
Department of Mathematics, Radboud University, Netherlands, scarabos@ma.tum.de

Barbara Wohlmuth
Department of Mathematics, Technical University of Munich, Germany,

J. Tinsley Oden
Oden Institute for Computational Engineering and Sciences, UT Austin, TX, US,

Danial Faghihi
Department of Mechanical and Aerospace Engineering, University at Buffalo, NY, US,

Uncertainty quantification for fine scale models of random heterogeneous materials is computationally challenging, because, in principle, one needs to resolve the small scale variations for every realization. For some local quantities of interest, however, a good approximation for each sample can be obtained by resolving the microstructure only in some parts of the computational domain and using an upscaled model elsewhere [1]. In this talk, we show an error estimator-driven procedure that exploits this fact to construct a sequence of surrogate models for a given local quantity of interest. These models are then combined in a multilevel framework to accelerate Monte Carlo sampling [2]. Numerical experiments for steady-state heat conduction and linear elasticity on a microstructure generated via a hierarchical procedure show the effectiveness of the proposed algorithm.

[1] J. T. Oden and K. S. Vemaganti, Estimation of local modeling error and goal-oriented adaptive modeling of heterogeneous materials: I. error estimates and adaptive algorithms, Journal of Computational Physics, 164(1): 22–47, 2000.

[2] L. Scarabosio, B. Wohlmuth, J. T. Oden and D. Faghihi, Goal-oriented adaptive modeling of random heterogeneous media and model-based multilevel Monte Carlo methods, Computers & Mathematics with Applications, 78(8): 2700–2718, 2019.

Minisymposium

Probabilistic Numerical and Kernel-Based Methods (Part 1 of 2)

Organizer(s): Toni Karvonen, Jon Cockayne

Chair: Toni Karvonen

Probabilistic numerical methods (PNMs) are a class of numerical methods that use ideas from probability and statistics in their construction. Since their inception there has been a focus on the application of these methods to integration, to quantify uncertainty in the value of the integral or for other aspects of algorithm design. Most PNMs for integration are based on Gaussian process regression and as such are closely related to kernel-based interpolation and worst-case optimal approximation in the reproducing kernel Hilbert space of the covariance kernel. This two-part session will present recent advances from the literature on PNMs and kernel-based methods for integration. Part 1 focuses on the probabilistic perspective, opening with an introduction to PNMs, while kernel-based methods are the topic of Part 2, the first talk of which contains a review of the connections between the two approaches.

Online **Youtube**

Bayesian Probabilistic Numerical Methods in Integration

Speaker: Jonathan Cockayne

Jonathan Cockayne
The Alan Turing Institute, UK, jcockayne@turing.ac.uk

Probabilistic numerical methods are a class of numerical methods for solving intractable problems whose output is a probabilistic distribution. The probability here is a tool used to quantify uncertainty in the true solution of the problem given that finite computational effort was expended to obtain the solution. This talk will present an introduction to probabilistic numerical methods and discuss their use in integration. We will discuss the Bayesian approach to numerics [1] and how it is applied in quadrature problems, then present the link between probabilistic numerical methods and kernel methods to provide context for the other talks in the session.

[1] J. Cockayne, C.J. Oates, T.J. Sullivan, M. Girolami Bayesian Probabilistic Numerical Methods *SIAM Review*, 61 (4): 756–789, 2019.

Online **Youtube**

Adaptive Algorithms in Bayesian Quadrature

Speaker: Matthew Fisher

Matthew Fisher
School of Mathematics, Statistics and Physics, Newcastle University, United Kingdom,
m.fisher1@newcastle.ac.uk

Bayesian Cubature is a popular probabilistic approach to numerical integration where a stochastic process model is posited for the integrand. After conditioning on data, the integral of the process provides us with a representation

of our uncertainty for the numerical integral. Several approaches have been put forward to encode sequential adaptivity (i.e. dependence on previous integrand evaluations) into this method. However, existing proposals have been limited to either estimating the parameters of a stationary covariance model or focusing computational resources in spatial regions where large values are taken by the integrand. In contrast, classical adaptive methods are more direct and focus computational resources on regions where local error estimates are largest, thus potentially reducing the total number of integrand evaluations required to obtain a prescribed error tolerance. In this talk we will demonstrate that, unlike the case for classical non-adaptive cubature methods, there are not direct Bayesian analogues of classical adaptive cubature methods in general. Motivated by this result, we develop a novel adaptive Bayesian cubature method that demonstrates empirically similar behaviour to classical adaptive methods.

Online **Youtube**

Design of Computer Experiments based on Bayesian Quadrature

Speaker: Luc Pronzato

Luc Pronzato
Laboratoire I3S, CNRS, Université Côte d'Azur, France, Luc.Pronzato@cnrs.fr

Anatoly Zhigljavsky
School of Mathematics, Cardiff University, UK, ZhigljavskyAA@cardiff.ac.uk

A standard objective in computer experiments is to predict/interpolate the behaviour of an unknown function f on a compact domain from a few evaluations inside the domain. When little is known about the function, space-filling design is advisable: typically, points of evaluation spread out across the available space are obtained by minimizing a geometrical criterion such as the covering or packing radius, or a discrepancy criterion measuring distance to uniformity. Sequential constructions, for which design points are added one at a time, are of particular interest. Our work is motivated by recent results [2] indicating that the sequence of design points generated by a vertex-direction algorithm applied to the minimization of a convex functional of a design measure can have better space filling properties than points generated by the greedy minimization of a supermodular set function. The presentation is based on the survey [3] and builds on several recent results [1, 4, 5] that show how energy functionals can be used to measure distance to uniformity.

[1] S.B. Damelin, F.J. Hickernell, D.L. Ragozin, and X. Zeng. On energy, discrepancy and group invariant measures on measurable subsets of Euclidean space. *J. Fourier Anal. Appl.*, 16:813–839, 2010.

[2] L. Pronzato and A.A. Zhigljavsky. Measures minimizing regularized dispersion. *J. Scientific Computing*, 78(3):1550–1570, 2019.

[3] L. Pronzato and A.A. Zhigljavsky. Bayesian quadrature, energy minimization and space-filling design. *SIAM/ASA J. Uncertainty Quantification*, 2020. (to appear) arXiv preprint arXiv:1808.10722, HAL preprint hal-01864076.

[4] S. Sejdinovic, B. Sriperumbudur, A. Gretton, and K. Fukumizu. Equivalence of distance-based and RKHS-based statistics in hypothesis testing. *The Annals of Statistics*, 41(5):2263–2291, 2013.

[5] B.K. Sriperumbudur, A. Gretton, K. Fukumizu, B. Schölkopf, and G.R.G. Lanckriet. Hilbert space embeddings and metrics on probability measures. *Journal of Machine Learning Research*, 11:1517–1561, 2010.

Online Youtube

Quadrature of Bayesian Neural Networks

Speaker: Takuo Matsubara

Takuo Matsubara
School of Mathematics, Statistics and Physics, Newcastle University / The Alan Turing Institute, UK,
tmatsubara@turing.ac.uk

A mathematical theory called the *ridgelet transform* [1, 2], which has been developed in the context of harmonic analysis of two layer neural networks, enables constructing a neural network by quadrature methods. The construction via quadrature methods is advantageous not only for the priori convergence analysis of neural networks but also for applications such as a parameter initialisation. Probabilistic numerics [3] aiming to establish a better estimation and a probabilistic interpretation of numerical methods is directly applicable for this construction in order to pursue the better accuracy and efficiency. Probabilistic numerics consider numerical methods as statistical inference and hence itself has an aspect as learning machines. In this talk, we will discuss the intriguing case where probabilistic numerical methods are applied to obtain a learning algorithm.

[1] E. Candès Ridgelets: Theory and Applications *Doctoral Dissertation*, Stanford University, 1998

[2] N. Murata An integral representation of functions using three-layered networks and their approximation bounds *Neural Networks* 9 (6): 947–956, 1996

[3] P. Hennig, M. Osborne, M. Girolami Probabilistic Numerics and Uncertainty in Computations *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences* 471 (2179), 2015.

Minisymposium

Probabilistic Numerical and Kernel-Based Methods (Part 2 of 2)

Organizer(s): Toni Karvonen, Jon Cockayne

Chair: Toni Karvonen

Online **Youtube**

Kernel Methods, Gaussian Processes and Uncertainty Quantification for Bayesian Quadrature

Speaker: Toni Karvonen

Toni Karvonen
The Alan Turing Institute, UK, tkarvonen@turing.ac.uk

George Wynne
Imperial College London, UK, g.wynne18@imperial.ac.uk

Filip Tronarp
University of Tübingen, Germany, filip.tronarp@uni-tuebingen.de

Chris J. Oates
Newcastle University & The Alan Turing Institute, UK, chris.oates@ncl.ac.uk

Simo Särkkä
Aalto University, Finland, simo.sarkka@aalto.fi

This talk discusses the equivalence of kernel-based approximation and Gaussian process (GP) regression, in particular how Bayesian quadrature rules can be viewed both as conditional Gaussian distributions over integrals and worst-case optimal integration rules in the reproducing kernel Hilbert space of the covariance kernel of the GP prior. These equivalences are used to prove results about uncertainty quantification properties of Bayesian quadrature rules when a covariance scaling parameter is estimated from data using maximum likelihood. It is shown that for a variety of kernels and *fixed* and *deterministic* data-generating functions Bayesian quadrature rules can become at most "slowly" overconfident in that their conditional standard deviations decay at most with a rate $\mathcal{O}(N^{-1/2})$ (up to logarithmic factors) faster than the true integration error, where N is the number of integration points. The latter part of the talk is based on recent work by Karvonen, Wynne, Tronarp, Oates, and Särkkä [1].

[1] T. Karvonen, G. Wynne, F. Tronarp, C. J. Oates, and S. Särkkä. Maximum likelihood estimation and uncertainty quantification for Gaussian process approximation of deterministic functions. *SIAM/ASA Journal on Uncertainty Quantification*, 2020. To appear.

Online Youtube

On the Positivity of Bayesian Quadrature Weights

Speaker: Motonobu Kanagawa

Motonobu Kanagawa
Data Science Department, EURECOM, France, motonobu.kanagawa@eurecom.fr

Toni Karvonen
The Alan Turing Institute, UK, tkarvonen@turing.ac.uk

Simo Särkkä
Aalto University, Finland, simo.sarkka@aalto.fi

In this talk I discuss the properties of Bayesian quadrature weights, which strongly affect stability and robustness of the quadrature rule. Specifically, I talk about conditions that are needed to guarantee that the weights are positive. It is shown that the weights are positive in the univariate case if the design points locally minimize the posterior integral variance and the covariance kernel is totally positive (e.g., Gaussian and Hardy kernels). This suggests that gradient-based optimization of design points may be effective in constructing stable and robust Bayesian quadrature rules. Numerical experiments demonstrate that significant generalizations and improvements appear to be possible, manifesting the need for further research.

Online Youtube

Variations around kernel quadrature with DPPs

Speaker: Ayoub Belhadji

Ayoub Belhadji
Ecole Centrale de Lille, CRIStAL, ayoub.belhadji@centralelille.fr

Determinantal Point Processes (DPP) are probabilistic models of negatively dependent random variables that arise in theoretical quantum optics and random matrix theory. We study quadrature rules, for smooth functions living in a reproducing kernel Hilbert space, using random nodes that follow the distribution of a DPP [1] or a mixture of DPPs [2]. The definition of these DPPs is tailored to the RKHS so that the corresponding quadratures converge at fast rates that depend on the eigenvalues of the corresponding integration operator. This unified analysis gives new insights on the experimental design of kernel-based quadrature rules.

[1] A. Belhadji, R. Bardenet, P. Chainais Kernel quadrature using DPPs *In Advances in Neural Information Processing Systems 32*, pages 12927-12937, 2019

[2] A. Belhadji, R. Bardenet, P. Chainais Kernel interpolation with continuous volume sampling *International Conference on Machine Learning*, 2020

Online Youtube

Generation of Point Sets by Global Optimization for Kernel-Based Numerical Integration

Speaker: Ken'ichiro Tanaka

Ken'ichiro Tanaka
Department of Mathematical Informatics, Graduate School of Information Science and Technology, The University of Tokyo, Japan, kenichiro@mist.i.u-tokyo.ac.jp

We propose methods for generating nodes (point sets) for Bayesian quadrature. Finding good nodes for Bayesian quadrature has been an important problem. To address this problem, we consider the Gaussian kernel and truncate its expansion to provide tractable optimization problems generating nodes.

1. First, we begin with the 1-dimensional case ($d = 1$).
   a) In this case, we use the technique proposed in [1] generating nodes for approximating functions. The negative logarithm of the determinant of the truncated kernel matrix becomes *a logarithmic energy with an external field*, which is a convex function with respect to the nodes. The nodes given by its minimizer are called approximate Fekete points. Since this technique yields a convex optimization problem with respect to the nodes, we can generate effectively them.
   b) We use the nodes for Bayesian quadrature and observe their good properties via numerical experiments.
2. Second, we consider the higher-dimensional cases ($d \geq 2$).
   a) In these cases, we have not obtained a concise expression of the logarithmic energy as opposed to the 1-dimensional case.
   b) Therefore we directly deal with the approximated determinant given by the truncation of the Gaussian kernel in this article. By numerical experiments, we can observe that higher-dimensional approximate Fekete points are found by minimizing this *determinantal logarithmic energy*, although there is no mathematical guarantee that this is always the case. We observe similar good properties of the nodes to the 1-dimensional case.

[1] T. Karvonen, S. Särkkä, and K. Tanaka: Kernel-based interpolation at approximate Fekete points, Numerical Algorithms (2020).
https://doi.org/10.1007/s11075-020-00973-y

Minisymposium

Piecewise Deterministic Markov Chain Monte Carlo Methods and Hypocoercivity

Organizer(s): Daniel Paulin

Chair: Daniel Paulin

Non-reversible Piecewise Deterministic MCMC methods and Hypocoercivity are active research areas, with many important methodological and theoretical developments in the last few years. This mini-symposium will showcase some interesting recent developments in these areas by 4 researchers: Ben Leimkuhler, Manon Michel, George Deligiannidis and Pierre Monmarché.

Online **Youtube**

Stochastic models and numerical methods with applications to machine learning

Speaker: Benedict Leimkuhler

Benedict Leimkuhler
School of Mathematics, University of Edinburgh, UK, b.leimkuhler@ed.ac.uk

I will describe SDE-based stochastic sampling algorithms we have been exploring. Most of these schemes have been conceived in the context of molecular modelling, but they have relevance to a wide variety of statistical computations, and for the training of hierarchical models such as neural networks. In particular, I will discuss the use of adaptive thermostats and Langevin schemes to control the properties of the statistical ensemble and describe the use of these schemes within a partitioned framework applicable to hierarchical models. I will also mention the use of constraints as regularization strategies in machine learning which raises additional mathematical challenges. This talk touches on projects involving a number of collaborators, including Charlie Matthews, Timothée Pouchon, Matthias Sachs, Gabriel Stoltz, Amos Storkey and Tiffany Vlaar.

Online **Youtube**

Sampling with Piecewise Deterministic Markov Processes: breaking free from reversibility through symmetry

Speaker: Manon Michel

Manon Michel
Laboratoire de Mathématiques Blaise Pascal, CNRS, Université Clermont-Auvergne, France,
manon.michel@uca.fr

During this talk, I will discuss the main concepts and ideas that have been underlying and pushing the development of PDMP-based MCMC sampling. Since their first implementations in multiparticle systems [1], their evolution and generalization can be framed into a constant search of replacing the usually-enforced time reversibility by symmetries of the sampled probability distribution itself [1]. Recently, their implementation to Bayesian inference problems [3] have raised new challenges but also new perspectives [4].

[1] M. Michel, S. Kapfer and W. Krauth. Generalized event-chain Monte Carlo: Constructing rejection-free global-balance algorithms from infinitesimal steps. *Journal of Chemical Physics*, 140 : 054116, 2014.

[2] J. Harland, M. Michel, T. A. Kampmann and J. Kierfeld. Event-chain Monte Carlo algorithms for three- and many-particle interactions. *EPL*, 117 : 30001, 2017.

[3] A. Bouchard-Côté, S. Vollmer and A. Doucet. The Bouncy Particle Sampler: A Nonreversible Rejection-Free Markov Chain Monte Carlo Method. *Journal of the American Statistical Association*, 113(522) : 855-867, 2018.

[4] M. Michel, A. Durmus and S. Sénécal. Forward Event-Chain Monte Carlo: Fast Sampling by Randomness Control in Irreversible Markov Chains. *Journal of Computational and Graphical Statistics*, 0 (0): 1–14, 2020.

Online **Youtube**

Hypocoercivity for Randomized Hamiltonian Monte Carlo

Speaker: George Deligiannidis

George Deligiannidis
Department of Statistics, University of Oxford, UK, deligian@stats.ox.ac.uk

Daniel Paulin
School of Mathematics, University of Edinburgh, UK, dpaulin@ed.ac.uk

Alexandre Bouchard-Côté
Department of Statistics, University of British Columbia, Canada, bouchard@stat.ubc.ca

Arnaud Doucet
Department of Statistics, Oxford, UK, doucet@stats.ox.ac.uk

The Bouncy Particle Sampler is a Markov chain Monte Carlo method based on a nonreversible piecewise deterministic Markov process. In this scheme, a particle explores the state space of interest by evolving according to a linear dynamics which is altered by bouncing on the hyperplane tangent to the gradient of the negative log-target density at the arrival times of an inhomogeneous Poisson Process (PP) and by randomly perturbing its velocity at the arrival times of a homogeneous PP. Under regularity conditions, we show here that the process corresponding to the first component of the particle and its corresponding velocity converges weakly towards a Randomized Hamiltonian Monte Carlo (RHMC) process as the dimension of the ambient space goes to infinity. RHMC is another piecewise deterministic non-reversible Markov process where a Hamiltonian dynamics is altered at the arrival times of a homogeneous PP by randomly perturbing the momentum component. We then establish dimension-free convergence rates for RHMC for strongly log-concave targets with bounded Hessians using coupling ideas and hypo-coercivity techniques. We use our understanding of the mixing properties of the limiting RHMC process to choose the refreshment rate parameter of BPS. This results in significantly better performance in our simulation study than previously suggested guidelines.

Online **Youtube**

Velocity jumps: an alternative to multi-time-step integrators

Speaker: Pierre Monmarché

Pierre Monmarché
LJLL and LCT, Sorbonne Université, France, pierre.monmarche@sorbonne-universite.fr

When simulating trajectories of classical MCMC samplers like the overdamped or kinetic Langevin diffusions, most of the numerical cost is due to the computation of the gradient of the energy (or log-likelihood). In molecular dynamics (MD) applications, standard multi-time-step methods address this issue by separating short-range forces from long-range forces and using different time-steps for each (the long-range, which are more expensive, evolve at lower frequencices, and can thus be evaluated less often). Nevertheless, this worsens the numerical bias due to time discretization, and the size of the time-steps is limited by resonance problems. As an alternative to these methods, we use a kinetic sampler for which short-range forces are treated in a standard Langevin diffusion while the long-range forces are taken into account by a velocity bounce mechanism, such as in the piecewise deterministic kinetic processes that have recently drawned interest in the MCMC community. Indeed, velocity jumps do not suffer from time discretization bias as they can be sampled exactly; however, this exact simulation requires bounds on the forces, which makes the method very efficient for long-range forces (which are bounded) but much less suitable for short-range ones (which are singular). A proof of concept implementation in the MD code Tinker HP for pair interactions shows a speed-up up to a factor 4. This is a joint work with Jérémy Weisman, Louis Lagardère and Jean-Philip Piquemal [1]. We will also briefly discuss the continuous interpolation between bounces and continuous drift (joint work with Mathias Rousset and Pierre-André Zitt).

[1] Pierre Monmarché, Jérémy Weisman, Louis Lagardère, and Jean-Philip Piquemal. Velocity jump processes: An alternative to multi-timestep methods for faster and accurate molecular dynamics simulations. *The Journal of Chemical Physics*, 153, 024101, 2020.

Minisymposium

Output Analysis for Markov Chain Monte Carlo

Organizer(s): James Flegal

Chair: James Flegal

Markov chain Monte Carlo (MCMC) is a sampling-based method for estimating features of probability distributions. MCMC methods produce a serially correlated, yet representative, sample from the desired distribution. As such it can be difficult to assess when the MCMC method is producing reliable results. This session presents some fundamental methods for ensuring a reliable simulation experiment. In particular, session speakers will present multivariate output analysis techniques, new estimators for the time-average covariance matrices in Markov chain central limit theorems, convergence properties of a collapsed Gibbs sampler that we propose for Bayesian vector autoregressions with predictors, and limitations of conventional techniques for bounding the convergence rates of general state space Markov chains.

Online **Youtube**

Multivariate Output Analysis for Markov Chain Monte Carlo

Speaker: James Flegal

Dootika Vats
Department of Mathematics and Statistics, Indian Institute of Technology Kanpur, India, dootika@iitk.ac.in

James Flegal
Department of Statistics, University of California Riverside, USA, jflegal@ucr.edu

Galin Jones
School of Statistics, University of Minnesota, galin@umn.edu

Markov chain Monte Carlo produces a correlated sample which may be used for estimating expectations with respect to a target distribution. A fundamental question is: when should sampling stop so that we have good estimates of the desired quantities? The key to answering this question lies in assessing the Monte Carlo error through a multivariate Markov chain central limit theorem. The multivariate nature of this Monte Carlo error has been largely ignored in the literature. We present a multivariate framework for terminating a simulation in Markov chain Monte Carlo. We define a multivariate effective sample size, the estimation of which requires strongly consistent estimators of the covariance matrix in the Markov chain central limit theorem, a property we show for the multivariate batch means estimator. We then provide a lower bound on the number of minimum effective samples required for a desired level of precision. This lower bound does not depend on the underlying stochastic process and can be calculated a priori. This result is obtained by drawing a connection between terminating simulation via effective sample size and terminating simulation using a relative standard deviation fixed-volume sequential stopping rule, which we demonstrate is an asymptotically valid procedure. The finite-sample properties of the proposed method are demonstrated in a variety of examples.

Online Youtube

Lug Sail Lag Windows for Estimating time-average Covariance Matrices

Speaker: Dootika Vats

Dootika Vats
Department of Mathematics and Statistics, Indian Institute of Technology Kanpur, India,
dootika@iitk.ac.in

James Flegal
Department of Statistics, University of California Riverside, USA, jflegal@ucr.edu

Lag windows are commonly used in econometrics and Markov chain Monte Carlo (MCMC) to estimate time-average covariance matrices. In the presence of high correlation, estimators of this matrix almost always exhibit significant negative bias, leading to undesirable finite-sample properties. We propose a new family of lag windows specifically designed to improve finite-sample performance by offsetting this negative bias in the opposite direction. We use these lag windows in weighted batch means estimators to produce an efficient lugsail batch means estimator. We obtain the bias and variance results for these multivariate estimators. Superior finite-sample properties are illustrated in an example.

Online Youtube

Convergence Analysis of a collapsed Gibbs Sampler for Bayesian Vector Autoregressions

Speaker: Karl Oskar Ekvall

Karl Oskar Ekvall
Division of Biostatistics, Institute of Environmental Medicine, Karolinska Institute,
karl.oskar.ekvall@ki.se

Galin Jones
School of Statistics, University of Minnesota, galin@umn.edu

We discuss the convergence properties of a collapsed Gibbs sampler that we propose for Bayesian vector autoregressions with predictors. After a brief introduction, the talk focuses on what is needed for the algorithm to perform well on large datasets. Specifically, we discuss assumptions which ensure the Markov chain generated by our algorithm converges to its stationary distribution at a geometric convergence rate bounded away from one as the sample size increases.

Online Youtube

Limitations of single-step Drift and Minorization in Markov Chain Convergence ANALYSIS

Speaker: Qian Qin

Qian Qin
School of Statistics, University of Minnesota, USA, qqin@umn.edu

James P. Hobert
Department of Statistics, University of Florida, USA, jhobert@stat.ufl.edu

Over the last three decades, there has been a considerable effort within the applied probability community to develop techniques for bounding the convergence rates of general state space Markov chains. Most of these results assume the existence of drift and minorization (d&m) conditions. It has often been observed that convergence rate bounds based on single-step d&m tend to be overly conservative, especially in high-dimensional situations. We build a frame-work for studying this phenomenon. It is shown that any convergence rate bound based on a set of d&m conditions cannot do better than a certain unknown optimal bound. Strategies are designed to put bounds on the optimal bound itself, and this allows one to quantify the extent to which a d&m-based convergence rate bound can be sharp. The new theory is applied to several examples. The results strongly suggest that convergence rate bounds based on single-step d&m conditions are quite inadequate in high-dimensional settings.

Minisymposium

Scalable Markov Chain Monte Carlo Algorithms (Part 1 of 2)

Organizer(s): Deborshee Sen

Chair: Deborshee Sen

Markov chain Monte Carlo (MCMC) algorithms form the backbone of Bayesian statistics. These range from Metropolis-Hastings based algorithms to more recent piecewise deterministic Markov processes. Unfortunately, current MCMC algorithms for posterior computation are inefficient as the number of observations or the dimension of the target space increases due to worsening computational time per step and mixing rates. This has motivated a rich variety of algorithms to tackle such problems including sub-sampling based algorithms and divide-and-conquer based algorithms. This session is about scalable MCMC algorithms for big data.

Part 1 focuses more specifically on non-reversible methods including piecewise deterministic Markov processes (PDMP). Part 2 is more general.

Online **Youtube**

Adaptive Piecewise Deterministic Monte Carlo Algorithms

Speaker: Andrea Bertazzi

Andrea Bertazzi
DIAM, Delft University of Technology, the Netherlands, a.bertazzi@tudelft.nl

Joris Bierkens
DIAM, Delft University of Technology, the Netherlands, joris.bierkens@tudelft.nl

The Bouncy Particle sampler (BPS) and the Zig-Zag sampler (ZZS) are continuous time, non reversible Monte Carlo methods based on piecewise deterministic Markov processes. Numerical experiments show that the mixing properties of these samplers can deteriorate when the target density is anisotropic. In principle this issue can be tackled by applying to the state space a suitable linear transformation that removes the correlations in the target and/or rescales the axes. This requires knowledge of the covariance matrix of the target, which in applications is usually unavailable beforehand. Therefore we propose an adaptive scheme for the BPS and ZZS in which the samplers learn (parts of) the covariance matrix on-the-fly. We discuss the theoretical properties of these adaptive MCMC methods, which are ergodic when the target density satisfies reasonable growth conditions. Numerical experiments show that such adaptive schemes can lead to remarkable improvements over the standard BPS and ZZS. Finally, we discuss how subsampling techniques can be incorporated in the proposed algorithms, thus making them applicable to Bayesian analysis of big data.

[1] A. Bouchard-Côté, S.J. Vollmer and A. Doucet. The Bouncy Particle Sampler: A Nonreversible Rejection-Free Markov Chain Monte Carlo Method. *Journal of the American Statistical Association*, 113:522, 855-867

[2] J. Bierkens, P. Fearnhead and G.O. Roberts. The Zig-Zag Process and Super-Efficient Sampling for Bayesian Analysis of Big Data. *Annals of Statistics*, 47 (3): 1288–1320, 2019.

Online Youtube

Non-Reversible Parallel Tempering

Speaker: Alexandre Bouchard

Alexandre Bouchard
Department of Statistics, University of British Columbia, Canada., bouchard@stat.ubc.ca

I will present an adaptive, non-reversible Parallel Tempering (PT) allowing MCMC exploration of challenging problems such as single cell phylogenetic trees. A sharp divide emerges in the behaviour and performance of reversible versus non-reversible PT schemes: the performance of the former eventually collapses as the number of parallel cores used increases whereas non-reversible benefits from arbitrarily many available parallel cores. These theoretical results are exploited to develop an adaptive scheme approximating the optimal annealing schedule. My group is also interested in making these advanced non-reversible Monte Carlo methods easily available to data scientists. To do so, we have designed a Bayesian modelling language to perform inference over arbitrary data types using non-reversible, highly parallel algorithms.

[1] Saifuddin Syed, Alexandre Bouchard-Côté, George Deligiannidis, Arnaud Doucet, Non-Reversible Parallel Tempering: a Scalable Highly Parallel MCMC Scheme.
https://arxiv.org/pdf/1905.02939.pdf.

[2] Software: https://www.stat.ubc.ca/∼bouchard/blang/

Online Youtube

Infinite Dimensional Piecewise Deterministic Markov Processes

Speaker: Paul Dobson

Paul Dobson
Delft Institute of Applied Mathematics, Technische Universiteit Delft, Netherlands, p.dobson@tudelft.nl

Joris Bierkens
Delft Institute of Applied Mathematics, Technische Universiteit Delft, Netherlands,
joris.bierkens@tudelft.nl

Andrew Duncan
Department of Mathematics, Imperial College London, UK, a.duncan@imperial.ac.uk

Michela Ottobre
Department of Mathematics, Heriot-Watt University, UK, m.ottobre@hw.ac.uk

Recently there has been a lot of work showing that irreversible algorithms may be advantageous for MCMC. A class of irreversible algorithms which have been introduced is that of Piecewise Deterministic Markov Processes (PDMP) which includes Zig Zag sampler and Bouncy Particle sampler. For this talk we wish to explore what we can say about PDMP in infinite dimensions. We shall discuss how to construct infinite dimensional PDMP

and some properties of such processes. In particular we will investigate two examples of such processes the infinite dimensional Zig Zag sampler and the Boomerang Sampler.

Online **Youtube**

Posterior Computation with the Gibbs Zig-Zag Zampler

Speaker: Matthias Sachs

Matthias Sachs
Department of Mathematics, Duke University, U.S.A., msachs@math.duke.edu

Deborshee Sen
Department of Statistical Science, Duke University, U.S.A., deborshee.sen@duke.edu

Jianfeng Lu
Department of Mathematics, Duke University, U.S.A., jianfeng@math.duke.edu

David Dunson
Department of Statistical Science, Duke University, U.S.A., dunson@duke.edu

Markov chain Monte Carlo (MCMC) sampling algorithms have dominated the literature on posterior computation. However MCMC faces substantial hurdles in performing efficient posterior sampling for challenging Bayesian models particularly in high-dimensional and large data settings. Motivated in part by such hurdles an intriguing new class of piecewise deterministic Markov processes (PDMPs) has recently been proposed as an alternative to MCMC. One of the most popular types of PDMPs is known as the zig-zag (ZZ) sampler. Such algorithms require a computational upper bound in a Poisson thinning step with performance improving for tighter bounds. In order to facilitate scaling to larger classes of problems we propose a general class of Gibbs zig-zag (GZZ) samplers. GZZ allows parameters to be updated in blocks with ZZ applied to certain parameters and traditional MCMC style updates to others. This provides a flexible framework to combine PDMPs with the rich literature on MCMC algorithms. We prove appealing theoretical properties of GZZ and demonstrate it on posterior sampling for logistic models with shrinkage priors for high-dimensional regression and random effects.

[1] M. Sachs, D. Sen, J. Lu, D. B. Dunson, Posterior computation with the Gibbs zig-zag sampler arXiv preprint arXiv:2004.04254, 2020.

Minisymposium

Scalable Markov Chain Monte Carlo Algorithms (Part 2 of 2)

Organizer(s): Deborshee Sen

Chair: Deborshee Sen

Markov chain Monte Carlo (MCMC) algorithms form the backbone of Bayesian statistics. These range from Metropolis-Hastings based algorithms to more recent piecewise deterministic Markov processes. Unfortunately, current MCMC algorithms for posterior computation are inefficient as the number of observations or the dimension of the target space increases due to worsening computational time per step and mixing rates. This has motivated a rich variety of algorithms to tackle such problems including sub-sampling based algorithms and divide-and-conquer based algorithms. This session is about scalable MCMC algorithms for big data.

Part 1 focuses more specifically on non-reversible methods including piecewise deterministic Markov processes (PDMP). Part 2 is more general.

Online **Youtube**

Bayesian Fusion

Speaker: Gareth Roberts

Hongsheng Dai
Department of Mathematics, University of Essex, UK, hdaia@essex.ac.uk

Murray Pollock
Department of Mathematics, Newcastle University, UK, murray.pollock@newcastle.ac.uk

Gareth Roberts
Department of Statistics, University of Warwick, UK, gareth.o.roberts@warwick.ac.uk

Suppose we can readily access samples from $\pi_i(x)$, $1 \leq i \leq n$, but we wish to obtain samples from $\pi(x) = \prod_{i=1}^{n} \pi_i(x)$. The so-called Bayesian Fusion problem comes up within various areas of modern Bayesian analysis, for example in the context of big data or privacy constraints, as well as more traditional areas such as meta-analysis. Many approximate solutions to this problem have been proposed. However this talk will present an exact solution based on rejection sampling in an extended state space, where the accept/reject decision is carried out by simulating the skeleton of a suitably constructed auxiliary collection of Brownian bridges. An early version of this work is available at [1].

[1] P. Name. Paper Bayesian Fusion. http://repository.essex.ac.uk/25975/1/BayesianFusion.pdf

Online **Youtube**

Removing the Mini-Batching Error in Bayesian Inference using Adaptive Langevin Dynamics

Speaker: Inass Sekkat

Inass Sekkat
Ecole des Ponts ParisTech, inass.sekkat@enpc.fr

The computational cost of usual Monte Carlo methods for sampling a posteriori laws in Bayesian inference scales linearly with the number of data points, and one wants to reduce it to a fraction of this cost. One option is to resort to mini-batching in conjunction with unadjusted discretizations of Langevin dynamics, in which case only a random fraction of the data is used to estimate the gradient. However, this leads to an additional noise in the dynamics and hence a bias on the invariant measure which is sampled by the Markov chain. We advocate using the so-called Adaptive Langevin dynamics, which is a modification of standard inertial Langevin dynamics with a dynamical friction which automatically corrects for the increased noise arising from mini-batching. We investigate in particular the practical relevance of the assumptions underpinning Adaptive Langevin (constant covariance for the estimation of the gradient), which are not satisfied in typical models of Bayesian inference; and discuss how to extend the approach to more general situations

[1] A. Jones and B. Leimkuhler, Adaptive stochastic methods for sampling driven molecular systems. *The Journal of Chemical Physics*, 135(8):084125, August 2011.

[2] M. Welling and Y. W. Teh, Bayesian learning via stochastic gradient Langevin dynamics. *Proceedings of the 28th International Conference on International Conference on Machine Learning*, pages 681–688, USA, 2011.

Online **Youtube**

Efficient sub-sampling for stochastic gradient MCMC for hidden Markov models

Speaker: Deborshee Sen

Rihui Ou
Department of Statistical Science, Duke University, rihui.ou@duke.edu

Deborshee Sen
Department of Statistical Science, Duke University, deborshee.sen@duke.edu

Alexander Young
Department of Statistics, Harvard University, alexander_young@fas.harvard.edu

David Dunson
Department of Statistical Science, Duke University, dunson@duke.edu

Markov chain Monte Carlo (MCMC) algorithms for hidden Markov models often rely on the forward-backward sampler. This makes them computationally slow as the length of the time series increases. Sub-sampling based approaches have recently been developed for posterior inference. These make use of the mixing of the hidden Markov process to approximate the full posterior by using small chunks of the data, an idea related to stochastic gradient MCMC. In the presence of imbalanced data resulting from rare latent states, minibatches often exclude rare state data, leading to inaccurate inference and prediction/detection of rare events. In this article, we propose a targeted sub-sampling approach that over-samples the rare states when calculating the stochastic gradient of parameters associated to them. This is achieved by using an initial clustering on the entire observation sequence. This reduces the variance in gradient estimation within stochastic gradient MCMC, which leads to improved sampling efficiency. We demonstrate substantial gains in predictive and inferential accuracy on real and synthetic examples.

Minisymposium

Variance-Reduced Estimators for Expected Information Gains in Bayesian Optimal Experimental Design

Organizer(s): Joakim Beck, Raúl Tempone

Chair: Joakim Beck

This minisymposium is devoted to the efficient estimation of the information-theoretic expected information gain criterion in the context of the Bayesian design of experiments. In the past few years, an explosion of activity has focused on variance reduction techniques, including control variates and antithetic variates within a multilevel framework, and importance sampling, for improving the computational efficiency of nested/double-loop Monte Carlo methods for estimating expected information gains. This minisymposium aims to present recent theoretical and algorithmic contributions to Bayesian experimental design with the expected information gain estimated by Monte Carlo using variance reduction.

Online **Youtube**

Multilevel Double-loop Monte Carlo to Simulation-based Bayesian Optimal Experimental Design

Speaker: Joakim Beck

Joakim Beck
CEMSE, KAUST, Saudi Arabia, joakim.beck@kaust.edu.sa

Ben Mansour Dia
CPG, KFUPM, Saudi Arabia, mansourben2002@yahoo.fr

Luis Espath
Department of Mathematics, RWTH Aachen, Germany, espath@uq.rwth-aachen.de

Raúl Tempone
CEMSE, KAUST, Saudi Arabia, and Chair of Mathematics for Uncertainty Quantification, RWTH Aachen, Germany, raul.tempone@kaust.edu.sa

We consider the problem of estimating the expected information gain in Bayesian optimal experimental design for nonlinear models. In the case of nonlinear models governed by partial differential equations, we present a Multilevel Double-Loop Monte Carlo (MLDLMC) estimator [1] that is a Double-Loop Monte Carlo (DLMC) with variance reduction through generalized control variates using a multilevel construction and Laplace-based importance sampling in the inner Monte Carlo. For sufficiently large error tolerances, the MLDLMC estimator achieves a lower computational complexity than single-level DLMC with Laplace-based importance sampling [2]. The approach aims to minimize the computational work of the estimator subject to the accuracy requirement of satisfying a specified error tolerance with a high probability. We demonstrate the computational efficiency of this multilevel estimator for an electrical impedance tomography design problem where the goal is to infer the fiber orientations in a composite laminate material.

[1] Beck, J., Dia, B.M., Espath, L.F.R., and Tempone, R. Multilevel Double Loop Monte Carlo and Stochastic Collocation Methods with Importance Sampling for Bayesian Optimal Experimental Design. *International Journal for Numerical Methods in Engineering*, 121(15):3482–3503, 2020.

[2] Beck, J., Dia, B.M., Espath, L.F.R., Long, Q., and Tempone, R. Fast Bayesian experimental design: Laplace-based importance sampling for the expected information gain. *Computer Methods in Applied Mechanics and Engineering*, 334:523–553, 2018.

Online **Youtube**

Two Applications of Multilevel Monte Carlo Methods to Bayesian Experimental Designs

Speaker: Tomohiko Hironaka

Tomohiko Hironaka
School of Engineering, the University of Tokyo, Japan, hironaka-tomohiko@g.ecc.u-tokyo.ac.jp

Takashi Goda
School of Engineering, the University of Tokyo, Japan, goda@frcer.t.u-tokyo.ac.jp

Estimating the expected information gain (EIG) has been considered computationally challenging, since it is defined as a nested expectation. In our study we developed an efficient algorithm to estimate the EIG using Multilevel Monte Carlo (MLMC)[1]. MLMC can be also applied to the estimation of the gradient of the expected information gain with respect to experimental design parameters[2]. Since the proposed estimator in [2] is unbiased, it can be combined with stochastic gradient descent algorithms, and therefore can be used to search for an optimal Bayesian experimental design. In this talk we explain these two applications of MLMC. This talk is based on two works[1, 2]. [1] is a joint work with Takeru Iwamoto, [2] with Wataru Kitade.

[1] Takashi Goda, Tomohiko Hironaka, and Takeru Iwamoto. Multilevel Monte Carlo estimation of expected information gains. *Stochastic Analysis and Applications*, 38(4):581–600, 2020.

[2] Takashi Goda, Tomohiko Hironaka, and Wataru Kitade. Unbiased MLMC stochastic gradient-based optimization of Bayesian experimental designs. *arXiv preprint arXiv:2005.08414*, 2020.

Online **Youtube**

Stochastic Optimization for Bayesian Design of Experiments

Speaker: André Gustavo Carlon

André Gustavo Carlon
Computer, Electrical and Mathematical Science and Engineering Division (CEMSE), King Abdullah University of Science and Technology (KAUST), Saudi Arabia, agcarlon@gmail.com

Ben Mansour Dia
Center for Integrative Petroleum Research (CIPR), King Fahd University of Petroleum and Minerals (KFUPM), Saudi Arabia, mansourben2002@yahoo.fr

Luis Espath
Department of Mathematics, RWTH Aachen University, Germany, espath@gmail.com

Rafael Holdorf Lopez
Department of Civil Engineering, Federal University of Santa Catarina (UFSC), Brazil,
rafaelholdorf@gmail.com

Raúl Tempone
Computer, Electrical and Mathematical Science and Engineering Division (CEMSE), King Abdullah University of Science and Technology (KAUST), Saudi Arabia; Alexander von Humboldt Professor in Mathematics and Uncertainty Quantification, RWTH Aachen University, Germany, raul.tempone@kaust.edu.sa

Finding the optimal setup for an experiment is a simultaneous uncertainty quantification and stochastic optimization task. Here, we consider the general case where we do not assume linearity of the experiment model nor Gaussianity of the prior information about the parameters of interests. As a measure of the quality of experimental setups in this setting, we use the Expected Information Gain (EIG) about the parameters of interest. To estimate the EIG, we use a Monte Carlo with Laplace approximation of the posterior (MCLA) [1] and a double-loop Monte Carlo with Laplace-based importance sampling (DLMCIS) [2]. The EIG maximization is performed using stochastic gradient descent (SGD) and accelerated stochastic gradient descent (ASGD) using gradient estimators based on MCLA and DLMCIS. We validate our methodology on three numerical examples, one of which with practical application on the industry consisting of finding the optimal currents to be imposed on an electrical impedance tomography experiment. This problem has a partial differential equations model that requires the use of the finite element method for its simulation. In all the cases tested, both SGD and ASGD improved the EIG about the parameters of interest. Moreover, the biases of the optima found using the Laplace approximation are numerically evaluated and shown to be negligible in the cases tested.

[1] Q. Long, M. Scavino, R. Tempone, S. Wang. Fast estimation of expected information gains for Bayesian experimental designs based on Laplace approximations *Computer Methods in Applied Mechanics and Engineering*, 259: 24–39, 2013.

[2] J. Beck, B. M. Dia, L. F. R. Espath, Q. Long, R. Tempone. Fast Bayesian experimental design: Laplace-based importance sampling for the expected information gain, *Computer Methods in Applied Mechanics and Engineering* 334, 523–553, 2018.

Minisymposium

Random Points: Quality Criteria and Applications

Organizer(s): Michael Gnewuch, Mario Ullrich

Chair: Mario Ullrich

Since the invention of the Monte Carlo method by Metropolis, Ulam, and von Neumann, random and pseudo-random point sets play an important role in stochastic simulation, numerical integration, optimization and other areas of applied mathematics. In all these areas and in applications in the natural sciences and in computer science there are usually different requirements that "good point sets" should satisfy. Quality criteria of common interest comprise, e.g., small variance (in stochastic simulation), low discrepancy (in quasi-Monte Carlo (QMC) integration) or small dispersion (in global optimization as, e.g., hyperparameter optimization in Deep Learning). Rather new criteria in stochastic simulation and discrepancy theory are based on notions of negative dependent random variables. In this special session we want to discuss different quality criteria for random or pseudo-random point sets, the construction of good point sets, and their performance in applications.

Online **Youtube**

Lower Bounds for the Error of Quadrature Formulas for Hilbert Spaces

Speaker: Aicke Hinrichs

Aicke Hinrichs
Johannes-Kepler-Universität Linz, Austria, aicke.hinrichs@jku.at

David Krieg
Johannes-Kepler-Universität Linz, Austria, david.krieg@jku.at

Erich Novak
Friedrich-Schiller-Universität Jena, Germany, erich.novak@uni-jena.de

Jan Vybiral
Czech Technical University Prague, Czech Republic, jan.vybiral@fjfi.cvut.cz

We prove lower bounds for the worst case error of quadrature formulas that use given sample points of a certain size. We are mainly interested in optimal point sets. We also prove lower bounds that hold for most randomly selected sets. As a tool, we use a recent result (and extensions thereof) of Vybiral on the positive semi-definiteness of certain matrices related to the product theorem of Schur. The new technique also works for spaces of analytic functions where known methods based on decomposable kernels cannot be applied.

Online **Youtube**

Revisiting One-Shot-Optimization

Speaker: Laurent Meunier

Laurent Meunier
Université Paris-Dauphine and Facebook AI Research, Paris, France, laurentmeunier@fb.com

Olivier Teytaud
Facebook AI Research, Paris, France, oteytaud@fb.com

One shot optimisation, widely used in hyperparameter tuning, consists in estimating the minimum of a function f having only access to parallel evaluations $f(x_1), \ldots, f(x_\lambda)$. More precisely, all evaluations are done simultaneously; one must make a decision based solely on this batch of results. When we know the prior distribution of the optimum, a simple solution is to randomly (or quasi-randomly) draw λ points from this distribution, get the values of the function on these points and pick up the best. In this talk, we present results for one-shot optimization in continuous domains, for simple objective functions. First, contrary to intuition, one should rescale the variance of the sampling distribution (compared to the known prior) to lower the regret. Second, we show that the average of the μ best points has a lower regret than the best point.

Online **Youtube**

An Application of Faulhaber's Formula to Star-Discrepancy

Speaker: Christian Weiß

Michael Gnewuch
Universität Osnabrück, Germany, michael.gnewuch@uni-osnabrueck.de

Hendrik Pasing
Hochschule Ruhr West, Mülheim a.d. Ruhr, Germany, Hendrik.Pasing@hs-ruhrwest.de

Christian Weiß
Hochschule Ruhr West, Mülheim a.d. Ruhr, Germany, Christian.Weiss@hs-ruhrwest.de

Bracketing numbers may be regarded as a tool to discretize star-discrepancy. Unfortunately, it is very hard to calculate them explicitly for dimensions greater than one. Therefore, research has so far concentrated on giving bounds. All these bounds depend on the dimension in an exponential way. In this talk, we aim to improve this dependence by using a generalized Faulhaber's formula. This leads to interesting applications in discrepancy theory.

Online **Youtube**

The Dependence Structure of Scrambled (t, m, s)-Nets

Speaker: Jaspar Wiart

Jaspar Wiart
Johannes-Kepler-Universität Linz, Austria, jaspar.wiart@jku.at

Christiane Lemieux
University of Waterloo, Canada, clemieux@uwaterloo.ca

Gracia Y. Dong
University of Waterloo, Canada, gracia.dong@uwaterloo.ca

We study the dependence structure of scrambled digital (t, m, s)-nets and show that they have a negative lower/upper orthant dependence structure if and only if $t = 0$. This study allows us to gain a deeper understanding about the classes of functions for which the variance of estimators based on scrambled $(0, m, s)$-nets can be proved to be no larger than that of a Monte Carlo estimator.

Minisymposium

Monte Carlo Methods for Particle Systems

Organizer(s): Abdul-Lateef Haji-Ali, Raúl Tempone

Chair: Abdul-Lateef Haji-Ali

Particle systems are versatile modeling tools that are easy to build starting from simple ODEs or SDEs, but can have complicated emergent properties. One drawback of these systems is the involved computational complexity since hundreds of thousands of coupled ODEs or SDEs have to be solved at sufficient accuracy. Continuous methods are usually employed to alleviate this complexity, but since they assume an infinite number of particles, they introduce a modeling error. More recently, using Monte Carlo methods for particle systems has become more desirable. This is due to the availability of computational resources, especially parallel architectures, and recent advancement in Monte Carlo methods that exploit the properties of the underlying systems. In this minisymposium we intend to present the latest algorithmic and theoretical contributions to of Monte Carlo methods when applied to particle systems.

Online **Youtube**

Convergence of a time-stepping scheme to the free boundary in the supercooled Stefan problem

Speaker: Christoph Reisinger

Vadim Kaushansky
Department of Mathematics, University of California Los Angeles, US, vadim.kaushansky@gmail.com

Christoph Reisinger
Mathematical Institute, University of Oxford, UK, christoph.reisinger@maths.ox.ac.uk

Mykhaylo Shkolnikov
ORFE Department, Bendheim Center for Finance, and Program in Applied & Computational Mathematics, Princeton University, US, mshkolni@gmail.com

Zhuo Qun Song
Department of Mathematics, Princeton University, US, zhuoqunsong@gmail.com

The supercooled Stefan problem and its variants describe the freezing of a supercooled liquid in physics, as well as the large system limits of systemic risk models in finance and of integrate-and-fire models in neuroscience. Adopting the physics terminology, the supercooled Stefan problem is known to feature a finite-time blow-up of the freezing rate for a wide range of initial temperature distributions in the liquid. Such a blow-up can result in a discontinuity of the liquid-solid boundary. In [2], Kaushansky and Reisinger propose a time-stepping scheme, whose convergence to the liquid-solid boundary before the time of the first discontinuity is then shown. On the other hand, the recent work [1] by Delarue, Nadtochiy and Shkolnikov describes the unique physical solution to the supercooled Stefan problem for all times. In this talk, we give convergence results for the time-stepping scheme of [2] to the liquid-solid boundary in the physical solution of [1] globally in time, irrespectively of the discontinuities exhibited by the liquid-solid boundary and the freezing rate blow-ups. Moreover, we give an explicit bound on the rate of local convergence for the time-stepping scheme. We also show numerical tests to compare our theoretical results to the practically observed convergence behaviour.

[1] F. Delarue, S. Nadtochiy, and M. Shkolnikov. Global solutions to the supercooled Stefan problem with blow-ups: Regularity and Uniqueness. arXiv preprint arXiv:1902.05174.

[2] V. Kaushansky and C. Reisinger. Simulation of particle systems interacting through hitting times. *Discret. Contin. Dyn. S. – B*, 24(10), 2019.

Online **Youtube**

Multilevel ensemble Kalman filtering algorithms

Speaker: Håkon Hoel

Alexey Chernov
Institute for Mathematics, Carl von Ossietzky University Oldenburg, Germany,
alexey.chernov@uni-oldenburg.de

Håkon Hoel
Department of Mathematics, RWTH, Germany, hoel@uq.rwth-aachen.de

Kody Law
School of Mathematics, University of Manchester, United Kingdom, kody.law@manchester.ac.uk

Fabio Nobile
Mathematics Institute of Computational Science and Engineering, EPFL, Switzerland,
fabio.nobile@epfl.ch

Raul Tempone
Department of Mathematics, RWTH, Germany, tempone@uq.rwth-aachen.de

Gaukhar Shaimerdenova
CEMSE, KAUST, Saudi-Arabia, gaukhar.shaimerdenova@kaust.edu.sa

The ensemble Kalman filter (EnKF) is a Monte-Carlo-based filtering method that is often both robust and efficient, but its performance may suffer when the computational cost of simulating the forward model is high. We present recent results [1, 2, 3] on marrying the multilevel Monte Carlo method with EnKF to construct the multilevel ensemble Kalman filter (MLEnKF). MLEnKF is applicable in settings with either finite- or infinite-dimensional state space for the dynamics and discrete-time finite-dimensional observations, and it often approximates the (large-ensemble limit) mean-field filter more efficiently than standard EnKF.

[1] H. Hoel, K. JH Law, and R. Tempone. *Multilevel ensemble Kalman filtering*, SIAM J. Numer. Anal., 54(3), pp. 1813–1839 (2016).

[2] A. Chernov, H. Hoel, K. Law, F. Nobile, and R. Tempone. *Multilevel ensemble Kalman filtering for spatio-temporal processes*, ArXiv e-prints 1710.07282v2 (2020).

[3] H. Hoel, G. Shaimerdenova, and R. Tempone. *Multilevel ensemble kalman filtering with local-level kalman gains*, ArXiv e-prints 2002.00480 (2020).

Online **Youtube**

Mean-field Particle Systems and Rare Event Simulation

Speaker: Stefan Grosskinsky

Letizia Angeli
Mathematics Institute, University of Warwick, UK, L.Angeli@warwick.ac.uk

Stefan Grosskinsky
Mathematics Institute, University of Warwick, UK, S.W.Grosskinsky@warwick.ac.uk

Adam M. Johansen
Statistics Department, University of Warwick, UK, A.M.Johansen@warwick.ac.uk

Dynamic rare events of time-additive observables in Markov processes can be cast in terms of Feynman-Kac semigroups generated by a tilted version of the original generator. The McKean interpretations of those semigroups lead to non-linear Markov processes, which are numerically accessible by Monte Carlo sampling via particle approximations, i.e. ensembles of processes evolving in parallel subject to a mean-field selection interaction. We discuss several choices of McKean models and particle filters, including cloning algorithms [1, 2] which have attracted interest in the theoretical physics literature, and provide a mathematical framework for comparison based on the martingale characterization of (Feller) Markov processes. We adapt results from the sequential Monte Carlo literature [3, 4] on convergence rates and asymptotic variances of such algorithms [5], and illustrate those for stochastic lattice gases such as zero-range or exclusion processes [6].

[1] C. Giardiná, J. Kurchan and L. Peliti. Direct evaluation of large-deviation functions. *Physical Review Letters*, 96(12): 120603, 2006.

[2] V. Lecomte and J. Tailleur. A numerical approach to large deviations in continuous time. *Journal of Statistical Mechanics*, 2007(03): P03004, 2007.

[3] P. Del Moral and L. Miclo. Branching and interacting particle systems approximations of Feynman-Kac formulae with applications to non-linear filtering. In *Seminaire de probabilites XXXIV*, pages 1–145. Springer, 2000.

[4] M. Rousset. On the control of an interacting particle estimation of Schrödinger ground states. *SIAM Journal on Mathematical Analysis*, 38(3):824–844, 2006.

[5] L. Angeli, S. Grosskinsky, A.M. Johansen. Limit theorems for cloning algorithms. (submitted) arXiv:1902.00509

[6] L. Angeli, S. Grosskinsky, A.M. Johansen, A. Pizzoferrato. Rare event simulation for stochastic dynamics in continuous time. *Journal of Statistical Physics*, 176(5): 1185–1210, 2019.

Minisymposium

Stochastic Computation and Complexity

Organizer(s): Larisa Yaroslavtseva, Thomas Müller-Gronbach, Stefan Heinrich

Chair: Larisa Yaroslavtseva

The session is devoted to algorithms and complexity for

- quadrature of SDEs and SPDEs, in particular under nonstandard assumptions,
- approximation of stochastic processes

including aspects of

- lower bounds
- connections to functional analysis.

A Zoom meeting has been scheduled to discuss the presentations of the session. Please consider watching the videos on Youtube and their respective PDFs before the discussion.

Time: August 14, 10:00 CET (Amsterdam, Berlin, Rom, Stockholm, Wien)

Link: https://uni-passau.zoom.us/j/93010776362?pwd=OG5WYStlM2Y2dk5kT3lpN0hHendXZz09

Meeting-ID 930 1077 6362 and passcode 155174

Online **Youtube**

Backward Euler–Maruyama method for SDEs with multi-valued drift coefficient

Speaker: Monika Eisenmann

Mihály Kovács
Pázmány Péter Catholic University, Hungary, kovacs.mihaly@itk.ppke.hu

Raphael Kruse
Martin-Luther-Universität Halle-Wittenberg, Germany, raphael.kruse@mathematik.uni-halle.de

Stig Larsson
Chalmers University of Technology and University of Gothenburg, Sweden, stig@chalmers.se

We consider the numerical approximation of a multivalued SDE

$$\begin{cases} \mathrm{d}X(t) + f(X(t))\,\mathrm{d}t \ni b(X(t))\,\mathrm{d}t + g(X(t))\,\mathrm{d}W(t), \quad t \in (0,T], \\ X(0) = X_0, \end{cases}$$

where $f\colon \mathrm{R}^d \to 2^{\mathrm{R}^d}$ is maximal monotone, of at most polynomial growth, coercive and fulfills the condition

$$\langle f_v - f_z, z - w\rangle \leq \langle f_v - f_w, v - w\rangle,$$

for every $v, w, z \in \mathrm{R}^d$, $f_v \in f(v)$, $f_w \in f(w)$, and $f_z \in f(z)$ as proposed in [1]. Under these low regularity assumptions on the drift coefficient, we can prove well definedness of the backward Euler method as well as the strong convergence with a rate of $\frac{1}{4}$ for Lipschitz continuous b and g.

[1] R. H. Nochetto, G. Savaré, and C. Verdi. A posteriori error estimates for variable time-step discretizations of nonlinear evolution equations. *Comm. Pure Appl. Math.*, 53(5):525–589, 2000.

Online **Youtube**

Approximation of SDEs – a stochastic sewing approach

Speaker: Máté Gerencsér

Oleg Butkovsky
WIAS Berlin, Germany, oleg.butkovskiy@gmail.com

Konstantinos Dareiotis
University of Leeds, United Kingdom, k.dareiotis@leeds.ac.uk

Máté Gerencsér
TU Wien, Austria, mate.gerencser@tuwien.ac.at

We give a new take on the error analysis of approximations of stochastic differential equations (SDEs), utilizing and developing the stochastic sewing lemma of Lê (2020). As an alternative to earlier PDE-based works, this approach allows one to go beyond Markovian settings. We discuss the first results on convergence rates of the Euler-Maruyama scheme for SDEs driven by additive fractional noise and irregular drift, as well as the derivation of optimal convergence rates for SDEs driven by multiplicative standard Brownian noise and arbitrary Hölder-continuous drift.

Online **Youtube**

Sampling scheme for intractable copula function, application to the computation of tail events in factor copula model

Speaker: Emmanuel Gobet

Cyril Benezet
Center for Applied Mathematics (CMAP), Ecole polytechnique, France,
cyril.benezet@polytechnique.edu

Emmanuel Gobet
Center for Applied Mathematics (CMAP), Ecole polytechnique, France,
emmanuel.gobet@polytechnique.edu

Rodrigo Targino
School of Applied Mathematics (EMAp), Getulio Vargas Foundation (FGV), Brazil,
rodrigo.targino@fgv.br

In risk management, modeling dependency between variables X is crucial and a standard approach is the use of copula model. Say the copula model can be sampled through realization of Y having the copula function C: had the marginals of Y been known, sampling X_i would directly follow by composing with the cdf of Y_i

and the inverse cdf of X_i. In this work, the marginals of Y are not explicit: this is the case for instance in factor copula model. We design an algorithm which samples X through some empirical approximation of the cdf of the Y-marginals. We allow the sampling of Y to be provided by a MCMC sampler. We establish convergence results which rates depend on the tails of X, Y and the Lyapunov function of the MCMC sampler. We illustrate this sampling scheme with some numerical experiments coming from quantifying tail events.

Online -

Monte-Carlo Algorithms with Restricted Access to Randomness

Speaker: Stefan Heinrich

Stefan Heinrich
Department of Computer Science, University of Kaiserslautern, Germany, heinrich@informatik.uni-kl.de

We introduce a general notion of restricted Monte Carlo algorithms which generalizes previous notions like bit Monte Carlo algorithms in two ways: it includes full adaptivity and general (i.e. not only bit) restrictions. In particular, also the standard Monte Carlo setting is included, which assumes that independent uniformly distributed on $[0,1]$ random variables are available. In contrast to these, the unrestricted randomized setting assumes that arbitrary randomness can be used by the algorithm.

We show that for each restricted setting there is a computational problem that can be solved in the unrestricted randomized setting but not under the restriction.

Furthermore, we prove a general lower bound for minimal errors in the setting of Monte Carlo algorithms with finite range restrictions (which includes bit Monte Carlo) in terms of deterministic minimal errors. This extend a result from [1] to adaptive algorithms. We discuss some applications.

[1] S. Heinrich, E. Novak, and H. Pfeiffer, How many random bits do we need for Monte Carlo integration? *Monte Carlo and quasi-Monte Carlo methods 2002*, H. Niederreiter (editor), 27–49, Springer, Berlin, 2004.

Online Youtube

Weak convergence rates of semi-exact discretization schemes for the Heston model

Speaker: Andreas Neuenkirch

Annalena Mickel
Mathematical Institute and DFG Research Training Group 1953, University of Mannheim, Germany, amickel@mail.uni-mannheim.de

Andreas Neuenkirch
Mathematical Institute, University of Mannheim, Germany, aneuenki@mail.uni-mannheim.de

Inspired by the article *Weak Convergence Rate of a Time-Discrete Scheme for the Heston Stochastic Volatility Model, Chao Zheng, SIAM Journal on Numerical Analysis 2017, 55:3, 1243-1263*, we study the weak error of discretization schemes for the Heston model, which are based on exact simulation of the underlying volatility process.

Both for an Euler- and a trapezoidal-type scheme for the log-asset price we establish weak order order one for smooth payoffs without any assumptions on the Feller index of the volatility process. In our analysis, we also observe the usual tradeoff between the smoothness assumption on the payoff and the restriction on the Feller index. Moreover, we provide error expansions, which can be used to construct second order schemes via extrapolation.

We illustrate our theoretical findings by several numerical examples.

Online -

Stability and Convergence of Randomized Runge-Kutta Method Under Inexact Information

Speaker: Paweł Przybyłowicz

Tomasz Bochacik
Faculty of Applied Mathematics, AGH University of Science and Technology, Poland, bochacik@agh.edu.pl

Maciej Goćwin
Faculty of Applied Mathematics, AGH University of Science and Technology, Poland, gocwin@agh.edu.pl

Paweł M. Morkisz
Faculty of Applied Mathematics, AGH University of Science and Technology, Poland, morkiszp@agh.edu.pl

Paweł Przybyłowicz
Faculty of Applied Mathematics, AGH University of Science and Technology, Poland, pprzybyl@agh.edu.pl

We present recent results concerning approximation of ODEs under local Lipschitz condition and inexact standard information about the right-hand side functions, see [1]. We show that the randomized two-stage Runge-Kutta scheme is optimal among all randomized algorithms based on standard noisy information. In particular, we adopt model of noisy information from [3] and we extend the results obtained in [1]. Moreover, we report some properties of regions of stability defined for the optimal method.

[1] T. Bochacik, M. Goćwin, P. M. Morkisz, P. Przybyłowicz. Randomized Runge-Kutta method – stability and convergence under inexact information, submitted, see also https://arxiv.org/pdf/2006.12131.pdf

[2] R. Kruse, Y. Wu. Error analysis of randomized Runge-Kutta methods for differential equations with time-irregular coefficients, *Comput. Methods Appl. Math.*, 17 (2017), 479–498.

[3] P. M. Morkisz, P. Przybyłowicz. Randomized derivative-free Milstein algorithm for efficient approximation of solutions of SDEs under noisy information, accepted in 2020 in *J. Comput. and Appl. Math.*

Online **Youtube**

Spectral gap of slice sampling

Speaker: Daniel Rudolf

Viacheslav Natarovskii
Institute for Mathmatical Stochastics, University of Göttingen, Germany, vnataro@uni-goettingen.de

Daniel Rudolf
Institute for Mathmatical Stochastics, University of Göttingen, Germany,
daniel.rudolf@uni-goettingen.de

Björn Sprungk
Faculty of Mathematics and Computer Science, TU Bergakademie Freiberg, Germany,
bjoern.sprungk@math.tu-freiberg.de

We provide Wasserstein contraction results 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.

[1] V. Natarovskii, D. Rudolf and B. Sprungk. Quantitative spectral gap estimate andWasserstein contraction of simple slice sampling. *Ann. Appl. Probab.*, accepted for publication, 2020.

[2] K. Łatuszyński and D. Rudolf. Convergence of hybrid slice sampling via spectral gap. arXiv preprint arXiv:1409.2709, 2014.

Online **Youtube**

A fully data-driven approach to minimizing CVaR for portfolio of assets via SGLD with discontinuous updating

Speaker: Sotirios Sabanis

Sotirios Sabanis
Mathematics, University of Edinburgh, Scotland, UK
Alan Turing Institute, London, UK, s.sabanis@ed.ac.uk

Ying Zhang
Mathematics, University of Edinburgh, Scotland, UK, ying.zhang@ed.ac.uk

A new approach in stochastic optimization via the use of stochastic gradient Langevin dynamics (SGLD) algorithms, which is a variant of stochastic gradient decent (SGD) methods, allows us to efficiently approximate global minimizers of possibly complicated, high-dimensional landscapes. With this in mind, we extend here the non-asymptotic analysis of SGLD to the case of discontinuous stochastic gradients. We are thus able to provide theoretical guarantees for the algorithm's convergence in (standard) Wasserstein distances for both convex and non-convex objective functions. We also provide explicit upper estimates of the expected excess risk associated with the approximation of global minimizers of these objective functions.

All these findings allow us to devise and present a fully data-driven approach for the optimal allocation of weights for the minimization of CVaR of portfolio of assets with complete theoretical guarantees for its performance. Numerical results illustrate our main findings.

[1] Sabanis, Sotirios and Zhang, Ying. A fully data-driven approach to minimizing CVaR for portfolio of assets via SGLD with discontinuous updating. *ArXiv*, 2020.

Online Youtube

Multi-dimensional Avikainen's estimate

Speaker: Dai Taguchi

Dai Taguchi
Okayama University, Japan, dai.taguchi.dai@gmail.com

Akihiro Tanaka
Osaka University, Japan, tnkaki2000@gmail.com

Tomooki Yuasa
Ritsumeikan University, Japan, to-yuasa@fc.ritsumei.ac.jp

Let X be a real-valued random variable with bounded density p_X with respect to Lebesgue measure. Then Avikainen proved [1] that for any real-valued random variable $\widehat{X}$, function of bounded variation $f : \mathbf{R} \to \mathbf{R}$ and $p, q \in [1, \infty)$, it holds that

$$\mathbf{E}\left[\left|f(X) - f(\widehat{X})\right|^q\right] \leq 3^{q+1} V(f)^q \left(\sup_{x \in \mathbf{R}} p_X(x)\right)^{\frac{p}{p+1}} \mathbf{E}\left[\left|X - \widehat{X}\right|^p\right]^{\frac{1}{p+1}},$$

where $V(f)$ is the total variation of f. In this talk, we will talk about the multi-dimensional analogue of this estimate ([2]). We will apply this estimate to the numerical analysis on irregular functions of stochastic differential equations based on the Euler–Maruyama scheme and the multilevel Monte Carlo method, and L^2-time regularity of forward–backward stochastic differential equations.

[1] Avikainen, R. On irregular functionals of SDEs and the Euler scheme. *Finance Stoch.* 13(3): 381–401, 2009.

[2] Taguchi, D., Tanaka, A. and Yuasa, T. Multi-dimensional Avikainen's estimates. *arXiv:2005.03219*.

Online Youtube

Semi-implicit Taylor schemes for stiff rough differential equations

Speaker: Yue Wu

Raphael Kruse
Institute for Mathematics, Martin-Luther-University Halle-Wittenberg, Halle, Germany,
raphael.kruse@mathematik.uni-halle.de

Sebastian Riedel
Institut für Mathematik, Technische Universität Berlin, Germany and Weierstraß-Institut, Berlin, Germany,
riedel@math.tu-berlin.de

Yue Wu
Mathematical Institute, University of Oxford, Oxford, UK and Alan Turning Institute, London, UK,
yue.wu@maths.ox.ac.uk

We study a class of semi-implicit Taylor-type numerical methods that are easy to implement and designed to solve multidimensional stochastic differential equations driven by a general rough noise, e.g. a fractional Brownian motion. In the multiplicative noise case, the equation is understood as a rough differential equation in the sense of T. Lyons. We focus on equations for which the drift coefficient may be unbounded and satisfies a one-sided Lipschitz condition only. We prove well-posedness of the methods, provide a full analysis, and deduce their convergence rate. Numerical experiments show that our schemes are particularly useful in the case of stiff rough stochastic differential equations driven by a fractional Brownian motion.

Online **Youtube**

On sharp lower error bounds for strong approximation of SDEs with discontinuous drift coefficient

Speaker: Larisa Yaroslavtseva

Thomas Müller-Gronbach
Faculty of Computer Science and Mathematics, University of Passau, Germany,
thomas.mueller-gronbach@uni-passau.de

Larisa Yaroslavtseva
Faculty of Mathematics and Economics, University of Ulm, Germany, larisa.yaroslavtseva@uni-ulm.de

In the past decade, an intensive study of strong approximation of stochastic differential equations (SDEs) with a drift coefficient that has discontinuities in space has begun. All investigations carried out so far study the performance of classical numerical methods for such equations or present new numerical methods and provide corresponding upper error bounds. In the majority of these results it is assumed that the drift coefficient satisfies piecewise regularity conditions and that the diffusion coefficient is globally Lipschitz continuous and non-degenerate at the discontinuities of the drift coefficient. Under this type of assumption the best L_p-error rate obtained so far for approximation of scalar equations at the final time is $3/4$ in terms of the number of evaluations of the underlying driving Brownian motion, see [1], [2]. In this talk we show that for a huge class of scalar SDEs of this type the error rate $3/4$ can not be improved.

[1] T.Müller-Gronbach, L.Yaroslavtseva. A strong order $3/4$ method for SDEs with discontinuous drift coefficient. *arXiv:1904.09178*, 2019.

[2] A.Neuenkirch, M.Szolgyenyi. The Euler-Maruyama scheme for SDEs with irregular drift: Convergence rates via reduction to a quadrature problem. *arXiv:1904.07784*, 2019.

Abstracts of Special Sessions

Special Session

MCQMC and Machine Learning

Chair: First Last

Online **Youtube**

Mean dimension of ridge functions

Speaker: Art Owen

Christopher Hoyt
Institute for Computational Mathematics, Stanford University, USA, crhoyt@stanford.edu

Art Owen
Department of Statistics, Stanford University, USA, owen@stanford.edu

Quasi-Monte Carlo (QMC) can greatly improve on plain Monte Carlo when these three properties hold: **a)** the integrand is dominated by low dimensional parts, **b)** those parts are regular enough to benefit from QMC sampling, and **c)** the QMC points have low discrepancy in their low dimensional coordinate projections. Condition **c** is a usual property of QMC points, and condition **b** is also frequently obtained [2] for low dimensional ANOVA components, even of nonsmooth integrands. This talk examines point **a** using ridge functions.

A ridge function in R^d takes the form $f(x) = g(\Theta^T x)$ where $\Theta \in R^{d\times r}$ is an orthonormal matrix. Paul Constantine's work [1] on active subspaces shows that many functions appearing in the physical sciences and engineering are well approximated by ridge functions with quite small r. Using the notion of 'mean dimension', $\nu(f)$, from an ANOVA decomposition of f, we study when ridge functions are nearly a superposition of functions of only a few of their inputs at a time. For $x \sim N(0, I_d)$ we find [4] that the mean dimension of f remains bounded as the nominal dimension $d \to \infty$ when $g : R^r \to R$ is Lipschitz continuous. For discontinuous g, we show examples where $\nu(f)$ grows proportionally to $\sqrt{d}$. Pre-integration [3] of just one component of x, from such discontinuous ridge functions can yield mean dimension $O(1)$ instead of $O(\sqrt{d})$, if the importance of that one component is bounded away from zero as $d \to \infty$ [4]. Work in progress finds low mean dimension for some popular radial basis functions too.

[1] P. G. Constantine, *Active subspaces: Emerging ideas for dimension reduction in parameter studies*, SIAM, Philadelphia, 2015.

[2] M. Griebel, F. Y. Kuo, and I. H. Sloan, *The smoothing effect of integration in R^d and the ANOVA decomposition*, Mathematics of Computation, 82 (2013), pp. 383–400.

[3] A. Griewank, F. Y. Kuo, H. Leövey, and I. H. Sloan, *High dimensional integration of kinks and jumps–Smoothing by preintegration*, Journal of Computational and Applied Mathematics, 344 (2018), pp. 259–274.

[4] C. R. Hoyt and A. B. Owen Mean dimension of ridge functions. *arXiv*:1907.01942 (to appear in SINUM)

Online Youtube

Quasi-Monte Carlo for Multivariate Distributions via Generative Neural Networks

Speaker: Avinash Prasad

Marius Hofert
Department of Statistics and Actuarial Science, University of Waterloo, Canada,
marius.hofert@uwaterloo.ca

Avinash Prasad
Department of Statistics and Actuarial Science, University of Waterloo, Canada, a2prasad@uwaterloo.ca

Mu Zhu
Department of Statistics and Actuarial Science, University of Waterloo, Canada, mu.zhu@uwaterloo.ca

In this talk, generative moment matching networks (GMMNs) are introduced as quasi-random variate generators (QRVGs) for multivariate models with any underlying copula in order to estimate expectations with variance reduction. So far, QRVGs for multivariate distributions required a careful design, exploiting specific properties (such as conditional distributions) of the implied parametric copula or the underlying quasi-Monte Carlo (QMC) point set, and were only tractable for a small number of models. Utilizing GMMNs allows one to construct QRVGs for a much larger variety of multivariate distributions without such restrictions — including for empirical distributions with dependence structures not adequately captured by parametric copulas. Once trained with a pseudo-random sample or an empirical data set, these neural networks only require a multivariate standard uniform randomized QMC point set as input and are thus fast in estimating expectations of interest under dependence with variance reduction. Numerical examples are considered to demonstrate the approach, including applications inspired by risk management practice.

Online Youtube

How to Train Your Samples?

Speaker: Gurprit Singh

Gurprit Singh
Max Planck Institute for Informatics, Saarbrücken, gsingh@mpi-inf.mpg.de

Samples are the basic building block in many domains including computer graphics, computer vision and financial mathematics. Designing sampling patterns with desired properties/correlations can require substantial effort, both in hand-crafting coding and mathematical derivation. Retaining these properties in multiple dimensions or for a substantial number of points can be challenging and computationally expensive. Tackling these issues, we present a deep learning approach [1] to automatically generate point patterns from design goals using a convolutional neural network architecture.

We phrase pattern generation as a deep composition of weighted distance-based unstructured filters. The proposed architecture optimize over different compositions defined according to a user-provided point-correlation loss function: a small program which measures a pattern's fidelity in respect to its spatial or spectral statistics, linear or non-linear (e.g., radial) projections, or any arbitrary combination thereof. We primarily focus on stochastic sampling patterns with different Fourier characteristics (blue, green, step, stair,etc.-noise), generalize them to countless new combinations in a systematic way and leverage existing Monte Carlo error estimation formulations [2] to generate *novel* point patterns for a user-provided class of integrand functions. Ultimately, we hope that this work would encourage machine learning tools in the design and generation of novel sampling patterns (including randomized-Quasi Monte Carlo), that would make their application easier in practice and would help move forward their theoretical understanding.

†Joint work with Thomas Leimkühler, Karol Myszkowski, Hans-Peter Seidel and Tobias Ritschel

[1] Thomas Leimkühler, Gurprit Singh, Karol Myszkowski, Hans-Peter Seidel, and Tobias Ritschel. Deep Point Correlation Design. *ACM Transaction on Graphics (Proceedings of SIGGRAPH Asia 2019)*, 38(6), November 2019.

[2] Gurprit Singh, Cengiz Öztireli, Abdalla G.M. Ahmed, David Coeurjolly, Kartic Subr, Oliver Deussen, Victor Ostromoukhov, Ravi Ramamoorthi, and Wojciech Jarosz. Analysis of Sample Correlations for Monte Carlo Rendering. *Computer Graphics Forum*, 38(2), May 2019.

Online Youtube

Quasi-Monte Carlo Methods and Neural Networks

Speaker: Alexander Keller

Alexander Keller
NVIDIA, Germany, akeller@nvidia.com

Matthijs Van keirsbilck
NVIDIA, Germany, matthijsv@nvidia.com

The average human brain has about 10^{11} nerve cells, where each of them may be connected to up to 10^4 others. In contrast, artificial neural networks use fully connected sets of neurons. We therefore investigate the question whether there are algorithms for artificial neural networks that are linear in the number of neurons. Representing artificial neural networks by paths, we offer two approaches to answer this question: First, we derive an algorithm that quantizes a trained artificial neural network such that the resulting complexity is linear [1]. Second, we demonstrate that training artificial neural networks, whose connections are determined by a set of random walks, can achieve a accuracy similar to fully connected layers. Due to their structural sparsity, these networks can be trained much faster and allow for deterministic initialization. Generating the paths using quasi-Monte Carlo methods, especially the Sobol' low discrepancy sequence, leads to a new parallel hardware architecture for artificial neural networks.

[1] G. Mordido, M. Van keirsbilck, and A. Keller. Instant Quantization of Neural Networks using Monte Carlo Methods. https://arxiv.org/abs/1905.12253, 2019.

Special Session

Sampling

Chair: First Last

Online **Youtube**

New strategy for dynamical sampling from multi-modal distribution and calculating the partition-function

Speaker: DanHua ShangGuan

DanHua ShangGuan
Institute of Applied Physics and Computational Mathematics, Beijing, China, sgdh@iapcm.ac.cn

A new strategy based on the Langevin simulation is proposed to sampling from the Gibbs-Boltzmann distribution which has an unknown partition-function. This strategy is efficient for multi-modal distribution which is usually hard to get satisfactory samples. At the same time, an unbiased estimator based on the kernel density estimator is suggested to get the unknown partition-function. This strategy can also be used to realize the zero-variance importance sampling method asymptotically. Some one dimensional and multidimensional examples are utilized to test the strategy mentioned above and the results are encouraging.

Online **Youtube**

Computation and inversion of cumulative distribution functions

Speaker: Amparo Gil

Amparo Gil
Departmento de Matemática Aplicada y CC de la Computación, Universidad de Cantabria, Spain,
amparo.gil@unican.es

Javier Segura
Departmento de Matemáticas, Estadística y Computación, Universidad de Cantabria, Spain,
javier.segura@unican.es

Nico M. Temme
CWI, The Netherlands, nico.temme@cwi.nl

Some special functions are particularly relevant in Applied Probability and Statistics. For example, the incomplete gamma and beta functions are (up to normalization factors) the cumulative central gamma and beta distribution functions respectively. The corresponding noncentral distributions (the Marcum-Q function and the cumulative noncentral beta distribution function) also play a significant role in several applications, and the inversion of these cumulative distribution functions (CDFs) are useful in hypothesis testing as well as for generating random samples distributed according to the corresponding probability density functions. We describe developments in the asymptotic and numerical computation and inversion of the gamma [1, 2] and beta CDFs (central and non-central) [3, 4, 5]. The performance of the methods will be illustrated with numerical examples.

[1] A. Gil, J. Segura, N.M. Temme. Efficient and accurate algorithms for the computation and inversion of the incomplete gamma function ratios. *SIAM J. Sci. Comput.* , 34(6): A2965–A2981, 2012.

[2] A. Gil, J. Segura, N.M. Temme. The asymptotic and numerical inversion of the Marcum Q-function. *Stud. Appl. Math.*, 133(2): 257–278, 2014

[3] A. Gil, J. Segura, N.M. Temme. Efficient algorithms for the inversion of the cumulative central beta distribution. *Numer. Algorithms*, 74(1): 77–91, 2017.

[4] A. Gil, J. Segura, N.M. Temme. On the computation and inversion of the cumulative noncentral beta distribution. *Appl. Math. Comput.*, 361: 74–86, 2019.

[5] A. Gil, J. Segura, N.M. Temme. Asymptotic inversion of the binomial and negative binomial cumulative distribution functions. *Electron. Trans. Numer. Anal.*, 2020.

Online Youtube

Particle swarm sampling

Speaker: Grégoire Clarté

Grégoire Clarté
CEREMADE, Université Paris Dauphine, France, clarte@ceremade.dauphine.fr

Antoine Diez
Imperial College London, UK, antoine.diez18@imperial.ac.uk

Jean Feydy
Imperial College London, UK, j.feydy@imperial.ac.uk

Over the last decades, most extensions to the seminal Metropolis-Hastings (MH) algorithm have focused on the ergodic properties of a single Markov chain and do not induce parallel algorithms that could leverage parallel implementations. In order to benefit fully from the computational power of modern hardware (GPU, clusters, etc.), a sensible strategy is therefore to study the evolution of a population of (many) particles: by letting simultaneous samples interact with each other, we can improve the convergence speed and the mixing properties of our samplers while preserving the versatility of the MH algorithm. The convergence of such Collective Monte Carlo methods can be analysed with tools borrowed from the study of many-particle systems (mean-field limit, propagation of chaos, links with PDE etc.). Applications to sampling problems over Euclidean spaces and manifolds will be presented.

Online Youtube

Maximum Entropy Distributions with Applications to Graph Simulation

Speaker: Enrique Lelo de Larrea

Paul Glasserman
Graduate School of Business, Columbia University, USA, pg20@columbia.edu

Enrique Lelo de Larrea
Department of IE&OR, Columbia University, USA, el2805@columbia.edu

We study the problem of sampling uniformly from discrete or continuous product sets subject to linear constraints (target set). This family of problems includes sampling bipartite (weighted) graphs with given degree sequences. We compare two candidate distributions to sample from the target set. The first one is the distribution which maximizes the entropy subject to satisfying the constraints in expectation. The second one is the distribution from an exponential family that maximizes the minimum probability over the target set. Our main result gives a condition under which the maximum entropy and the max-min distributions are actually the same. For the discrete case, we also develop a sequential procedure that updates the maximum entropy distribution after some components have been sampled. This procedure sacrifices the uniformity of the samples, in exchange for always sampling a valid point in the target set. To address the loss of uniformity, we use importance sampling weights. The quality of these weights is affected by the order in which the components are simulated. We thus propose an adaptive rule for this order which appears to improve the weights' quality in practice. This talk is based on [1].

[1] Paul Glasserman and Enrique Lelo de Larrea. Maximum entropy distributions with applications to graph simulation. Working paper, Columbia University, New York, 2020.

Online Youtube

Theoretical Analysis on Visible Flaws of Xorshift128+: a Newly Proposed Pseudorandom Number Generator

Speaker: Hiroshi Haramoto

Hiroshi Haramoto
Faculty of Education, Ehime University, Japan, haramoto@ehime-u.ac.jp

Makoto Matsumoto
Department of Mathematics, Graduate School of Science, Hiroshima University, Japan,
m-mat@math.sci.hiroshima-u.ac.jp

Mutsuo Saito
Department of Mathematics, Graduate School of Science, Hiroshima University, Japan,
sai10@hiroshima-u.ac.jp

Recently introduced xorshift128+ generators [1] became one of the most popular pseudorandom number generators. For example, some of the xorshift128+ generators are selected as standard generators in Google V8 JavaScript Engine. As a result, all the browsers based on this engine (e.g., Google Chrome) use the xorshift128+ generators. The xorshift128+ generators pass a stringent test suite TestU01, but we pointed out visible flaws of those: points in the three-dimensional unit cube generated by three consecutive outputs of the generators concentrate on particular planes. Such deviation may ruin a simulation. In this talk, we explain how we reach to these experiments, and give a mathematical explanation on these phenomena. The key observation is that the bit-wise exclusive-or $x \oplus y$ for non-negative integers x and y coincides with one of the arithmetic addition or subtraction $x + y$, $x - y$ or $y - x$ with non negligible probability. As a result, for three consecutive output 64-bit integers (x, y, z) of an xorshift128+, with non-negligible probability an approximation

$$z \approx \pm 2^a x \pm y \pmod{2^{64}}$$

holds, where a is one of the three parameters of the xorshift128+.

[1] S. Vigna, "Further Scramblings of Marsaglia's Xorshift Generators," *Journal of Computational and Applied Mathematics*, 315: 175–181, 2017.

Online Youtube

Algorithms and Software for Custom Digital Net Constructions

Speaker: Pierre Marion

Pierre Marion
LPSM, Sorbonne University, France, pierre.marion@mines.org

Maxime Godin
CMAP, Ecole Polytechnique, France, maxime.godin@polytechnique.org

Pierre L'Ecuyer
DIRO, Université de Montréal, Canada, lecuyer@iro.umontreal.ca

We present LatNet Builder [2], a C++ software for constructing highly-uniform point sets for quasi-Monte Carlo and randomized quasi-Monte Carlo. It extends and replaces LatticeBuilder [1]. The software now handles digital nets in base 2, including Sobol' nets, which are among the most successful methods to construct low-discrepancy point sets, in addition to rank-1 ordinary and polynomial lattice rules. In this talk, we will demonstrate the ease-of-use and flexibility of the software for concrete examples. The search for good point sets can be customized according to the number of points, the dimension, the quality criterion (so-called figure of merit), and the exploration method. The software also covers more advanced features (embedded nets, interlaced digital nets). We also describe an efficient algorithm implemented in LatNet Builder to compute the t-values of several projections over subsets of coordinates, as well as the t-values of embedded nets obtained by taking subsets of the points, in order to use a weighted average (or some other function) of these values as a figure of merit [3].

[1] P. L'Ecuyer and D. Munger. Algorithm 958: LatticeBuilder: A General Software Tool for Constructing Rank-1 Lattice Rules. *ACM Transactions on Mathematical Software*, 42 (2): Article 15, 2016.

[2] P. L'Ecuyer, M. Godin, A. Jemel, P. Marion, D. Munger. LatNet Builder: A general software tool for constructing highly uniform point sets. Available at https://github.com/umontreal-simul/latnetbuilder, 2019

[3] P. Marion, M. Godin and P. L'Ecuyer. An algorithm to compute the t-value of a digital net and of its projections. *Journal of Computational and Applied Mathematics*, 371, 2020.

Online **Youtube**

Improving Perfect Simulation for the Strauss Process Using Stitching

Speaker: Mark Huber

Mark Huber
Department of Mathematical Sciences, Claremont McKenna College, USA, mhuber@cmc.edu

A Strauss process is a repulsive point process that adds a factor of $\gamma \in (0,1)$ for each pair of points within distance R of each other to the density of a Poisson point process of rate λ. Basic acceptance rejection (AR) can be used to sample from this distribution for finite regions, but can be very slow. For a process in a square of side length s, basic AR requires time of the form $c_1 \exp(c_2 s^2)$, where the constants depend on λ, R, and γ. We introduce here *AR Stitching*, which gives an algorithm that requires time $c_3 \exp(c_4 s)$, making the method viable for problems of reasonable size. Simultaneously, this gives a method of estimating the normalizing constant for the distribution with a much smaller variance than basic AR. There do exist other perfect simulation methods that draw exactly from this process, in particular there is a method for the $\gamma = 0$ case introduced by Jerrum and Guo [1]. For the Jerrum-Guo algorithm, AR Stitching can still be used to increase the speed of this type of algorithm. These algorithms are typically provably near linear time when $\lambda > c/R^2$ for some constant c. We will look at how c changes with the improvements, both theoretically and experimentally.

[1] M. Jerrum and H. Guo. Perfect Simulation of the Hard Disks Model by Partial Rejection Sampling. *Annales de l'Institut Henri Poincaré D (AIHPD)*, June 5, 2019

Online **Youtube**

Construction Algorithm for Polynomial Lattice Rules in Weighted Spaces

Speaker: Onyekachi Osisiogu

Adrian Ebert
RICAM, Austrian Academy of Sciences, Austria, adrian.ebert@ricam.oeaw.ac.at

Peter Krizter
RICAM, Austrian Academy of Sciences, Austria, peter.kritzer@oeaw.ac.at

Onyekachi Osisiogu
RICAM, Austrian Academy of Sciences, Austria, onyekachi.osisigu@.oeaw.ac.at

Tetiana Stepaniuk
University of Lübeck, Germany, stepaniuk.tet@gmail.com

Polynomial lattice point sets are special types of digital (t, m, s)-nets. Quasi-Monte Carlo rules using them as underlying nodes are called polynomial lattice rules. In this talk, we describe a modified component-by-component method (CBC) for polynomial lattice rules such that the quality function is independent of the smoothness parameter α. The CBC constructed polynomial lattice rules achieve almost optimal order of convergence and we show that with suitable weights conditions the error bounds can be made independent of the dimensions. Futhermore, we present fast implementation of the construction such that the construction only requires (with the help of the techniques introduced by Dirk Nuyens and Ronald Cools [2]) and analyze the computational complexity. We further show numerical experiments.

[1] A. Ebert, P. Kritzer, D. Nuyens, O. Osisiogu. *Digi-by-digit and component-by-component constructions of lattice rules for periodic functions with unknown smoothness*. ArXiv, 2019.

[2] D. Nuyens, R. Cools. *Fast component-by-component construction of rank-1 lattice rules with a non-prime number of points*. J. Complexity 22(1), 4–28, 2006.

Special Session

Multi-Level Monte Carlo

Chair: First Last

Online **Youtube**

Multilevel Monte Carlo with improved correlation for kinetic equations in the diffusive scaling

Speaker: Emil Løvbak

Emil Løvbak
Dept. of Computer Science, KU Leuven, Belgium, emil.loevbak@kuleuven.be

Bert Mortier
Dept. of Computer Science, KU Leuven, Belgium, bert.mortier@kuleuven.be

Giovanni Samaey
Dept. of Computer Science, KU Leuven, Belgium, giovanni.samaey@kuleuven.be

Stefan Vandewalle
Dept. of Computer Science, KU Leuven, Belgium, stefan.vandewalle@kuleuven.be

In many applications it is necessary to compute the time-dependent distribution of an ensemble of particles subject to transport and collision phenomena. Kinetic equations are PDEs that model such particles in a position-velocity phase space. In the low collisional regime explicit particle-based Monte Carlo methods simulate these high dimensional equations efficiently, but, as the collision rate increases, these methods suffer from severe time-step constraints. In the high collision regime, the asymptotic-preserving particle scheme presented in [1] is able to produce stable results. However, this stability comes at the cost of a bias in the computed results. The multilevel Monte Carlo method [2, 3] can be used to reduce this bias by combining simulations with large and small time steps. In [4], a multilevel Monte Carlo method was introduced which reduces this bias by combining simulations with large and small time steps. In this talk, we present an improved correlation approach that decreases the variance when bridging the gap from large time steps to time steps of the order of magnitude of the collision rate. We also show the significant speedup that is achieved by using this new approach.

[1] G. Dimarco, L. Pareschi and G. Samaey. Asymptotic-Preserving Monte Carlo methods for transport equations in the diffusive limit. *SIAM Journal on Scientific Computing*, 40: A504–A528, 2018.

[2] M.B. Giles. Multilevel Monte Carlo Path Simulation. *Operations Research*, 56 (3): 607–617, 2008.

[3] S. Heinrich Multilevel Monte Carlo Methods. *Lecture Notes in Computer Science (Multigrid Methods)* 2179: 58–67, 2001.

[4] E. Løvbak, G. Samaey, S. Vandewalle. A Multilevel Monte Carlo Asymptotic-Preserving Particle Method for Kinetic Equations in the Diffusion Limit. In B. Tuffin, P. L'Ecuyer (Eds.), *Monte Carlo and Quasi-Monte Carlo Methods 2018*, In press, Springer, 2020.

Online **Youtube**

Multilevel Monte Carlo for LIBOR Market Model

Speaker: Giray Ökten

Arun Kumar Polala
Department of Mathematics, Florida State University, USA, apolala@math.fsu.edu

Giray Ökten
Department of Mathematics, Florida State University, USA, okten@math.fsu.edu

The LIBOR market model is a popular interest rate model used for pricing interest rate derivatives like caplets, caps, and swaptions. Recently, long-dated interest rate derivatives have been popular in the interest rate derivatives market, and the practitioners typically price them using the standard Monte Carlo method. The pricing problem is computationally demanding: the dimension of the problem could be as high as 360, and prices of hundreds of financial derivatives have to be estimated using Monte Carlo quickly. To overcome this computational burden, practitioners price these derivatives using very few number of samples, typically in low hundreds. We use multilevel Monte Carlo, low-discrepancy sequences, and path-generation techniques to develop fast and accurate algorithms for pricing long-dated interest rate derivatives in the LIBOR market model framework.

Online **Youtube**

Efficient Multi-Level Monte Carlo Estimators for Risk-Averse Engineering Design

Speaker: Sundar Ganesh

Fabio Nobile
CSQI, École Polytechnique Fédérale de Lausanne, Switzerland, fabio.nobile@epfl.ch

Quentin Ayoul-Guilmard
CSQI, École Polytechnique Fédérale de Lausanne, Switzerland, quentin.ayoul-guilmard@epfl.ch

Sundar Ganesh
CSQI, École Polytechnique Fédérale de Lausanne, Switzerland, sundar.ganesh@epfl.ch

Uncertainties in wind conditions have a significant effect on the risk-averse design of civil engineering structures. These uncertainties can be quantified using risk-measures of an underlying random quantity of interest. An Optimization-Under-Uncertainty (OUU) problem can then be set up such that the optimal design minimizes this risk-measure subject to design constraints. The solution of this OUU problem using gradient-based techniques typically requires the efficient and accurate estimation of these risk-measures and their sensitivities with respect to design parameters. We propose Multi-Level Monte Carlo estimators for parametric expectations [2] from which common risk measures, such as the mean-variance risk-measure or the conditional-value-at-risk, etc. can be computed as a post-processing step. We present novel error estimators and algorithms based on which these estimators are adaptively callibrated [1]. The estimators and the simulations required to construct them are implemented in the parallelized python library XMC [1] using an efficient task scheduler [3]. Finally, we demonstrate the performance of the estimators using numerical examples inspired by civil engineering, featuring fluid-flow problems and uncertain wind profiles.

[1] Nathan Collier, Abdul-Lateef Haji-Ali, Fabio Nobile, Erik von Schwerin, and Raúl Tempone. A continuation multilevel Monte Carlo algorithm. *BIT Numerical Mathematics*, 55(2):399–432, 6 2015.

[2] Sebastian Krumscheid and Fabio Nobile. Multilevel Monte Carlo approximation of functions. *SIAM/ASA Journal on Uncertainty Quantification*, 6(3):1256–1293, 2018.

[3] Enric Tejedor, Yolanda Becerra, Guillem Alomar, Anna Queralt, Rosa M. Badia, Jordi Torres, Toni Cortes, and Jesús Labarta. PyCOMPSs: Parallel computational workflows in Python. *The International Journal of High Performance Computing Applications*, 31(1):66–82, 2017.

[1] DOI:10.5281/zenodo.3235833

Online Youtube

p-refined Multilevel Quasi-Monte Carlo for Galerkin Finite Element Methods with applications in Geotechnical Engineering

Speaker: Philippe Blondeel

Philippe Blondeel
Dep. Computer Science, KU Leuven, Belgium, philippe.blondeel@kuleuven.be

Pieterjan Robbe
Dep. Computer Science, KU Leuven, Belgium, pieterjan.robbe@kuleuven.be

Stijn François
Dep. Civil Engineering, KU Leuven, Belgium, stijn.francois@kuleuven.be

Geert Lombaert
Dep. Civil Engineering, KU Leuven, Belgium, geert.lombaert@kuleuven.be

Stefan Vandewalle
Dep. Computer Science, KU Leuven, Belgium, stefan.vandewalle@kuleuven.be

The soil's material parameters in practical geotechnical engineering problems are characterized by significant uncertainty as for example, the Young's modulus, the cohesion and the friction angle. The representation of these parameters is typically done by means of random fields. Discretization of the governing problem equations with the Galerkin Finite Element method allows one to take into account the aforementioned material uncertainty. This is accomplished by assigning discrete values resulting from the random field to the elements of the discretized domain. Computation of the stochastic responses remains very costly, even when state-of-the-art Multilevel Monte Carlo (MLMC) [2] is used. A significant cost reduction can be achieved by using a recently developed multilevel method: p-refined Multilevel Quasi-Monte Carlo (p-MLQMC). This method is based on the idea of variance reduction by employing a hierarchical discretization of the problem based on a p-refinement scheme instead of an h-refinement scheme. It is combined with a rank-1 Quasi-Monte Carlo (QMC) lattice rule, which yields faster convergence compared to the use of random Monte Carlo points. A comparison between MLMC and MLQMC for civil engineering problems can be found in [1]. We use p-MLQMC for the assessment of the stability of slopes, a problem that arises in geotechnical engineering, and typically suffers from large parameter uncertainty. The uncertainty resides in the cohesion of the soil and is represented by means of a lognormal random field resulting from a Karhunen–Loève expansion. We achieve a high computational gain for p-MLQMC with respect to MLMC.

[1] Philippe Blondeel, Pieterjan Robbe, Cédric Van hoorickx, Geert Lombaert, and Stefan Vandewalle. Multilevel sampling with Monte Carlo and Quasi-Monte Carlo methods for uncertainty quantification in structural engineering. *13th International Conference on Applications of Statistics and Probability in Civil Engineering, ICASP13, Seoul, South Korea*, 2019.

[2] Michael B. Giles. Multilevel Monte Carlo path simulation. *Operations Research*, 56(3):607–617, 2008.

Special Session

Markov Chain Monte Carlo

Chair: First Last

Online **Youtube**

Bayesian Estimation of Time-Trees

Speaker: Luiz Max Carvalho

Luiz Max Carvalho
School of Applied Mathematics (EMAp), Getúlio Vargas Foundation (FGV), Brazil, luiz.fagundes@fgv.br

Marc Suchard
Department of Biostatistics, University of California Los Angeles (UCLA), United States of America, msuchard@ucla.edu

Guy Baele
Department of Microbiology, Immunology and Transplantation, KU Leuven, Belgium, guy.baele@kuleuven.be

Andrew Rambaut
Institute of Evolutionary Biology, University of Edinburgh, United Kingdom, a.rambaut@ed.ac.uk

Bayesian phylogenetics is concerned with the problem of given data on n species/individuals, reconstructing the – rooted, binary – tree t that depicts the evolutionary relationships between them. In addition, one is interested in estimating the lengths of the $d = 2n - 2$ edges of t, henceforth denoted $\boldsymbol{b}_t$. The object $\tau = (t, \boldsymbol{b}_t)$ is then called a phylogeny. One is usually interested in computing a posterior distribution of the form

$$p(t, \boldsymbol{b}_t, \boldsymbol{\theta}|D) = \frac{f(D|t, \boldsymbol{b}_t, \boldsymbol{\theta})\pi(t, \boldsymbol{b}_t, \boldsymbol{\theta})}{\sum_{t\in\boldsymbol{T}} \int_{\boldsymbol{B}_t} \int_{\boldsymbol{\Theta}} f(D|t, \boldsymbol{b}_t, \boldsymbol{\theta})\pi(t, \boldsymbol{b}_t, \boldsymbol{\theta}) d\boldsymbol{\theta} d\boldsymbol{b}_t}\,, \tag{7.1}$$

where D is observed data and $\boldsymbol{T}$ is the set of all binary rooted trees on n. Finally, $\boldsymbol{\theta}$ is a set of parameters of interest such as substitution model parameters, migration rates, heritability coefficients, etc, in which the analyst is usually interested in as well as in the phylogeny – meaning these are not nuisance parameters. The target distribution in (7.1) is intractable for all but the simplest settings, and hence computation of expectations and other summaries needs to be approximated, usually via Markov Chain Monte Carlo (MCMC) techniques. Adaptive MCMC seeks to adjust the scale of proposals as the chain progresses in order to strike an optimal balance between conservative and bold proposals. While this is a well-studied class of algorithms, the non-standard nature of the space of phylogenetic trees makes it difficult to design efficient transition kernels. The talk will be organised as follows: we will (i) frame phylogenetics as a fundamentally statistical problem and give a brief overview of the necessary objects (e.g. cubic complexes, CAT(0) spaces); (ii) discuss the specific problem of designing adaptive candidate-generating mechanisms for phylogenetic MCMC that adhere to constraints; (iii) discuss the hSPR class of phylogeny-generating mechanisms, which jointly update tree structure and edge (branch) lengths while preserving time-precedence constraints and (iv) discuss sufficient conditions for the proposed class to lead to ergodic chains. We conclude with (v) a few preliminary empirical results.

Online Youtube

Hamiltonian Monte Carlo Applied to Inverse Problems in Imaging the Earth's Subsurface

Speaker: Florian Puchhammer

Elena Akhmatskaya
IKERBASQUE and Basque Center for Applied Mathematics, Spain, akhmatskaya@bcamath.org

Julen Alvarez-Aramberri
Earth Institute, Columbia University, USA, julen@ldeo.columbia.edu

David Pardo
IKERBASQUE and Universidad del Pais Vasco, Spain, dzubiaur@gmail.com

Florian Puchhammer
Basque Center for Applied Mathematics, Spain, fpuchhammer@bcamath.org

Tijana Radivojević
Lawrence Berkeley National Laboratory, USA, tradivojevic@lbl.gov

Wayne I. T. Uy
Courant Institute of Mathematical Sciences, NYU, USA, wayne.uy@cims.nyu.edu

The magnetotelluric (MT) method is a technique to infer the structure of the Earth's subsurface from electromagnetic measurements governed by Maxwell's equations. More precisely, receivers are placed a few centimeters below the surface which measure at various frequencies the impedance caused by the interplay of natural electromagnetic source fields with the resistivity of different materials in the ground. In contrast to many other popular methods, the MT technique uses natural sources and is non-invasive, which made it one of the most important tools in deep Earth's research. To obtain an image of the Earth's subsurface, one needs to solve the inverse MT problem, i.e., infering the resistivities of sub-surface layers from impedance measurements. In this talk we consider one-dimensional media which are composed of horizontal layers and seek solutions of the inverse MT problem by Bayesian inference with Markov chain Monte Carlo (MCMC). Traditional Metropolis random walk loses its efficiency in higher dimensions. For this problem, we therefore study the performance of more advanced samplers which typically achieve higher sampling efficiency than MCMC through the use of gradient information of the posterior distribution: Hybrid/Hamiltonian Monte Carlo (HMC) [2, 6], Generalized Hybrid Monte Carlo [4, 5], and the recently proposed HMC importance sampling (Mix and Match Hamiltonian Monte Carlo) [7]. A crucial task for these methods is the selection of an integration scheme for Hamilton's equations, as its accuracy has a direct impact on acceptance rates, and thus on sampling performance. Recently proposed adaptive splitting integrators [3, 1] proved to be a successful option for HMC methods in molecular simulation applications. We extend these ideas to a statistical setting and apply them to the inverse MT problem.

[1] E. Akhmatskaya, M. Fernández-Pendás, T. Radivojević. Adaptive splitting integrators for enhancing sampling efficiency of modified hamiltonian monte carlo methods in molecular simulation. *Langmuir*, 33(42): 11530–11542, 2017.

[2] S. Duane, A. D. Kennedy, B. J. Pendleton, D. Roweth Hybrid Monte Carlo. *Phys. Lett. B*, 195(2), 216–222, 1987.

[3] M. Fernández-Pendás, E. Akhmatskaya, J.M. Sanz-Serna. Adaptive multi-stage integrators for optimal energy conservation in molecular simulations. *Journal of Computational Physics*, 327, 434–449, 2016.

[4] A. M. Horowitz. A generalized guided Monte Carlo algorithm. *Phys. Lett. B*, 268, 247–252, 1991.

[5] A. D. Kennedy, B. J. Pendleton. Cost of the generalised hybrid Monte Carlo algorithm for free field theory. *Nucl. Phys. B*, 607, 456–510, 2001.

[6] R. M. Neil. Bayesian learning for neural networks. *Ph.D. Thesis*, Department of Computer Science, University of Toronto, 1994.

[7] R. Radivojević, E. Akhmatskaya. Modified Hamiltonian Monte Carlo for Bayesian inference. *Statistics and Computing*, 30: 377–404, 2020.

Online **Youtube**

Component-wise approximate Bayesian computation via Gibbs-like steps

Speaker: Robin J. Ryder

Grégoire Clarté
CEREMADE, Université Paris-Dauphine, Université PSL, CNRS, clarte@ceremade.dauphine.fr

Christian P. Robert
CEREMADE, Université Paris-Dauphine, Université PSL, CNRS, xian@ceremade.dauphine.fr

Robin J. Ryder
CEREMADE, Université Paris-Dauphine, Université PSL, CNRS, ryder@ceremade.dauphine.fr

Julien Stoehr
CEREMADE, Université Paris-Dauphine, Université PSL, CNRS, stoehr@ceremade.dauphine.fr

Approximate Bayesian computation methods are useful for generative models with intractable likelihoods. These methods are however sensitive to the dimension of the parameter space, requiring exponentially increasing resources as this dimension grows. To tackle this difficulty, we explore a Gibbs versionof the Approximate Bayesian computation approach that runs component-wise approximate Bayesiancomputation steps aimed at the corresponding conditional posterior distributions, and based on summarystatistics of reduced dimensions. While lacking the standard justifications for the Gibbs sampler, the resulting Markov chain is shown to converge in distribution under some partial independence conditions.The associated stationary distribution can further be shown to be close to the true posterior distributionand some hierarchical versions of the proposed mechanism enjoy a closed form limiting distribution. Experiments also demonstrate the gain in efficiency brought by the Gibbs version over the standard solution.

Online Youtube

Multifidelity Sequential Tempered Markov Chain Monte Carlo for Bayesian Inference

Speaker: Thomas Catanach

Thomas Catanach
Computing and Information Science, Sandia National Laboratories, USA, tacatan@sandia.gov

Huy Vo
Chemical and Biological Engineering, Colorado State University, USA, Huy.Vo@colostate.edu

Brian Munsky
Chemical and Biological Engineering, Colorado State University, USA, Brian.Munsky@colostate.edu

Bayesian updating for model calibration is held back by the computational cost of Markov Chain Monte Carlo (MCMC). Sequential Monte Carlo (SMC) methods have parallelized MCMC for Bayesian updating. These methods transform a sample population from the prior to the posterior using a series of annealing levels that gradually introduce the likelihood. While faster, SMC is still costly for expensive models. To speed up inference, the Multilevel Sequential^2 Monte Carlo[1] algorithm demonstrates that early annealing levels may leverage, lower accuracy, but faster, models. As the annealing factor increases, more expensive but higher accuracy models are used. The main challenge is determining when to increase model fidelity. They proposes an effective sample size statistic. Alternatively, we introduce an information-theoretic criteria that seeks to extract the most information about the posterior from each model in the hierarchy without inducing significant bias[2]. This criteria uses a limited number of full-fidelity model evaluations at each annealing level to estimate whether increasing the annealing factor with the current model fidelity gains information about the full-fidelity posterior. We demonstrate this approach on computationally expensive inference problems for single cell stochastic gene regulatory networks. We reformulate the finite state projection algorithm, a method for solving the chemical master equation, to produce a multifidelity hierarchy to be used in inference. SNL is managed and operated by NTESS under DOE NNSA contract DE-NA0003525, SAND2020-3305 A.

[1] J. Latz, I. Papaioannou, E. Ullmann. Multilevel Sequential^2 Monte Carlo for Bayesian Inverse Problems. *Journal of Computational Physics*, 368: 154–178, 2018.

[2] T. Catanach, H. Vo, B. Munsky. Bayesian Inference of Stochastic Reaction Networks using Multifidelity Sequential Tempered Markov Chain Monte Carlo. *arXiv*, 2001.01373, 2020.

Online Youtube

A Note on the Metropolis-Hastings Acceptance Probabilities for Mixture Spaces

Speaker: Tobias Siems

Tobias Siems
Ruprecht-Karls-University Heidelberg Institute for Applied Mathematics and Interdisciplinary Center for Scientific Computing (IWR), Heidelberg, Germany, tobias.siems@uni-heidelberg.de

Lisa Koeppel
Ruprecht-Karls-University Heidelberg Sektion Klinische Tropenmedizin, Heidelberg, Germany, l.koeppel@uni-heidelberg.de

This work is driven by the ubiquitous dissent over the abilities and contributions of the Metropolis-Hastings [3, 2] and reversible jump algorithm [1] within the context of trans dimensional sampling. We demystify this topic by taking a deeper look into the implementation of Metropolis-Hastings acceptance probabilities with regard to general mixture spaces. Whilst unspectacular from a theoretical point of view, mixture spaces gave rise to challenging demands concerning their effective exploration. An often applied but not extensively studied tool for transitioning between distinct spaces are translation functions. We make use of precise measure theoretic formulations in order to give an enlightening treatment of this topic in practical terms. As a result, we come across a slight generalization of the reversible jump algorithm and unveil another promising translation technique, so-called post-hoc and ad-hoc translations. We underpin our findings and compare the performances of our approaches by means of a change point example. Thereafter, in a more theoretical context, we revitalize the somewhat forgotten concept of maximal acceptance probabilities [4, 5]. This yields an important classification of post-hoc translations, and thus reversible jump instances, into two fundamental classes. These two can be used to achieve very different goals. Together with ad-hoc translations a large variety of feasible Metropolis-Hastings proposals becomes amenable not only to the experienced, but also to the average user.

[1] P. J. Green. Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika*, 82(4):711–732, 1995.

[2] W. K. Hastings. Monte Carlo Sampling Methods Using Markov Chains and Their Applications. *Biometrika*, 1970.

[3] Nicholas Metropolis, Arianna W Rosenbluth, Marshall N Rosenbluth, Augusta H Teller, and Edward Teller. Equation of state calculations by fast computing machines. *The Journal of Chemical Physics*, 21(6):1087–1092, 1953.

[4] Peter H Peskun. Optimum Monte-Carlo sampling using Markov chains. *Biometrika*, 60(3):607–612, 1973.

[5] Luke Tierney. A note on Metropolis-Hastings kernels for general state spaces. *The Annals of Applied Probability*, 8(1):1–9, 1998.

Online Youtube

A Bayesian Robust Nonlinear Multivariate Time Series Model with Autoregressive and t-Distributed Errors – A Case Study for GNSS Data

Speaker: Alexander Dorndorf

Alexander Dorndorf
Leibniz University Hannover, Germany, dorndorf@gih.uni-hannover.de

Boris Kargoll
Anhalt University of Applied Sciences, Germany, boris.kargoll@hs-anhalt.de

Jens-André Paffenholz
Clausthal University of Technology, Germany, jap19@tu-clausthal.de

Hamza Alkhatib
Leibniz University Hannover, Germany, alkhatib@gih.uni-hannover.de

The estimation of spatio-temporal model parameters and their stochastic information from measurements is one of the main tasks in geodesy and related fields dealing with spatial data. Exemplary measurement data are obtained by the global navigation satellite system (GNSS). These data can be modelled as a multivariate time series consisting of a deterministic ("functional") model describing the trend and a stochastic model of the correlated noise. The GNSS 3D coordinate time series of the x-, y- and z-components are often affected by outliers and their stochastic properties can vary significantly. The functional model of the time series (e.g., a circle in 3D) is usually nonlinear regarding the trend parameters. To deal with these characteristics, a time series model which can generally be explained as the additive combination of a multivariate, nonlinear regression model with multiple univariate, covariance-stationary autoregressive (AR) processes whose white noise components obey independent, scaled t-distributions was proposed in [1]. The parameters of that model were estimated via maximum likelihood estimation, which was achieved by means of a generalized expectation maximization (GEM) algorithm. In this paper, we extend the aforementioned model to include prior knowledge regarding various model parameters, which information is often available in practical situations. We develop an algorithm based on Bayesian statistics that provides a robust and reliable estimation of the functional parameters, the coefficients of the AR process and the parameters of the underlying t-distribution (i.e., the scale factor and the degree of freedom). We approximate the resulting posterior density using Markov chain Monte Carlo techniques consisting of Gibbs samplers and Metropolis-Hastings algorithms. An advantage of this procedure compared to the GEM algorithm, besides the capability of processing additional prior knowledge, is that the approximation of the posterior model parameters is feasible without linearisation of the functional model. Furthermore, the approximation of the variance-covariance matrix of the estimated parameters can be derived directly from the generated chains.

[1] Kargoll, B., Kermarrec, G., Korte, J. et al. Self-tuning robust adjustment within multivariate regression time series models with vector-autoregressive random errors. J Geod 94, 51 (2020). https://doi.org/10.1007/s00190-020-01376-6

Online Youtube

Implementing short-period Tausworthe generators for Markov chain quasi-Monte Carlo

Speaker: Shin Harase

Shin Harase
College of Science and Engineering, Ritsumeikan University, Japan, harase@fc.ritsumei.ac.jp

We consider the problem of estimating expectations by using Markov chain Monte Carlo methods and improving the accuracy by replacing IID uniform random points with quasi-Monte Carlo (QMC) points. Chen et al. [1] proved that Markov chain QMC remains consistent when the driving sequences are completely uniformly distributed (CUD). A sequence $u_0, u_1, u_2, \ldots \in [0,1)$ is said to be CUD if overlapping s-blocks $(u_i, u_{i+1}, \ldots, u_{i+s-1})$, $i = 0, 1, 2, \ldots$ are uniformly distributed for every dimension $s \geq 1$. To construct CUD sequences approximately, Chen et al. [2] implemented short-period Tausworthe generators (i.e., linear feedback shift register generators over the two-element field) optimized in terms of the equidistribution property, which is a coarse criterion used in the area of pseudorandom number generation. In this talk, we conduct an exhaustive search of short-period Tausworthe generators for Markov chain QMC in terms of the t-value, which is a criterion of uniformity widely used in the study of QMC methods. We describe an algorithm for finding parameters whose t-values are zero for $s = 2$ and as small as possible for $s \geq 3$, and discuss recent progress. The talk is based on the preprint [3] and some new material.

[1] S. Chen, J. Dick, and A. B. Owen. Consistency of Markov chain quasi-Monte Carlo on continuous state spaces. *Ann. Statist.,* 39 (2): 673–701, 2011.

[2] S. Chen, M. Matsumoto, T. Nishimura, and A. B. Owen. New inputs and methods for Markov chain quasi-Monte Carlo. *Monte Carlo and quasi-Monte Carlo methods 2010*: 313–327, Springer Proc. Math. Stat., 23, Springer, Heidelberg, 2012.

[3] S. Harase. A table of short-period Tausworthe generators for Markov chain quasi-Monte Carlo. *arXiv:2002.09006*, 2020.

Online Youtube

Manifold MCMC Methods for Efficient Inference in a Wide Class of Diffusion Models

Speaker: Matthew Graham

Matthew Graham
School of Mathematics, Statistics and Physics, Newcastle University, UK, matthew.graham@ncl.ac.uk

Alexandre Thiery
Department of Statistics and Applied Probability, National University of Singapore, Singapore, a.h.thiery@nus.edu.sg

Alexandros Beskos
Department of Statistical Science, University College London, UK, a.beskos@ucl.ac.uk

Bayesian inference for partially observed, nonlinear diffusion models is a challenging task that has led to the development of several important methodological advances. We propose a novel framework for inferring the posterior distribution on both a time discretization of the diffusion process and any unknown model parameters, given partial observations of the process. The set of joint configurations of the noise increments and parameters which map to diffusion paths consistent with the observations form an implicitly defined manifold. By using a constrained Hamiltonian Monte Carlo algorithm for constructing Markov kernels on embedded manifolds, we are able to perform computationally efficient inference in a wide class of partially observed diffusions. Unlike other approaches in the literature, that are often limited to specific model classes, our approach allows full generality in the choice of observation and diffusion models, including complex cases such as hypoelliptic systems with degenerate diffusion coefficients. By exploiting the Markovian structure of diffusions, we propose a variant of the approach with a complexity that scales linearly in the time resolution of the discretization and quasi-linearly in the number of observation times.

Online **Youtube**

The Barker Proposal: Robust, Gradient-based MCMC

Speaker: Samuel Livingstone

Samuel Livingsotne
Department of Statistical Science, UCL, UK, samuel.livingstone@ucl.ac.uk

Giacomo Zanella
Department of Decision Sciences, BIDSA and IGIER, Bocconi University, Italy,
giacomo.zanella@unibocconi.it

We consider the issue of robustness of MCMC algorithms with respect to heterogeneity in the target and their sensitivity to tuning. We show that the spectral gap of the Markov chains induced by classical gradient-based MCMC schemes (e.g. Langevin and Hamiltonian Monte Carlo) decays exponentially fast in the degree of mismatch between the scales of the proposal and target distributions, while for the random walk Metropolis (RWM) the decay is polynomial. We propose a novel and simple to implement gradient-based MCMC algorithm, inspired by the classical Barker accept-reject rule, with improved robustness properties. With some theory and simulation studies we illustrate how this type of robustness is particularly beneficial in the context of adaptive MCMC, giving examples in which the new scheme gives orders of magnitude improvements in efficiency over state-of-the-art alternatives.

Special Session

Differential Equations

Chair: First Last

Online **Youtube**

Construction of QMC Finite Element Methods for Elliptic PDEs with Random Coefficients by a Reduced CBC Construction Algorithm

Speaker: Adrian Ebert

Adrian Ebert
Johann Radon Institute for Computational and Applied Mathematics (RICAM), Austria,
adrian.ebert@ricam.oeaw.ac.at

Peter Kritzer
Johann Radon Institute for Computational and Applied Mathematics (RICAM), Austria,
peter.kritzer@ricam.oeaw.ac.at

Dirk Nuyens
Department of Computer Science, KU Leuven, Belgium, dirk.nuyens@cs.kuleuven.be

In this talk, we consider the use of quasi-Monte Carlo (QMC) methods to approximate expectations of a linear functional applied to the solution of an affine-parametric, elliptic PDE with random diffusion coefficient. The sensitivity w.r.t. the parameters is often stated in terms of product-and-order-dependent (POD) weights. The (offline) fast component-by-component (CBC) construction of an N-point QMC method making use POD weights leads to a cost of $\mathcal{O}(sN\ln N + s^2N)$ with s the parameter truncation dimension. When s is large this cost is prohibitive. As an alternative Gantner, Herrmann and Schwab [2] introduced an analysis resulting in product weights to reduce the construction cost to $\mathcal{O}(sN\ln N)$. Here, we present how the reduced CBC method [1] can be used for POD weights to reduce the cost to $\mathcal{O}(\sum_{j=1}^{\min\{s,s^*\}}(m-w_j+j)\,b^{m-w_j})$, where $N=b^m$ with prime b, $w_1\le\cdots\le w_s$ are nonnegative integers and s^* can be chosen much smaller than s depending on the regularity of the random field expansion as such making it possible to use the POD weights directly. We show a total error estimate for using randomly shifted lattice rules constructed by the reduced CBC method. Numerical experiments additionally demonstrate the effectiveness of the reduced CBC construction.

[1] J. Dick, P. Kritzer, G. Leobacher, F. Pillichshammer. A reduced fast component-by-component construction of lattice points for integration in weighted spaces with fast decreasing weights. *Journal of Computational Applied Mathematics*, 276:1–15, 2015.

[2] R.N. Gantner, L. Herrmann, Ch. Schwab. Quasi-Monte Carlo integration for affine-parametric, elliptic PDEs: Local supports and product weights. *SIAM Journal on Numerical Analysis*, 56:111–135, 2018.

Online Youtube

Infinite Parallel Plates Algorithm

Speaker: Chi-Ok Hwang

Chi-Ok Hwang
Division of Liberal Arts and Sciences, Gwangju Institute of Science and Technology, South Korea,
chwang@gist.ac.kr

Mingyo Kim
School of Materials Science and Engineering, Gwangju Institute of Science and Technology, South Korea,
kmg129@gist.ac.kr

Based on isomorphism between the electrostatic Poisson problem and the corresponding diffusion motion expectation of the first-passage, we developed a new parallel plates algorithm. using a series solution for the induced-charge density on the parallel plates by a charge at the center between the parallel plates, combined with the acceptance-rejection sampling method. We verified that the proposed parallel plates algorithm was significantly more efficient than the current "Walks-on-Spheres" algorithm.

Online Youtube

Stochastic Runge–Kutta methods based on Markov jump processes and applications to deterministic systems of differential equations

Speaker: Flavius Guiaș

Flavius Guiaș
Dortmund University of Applied Sciences and Arts, Germany, flavius.guias@fh-dortmund.de

In this presentation we will discuss the stochastic Picard–Runge–Kutta solvers for systems of ordinary differential equations introduced in [1] and [2] and propose a further improvement of them. These solvers deliver best results for systems with a sparse incidence matrix, for example in the case of finite–difference discretization of partial differential equations. Their principle is based on the connection between the infinitesimal generators of Markov jump processes and corresponding differential equations. The step function $\tilde{X}$ computed by simulating the jump processes can serve as a predictor which is further improved by suitable correction steps. Given the improved approximation $X^*(t)$ at time t, we compute the corresponding approximation at time $t+h$ by an integral scheme of the form $X^*(t+h) = X^*(t) + \int_t^{t+h} Q(s)\,ds$. For computing the improved approximations $X^*(\cdot)$ we take for the integrand Q a polynomial which interpolates some equidistant intermediate values of $F(\tilde{X}(\cdot))$ between t and $t+h$. According to [1] and [2], by using an exact quadrature formula in order to compute the integral above, we can employ in this stochastic context the same principle of the deterministic Runge–Kutta method in order to compute a better approximation for the solution of the equation. In addition to the mentioned works, we will show how the precision of the intermediate values can be improved by using first a stochastic counterpart of a second-order Runge–Kutta method. The final result is a high precision scheme with several layers, which starts from the crude approximation delivered by the standard jump process, and based on this data it computes several steps in which the approximations are successively refined.

[1] F. Guiaș, P. Eremeev. Improving the stochastic direct simulation method with applications to evolution partial differential equations. *Appl. Math. Comput.*, 289 : 353–370, 2016.

[2] F. Guiaș, Stochastic Picard–Runge–Kutta Solvers for Large Systems of Autonomous Ordinary Differential Equations. *2017 Fourth International Conference on Mathematics and Computers in Sciences and in Industry (MCSI)*, DOI 10.1109/MCSI.2017.55 : 298–302, 2017.

Special Session

Variance Reduction

Chair: First Last

Online **Youtube**

Polynomial Chaos as a Control Variate Method

Speaker: Jamie Fox

Jamie Fox
Department of Mathematics, Florida State University, USA, jfox@math.fsu.edu

Giray Ökten
Department of Mathematics, Florida State University, USA, okten@math.fsu.edu

We investigate the use of polynomial chaos as a control variate method for Monte Carlo simulation. First, we analyze the mean square error of the control variate estimator when the coefficients of the polynomial chaos approximation are obtained from Monte Carlo simulation. For a fixed computational cost, our objective amounts to determining the optimal allocation of cost between approximating the polynomial chaos coefficients, and estimating the expectation of the function with the control variate estimator. Next, we examine the effects of setting the control to a polynomial chaos approximation of a reduced model, formed by freezing the insignificant inputs of the original model using global sensitivity analysis. This approach not only reduces the computational cost associated with the calibration and evaluation of the polynomial, but can also reduce the mean square error of the control variate estimator. Finally, we propose two different polynomial-based control variate methods for the calculation of Sobol' sensitivity indices. Then, we compare our methods numerically against crude Monte Carlo and an approach that only uses polynomial chaos.

Online Youtube

On the Use of Global Sensitivity Analysis for Optimisation of Numerical Schemes for Evaluation of Option Prices and Greeks

Speaker: Emanouil Atanassov

Emanouil Atanassov
IICT-BAS, Bulgaria, emanouil@parallel.bas.bg

Sergei Kucherenko
Imperial College London, UK, s.kucherenko@imperial.ac.uk

Sofiya Ivanovska
IICT-BAS, Bulgaria, sofia@parallel.bas.bg

The quasi-Monte Carlo methods are an established tool for approximating option prices as well as the relevant Greeks. The Global Sensitivity Analysis can be applied to such algorithms in order to quantify the interactions of the different variables that are sampled from low-discrepancy sequences. The numerical schemes that are developed for the more complex option pricing models that include stochastic volatility, e.g., the Heston model, have a more involved structure of the interactions between variables. Nevertheless, through the use of GPGPU computing, we are able to estimate efficiently the Sobol' sensitivity coefficients and to evaluate the effective and average dimensions for these schemes. Based on this information, we undertook an optimization of the specific way the Sobol' sequences are used in the numerical schemes as well as their directional numbers. Through extensive numerical experiments we demonstrate the viability of our approach for speeding-up the computation of option prices and Greeks.

Online Youtube

Randomized Dimension Reduction for Monte Carlo Simulations

Speaker: Nabil Kahalé

Nabil Kahalé
Department of Finance, ESCP Business School, France, nkahale@escp.eu

We present a new unbiased algorithm that estimates the expected value of $f(U)$ via Monte Carlo simulation, where U is a vector of d independent random variables, and f is a function of d variables. We assume that f does not depend equally on all its arguments. Under certain conditions we prove that, for the same computational cost, the variance of our estimator is lower than the variance of the standard Monte Carlo estimator by a factor of order d. Our method can be used to obtain a low-variance unbiased estimator for the expectation of a function of the state of a Markov chain at a given time-step. We study applications to volatility forecasting and time-varying queues. Numerical experiments show that our algorithm dramatically improves upon the standard Monte Carlo method for large values of d, and is highly resilient to discontinuities.

Online **Youtube**

Rare-Event Analysis and Simulation of Queues with Time-Varying Rates

Speaker: Ad Ridder

Ad Ridder
Department of EOR, Vrije Universiteit Amsterdam, Netherlands, ad.ridder@vu.nl

In this paper we study rare-event probabilities in Markovian $M_t/M_t/1$ queues with time-varying arrival rates (nonhomogeneous Poisson arrivals) and time-varying service rates. As running examples, we take periodic arrival rate functions $\alpha(t) = \lambda + \delta(t)$, $t \geq 0$, with sinusoidal or sawtooth deviation functions

$$\delta(t) = A\sin(2\pi t/\tau), \text{ or, } \delta(t) = -A + 2A\big(t/\tau - \lfloor t/\tau \rfloor\big),$$

where τ is the period. Similarly, the service rate functions $\beta(t) = \mu + \epsilon(t)$, $t \geq 0$ are sinusoidal or sawtooth deviation functions, with the same period and a shift,

$$\epsilon(t) = B\sin\big(2\pi(t-\sigma)/\tau\big), \text{ or, } -B + 2B\big((t-\sigma)/\tau - \lfloor (t-\sigma)/\tau \rfloor\big).$$

We consider transient level-crossing probabilities, and busy cycle level-crossing probabilities. These become rare in the fluid scaling regime, which goes as follows (in case of transient level-crossing). Define $\{X_n(t) : 0 \leq t \leq T; n = 1, 2, \ldots\}$ to be $M_t/M_t/1$ queues with arrival rate $\alpha_n(t) = n\alpha(t)$, and service rate $\beta_n(t) = n\beta(t)$. T is a finite, fixed horizon. Fix $\overline{x} > 0, \overline{y} > 0$, and T; then the rare-event transient level-crossing probability is

$$\ell_n = P\big(X_n(T)/n \geq \overline{y} \,|\, X_n(0)/n = \overline{x}\big),$$

where we let $n \to \infty$. We shall analyse the most likely behaviour for this rare event, and we obtain the most likely paths to the rare event. Then we discuss an importance sampling sampling simulation algorithm to efficiently estimate these probabilities, and analyse its complexity. The analysis is illustrated by numerical results.

Online Youtube

Solving Integral Equations in real-time

Speaker: Nikolaus Binder

Nikolaus Binder
NVIDIA, Germany, nbinder@nvidia.com

Alexander Keller
NVIDIA, Germany, akeller@nvidia.com

For the setting of computer graphics, we present an algorithm that can be considered a step into the direction of computing functionals of complex Fredholm integral equations in real-time: Restricting path tracing to a small number of paths per pixel for performance reasons rarely achieves a satisfactory image quality for scenes of interest. However, path space filtering [1] may dramatically improve the visual quality by sharing information across vertices of paths classified as proximate. Unlike screen space-based approaches, these paths neither need to be present on the screen, nor is filtering restricted to the first intersection with the scene. While searching proximate vertices had been more expensive than filtering in screen space, we greatly improve over this performance penalty by storing and looking up the required information in a hash table, optimized for high performance on massively parallel processors. The keys are constructed from jittered and quantized information, such that only a single query very likely replaces costly neighborhood searches. A massively parallel implementation of the algorithm is demonstrated on a GPU and a preprint is available [2].

[1] A. Keller, K. Dahm, and N. Binder. Path Space Filtering. In *Monte Carlo and Quasi-Monte Carlo Methods 2014*, R. Cools and D. Nuyens (Eds.). Springer, 423–436, 2016.

[2] N. Binder, S. Fricke, and A. Keller Massively Parallel Path Space Filtering. https://arxiv.org/pdf/1902.05942.pdf, 2019.

Special Session

Optimization and Optimality

Chair: First Last

Online **Youtube**

Infinite-Dimensional Alpha-Divergence Minimisation for Variational Inference

Speaker: Kamélia Daudel

Kamélia Daudel
LTCI, Télécom Paris, Institut Polytechnique de Paris, France, kamelia.daudel@telecom-paris.fr

Randal Douc
SAMOVAR, Télécom SudParis, Institut Polytechnique de Paris, France,
randal.douc@telecom-sudparis.eu

François Portier
LTCI, Télécom Paris, Institut Polytechnique de Paris, France, francois.portier@telecom-paris.fr

In this work, we introduce the (α, Γ)-descent [1], an iterative algorithm which operates on measures and performs α-divergence minimisation in a Bayesian framework. This gradient-based procedure extends the commonly-used variational approximation by adding a prior on the variational parameters in the form of a measure. In particular, it allows to optimise the mixture weights of any given mixture model without any information on the underlying distribution of the variational parameters.

We prove that for a rich family of functions Γ, this algorithm leads to a systematic decrease in the α-divergence at each step. Our framework recovers the Entropic Mirror Descent (MD) algorithm with improved $O(1/N)$ convergence results and provides a novel alternative to the Entropic MD that we call the Power Descent and for which we prove convergence to an optimum. As our method involves computing an integral which might be intractable, we resort to Monte Carlo approximations to obtain a practical version of our algorithm. We demonstrate empirically on both toy and real-world examples the benefit of using the Power Descent and going beyond the Entropic MD framework.

[1] Kamélia Daudel, Randal Douc and François Portier. Infinite-dimensional gradient-based descent for alpha-divergence minimisation, https://hal.telecom-paris.fr/hal-02614605, 2020.

Online **Youtube**

Monte Carlo Optimization Including Weights

Speaker: Stefka Fidanova

Stefka Fidanova
Institute of Information and Communication Technologies, Bulgaria, stefka@parallel.bas.bg

Krassimir Atanassov
Institute of Biophysics and Biomedical Engineering, Bulgaria, krat@bas.bg

NP-hard problems need huge amount of computational resources,which are growing exponentially, when the problem grows. Traditional numerical methods are not applicable on these kind of problems, therefore Monte Carlo methods are used to be solved. A nature inspired Monte Carlo methods and algorithms are widely used to solve HP-hard optimization problems. Examples are simulated annealing, bee colony optimization, bat algorithm, fire fly algorithm, particle swarm optimization, gray wolf algorithm and so on. Ant Colony Optimization (ACO) [1] is one of the most successful Monte Carlo methods for solving discrete optimization problems. The idea comes from real ants behavior, which manage to establish the shortest routes to feeding sources and back. The ants in a nature mark their way back with a chemical substance called pheromone. The problem is represented by a graph and the solutions are represented by path in a graph. Thus finding the optimal solutions is equivalent to finding the shorter path in a graph, according some constraints. The artificial ants start from random node. They include new nodes in their solution applying probabilistic rule. The probabilistic rule is a product of the quantity of the pheromone and heuristic information related to the problem. In the algorithm a numerical information is related to the elements of the graph imitating the pheromone. The elements of better solutions, receive more pheromone than others and they become more desirable. Sometimes some elements of the graph accumulates more pheromone than others. In this paper we include weights related with the nodes of the graph. At the beginning the weight of the nodes is the same for hall nodes. If some node is included in some solution, its weight decreases. In our algorithm the transition probability is a product of three elements: quantity of the pheromone; heuristic information and weight. As a results the regions with best so far solutions and unexplored regions in a search space are more desirable, than explored regions with worse solutions. This new element in the transition probability leads to enlarging diversification of the search and giving more chance to unexplored regions of the search space to be explored. Our idea is tested on Multiple Knapsack Problem as a representative of subset problems. It is capital budgeting problem and a lot of real life and industrial problems can be defined as a knapsack problem.

[1] M. Dorigo, T. Stutzle. *Ant colony optimization*. MIT Press, 2004.

Online **Youtube**

Monte Carlo integration using designs obtained via Periodic Maximin and ϕ_p (Phi) Criteria

Speaker: Miroslav Vořechovský

Miroslav Vořechovský
Brno University of Technology, Czech Republic, vorechovsky.m@vut.cz

Jan Eliáš
Brno University of Technology, Czech Republic, elias.j@fce.vutbr.cz

Jan Mašek
Brno University of Technology, Czech Republic, masek.j@fce.vutbr.cz

The presentation, which is a promotion of a recent Technometrics paper, proposes an alternative sampling technique that delivers robust designs selected from a design domain in the shape of a unit hypercube. The designs are guaranteed to provide a *statistically uniform* point distribution, meaning that every location has the same probability of being selected. Moreover, the designs are *sample-uniform*, meaning that each individual design has its points spread evenly throughout the domain. The *sample uniformity* (often measured via a *discrepancy* criterion) is achieved using distance-based criteria (ϕ_p or Maximin), i.e. criteria normally used in space-filling designs. We show that the standard intersite metrics employed in distance-based criteria (Maximin and ϕ_p (phi)) do *not* deliver statistically uniform designs. Similarly, designs optimized via centered L_2 discrepancy or support points are also not statistically uniform. When these designs (after optimization based on intersite distances) are used for Monte Carlo type of integration, their statistical nonuniformity is a serious problem as it may lead to a systematic bias. This paper proposes using a periodic metric to guarantee the statistical uniformity of the family of distance-based designs. The ability to integrate smooth functions will be documented. Additionally, designs obtained with the proposed criteria are compared to other known designs with respect to various optimality criteria. We suggest that the proposed distance-based periodic ϕ_p (Phi) criterion be a suitable figure of merit regarding efficiency in numerical integration. The talk will also introduce methods for construction of the designs. One of the ways is to exploit recently proposed analogy between the proposed design criterion and a dynamical system of interacting particles. The possibility of viewing the above distance-based optimality criteria as formulas representing the potential energy of a system of charged particles will be discussed. The potential energy is employed in deriving the equations of motion of the particles. The particles are either attracted to all points in space or mutually repelled and dissipative dynamical systems can be simulated to achieve optimal and near-optimal arrangements of points.

Online Youtube

On the Evolution of Minimal-Volume, Sufficient-Probability Sets for Stochastic Paths

Speaker: Ryan White

Ryan White
Mathematical Sciences, Florida Institute of Technology, United States, rwhite@fit.edu

In analysis a stochastic process $A(t)$ valued in $\mathbb{R}^n$, of great interest is a minimal volume set M_t (in the Lebesgue sense) with a given fixed probability $1-\alpha$ of containing $A(t)$. We propose a Monte Carlo method to track the evolution of such minimal-volume-sufficient-probability sets M_t. Garcia, et. al. [1] provide some results that ensure the problem reduces to predicting the evolution of level sets of the probability density of $A(t)$. We find these efficiently through a Monte Carlo method where the probability density is known (multidimensional Brownian motion and a continuous marked random walk process studied by Dshalalow and White in [2], [3]) and a jump process whose probability density is unknown but approximated empirically on a mesh.

[1] Javier Nuñez Garcia, Zoltan Kutalik, Kwang-Hyun Cho, Olaf Wolkenhauer. Level sets and minimum volume sets of probability density functions. *International Journal of Approximate Reasoning*, 34 (1): 25-47, 2003.

[2] Ryan T. White and Jewgeni H. Dshalalow. Characterizations of random walks on random lattices and their ramifications. *Journal Stochastic Analysis and Applications*, 38 (2): 307-342, 2019.

[3] Ryan T. White. On Reliability of Stochastic Networks with Empirically Distributed Failures (manuscript in progress)

Special Session

Methods for Engineering and Finance

Chair: First Last

Online **Youtube**

Higher-order weak schemes for the Heston stochastic volatility model by extrapolation

Speaker: Chao Zheng

Chao Zheng
Department of Data Sciences, Zhejiang University of Finance and Economics, China,
chao.zheng12@gmail.com

The Heston stochastic volatility model is one of the most important models in mathematical finance. Although many discretization schemes have been proposed for this model, the analysis of their weak convergence rates is quite limited. We consider a time-discrete scheme for the Heston model, which employs the stochastic trapezoidal rule to discretize the logarithmic asset process, provided that the variance process is simulated exactly. Zheng [1] has proved that this scheme is of weak order two for the full parameter regime, for any polynomial payoff function of the log-asset process. We extend this result by proving that the weak error can be expanded in arbitrarily high powers of step size. This property allows us to build a weak scheme of an arbitrary order by extrapolation, and the result is also free of parameter restrictions. Furthermore, our analysis is purely deterministic rather than stochastic, where the problem can be transferred to a deterministic numerical integration problem with some singularities. Finally, we extend the analysis for more general SDEs.

[1] C. Zheng Weak convergence rate of a time-discrete scheme for the Heston stochastic volatility model. *SIAM Journal on Numerical Analysis*, 55 (3): 1243-1263, 2017.

Online Youtube

Statistical Analysis of Progressive Type-I Interval Censored Data under Competing Risks

Speaker: Soumya Roy

Soumya Roy
Quantitative Methods and Operations Management Area, Indian Institute of Management Kozhikode, India,
soumya@iimk.ac.in

In reliability engineering literature, Type-I and Type-II censoring schemes have received major attention from the researchers. This is primarily due to the fact that these two censoring schemes are easy to implement in practical applications. However, a major drawback with these traditional censoring schemes is that they assume continuous inspection during the life-test experiment. Furthermore, these censoring schemes do not allow intermediate withdrawals from the ongoing experiment. In view of this, Progressive Type-I Interval Censoring (henceforth, PIC-I) scheme is proposed in the literature. A PIC-I scheme is essentially an extension of the traditional Interval Censoring scheme, in the sense that it permits intermediate withdrawals from the experiment. In the existing literature on PIC-I data, it is commonly assumed that that the systems have only one failure mode. However in practice, a system often has more than one failure mode. Moreover, the system may stop functioning as soon as one of the failure modes occurs. This article works with such systems. It is assumed that the lifetime corresponding to each failure mode follows an independent Weibull model with distinct shape parameter. Classical and Bayesian inference for the unknown model parameters are then presented based on PIC-I data. A detailed Monte Carlo simulation study is conducted to evaluate the performance of various Classical and Bayesian estimators. A real data set is also analyzed to illustrate the developed methodology, after presenting necessary model comparisons through Bayes Factors. An extension to the dependent set-up is also discussed.

Online **Youtube**

A new combined kernel-projection statistical estimator with applications to the study of polarized radiation intensity

Speaker: Natalya Tracheva

Gennady Mikhailov
Laboratory of Monte Carlo methods, Institute of Computational Mathematics and Mathematical Geophysics SB RAS,
Mechanics and Mathematics Departmen, Novosibirsk State University, Russia, gam@sscc.ru

Natalya Tracheva
Laboratory of Monte Carlo methods, Institute of Computational Mathematics and Mathematical Geophysics SB RAS,
Mechanics and Mathematics Departmen, Novosibirsk State University, Russia, tnv@osmf.sscc.ru

Sergey Ukhinov
Laboratory of Monte Carlo methods, Institute of Computational Mathematics and Mathematical Geophysics SB RAS,
Mechanics and Mathematics Departmen, Novosibirsk State University, Russia, sau@sscc.ru

Kernel statistical estimator in the Monte Carlo method is usually optimized based on the preliminary construction of a micro-grouped sample of the studied variable values [1]. This approach works in the case of the one-dimensional problem, but even in the two-dimensional problem case, such optimization is quite difficult. In this work, we propose a combined kernel-projection statistical estimator of the two-dimensional distribution density. We construct it in the following manner: for one of the variables the classical one-dimensional kernel estimator is formed and for the other – the projection estimator. According to this approach, for each kernel bandwidth, pre-defined with a micro-sample, the coefficients of the certain orthogonal expansion of the conditional density are statistically estimated. Provided the assumptions made about the convergence rate of the orthogonal decomposition in use, we obtained optimal parameters for such a combined kernel-projection statistical estimator. This approach was implemented for estimating a bidirectional angular distribution of the polarized radiation flux transmitted through and backscattered by the scattering layered substances [2]. Results were verified with the local Monte Carlo estimator and the direct Monte Carlo simulation. The reported study was partially funded by RFBR according to the research project 18-01-00356.

[1] G. A. Mikhailov, S. M. Prigarin, S. A. Rozhenko. Comparative analysis of vector algorithms for statistical modelling of polarized radiative transfer process. *Russian Journal of Numerical Analysis and Mathematical Modelling*, 33 (4): 253–263, 2018.

[2] N. V. Tracheva, S. A. Ukhinov. On the evaluation of spatial–angular distributions of polarization characteristics of scattered radiation. *Statistical Papers*, 59 (4): 1541–1557, 2018.

Online Youtube

Efficient Monte Carlo methods for forecasting returns from large-scale consumer debt portfolios

Speaker: Edmund Ryan

Edmund Ryan
Department of Mathematics, University of Manchester, UK, edmund.ryan@manchester.ac.uk

Simon Cotter
Department of Mathematics, University of Manchester, UK, simon.cotter@manchester.ac.uk

Timothy Waite
Department of Mathematics, University of Manchester, UK, timothy.waite@manchester.ac.uk

Sam Baynes
Analytics, Arrow Global Ltd, UK, sbaynes@arrowglobal.net

Paul Russell
Analytics, Arrow Global Ltd, UK, prussell@arrowglobal.net

In the debt recovery industry, financial models exist to produce long-term account-level cash flows, which are key to the analysis of strategies and valuations. At Arrow Global, a stochastic model is used to simulate the collections for each of the $\sim$4 million customers. In Arrow's current implementation only a small number of Monte Carlo repeats for each customer are carried out due to the high computational cost. In this talk we will present more computationally efficient methods that use Gaussian processes and subsampling to optimize the number of required repeats per customer.

Special Session

Discrepancy and Quasi-Monte Carlo Methods

Chair: First Last

Online -

On the asymptotic behaviour of the Sudler product of sines

Speaker: Mario Neumüller

Sigrid Grepstad
Department of Mathematical Sciences, Norwegian University of Science and Technology, Norway,
sigrid.grepstad@ntnu.no

Lisa Kaltenböck
Department of Financial Mathematics and Applied Number Theory, Johannes Kepler University, Austria,
lisa.kaltenboeck@jku.at

Mario Neumüller
Department of Financial Mathematics and Applied Number Theory, Johannes Kepler University, Austria,
mario.neumueller@jku.at

We study the growth of the following trigonometric product

$$P_N(\alpha) = \prod_{r=1}^{N} |2\sin(\pi r\alpha)|,$$

where α is real and irrational. These kind of products, sometimes referred to as Sudler product, appear in a variety of different fields of pure and applied mathematics e.g. partition theory, Pade approximation, continued fraction theory or topics related to mathematical physic. Moreover, these products have interesting connections to the discrepancy of certain hybrid sequences (see [3]) and uniform distribution theory (see [1]).
More precise, we are going to focus on the asymptotic behaviour of the subsequence $(P_{q_n}(\alpha))_{n\geq 1}$, where q_n is the nth best approximation denominator of α and α is a fixed irrational number fulfilling some additional properties e.g. being a quadratic irrational or having bounded continued fraction coefficients. Additionally, we will investigate the quantity $\liminf_{N\to\infty} P_N(\alpha)$ and present interesting results concerning unresolved questions posed by Erdős and Szekeres more than 60 years ago (see [2]).

[1] C. Aistleitner, G. Larcher, F. Pillichshammer, S. Eddin and R. Tichy. On Weyl products and uniform distribution modulo one. *Monatsh. Math.*, 185 (3): 365–395, 2018.

[2] P. Erdős and G. Szekeres. On the product $\prod_{k=1}^{n}(1-z^{a_k})$. *Acad. Serbe Sci. Publ. Inst. Math.*, 13: 29–34, 1959.

[3] R. Hofer and F. Puchhammer. On the discrepancy of two-dimensional perturbed Halton-Kronecker sequences and lacunary trigonometric products. *Acta Arith.*, 180: 365–392, 2017.

Online Youtube

Multivariate approximation based on transformed rank-1 lattices

Speaker: Robert Nasdala

Robert Nasdala
Faculty of Mathematics, Chemnitz University of Technology, Germany,
robert.nasdala@math.tu-chemnitz.de

Daniel Potts
Faculty of Mathematics, Chemnitz University of Technology, Germany, potts@math.tu-chemnitz.de

For the approximation of multivariate non-periodic functions on high-dimensional domains, such as R^d and the cube $\left[-\frac{1}{2}, \frac{1}{2}\right]^d$, we combine a periodization strategy for weighted L_2-integrands with efficient approximation methods. We prove sufficient conditions on transformations to the d-variate domains and on the non-negative weight function such that the composition of a possibly non-periodic function with such a transformation yields a smooth function in the Sobolev space of functions on the torus with mixed smoothness of natural order. In this framework we adapt certain approximation error estimates for single rank-1 lattice approximation methods as well as algorithms for the evaluation and reconstruction of multivariate trigonometric polynomials on the torus. Various numerical tests confirm the obtained theoretical results for the transformed approximation methods.

[1] R. Nasdala und D. Potts Transformed rank-1 lattices for high-dimensional approximation. *Electron. Trans. Numer. Anal.*, 53, 239-282, 2020.

[2] R. Nasdala und D. Potts Efficient multivariate approximation on the cube. arXiv: 1912.03090, 2019.

Online **Youtube**

Reproducing Kernel Banach Spaces and QMC Integration

Speaker: Marcin Wnuk

Marcin Wnuk
Institute of Mathematics, University of Osnabrück, Germany, marcin.wnuk@uni-osnabrueck.de

Michael Gnewuch
Institute of Mathematics, University of Osnabrück, Germany, michael.gnewuch@uni-osnabrueck.de

Aicke Hinrichs
Department of Mathematics, Johannes Kepler University Linz, Austria, aicke.hinrichs@jku.at

Alexander Lindenberger
Department of Mathematics, Johannes Kepler University Linz, Austria, alexander.lindenberger@jku.at

Reproducing kernel Hilbert spaces are a typical setting in which QMC integration problems are studied. One of their greatest advantages is that worst-case error can be explicitly written down in terms of the reproducing kernel. In recent years there has been a growing interest, above all in the Machine Learning community, in generalizing the concept of reproducing kernels to the Banach space context, see e.g. [1]. In this talk first we present the notion of reproducing kernel Banach spaces introduced in [1], and then give examples of scales of Banach spaces interesting for MCQMC community which fit into the framework. Those encompass function spaces defined by decay of weighted (generalized) Fourier coefficients (e.g. wavelet spaces) and spaces that may be defined with the help of integral kernels in L^p spaces (e.g. anchored and ANOVA Sobolev spaces). In the end we present Koksma-Hlawka type inequalities and error formulae valid for the spaces of interest.

[1] P. Georgiev, L. Sanchez-Gonzalez, P. M. Pardalos, Construction of Pairs of Reproducing Kernel Banach Spaces. *Constructive Nonsmooth Analysis and its Applications* ed. V.F. Demyanov et al., Springer Science+Business Media New York,2014.

Online Youtube

Optimal cubature rules on Haar wavelet spaces and on spaces with fractional smoothness

Speaker: Michael Gnewuch

Josef Dick
School of Mathematics, University of New South Wales, Australia, josef.dick@unsw.edu.au

Michael Gnewuch
Institut für Mathematik, Universität Osnabrück, Germany, michael.gnewuch@uni-osnabrueck.de

Lev Markhasin
Institut für Stochastik und Anwendungen, Universität Stuttgart, Germany,
lev.markhasin@mathematik.uni-stuttgart.de

Winfried Sickel
Mathematisches Institut, Friedrich-Schiller-Universität Jena, Germany, winfried.sickel@uni-jena.de

We consider Haar wavelet spaces that are defined by a decay parameter α and by summability parameters $1 \leq p, q \leq \infty$, where $p > 1/\alpha$. By proving upper and lower bounds we show that quasi-Monte Carlo cubatures based on (arbitrary) (t, m, s)-nets achieve the optimal rate of convergence. These findings generalize and extend results from [2] and [3]. Furthermore, we consider spaces of functions with integrability/summability parameters $1 \leq p, q \leq \infty$ and fractional smoothness $1/p < \alpha \leq 1$, defined with the help of fractional derivatives in the sense of Riemann and Liouville. For a range of parameters we are able to embed these spaces first into corresponding Besov spaces and then the Besov spaces into the corresponding Haar wavelet spaces. This implies, in particular, that quasi-Monte Carlo cubatures based on (arbitrary) (t, m, s)-nets achieve (at least) the same rate of convergence in the spaces of fractional smoothness as in the corresponding Haar wavelet spaces. The last mentioned finding complements the results in [1].

[1] J. Dick. Koksma-Hlawka type inequalities of fractional order. *Ann. Mat. Pura Appl.*, 187: 385–403, 2008.

[2] K. Entacher. Quasi-Monte Carlo methods for numerical integration of multivariate Haar series. *BIT*, 37: 846–861, 1997.

[3] S. Heinrich, F. J. Hickernell, R. X. Yue. Optimal quadrature for Haar wavelet spaces. *Math. Comput.*, 73: 259–277, 2004.

Online Youtube

A-posteriori QMC error estimation

Speaker: Marcello Longo

Josef Dick
University of New South Wales, Sydney, josef.dick@unsw.edu.au

Marcello Longo
SAM, ETH Zürich, Switzerland, marcello.longo@sam.math.ethz.ch

Christoph Schwab
SAM, ETH Zürich, Switzerland, christoph.schwab@sam.math.ethz.ch

A-posteriori estimates are computable quantities that give information about a numerical approximation. In many practical applications, these are used to devise a stopping criterion such that the error is bounded by given tolerance, without prior knowledge of the exact solution and without waste of computational time. While there are several results on the a-priori approximation power of Quasi-Monte Carlo sampling, very little is known about a-posteriori estimation.

We introduce a computable QMC error estimator [1], using the CBC construction of polynomial lattice rules from [2]; further, we show that the estimator is asymptotically exact, for a class of functions in weighted unanchored Sobolev spaces with dominating mixed smoothness.

The asymptotic exactness of the QMC error estimator does not incur in the curse of dimensionality and our approach can be also applied for general weights as shown in [1].

[1] J. Dick, M. Longo, and Ch. Schwab. Extrapolated lattice rule integration in computational uncertainty quantification. Technical Report 2020-29, Seminar for Applied Mathematics, ETH Zürich, Switzerland, 2020.

[2] Josef Dick, Takashi Goda, and Takehito Yoshiki. Richardson extrapolation of polynomial lattice rules. *SIAM J. Numer. Anal.*, 57(1):44–69, 2019.

List of Speakers

List of Chairs