Mathematisches Forschungsinstitut Oberwolfach

Report No. 21/2025

DOI: 10.4171/OWR/2025/21

Uncertainty Quantification

Organized by
Oliver G. Ernst, Chemnitz
Fabio Nobile, Lausanne
Claudia Schillings, Berlin
Tim J. Sullivan, Coventry/London

20 April – 25 April 2025

Abstract. Uncertainty quantification (UQ) is concerned with including and characterising uncertainties in mathematical models. Major steps include the proper description of system uncertainties, analysis and efficient quantification of uncertainties in predictions and design problems, and statistical inference on uncertain parameters starting from available measurements. Research in UQ addresses fundamental mathematical and statistical challenges, but has also wide applicability in areas such as engineering, environmental, physical and biological applications. This workshop focussed on mathematical challenges at the interface of applied mathematics, probability and statistics, numerical analysis, scientific computing and application domains such as machine learning, modelling of energy production, and bifurcations in climate models. The workshop brought together experts from those disciplines to enhance their interaction, to exchange ideas and to develop new, powerful methods for UQ.

Mathematics Subject Classification (2020): 49-XX, 60-XX, 62-XX, 65-XX, 90-XX.

Introduction by the Organizers

This workshop on uncertainty quantification (UQ) was the sequel to one held in March 2019. UQ is a key area of research at the interface of applied mathematics, statistics, and computational science and engineering (CSE). It plays a central role in predictive modelling by ensuring that uncertainty—arising from incomplete

knowledge, variability, or measurement error—is rigorously characterised and incorporated into simulation, inference, and decision-making. UQ encompasses several interconnected challenges: modelling and representing system uncertainties; analysing sensitivities and quantifying their impact on predictions; inferring parameters from partial, noisy, and indirect data; and enabling experimental design and optimisation under uncertainty. This workshop aimed to survey advances across the broad front of UQ activity over the last few years.

The Bayesian approach to inverse problems is a topic of long-standing importance in UQ. Since the unknowns of interest are often functions and fields, there is a strong case—nowadays commonly attributed to the "Finnish school" of inverse problems c. 2005 and seminal contributions of Andrew Stuart c. 2010—that Bayesian inference for such objects should be analysed in the corresponding infinite-dimensional spaces and that the finite-dimensional computations should be formulated in a dimension-independent fashion. These aspects were discussed by many speakers, but were particularly prominent in the joint talk of Andrea Barth and Oliver König on Bayesian inversion with stochastic forward maps, Hefin Lambley's talk on autoencoders in function space, the talk of Björn Sprungk on dimension-independent Markov chain Monte Carlo on the sphere, and Hanne Kekkonen's proposal of random-tree Besov priors for edge preservation. Tapio Helin presented results on approximate optimal experimental design in Bayesian inverse problems.

A commonly-used summary statistic for Bayesian inverse problems in the maximum a posteriori estimator, which may be heuristically understood as a "most likely point" under the posterior distribution and as a minimiser of the prior-regularised misfit functional. Ilja Klebanov gave a blackboard talk on rigorous generalisations of these ideas to infinite-dimensional and/or metric spaces.

The approximation of Bayesian posterior distributions using variational Bayes approaches came up in several talks, both their general ideas (Håvard Rue) and their use within variational autoencoders (Hefin Lambley). Other approximation approaches discussed included low-rank structure and other factorisations, as in the talks of Colin Fox and Han Cheng Lie.

Sampling-based approximations to target distributions are also workhorse tools of Bayesian inference and UQ more generally. In this vein, Youssef Marzouk—who has long worked on methods for transporting samples from a "simple" reference distribution to a complicated target—gave a talk on the optimal scheduling of these dynamic transport schemes. Some results on the near-optimality of quasi-Monto Carlo schemes were presented by Yoshihito Kazashi. Björn Sprungk's talk on dimension-independent MCMC again deserves mention in this context. However, one connection that sprang into prominence at this workshop (compared to 2019) was the interaction between control theory and sampling, with the control objective being to drive the sampling distribution towards the target; Raúl Tempone, Sebastian Reich, Georg Stadler, and Philipp Guth all talked on aspects of this promising approach. Tangentially to this area, Caroline Geiersbach talked

about some stochastic optimisation problems in Banach spaces, in particular those with random or almost sure state constraints.

Numerical methods for deterministic and stochastic partial differential equations remain of considerable relevance to the UQ community. In this connection, Kristin Kirchner presented some numerical methods for space-time SPDEs. Hanno Gottschalk addressed an extension of the familiar diffusion problem to random fields with Lévy coefficients, where there are challenges both in formulating an appropriate Karhunen–Loève expansion and in calculating quadrature estimates. Mattieu Dolbeault also addressed a variation on the familiar elliptic problem, in this case with high-contrast diffusion coefficients.

Building upon the PDE problems that are familiar to the UQ problem, Ana Djurdevac was able to address some problems in the field of shape uncertainty. Further broadening the kinds of uncertainties amenable to UQ analysis, Kerstin Lux-Gottschalk addressed UQ for bifurcations, with a particular application to tipping points in climate models.

Sven Wang and Imma Curato both addressed the problem of learning from high-dimensional data and processes, Wang in the case of low-frequency data from an underlying diffusion, and Curato the use of mixed moving-average fields to learn from rasterised spatio-temporal fields.

The connections between UQ and machine learning (ML) were strongly represented in this workshop, much more so than in 2019; this reflected the explosion in ML in recent years, and the consequent importance of UQ for ML applications. Eyke Hüllermeier gave an excellent survey of different forms of uncertainty in ML, along with open challenges and fundamental "no-go" theorems. In the opposite direction, Claudia Strauch discussed some statistical guarantees for the performance of denoising diffusion models as generative models in ML. Kernel methods for the learning of differential equations were discussed by Houman Owhadi, with a particular emphasis on data efficiency and quantitative error bounds. Hefin Lambley offered a talk on the extension of autoencoders and generative modelling to function spaces.

On Wednesday evening, after returning from the traditional hike to Sankt Roman, four workshop participants treated the group to a special musical programme. Alexey Chernov (guitar), Caroline Geiersbach (violin), Youssef Marzouk (piano), and Sven Wang (piano) performed a selection of solo and duet pieces, ranging from lively folk dances to classical works by Beethoven, Brahms, and Chopin.

Acknowledgement: The MFO and the workshop organizers would like to thank the National Science Foundation for supporting the participation of junior researchers in the workshop by the grant DMS-2230648, "US Junior Oberwolfach Fellows".

Workshop: Uncertainty Quantification

Table of Contents

Abstracts

Conic-constrained stochastic optimization: optimality conditions and sample-based consistency

Caroline Geiersbach
(joint work with Johannes Milz)

This talk is concerned with a general class of risk-neutral stochastic optimisation problems defined on a Banach space with almost sure conic-type constraints of the form

$$\min_{u\in U} \quad \mathbb{E}[J(u,\xi)] + \psi(u) \qquad \text{subject to} \qquad G(u,\xi) \in K \quad \text{almost surely.} \tag{1}$$

Here, ξ is a vector-valued random variable defined on a complete probability space and the set U is a Banach space. The objective is possibly non-convex and is allowed to contain a non-smooth convex term ψ, and the parametrised constraint function G should be contained in a convex cone K in Banach space R. This setting is motivated by applications from optimal control under uncertainty where the control-to-state operator is not necessarily linear and a further constraint is applied to the state. For this class of problems, we investigate the consistency of optimal values and solutions to the corresponding sample average approximation (SAA) problem

$$\min_{u\in U} \ \frac{1}{N}\sum_{k=1}^{N} J(u,\xi^k) + \psi(u) \qquad \text{subject to} \qquad G(u,\xi^i) \in K, \qquad i = 1,\dots,N$$

as the sample size N is taken to infinity. An assumption of compactness with respect to the infinite-dimensional optimisation variable u permits us to invoke epigraphical laws of large numbers following the arguments developed in [1].

In numerical simulations, a Moreau–Yosida-type regularisation of the constraint is often used to handle state constraints. A continuous function $\beta\colon R \to [0,\infty)$ is introduced that has the property whereby $\beta(k) = 0$ if and only if $k \in K$. We study the example where problem (1) is replaced by

$$\min_{u\in U} \quad \mathbb{E}[J(u,\xi)] + \psi(u) + \gamma\mathbb{E}[\beta(G(u,\xi))]. \tag{2}$$

Consistency of solutions and optimal values to the corresponding SAA problems can also be proven in this case when γ_N and N are taken to infinity.

In the second half of the talk, we consider optimality conditions corresponding to (1) and (2). The classical framework in which optimality conditions for these problems have been derived involves the assumption that the constraint function is essentially bounded with respect to the parameter ξ. This allows for the application of a constraint qualification, which relies on the existence of interior points. An alternative way to ensure the existence of Lagrange multipliers is to strengthen the assumptions to require continuity with respect to ξ as was done in [2]. This is the setting found in robust and semi-infinite optimisation and appears to provide a theoretical advantage in the sense that, provided R is separable, one can

work with sequential arguments for Lagrange multipliers as opposed to nets (see [3] for the latter). Certainly, this is desirable for the SAA problem in the study of the limiting case as N is taken to infinity. By assuming more regularity on the objective and constraint functions, we obtain the existence of Lagrange multipliers for problems (1) and (2) under Robinson's constraint qualification. For (1), consistency of Lagrange multipliers from the underlying optimality conditions can be shown; for this, the constraint qualification must be satisfied locally at an accumulation point of the sequence (u_N^*) of optimal solutions to the corresponding SAA problem. Similar arguments can be applied in the case of problem (2).

This analysis fills a theoretical gap for infinite-dimensional conic-constrained stochastic optimisation problems, providing the theoretical justification for the numerical computation of solutions using SAA.

References

[1] J. Milz, T. Surowiec. *Asymptotic consistency for nonconvex risk-averse stochastic optimization with infinite dimensional decision spaces.* Math. Oper. Res., 9 (2024), 933 1403–1418.

[2] C. Geiersbach, R. Henrion. *Optimality conditions in control problems with random state constraints in probabilistic or almost-sure form.* Math. Oper. Res., Articles in Advance (2024), 1–27.

[3] C. Geiersbach, M. Hintermüller. *Optimality conditions and Moreau–Yosida regularization for almost sure state constraints.* ESAIM Control Optim. Calc. Var. **28** (2022), Paper No. 80, 36.

Climate Tipping Points under Uncertainty – A Bifurcation Theoretic Perspective

Kerstin Lux-Gottschalk

(joint work with Christian Kuehn, Peter Ashwin, Richard Wood, Jonathan Baker, Björn Sprungk, and Oliver G. Ernst)

Greenland Ice Sheet, Atlantic Meridional Overturning Circulation and Amazon Rainforest – These are prominent examples of subsystems of the Earth that exhibit inherently nonlinear dynamics. Instead of only promoting gradual changes, in these systems, abrupt large and oftentimes irreversible changes can occur under the variation of an external forcing parameter and might alter equilibria and their stability properties significantly. This phenomenon is known as tipping. In this talk, I will shed light on Climate Tipping Points from the perspective of bifurcation theory by identifying tipping points as particular bifurcation points of the system. Whereas bifurcation theory is an established and well understood discipline in nonlinear dynamics, very interesting research questions appear at the interface with Uncertainty Quantification (UQ). How do uncertain model parameters affect the bifurcation behaviour of the system and which parameters contribute most to the tipping dynamics? What is the probability of a system to tip? Which role does time-correlation in noisy systems play? In this talk, I will focus on uncertainty in model parameters and present a combination of sensitivity analysis and bifurcation theory in the form of a probabilistic analysis of bifurcation points and curves. The

methodology will be illustrated in the context of tipping points of the Atlantic Meridional Ocean Circulation (AMOC). The latter plays an important role for the North Atlantic heat transport.

Figure 1 provides a visualisation of our methodology.

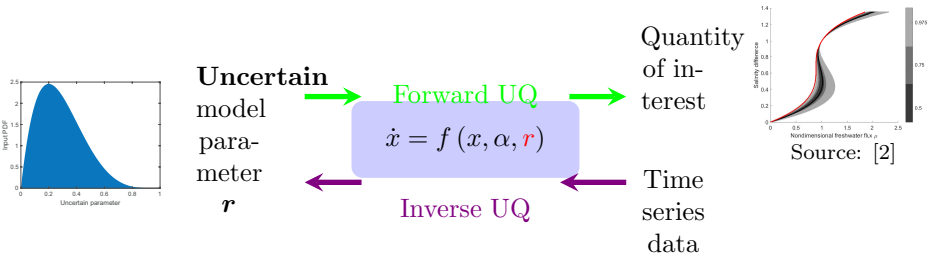


FIGURE 1. Visualization of our UQ Methodology.

Oftentimes, we do not have exact knowledge of the parameter values but rather physically reasonable ranges. Therefore, in a first inverse UQ step, we use a Bayesian inference technique to narrow down the uncertainty in these prior ranges based on time series data. As a proof of concept of our methodology, we start with a very simple two-box model of the AMOC [1] and use synthetic time series data that we obtained by a forward simulation of the simple conceptual model for the AMOC and including Gaussian additive noise. We calculate the posterior probability distribution of our model parameter of interest by using a Markov Chain Monte Carlo approach and derive a probabilistic representation of bifurcation curves [2]. We achieve a substantial narrowing of the range of tipping to occur within. This brings prospects that we can use knowledge of past behaviour to better understand likelihoods of future tipping events.

In our current ongoing work, as a second step, we work with a conceptual five-box model of the AMOC [3]. This enables us to use actual time series data from bigger General Circulation Models (GCMs) and infer corresponding parameter values based thereon. The idea is to provide a surrogate model for the computationally expensive GCMs. Based on the obtained posterior distribution for the model parameters, we perform a forward uncertainty propagation through the nonlinear dynamics to obtain the distribution of our output quantity of interest. We first focus on a particular bifurcation point. We calculate Sobol' indices to identify the most influential parameters regarding the location of this bifurcation point.

As a next step, we focus on bifurcation curves that describe the overall tipping behaviour of the system instead of just providing the location of a single tipping point. A challenge is the non-scalar valued nature of these objects. To tackle this, we have ongoing research starting from [4], where the authors present a sensitivity analysis of uncertainty propagation for differential equations with random inputs subject to perturbations of the input measures. To build upon these results, we

choose the Fréchet distance as a metric on the space of bifurcation curves. We transfer results from [4] to bifurcation theory. Thereby, we provide a worst case estimate on the distance between families of bifurcation curves for different input parameter distributions. Thereby, we contribute to a risk assessment of climate tipping phenomena.

References

[1] P. Cessi, *A simple box model of stochastically forced thermohaline flow*, J. Phys. Oceanogr. **24, no. 9** (1994), 1911–1920.

[2] K. Lux, P. Ashwin, R. Wood and C. Kuehn, *Assessing the impact of parametric uncertainty on tipping points of the Atlantic meridional overturning circulation*, Environ. Res. Lett. **17** (2022), no. 7, 075002.

[3] R.A. Wood, J.M. Rodríguez, R.S. Smith et al., *Observable, low-order dynamical controls on thresholds of the Atlantic meridional overturning circulation*, Clim. Dyn. **53** (2019), 6815–6834.

[4] O.G. Ernst, A. Pichler and B. Sprungk, *Wasserstein sensitivity of risk and uncertainty propagation*, SIAM/ASA J. Uncertain. Quantif. **10** (2022), no. 3, 915–948.

Numerical methods for the SPDE approach in space–time

Kristin Kirchner

(joint work with Joshua Willems)

Most environmental data sets contain measurements collected over space and time. It is the purpose of spatiotemporal statistical models to adequately describe the underlying uncertain spatially explicit phenomena evolving over time. In this context, Gaussian processes play a central role either as *prior distributions* or as components in *hierarchical models* to describe non-Gaussian dependencies.

Since a Gaussian process $(X(j))_{j\in\mathcal{I}}$ is fully characterised by its mean and its covariance function, second-order-based approaches focus on the construction of appropriate covariance classes. For processes indexed by a spatial domain in the Euclidean space $\mathcal{I} = \mathcal{D} \subseteq \mathbb{R}^d$, the *Matérn covariance class*,

$$\varrho(x,y) = 2^{1-\nu}\sigma^2[\Gamma(\nu)]^{-1}(\kappa\|x-y\|_{\mathbb{R}^d})^\nu K_\nu(\kappa\|x-y\|_{\mathbb{R}^d}), \quad x,y\in\mathcal{D}, \tag{1}$$

is an important and widely used model. Here, K_ν denotes the modified Bessel function of the second kind, and the three parameters $\nu,\kappa,\sigma^2\in(0,\infty)$ determine *smoothness*, *correlation length* and *variance* of the process. The interpretability of these parameters renders this covariance class particularly suitable for making inference about spatial data.

When considering *spatiotemporal* phenomena, the following difficulties occur:

- It is desirable to control the properties of the stochastic process named above (in particular, smoothness and correlation lengths) separately in space and time. For this reason, considering (1) in $d+1$ dimensions is not expedient and it is a difficult task to construct appropriate spatiotemporal covariance function classes.

- Second-order-based approaches require the factorisation of, in general, dense covariance matrices, causing computational costs which are cubic in the number of observations. The two common assumptions imposed on spatiotemporal covariance models to reduce the computational costs, namely separability (factorisation into merely spatial and temporal covariance functions) and stationarity (invariance under translations), have proven unrealistic in many situations.

In the past decade, the idea of representing Gaussian processes as solutions to appropriate stochastic partial differential equations (SPDEs) has gained popularity. More specifically, the (spatial) SPDE approach is motivated by the observation that a stationary process $(X(x))_{x\in\mathcal{D}}$ indexed by the entire Euclidean space $\mathcal{D} = \mathbb{R}^d$ which solves the SPDE

$$\left(\kappa^2 - \Delta\right)^{\beta} X(x) = \mathcal{W}(x), \quad x \in \mathcal{D}, \tag{2}$$

has a covariance function of Matérn type (1) with $\nu = 2\beta - \frac{d}{2}$. Here, Δ denotes the Laplacian and $\mathcal{W}$ is Gaussian white noise. This relation gave rise to the SPDE approach proposed by Lindgren, Rue, and Lindström [1], where the SPDE (2) is considered on a bounded domain $\mathcal{D} \subset \mathbb{R}^d$ and augmented with Dirichlet or Neumann boundary conditions. Besides enabling the applicability of efficient numerical methods available for (S)PDEs, such as finite element methods or wavelets, this approach has the advantage of allowing for more general domains, such as manifolds, and nonstationary or anisotropic generalizations by replacing $\kappa^2 - \Delta$ in (2) with more general strongly elliptic second-order differential operators,

$$(Lv)(x) = \kappa^2(x)v(x) - \nabla \cdot (a(x)\,\nabla v(x)), \quad x \in \mathcal{D}, \tag{3}$$

where $\kappa\colon \mathcal{D} \to \mathbb{R}$ and $a\colon \mathcal{D} \to \mathbb{R}^{d\times d}_{\mathrm{sym}}$ are functions.

In the SPDE (2) the fractional exponent β defines the (spatial) differentiability of its solution. A realistic description of spatiotemporal phenomena necessitates controllable differentiability in space and time. This motivates to introduce the space–time fractional SPDE model

$$\begin{cases} \left(\partial_t + L^{\beta}\right)^{\gamma} X(t,x) = \dot{\mathcal{W}}(t,x), & t \in [0,T], \quad x \in \mathcal{D}, \\ X(0,x) = 0, & x \in \mathcal{D}, \end{cases} \tag{4}$$

where L in (3) is augmented with boundary conditions on $\partial\mathcal{D}$, $\dot{\mathcal{W}}$ denotes space–time Gaussian white noise, and $T \in (0,\infty)$ is the time horizon. Notably, it is the interplay of the two fractional exponents β and γ that facilitates controlling spatial and temporal smoothness of the solution process.

We use the method of semigroups to interpret (4) as a fractional parabolic stochastic evolution equation, and correspondingly introduce solution concepts for it. To this end, we first give a meaning to negative fractional powers of a parabolic operator of the form $\mathcal{B} := \partial_t + A$, where $-A\colon \mathsf{D}(A) \subseteq H \to H$ generates a C_0-semigroup $(S(t))_{t\geq 0}$ on a separable Hilbert space H. More specifically, we exploit

the semigroup representation of $\mathcal{B}^{-\gamma}$ given by

$$\big[\mathcal{B}^{-\gamma} f\big](t) = \frac{1}{\Gamma(\gamma)} \int_0^t (t-s)^{\gamma-1} S(t-s) f(s)\,\mathrm{d}s, \quad f \in L^2(0,T;H),$$

to define, for $\gamma \in (\frac{1}{2}, \infty)$, mild solutions to problems of the form (4) via

$$Z_\gamma(t) = \frac{1}{\Gamma(\gamma)} \int_0^t (t-s)^{\gamma-1} S(t-s)\,\mathrm{d}W^Q(s), \quad t \in [0,T],$$

where $\dot{W}^Q$ is an H-valued Q-Wiener process, for some $Q \in \mathcal{L}(H)$. We then investigate their existence, uniqueness, regularity and covariance. Our main findings [2] show that the problem (4) is well-posed, and the properties of its solution with respect to smoothness and covariance structure generalise those of the spatial SPDE model (2) and relate to the parameters $\beta, \gamma \in (0, \infty)$ in the desired way.

Furthermore, we discuss the efficient approximation of covariance operators corresponding to the so-defined spatiotemporal Gaussian processes. The numerical methods are based on space-time finite element discretizations of fractional parabolic operators, supported by a rigorous error analysis using inf-sup theory.

References

[1] F. Lindgren, H. Rue, and J. Lindström, *An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach*, Journal of the Royal Statistical Society: Series B (Statistical Methodology) **73** (2011), 423–498.

[2] K. Kirchner and J. Willems, *Regularity theory for a new class of fractional parabolic stochastic evolution equations*, Stochastics and Partial Differential Equations: Analysis and Computations **12** (2024), 1805–1854.

Data-Efficient Kernel Methods for Learning Differential Equations and Their Solution Operators: Algorithms and Error Analysis

Houman Owhadi

(joint work with Yasamin Jalalian, Juan Felipe Osorio Ramirez, Alexander Hsu, and Bamdad Hosseini)

We introduce a novel kernel-based framework for learning differential equations and their solution maps that is efficient in data requirements, in terms of solution examples and amount of measurements from each example, and computational cost, in terms of training procedures. Our approach is mathematically interpretable and backed by rigorous theoretical guarantees in the form of quantitative worst-case error bounds for the learned equation. Numerical benchmarks demonstrate significant improvements in computational complexity and robustness while achieving one to two orders of magnitude improvements in terms of accuracy compared to state-of-the-art algorithms. In comparison to equivalent neural net methods, our approach is significantly more robust to the choice of hyperparameters and does not require close human supervision during training.

References

[1] Y. Jalalian, J. F. Osorio Ramirez, A. Hsu, B. Hosseini, and H. Owhadi, *Data-Efficient Kernel Methods for Learning Differential Equations and Their Solution Operators: Algorithms and Error Analysis*, arXiv:2503.01036 [stat.ML] (2025).

Why we should care about Variational Bayes

Håvard Rue

(joint work with M. E. Khan, J. van Niekerk, and S. Dutta)

In this talk I discussed Variational Bayes (VB), what it is, why we should care, and why it should be an integral part of any (modern) statistician's toolbox. I discussed our use of it in the R-INLA project, but also its role in the 'Bayesian Learning Rule' from which a wide range of algorithms can be derived from: ridge regression, Newton's method, Kalman filter, stochastic-gradient descent, RMSprop, and Dropout, Laplace's method, EM and so on.

The use of VB within the R-INLA project is discussed in [1] for the mean correction, while [2] discusses variance and skewness correction. The *Bayesian Learning Rule* is presented in [3].

References

[1] J. van Niekerk and H. Rue, *Low-rank variational Bayes correction to the Laplace method*, Journal of Machine Learning Research **25**(62) (2024), 1–25. URL `https://jmlr.org/papers/v25/21-1405.html`.

[2] S. Dutta, J. van Niekerk, and H. Rue, *Scalable skewed Bayesian inference for latent Gaussian model*, Submitted (2025). URL `https://arxiv.org/abs/2502.19083`.

[3] M. E. Khan and H. Rue, *The Bayesian learning rule*, Journal of Machine Learning Research **24** (2023), 1–46. URL `https://jmlr.org/papers/volume24/22-0291/22-0291.pdf`.

An approach to infinite-dimensional Bayesian inversion for stochastic problems

Andrea Barth, Oliver König

In numerous applications necessitating the estimation of a quantity of interest, direct observation is not a viable method. Consequently, an alternative quantity is observed that is associated with the quantity of interest. The problem of inferring the quantity of interest from this observation gives rise to the field of inverse problems. In a more general sense, inverse problems can be understood as data-driven model fitting problems, thus arising in numerous applications in the sciences, engineering, and finance. Inverse problems are formally defined as the inversion of well-posed problems, which are often referred to as forward problems. However, it has been observed that the process of inversion of a well-posed problem frequently results in an ill-posed problem. Given the ill-posedness of inverse problems, it is not possible to expect an exact reconstruction of the quantity of interest. Therefore, methods for quantifying the uncertainty associated with the reconstruction

are required. To address this issue, Bayesian methods construct a probability distribution that allows for the reconstruction of the quantity of interest and the quantification of the associated uncertainties. Here, we consider a generalisation of a Bayesian inverse problem to allow for infinite-dimensional input and output spaces that are necessary to consider inherently stochastic forward maps, lifting the standard assumption of additive, independent noise, allowing for a broader class of stochastic models of the forward problem capturing both aleatoric and epistemic uncertainties. This formulation allows the consideration of inverse problems based on stochastic processes and random fields as well as stochastic and random (partial) differential equations. The definition of a consistent finite-dimensional approximation is then the essential tool. This is achieved by the definition of a family of finite-dimensional projections of the forward model, that is associated to a filtration which may be interpreted as the information gain, leading to an approximate sequence of posterior measures.

Let the input space $\mathcal{X}$ and the output space $\mathcal{Y}$ be both (infinite dimensional) separable Banach spaces equipped with their Borel σ-algebras and let X resp. Y be random variables (defined on a complete probability space $(\Omega, \mathcal{A}, \mathbb{P})$) modelling the uncertain input resp. output taking values in $\mathcal{X}$ resp. $\mathcal{Y}$. In the absence of densities the posterior may still be defined using regular conditional probabilities although the posterior density can, in general, not be accessed by Bayes' theorem even if the posterior density exists. The goal is therefore to construct an approximation of the stochastic forward map to obtain an approximation of the posterior. To this end, we define for each $n \in \mathbb{N}$ by $\pi^{(n)} \colon \mathcal{Y} \to \mathcal{Y}$ a measurable mapping such that the sequence $(\pi^{(n)}, n \in \mathbb{N})$ converges pointwise to the identity mapping on $\mathcal{Y}$. This directly translates to the sequence $(Y^{(n)} := \pi^{(n)}(Y), n \in \mathbb{N})$ converging pointwise to Y. The family of projections $\pi := (\pi^{(n)}, n \in \mathbb{N})$ has to be chosen such that the family of generated σ-algebras $(\sigma(Y^{(n)}), n \in \mathbb{N})$ defines a filtration on $(\Omega, \mathcal{A}, \mathbb{P})$. Denote by $(\mathcal{F}_n, n \in \mathbb{N})$ the augmentation of that filtration, we then can show convergence of the corresponding posterior distributions. We deduct this from the fact that for every Borel set $A \in \mathcal{B}(\mathcal{X})$ it holds that for the conditional expectations

$$\mathbb{E}(1_A(X)|\mathcal{F}_n) \to \mathbb{E}(1_A(X)|\mathcal{F}_\infty)$$

$\mathbb{P}$-almost surely as $n \to \infty$. This is the foundation of the convergence of the approximated posterior distribution: Let $(\mathbb{P}(\cdot|y), y \in \mathcal{Y})$ denote the regular conditional probability of (Y, X) and for $n \in \mathbb{N}$, let $(\mathbb{P}^{(n)}(\cdot|\pi^{(n)}(y)), y \in \mathcal{Y})$ denote the regular conditional probability of $(Y^{(n)}, X)$. Then, for every Borel set $A \in \mathcal{B}(\mathcal{X})$, it holds

$$\mathbb{P}^{(n)}(A|\pi^{(n)}(y)) \to \mathbb{P}(A|y) \qquad \text{for } n \to \infty.$$

This proves convergence of the approximation in the number of observation points and, furthermore, the existence of moments can be guaranteed as well as stability in the most common distances. Numerous examples encompassing the reconstruction of coefficient functions of a stochastic (partial) differential equation from path observations and finding the parameters of a covariance kernel of a

Gaussian random field from a realization with a varying number of observations are presented.

PDEs on Gevrey regular random domain deformations: forward and inverse problems

Ana Djurdjevac
(joint work with V. Kaarnioja, M. Orteu, C. Schillings, and A. Zepernick)

We study uncertainty quantification for partial differential equations (PDEs) subject to domain uncertainty. Random domains naturally appear in various applications such as biology and imaging. The numerical consideration of these type of problems has been very popular [6, 9, 1, 7].

The starting point of the problem is to specify how is the random domain defined. In this work we consider the so-called domain mapping method, where it is assumed that the random domain is given via a random flow that connect the fixed domain and its realization. Next the problem on the random domain is pulled-back to the fixed deterministic domain, resulting in a PDE on a fixed domain with random coefficients.

The first question that appears is what type of random deformations do we consider, in particular with respect to the random parameter $y \in U$. Motivated by the recent work [2, 3], we consider the Gevrey class of deformations that contains smooth, but not necessarily holomorphic, functions with a growth condition on the higher-order partial derivatives. More precisely, the *Gevrey regularity of the perturbation field* $V(\cdot, y)$ assumes that it is infinitely many times continuously differentiable with respect to parameter $y \in U$ and there exists a constant $C \geq 1$, an exponent $\beta \geq 1$, and a sequence $b = (b_j)_{j \geq 1}$ of non-negative numbers such that

$$\|\partial_y^\nu V(\cdot, y)\|_{W^{1,\infty}(D_{\mathrm{ref}})} \leq C(|\nu|!)^\beta b^\nu \quad \text{for all } \nu \in \mathcal{F}, y \in U,$$

where $\mathcal{F}$ is the set of finitely-supported multi-indices. Harbrecht et al. [8] showed that such Gevrey regularity is preserved for a broad class of operator equations. While some of their analysis applies to the stationary PDE problem we consider, their focus is not on numerical methods, and certain constants in their estimates are not explicitly tracked. This approach has the advantage of being substantially more general than models which assume a particular parametric representation of the input random field such as a Karhunen–Loève series expansion.

We consider both the Poisson equation as well as the heat equation for which we consider the space–time weak formulation. Next we design randomly shifted lattice quasi-Monte Carlo (QMC) cubature rules for the computation of the expected solution under domain uncertainty and consider approximation errors. We develop a novel parametric regularity analysis for these problems, which is used to design tailored QMC cubature rules with essentially linear convergence rates independently of the truncation dimension. In particular, for the Poisson equation

we obtain the regularity of the type

$$\|\partial_y^\nu \hat{u}(\cdot, y)\|_{H_0^1(D_{\mathrm{ref}})} \leq C_{\hat{u},1} C_{\hat{u},2}^{|\nu|} (|\nu|!)^\beta b^\nu,$$

where $C_{\hat{u},1}$ and $C_{\hat{u},2}$ are constants that we explicitly compute. The analogue result holds for the parabolic case. Based on this regularity results, we prove the main result that gives a choice of product and order dependent weights as well as a QMC error bound in the sense of the root-mean square error independent of the dimension of parameter truncation. These results are joint work with V. Kaarnioja, C. Schillings and A. Zepernick, presented in [4].

Based on these results we also consider Bayesian shape inversion subject to the Poisson equation under Gevrey regular parameterisations of domain uncertainty. More precisely, we consider the measurement model that is described by the solution u of the Poisson problem on the random domain and we consider Bayesian inverse problem of inferring the domain shape based on measurements of certain observable quantity of interest of u. We study the parametric regularity of the associated posterior distribution and construct randomly shifted rank-1 lattice rules that yield dimension-independent convergence rates faster than standard Monte Carlo for high-dimensional integrals over the posterior. Additionally, we examine the impact of dimension truncation and finite element discretisation errors within this model. This is a joint work with V. Kaarnioja, M. Orteu and C. Schillings [5].

References

[1] J. E. Castrillón-Candás, F. Nobile, R. Tempone, *Analytic regularity and collocation approximation for elliptic PDEs with random domain deformations*, Comput. Math. Appl. **71(6)** (2016), 1173–1197.

[2] A. Chernov, T. Lê, *Analytic and Gevrey class regularity for parametric elliptic eigenvalue problems and applications*, SIAM J. Numer. Anal. **S4**, 62 (2024), 1874–1900.

[3] A. Chernov, T. Lê, *Analytic and Gevrey class regularity for parametric semilinear reaction-diffusion problems and applications in uncertainty quantification*, Comput. Math. Appl. **164** (2024), 116–130.

[4] A. Djurdjevac, V. Kaarnioja, C. Schillings, A. Zepernick, *Uncertainty quantification for stationary and time-dependent PDEs subject to Gevrey regular random domain deformations*, arXiv preprint arXiv:2502.12345 (2025).

[5] A. Djurdjevac, V. Kaarnioja, M. Orteu, C. Schillings, *Quasi-Monte Carlo for Bayesian shape inversion governed by the Poisson problem subject to Gevrey regular domain deformations*, arXiv preprint arXiv:2502.14661 (2025).

[6] J. Dölz, H. Harbrecht, C. Jerez-Hanckes, M. Multerer, *Isogeometric multilevel quadrature for forward and inverse random acoustic scattering*, Comput. Methods Appl. Mech. Engrg. **388**, (2022), 114–242.

[7] H. Harbrecht, M. Peters, M. Siebenmorgen, *Analysis of the domain mapping method for elliptic diffusion problems on random domains*, Numer. Math. **134(4)** (2016), 823–856.

[8] H. Harbrecht, M. Schmidlin, Ch. Schwab, *The Gevrey class implicit mapping theorem with application to UQ of semilinear elliptic PDEs*, Math. Models Methods Appl. Sci. **34(5)** (2024), 881–917.

[9] D. Xiu, D. M. Tartakovsky, *Numerical methods for differential equations in random domains*, SIAM J. Sci. Comput. **28(3)** (2006), 1167–1185.

Uncertainty quantification in machine learning: from aleatoric to epistemic

Eyke Hüllermeier
(joint work with Viktor Bengs and Willem Waegeman)

Questions regarding the representation and adequate handling of (predictive) uncertainty have recently received increasing attention in machine learning research. In this regard, a particular focus is put on the distinction between two important types of uncertainty, often referred to as aleatoric and epistemic, and the question of how to quantify these uncertainties in terms of appropriate numerical measures. Roughly speaking, while aleatoric uncertainty is due to the randomness inherent in the data-generating process, epistemic uncertainty is caused by the learner's ignorance of the true underlying model [4].

Consider a basic task of learning a classifier in a supervised manner: Given training data

$$\mathcal{D} = \big\{(\boldsymbol{x}_i, y_i)\big\}_{i=1}^{N} \subset \mathcal{X} \times \mathcal{Y}\,,$$

typically assumed to be i.i.d. according to some (unknown) measure P on $\mathcal{X} \times \mathcal{Y}$, a hypothesis space $\mathcal{H}$ of predictors $\mathcal{X} \longrightarrow \mathbb{P}(\mathcal{Y})$ to choose from, and a loss function $L : \mathcal{Y} \times \mathbb{P}(\mathcal{Y}) \longrightarrow \mathbb{R}$, the learner seeks to find

$$h^* \in \operatorname*{argmin}_{h \in \mathcal{H}} R(h)\,,$$

i.e., a predictor $h : \mathcal{X} \longrightarrow \mathbb{P}(\mathcal{Y})$ with minimal risk (expected loss)

$$R(h) \doteq \mathbb{E}_{(\boldsymbol{x},y)\sim P} L(y, h(\boldsymbol{x})) = \int_{\mathcal{X}\times\mathcal{Y}} L(y, h(\boldsymbol{x}))\, d\, P(\boldsymbol{x}, y)\,.$$

Here, $\mathcal{X}$ is the so-called instance space, $\mathcal{Y}$ a finite class of categories, and $\mathbb{P}(\mathcal{Y})$ denotes the set of probability distributions on $\mathcal{Y}$.

Probabilistic predictors $h : \mathcal{X} \longrightarrow \mathbb{P}(\mathcal{Y})$ are able to capture aleatoric but no epistemic uncertainty. Therefore, the learning of second-order predictors $H : \mathcal{X} \longrightarrow \mathbb{P}(\mathbb{P}(\mathcal{Y}))$ has recently been considered. By predicting a probability distribution over probability distributions (on the outcome space $\mathcal{Y}$), a second-order predictor H can express uncertainty about the "right" (ground-truth) first-order distribution, and thereby represent epistemic uncertainty.

In principle, second-order predictors could be learned in a Bayesian way, which, however, requires the specification of a prior (over extremely complex hypothesis spaces) and is often computationally intractable. As an alternative, the idea of learning a second-order predictor in a more "direct" way has recently been advocated, namely by following the standard principle of empirical risk minimisation [6]: Given training data $\mathcal{D} = \big\{(\boldsymbol{x}_i, y_i)\big\}_{i=1}^{N} \subset \mathcal{X} \times \mathcal{Y}$, the idea is to train a second-order predictor by minimising the empirical risk (loss on the training data)

$$R_{emp}(H) = \sum_{i=1}^{N} L_E\big(H(\boldsymbol{x}_i), y_i\big)\,,$$

with a suitable second-order (epistemic) loss function

$$L_E : \mathbb{P}(\mathbb{P}(\mathcal{Y})) \times \mathcal{Y} \longrightarrow \mathbb{R} \,.$$

This approach can be motivated by its first-order counterpart: Training a (first-order) probabilistic predictor h via empirical risk minimisation, i.e.,

$$h \in \operatorname*{argmin}_{g \in \mathcal{H}} \sum_{i=1}^{N} L_A \left(g(\boldsymbol{x}_i), y_i \right) \,,$$

yields (theoretically) unbiased predictors if L_A is a (strictly) proper scoring rule [3]. A loss L_A is a proper scoring rule if the following holds for $Y \sim p$:

$$\mathbb{E}_{Y \sim p}[\, L_A(p, Y) \,] \leq \mathbb{E}_{Y \sim p}[\, L_A(\hat{p}, Y) \,]$$

for all distributions $\hat{p} \in \mathbb{P}(\mathcal{Y})$ (and a strictly proper scoring rule if the inequality is strict for $\hat{p} \neq p$). This can be interpreted as follows: If the learner knows (or believes) that the outcome Y is sampled according to a distribution p, and furthermore that predictions are penalised according to the loss L_A, then it must predict p in order to minimise loss in expectation. In other words, a proper scoring rule L_A incentivises the learner to predict the ground-truth (conditional) probabilities.

Despite this motivation, we recently obtained a series of negative results, showing that second-order loss minimisation is theoretically flawed [1, 2, 5]. Somewhat simplified, a loss L_E incentivising the learner to make meaningful second-order predictions Q cannot exist. To make this more rigorous, we generalised the notion of proper scoring rule to proper second-order scoring rule: A second-order loss L_E (such that $L_E(Q, \cdot)$ is $\mathbb{P}(\mathbb{P}(\mathcal{Y}))$-quasi-integrable for all $Q \in \mathbb{P}(\mathbb{P}(\mathcal{Y}))$) is a proper second-order scoring rule if, for all $\hat{Q}, Q \in \mathbb{P}(\mathbb{P}(\mathcal{Y}))$,

$$\mathbb{E}_{p \sim Q}\Big[\, \mathbb{E}_{Y \sim p}[\, L_E(Q, Y) \,] \Big] \leq \mathbb{E}_{p \sim Q}\Big[\, \mathbb{E}_{Y \sim p}[\, L_E(\hat{Q}, Y) \,] \Big] \,.$$

This can be interpreted as follows: If the learner holds "second-order belief" Q and is penalised according to L_E, then it should report $\hat{Q} = Q$ as the (double-)expected loss-minimising prediction. We could show, however, that a second-order scoring rule L_E that is proper in this sense (i.e., guaranteed to satisfy the above inequality) cannot exist [2].

An alternative second-order representation is in terms of a set (instead of distribution) of probability distributions (on $\mathcal{Y}$), a so-called credal set [7]. It is an interesting question to what extent our results for second-order distributions carry over to credal sets, and this is something that we will explore in future work.

References

[1] V. Bengs, E. Hüllermeier, and W. Waegeman. *Pitfalls of epistemic uncertainty quantification through loss minimisation.* Proc. NeurIPS, 2022.

[2] V. Bengs, E. Hüllermeier, and W. Waegeman. *On second-order scoring rules for epistemic uncertainty quantification.* Proc. ICML, 2023.

[3] *Strictly Proper Scoring Rules, Prediction, and Estimation.* Technical report 463R. Department of Statistics, University of Washington, 2005.

[4] E. Hüllermeier and W. Waegeman. *Aleatoric and epistemic uncertainty in machine learning: An introduction to concepts and methods*, Machine Learning, 110(3):457–506, 2021.
[5] M. Jürgens, V. Bengs, N. Meinert, E. Hüllermeier, and W. Waegeman. *Is epistemic uncertainty faithfully represented by evidential deep learning methods?*, Proc. ICML 2024.
[6] M. Sensoy and L. Kaplan and M. Kandemir. *Evidential Deep Learning to Quantify Classification Uncertainty*, Proc. NeurIPS, 2018.
[7] M.H. Shaker and E. Hüllermeier. *Ensemble-based Uncertainty Quantification: Bayesian versus Credal Inference*, arXiv:2107.10384, 2021.

Statistical Algorithms for Low-Frequency Diffusion Data: A PDE Approach

Sven Wang
(joint work with Matteo Giordano)

This extended abstract summarises the main contributions of our work [1], which develops new computational techniques for statistical inference from low-frequency diffusion data. We consider the problem of statistical inference for multi-dimensional diffusion processes from low-frequency observations. In this setting, traditional likelihood-based methods are notoriously difficult to implement due to the intractability of the transition densities and their gradients. Motivated by these challenges, we develop a novel computational approach that builds on the theory of partial differential equations (PDEs) and leverages spectral techniques for elliptic operators.

Our approach is based on the characterisation of the transition densities of the underlying reflected diffusion process as solutions of the associated Fokker–Planck equation with Neumann boundary conditions. Using regularity results from parabolic PDE theory [5], we derive a new representation for the gradient of the likelihood with respect to the unknown diffusivity function. This representation expresses the derivative through a variation-of-constants formula, see also [3], and allows us to avoid the need for costly data augmentation schemes often employed in the analysis of low-frequency diffusion data.

Crucially, both the transition densities and their gradients can be approximated via the spectral decomposition of the elliptic generator of the diffusion, a self-adjoint operator in divergence form. This reduces the problem to the numerical solution of standard elliptic eigenvalue problems, for which efficient finite element solvers are available. Our approach thus enables the use of a wide range of statistical algorithms, including gradient-based optimisation methods and gradient-informed Markov chain Monte Carlo (MCMC) samplers.

We demonstrate these developments in a nonparametric Bayesian framework using Gaussian process priors [4]. The resulting algorithms allow for the computation of maximum likelihood and maximum a posteriori estimates, posterior means, and quantiles, all without resorting to trajectory simulation or latent variable augmentation. In extensive simulation studies on a two-dimensional domain, our methods show excellent empirical performance, providing accurate reconstruction of the diffusivity function and competitive runtimes even at high sample sizes.

Our work opens up several avenues for future research. These include extensions to diffusions with non-divergence form structure, models with noisy observations, and sampling on unbounded domains. Moreover, the PDE-based gradient characterisation may pave the way for a theoretical analysis of the computational complexity of the employed statistical algorithms, such as proving stability bounds and polynomial-time computability [2].

References

[1] Giordano, M. and Wang, S. (2025). Statistical algorithms for low-frequency diffusion data: A PDE approach. *The Annals of Statistics*, to appear.

[2] Nickl, R. and Wang, S. (2024). On polynomial-time computation of high-dimensional posterior measures by Langevin-type algorithms. *Journal of the European Mathematical Society*, **26**, 1031–1112.

[3] Wang, S. (2019). The nonparametric LAN expansion for discretely observed diffusions. *Electronic Journal of Statistics*, **13**, 1329–1358.

[4] Nickl, R. (2024). Consistent inference for diffusions from low frequency measurements. *The Annals of Statistics*, **52**, 519–549.

[5] Lunardi, A. (1995). *Analytic Semigroups and Optimal Regularity for Parabolic Problems*. Birkhäuser.

Defining Modes and MAP Estimators: A Systematic Approach

Ilja Klebanov

(joint work with Hefin Lambley and T. J. Sullivan)

What does it mean to be the "most likely" outcome of a probability measure $\mu \in \mathcal{P}(X)$ on a metric space (X, d)? Such "modes" naturally arise as MAP estimators in Bayesian inverse problems, yet multiple competing definitions exist. These definitions reflect different ways of characterising the asymptotic behaviour of probability mass in vanishingly small neighborhoods. Instead of examining specific mode definitions, we introduce a systematic framework based on *mode maps* $\mathfrak{M}\colon \mathsf{MetProb} \to \mathsf{Set}$, which map metric probability spaces to sets of modes: $\mathfrak{M}(\mu) = \mathfrak{M}(X, d, \mu) = \{\text{modes of } \mu\} \subseteq X$. Guided by intuitive axioms that ensure consistency across discrete and continuous settings, we identify exactly ten well-justified definitions and explore their interactions, coincidences, and simplifications in well-behaved cases. The details of this work can be found in [4].

Definition 1 (Axioms for mode maps)**.** A mode map $\mathfrak{M}\colon \mathsf{MetProb} \to \mathsf{Set}$ satisfies

(LP) the **Lebesgue property** if, whenever $X = \mathbb{R}^m$ and $\mu \in \mathcal{P}(\mathbb{R}^m)$ has continuous Lebesgue density ρ, $\mathfrak{M}(\mu) = \arg\max_{x \in X} \rho(x)$;

(AP) the **atomic property** if, for any metric space X, $K \in \mathbb{N}$, non-atomic measure $\mu_0 \in \mathcal{P}(X)$, pairwise distinct $x_1, \ldots, x_K \in X$, $\alpha_0, \ldots, \alpha_K \in [0, 1]$ with $\sum_{k=0}^{K} \alpha_k = 1$ and $\alpha_0 \neq 1$,

$$\mathfrak{M}\left(\alpha_0 \mu_0 + \sum_{k=1}^{K} \alpha_k \delta_{x_k}\right) = \left\{ x_k \,\middle|\, k = 1, \ldots, K \text{ with } \alpha_k = \max_{j=1,\ldots,K} \alpha_j \right\};$$

(CP) the **cloning property** if, for any normed space X, $\mu \in \mathcal{P}(X)$, $\alpha \in [0,1]$ and $b \in X$ such that $\operatorname{dist}(\operatorname{supp}(\mu), b + \operatorname{supp}(\mu)) > 0$,

$$\mathfrak{M}\big(\alpha\mu + (1-\alpha)\mu(\,\cdot\, - b)\big) = \begin{cases} \mathfrak{M}(\mu), & \text{if } \alpha > \frac{1}{2}, \\ \mathfrak{M}(\mu) \cup (\mathfrak{M}(\mu) + b), & \text{if } \alpha = \frac{1}{2}, \\ \mathfrak{M}(\mu) + b, & \text{if } \alpha < \frac{1}{2}. \end{cases}$$

If a measure lacks atoms or a Lebesgue density (in particular, in infinite-dimensional spaces X), defining a mode becomes difficult and ambiguous. Instead of using densities, a common approach is to compare the mass of balls $B_r(x)$ as the radius $r > 0$ shrinks to zero. The ratio of such ball masses,

$$\mathfrak{R}^\mu_r(u,v) \doteq \frac{\mu(B_r(u))}{\mu(B_r(v))}, \qquad \mathfrak{R}^\mu_r(u,\sup) \doteq \frac{\mu(B_r(u))}{\sup_{x\in X} \mu(B_r(x))}, \qquad u, v \in X,$$

is then a crucial concept. Modes can be defined as points with the "heaviest small balls", leading to the following well-established definitions:

Definition 2. For $\mu \in \mathcal{P}(X)$, a point $u \in X$ is called a

(a) **strong mode** ($\mathfrak{s}$-mode, [2]) if $\liminf_{r\to 0} \mathfrak{R}^\mu_r(u,\sup) \geq 1$;
(b) **weak mode** ($\mathfrak{w}$-mode, [3]) if, for every **comparison point** (cp) $v \neq u$, $\liminf_{r\to 0} \mathfrak{R}^\mu_r(u,v) \geq 1$;
(c) **generalized strong mode** ($\mathfrak{gs}$-mode, [1]) if, for every **null sequence** (ns) $r_n \to 0$, there exists an **approximating sequence** (as) $u_n \to u$ such that $\liminf_{n\to\infty} \mathfrak{R}^\mu_r(u_n,\sup) \geq 1$.

For strong modes, the μ-probability of the ball $B_r(u)$ asymptotically dominates the supremal ball mass $\sup_{x\in X} \mu(B_r(x))$. In contrast, weak modes dominate $B_r(v)$ for every *comparison point* (cp) $v \neq u$ separately. Generalised strong modes allow the ball mass to dominate along some *approximating sequence* (as) $u_n \to u$.

Similarly, one can define generalised weak modes ($\mathfrak{gw}$-modes). However, since the adjectives *weak* and *generalised* correspond to quantifiers $\forall\mathsf{cp}$ and $\exists\mathsf{as}$, their order becomes an issue. We therefore distinguish between generalised weak ($\mathfrak{gw}$) and weak generalised ($\mathfrak{wg}$) modes, leaving the precise definitions to the reader. Given our definitions so far, several natural questions arise:

- Should the comparison point v have its own **comparison sequence** (cs) $v_n \to v$, introducing the quantifier *"approximating"* $= \mathfrak{a} = \forall\mathsf{cs}$?
- Should u be required to be dominant only along *some* ns r_n, rather than every ns, leading to the quantifier *"partial"* $= \mathfrak{p} = \exists\mathsf{ns}$?

While these new quantifiers lead to seemingly meaningful definitions, they introduce a whole zoo of notions due to possible permutations of quantifier order, making the relationships between them complex. Moreover, this "letter notation" [4, Definition 4.1], based on the alphabet $\{\mathfrak{s}, \mathfrak{w}, \mathfrak{g}, \mathfrak{a}, \mathfrak{p}\}$, still lacks some potentially meaningful definitions:

- What about the quantifiers $\forall\mathsf{as}$ and $\exists\mathsf{cs}$?

- While the quantifier $\mathfrak{p} = \exists\mathsf{ns}$ can appear at different positions (e.g., $\mathfrak{wpag}$ or $\mathfrak{wgap}$), $\forall\mathsf{ns}$ (implicitly representing the absence of $\mathfrak{p}$) is fixed as the first quantifier, limiting flexibility.

To address these issues, we propose replacing the cumbersome "letter notation" with a more structured system that directly incorporates quantifiers into the mode (map) name:

Definition 3. For $\mu \in \mathcal{P}(X)$, $u \in X$ is called a $[Q_1 \dots Q_K]$**-mode** for a sequence $\mathcal{Q} = [Q_1 \dots Q_K]$ of quantifiers $Q_k \in \{\forall\mathsf{ns}, \exists\mathsf{ns}, \forall\mathsf{cp}, \exists\mathsf{cp}, \forall\mathsf{as}, \exists\mathsf{as}, \forall\mathsf{cs}, \exists\mathsf{cs}\}$, if

$$Q_1 \dots Q_K : \qquad \liminf_{n\to\infty} \mathfrak{R}^{\mu}_{r_n}(\star_1, \star_2) \geq 1,$$

where, recalling that the quantifiers $Q_1 \dots Q_K$ may define $(u_n)_{n\in\mathbb{N}}$, v, and $(v_n)_{n\in\mathbb{N}}$, $\star_1 \in \{u, u_n\}$ and $\star_2 \in \{v, v_n, \sup\}$ depend on whether as , cs and cp appeared in $\mathcal{Q}$. The sequence $\mathcal{Q}$ of quantifiers obeys certain rather obvious rules [4, Definition 4.9], and the corresponding mode map is denoted by $\mathfrak{M}[Q_1 \dots Q_K]$.

We reduce the large number of resulting definitions to a manageable set as follows [4, Propositions 4.13, 4.14, 4.15, 4.16; and Section 6], and illustrate the resulting Hasse diagram in Figure 1.

- Of the 282 definitions from Definition 3, many are trivially equivalent, as subsequent $\forall$-quantifiers (or $\exists$-quantifiers) commute (e.g., $[\forall\mathsf{ns}\forall\mathsf{cp}]$-modes and $[\forall\mathsf{cp}\forall\mathsf{ns}]$-modes coincide).
- Of the remaining 144 definitions, many violate at least one of the axioms (LP), (AP), (CP).
- Among the remaining 21 definitions, some are (non-trivially) equivalent.
- The final 10 definitions satisfy our axioms (LP), (AP), (CP) and are provably distinct (cf. Figure 1).
- All but one of these 10 definitions fit the letter notation above. The exception is the $[\forall\mathsf{cp}\forall\mathsf{cs}\exists\mathsf{as}\forall\mathsf{ns}]$-mode, termed the **exotic mode** ($\mathfrak{e}$-mode).

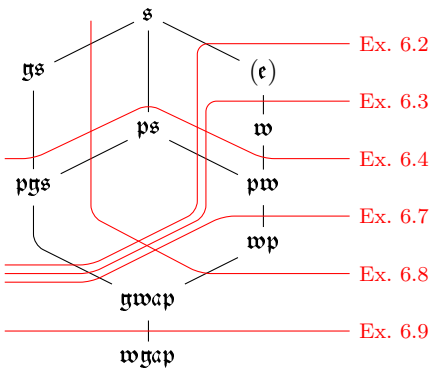


FIGURE 1. The remaining ten meaningful small-ball mode maps, where descending black lines indicate the implication structure. Each counterexample from [4, Section 6] separates the lattice into two subsets: the indicated example is a mode for all types below the red line but not for those above, ruling out further implications among the ten types.

In [4, Section 5], we study conditions under which certain mode maps coincide. As an example, we state the following result:

Theorem 1 (MAP estimators for Gaussian prior and continuous potential)**.** *Let X be a separable Banach space and $\mu \in \mathcal{P}(X)$ have a continuous potential with respect to a non-degenerate Gaussian prior μ_0. Then the ten meaningful small-ball modes of μ fall into three equivalence classes, namely*

$$\{\mathfrak{s}, \mathfrak{gs}\}, \quad \{\mathfrak{ps}, \mathfrak{pgs}\}, \quad \textit{and} \quad \{\mathfrak{e}, \mathfrak{w}, \mathfrak{pw}, \mathfrak{wp}, \mathfrak{gwap}, \mathfrak{wgap}\}.$$

These three classes are generally distinct, but coincide if the potential is either uniformly continuous or quadratically lower bounded.

Future work may include exploring additional axioms, order-theoretic properties of modes, and alternative mode definitions not covered by Definition 3.

References

[1] C. Clason, T. Helin, R. Kretschmann, and P. Piiroinen, *Generalized Modes in Bayesian Inverse Problems*, SIAM/ASA J. Uncertain. Quantification **7** (2019), no. 2, 652–684.

[2] M. Dashti, K. J. H. Law, A. M. Stuart, and J. Voss, *MAP estimators and their consistency in Bayesian nonparametric inverse problems*, Inverse Problems **29** (2013), no. 9, 095017.

[3] T. Helin and M. Burger, *Maximum a posteriori probability estimates in infinite-dimensional Bayesian inverse problems*, Inverse Problems **31** (2015), no. 8, 085009.

[4] I. Klebanov, H. Lambley, and T. J. Sullivan, *Classification of small-ball modes and maximum a posteriori estimators*, preprint, arXiv:2306.16278 [math.PR], 2025.

Bayesian Inference (for Inverse Problems) by Factorisation and Function Approximation, without Sampling

Colin Fox

(joint work with Lennart Golks)

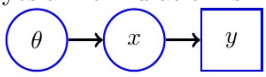
We present our ongoing work on evaluating posterior expectations, in Bayesian hierarchical formulations of inverse problems, using deterministic algorithms. In particular, we avoid sample-based Monte Carlo estimates of posterior expectations, commonly used in the fields of UQ (Uncertainty Quantification) and Statistics.

In doing so we are heeding the advice of the giants of the subject who preceded us; John von Neumann said that anyone implementing Monte Carlo is "living in a state of sin", while Hammersley & Handscomb advised that "it will usually pay to scrutinize each part of a Monte Carlo experiment to see whether that part cannot be replaced by exact theoretical analysis contributing no uncertainty". More recently, Alan Sokal emphasised that "Monte Carlo is an extremely bad method; it should be used only when all alternative methods are worse." We have taken these wise words to heart, and for some time we, with colleagues, have been building the tools required to allow posterior expectations to be evaluated in realistic, high-dimensional inverse problems, using only deterministic calculations. We report those methods and an example here.

Our experience in solving inverse problems, particularly in industrial settings where quantitative performance is paramount, and checked (!), has lead us to a view of what constitutes an 'inverse problem'. In common with all formulations, the observation process involves a complex physical mapping for which analytic

inversion presents difficulties, and the 'primary unknowns' are the unobserved physical properties of the system under study. When a stochastic model is used to represent the allowable physical properties, the unknowns are a 'latent field'. But then, in most current UQ analyses the latent field is modelled as a Gaussian random field, noise is additive Gaussian, forming the complete Bayesian model that is analysed. We have found that quantitative accuracy also requires Physics-based hierarchical modelling of hyperparameters appearing in the stochastic model for the latent field, to ensure consistency with a plausible physical process, and have also found that directly modelling objects or boundaries using mid- and high-level spatial models [6], such as as developed in the field of stochastic geometry [10], can dramatically improve quantitative performance and interpretability of recovered fields, as well reducing ill-posedness. Unfortunately, high-level representations are beyond our current ability to compute deterministically, but comprehensive hierarchical modelling is readily included, as we will see.

A simplified, though ubiquitous, DAG (directed acyclic graph) showing conditional dependencies in the Bayesian formulation is



where θ are hyper-parameters with hyperprior distribution $[\theta]$[1], x is the latent field with prior distribution $[x|\theta]$, and y is observed data with likelihood function $[y|x]$. Such a DAG is sufficient to explain our methods, while accommodating a physically realistic prior representation, a physical hierarchical model for hyperparameters, and validated/noninformative prior and hyperprior distributions.

The focus of inference is the posterior distribution

$$[x, \theta|y] = \frac{[y|x]\,[x|\theta]\,[\theta]}{[y]}$$

and we will assume throughout that the normalizing constant $[y]$ is finite. Our goal is often to compute posterior expectations

$$\mathbb{E}_{x,\theta|y}[f(x)] = \int f(x)[x, \theta|y]\,\mathrm{d}x\,\mathrm{d}\theta$$

for suitable functions f. The obvious difficulty is performing integration over high-dimensional latent field x and also hyperparameters θ.

As mentioned above, the most common current route is to approximate the integral, that defines the expectation, by the Monte Carlo estimate

$$\int f(x)[x, \theta|y)\,\mathrm{d}x\mathrm{d}\theta \approx \frac{1}{N}\sum_{i=1}^{N} f(x_i)$$

where $(x_i, \theta_i) \sim [x, \theta|y]$ are samples drawn from the posterior distribution, or, more generally $\{(x_i, \theta_i)\}$ is a sequence that is ergodic for $[x, \theta|y]$.

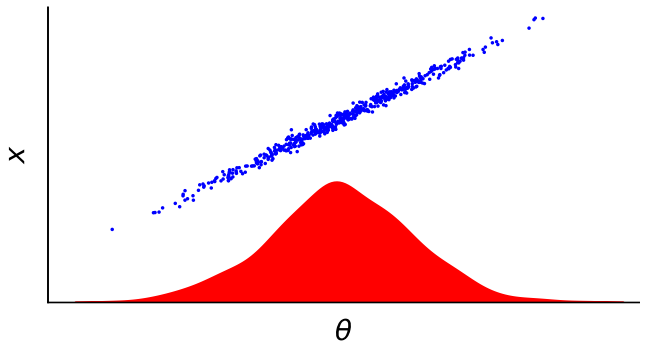
[1]We learned this notation from Alan Gelfand: read $[a]$ as "the distribution over a", and $[a|b]$ as "the distribution over a given b". We will shortly abuse this notation to also denote the density function.

There are many choices of algorithms to generate ergodic sequences, under the moniker of MCMC (Markov chain Monte Carlo) algorithms. Most common[2] are the random-walk Metropolis algorithms, including off-the-shelf Adaptive Metropolis (AM) and Hamiltonian Monte Carlo (HMC), that can be very slow for large-scale inverse problems, because of the ill-posedness that causes high posterior correlations, or weird geometry of state space.

A representative MCMC scheme is the block, or partially-collapsed, Gibbs sampler that updates (x, θ) to (x', θ') by alternating drawing from full conditionals

- Draw $\theta' \sim [\theta|x]$
- Draw $x' \sim [x|\theta', y]$

producing a transition kernel that targets the posterior $[x, \theta|y]$. By treating the latent field as a single block, block Gibbs is immune to the high correlations within the latent field, but remains very slow due to high correlation between the latent field x and hyperparameters θ that leads to slow mixing, because full conditionals are narrow, as indicated by the following stylised scatter plot of $[x, \theta|y]$.



As noted in [8, 4], better mixing is achieved by moving θ according to the marginal posterior distribution over hyperparameters, $[\theta|y] = \int_X [x, \theta|y]\, \mathrm{d}x$, also shown in the figure. Then one utilises the factorisation of the posterior density

$$[x, \theta|y] = [x|\theta, y]\, [\theta|y]$$

i.e. into the full conditional for x and the marginal posterior over θ.

The following Lemma shows how to sample from the posterior distribution.

Lemma 1. *Sampling $\theta' \sim [\theta|y]$ then $x' \sim [x|\theta', y]$ generates a sample from the posterior distribution, i.e.,*

$$(x', \theta') \sim [x, \theta|y].$$

As can be seen, one samples first from the marginal posterior over hyperparameters, then the full conditional over the latent field, a scheme that we call marginal then conditional (MTC) sampling when the first step draws independent θ'.

[2] as measured by papers I review

Various schemes may be used to sample $\theta \sim [\theta|y]$ (see [4]):

- Linchpin [1] updates θ using one step of a geometrically ergodic MCMC; convergence rate of the chain in (x, θ) is the same as the chain in θ.
- One-block [8] is a linchpin algorithm though requires x in the marginal posterior calculation, so has expensive cost per step.
- MTC [4] is a linchpin with independent θ drawn by many steps of a cheap MCMC, hence draws independent posterior samples.

More efficient again is using the function approximation methods developed in [9] to draw independent samples from $[\theta|y]$. Later we will use these same function representations to perform efficient *quadrature* over hyperparameters, thereby avoiding sampling completely.

The marginal posterior distribution over hyperparameters is defined by the integral $[\theta|y] = \int_X [x, \theta|y] \, \mathrm{d}x$, as mentioned above, but this calculation is to be avoided because the integral is over the high-dimensional latent field x. A cheap algebraic calculation is available when the full conditional for x

$$[x|\theta, y] = \frac{[y|x)[x|\theta]}{[y|\theta)]}$$

has known form, implying that the normalising constant $[y|\theta]$ has known θ dependence, and hence one can evaluate the marginal posterior over θ

$$[\theta|y] \propto [y|\theta][\theta].$$

See [7] for details, and applications in several (nonlinear) models from Statistics.

The typically moderate dimension of hyperparameters θ makes it feasible to represent the full marginal posterior distribution $[\theta|y]$ in the tensor-train (TT) representation [9], while the algebraic calculation, above, makes this step computationally cheap. Cui and Dolgov [2] recently developed the 'SIRT' TT representation that has improved regularity. These TT representations have many similarities to the transport map methods [3], though the TT representations also allow efficient quadrature against the reference measure, as noted in [9].

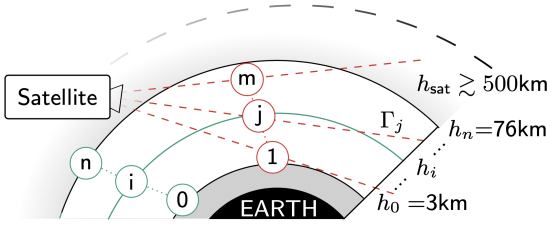
The final piece required for deterministic calculation is the expansion of the posterior expectation of any function $h(x)$ using the 'law of iterated expectation'

$$\mathrm{E}_{x,\theta|y} \left[h\left(x\right) \right] = \mathrm{E}_{\theta|y} \left[\mathrm{E}_{x|\theta,y} \left[h\left(x\right) \right] \right].$$

Using TT representation, the outer expectation on the RHS may be computed by quadrature, in those cases where the inner expectation is sufficiently regular. A common case where the inner expectation may be computed directly, that is without sampling, is where the full conditional $[x|\theta, y]$ is Gaussian, as occurs in the linear-Gaussian inverse problem, or in the weakly non-linear example that we consider later. Then, the inner expectation is a task in numerical linear algebra [4, 5], and with the outer expectation evaluated by quadrature, we may evaluate the posterior mean using only deterministic calculations. When the inner expectation evaluates variances, it is inconvenient, and somewhat pointless, to evaluate the full

conditional covariance matrices by deterministic methods, and instead we evaluate estimates of the posterior variance from a handful of independent posterior samples, drawn using the MTC method, also avoiding MCMC.

Our first application of this deterministic calculation of posterior expectations, used for ironing out computational issues and bugs, has been to the recovery of the stratospheric ozone profile from limb data. In this application, passive radiation from stratospheric ozone is measured by a radiometer mounted on a satellite at a height of 500km to 800km, as depicted in the following figure.



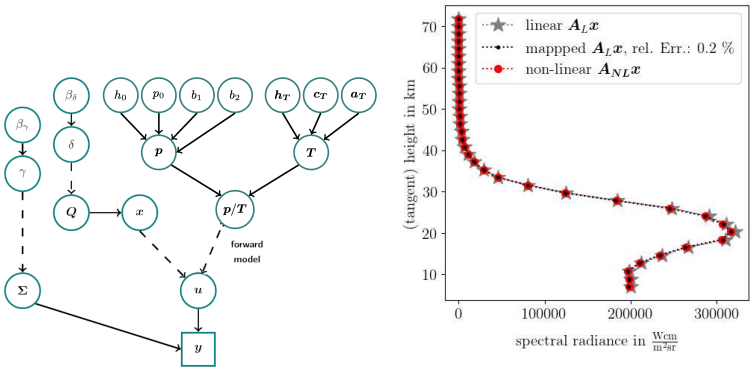
Measurements are made in multiple pointing directions of the radiometer, defining multiple lines of sight from the radiometer. This measurement geometry is called 'limb sounding', with the 'limb' being the shell of the atmosphere tangent to the line of sight. Each measurement corresponds to a path integral of thermal radiation in the stratosphere, reduced by re-absorption, along the line of sight of the radiometer, leading to the weakly nonlinear radiative transfer equation

$$y_j = \int_{\Gamma_j} \underbrace{k(\nu, T) \frac{p(T)}{k_{\mathrm{B}} T(r)} B(\nu, T)}_{A_j} x(r) \tau(r) \mathrm{d}r$$

$$\tau(r) = \exp\left\{ - \int_{r_{\mathrm{obs}}}^{0} k(\nu, T) \frac{p(T)}{k_b T(r')} x(r') \mathrm{d}r' \right\}$$

where x is ozone profile (radiance), p is pressure, T is temperature, and k_{B} is Boltzmann's constant. The nonlinearity occurs because the absorption term τ depends on unknown x, though this typically only reduces measurements by a few percent.

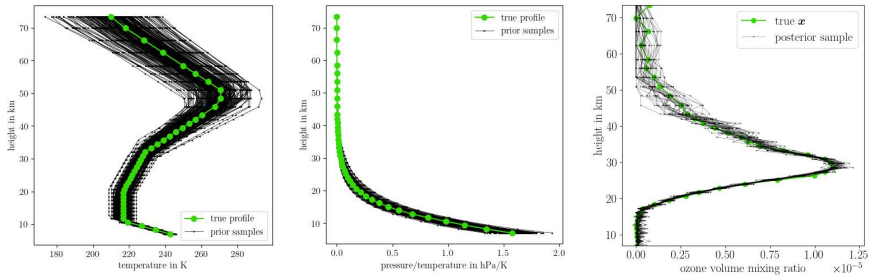
The following DAG shows the dependence of 16 hyperparameters introduced to model physically-realistic unknown pressure and temperature profiles, and also a non-stationary precision matrix Q defining the GMRF (Gauss–Markov random field) over latent profile x.



We do not present details of those models here, but suffice it to say that a comprehensive hierarchical model may be accommodated in the deterministic calculation of posterior expectations. We simulated noisy data for a typical ozone profile. The right subfigure shows the true (nonlinear) data, data simulated using the approximate linear map that ignores absorption, and the result of an affine approximation to the forward map showing that the approximation allows data simulation to well within observation errors.

Since an affine map preserves Gaussian fields, the full conditional over the latent field is Gaussian, hence has known form, allowing the efficient representation of the marginal posterior distribution over all 16 hyperparameters as outlined above. Further, the mean of the full conditional distribution over the latent field may be evaluated by numerical linear algebra, for each value of hyperparameters. These are the components needed to evaluate the posterior mean via the law of iterated expectation, using only deterministic calculation.

We also evaluated posterior variances by drawing a few dozen *independent* posterior samples, requiring just a few dozen linear solves. Some diagnostics for those models and calculations are presented in the following figure.



The left two subfigures show *prior* samples over pressure and temperature, indicating the range of physically-plausible profiles allowed. The right-most subfigure show *posterior* samples and the 'true' unknown ozone profile x showing that the posterior mean recovers the unknown ozone profile, while the cloud of sampled

profiles indicates posterior variances and covariances. This whole calculation took a few minutes on a standard laptop computer, though we have to admit that adjusting code took a lot longer, particularly for the TT representation of $[\theta|y]$.

Development of robust code for that TT representation step is currently underway. We are optimistic that the majority of inverse problems currently treated by MCMC methods will soon be amenable to completely deterministic evaluation of posterior expectations, as we have described here.

References

[1] F. Acosta, M. L. Huber, and G. L. Jones. Markov chain Monte Carlo with linchpin variables. Retrieved on 20/04/2015 from `http://users.stat.umn.edu/~acosta/files/AHJ.pdf`, 2015.

[2] T. Cui and S. Dolgov. Deep composition of tensor-trains using squared inverse Rosenblatt transports. *Foundations of Computational Mathematics*, 22(6):1863–1922, 2022.

[3] C. Fox, L.-J. Hsiao, and J.-E. Lee. Solutions of the multivariate inverse Frobenius–Perron problem. *Entropy*, 23(7):838, 2021.

[4] C. Fox and R. A. Norton. Fast sampling in a linear-Gaussian inverse problem. *SIAM/ASA Journal on Uncertainty Quantification*, 4(1):1191–1218, 2016.

[5] C. Fox and A. Parker. Accelerated Gibbs sampling of normal distributions using matrix splittings and polynomials. *Bernoulli*, 23(4B):3711–3743, 2017.

[6] M. A. Hurn, O. Husby, and H. Rue. Advances in Bayesian image analysis. In P. J. Green, N. L. Hjort, and S. Richardson, editors, *Highly Structured Stochastic Systems*, pages 301–322. Oxford University Press, 2003.

[7] R. A. Norton, J. A. Christen, and C. Fox. Sampling hyperparameters in hierarchical models: improving on Gibbs for high-dimensional latent fields and large datasets. *Communications in Statistics-Simulation and Computation*, 47(9):2639–2655, 2018.

[8] H. Rue and L. Held. *Gaussian Markov random fields: Theory and applications*. Chapman Hall, New York, 2005.

[9] S. Dolgov, K. Anaya-Izquierdo, C. Fox, and R. Scheichl. Approximation and sampling of multivariate probability distributions in the tensor train decomposition. *Statistics and Computing*, 30(3):603–625, 2020.

[10] D. Stoyan, W. S. Kendall, S. N. Chiu, and J. Mecke. *Stochastic geometry and its applications*. John Wiley & Sons, 2013.

Approximative Bayesian optimal design in inverse problems

Tapio Helin

(joint work with Youssef Marzouk and Rodrigo Rojo-Garcia)

Bayesian optimal experimental design provides a principled methodology for choosing experimental configurations that maximise the information gained about an unknown parameter of interest. More precisely, suppose X denotes our unknown parameter, Y stands for the observation and θ is the design parameter. The expected utility U is given by

$$U(\theta) = \mathbb{E}^{\nu} u(X, Y; \theta),$$

where $u(X, Y; \theta)$ denotes the utility of an estimate X given observation Y with design θ, and the expectation is taken w.r.t. the joint distribution ν of X and Y. Notice carefully that ν depends on θ.

In the inverse problem context, the utility u is traditionally chosen as the expected information gain (EIG), based on the Kullback–Leibler divergence between prior and posterior distributions, or the negative squared distance between X and the posterior mean, which leads to a utility proportional to the trace of the posterior covariance. While EIG is grounded in information theory and satisfies key theoretical criteria such as sufficiency ordering and full posterior dependence, the latter is tempting in practise as it reflects expected improvement in estimation through a more cumulative and averaged lens.

Inspired by this tradeoff, we propose a new class of utility functions based on the Wasserstein distance between the posterior and prior measures in [1]. The Wasserstein distance provides a geometric measure of discrepancy between probability distributions by quantifying the optimal transport cost required to transform one distribution into another. Moreover, it remains well-defined even when the two measures are mutually singular, making it particularly suitable for the high-dimensional and computationally intensive setting of Bayesian inverse problems. We define the expected Wasserstein utility by setting

$$U_p(\theta) = \mathbb{E}^Y W_p^p(\mu, \mu^Y),$$

where W_p is the p-Wasserstein distance between the prior μ and the posterior μ^Y. In the case of Gaussian priors and linear observation operators, the Wasserstein-2 utility admits a closed-form expression involving only the mean and covariance operators and leads to efficient computation and a transparent interpretation analogous to A-optimality criteria.

We show that the Wasserstein utility satisfies the sufficiency ordering condition in the sense of Ginebra [3] in finite parameter spaces. In addition, we establish the well-posedness of the utility in infinite-dimensional Hilbert spaces, ensuring its applicability in nonparametric Bayesian inverse problems. The main contribution of this work, however, lies in the stability analysis of the Wasserstein utility, especially for $p = 1$. A key result is the stability with respect to perturbations in the prior distribution. Namely, suppose two different prior distributions μ and $\tilde{\mu}$ give rise to two expected utilities U_1 and $\widetilde{U}_1$, respectively. Then, under suitable conditions, there exists a constant $C > 0$ such that

$$|U_1(\theta) - \widetilde{U}_1(\theta)| \leq C W_2(\mu, \tilde{\mu}).$$

It follows in particular that U_1 is stable under empirical approximations of the prior. In our earlier work [2], we derived stability result for the expected information gain under likelihood perturbations, but prior perturbations were not analysed. In the present study, analogous likelihood stability is proven for W_1-utility.

Finally, we demonstrate the computability of the Wasserstein utility through examples. Nonetheless, the development of efficient computational methods for high-dimensional problems remains an important part of the future work.

References

[1] T. Helin, Y. Marzouk and J.R. Rojo-Garcia, *Bayesian optimal experimental design with Wasserstein information criteria*, arXiv:2504.10092.

[2] D.L. Duong, T. Helin and J.R. Rojo-Garcia, *Stability estimates for the expected utility in Bayesian optimal experimental design*, Inverse Problems **39** (2023), 125008.

[3] J. Ginebra, *On the measure of the information in a statistical experiment*, Bayesian Analysis **2** (2007), 167–211.

MMAF-guided learning for spatio-temporal data

Imma Curato

(joint work with Lorenzo Proietti, Jasmin Sternkopf, and Leonardo Bardi)

Mixed moving average fields are a class of stationary models defined as

$$\boldsymbol{Z}_t(x) = \int_S \int_{\mathbb{R}\times\mathbb{R}^d} f(A, x-\xi, t-s)\ \Lambda(dA, d\xi, ds),\ (t,x) \in \mathbb{R}\times\mathbb{R}^d, \tag{1}$$

where f is a deterministic function called *kernel*, S is denoting a non-empty topological space and Λ is a (homogeneous) Lévy basis, i.e. an independently scattered random measure on $S\times\mathbb{R}\times\mathbb{R}^d$. Loosely speaking, the measure associates a random number with any bounded subset B belonging to the Borel sets $\mathcal{B}(S\times\mathbb{R}\times\mathbb{R}^d)$. Whenever two subsets are disjoint, the associated measures are independent, and the measure of a disjoint union of sets almost certainly equals the sum of the measures of the individual sets. The dependence on the random parameter A in the kernel function is a key feature of MMAF, which allows versatile modelling of *short and long-range* temporal and spatial dependence. Despite their versatility, the use of MMAF is hindered by the fact that its predictive distribution is not generally known as part for the Gaussian case, see [5]. More in details, let us assume to have n observations at different space-time locations $\{\boldsymbol{Z}_i = \boldsymbol{Z}_{t_i}(x_i) : i = 1,\dots,n\}$, and we want to predict $\boldsymbol{Z}_0 = \boldsymbol{Z}_{t_0}(x_0)$ at a new space-time location. Then, the predictive distribution is defined as $\boldsymbol{Z}_0|\boldsymbol{Z}$, where $\boldsymbol{Z} = (\boldsymbol{Z}_1,\dots,\boldsymbol{Z}_n)$, and it is not available in closed form for non-Gaussian distributed Lévy basis in (1).

MMAF-guided learning is a theory-guided machine learning methodology that allows for determining a one-time ahead **ensemble forecast** in a given spatial position and extends the applicability of MMAF to forecast non-Gaussian distributed data. Our methodology is based solely on moment assumptions, an opportune spatio-temporal embedding, and a generalised Bayesian algorithm [4].

So far, MMAF-guided learning has been designed for handling spatio-temporal data called **raster data cubes**. Nowadays, the latter is used for social and demographic analysis, environmental monitoring, and satellite image time series analysis. A data set observed on a regular lattice $\mathbb{L}\subset\mathbb{R}^2$ across time $\mathbb{T} = \{1,\dots,N\}$ is a general example of a raster data cube, see Figure 1-(a).

We assume that such data are generated by an MMAF defined on the set

$$A_t(x) := \big\{(s,\xi)\in\mathbb{R}\times\mathbb{R}^d : s\le t \text{ and } \|x-\xi\|\le c|t-s|\big\}. \tag{2}$$

Note that the model (1) defined on the set $A_t(x)$ is also called an Ambit field in the literature; see [1]. The cone (2) is a model of the causal relationship between space-time points that influence the values of the field in a given spatial point (t, x).

The first step in MMAF-guided learning is selecting a training data set from a raster data cube to enable one-time ahead prediction in a given spatial position x^*, pictured as a red pixel in Figure 1-(b). We follow a supervised learning framework and extract as input data the ones corresponding to the green pixels in Figure 1-(b), namely,

$$\mathcal{I}(t, x^*) := \{(i_s, \xi_s) : \|x^* - \xi_s\| \leq c\,(t - i_s) \text{ for } 0 < t - i_s \leq p, \\ \text{and } (i_s, \xi_s) <_{lex} (i_{s+1}, \xi_{s+1})\}.$$

We then define a training data set, as shown in Figure 1-(c). The spatio-temporal

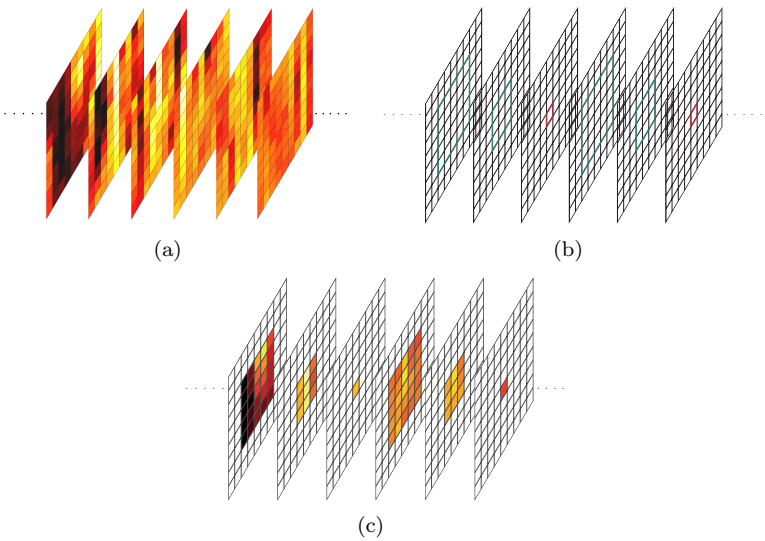


FIGURE 1. (a) Observed raster data cube. (b) The index set of the spatio-temporal embedding is determined as follows. The outputs in each example are determined by the red pixels, which identify the spatial point x^*, and the inputs by the sets $\mathcal{I}(t, x^*)$, which are represented by the pixels in the green boxes. (c) The training data set is then defined by sampling the observed field on the index set defined in (b).

embedding is defined such to preserve the dependence and causal structure of an MMAF defined on the set $A_t(x)$, for more details see [2, Section 3]. We then train a (stochastic) neural network using a PAC Bayesian bound designed to hold for θ-lex weakly dependent data [3]. The proof of such result can be found in [2, Section 3], as the detailed methodology which brings at the end to select the distribution of the weights of the neural network—what we call **a generalised posterior**.

Finally, for each given spatial position x^*, we sample from the (generalised) posterior previously selected and obtain different one-time ahead forecasts, which we collect in an **ensemble**. In [2], we test the learning procedure for a Gibbs posterior distribution and a Gaussian reference distribution on the class of linear models. Our posterior has a convergence rate of $\mathcal{O}(m^{-\frac{1}{2}})$, where m is the number of examples in the training data set, with respect to our PAC Bayesian bound framework. We simulate a set of six data sets from an STOU process with Gaussian and NIG Lévy seed and determine (50 members) ensemble forecasts. We obtain that the inter-quartile ranges of our forecasts always contain the test set and are narrower when the number of observations in the training data set increases. Moreover, our forecasts have a causal interpretation induced by the sets $A_t(x)$ of the data-generating process known as Rubin's potential outcomes framework [6].

Our actual and future work on MMAF-guided learning focuses on investigating the performance of different (generalised) posterior distributions for deep neural network architecture and establishing a theoretical framework for analysing the relationship between the generated ensemble forecast and the true predictive distribution of an MMAF.

References

[1] O. E. Barndorff-Nielsen, F. E. Benth and A. E. D. Veraart *Ambit Stochastics*, Springer, Cham, (2018).

[2] I. V. Curato, O. Furat, Lorenzo Proietti and B. Stroeh, *Mixed moving average field guided learning for spatio-temporal data*, Electron. J. Stat. **19** (2025), 519–592.

[3] I. V. Curato, R. Stelzer, and B. Ströh, *Central limit theorems for stationary random fields under weak dependence with application to ambit and mixed moving average fields*, Ann. Appl. Probab. **32** (2022), 1814–1861.

[4] B. Guedj, *A primer on PAC-Bayesian learning*, arXiv:1901.05353v3 (2019).

[5] M. Nguyen, and A. E. D. Veraart, *Spatio-temporal Ornstein–Uhlenbeck processes: theory, simulation and statistical inference*, Scand. J. Stat. **44** (2017), 46–80.

[6] D. Rubin. *Estimating casual effects of treatments in randomized and non-randomized studies*, J. Educ. Psychol. **66(5)**(1972) 688–701.

Constructing and optimizing dynamic transport schemes

Youssef Marzouk

(joint work with Aimee Maurais, Zhi Robert Ren, Panos Tsimpos, and Jakob Zech)

We discuss constructions for sampling a probability distribution of interest via *transportation of measure*. The essential idea of these constructions is to couple the target probability distribution with a simple, tractable 'reference' distribution, and to use this coupling to generate new samples. Our focus here is on deterministic couplings, induced by transport maps, and in particular on *dynamic* transport schemes that realise such maps incrementally, via the flow map of a system of ordinary differential equations (ODEs). Dynamic transport schemes underpin many

modern methods for Bayesian computation and generative modeling, but the considerable 'design' freedom they offer—centering on how to make effective use of the time axis—has not been rigorously understood.

A way of understanding this freedom is as follows. Suppose that one has prescribed a reference measure π_0 and a target measure π_1 on $\mathbb{R}^d$, and let both be absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}^d$. Then there exist, in general, infinitely many transport maps $T : \mathbb{R}^d \to \mathbb{R}^d$ achieving $T_\sharp \pi_0 = \pi_1$. One such map is the Brenier map, i.e., the optimal map with respect to L^2 cost. Another such map is the triangular Knothe–Rosenblatt rearrangement. But there are (infinitely) many more. A different way to think of the problem is to consider paths through the space of (absolutely continuous) probability measures on $\mathbb{R}^d$, $(\pi_t)_{t\in[0,1]}$, indexed by a 'time' parameter t. Infinitely many paths can have π_0 and π_1 as endpoints. If one prescribes a 'velocity' field $v : \mathbb{R}^d \times [0,1] \to \mathbb{R}$ and solves the ODE initial value problem,

$$\begin{cases} \frac{d}{dt}X(x,t) &= v(X(x,t),t), \qquad t \in [0,1], \\ X(x,0) &= x, \end{cases} \tag{1}$$

then both the transport T and the path of measures $(\pi_t)_{t\in[0,1]}$ are prescribed. In other words, v determines T and $(\pi_t)_{t\in 0,1}$ (with the initial condition $x \sim \pi_0$), but the latter two do not fully specify the first. There are in general infinitely many velocity fields that realise a given transport map T as the time-one flow map of (1), i.e., achieving $T(x) = X(x,1)$. Similarly, there are infinitely many velocity fields that produce the path of marginal distributions $(\pi_t)_{t\in[0,1]}$ (consider two velocities that differ by a solenoidal vector field). These considerations frame the two main topics of the talk.

First, we propose a method for realising (1) given access only to samples from π_0 and the unnormalised density ratio π_1/π_0. This setting arises in Bayesian inference (where π_1 represents the posterior distribution) and in ensemble data assimilation schemes, where often one does not have the ability to evaluate gradients of $\log(\pi_1/\pi_0)$. Our method construct a mean-field analogue of (1), where the velocity $v(\cdot,t)$ depends on the law of X_t, and then a corresponding interacting particle system (IPS) for sampling. The mean-field ODE is obtained by solving a Poisson equation for a velocity field that transports samples along the geometric mixture of the two densities, which is the path of a particular Fisher–Rao gradient flow. We employ a RKHS ansatz for the velocity field, which makes the Poisson equation tractable and enables discretisation of the resulting mean-field ODE over finite samples. The mean-field ODE can be additionally be derived from a discrete-time perspective as the limit of successive linearizations of the Monge–Ampère equations, within a framework known as sample-driven optimal transport. We illustrate both the successes and pitfalls of this construction, and discuss how choices of path other than the geometric mixture could yield better performance in some settings.

Second, we address the question of optimal *scheduling* of dynamic transport, i.e., with what speed should one proceed along a prescribed path of probability

measures. Though many popular methods seek straight line trajectories, i.e., trajectories with zero acceleration in a Lagrangian frame, we show how a specific class of 'curved' trajectories can significantly improve approximation and learning. In particular, we consider the unit-time interpolation of a given transport map T with the identity map, i.e., $T_t(x) = tT(x) + (1-t)x$ for $t \in [0,1]$. A modified version of this one-parameter family of transports is given by $x \mapsto T_{\tau(t)}(x)$, where $\tau : [0,1] \to [0,1]$ is called a 'schedule' and satisfies $\tau'(t) > 0$, $\tau(0) = 0$, $\tau(1) = 1$. A velocity field v that realises $x \mapsto T_{\tau(t)}(x)$ as the time-t flow map of (1) can be written down quite easily. Now we seek the the τ that minimises the spatial Lipschitz constant of this $v(\cdot, t)$, uniformly over times $t \in [0,1]$. Crucially, this aspect of regularity in the velocity controls distribution approximation error when the velocity field is learned from data. We show that, for a broad class of source/target measures and transport maps T, the optimal schedule can be computed in closed form, and that the resulting optimal Lipschitz constant is *exponentially smaller* than that induced by an identity schedule (corresponding to, for instance, the Wasserstein geodesic). Our proof technique relies on the calculus of variations and Γ-convergence, allowing us to approximate the aforementioned degenerate objective by a family of smooth, tractable problems. We close the talk by discussing how this variational formulation could be applied to other dynamic transport schemes, encompassing broader design choices than the schedule τ above.

References

[1] A. Maurais, Y. Marzouk *Sampling in unit time with kernel Fisher–Rao flow*, International Conference on Machine Learning (ICML), 2024.

[2] P. Tsimpos, Z. Ren, J. Zech, Y. Marzouk, *Optimal scheduling of dynamic transport*, Conference on Learning Theory (COLT), 2025.

On hierarchical low-rank structure of posteriors for linear Gaussian inverse problems

Han Cheng Lie

(joint work with Giuseppe Carere)

We consider linear Gaussian inverse problems, i.e. Bayesian inverse problems given by the observation model

$$Y = GX + \zeta,$$

for $G \in \mathcal{B}(\mathcal{H}, \mathbb{R}^n)$, a separable Hilbertian parameter space $\mathcal{H}$, finite-dimensional data space $\mathbb{R}^n$, nondegenerate $\mathbb{R}^n$-valued Gaussian noise $\zeta \sim \mathcal{N}(0, \mathcal{C}_{\text{obs}})$, and non-degenerate Gaussian prior for X given by $\mu_{\text{pr}} = \mathcal{N}(0, \mathcal{C}_{\text{pr}})$ on $\mathcal{H}$. Thus, $\mathcal{C}_{\text{pr}}$ is a linear, positive, self-adjoint trace-class operator on $\mathcal{H}$. Given a realisation y of the data random variable Y, the corresponding solution to the Bayesian inverse

problem is the Gaussian posterior measure $\mu_{\text{pos}}(y) = \mathcal{N}(m_{\text{pos}}(y), \mathcal{C}_{\text{pos}})$, where

$$m_{\text{pos}}(y) = \mathcal{C}_{\text{pos}} G^* \mathcal{C}_{\text{obs}}^{-1} y, \tag{1a}$$

$$\mathcal{C}_{\text{pos}} = \mathcal{C}_{\text{pr}} - \mathcal{C}_{\text{pr}} G^* (\mathcal{C}_{\text{obs}} + G\mathcal{C}_{\text{pr}} G^*)^{-1} G\mathcal{C}_{\text{pr}}, \tag{1b}$$

$$\mathcal{C}_{\text{pos}}^{-1} = \mathcal{C}_{\text{pr}}^{-1} + H, \tag{1c}$$

and $H \coloneqq G^* \mathcal{C}_{\text{obs}}^{-1} G$ denotes the Hessian of the negative-log likelihood.

In [4], a numerical method was proposed for efficiently computing approximations of $\mathcal{C}_{\text{pos}}$, in the case where $\mathcal{H} = \mathbb{R}^d$ for some $d \in \mathbb{N}$. This method was motivated by large-scale PDE inverse problems. The starting point of the method was to observe that by (1c),

$$\mathcal{C}_{\text{pos}} = (\mathcal{C}_{\text{pos}}^{-1})^{-1} = (\mathcal{C}_{\text{pr}}^{-1} + H)^{-1} = \mathcal{C}_{\text{pr}}^{1/2} (I + \mathcal{C}_{\text{pr}}^{1/2} H \mathcal{C}_{\text{pr}}^{1/2})^{-1} \mathcal{C}_{\text{pr}}^{1/2},$$

where $\mathcal{C}_{\text{pr}}^{1/2} H \mathcal{C}_{\text{pr}}^{1/2}$ is known as the 'prior-preconditioned Hessian'. By using a rank-r truncated singular value decomposition (SVD)

$$\mathcal{C}_{\text{pr}}^{1/2} H \mathcal{C}_{\text{pr}}^{1/2} = W \operatorname{diag}((\tfrac{-\lambda_i}{1+\lambda_i})_{i=1}^{d})\, W^\top \approx W_r \operatorname{diag}((\tfrac{-\lambda_i}{1+\lambda_i})_{i=1}^{r})\, W_r^\top$$

and by applying the Sherman–Morrison–Woodbury formula, one obtains

$$\mathcal{C}_{\text{pos}} = \mathcal{C}_{\text{pr}}^{1/2} (I + \mathcal{C}_{\text{pr}}^{1/2} H \mathcal{C}_{\text{pr}}^{1/2})^{-1} \mathcal{C}_{\text{pr}}^{1/2} = \mathcal{C}_{\text{pr}} - \mathcal{C}_{\text{pr}}^{1/2} W_r \operatorname{diag}((\lambda_i)_{i=1}^{r}) W_r^\top \mathcal{C}_{\text{pr}}^{1/2}, \tag{2}$$

i.e. an approximation of $\mathcal{C}_{\text{pos}}$ in terms of a rank-r negative update of $\mathcal{C}_{\text{pr}}$.

In [5], the optimality of the low-rank approximations of $\mathcal{C}_{\text{pos}}$ in (2) was analysed in terms of finding best approximations to the exact Gaussian posterior. Let $D(\cdot\|\cdot)$ denote a measure of dissimilarity on the space of probability measures on $\mathcal{H}$, e.g. the Hellinger metric or the Kullback–Leibler divergence. Since $G \in \mathcal{B}(\mathcal{H}, \mathbb{R}^n)$, it follows that the operator $\mathcal{C}_{\text{pr}} G^* (\mathcal{C}_{\text{obs}} + G\mathcal{C}_{\text{pr}} G^*)^{-1} G\mathcal{C}_{\text{pr}}$ in (1b) is a bounded self-adjoint operator of rank at most n, and hence can be represented as

$$\mathcal{C}_{\text{pr}} G^* (\mathcal{C}_{\text{obs}} + G\mathcal{C}_{\text{pr}} G^*)^{-1} G\mathcal{C}_{\text{pr}} = K_n K_n^*$$

for some $K_n \in \mathcal{B}(\mathbb{R}^n, \mathcal{H})$. This suggests a family of approximation problems:

$$\min \left\{ D\big(\mathcal{N}(m_{\text{pos}}(y), \mathcal{C}_{\text{pos}}) \| \mathcal{N}(m_{\text{pos}}(y), \mathcal{C}_{\text{pr}} - KK^*)\big) \;:\; K \in \mathcal{B}(\mathbb{R}^r, \mathcal{H}) \right\}, \tag{3}$$

where the family is indexed by the rank parameter $r \in \{1, \dots, n\}$. As r increases, one expects the corresponding solutions to (3) to yield a hierarchy of increasingly accurate low-rank approximations, with zero error when $r = n$.

We report some our recent results in [2, 3] on generalisations of the results of [5] to the case where $\mathcal{H}$ can be $\mathbb{R}^d$ or an infinite-dimensional separable Hilbert space.

By the Feldman–Hajek theorem, if $\mu_i \coloneqq \mathcal{N}(m_i, \mathcal{C}_i)$, $i = 1, 2$ are Gaussian measures on a separable Hilbert space $\mathcal{H}$, then either μ_1 and μ_2 are either mutually singular or mutually absolutely continuous, and the latter holds if and only if the following conditions all hold: $\operatorname{ran} \mathcal{C}_1^{1/2} = \operatorname{ran} \mathcal{C}_2^{1/2}$, $m_1 - m_2 \in \operatorname{ran} \mathcal{C}_1^{1/2}$, and

$$R(\mathcal{C}_1 \| \mathcal{C}_2) \coloneqq (\mathcal{C}_1^{-1/2} \mathcal{C}_2^{1/2})(\mathcal{C}_1^{-1/2} \mathcal{C}_2^{1/2})^* - I$$

is Hilbert–Schmidt on $\mathcal{H}$. Our first key result is to show a connection between the Feldman–Hajek operator $R(\mathcal{C}_{\text{pos}} \| \mathcal{C}_{\text{pr}})$ and the prior-preconditioned Hessian.

Proposition 1 ([2, Proposition 3.7]). *There exists a nondecreasing sequence* $(\lambda_i)_i$ *in* $\ell^2((-1,0])$ *with exactly* $\operatorname{rank}(H)$ *nonzero entries and an orthonormal basis* $(w_i)_i$ *of* $\mathcal{H}$*, such that* $(w_i)_i \subset \operatorname{ran} \mathcal{C}_{\mathrm{pr}}^{1/2}$*, and*

$$R(\mathcal{C}_{\mathrm{pos}}\|\mathcal{C}_{\mathrm{pr}}) = (\mathcal{C}_{\mathrm{pos}}^{-1/2}\mathcal{C}_{\mathrm{pr}}^{1/2})(\mathcal{C}_{\mathrm{pos}}^{-1/2}\mathcal{C}_{\mathrm{pr}}^{1/2})^* - I = \sum_i \lambda_i w_i \otimes w_i$$

$$\mathcal{C}_{\mathrm{pr}}^{1/2} H \mathcal{C}_{\mathrm{pr}}^{1/2} = (\mathcal{C}_{\mathrm{pos}}^{-1/2}\mathcal{C}_{\mathrm{pr}}^{1/2})^*(\mathcal{C}_{\mathrm{pos}}^{-1/2}\mathcal{C}_{\mathrm{pr}}^{1/2}) - I = \sum_i \frac{-\lambda_i}{1+\lambda_i} w_i \otimes w_i.$$

The proposition above shows that the Feldman–Hajek operator $R(\mathcal{C}_{\mathrm{pos}}\|\mathcal{C}_{\mathrm{pr}})$ and the prior-preconditioned Hessian are simultaneously diagonalisable by $(w_i)_i$, and that their eigenvalues are related via the involution $(-1,\infty) \ni x \mapsto \frac{-x}{1+x}$. We use the proposition to obtain the following.

Theorem 1 ([2, Theorem 4.21]). *Let* $(\lambda_i)_i \in \ell^2((-1,0])$ *and* $(w_i)_i$ *be as in Proposition 1. For* $r \le n$*, an optimal solution of* (3) *is*

$$\mathcal{C}_r^{\mathrm{opt}} \coloneqq \mathcal{C}_{\mathrm{pr}} - \mathcal{C}_{\mathrm{pr}}^{1/2}\Big(\sum_{i=1}^{r} -\lambda_i w_i \otimes w_i\Big)\mathcal{C}_{\mathrm{pr}}^{1/2}.$$

This solution is unique if and only if $\lambda_{r+1} = 0$ *or* $\lambda_r < \lambda_{r+1}$.

When the measure of dissimilarity $D(\cdot\|\cdot)$ in (3) is the Kullback–Leibler divergence, Hellinger distance, or some ρ-Rényi divergence for $\rho \in (0,1)$, there is an explicit formula for the corresponding approximation error; see [2, Lemma 4.2].

Theorem 1 shows that the approximation (2) that was proposed in [4] for PDE inverse problems applies more generally to all linear Gaussian inverse problems with Hilbertian parameter spaces. In particular, the optimality of the approximations proposed in [4] is independent of the numerical method used to solve the PDE inverse problem and the dimension of the discretisation.

In addition to the low-rank posterior covariance approximation problem (3), it is natural to consider a family of low-rank mean approximation problems of the posterior mean, by searching for rank-r approximations of the data-to-posterior mean map $\mathcal{C}_{\mathrm{pos}} G^* \mathcal{C}_{\mathrm{obs}}^{-1}$ in (1a) that are optimal when averaged over the data Y:

$$\min\left\{\mathbb{E}[D(\mathcal{N}(m_{\mathrm{pos}}(Y),\mathcal{C}_{\mathrm{pos}})\|\mathcal{N}(AY,\mathcal{C}_{\mathrm{pos}})] \;:\; A \in \mathcal{B}(\mathbb{R}^n,\mathcal{H}),\ \operatorname{rank}(A) \le r\right\}, \tag{5}$$

see [5, Section 4]. To solve (5), we consider the square root factorisation

$$\mathcal{C}_{\mathrm{pr}}^{1/2} H \mathcal{C}_{\mathrm{pr}}^{1/2} = \mathcal{C}_{\mathrm{pr}}^{1/2} G^* \mathcal{C}_{\mathrm{obs}}^{-1} G \mathcal{C}_{\mathrm{pr}}^{1/2} = (\mathcal{C}_{\mathrm{pr}}^{1/2} G^* \mathcal{C}_{\mathrm{obs}}^{-1/2})(\mathcal{C}_{\mathrm{pr}}^{1/2} G^* \mathcal{C}_{\mathrm{obs}}^{-1/2})^*$$

of the prior-preconditioned Hessian, and then compute the SVD of the square root. Recalling that $G \in \mathcal{B}(\mathcal{H},\mathbb{R}^n)$, we have

$$\mathcal{C}_{\mathrm{pr}}^{1/2} G^* \mathcal{C}_{\mathrm{obs}}^{-1/2} = \sum_{i=1}^{n} \sqrt{\frac{-\lambda_i}{1+\lambda_i}}\, w_i \otimes \varphi_i, \tag{6}$$

for $(\lambda_i, w_i)_i$ from Proposition 1 and an orthonormal basis $(\varphi_i)_{i=1}^n$ of $\mathbb{R}^n$.

Theorem 2 ([3, Theorem 5.10]). *Let $(\lambda_i, w_i)_i$ and $(\varphi_i)_{i=1}^n$ be as in (6) and $r \leq n$. An optimal solution of (5) is*

$$A_r^{\mathrm{opt}} = \mathcal{C}_{\mathrm{pr}}^{1/2}\Big(\sum_{i=1}^{r}\sqrt{-\lambda_i(1+\lambda_i)}w_i \otimes \varphi_i\Big)\mathcal{C}_{\mathrm{obs}}^{-1/2}.$$

This solution is unique if and only if $\lambda_{r+1} = 0$ or $\lambda_r < \lambda_{r+1}$.

The result above generalises [5, Theorem 4.1] to infinite-dimensional parameter spaces $\mathcal{H}$, and uses our work on rank-constrained operator approximation [1].

We can combine the optimal solutions from Theorems 1 and 2 to find low-rank *joint* approximations of the mean and covariance that are optimal in a Kullback–Leibler divergence; see [3, Proposition 6.1]. We characterise the resulting approximations $\mathcal{N}(Ay, \mathcal{C}_{\mathrm{pr}} - KK^*)$ of the exact Gaussian posterior $\mathcal{N}(m_{\mathrm{pos}}(y), \mathcal{C}_{\mathrm{pos}})$ as the posteriors corresponding to projected Bayesian inverse problems, i.e. inverse problems with the observation model $Y = GPX + \zeta$ for some projection P to a r-dimensional subspace of the parameter space $\mathcal{H}$ [3, Section 7]. These results are new, also for the case of finite-dimensional parameter spaces $\mathcal{H}$.

Our results reveal the importance of the Feldman–Hajek theorem and the finite dimensionality of the data for the low-rank structure of posteriors associated to linear Gaussian inverse problems.

References

[1] G. Carere and H.C. Lie, *Generalised rank-constrained approximations of Hilbert–Schmidt operators on separable Hilbert spaces and applications*, arXiv:2408.05104

[2] G. Carere and H.C. Lie, *Optimal low-rank approximations for linear Gaussian inverse problems on Hilbert spaces, Part I: posterior covariance approximation*, arXiv:2411.01112

[3] G. Carere and H.C. Lie, *Optimal low-rank approximations for linear Gaussian inverse problems on Hilbert spaces, Part II: posterior mean approximation*, arXiv:2503.24209

[4] H. Flath, L. Wilcox, V. Akcelik, J. Hill, B. van Bloemen Waanders, O. Ghattas, *Fast algorithms for Bayesian uncertainty quantification in large-scale linear inverse problems based on low-rank partial Hessian approximations*, SIAM Journal of Scientific Computing **33** (2011), 407–432.

[5] A. Spantini, A. Solonen, T. Cui, J. Martin, L. Tenorio, Y. Marzouk, *Optimal low-rank approximations of Bayesian linear inverse problems*, SIAM Journal of Scientific Computing **37** (2015), A2451-A2487

A Pontryagin minimum principle for stochastic optimal control

Sebastian Reich

(joint work with Manfred Opper)

In this talk, a deterministic mean-field formulation of the Pontryagin minimum principle for stochastic optimal control problems has been sketched out for the first time. Contrary to the well-known forward and backward SDE formulation of the stochastic Pontryagin minimum principle [1], the proposed mean-field approach leads to a gauge variable which can be freely chosen and can be used to decouple the arising forward and reverse time mean-field ODEs.

Problem statement. We consider the optimal control problem for a controlled SDE of the form

$$\mathrm{d}X_t = b(X_t)\mathrm{d}t + GU_t\mathrm{d}t + \Sigma^{1/2}\mathrm{d}B_t, \qquad X_0 = a, \tag{1}$$

under cost function

$$J_T(a, U_{0:T}) = \mathbb{E}\left[\int_0^T \left(c(X_t) + \frac{1}{2}U_t^{\mathrm{T}} R^{-1} U_t\right)\mathrm{d}t + f(X_T)\right]. \tag{2}$$

Here B_t denotes d_x-dimensional Brownian motion, $\Sigma \in \mathbb{R}^{d_x \times d_x}$ the symmetric positive definite diffusion matrix, $R \in \mathbb{R}^{d_u \times d_u}$ a symmetric positive definite weight matrix, $G \in \mathbb{R}^{d_x \times d_u}$ the control matrix, $c(x)$ the running cost, and $f(x)$ the terminal cost. See, for example, reference [1] for more details. We also introduce the weighted norm $\|\cdot\|_R$ via $\|u\|_R^2 = u^{\mathrm{T}} R^{-1} u$.

The aim is to find the closed loop control law $U_t = u_t(X_t)$ that minimises $J_T(a, U_{0:T})$ over the set of admissible control laws. It is well-known [1] that, assuming sufficient regularity, the desired closed loop control law is provided by

$$u_t(x) = -RG^{\mathrm{T}}\nabla_x v_t(x) \tag{3}$$

with the optimal value function $v_t(x)$ satisfying the Hamilton–Jacobi–Bellman (HJB) equation

$$-\partial_t v_t = b \cdot \nabla_x v_t + \frac{1}{2}\Sigma : D_x^2 v_t + c + \min_u \left(Gu \cdot \nabla_x v_t + \frac{1}{2}\|u\|_R^2\right), \quad v_T = f. \tag{4}$$

Deterministic Hamiltonian mean-field formulation. We now formulate the proposed mean-field Pontryagin minimum principle. The initial conditions $a \in \mathbb{R}^{d_x}$ may be viewed as a label in the sense of Lagrangian fluid dynamics, which we assume to be distributed according to a probability density function π_0. We therefore consider functions $x(a)$, $p(x)$, $u(a)$, and $\beta(a)$ and introduce the Hamiltonian functional

$$\mathcal{H}(x, p, u, \beta) = \int_{\mathbb{R}^{d_x}} H(x, p, u, \beta)(a)\, \pi_0(a)\, \mathrm{d}a \tag{5}$$

with Hamiltonian density

$$H(x, p, u, \beta)(a) := p(a)^{\mathrm{T}} \left(b(x(a)) + Gu(a)\right) + \frac{1}{2}\nabla_x \cdot \left(\Sigma\phi(x(a))\right) + \tag{6a}$$

$$\beta(a)^{\mathrm{T}} \left(p(a) - \phi(x(a))\right) + c(x(a)) + \frac{1}{2}\|u(a)\|_R^2. \tag{6b}$$

Here $\beta(a) \in \mathbb{R}^{d_x}$ takes the role of a gauge variable [2], which does not appear in the classical Pontryagin minimum principle [3]. We also note the occurrence of the function $\phi(x)$, which will be determined in terms of the non-holonomic constraint arising from variations with respect to $\beta(a)$ [2]. More specifically, the desired equations of motion are induced by the phase space action principle [2] applied to

$$\mathcal{S} = \int_{\mathbb{R}^{d_x}} \left\{\int_0^T \left(P_t^{\mathrm{T}} \dot{X}_t - H(X_t, P_t, U_t, \beta_t)\right)\mathrm{d}t - f(X_T)\right\} \pi_0 \mathrm{d}a. \tag{7}$$

Taking variations with respect to U_t, we find that the optimal control satisfies

$$\nabla_u H(X_t(a), P_t(a), U_t(a), \beta_t(a)) = R^{-1}U_t(a) + G^{\mathrm{T}}P_t(a). \tag{8}$$

Variations with respect to β_t lead on the other hand to the constraint

$$\phi_t(X_t(a)) = P_t(a), \tag{9}$$

which defines the function $\phi_t(x)$ in terms of $X_t(a)$ and $P_t(a)$. Using the thus specified $\phi_t(x)$, we obtain the closed loop control

$$u_t(x) = -RG^{\mathrm{T}}\phi_t(x). \tag{10}$$

Finally, variations with respect to X_t and P_t lead to the Hamiltonian evolution equations in (X_t, P_t); i.e.,

$$\dot{X}_t(a) = +\nabla_p H(X_t(a), P_t(a), U_t(a), \beta_t(a)), \tag{11a}$$

$$\dot{P}_t(a) = -\nabla_x H(X_t(a), P_t(a), U_t(a), \beta_t(a)) \tag{11b}$$

for each $a \in \mathbb{R}^{d_x}$. The boundary conditions are $X_0(a) = a \sim \pi_0$ and $P_T(a) = \nabla_x f(X_T(a))$. Dropping the label $a \in \mathbb{R}^{d_x}$ from now on, the Hamiltonian equations of motion (11) therefore become

$$\dot{X}_t = b(X_t) + GU_t + \beta_t, \tag{12a}$$

$$\dot{P}_t = (D_x\phi_t(X_t))^{\mathrm{T}}\,\beta_t - (D_x b(X_t))^{\mathrm{T}}\,P_t - \frac{1}{2}\nabla_x\nabla_x\cdot(\Sigma\phi_t(X_t)) - \nabla_x c(X_t). \tag{12b}$$

The following theorem provides the key result with regard to the gauge variable β_t and demonstrates that (12) indeed delivers the desired extension of the Pontryagin minimum principle to stochastic optimal control problems.

Theorem 1. *For any choice of the gauge variable β_t, the resulting function $\phi_t(x)$ satisfies*

$$\phi_t(x) = \nabla_x v_t(x) \tag{13}$$

where $v_t(x)$ is the value function satisfying the HJB equation (4).

Proof. Let us derive the evolution equation for $\phi_t(x)$ implied by (9):

$$-\partial_t\phi_t(X_t) = D_x\phi_t(X_t)\dot{X}_t - \dot{P}_t \tag{14a}$$

$$= D_x\phi_t(X_t)\,(b(X_t) + GU_t) + \frac{1}{2}\nabla_x\nabla_x\cdot(\Sigma\phi_t(X_t)) + \tag{14b}$$

$$(D_x b(X_t))^{\mathrm{T}}\,\phi_t(X_t) + \nabla_x c(X_t). \tag{14c}$$

Here we have used that $D_x\phi_t(x)$ is symmetric since $\phi_t(x)$ itself is the gradient of the value function $v_t(x)$. Hence, $\phi_t(x)$ satisfies the reverse time PDE

$$-\partial_t\phi_t = D_x\phi_t\,(b + GU_t) + \frac{1}{2}\nabla_x\nabla_x\cdot(\Sigma\phi_t) + (D_x b)^{\mathrm{T}}\,\phi_t + \nabla_x c \tag{15}$$

subject to the terminal condition $\phi_T = \nabla_x f$, which also follows from (4) by taking the gradient. Hence $\phi_t(x) = \nabla_x v_t(x)$ independent of β_t. □

A natural choice for the gauge function β_t is

$$\beta_t = -\frac{1}{2}\Sigma\nabla_x \log \pi_t(X_t), \tag{16}$$

where $\pi_t(x)$ denotes the law of X_t. Alternatively, consider

$$\beta_t = GRG^{\mathrm{T}}\phi_t(X_t), \tag{17}$$

which eliminates the control from the forward evolution equation in X_t since

$$GU_t = -GRG^{\mathrm{T}}\phi_t(X_t) = -\beta_t. \tag{18}$$

Both choices for the gauge variable can also be combined into

$$\beta_t = GRG^{\mathrm{T}}\phi_t(X_t) + Gu_t^{\mathrm{ref}}(X_t) - \frac{1}{2}\Sigma\nabla_x \log \pi_t(X_t), \tag{19}$$

where $u_t^{\mathrm{ref}}(x)$ denotes a known reference control; if available.

Open questions. Similar to the well-known forward and backward SDE formulation of the stochastic Pontryagin principle [1], a linear regression problem arises from (9) when discretised as an interacting particle systems. The accuracy and stability of numerical approximations needs to be investigated and compared to forward and backward SDE formulations. Furthermore, the freedom in the choice of the gauge variable β_t, such as (19), allows for a decoupling of the forward and reverse time mean-field ODEs in X_t and P_t. Again, implications on algorithmic implementations need to be investigated. Application of the proposed methodology to model predictive control [4] should be of particular interest.

References

[1] René Carmona. *Lectures on BSDEs, Stochastic Control, and Stochastic Differential Games with Financial Applications*. SIAM, Philadelphia, 2016.

[2] Paul A.M. Dirac. Generalized Hamiltonian dynamics. *Can. J. Math.*, 2:129–148, 1950.

[3] L.S. Pointryagin, V.G. Boltyanskii, R.V. Gamkrelidze, and E.F. Mihchenko. *The Mathematical Theory of Optimal Processes*. John Wiley & Sons, New York, 1962.

[4] James B. Rawlings, David Q. Mayne, and Moritz M. Diehl. *Model Predictive Control: Theory, Computation, and Design*. Nob Hill Publishing, Madison, 2nd edition, 2018.

Autoencoders in Function Space

Hefin Lambley

(joint work with Justin Bunker, Mark Girolami, Andrew M. Stuart, and T. J. Sullivan)

Deep learning is widely used in uncertainty quantification, e.g., in the construction of surrogate models, but traditional architectures must operate at fixed input and output resolutions. In recent years there has been much interest in deep-learning methods for functions, with architectures such as DeepONet [1] and Fourier neural operators [2] that can be discretised, trained, and evaluated at any resolution. Much of this work has focussed on supervised learning, e.g., in approximating the

solution operator to a differential equation from input–output pairs, and unsupervised tasks such as dimension reduction and generative modelling have only recently started to receive attention in, e.g., [3] and [4].

In this work, we focus on a class of methods known as autoencoders. Autoencoders solve two tasks in unsupervised learning: *dimension reduction* and *generative modelling*. More precisely, autoencoders assume access to samples from an unknown distribution Υ on the data space $\mathcal{U}$; then, a latent space $\mathcal{Z}$ is chosen, typically of much lower dimension than $\mathcal{U}$, and an autoencoder seeks to learn two transformations: an *encoder* mapping from $\mathcal{U}$ to $\mathcal{Z}$, and a *decoder* mapping from $\mathcal{Z}$ to $\mathcal{U}$. The goal is to choose the encoder and decoder so that their composition is approximately the identity; doing so leads to dimension reduction, because the encoded representation of $u \sim \Upsilon$ is forced to capture meaningful features. Moreover, a generative model can be obtained by sampling from the latent space $\mathcal{Z}$ appropriately and then decoding to obtain a distribution approximating Υ.

We formulate function-space versions of autoencoders, in both their deterministic (FAE) and variational (FVAE) forms, and deploy them on scientific data sets including path distributions of stochastic differential equations (SDEs) as arising in molecular dynamics, and vorticity fields of Navier–Stokes fluid flows [5].

FVAE is an extension of the variational autoencoder (VAE) of Kingma and Welling [6], which we formulate as a problem of matching two joint distributions in function space. Take $\mathcal{U}$ to be a separable Banach space, possibly of infinite dimension, and select a latent space $\mathcal{Z} = \mathbb{R}^{d_Z}$ and an easily sampled latent distribution $\mathbb{P}_z$ on $\mathcal{Z}$. Under FVAE, the encoder and decoder are transformations taking points to probability distributions:

$$\text{(1a)} \qquad \text{(encoder) } \mathcal{U} \ni u \mapsto \mathbb{Q}^{\theta}_{z|u} \in \mathscr{P}(\mathcal{Z}),$$

$$\text{(1b)} \qquad \text{(decoder) } \mathcal{Z} \ni z \mapsto \mathbb{P}^{\psi}_{u|z} \in \mathscr{P}(\mathcal{U}).$$

Here $\mathscr{P}(X)$ denotes the set of probability measures on X, and $\mathbb{Q}^{\theta}_{z|u}$ and $\mathbb{P}^{\psi}_{u|z}$ will be chosen to lie in parametric classes depending on parameters $\theta \in \Theta$ and $\psi \in \Psi$.

We wish to approximately enforce that composing (1a) and (1b) gives the identity for $u \sim \Upsilon$. To do this we specify two joint models on the product space $\mathcal{Z} \times \mathcal{U}$:

$$\text{(2a)} \qquad \text{(joint encoder model) } \mathbb{Q}^{\theta}_{z,u}(\mathrm{d}z, \mathrm{d}u) = \mathbb{Q}^{\theta}_{z|u}(\mathrm{d}z)\Upsilon(\mathrm{d}u),$$

$$\text{(2b)} \qquad \text{(joint decoder model) } \mathbb{P}^{\psi}_{z,u}(\mathrm{d}z, \mathrm{d}u) = \mathbb{P}^{\psi}_{u|z}(\mathrm{d}u)\mathbb{P}_z(\mathrm{d}z).$$

This suggests a natural objective functional to be minimised, which proves to be an appropriate generalisation of the VAE objective to function space:

$$\text{(3)} \qquad \text{(FVAE objective) } \operatorname*{arg\,min}_{\theta \in \Theta,\, \psi \in \Psi} D_{\mathrm{KL}}\big(\mathbb{Q}^{\theta}_{z,u} \,\|\, \mathbb{P}^{\psi}_{z,u}\big).$$

This objective simultaneously trains an autoencoder as well as a generative model $\mathbb{P}^{\psi}_{u}$, the u-marginal of (2b). When $\mathcal{U}$ has finite dimension, the objective (3) is equal, up to a finite constant, to the evidence lower bound (ELBO) usually taken

as the VAE training objective, but our approach makes no appeal to Lebesgue densities, which do not exist in infinite dimensions.

The most common VAE model in finite dimensions is to take all relevant distributions to be Gaussian, with the mean and covariance of the encoder determined by learnable maps f and Σ, with the mean of the decoder determined by a learnable map g, and with the decoder covariance fixed as $\beta I_{\mathcal{U}}$:

$$\mathbb{P}_z = N\big(0, I_{\mathcal{Z}}\big) \tag{4a}$$

$$\mathbb{Q}^{\theta}_{z|u} = N\big(f(u;\theta), \Sigma(u;\theta)\big) \tag{4b}$$

$$\mathbb{P}^{\psi}_{u|z} = N\big(g(z;\psi), \beta I_{\mathcal{U}}\big). \tag{4c}$$

To illustrate the difficulties that arise in the infinite-dimensional setting, take $\mathcal{U} = L^2(0,1)$ and $\Upsilon \in \mathscr{P}(\mathcal{U})$; then consider the model (4a)–(4c), noting that, for each latent vector $z \in \mathcal{Z}$, the decoder $\mathbb{P}^{\psi}_{u|z}$ now returns a *function* $g(z;\psi)$ corrupted by additive white noise. This is the setting adopted by the variational autoencoding neural operator [7]. One key finding of our work is that, since realisations from the decoder contain additive white noise that never has L^2-regularity, draws from the decoder distribution and data distribution have different levels of smoothness. Owing to this incompatibility, there is no absolute continuity between the joint models (2a) and (2b), and thus the joint divergence in (3) is identically infinite for all θ and ψ. This renders the FVAE objective meaningless; moreover this issue is not purely theoretical, and manifests in practice through increasing instability and a divergent training objective as the resolution of the training data is refined.

In order to obtain a meaningful extension of VAEs to infinite dimensions, we restrict attention to problem classes where the decoder distribution can be chosen to be compatible with the data distribution. This is a stringent condition in infinite dimensions, but there are many problem classes for which this compatibility can be established, e.g., when Υ is the path distribution of an SDE. In such settings, FVAE performs well as a probabilistic generative model and an autoencoder, with the advantage that uncertainty quantification is inherently built in to the model.

To address the fact that FVAE is limited to specific classes of data, we also generalise regularised autoencoders to yield FAE. The FAE objective works "out of the box" for a broad class of data distributions, but has the disadvantage that it merely trains an autoencoder with no inherent uncertainty quantification. However, we show that a generative model can be established through a two-step procedure in which one first trains an autoencoder and then trains a finite-dimensional generative model on the autoencoder latent space.

Using neural operators in the encoder and decoder of FVAE and FAE enables training and evaluation across resolutions; this permits new applications to inpainting, superresolution, and generative modelling. For example, after training on a data set of vorticity fields of fluid flows, FAE can reconstruct the vorticity field from sparse measurements with 95% of the original training mesh missing.

References

[1] L. Lu, P. Jin, G. Pang, Z. Zhang, and G. E. Karniadakis, *Learning nonlinear operators via DeepONet based on the universal approximation theorem of operators*, Nature Machine Intelligence **3**:3 (2021), 218–229.

[2] N. Kovachki, Z. Li, B. Liu, K. Azizzadenesheli, K. Bhattacharya, A. Stuart, and A. Anandkumar, *Neural operator: Learning maps between function spaces with applications to PDEs*, Journal of Machine Learning Research **24** (2023), 1–97.

[3] J. H. Lim, N. B. Kovachki, R. Baptista, C. Beckham, K. Azizzadenesheli, J. Kossaifi, V. Voleti, J. Song, K. Kreis, J. Kautz, C. Pal, A. Vahdat, and A. Anandkumar, *Score-based diffusion models in function space* (2023), arXiv:2302.07400.

[4] J. Pidstrigach, Y. Marzouk, S. Reich, and S. Wang, *Infinite-dimensional diffusion models for function spaces* (2023), arXiv:2302.10130.

[5] J. Bunker, M. Girolami, H. Lambley, A. M. Stuart, and T. J. Sullivan, *Autoencoders in function space* (2024), arXiv:2408.01362.

[6] D. P. Kingma and M. Welling, *Auto-encoding variational Bayes*, in Y. Bengio and Y. LeCun, editors, *The Second International Conference on Learning Representations* (2014).

[7] J. H. Seidman, G. Kissas, G. J. Pappas, and P. Perdikaris, *Variational autoencoding neural operators*, in A. Krause, E. Brunskill, K. Cho, B. Engelhardt, S. Sabato, and J. Scarlett, editors, *Proceedings of the 40th International Conference on Machine Learning*, Proceedings of Machine Learning Research **202** (2023), 30491–30522.

Statistical guarantees for denoising reflected diffusion models

Claudia Strauch

(joint work with Asbjørn Holk and Lukas Trottner)

Denoising diffusion models represent a prominent class of generative models, in which a forward stochastic process gradually perturbs data by noise and a backward process is trained to reverse this transformation. The backward dynamics are typically characterised via the so-called *score*, the gradient of the log-density of the noised data distribution. While these models have shown impressive empirical performance in various application domains, an in-depth study of their statistical properties as estimators of data-generating distributions has only recently begun [1, 5, 6].

In our talk, based on [3], we presented a constrained variant of such models, referred to as *denoising reflected diffusion models* (DRDMs), in which the forward and backward processes are confined to a bounded domain via normal reflection at the boundary. This setting naturally arises when the support of the data-generating distribution is compact, or when physical constraints impose spatial boundaries. While recent empirical studies [2, 4] have demonstrated the potential of such models, a statistical analysis investigating the model in terms of distributional learning has so far been lacking.

The focus of the talk was on the *statistical properties* of DRDMs as distribution estimators. Implementation and algorithmic aspects were not addressed. Instead, the analysis concentrated on quantifying the effect of estimation errors in the score function on the discrepancy between the true data distribution and the distribution induced by the generative process. The forward process in our model is defined

by the reflected stochastic differential equation

$$\mathrm{d}X_t = \nabla f(X_t)\,\mathrm{d}t + \sqrt{2f(X_t)}\,\mathrm{d}W_t + \nu(X_t)\,\mathrm{d}\ell_t^D,$$

where W is a d-dimensional Brownian motion, $f\colon \mathbb{R}^d \to [f_{\min}, \infty) \subset (0, \infty)$ is a smooth potential, D is an open and bounded domain with smooth boundary, ν is the inward-pointing normal vector field on ∂D, and $\ell_\cdot^D$ is the local time at the boundary. This process is constrained to $\overline{D}$ through normal reflections at the boundary and is analytically characterised by the divergence-form generator $\mathcal{A} = \nabla \cdot f \nabla$, subject to Neumann boundary conditions. The specific form of the generator implies time-reversibility of the process with respect to its uniform stationary distribution, and the spectral gap of $\mathcal{A}$ yields an exponential convergence rate. This combination of a fast speed of convergence and an easy-to-sample-from limiting distribution makes this class of processes particularly suitable for generative modelling purposes.

The backward process reverses the forward dynamics and requires the time- and space-dependent score function $s^\circ(x,t) = \nabla \log p_t(x)$, where for the forward transition densities $q_t(x,y)$ and the underlying data distribution p_0, the forward density $p_t(x)$ is given by $p_t(x) = \int_0^t q_t(y,x) p_0\,(\mathrm{d}y)$. Since it depends on p_0, which is generally unknown, this score must be estimated from data via a variant of *score matching*, adapted to the reflected setting. For this purpose, a spectral representation of $s^\circ(x,t)$ in terms of the eigenfunctions of $\mathcal{A}$ is derived. This representation facilitates the construction of neural network estimators, the approximation and generalisation abilities of which are studied in [3]. More precisely, our investigation addresses the estimation error in total variation between the true and generated data distributions. The score estimator is obtained via empirical risk minimisation of the *denoising score matching loss* over a suitably regularised class of ReLU networks. The generalisation error is controlled via uniform bounds and the metric entropy of the induced loss class, while the approximation error is analysed using Sobolev smoothness assumptions on p_0 and spectral truncation of the score expansion. The main result establishes that, under appropriate lower bounds and Sobolev regularity conditions on the data-generating density, the expected total variation distance between the true distribution and the one induced by the learned generative process converges at a minimax optimal rate, up to small logarithmic factors. This indicates that DRDMs can achieve statistically optimal performance in a constrained setting.

The framework presented in the talk extends the statistical theory of denoising diffusion models to settings with spatial constraints, providing a detailed account of the interplay between generalisation and approximation in score-based generative modelling. While the present work abstracts from numerical and implementation issues, it suggests further research directions, particularly with regard to uncertainty quantification in reflected generative models, where the propagation of score estimation uncertainty through the generative process remains an open question.

References

[1] I. Azangulov, G. Deligiannidis, and J. Rousseau, *Convergence of Diffusion Models Under the Manifold Hypothesis in High-Dimensions*, arXiv preprint arXiv:2409.18804, 2024.

[2] N. Fishman, L. Klarner, V. De Bortoli, E. Mathieu, and M. Hutchinson, *Diffusion Models for Constrained Domains*, Transactions on Machine Learning Research, PMLR, 2023.

[3] A. Holk, C. Strauch and L. Trottner, *Statistical guarantees for denoising reflected diffusion models*, arXiv preprint arXiv:2411.01563v2, 2024.

[4] A. Lou, and S. Ermon, *Reflected Diffusion Models*, International Conference on Machine Learning, pp. 22675–22701, PMLR, 2023.

[5] K. Oko, S. Akiyama, and T. Suzuki, *Diffusion Models are Minimax Optimal Distribution Estimators*, International Conference on Machine Learning, pp. 26517–26582, PMLR, 2023.

[6] R. Tang, and Y. Yang, *Adaptivity of Diffusion Models to Manifold Structures*, International Conference on Artificial Intelligence and Statistics, pp. 1648–1656, PMLR, 2024.

Dimension-independent Markov chain Monte Carlo on the sphere

Björn Sprungk

(joint work with Han Cheng Lie, Daniel Rudolf, and T. J. Sullivan)

We consider Bayesian analysis on high-dimensional unit spheres $\mathbb{S}^{d-1} \subset \mathbb{R}^d$ with angular central Gaussian prior distribution. This prior can be defined as the push-forward of a centred Gaussian distribution in $\mathbb{R}^d$ under the radial projection. It models antipodally symmetric directional data, is easily defined in Hilbert spaces, and occurs, for instance, in Bayesian density estimation and binary level set inversion. In [2] we derive efficient Markov chain Monte Carlo methods for approximate sampling of posteriors with respect to this kind of priors. Our approach relies on lifting the sampling problem to the ambient Hilbert space and exploit existing dimension-independent samplers in $\mathbb{R}^d$ such as the pCN Metropolis algorithm [1] and the elliptical slice sampler [3]. By a push-forward Markov kernel construction [4] we then obtain Markov chains on the sphere which inherit reversibility and spectral gap properties from samplers in linear spaces. In particular, we can show that given only the boundedness of the likelihood the obtained reprojected pCN Metropolis on the sphere possesses a dimension-independent spectral gap [2]. This is verified in numerical experiments. Also the reprojected elliptical slice sampler performs dimension-independent in our experiments. Moreover, we observe that the performance in terms of integrated autocorrelation time (IAT) of classical random walk-like Metropolis algorithms on the sphere (e.g., [5]) deteriorate very quickly with increasing dimension $d \to \infty$. Our numerical experiments suggest that the IAT grows like $\mathcal{O}(d^5)$ which is in contrast to the well-known rate of $\mathcal{O}(d^1)$ in $\mathbb{R}^d$. This behaviour on $\mathbb{S}^{d-1}$ is currently unexplained and left as an open question for future research on optimal scalings of random walk-like Metropolis algorithms on $\mathbb{S}^{d-1}$.

References

[1] S. L. Cotter, G. O. Roberts, A. M. Stuart, D. White, *MCMC methods for functions: Modifying old algorithms to make them faster*, Statistical Science **28** (2013), 424–446.

[2] H. C. Lie, D. Rudolf, B. Sprungk, T. J. Sullivan, *Dimension-independent Markov chain Monte Carlo on the sphere*, Scandandinavian Journal of Statistics **50** (2023), 1818–1858.
[3] I. Murray, R. Adams, D. MacKay, *Elliptical slice sampling*, Proceedings of machine learning research **9** (2010), 541–548).
[4] D. Rudolf, B. Sprungk, *Robust random walk-like Metropolis–Hastings algorithms for concentrating posteriors*, arXiv:2202.12127 (2022).
[5] E. Zappa, M. Holmes-Cerfon, J. Goodman, *Monte Carlo on manifolds: Sampling densities and integrating functions*, Communications on Pure and Applied Analysis **71** (2018), 2609–2647.

Edge preserving random tree Besov priors

Hanne Kekkonen
(joint work with Matti Lassas, Eero Saksman, Samuli Siltanen, and Andreas Tataris)

Robust methods for solving inverse problems rely on combining measurement data with *a priori* information about the unknown function. One of the key challenges in inverse problems research is to formulate such prior knowledge in a mathematically meaningful and computationally tractable way. Popular models promote global smoothness, piecewise regularity, or sparsity. These models can be implemented via variational regularisation, yielding stable solutions but offering little information about its uncertainties. Bayesian inversion provides an attractive alternative by delivering not only point estimates but also a characterisation of the uncertainties arising from noise and model error.

In practice measurements are discrete and corrupted by noise, which can in many cases be reasonably modelled using independent Gaussian random variables. This gives rise to the measurement model

$$M_i = (Af)_i + w_i, \quad i = 1, \dots, n, \ w_i \overset{\text{iid}}{\sim} \mathcal{N}(0,1), \tag{1}$$

where A describes the forward process. Solving the inverse problem computationally also requires a finite-dimensional model for the unknown f. It is advisable to construct models for f in a discretisation-invariant way, for instance, by defining a continuous model and discretising it as late as possible in the inference procedure [5].

In the Bayesian framework, we place a prior probability measure Π on f; the solution to the inverse problem is then the posterior distribution, i.e. the conditional distribution of f given the observed data M. Most existing theory for infinite-dimensional Bayesian inverse problems assumes a Gaussian prior. While Gaussian priors have computational advantages, they fail to capture sharp features such as edges and interfaces, which are critical in many signal and image reconstruction tasks.

A popular method for achieving edge-preserving reconstructions in image analysis is using the so-called total variation (TV) prior and the mode of the posterior as a point estimate. In practice this means minimising

$$\|Af - M\|_{L^2}^2 + \alpha \|\nabla f\|_{L^1}. \tag{2}$$

However, despite extensive study, there is no known infinite-dimensional prior for which (2) would be the maximum a posteriori (MAP) estimate. In other words, the commonly used formal prior

$$\pi(f) \underset{\text{formally}}{\propto} \exp\left(-\alpha\|\nabla f\|_{L^1}\right) \tag{3}$$

is not known to correspond to any well-defined random variable. Furthermore, standard discrete TV priors can converge to Gaussian smoothness priors under mesh refinement, see [4].

To address this, [3] proposed replacing (3) with a well-defined prior

$$\pi(f) \underset{\text{formally}}{\propto} \exp\left(-\|\nabla f\|^p_{B^0_{pp}}\right),$$

where the Besov spaces B^0_{pp} are closely related to L^p spaces. For $p = 1$, the above prior has similar properties to the total variation prior, but corresponds to a well-defined infinite-dimensional random variable. The construction of the Besov prior uses the Karhunen-Loève expansion with wavelet basis. This is particularly suitable when modelling smooth functions with few local irregularities since such functions have sparse expansion in the wavelet basis unlike, e.g., in the Fourier basis. Truncating the expansion yields practical finite-dimensional approximations. The well-definedness and well-posedness of the posterior measure were extended to non-linear cases in [1].

The idea of the random tree Besov priors, presented in [2], is to use the Karhunen-Loève expansion to create priors similar to [3], but to choose the non-zero wavelets in the sum in a systematic way, so that draws are mainly smooth with few large jumps. This can be done by introducing a new random variable T that takes values in the space of 'trees'. The wavelet coefficients can be arranged into a tree with the coarser scales at the top, and finer details further down. A coefficient at a finer scale can be non-zero only if all its ancestors in the tree are also non-zero. This structure models persistence of features across scales. A jump in a signal results in large wavelet coefficients across many levels, while white noise yields large coefficients primarily at fine scales. The density of the non-zero coefficients, and consequently the sparsity of the reconstruction, is controlled by a parameter β.

In [2] the wavelet density index β was chosen manually for the considered denoising problems but we would like to choose β automatically from data. This also enables the use of level-dependent $\beta = (\beta_j)_j$, allowing the sparsity to vary across scales. This is a natural extension, as wavelet coefficients of a signal or image are typically sparser at finer resolutions. We place a hyperprior on β and estimate it via levelwise MAP estimation.

We consider first the non-parametric regression problem where $A = I$ in (1), and set the prior $\pi(\beta_j) \propto (0.5 - \beta_j)^c$, $\beta_j \in [0, 0.5]$, $c \geq 0$ for the wavelet density index. This leads to a problem that can be solved analytically; starting from the finest level, we estimate β_j iteratively up to the coarsest level. The resulting reconstructions outperform those obtained with the best fixed β. For general

linear inverse problems, a plug-and-play approach leads to similarly promising edge preserving results in deconvolution.

An interesting open question is how to sample from the posteriors arising from random tree Besov priors. While tree-based methods have proven highly successful in regression and density estimation, they do not readily extend to inverse problems. Developing efficient sampling methods for such priors could open new directions in both inverse problems and Bayesian statistics.

References

[1] M. Dashti, S. Harris and A. Stuart, *Besov priors for Bayesian inverse problems*, Inverse Problems and Imaging, **6** (2012), 183–200.

[2] H. Kekkonen, M. Lassas, E. Saksman, and S. Siltanen, *Random tree Besov priors - towards fractal imaging*, Inverse Problems and Imaging **17** (2023) 507–531.

[3] M. Lassas, E. Saksman and S. Siltanen, *Discretization-invariant Bayesian inversion and Besov space priors*, Inverse Problems and Imaging, **3** (2009), 87–122.

[4] M. Lassas and S. Siltanen, *Can one use total variation prior for edge-preserving Bayesian inversion?*, Inverse Problems, **20** (2004), 1537–1563.

[5] A. M. Stuart, *Inverse problems: A Bayesian perspective*, Acta Numerica, **19** (2010), 451–559.

Joint chance constraints in stochastic optimal control and Gaussian processes

Georg Stadler

(joint work with Kewei Wang)

Chance constraints are the generalisation of inequality constraints to stochastic optimisation, where constraints for random variables can often only be enforced with a given, typically high probability. *Joint* chance constraints generalise point-wise bound constraints and require that the realisations of a random variable satisfy a pointwise bound constraint *everywhere* with high probability. Such probabilistic constraints have been studied in the context of finite-dimensional optimisation for several decades, but their analysis in the context of infinite-dimensional optimisation and optimal control is more recent [1, 2, 3]. The challenges are both theoretical and computational. An example of an optimal control problem with joint chance state constraints is given by

$$\min_{u,y} \mathcal{J}(y,u) \quad \text{s.t.} \quad e(y(\omega),u,\xi(\omega)) = 0, \tag{1}$$

where u is the deterministic control and $\xi(\omega)$ is a random variable depending on an event ω. Through the PDE constraint $e(y(\omega),u,\xi(\omega)) = 0$, the state $y(\omega)$ depends not only on u but also on $\xi(\omega)$ and is thus a random variable. The objective $\mathcal{J}$ typically involves an expectation over ω. Assume that the PDE is defined over a domain $\mathcal{D} \subset \mathbb{R}^d$, $d \in \{1,2,3\}$, such that the state y is also a function of $x \in \mathcal{D}$. A joint chance state constraint for given lower bound $\underline{y} : \mathcal{D} \mapsto \mathbb{R}$ is then given by

$$\mathbb{P}(\omega \,|\, \underline{y}(x) \le y(x,\omega) \text{ for a. a. } x \in \mathcal{D}) \ge p, \tag{2}$$

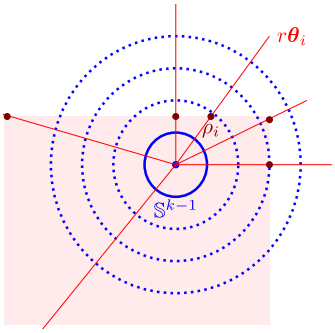


FIGURE 1. Illustration of spherical-radial decomposition to estimate a two-dimensional joint chance constraint probability, given by a Gaussian measure integrated over the red shaded area. Shown are level sets of the Gaussian density in blue (dotted), the unit sphere $\mathbb{S}^1$ in blue (solid), the rays $r\boldsymbol{\theta}_i$, $r > 0$ in red, as well as the radius ρ_i at which the ray exits the feasible area.

where $0 < p < 1$ is given (typically close to 1), and almost everywhere bounds are necessary if the states $y(\omega)$ are not guaranteed to be continuous functions in x. For the case where the governing PDE (2) is linear or bilinear, theoretical and computational results for problems of the form (1), (2) are presented in [3].

For the numerical approximation of chance constraints, the *spherical-radial decomposition* (SRD) of elliptically distributed (e.g., Gaussian) multivariate distributions has proven to be a useful tool. The SRD for an n-dimensional random vector $\boldsymbol{\zeta}$ is

$$\boldsymbol{\zeta} = \boldsymbol{m} + \tau L\boldsymbol{\theta}, \tag{3}$$

where $\boldsymbol{m} \in \mathbb{R}^n$, $L \in \mathbb{R}^{n\times k}$ with $k \leq n$, τ is a one-dimensional non-negative random variable, and $\boldsymbol{\theta}$ is a uniformly distributed random vector on the unit sphere $\mathbb{S}^{k-1}$ of $\mathbb{R}^k$. When $\boldsymbol{\zeta}$ follows a multivariate normal distribution, then $\boldsymbol{m}$ is the mean of this distribution, L is a square root of the covariance matrix, and τ follows a one-dimensional χ-distribution with k degrees of freedom. The decomposition (3) not only provides an alternative to standard Monte Carlo sampling from the distribution, but additionally facilitates the computation of probabilities arising in joint chance constraints. To be more precise, one can draw samples $\boldsymbol{\theta}_i$ from the uniform distribution over $\mathbb{S}^{k-1}$ and then, for problems where the random variable enters linearly, compute the radius ρ_i, at which the ray $r\boldsymbol{\theta}_i$ ($r \geq 0$) leaves the feasible set, as illustrated in Figure 1. Compared to standard Monte Carlo sampling of (2), a Monte Carlo estimator based on the SRD has a reduced variance and provides derivatives of the probability with respect to the control [3].

Currently, we extend SRD-based methods to enforce that realisations of a Gaussian process satisfy a bound constraint with high probability [4, 5]. Consider a (prior) Gaussian process $\xi_{\text{pr}} \sim \mathcal{N}(\xi_0, \mathcal{K})$, where the covariance function $\mathcal{K}$

is

$$\mathcal{K}(x,x') = \sigma^2 \exp\left(\frac{-\|x-x'\|_2^2}{2l^2}\right) + \sigma_n^2 \delta_{x,x'}, \tag{4}$$

with to-be-determined kernel parameters (l, σ, σ_n). Suppose that we have observations $\boldsymbol{y} = (y^{(1)}, \ldots, y^{(N)})$ at locations $\boldsymbol{X} = (x^{(1)}, \ldots, x^{(N)})$ and use a Gaussian process to fit these data. To tailor the kernel function, we minimise the negative log-likelihood as a function of the kernel parameters, i.e.,

$$\min_{l,\sigma,\sigma_n} \frac{1}{2}\left((\boldsymbol{y}-\boldsymbol{\xi}_0)^T K^{-1}(\boldsymbol{y}-\boldsymbol{\xi}_0) + \log|K| + N\log(2\pi)\right), \tag{5}$$

where $\boldsymbol{\xi}_0 = (\xi_0(x^{(1)}), \ldots, \xi_0(x^{(N)}))$, and K denotes the covariance matrix for the points in $\boldsymbol{X}$. We also require that the posterior ξ_{post} satisfies a joint chance constraint for the lower bound $\underline{\xi}$ with high probability p, $0 < p < 1$, i.e.,

$$\mathbb{P}(\omega | \underline{\xi}(x) \leq \xi_{\text{post}}(x,\omega) \text{ for a.a. } x) \geq p. \tag{6}$$

This problem resembles the stochastic optimal control problem (1), (2), and we again apply the SRD to enforce the joint chance constraint (6). In (3), we now use a Karhunen-Loève (KL) expansion of the covariance operator corresponding to the kernel function (4), and truncate this KL expansion after m terms. The variable $\boldsymbol{\theta}$ then follows a uniform distribution on $\mathbb{S}^{m-1}$, and we use a large number of grid points at which we enforce the bound constraint. For our method, it is necessary to make an explicit choice for m and use the m-variable χ-distribution in the SRD. Although the SRD method is not very sensitive to the choice of m, we would prefer to formulate and use the SRD in infinite dimensions to avoid making an explicit choice for m. However, this is challenging, as currently the SRD requires the transformation of a Gaussian random variable into a random variable with a unit covariance matrix (see (3)). Choosing $m = 40$, the numerical results obtained with our SRD-based method applied to Example 1 from [4] can be seen in Figure 2. In these calculations, the derivatives with respect to the kernel parameter, which are needed for the minimisation of (5) are calculated using automatic differentiation as provided by JAX.

References

[1] M. H. Farshbaf-Shaker, R. Henrion, and D. Hömberg, *Properties of chance constraints in infinite dimensions with an application to PDE-constrained optimization*, Set-Valued and Variational Analysis, **26** (2018), 821–841.

[2] A. Geletu, A. Hoffmann, P. Schmidt, and P. Li, *Chance constrained optimization of elliptic PDE systems with a smoothing convex approximation*, ESAIM: COCV, **26** (2020), 70.

[3] R. Henrion, G. Stadler, and F. Wechsung, *Optimal control under uncertainty with joint chance state constraints: almost-everywhere bounds, variance reduction, and application to (bi-)linear elliptic PDEs*, arXiv:2412.05125, to appear in JUQ (2025).

[4] A. Pensoneault, X. Yang, and X. Zhu, *Nonnegativity-enforced Gaussian process regression*, Theoretical and Applied Mechanics Letters, **10** (2020), 182–187.

[5] L. Swiler, M. Gulian, A. Frankel, C. Safta, and J. D. Jakeman, *A survey of constrained Gaussian process regression: Approaches and implementation challenges*, Journal of Machine Learning for Modeling and Computing **1(2)** (2020), 119–156.

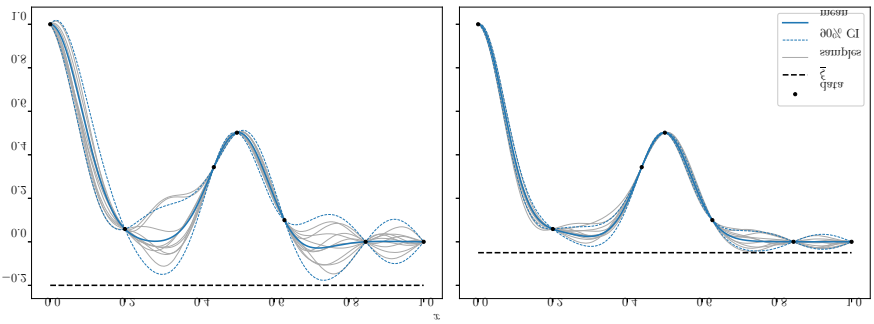


FIGURE 2. Gaussian process with optimised kernel parameters such that realizations from the posterior satisfy the joint chance bound constraint with probability $p = 0.95$. Shown are the data points (black), the mean of the posterior (blue) and random draws from the posterior (in grey). The right figure uses a tighter lower bound $\underline{\xi}$, requiring kernel parameter resulting in lower variance of the posterior samples.

Learning to integrate the uncertainty of diffusion equations with Lévy coefficients

HANNO GOTTSCHALK

(joint work with Oliver Ernst, Toni Kowalewitz, and P. Krüger)

Uncertainty quantification often requires multiple simulations of physical systems at a high computational cost [11]. A well studied example is the case of the diffusion equation with random conductivity modeled by a random field.

$$-\nabla \cdot (a(x)\nabla)u(x) = 0,\ a(x) = \exp(Z_K(x)),$$

where mixed Neumann and Dirichlet boundary conditions are applied on different portions of the boundary of the computational domain D, which is assumed to be bounded and to have sufficiently regular boundary.

Under the hypothesis that the random field $Z_K(x)$ is Gaussian, one can apply the Karhunen–Loève (KL) expansion [11] and approximate the law of the Gaussian random field with a dimensional normal distributions. For these, quadrature rules like Smolyak's sparse grid [9] allow an evaluation of the expected values on the number of quadrature points that scales favourably with respect to the dimension of the KL expansion.

In this extended abstract, we reconsider the diffusion equation with Lévy random fields [1, 2] in place of the Gaussian random fields. We therefore have two crucial problems to solve. First, we have to develop the KL expansion for Lévy random fields, and, second, we have to design quadrature rules, which work for this extended class of Lévy distributions.

Starting with the first problem, we consider a solution $Z_K(x)$ of stochastic pseudo partial differential equation driven by Lévy noise $Z(x)$ [1, 5],

$$(-\Delta + m^2)^\alpha Z_K(x) = Z(x),$$

where Δ the Laplace operator and m^2, α positive constants.

The solution of this equation can be can be obtained applying the green function of pseudo differential operator $(-\Delta + m^2)^\alpha$ to the Lévy noise generalised random field [6]. As Lévy random fields are not determined by their co-variance function, we have to apply the expansion directly to the integral kernel of the Green's function. Unfortunately, this is not possible as pseudo differential operator has continuous spectrum.

We therefore first apply a circular embedding on a torus which leads to a Laplace operator with discrete spectrum and hence the Green's function can now be expanded into an eigenvalue decomposition where the eigenfunctions are the trigonometric functions $\sin(\kappa \cdot x)$ and $\cos(\kappa \cdot x)$ with wave vectors κ inside a discrete grid $\pi\mathbb{Z}^2$. The eigenvalues are then obtained by inserting the wave vectors into the symbol $(|\kappa|^2 + m^2)^{-\alpha}$ of the Green's function. It has been shown in prior work that under this expansion the solution to the solution of the diffusion equation converges to the solution with the non-truncated Lévy random field [5].

The second problem, how to evaluate the expected value of some quantity of interest that depends on the random solution $u(x)$, however remains, as there are no efficient quadrature rules for high-dimensional Lévy distributions.

In our work, we tackle this problem with recent machine learning models from the field of generative AI. In generative AI, some complex distribution of data is transformed into a simple distribution often given by a standard normal distribution. In other words, generative AI learns a transport map [10] that normalizes complex distributions to multivariate standard normal distribution. There are many variants [4, 3, 7], how such normalising flows can be modeled by your networks. In our work, we use affine coupling flows, flow matching and optimal transport flow matching as algorithms and apply them to the given high-dimensional Lévy distribution. All these normalising flow networks are easy to invert and admit an easy to evaluate representation of the log-likelihood. Alternatively, the flows are modeled in continuous time and the training objective is to establish a similarity of the temporal derivatives of the flow with some pre-defined transport vector field [7]. Hence, after appropriate likelihood maximisation or vector field matching based training on the Lévy data, we can not only approximately transform the Levy distribution into Gaussian noise, but we can also transform Gaussian noise approximately back into the Lévy distribution.

This enables us to use Smolyak's [9] sparse grid quadrature points and weights for the standard normal distribution to evaluate the quantity of interest composed with the generative direction of our machine learning model. Equivalently, we can transport the quadrature points of the standard normal distribution into quadrature points of the involved Lévy distributions by application of the generative map to the quadrature points, keeping the weights fixed.

In this way, we obtain learned quadrature rules on which we then can evaluate the quantity of interest by first performing the simulations and then evaluating the quantity of interest on the solution with the same computational complexity as for original sparse grids.

We also provide extensive numerical experiments that show that the evaluations of the expected values nicely reproduce the results of a brute-force Monte Carlo simulation with hundred thousand samples, provided the approximate level distributions do not contain discrete components, or have restricted support.

References

[1] S. Albeverio, H. Gottschalk, and J.-L. Wu, *Convoluted generalized white noise, Schwinger functions and their analytic continuation to Wightman functions*, Reviews in Mathematical Physics, **8** (1996), 763–817.

[2] D. Applebaum, *Lévy processes and stochastic calculus*, Cambridge University Press, 2009.

[3] T. Chen, Y. Rubanova, J. Bettencourt, and D. K. Duvenaud, *Neural ordinary differential equations*, Advances in neural information processing systems, **31** (2018).

[4] L. Dinh, J. Sohl-Dickstein, and S. Bengio, *Density estimation using Real NVP*, International Conference on Learning Representations, (2017).

[5] O. G. Ernst, H. Gottschalk, T. Kalmes, T. Kowalewitz, and M. Reese, *Integrability and approximability of solutions to the stationary diffusion equation with Lévy coefficient*, 2021, arXiv:2010.14912v3.

[6] K. Itô, *Foundations of stochastic differential equations in infinite dimensional spaces*, SIAM, 1984.

[7] Y. Lipman, R. T. Chen, H. Ben-Hamu, M. Nickel, and M. Le, *Flow matching for generative modeling*, International Conference on Learning Representations, (2023).

[8] X. Liu, C. Gong, and Q. Liu, *Flow straight and fast: Learning to generate and transfer data with rectified flow*, International Conference on Learning Representations, (2023).

[9] E. Novak and K. Ritter, *High dimensional integration of smooth functions over cubes*, Numerische Mathematik **75** (1996), 79–97.

[10] F. Santambrogio, *Optimal transport for applied mathematicians*, Springer, (2015).

[11] T. J. Sullivan, *Introduction to uncertainty quantification*, Springer, (2015).

[12] A. Tong, K. Fatras, N. Malkin, G. Huguet, Y. Zhang, J. Rector-Brooks, G. Wolf, and Y. Bengio, *Improving and generalizing flow-based generative models with minibatch optimal transport*, Transactions on Machine Learning Research, (2024).

Approximation to elliptic PDEs with high contrast diffusion coefficients

Matthieu Dolbeault

(joint work with Albert Cohen, Wolfgang Dahmen, and Agustin Somacal)

We consider the elliptic diffusion equation

$$\begin{cases} -\mathrm{div} \cdot a\nabla u &= f \qquad \text{in } \Omega \\ u &= 0 \qquad \text{on } \partial\Omega \end{cases}$$

for some domain $\Omega \subset \mathbb{R}^d$, a fixed source term $f \in H^{-1}(\Omega)$, and a piecewise constant diffusion coefficient $a : \Omega \to \mathbb{R}_+^*$. This setting occurs for instance when modelling the thermal properties of a heterogeneous material: a takes large values on conductive components of Ω, and small values in insulating parts.

Our interest lies in understanding how the solution u depends on the values taken by a. This question arises in uncertainty quantification, when the material properties are not precisely known, and one wishes to estimate the statistical distribution of a quantity of interest, such as the maximal temperature or the heat flow through a part of the boundary. It is also relevant for the construction of preconditioners, since classical multilevel methods [1] fail to converge in the high-contrast regime

$$\frac{\max_{x\in\Omega} a(x)}{\min_{x\in\Omega} a(x)} \gg 1.$$

To describe the similarity between solutions, we resort to reduced order modelling: given the partition of Ω into subdomains $\Omega_1, \dots, \Omega_p$, we consider the set of all solutions

$$\mathcal{K} = \left\{ u(a) \ : \ a = \sum_{j=1}^{p} a_j \mathbb{1}_{\Omega_j} \ , \ (a_1, \dots, a_p) \in (0,\infty)^p \right\}$$

and estimate its size by bounding the Kolmogorov widths

$$d_n(\mathcal{K})_{H_0^1} = \inf_{\dim(V_n)=n} \sup_{u\in\mathcal{K}} \min_{v\in V_n} \|u-v\|_{H_0^1},$$

that is, the distance between class $\mathcal{K}$ and the linear subspace $V_n \subset H_0^1(\Omega)$ that best approximates it.

When the values of a are bounded from above and below, it is known from [2] that the Kolmogorov widths decay as $\exp(-cn^{1/p})$ when n goes to infinity. In the unbounded case, we prove in [3] that

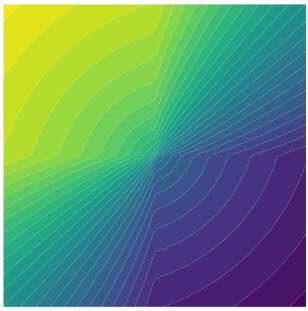
$$d_n(\mathcal{K})_{H_0^1} \leq C \exp(-cn^{1/2p})$$

for some constants C, c that depend on the geometry of the partition.

In fact, the above result only holds under a geometric assumption, which is also required in the construction of multilevel preconditioners adapted to the high contrast regime [4]. In an ongoing work with Markus Bachmayr [5], we try to single out the cases where this geometric assumption is not satisfied, in order to treat them separately. It turns out that when Ω is of dimension $d = 2$, the only difficulty is the situation where four subdomains join at a point. Simplifying the geometry even further, we analyse the case where $\Omega = [-1,1]^2$ is split along the axes into four smaller squares. Leveraging the homogeneity and the symmetries with respect to the vertical and horizontal axes, we only need to study a one-parameter family of solutions, depicted on the image below.

Writing the equation in radial coordinates allows to construct explicit a-harmonic solutions [6], and the set of these solutions has exponentially decaying Kolmogorov widths. Therefore, we expect this rate of convergence to also hold for more general geometries, which would lead to improvements in reduced-order models and preconditioners for the diffusion equation.

Ω_1 $a_1 \gg 1$	Ω_2 $a_2 = 1$
Ω_3 $a_3 = 1$	Ω_4 $a_4 = a_1$



References

[1] J. H. Bramble, J. E. Pasciak and J. Xu, *Parallel multilevel preconditioners*, Mathematics of computation, **55** (1990), 1–22.

[2] M. Bachmayr and A. Cohen, *Kolmogorov widths and low-rank approximations of parametric elliptic PDEs*, Mathematics of Computation, **86** (2017), 701–724.

[3] A. Cohen, W. Dahmen, M. Dolbeault and A. Somacal, *Reduced order modeling for elliptic problems with high contrast diffusion coefficients*, ESAIM: Mathematical Modelling and Numerical Analysis, **57** (2023), 2775–2802.

[4] P. Oswald, *On the Robustness of the BPX-Preconditioner with Respect to Jumps in the Coefficients*, Mathematics of Computation, **68** (1999), 633–650.

[5] M. Bachmayr and M. Dolbeault, *High contrast diffusion with singular geometries*, in preparation.

[6] B. R. Kellogg, *On the Poisson equation with intersecting interfaces*, Applicable Analysis, **4** (1974), 101–129.

Feedback control under uncertainty

Philipp A. Guth

(joint work with Peter Kritzer and Karl Kunisch)

The approximation of a Riccati-based feedback control law for optimal control problems constrained by parametric partial differential equations (PDEs) is studied. Parametric PDEs arise, for instance, from countably infinite series expansions of random or uncertain coefficients in PDEs. To do so, given a sequence of parameters $\boldsymbol{\sigma} = (\sigma)_{j\geq 1} \in [-1,1]^{\mathbb{N}}$ with i.i.d. uniformly distributed components, let us consider the parameter-dependent optimal control problem

$$\min_{u,y} \mathcal{J}(u,y), \quad \mathcal{J}(u,y) := \frac{1}{2}\Bigg(\int_0^T \left(\mathcal{R}\left(\|C(y(\cdot;t) - g(t))\|_H^2\right) + \|u(t)\|_U^2\right) \mathrm{d}t + \mathcal{R}\left(\|P(y(\cdot;T) - g_T)\|_H^2\right)\Bigg), \tag{1}$$

subject to

$$\dot{y}(\boldsymbol{\sigma};t) = A(\boldsymbol{\sigma})y(\boldsymbol{\sigma};t) + Bu(t) + f(t) \quad y(\boldsymbol{\sigma};0) = y_0. \tag{2}$$

Given a time horizon $T > 0$, an initial condition $y(\boldsymbol{\sigma};0) = y_0 \in H$, and an external forcing $f \in L^2(0,T;V')$, the goal is to find a control input $u \in L^2(0,T;U)$ which steers the parameter-dependent state $y(\boldsymbol{\sigma})$ as close as possible to the targets $g \in \mathcal{C}([0,T];H)$ and $g_T \in H$. The time evolution of the system is described by the parameter-dependent $A(\boldsymbol{\sigma}) \in \mathcal{L}(V,V')$ acting on the current state, and $B \in \mathcal{L}(U,H)$ acting on the input control. The operator $C \in \mathcal{L}(H)$ is an observation operator and $P \in \mathcal{L}(H)$.

The following two risk measures are studied: the risk-neutral expected value $\mathcal{R} = \mathbb{E}$, and the risk-averse entropic risk measure $\mathcal{R} = \mathcal{R}_\theta$, defined as

$$\mathcal{R}_\theta(X) := \frac{1}{\theta}\log\left(\mathbb{E}\left[\mathrm{e}^{\theta X}\right]\right), \quad \theta > 0.$$

The expected values are given as infinite-dimensional parametric integrals $\mathbb{E}[X] = \int_{[-1,1]^{\mathbb{N}}} X(\boldsymbol{\sigma})\,\mathrm{d}\boldsymbol{\sigma}$, with respect to the product probability measure $\mathrm{d}\boldsymbol{\sigma} := \bigotimes_{j\in\mathbb{N}} \frac{\mathrm{d}\sigma_j}{2}$.

In the risk-averse case, a quadratic approximation of the entropic risk measure about $\bar{y} \in \mathcal{C}([0,T];L^\infty(\mathfrak{S};H))$ leads to the linear quadratic subproblem $\min_{u,y} J_{\mathrm{quad}}$, subject to (2), where

$$\begin{aligned} J_{\mathrm{quad}}(u,y) = \frac{1}{2}\Bigg(\int_0^T \Big(\mathbb{E}\left[\|\mathcal{Q}_{C,\omega}(\bar{y}(\cdot;t))^{1/2}(y(\cdot;t) - \widetilde{g}(\cdot;t))\|_H^2 \right] + \|u(t)\|_U^2 \Big)\,\mathrm{d}t, \\ + \mathbb{E}\left[\|\mathcal{Q}_{P,\omega}(\bar{y}(\cdot;T))^{1/2}(y(\cdot;T) - \widetilde{g}_T(\cdot))\|_H^2 \right] \Bigg) \end{aligned}$$

with target $\widetilde{g}(\boldsymbol{\sigma};t) = \bar{y}(\boldsymbol{\sigma};t) - \mathcal{Q}_{C,\omega}(\bar{y}(\boldsymbol{\sigma};t))^{-1}\mathscr{R}_C(\bar{y}(\boldsymbol{\sigma};t);t)$ and terminal target $\widetilde{g}_T(\boldsymbol{\sigma}) = \bar{y}(\boldsymbol{\sigma};T) - \mathcal{Q}_{P,\omega}(\bar{y}(\boldsymbol{\sigma};T);T)^{-1}\mathscr{R}_P(\bar{y}(\boldsymbol{\sigma};T);T)$. The operators $\mathcal{Q}_{C,\omega}(\bar{y};t) \in \mathcal{L}(L^2(\mathfrak{S};H))$ and $\mathcal{Q}_{P,\omega}(\bar{y};T) \in \mathcal{L}(L^2(\mathfrak{S};H))$ are given as

$$\begin{aligned} \langle \mathcal{Q}_{C,\omega}(\bar{y}(\cdot;t);t)\delta_1,\delta_2\rangle_{L^2(\mathfrak{S};H)} &= \mathbb{E}_{\omega_{\theta,C}(\bar{y}(\boldsymbol{\sigma};t))}\left[\langle C^*C\delta_2(\cdot;t),\delta_1(\cdot;t)\rangle_H\right] \\ &+ 2\theta\,\mathrm{Cov}_{\omega_{\theta,C}(y(\boldsymbol{\sigma};t))}\Big(\langle\overline{\mathscr{C}}(\cdot;t),\delta_2(\cdot;t)\rangle_H, \langle\overline{\mathscr{C}}(\cdot;t),\delta_1(\cdot;t)\rangle_H\Big) \end{aligned}$$

for almost every $t \in [0,T]$ and

$$\begin{aligned} \langle \mathcal{Q}_{P,\omega}(\bar{y}(\cdot;T);T)\delta_1,\delta_2\rangle_{L^2(\mathfrak{S};H)} &= \mathbb{E}_{\omega_{\theta,P}(\bar{y}(\boldsymbol{\sigma};T))}\left[\langle P^*P\delta_2(\cdot;T),\delta_1(\cdot;T)\rangle_H\right] \\ &+ 2\theta\,\mathrm{Cov}_{\omega_{\theta,P}(y(\boldsymbol{\sigma};T))}\Big(\langle\overline{\mathscr{P}}(\cdot;T),\delta_2(\cdot;T)\rangle_H, \langle\overline{\mathscr{P}}(\cdot;T),\delta_1(\cdot;T)\rangle_H\Big), \end{aligned}$$

where we use $\overline{\mathscr{C}}(\boldsymbol{\sigma};t) := C^*C(\bar{y}(\boldsymbol{\sigma};t) - g(t))$ and $\overline{\mathscr{P}}(\boldsymbol{\sigma};T) := P^*P(\bar{y}(\boldsymbol{\sigma};T) - g_T)$, $\mathscr{R}_C(\bar{y}(\boldsymbol{\sigma};t);t) := \overline{\mathscr{C}}(\cdot;t)\omega_{\theta,C}(\bar{y}(\boldsymbol{\sigma};t))$, $\mathscr{R}_P(\bar{y}(\boldsymbol{\sigma};T);T) := \overline{\mathscr{P}}(\cdot;T)\omega_{\theta,P}(\bar{y}(\boldsymbol{\sigma};T))$, and the expected values as well as the covariances are respectively weighted by

$$\omega_{\theta,C}(y(\boldsymbol{\sigma};t)) := \frac{\mathrm{e}^{\theta\|C(y(\boldsymbol{\sigma};t)-g(t))\|_H^2}}{\mathbb{E}\left[\mathrm{e}^{\theta\|C(y(\cdot;t)-g(t))\|_H^2}\right]}, \text{ and } \omega_{\theta,P}(y(\boldsymbol{\sigma};T)) = \frac{\mathrm{e}^{\theta\|P(y(\boldsymbol{\sigma};T)-g_T)\|_H^2}}{\mathbb{E}\left[\mathrm{e}^{\theta\|P(y(\cdot;T)-g_T)\|_H^2}\right]}.$$

For $\theta \to 0$, the quadratic approximation coincides with the risk-neutral case.

Furthermore, the optimality system of the quadratic approximation is a Newton step for the original problem (1)–(2). Thus, repeatedly solving the quadratic

approximation with updated expansion points results in a so-called sequential quadratic programming (SQP) method [4] for solving the original problem. More precisely, the sequence of minimisers $\{(u_\star^{(k)}, y_\star^{(k)})\}_{k\geq 0}$ generated by repeatedly solving the linear quadratic subproblems with updated expansion points $\bar{y}^{(k)} = y_\star^{(k-1)}$ converges locally quadratically to the unique minimiser of the original problem (1)–(2) provided that the initial guess is sufficiently close to the minimiser. Particularly, this procedure finds an approximation of the risk-averse optimal control in feedback form.

In [5] it has been shown that the optimal control and optimal state-adjoint-state pair of a parametric linear quadratic optimal control problem are real analytic functions with respect to the parameter sequence under the assumptions that the family $\{A(\boldsymbol{\sigma}) \mid \boldsymbol{\sigma} \in [-1,1]^{\mathbb{N}}\}$ has a uniformly bounded inverse and that $\|\partial_{\boldsymbol{\sigma}}^{\boldsymbol{\nu}} A(\boldsymbol{\sigma})\|_{\mathcal{L}(V,V')} \leq \boldsymbol{b}^{\boldsymbol{\nu}}$ for a monotonically decreasing sequence $\boldsymbol{b} = (b_j)_{j\in\mathbb{N}}$, which is p-summable for some $p \in (0,1)$. Here the following notation is used: for a sequence $\boldsymbol{b} := (b_j)_{j\in\mathbb{N}}$ of real numbers and $\boldsymbol{\nu} \in \{\boldsymbol{m} \in \mathbb{N}_0^{\mathbb{N}} \mid \sum_{j\geq 1} m_j < \infty\}$, define $\partial_{\boldsymbol{\sigma}}^{\boldsymbol{\nu}} := \frac{\partial^{\nu_1}}{\partial\sigma_1}\frac{\partial^{\nu_2}}{\partial\sigma_2}\cdots$, and $\boldsymbol{\sigma}^{\boldsymbol{\nu}} := \prod_{j=1}^{\infty} \sigma_j^{\nu_j}$, with the convention $0^0 := 1$. Based on this result, in [8] it is shown that a Riccati-based feedback law for an autonomous system $(A(\cdot), B, C, P)$ admits the same analytic regularity. Such regularity results are frequently obtained in the context of random field expansions, and can be used for the numerical analysis of the problem; this includes the so-called dimension truncation [6] (i.e., the truncation of the parameter sequence to $\boldsymbol{\sigma}_{\leq s} := (\sigma_1, \ldots, \sigma_s, 0, 0, \ldots)$), the approximation of the high-dimensional integrals over the parameters using higher-order methods, such as quasi-Monte Carlo methods [3] or sparse-grid methods [1], as well as approximations by generalised polynomial chaos expansions [2] or deep neural networks [9].

While the parametric regularity for the problem (1)–(2) with the entropic risk measure is analysed in [7], the analysis of the quadratic approximation of this problem, with a non-autonomous Riccati equation as well as a parameter-dependent target, is an interesting open research direction and would enable the design of efficient algorithms to approximate a feedback law in the risk-averse formulation.

References

[1] H.-J. Bungartz and M. Griebel, *Sparse Grids*, Acta Numer. **13** (2004), 147–269.

[2] A. Cohen and R. DeVore, *Approximation of high-dimensional parametric PDEs*, Acta Numer. **24** (2015), 1–159.

[3] J. Dick, F. Y. Kuo, I. H. Sloan, *High-dimensional integration: The quasi-Monte Carlo way*, Acta Numer. **22** (2013), 133–288.

[4] K. Ito and K. Kunisch, *Lagrange Multiplier Approach to Variational Problems and Applications*, Society for Industrial and Applied Mathematics, USA, 2008.

[5] A. Kunoth and Ch. Schwab, *Analytic regularity and GPC approximation for control problems constrained by linear parametric elliptic and parabolic PDEs*, SIAM J. Control Optim. **51** (2013), 2442–2471.

[6] P. A. Guth and V. Kaarnioja, *Generalized Dimension Truncation Error Analysis for High-Dimensional Numerical Integration: Lognormal Setting and Beyond*, SIAM J. Numer. Anal. **62** (2024), 872–892.

[7] P. A. Guth and V. Kaarnioja and F. Y. Kuo and C. Schillings and I. H. Sloan, *Parabolic PDE-constrained optimal control under uncertainty with entropic risk measure using quasi-Monte Carlo integration*, Numer. Math. **156** (2024), 565–608.
[8] P. A. Guth and P. Kritzer and K. Kunisch, *Quasi-Monte Carlo integration for feedback control under uncertainty*, preprint at arXiv:2409.15537 [math.OC], 2024.
[9] Ch. Schwab and J. Zech, *Deep learning in high dimension: Neural network expression rates for generalized polynomial chaos expansions in UQ*, Anal. Appl. **17** (2019), 19–55.

(Near-)Optimality of Quasi-Monte Carlo Methods and Suboptimality of the Sparse-Grid Gauss–Hermite Rule in Gaussian Sobolev Spaces

Yoshihito Kazashi

(joint work with Takashi Goda and Yuya Suzuki)

We study the numerical integration of multivariate functions over $\mathbb{R}^d$ in high dimensions, with particular focus on integration with respect to the standard Gaussian measure:

$$I(f) := \int_{\mathbb{R}^d} f(\boldsymbol{x})\rho(\boldsymbol{x})\mathrm{d}\boldsymbol{x} \approx Q_N(f),$$

where $\rho(\boldsymbol{x}) := \frac{\mathrm{e}^{-|\boldsymbol{x}|^2/2}}{(2\pi)^{d/2}}$, and $|\boldsymbol{x}| = \sqrt{x_1^2 + \cdots + x_d^2}$ for $\boldsymbol{x} \in \mathbb{R}^d$ denotes the Euclidean norm, and Q_N is a suitable numerical integration rule using N evaluations of f.

In this work, we show that, in terms of convergence rate, sparse-grid Gauss–Hermite quadrature is suboptimal, whereas several quasi-Monte Carlo (QMC) methods are optimal. More precisely, for L^2-Sobolev spaces of an integer degree α, the sparse-grid Gauss–Hermite rule achieves a convergence rate $O(N^{-\alpha/2})$ up to a logarithmic factor; several QMC methods together with change of variables achieve $O(N^{-\alpha})$ up to a logarithmic factor. The rate $O(N^{-\alpha})$ is in fact optimal, as there exists a matching lower bound established by Dick et al. [1] for general numerical integration rules using N-point evaluations. In this sense, QMC methods considered herein are optimal. In contrast, we also show a lower bound of the rate $N^{-\alpha/2}$ for the sparse-grid Gauss–Hermite rule, showing that this rate is unimprovable more than a logarithmic factor.

Such a gap in convergence rates between the sparse-grid Gauss–Hermite rule and QMC methods has also been observed numerically by Dick et al. [1] and in subsequent work by Nuyens and one of the collaborators of the present work [3]. Our theoretical results provide a rigorous explanation of these empirical findings.

We work within the Sobolev space defined by the norm:

$$\|f\|_{H^\alpha_\rho} := \left(\sum_{|\boldsymbol{r}|_\infty \le \alpha} \|D^{\boldsymbol{r}} f\|_{L^2_\rho}^2\right)^{\frac{1}{2}}$$

where $\|g\|_{L^2_\rho} := \sqrt{\int_{\mathbb{R}^d} |g(\boldsymbol{x})|^2 \rho(\boldsymbol{x})\mathrm{d}\boldsymbol{x}}$ is the weighted L^2-norm. Note that the summation is taken over multi-indices $\boldsymbol{r}$ with respect to the maximum norm $|\cdot|_\infty$, in contrast to the more commonly used ℓ_1-norm $|\cdot|_1$. This space is also known as the Hermite space of finite smoothness.

Sparse-grid quadrature based on the Gauss–Hermite rule

Let S_{Λ_L} be the sparse–grid quadrature based on the Gauss–Hermite rule, associated with the index set $\Lambda_L = \{\boldsymbol{\ell} \in \mathbb{N}^d \mid \mathbf{1} \le \boldsymbol{\ell}, |\boldsymbol{\ell}|_1 \le L\}$. Let Q_ℓ^{uni}, for $\ell \in \mathbb{N}$, denote the Gauss–Hermite rule for univariate functions with n_ℓ quadrature points. With $\Delta_\ell(f) := Q_\ell^{\mathrm{uni}}(f) - Q_{\ell-1}^{\mathrm{uni}}(f)$ for $\ell \in \mathbb{N}$ and $\Delta_0 := 0$, the algorithm S_{Λ_L} is of the form

$$S_{\Lambda_L}(f) = \sum_{(\ell_1,\dots,\ell_d)\in\Lambda_L} \big(\Delta_{\ell_1} \otimes \cdots \otimes \Delta_{\ell_d}\big)(f),$$

where $\Delta_{\ell_1} \otimes \cdots \otimes \Delta_{\ell_d}$ denotes the product rule defined by the difference quadratures, i.e., each quadrature Δ_{ℓ_j} acts on f by treating it as a univariate function in the j-th variable, with all other variables held fixed.

Suppose that we choose n_ℓ as

$$n_\ell \asymp M^\ell \qquad \text{and} \qquad n_{L-d+1} \ge \mathrm{e}^{d-1},$$

for some $M > 1$. Let N denote the resulting total number of the function evaluations used in S_{Λ_L}. Then, we show

$$\sup_{0\neq f\in H_\rho^\alpha} \frac{|I(f) - S_{\Lambda_L}(f)|}{\|f\|_{H_\rho^\alpha}} \ge c_{\alpha,d} N^{-\alpha/2} (\log N)^{\frac{\alpha}{2}(d-1)}.$$

where the constant $c_{\alpha,d} > 0$ is independent of L and N.

We also show an analogous lower bound for more general, downward-closed index set, for which we have a lower bound of order $N^{-\alpha/2}$. These lower bounds are built upon out previous result in one dimension [2].

The polynomial factor $N^{-\alpha/2}$ is sharp. Indeed, we can show a matching upper bound up to a logarithmic factor. Namely, for $n_\ell \asymp M^\ell$ we have

$$\sup_{0\neq f\in H_\rho^\alpha} \frac{|I(f) - S_{\Lambda_L}(f)|}{\|f\|_{H_\rho^\alpha}} \le c_{\alpha,d} (\log N)^{(1+\frac{\alpha}{2})(d-1)} N^{-\alpha/2},$$

where the constant $c_{\alpha,d}$ is independent of f.

Quasi-Monte Carlo with Change of Variables

A quasi-Monte Carlo method is an equal-weight integration rule. Such methods are typically defined over the unit cube $[0,1]^d$, i.e.,

$$\frac{1}{N}\sum_{j=1}^N f(\boldsymbol{t}_j) \approx \int_{[0,1]^d} f(\boldsymbol{x})\mathrm{d}\boldsymbol{x},$$

where the numerical integration points $\boldsymbol{t}_j \in [0,1]^d$, $j = 1,\dots,N$, are chosen suitably. Popular classes for the choice of $\boldsymbol{t}_j$ include lattice points, higher-order digital nets, and low discrepancy points/sequences (these classes are not mutually exclusive).

As mentioned above, QMC points are typically defined on the unit cube $[0,1]^d$. To integrate functions over $\mathbb{R}^d$, we therefore apply a change of variables:

$$\begin{aligned}\int_{\mathbb{R}^d} f(\boldsymbol{x})\,\rho(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} &= \int_{[0,1]^d} f(\Psi(\boldsymbol{t}))\,\rho(\Psi(\boldsymbol{t}))\,|\det(D\Psi(\boldsymbol{t}))|\,\mathrm{d}\boldsymbol{t}\\ &\approx \frac{1}{N}\sum_{j=1}^{N} f\left(\Psi(\boldsymbol{t}_j)\right)\rho\left(\Psi(\boldsymbol{t}_j)\right)|\det(D\Psi(\boldsymbol{t}_j)| =: Q_N(f). \qquad (1)\end{aligned}$$

We work with non-negative component-wise mappings, $\Psi(\boldsymbol{t}) = (\Psi_1(t^1), ..., \Psi_d(t^d))$, in which case the Jacobian factor simplifies as

$$|\det(D\Psi(\boldsymbol{t}))| = \prod_{k=1}^{d} \Psi_k'(t^k).$$

In this work, we study two types of change of variables. The first is the affine map considered in [1]. For $\boldsymbol{b} = (b, \ldots, b) \in (0,\infty)^d$, define the isotropic affine transformation $\Psi_{\mathrm{affine},b} : [0,1]^d \to \prod_{j=1}^{d}[-b,b] \subset \mathbb{R}^d$ by

$$\Psi_{\mathrm{affine},b}(\boldsymbol{t}) = 2b\boldsymbol{t} - \boldsymbol{b}.$$

Dick et al. [1] showed that with $b = \sqrt{\alpha \log N}$, one can construct a higher order digital net such that the corresponding method (1) achieves the convergence rate $O(\frac{(\log N)^{s\frac{2\alpha+3}{4}-\frac{1}{2}}}{N^\alpha})$ for functions in H^α_ρ.

The other change of variables we consider is a co-tangent Möbius transformation considered in [4]. Here we define our change of variables $\Psi_{\mathrm{cotan}} : [0,1]^d \to \mathbb{R}^d$ by

$$\Psi_{\mathrm{cotan}}(\boldsymbol{t}) \coloneqq (\phi(t^1), \ldots, \phi(t^d)), \quad \phi(t) \coloneqq -\cot(2\pi t).$$

It can be shown that good rank-1 lattice rules attain the convergence rate $N^{-\alpha}(\ln N)^{d\alpha}$ and higher-order digital nets attain the rate

$$O(N^{-\alpha}(\log N)^{(d-1)/2}). \qquad (2)$$

We reiterate that Dick et al. [1] also showed that the polynomial factor $N^{-\alpha}$ is the best possible among any numerical integration using N-point evaluations as information about the integral. Consequently, these rates achieved by QMC methods with change of variables are not improvable beyond a logarithmic factor. In fact, the lower bound established therein is of order $N^{-\alpha}(\log N)^{(d-1)/2}$, so the rate (2) achieved by the higher-order digital nets with the co-tangent transform is optimal including the logarithmic factor.

References

[1] Dick, Josef, Christian Irrgeher, Gunther Leobacher, and Friedrich Pillichshammer, *On the optimal order of integration in Hermite spaces with finite smoothness*, SIAM Journal on Numerical Analysis **56**, no. 2 (2018): 684-707.

[2] Kazashi, Yoshihito, Yuya Suzuki, and Takashi Goda, *Suboptimality of Gauss–Hermite quadrature and optimality of the trapezoidal rule for functions with finite smoothness*, SIAM Journal on Numerical Analysis **61**, no. 3 (2023): 1426-1448.

[3] Nuyens, Dirk, and Yuya Suzuki, *Scaled lattice rules for integration on $\mathbb{R}^d$ achieving higher-order convergence with error analysis in terms of orthogonal projections onto periodic spaces*, Mathematics of Computation **92**, no. 339 (2023): 307-347.
[4] Suzuki, Yuya, Nuutti Hyvönen, and Toni Karvonen, *Möbius-transformed trapezoidal rule*, Mathematics of Computation (2025).

Reporter: Hefin Lambley

Participants

Prof. Dr. Andrea Barth
Institut für Angewandte Analysis und Numerische Simulation
Abteilung für Computational Methods for Uncertainty Quantification
Universität Stuttgart
Allmandring 5b
70569 Stuttgart
GERMANY

Dr. Joshua Chen
Oden Institute for Computational Engineering and Sciences
The University of Texas at Austin
1 University Station C1200
Austin TX 78712-1082
UNITED STATES

Prof. Dr. Alexey Chernov
Institut für Mathematik
Carl von Ossietzky Universität Oldenburg
26111 Oldenburg
GERMANY

Prof. Dr. Imma Valentina Curato
Fakultät für Mathematik
TU Chemnitz
Postfach 964
09009 Chemnitz
GERMANY

Prof. Dr. Ana Djurdjevac
Freie Universität Berlin
Arnimallee 6
14195 Berlin
GERMANY

Dr. Matthieu Dolbeault
Institut für Geometrie und Praktische Mathematik
RWTH Aachen
Templergraben 55
52062 Aachen
GERMANY

Prof. Dr. Oliver Ernst
Fakultät für Mathematik
Technische Universität Chemnitz
09107 Chemnitz
GERMANY

Prof. Dr. Colin Fox
Department of Physics
University of Otago
P.O. Box 56
Dunedin 9054
NEW ZEALAND

Prof. Dr. Caroline Geiersbach
Department Mathematik
Universität Hamburg
Bundesstr. 55
20146 Hamburg
GERMANY

Prof. Dr. Hanno Gottschalk
Fachbereich Mathematik, Sekr.MA 8-5
Technische Universität Berlin
Straße des 17. Juni 136
10623 Berlin
GERMANY

Dr. Philipp Guth
Johann Radon Institute for Computational and Applied Mathematics
Austrian Academy of Sciences
4040 Linz
AUSTRIA

Prof. Dr. Helmut Harbrecht
Departement Mathematik und Informatik
Universität Basel
Spiegelgasse 1
4051 Basel
SWITZERLAND

Prof. Dr. Tapio Helin
Department of Computational Engineering
LUT University
53851 Lappeenranta
FINLAND

Ly Duc Hoang
Rudower Chaussee 25, Room 1.211
Fachbereich Mathematik
Humboldt Universität Berlin
12161 Berlin
GERMANY

Prof. Dr. Eyke Hüllermeier
Ludwig-Maximilians-Universität München
Akademiestr. 7
80799 München
GERMANY

Dr. Yoshihito Kazashi
Department of Mathematics & Statistics
University of Strathclyde
Livingstone Tower
26, Richmond Street
Glasgow G1 1XH
UNITED KINGDOM

Dr. Hanne Kekkonen
Delft Institute of Applied Mathematics, TU Delft
Mekelweg 4
Delft 2628 CD
NETHERLANDS

Dr. Kristin Kirchner
Delft Institute of Applied Mathematics, Delft University of Technology
P.O. Box 5031
2600 GA Delft
NETHERLANDS

Dr. Ilja Klebanov
FB Mathematik u. Informatik
Freie Universität Berlin
14195 Berlin
GERMANY

Dr. Oliver Koenig
Institute of Applied Analysis and Numerical Simulation
Department of Mathematics
Universität Stuttgart
Allmandring 5b
70569 Stuttgart
GERMANY

Hefin Lambley
Mathematics Institute
University of Warwick
Gibbet Hill Road
Coventry CV4 7AL
UNITED KINGDOM

Dr. Jonas Latz
Department of Mathematics
University of Manchester
Alan Turing Building, Oxford Road
Manchester M13 9PL
UNITED KINGDOM

Prof. Dr. Han Cheng Lie
Institut für Mathematik
Universität Potsdam
Campus Golm
Karl-Liebknecht-Str 24-25
14476 Potsdam
GERMANY

Dr. Kerstin Lux-Gottschalk
Department of Mathematics &
Computer Science
Eindhoven University of Technology
De Groene Loper 5
5612 AZ Eindhoven
NETHERLANDS

Prof. Dr. Youssef Marzouk
Center for Computational Science and
Engineering
Laboratory for Information and Decision
Systems
Massachusetts Institute of Technology
77 Massachusetts Avenue
Cambridge MA 02139
UNITED STATES

Dr. Rebecca Morrison
Department of Computer Science
University of Colorado
Boulder 80309
UNITED STATES

Prof. Dr. Fabio Nobile
CSQI - MATH
École Polytechnique Fédérale de
Lausanne
Station 8
1015 Lausanne
SWITZERLAND

Prof. Dr. Houman Owhadi
Department of Computing and
Mathematical Sciences
California Institute of Technology
MC 9-94
1200 E California blvd
Pasadena CA 91125
UNITED STATES

Chiara Piazzola
TU München
School of Computation, Information and
Technology
Boltzmannstr. 3
85748 München
GERMANY

Moritz Poguntke
Fakultät für Mathematik
TU Chemnitz
Postfach 964
09009 Chemnitz
GERMANY

Prof. Dr. Sebastian Reich
Institut für Mathematik
Universität Potsdam
Karl-Liebknecht-Straße 24-25
14476 Potsdam
GERMANY

Prof. Dr. Markus Reiß
Institut für Mathematik
Humboldt-Universität Berlin
Unter den Linden 6
10117 Berlin
GERMANY

Prof. Dr. Daniel Rudolf
Fakultät für Mathematik
und Informatik
Universität Passau
Innstr. 33
94032 Passau
GERMANY

Dr. Havard Rue
4700 King Abdullah University of
Science
and Technology (KAUST)
Mathematical & Computer Sciences
P.O. Box 2360
23955-6900 Jeddah
SAUDI ARABIA

Dr. Florian Schäfer
School of Computational Science and Engineering
Georgia Institute of Technology
S1317 CODA
756 W Peachtree St
Atlanta GA 30332
UNITED STATES

Prof. Dr. Claudia Schillings
Freie Universität Berlin
Arnimallee 6
12159 Berlin
GERMANY

Prof. Dr. Christoph Schwab
Seminar für Angewandte Mathematik
ETH Zurich
ETH Zentrum, HG G 57.1
Rämistrasse 101
8092 Zürich
SWITZERLAND

Prof. Dr. Björn Sprungk
Technische Universität Bergakademie Freiberg
Fakultät für Mathematik und Informatik
Institut für Stochastik
09596 Freiberg
GERMANY

Prof. Dr. Georg Stadler
Courant Institute of Mathematical Sciences
New York University
251, Mercer Street
New York NY 10012-1110
UNITED STATES

Prof. Dr. Claudia Strauch
Institut für Mathematik
Universität Heidelberg
Im Neuenheimer Feld 205
69120 Heidelberg
GERMANY

Dr. Tim J. Sullivan
Mathematics Institute
University of Warwick
Gibbet Hill Road
Coventry CV4 7AL
UNITED KINGDOM

Dr. Lorenzo Tamellini
CNR-IMATI
Via Ferrata 1
27100 Pavia
ITALY

Prof. Dr. Raul F. Tempone
Chair of Mathematics for Uncertainty Quantification
RWTH Aachen
Pontdriesch 14-16
52062 Aachen
GERMANY

Prof. Dr. Elisabeth Ullmann
Department Mathematik
Technische Universität München
Boltzmannstraße 3
85748 Garching bei München
GERMANY

Prof. Dr. Sven Wang
Institut für Mathematik
Fachbereich Mathematik
Humboldt-Universität Berlin
10099 Berlin
GERMANY

Dr. Simon Weissmann
Universität Mannheim
68159 Mannheim
GERMANY

Dana Wrischnig
Fachbereich Mathematik & Informatik
Freie Universität Berlin
Arnimallee 6
14195 Berlin
GERMANY

Prof. Dr. Jakob Zech
Faculty of Mathematics and Computer Science
Heidelberg University
Im Neuenheimer Feld 205
69115 Heidelberg
GERMANY