Report from Dagstuhl Seminar 21432

Probabilistic Numerical Methods – From Theory to Implementation

Edited by

Philipp Hennig[1], Ilse C.F. Ipsen[2], Maren Mahsereci[3], and Tim Sullivan[4]

1 **Universität Tübingen, DE,** philipp.hennig@uni-tuebingen.de
2 **North Carolina State University – Raleigh, US,** ipsen@ncsu.edu
3 **Universität Tübingen, DE,** maren.mahsereci@uni-tuebingen.de
4 **University of Warwick – Coventry, GB,** t.j.sullivan@warwick.ac.uk

Abstract

Numerical methods provide the computational foundation of science, and power automated data analysis and inference in its contemporary form of machine learning. Probabilistic numerical methods aim to explicitly represent uncertainty resulting from limited computational resources and imprecise inputs in these models. With theoretical analysis well underway, software development is now a key next step to wide-spread success. This seminar brought together experts from the forefront of machine learning, statistics and numerical analysis to identify important open problems in the field and to lay the theoretical and practical foundation for a software stack for probabilistic numerical methods.

Seminar October 24–29, 2021 – http://www.dagstuhl.de/21432
2012 ACM Subject Classification Computing methodologies → Machine learning; Mathematics of computing → Numerical analysis
Keywords and phrases Machine learning, Numerical analysis, Probabilistic numerics
Digital Object Identifier 10.4230/DagRep.11.9.102
Edited in cooperation with Jonathan Wenger

1 Executive Summary

Philipp Hennig (Universität Tübingen, DE)
Ilse C.F. Ipsen (North Carolina State University – Raleigh, US)
Maren Mahsereci (Universität Tübingen, DE)
Tim Sullivan (University of Warwick – Coventry, GB)
Jonathan Wenger (Universität Tübingen, DE)

Probabilistic Numerical algorithms frame a numerical task as a statistical inference problem, expressed in the language of probabilistic inference. The key advantage of this approach is that it allows quantification of uncertainty arising from finite computational resources, and to combine thus with other forms of uncertainty, in particular those arising from model misspecification, finite observational data, and measurement errors. In recent years, algorithms arising from this formalism have repeatedly shown that they can enrich and improve upon classic methods in tasks where

Probabilistic Numerical Methods – From Theory to Implementation, *Dagstuhl Reports*, Vol. 11, Issue 09, pp. 102–119
Editors: Philipp Hennig, Ilse C.F. Ipsen, Maren Mahsereci, and Tim Sullivan

Dagstuhl Reports
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

- hyperparameter adaptation is not straightforward;
- computational stochasticity and low precision play a prominent role;
- limited data make uncertainty quantification a key functionality;
- related problems have to be solved repeatedly;
- and where extreme scale or tight budgets call for rough approximations at low cost.

Probabilistic Numerics lies at the intersection of machine learning within computer science and numerical analysis within applied mathematics. This interdisciplinary nature raises an exciting and challenging set of viewpoints with regards to goals and challenges of the field. The first goal of this seminar was to rekindle our community following two years of pandemic lockdown, to provide an opportunity to update others on one's own research, and to discuss new directions and ideas together. We were lucky to assemble – both in-person and remote – a diverse group of people from computer science, machine learning, from statistics, optimization, and from numerical analysis.

The second key goal of this seminar was to take the next step in the development of probabilistic numerical methods by focusing on their implementation. From `Lapack` to `SciPy` to `PyTorch`, open-source software libraries have driven scientific advancement in their respective domains. Such libraries accelerate research, enable benchmarking and promote the development of new methods via rapid prototyping. Most importantly, they are a necessary step towards their use in applications. While considerable advances in the theoretical understanding of probabilistic numerical methods have been made, the lack of high-quality implementations is holding back their adoption. In response, we recently started a community effort to develop an open-source framework named `ProbNum` (`http://probnum.org`).

A central theme of Dagstuhl seminars is the open, collaborative atmosphere with a focus on new ideas and tangible outcomes as opposed to existing work. The seminar stimulated multiple focussed discussions around software and additions to ProbNum. Examples included how to best include automatic differentiation functionality, or how to expand the package's Bayesian quadrature functionalities. It also set the starting point for potential new research collaborations on probabilistic linear solvers and probabilistic numerical methods for PDEs. Even at this point, shortly after the seminar's conclusion, two tangible products are already available: the seminar's participants jointly created a Probabilistic Numerics Wikipedia page `https://en.wikipedia.org/wiki/Probabilistic_numerics`, and the implementation sessions culminated with a preprint for the community library `ProbNum` [1].

As the organizers we want to thank all participants, both physical and virtual, for their interesting talks, the stimulating discussions and the collaborative overall atmosphere. We also want to thank Schloss Dagstuhl for their technical support that made the challenging hybrid format possible.

References

1 Jonathan Wenger, Nicholas Krämer, Marvin Pförtner, Jonathan Schmidt, Nathanael Bosch, Nina Effenberger, Johannes Zenn, Toni Karvonen Alexandra Gessner, François-Xavier Briol, Maren Mahsereci, and Philipp Hennig. ProbNum: Probabilistic numerics in Python. arxiv preprint, 2021. URL `http://arxiv.org/abs/2112.02100`

2 Table of Contents

3 Overview of Talks

3.1 Approximate Gaussian Process Regression as an Early Stopping Problem

Simon Bartels (University of Copenhagen, DK)

Joint work of Simon Bartels, Pablo Moreno-Munoz, Kristoffer Stensbo-Smidt, Wouter Boomsma, Jes Frellsen, Soren Hauberg

This talk is about a method to fit Gaussian process regression models to large datasets from only a subset of the data. The novelty of this approach is that the size of the subset is selected on the fly during inference with little computational overhead. This is achieved by monitoring probabilistic bounds on the model evidence that tighten as more data is processed. Remarkably, these bounds are largely composed of terms that appear in intermediate steps of the standard Cholesky decomposition, allowing to adaptively stop the decomposition once enough data have been observed.

3.2 ProbNumDiffEq.jl: Fast and Practical ODE Filters in Julia

Nathanael Bosch (Universität Tübingen, DE)

Joint work of Nathanael Bosch, Filip Tronarp, Philipp Hennig

ProbNumDiffEq.jl provides an implementation of ODE filters in Julia, building on top of the established DifferentialEquations.jl ecosystem. In this session, we discuss how software for probabilistic numerics (PN) can benefit from existing non-PN code to obtain higher performance, a larger set of features, and to reach a wider audience. This is facilitated by the composability of the Julia programming language, and the modular structure of DifferentialEquations.jl. We demonstrate the speed and ease of use of the resulting ODE solvers in a live code demo. Additionally, we show how ODE filters can be extended to include additional knowledge about the problem structure (such as second-order ODEs or energy conservation) and demonstrate how to solve such problems in practice.

3.3 Iterative Unbiased Linear Solvers for Gaussian Processes

Maurizio Filippone (EURECOM – Biot, FR)

Joint work of Maurizio Filippone, Raphael Engler
Main reference Maurizio Filippone, Raphael Engler: "Enabling scalable stochastic gradient-based inference for Gaussian processes by employing the Unbiased LInear System SolvEr (ULISSE)", in Proc. of the 32nd International Conference on Machine Learning, ICML 2015, Lille, France, 6-11 July 2015, JMLR Workshop and Conference Proceedings, Vol. 37, pp. 1015–1024, JMLR.org, 2015.
URL http://proceedings.mlr.press/v37/filippone15.html

In applications of Gaussian processes where quantification of uncertainty is of primary interest, it is necessary to accurately characterize the posterior distribution over covariance parameters. This paper proposes an adaptation of the Stochastic Gradient Langevin Dynamics algorithm

to draw samples from the posterior distribution over covariance parameters with negligible bias and without the need to compute the marginal likelihood. In Gaussian process regression, this has the enormous advantage that stochastic gradients can be computed by solving linear systems only. A novel unbiased linear systems solver based on parallelizable covariance matrix-vector products is developed to accelerate the unbiased estimation of gradients. The results demonstrate the possibility to enable scalable and exact (in a Monte Carlo sense) quantification of uncertainty in Gaussian processes without imposing any special structure on the covariance or reducing the number of input vectors.

3.4 Bayesian Cubature with Low Discrepancy Sequences in QMCPy

Fred J. Hickernell (Illinois Institute of Technology – Chicago, US)
Jagadeeswaran Rathinavel (Illinois Institute of Technology – Chicago, US)
Aleksei Sorokin (Illinois Institute of Technology – Chicago, US)

Bayesian cubature proceeds by constructing credible intervals for the integral based on a prior distribution for the integrand combined with sampled integrand values. We explain this approach has been developed using low discrepancy (digital net and lattice) sampling and matching covariance kernels to expedite the computation to be much faster than $\mathcal{O}(n^3)$, where n is the sample size. We tune the hyper-parameters of our covariance kernels using function data to increase the chance that the integrand is not an outlier. We also show how we have implemented these Bayesian cubature algorithms in QMCPy, a community supported Python library for quasi-Monte Carlo calculations. Some preliminary results also call into question the Bayesian assumption. This matter requires further study.

3.5 BayesCG: A probablistic numeric linear solver

Ilse C.F. Ipsen (North Carolina State University – Raleigh, US)

Joint work of Ilse C.F. Ipsen, Tim W. Reid, Jon Cockayne, Chris J. Oates
Main reference Tim W. Reid, Ilse C. F. Ipsen, Jon Cockayne, Chris J. Oates: "A Probabilistic Numerical Extension of the Conjugate Gradient Method", CoRR, Vol. abs/2008.03225, 2020.
URL https://arxiv.org/abs/2008.03225

We present the probabilistic linear solver BayesCG, an extension of the Conjugate Gradient method (CG) that relies on probability distributions to capture uncertainty due to early termination when solving linear systems with real symmetric positive definite coefficient matrices. We present a CG-based implementation of BayesCG with a structure-exploiting prior distribution. The BayesCG output consists of CG iterates and posterior covariances that can be propagated to subsequent computations. The covariances are low-rank and maintained in factored form. This allows easy generation of accurate samples to probe uncertainty in subsequent computations. We discuss the choice of efficient prior distributions, and end with speculation on how to propagate uncertainty through computational pipelines.

3.6 bayesquad: Bayesian quadrature in ProbNum

Toni Karvonen (University of Helsinki, FI)
Alexandra Gessner (Universität Tübingen, DE)

We present the current state of the implementation of Bayesian quadrature in ProbNum and discuss features that are to be included in the future. The talk contains examples of the use of the two main Bayesian quadrature functions, `bayesquad` and `bayesquad_from_data`, and their comparison to the corresponding SciPy functions.

3.7 Fun with ODE filters

Peter Nicholas Krämer (Universität Tübingen, DE)

Joint work of Peter Nicholas Krämer, Nathanael Bosch, Jonathan Schmidt, Philipp Hennig
Main reference Nicholas Krämer, Philipp Hennig: "Stable Implementation of Probabilistic ODE Solvers", CoRR, Vol. abs/2012.10106, 2020.
URL https://arxiv.org/abs/2012.10106

This talk surveys recent advances of probabilistic solvers for differential equations. At first, stable implementation of probabilistic, filtering-based ODE solvers is discussed; both, in low- and high-dimensional settings. Then, efficient probabilistic solvers for ODE boundary value problems and partial differnetial equations are explained.

3.8 ProbNum: Probabilistic Numerical Methods in Python

Maren Mahsereci (Universität Tübingen, DE)
Jonathan Wenger (Universität Tübingen, DE)

Joint work of Jonathan Wenger, Nicholas Krämer, Marvin Pförtner, Jonathan Schmidt, Nathanael Bosch, Nina Effenberger, Johannes Zenn, Alexandra Gessner, Toni Karvonen, François-Xavier Briol, Maren Mahsereci, Philipp Hennig
URL http://probnum.org

ProbNum is a Python library that provides probabilistic numerical solvers to a wider audience. In the talk, we describe the current state and functionality of ProbNum and highlight some benefits of open source collaboration for students and for the community. The second part of the talk contains a live demonstration of some of the ProbNum solvers.

3.9 Solving and Learning Differential Equations with Gaussian Processes

Houman Owhadi (California Institute of Technology – Pasadena, US)

Joint work of Houman Owhadi, Yifan Chen, Boumediene Hamzi, Bamdad Hosseini, Romit Maulik, Yoo Gene Ryan, Florian Schäfer, Clint Scovel, Andrew Stuart

Main reference Yifan Chen, Bamdad Hosseini, Houman Owhadi, Andrew M. Stuart: "Solving and learning nonlinear PDEs with Gaussian processes", J. Comput. Phys., Vol. 447, p. 110668, 2021.
URL https://doi.org/10.1016/j.jcp.2021.110668

Main reference B. Hamzi, R. Maulik, H. Owhadi: "Simple, low-cost and accurate data-driven geophysical forecasting with learned kernels", Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences, Vol. 477(2252), p. 20210326, 2021.
URL https://doi.org/10.1098/rspa.2021.0326

Main reference Boumediene Hamzi, Houman Owhadi: "Learning dynamical systems from data: A simple cross-validation perspective, part I: Parametric kernel flows", Physica D: Nonlinear Phenomena, Vol. 421, p. 132817, 2021.
URL https://doi.org/10.1016/j.physd.2020.132817

Main reference Houman Owhadi, Clint Scovel: "Operator-Adapted Wavelets, Fast Solvers, and Numerical Homogenization: From a Game Theoretic Approach to Numerical Approximation and Algorithm Design", Cambridge University Press, 2019.
URL https://doi.org/10.1017/9781108594967

Main reference Houman Owhadi, Gene Ryan Yoo: "Kernel Flows: From learning kernels from data into the abyss", J. Comput. Phys., Vol. 389, pp. 22–47, 2019.
URL https://doi.org/10.1016/j.jcp.2019.03.040

We present a simple, rigorous, and unified framework for solving and learning (possibly nonlinear) differential equations (PDEs and ODEs) using the framework of Gaussian processes/kernel methods. For PDEs the proposed approach: (1) provides a natural generalization of collocation kernel methods to nonlinear PDEs and Inverse Problems; (2) has guaranteed convergence for a very general class of PDEs, and comes equipped with a path to compute error bounds for specific PDE approximations; (3) inherits the state-of-the-art computational complexity of linear solvers for dense kernel matrices. For ODEs, we illustrate the efficacy of the proposed approach by extrapolating weather/climate time series obtained from satellite data and highlight the importance of using adapted/learned kernels.

3.10 Implementing BayesCG Under The Krylov Prior

Timothy Reid (North Carolina State University – Raleigh, US)

Joint work of Timothy Reid, Ilse C.F. Ipsen, Jon Cockayne, Chris J Oates

Main reference Tim W. Reid, Ilse C. F. Ipsen, Jon Cockayne, Chris J. Oates: "A Probabilistic Numerical Extension of the Conjugate Gradient Method", CoRR, Vol. abs/2008.03225, 2020.
URL https://arxiv.org/abs/2008.03225

We present solutions to the computational challenges associated with implementing BayesCG under the Krylov prior. We also have a discussion of possible solutions to open questions related to implementing probabilistic numerical linear solvers.

3.11 A Probabilistic State Space Model for Joint Inference from Differential Equations and Data

Jonathan Schmidt (Universität Tübingen, DE)

Joint work of Jonathan Schmidt, Nicholas Krämer, Philipp Hennig
Main reference Jonathan Schmidt, Nicholas Krämer, Philipp Hennig: "A Probabilistic State Space Model for Joint Inference from Differential Equations and Data", CoRR, Vol. abs/2103.10153, 2021.
URL https://arxiv.org/abs/2103.10153

This talk shows how different sources of information – mechanistic knowledge and empirical observations – can be combined to solve differential equations that are assumed to underlie observed data. The mismatch between both sources of knowledge is captured by introducing a latent force acting on the vector field. The method is showcased by fitting an SIRD-model to COVID-19 case counts.

3.12 A Probabilistic View on Sparse Cholesky Factorization

Florian Schäfer (Georgia Institute of Technology – Atlanta, US)

Joint work of Jiong Chen, Mathieu Desbrun, Jin Huang, Matthias Katzfuss, Houman Owhadi, Florian Schäfer, Tim J. Sullivan
Main reference Florian Schäfer, Houman Owhadi: "Sparse recovery of elliptic solvers from matrix-vector products", CoRR, Vol. abs/2110.05351, 2021.
URL https://arxiv.org/abs/2110.05351
Main reference Florian Schäfer, Timothy John Sullivan, Houman Owhadi: "Compression, inversion, and approximate PCA of dense kernel matrices at near-linear computational complexity", CoRR, Vol. abs/1706.02205, 2017.
URL http://arxiv.org/abs/1706.02205
Main reference Florian Schäfer, Matthias Katzfuss, Houman Owhadi: "Sparse Cholesky factorization by Kullback-Leibler minimization", CoRR, Vol. abs/2004.14455, 2020.
URL https://arxiv.org/abs/2004.14455
Main reference Jiong Chen, Florian Schäfer, Jin Huang, Mathieu Desbrun: "Multiscale cholesky preconditioning for ill-conditioned problems", ACM Trans. Graph., Vol. 40(4), pp. 81:1–81:13, 2021.
URL https://doi.org/10.1145/3450626.3459851

The guiding theme of this talk is the probabilistic interpretation of numerical linear algebra, in particular of the Cholesky factorization of kernel matrices and discretized elliptic partial differential equations. By using this interpretation, we derive simple, fast solvers with state-of-the-art complexity vs. accuracy guarantees for general elliptic differential- and integral equations. We furthermore derive an algorithm that allows the reconstruction, or learning, of elliptic solution operators from a number of solution pairs that scales only polylogarithmically in the target accuracy. Our methods come with rigorous error estimates, are easy to parallelize, and show good performance in practice.

3.13 Convergence and Robustness of Gaussian Process Regression

Aretha Teckentrup (University of Edinburgh, GB)

Main reference Aretha L. Teckentrup: "Convergence of Gaussian Process Regression with Estimated Hyper-Parameters and Applications in Bayesian Inverse Problems", SIAM/ASA J. Uncertain. Quantification, Vol. 8(4), pp. 1310–1337, 2020.
URL https://doi.org/10.1137/19M1284816

We are interested in the task of estimating an unknown function from data, given as a set of point evaluations. In this context, Gaussian process regression is often used as a Bayesian inference procedure, and we are interested in the convergence as the number of data points goes to infinity. Hyper-parameters appearing in the mean and covariance structure of the Gaussian process prior, such as smoothness of the function and typical length scales, are often unknown and learnt from the data, along with the posterior mean and covariance. We work in the framework of empirical Bayes', where a point estimate of the hyper-parameters is computed, using the data, and then used within the standard Gaussian process prior to posterior update. Using results from scattered data approximation, we provide a convergence analysis of the method applied to a fixed, unknown function of interest.

3.14 Black Box Probabilistic Numerics

Onur Teymur (University of Kent, GB)

Joint work of Onur Teymur, Christopher N. Foley, Philip G. Breen, Toni Karvonen, Chris. J. Oates
Main reference Onur Teymur, Christopher N. Foley, Philip G. Breen, Toni Karvonen, Chris J. Oates: "Black Box Probabilistic Numerics", CoRR, Vol. abs/2106.13718, 2021.
URL https://arxiv.org/abs/2106.13718

In many numerical algorithms, intermediate numerical outputs are nonlinearly related to the quantity of interest, rendering the proper conditioning of random variables in the probabilistic numerics paradigm difficult and limiting the range of numerical tasks that can be addressed by existing approaches. In this presentation we introduce an idea to construct probabilistic numerical methods based only on the final output from a traditional method. A convergent sequence of approximations to the quantity of interest constitute a dataset, from which the limiting quantity of interest can be extrapolated, in a probabilistic analogue of Richardson's deferred approach to the limit. This black box approach (1) massively expands the range of tasks to which probabilistic numerics can be applied, (2) inherits the features and performance of state-of-the-art numerical methods, and (3) enables provably higher orders of convergence to be achieved. We present several proof-of-concept applications, such as for nonlinear ordinary and partial differential equations, as well as for eigenvalue problems – the latter a setting for which no probabilistic numerical methods have yet been developed.

3.15 The MAP for ODEs

Filip Tronarp (Universität Tübingen, DE)

Joint work of Filip Tronarp, Simo Särkkä, Philipp Hennig
Main reference Filip Tronarp, Simo Särkkä, Philipp Hennig: "Bayesian ODE solvers: the maximum a posteriori estimate", Stat. Comput., Vol. 31(3), p. 23, 2021.
URL https://doi.org/10.1007/s11222-021-09993-7

There is a growing interest in probabilistic numerical solutions to ordinary differential equations. In this talk, the maximum a posteriori estimate is studied under the class of ν times differentiable linear time-invariant Gauss-Markov priors. The maximum a posteriori estimate corresponds to an optimal interpolant in the reproducing kernel Hilbert space associated with the prior, which in the present case is equivalent to a Sobolev space of smoothness ν. Subject to mild conditions on the vector field, convergence rates of the maximum a posteriori estimate are then obtained via methods from nonlinear analysis and scattered data approximation.

3.16 Priors in Probabilistic Numerics

Zi Wang (Google – Cambridge, US)

Joint work of Zi Wang, George E. Dahl, Kevin Swersky, Chansoo Lee, Zelda Mariet, Zackary Nado, Justin Gilmer, Jasper Snoek, Zoubin Ghahramani
Main reference Zi Wang, Caelan Reed Garrett, Leslie Pack Kaelbling, Tomás Lozano-Pérez. "Learning compositional models of robot skills for task and motion planning." Int. J. Robotics Res. 40(6-7) (2021).
Zi Wang, George E. Dahl, Kevin Swersky, Chansoo Lee, Zelda Mariet, Zackary Nado, Justin Gilmer, Jasper Snoek, and Zoubin Ghahramani. "Automatic prior selection for meta Bayesian optimization with a case study on tuning deep neural network optimizers." arXiv preprint arXiv:2109.08215 (2021).

How do we understand priors? Surrounding this question, I discuss some of my thoughts on different forms of priors and how they impact the design and performance of algorithms. In particular, I group priors into 3 categories: priors in nature, engineering priors and data priors. Data priors are very relevant to how we may be able to set priors better in probabilistic numerics. I use Bayesian optimization to illustrate how priors estimated from multi-task data can lead to better performance than hand-selected priors.

4 Working groups

4.1 Probabilistic Numerics Wikipedia Page

François-Xavier Briol (University College London, GB)

This working group focused on developing a Wikipedia page for the field of probabilistic numerics. The aim was to help raise the visibility of the field, but also to serve as a brief introduction for non-experts. Overall, this was very successful; less than a month after it was first created, the page has been viewed around 2000 times and has 30 distinct authors. A second page for the sub-field of Bayesian quadrature is also currently undergoing the Wikipedia approval process.

4.2 ProbNum Library Scope

Jonathan Wenger (Universität Tübingen, DE)

ProbNum implements probabilistic numerical methods in Python. Such methods solve numerical problems from linear algebra, optimization, quadrature and differential equations using probabilistic inference. This session discusses the exact use cases of a PN library and defines the separation from other libraries for classic numerics, probabilistic programming and uncertainty quantification.

4.3 From (Prior) Information to Usable Algorithm

Jonathan Wenger (Universität Tübingen, DE)

Prior information is often touted as a prime example of why probabilistic numerical methods have an advantage over classical methods. But what types of prior information are available for different numerical problems and what can we actually encode?

5 Panel discussions

5.1 Bayesian Optimization

Roman Garnett (Washington University – St. Louis, US)

Main reference Roman Garnett: "Bayesian Optimization", Cambridge University Press, 2022.
URL https://bayesoptbook.com

I discussed a number of issues related to Bayesian optimization (BayesOpt) and themes behind several of the success stories in that field. I identified one major theme in the BayesOpt literature – a focus on expensive functions – which has allowed research in that field consider correspondingly expensive approaches to modeling and policy design. In probabilistic numerics (PN), on the other hand, there seems to be a general trend away from this "expensive regime." I concluded by posing three questions to the group:

- To what extent is the fidelity of modeling in PN? Is the "expensive regime" of interest?
- To what extent is the sophistication/myopia of algorithms an issue in PN? Is the "expensive regime" of interest?
- How can we improve PN pipelines through user interaction?

An engaging discussion ensued.

5.2 Scientific Software Development

Andrei Paleyes (University of Cambridge, GB)
Maren Mahsereci (Universität Tübingen, DE)
Michael McKerns (Los Alamos National Laboratory, US)
Masha Naslidnyk (Amazon Research Cambridge, GB)
Geoff Pleiss (Columbia University – New York, US)
Jonathan Wenger (Universität Tübingen, DE)

Recent years saw the high amount of scientific software produced by research groups in academia and industry. A good software package can be an asset of a research group, increasing productivity of its members and enabling faster experimentation. However, development and maintenance of a high quality software can be a challenging task. Recent trends and challenges of scientific software development were discussed by the panel at the Dagstuhl seminar 21432 "Probabilistic Numerical Methods – From Theory to Implementation". This document is the summary of the discussion.

5.2.1 Introduction

Over the past 6 years the proportion of papers in machine learning that are released along with the codebase used to produce the results tripled, growing from under 10% in 2015 to nearly 30% in 2021[1]. While some of these codebases can remain unchanged since their release, there is also a growing tendency to turn them into reusable software components. This decision can bring benefits, but also comes at a cost. The panel discussed challenges of developing and maintaining software tools for academic and industry teams, associated trade offs and past experience panelists had with their software packages.

5.2.2 Should scientists worry about developing software or just use existing tools?

While development of a new scientific tool does not need to be a goal of each research project, it may bring certain benefits worth pursuing. Software that is intended to be reused over multiple research projects or over several years tends to be well tested, thus further ensuring quality of the scientific results produced. Long-lived software also tends to be well documented, thus enabling others to reproduce results from prior work, and to directly compare their own new results produced with the software. Finally, good scientific software can accelerate research and the development of other software produced by a research group. A research group can leverage good scientific software to provide a solid reusable software foundation to extend, with the benefit of several years of development, testing, and validation, rather than starting from scratch on each new project.

Since research groups are normally built around a certain research direction, a software package may also emerge naturally, even though the group never had such a goal, by accumulating reusable components from multiple projects done by the group. A research group may decide to produce and maintain their own software, however an equally viable (and potentially more fruitful) choice is to adopt an existing open source package, and contribute to it's development. Interestingly, the decision to contribute to an existing codebase as opposed to developing a new software package from scratch is unfortunately rare.

[1] According to `https://paperswithcode.com/trends`, accessed 21.11.2021

5.2.3 What are challenges in maintaining scientific software?

Scientific code does not always have to be production ready, and therefore does not always need to uphold high standards in terms of software design or test coverage. In that sense, the academic environment can be liberating. However, when this code is turned into a reusable toolbox, a certain level of quality is anticipated. This can be a challenge for researchers, as most of them do not have software engineering experience, and therefore might not be aware of best practices. Having access to a trained software developer might help, but again not every lab can afford hiring one. For that reason a lot of researchers in machine learning have to learn engineering skills on the fly in addition to producing research in their primary topic. This can be a steep and challenging learning curve.

In years past, a researcher learning good software development in an academic environment was often at the cost of an academic career, as the generation of software products was not seen as part of countable academic achievements. However, the ability to produce good software concurrent with academic research is now a highly-desired skill, even though a record of software contributions is still generally measured unfavorably against a record of journal publications. As journals and funding agencies have begun to expect research groups to develop and extend software when they produce new research results, the ability for an academic environment to attract and retain researchers with good software dvelopment skills has increased.

Maintaining software beyond the original funding source, or the original developer, can be challenging. Long-lived research software like mystic [1, 2] sustain steady development by being central to new proposals and new research funding. Software tied to project funding can pay researchers to serve as developers as part of their jobs, as opposed to requiring additional efforts outside of their funded work.

5.2.4 Can you hope to compete against industrial research labs when writing a software package?

At the first glance it may seem impossible for an academic research group to compete with industry when it comes to software development. Teams in industry are often well funded, have access to better infrastructure and attract significant software engineering talent. In fact, the competition is not as severe as it may seem. Companies normally only fund engineering effort that is able to support their business goals. As such, software engineering projects created by industrial researchers are aimed at deploying well established research ideas in practical applications. Groups in universities and research centers, on the other hand, are working on the cutting edge research ideas, and therefore are more likely to create a software project in a direction, which is not ready for production yet.

National research laboratories have traditionally been a source of reusable scientific software, as researchers often work on multiple well-funded projects with expected deliverables and shortened timelines, and thus have similar conditions to those found in industry. Projects at national laboratories often have industry partners, where the groups at the national laboratory are tasked with providing high-quality cutting-edge research tools.

5.2.5 How do you tread the line between spending too much time on maintenance, implementing features for others and moving your own research forward?

Navigating the balance between maintenance and novel development is always a trade off. However, while maintenance will always demand certain time investment from the package owner, there are several general rules that can alleviate the load.

Most importantly, time invested in software design always pays off. Simple high level API means less errors due to an improper use. An extensible software architecture means more straightforward contributions that are easier to review. Good test coverage means fewer bugs. All of these are good engineering practices that also help reduce the maintenance load. As an example, the author of dill[2] and mystic maintains a range of open source scientific packages, and the overall load is only manageable because each package was developed with accordance to good engineering practices where the reusability of the software enables it to be directly tied to new research funding.

Successful software packages can attract a strong community of users and contributors. Such a community can, to a certain extent, share the load of maintenance, as is the case now with GPyTorch [3]. Clearly such a community requires time and effort to gather, and therefore can be seen as a long-term investment. Even after the initial release of Emukit [4] its authors still continue to promote the package both in academia and in industry, which helps growing the community of its users.

However there can be cases when authors of the package are no longer interested in spending their time and effort in maintaining it. It is important to detect such situations when they happen, and identify the best way forward, which can be even discontinuing maintenance of the package, as is now the case with GpyOpt [5].

5.2.6 How do you scale a project from a small group to an open-source project with outside contributors?

As already mentioned, scientific packages can become a primary tool for a an research group. This link can also work in the other direction, where a successful research idea can be implemented as a new functionality of the package. This setup works well particularly for student and intern projects, and may even have an amplifying effect, where a scientist who contributed to a project joins another research group and may attract further users. Declaring contribution as an additional goal of a research project helps to accelerate growth during the early stages.

Once the package gains a certain level of maturity, it may attract interest from the industry partners. They can benefit by applying the scientific methods implemented in the package to their business problems. The level of contribution that companies can bring back varies widely, from feature requests and bug reports, to the contribution of individual features, to financial support for the library and its authors.

An ideal scenario is finding industrial collaborators who wish to build software on top of your package. For example, Facebook built the BoTorch Bayesian optimization package [6] on top of the GPyTorch package. As a result, the BoTorch team has a vested interest in ensuring that GPyTorch is maintained, yet ownership still ultimately resides with the original GPyTorch team. It can be challenging to find these collaborations, and companies are more likely to reach out if the software package is actively maintained by responsive developers.

However, at each stage of development it is important to define the purpose and goals of the package and therefore clearly separate what functionality is in- and out of scope.

References

1 M. McKerns, P. Hung, and M. Aivazis. mystic: highly-constrained non-convex optimization and UQ. `http://pypi.python.org/pypi/mystic`, 2009.

[2] https://dill.readthedocs.io/en/latest/dill.html

2 M. McKerns, L. Strand, T. J. Sullivan, A. Fang, and M. Aivazis. Building aframework for predictive science. In Proceedings of the 10th Python in Science Conference, pages 67-78, 2011. http://arxiv.org/pdf/1202.1056.
3 J. Gardner, G. Pleiss, K. Q. Weinberger, D. Bindel, and A. G. Wilson. GPyTorch: Blackbox matrix-matrix gaussian process inference with GPU acceleration. Advances in Neural Information Processing Systems, 31, 2018.
4 A. Paleyes, M. Pullin, M. Mahsereci, N. Lawrence, and J. González. Emulation of physical processes with Emukit. In Second Workshop on Machine Learning and the Physical Sciences, NeurIPS, 2019.
5 The GPyOpt authors. GPyOpt: A Bayesian optimization framework in Python. http://github.com/SheffieldML/GPyOpt, 2016
6 M. Balandat, B. Karrer, D. R. Jiang, S. Daulton, B. Letham, A. G. Wilson,and E. Bakshy. BoTorch: A Framework for Efficient Monte-Carlo Bayesian Optimization. In Advances in Neural Information Processing Systems, 2020.

6 Open problems

6.1 Calibration of PN methods

Giacomo Garegnani (EPFL – Lausanne, CH)

One of the main goals underlying the development of probabilistic numerical methods is that they help to quantify the uncertainty due to approximate computations. In order for this purpose to be fulfilled, probabilistic methods should yield consistent global or local information on the error, or in other words they should be well calibrated. In this session, I will try to define what a well-calibrated probabilistic numerical methods should be, in particular by highlighting the connections with traditional a posteriori error estimators.

Participants

- Simon Bartels
University of Copenhagen – Copenhagen, DK
- Nathanael Bosch
Universität Tübingen, DE
- François-Xavier Briol
University College London, GB
- Maurizio Filippone
EURECOM – Biot, FR
- Giacomo Garegnani
EPFL – Lausanne, CH
- Roman Garnett
Washington University – St. Louis, US
- Alexandra Gessner
Universität Tübingen, DE
- Philipp Hennig
Universität Tübingen, DE
- Toni Karvonen
University of Helsinki, FI
- Peter Nicholas Krämer
Universität Tübingen, DE
- Maren Mahsereci
Universität Tübingen, DE
- Katharina Ott
Bosch Center for AI – Renningen, DE
- Marvin Pförtner
Universität Tübingen, DE
- Jonathan Schmidt
Universität Tübingen, DE
- Tomas Teren
TU Dresden, DE
- Filip Tronarp
Universität Tübingen, DE
- Jonathan Wenger
Universität Tübingen, DE
- Stephen Wright
University of Wisconsin – Madison, US

Remote Participants

- Oksana Chkrebtii
Ohio State University – Columbus, US
- Jon Cockayne
University of Southampton, GB
- Yuhan Ding
Illinois Institute of Technology – Chicago, US
- Matthew Fisher
Newcastle University, GB
- Fred J. Hickernell
Illinois Institute of Technology – Chicago, US
- Nick Higham
Manchester University, GB
- Ilse C.F. Ipsen
North Carolina State University – Raleigh, US
- Motonobu Kanagawa
EURECOM – Biot, FR
- Hans Kersting
INRIA – Paris, FR
- Tadashi Matsumoto
University of Warwick – Coventry, GB
- Michael McKerns
Los Alamos National Laboratory – New Mexico, US
- Masha Naslidnyk
Amazon Research Cambridge, GB
- Chris Oates
Newcastle University, GB
- Michael A. Osborne
University of Oxford„ GB
- Houman Owhadi
California Institute of Technology – Pasadena, US
- Andrei Paleyes
University of Cambridge, GB
- Kamran Pentland
University of Warwick – Coventry, GB

- Geoff Pleiss
Columbia University – New York, US
- Jagadeeswaran Rathinavel
Illinois Institute of Technology – Chicago, US
- Timothy Reid
North Carolina State University – Raleigh, US
- Simo Särkkä
Aalto University – Helsinki, FI
- Florian Schäfer
Georgia Institute of Technology – Atlanta, US
- Aleksei Sorokin
Illinois Institute of Technology – Chicago, US
- Tim Sullivan
University of Warwick – Coventry, GB
- Aretha Teckentrup
University of Edinburgh, GB
- Onur Teymur
University of Kent – Canterbury, GB
- Zi Wang
Google – Cambridge, US