Mathematisches Forschungsinstitut Oberwolfach

Report No. 18/2024

DOI: 10.4171/OWR/2024/18

Mini-Workshop: Mathematics of Entropic AI in the Natural Sciences

Organized by
Susanne Gerber, Mainz
Illia Horenko, Kaiserslautern
Rupert Klein, Berlin
Terence O'Kane, Hobart

7 April – 12 April 2024

Abstract. The mathematical framework of approximate entropic learning introduced very recently promises to provide robust, cheap and efficient ways of machine learning in the so called "small data" regime, when the underlying learning task is highly-underdetermined, due to a large problem dimension and relatively small data statistics size. Such "small data" learning challenges are particularly common in the natural sciences (e.g., in geosciences, in climate research, in economics, and in biomedicine), imposing considerable difficulties for the numerics of common "data-hungry" Artificial Intelligence (AI) tools like Deep Learning (DL). The aim of this workshop will be to bring together experts in the emergent fields of entropic and DL mathematics/numerics, with some lead experts applying AI in the domain disciplines. The goal will be to detect and to discuss the commonalities in the challenges and in their mathematical solutions, as well as to discuss and fine-tune common mathematical problem formulations that are motivated by the AI applications in natural sciences. The establishment of a common mathematical framework for such small-data machine learning tasks would not only bolster future methodological developments but would also lay solid foundations to further in-depth rigorous analysis and theoretically founded interpretation of these methods and their results.

Mathematics Subject Classification (2020): 68T01, 68T99, 68Q32, 86A22, 92C50.

Introduction by the Organizers

The mini workshop *Mathematics of Entropic AI in the Natural Sciences*, jointly organised by Susanne Gerber (Mainz), Illia Horenko (Kaiserslautern), Rupert Klein (Berlin) and Terence O'Kane (Hobart) aimed at bringing together a group of experts in the field of entropic learning and leading specialists from different disciplines to identifiy and address common challenges. Throughout the miniworkshop, the participants had the chance to present their work, learn about (potentially unfamiliar) scientific fields, and also to engage in fruitful discussions about strengths and possible limitations of machine learning and AI methods in their practical use. The use of mathematics is an invaluable aid in proposing solutions to many of the scientific disciplines, which ranged from biomedicine to geosciences and climate research.

One topic that was considered relevant by all the participants was that of "small data", i.e., the condition that arises when the task of learning from the available data is highly underdetermined. This can occur when the data, from which learning is sought, is high-dimensional (i.e., the amount of measured features is particularly elevated) and/or is too sparse to allow for reliable statistical assessment. Both the aforementioned conditions tend to often occur in the natural sciences when applying traditional data-hungry Artificial Intelligence methods, such as Deep Learning.

Perspectives on the current state of mathematical foundations of entropic and DL AI numerics and mathematics were discussed, and promising future directions for mathematical research in this area have been identified, guided by the problems imposed by domain applications in the natural sciences. An important common issue that was identified for future research is the interpretability and explainability of the AI models – as well as a formulation of rigorous mathematical criteria that would allow controlling the sensitivity and computational scalability of AI tools in higher dimensions. Such criteria would also allow increasing the robustness of AI tools to the so-called adversarial attacks. Recently introduced mathematics-driven entropic AI methods were identified as a promising research direction, potentially providing efficient and interpretable way to learn from (noisy) and high-dimensional data.

Mini-Workshop: Mathematics of Entropic AI in the Natural Sciences

Table of Contents

Abstracts

AI and the brain

Davide Bassetti

This mini workshop was opened with this presentation, in which I presented a bird's eye view description of Artificial Neural Networks (ANNs) as representatives of Artificial Intelligence (AI), and compared them to the biological intelligence systems, i.e., the brain. It was highlighted how the recent expedited development of AI systems originates from their ability to scale efficiently. In particular, I put some observation forward:

a) Most of the models that achieve superhuman performance on a task are not able to generalize to multiple ones, effectively being a tool for a single type of task or for a selection of few related ones.
b) Increasing model size translates to elevated economic cost for training a model, which puts hard constraints in term of "affordability" to achieve a set level of performance. Thus, at present, only private entities with sufficient investment means are able to train large and complex models. Given the investment, the incentive to share openly with the scientific community the models (or sometimes even information about them) can quickly disappear. This point includes the source of training data, which is often not specified and poses questions relative to intellectual property.
c) The exponential rate of growth of the models in the last years is outpacing the growth of hardware capacity (e.g., GPU memory) [1]. Moreover, the carbon consumption of training such models should also be considered as a factor [2].
d) Providing tools for explainability of complex systems proves to be a challenging task, especially for large models, due to their design.

A single human brain is instead effectively able to perform a broad range of tasks (a), and does so with astounding energy (b) and carbon (c) efficiency. The comparison between the brain and ANNs, therefore, poses the question of whether there could be some additional factor in the brain that is not yet included in ANNs and could contribute to their performance. After introducing how an ANN is a collection of generalized linear models (GLMs) which perform a weighted sum of the inputs based on (learned) "connectivity" parameters and apply a nonlinearity to the result, then propagating the signal in a feed forward way, I described the fundamental unit of the brain's computation: the neuron. Particular focus was put on the modality of communication (Action Potentials) and the biophysical mechanisms that contribute to their generation (using the Hodgkin-Huxley model [3]). Another accent, besides the different level of complexity between a GLM and the Hodgkin-Huxley neuron, was on the heterogeneity of neurons in the brain, in terms of morphology, physiology and connectivity, which is structured and includes feedback and specialization between inhibition and excitation.

Next, I described how the connections between neurons are fundamentally far from being static weights, but are a non-linear and non-stationary system themselves, and can operate simultaneously at different time scales.

This leaves open questions in terms of:

- How far are current neural networks from having the level of complexity of the human brain
- Where is the line between the complexity that is necessary for the function and that which is a necessary part of the embodiment of intelligence (e.g., the necessity to maintain a metabolism which could be avoided when modeling the system). In other terms: how much should we worry about biological plausibility of neuronal networks?
- What could be the "missing steps" that could potentially bridge the gap between the systems?

My closing remarks contained a wish for future research to develop and use small efficient and scalable models, which are also interpretable. The advantage of this mathematical development would be models trainable by a significant portion of the scientific community, and could be shared along with the training data, and thus benefit research and promoting open scientific practice. Neuroscientific research, would greatly benefit from such tools, as they could help addressing many unanswered questions, whose answer can provide the basis for further development in AI tools, as it happened in the past.

References

[1] A. Gholami, Z. Yao, S. Kim, C. Hooper, M. Mahoney & K. Keutzer, *Ai and memory wall*, IEEE Micro (2024)
[2] https://aiindex.stanford.edu/report/
[3] https://en.wikipedia.org/wiki/Hodgkin-Huxley_model

Thoughts on Machine Learning

Rupert Klein

This is a repetition of the author's report contributed to
Workshop 2331 "Transport and Scale Interactions in Geophysical Flows".

Concerns: Techniques of machine learning (ML) and what is called "artificial intelligence" (AI) today find a rapidly increasing range of applications touching upon social, economic, and technological aspects of everyday life. They are also being used increasingly and with great enthusiasm to fill in gaps in our scientific knowledge by data-based modelling approaches. I have followed these developments over the past almost 20 years with interest and concern, and with mounting disappointment. This leaves me sufficiently worried to raise here a couple of pointed remarks.

Obviously, when these technologies are being employed to take over decisive functionality in safety-critical applications, we would like to exactly know how to guarantee their compliance with pre-defined guardrails and limitations. Moreover,

when ML techniques are utilized as building blocks in scientific research, it would violate scientific standards – in the authors opinion – if these building blocks were used without a thorough understanding of their functionality, including inaccuracies, uncertainties, and other pitfalls.

The most frequently used tools in ML and AI today are deep neural networks (DNNs) and, to the best of my knowledge, they currently constitute a particularly severe breach of what I postulate to be desirable for safety-critical applications and for their utilization in scientific research. In fact, I see the following related and further drawbacks:

a) Interpretability/Explainability: It is remarkable that the issue of how to reliably interpret the workings of DNN technology has become a topic of intense research only relatively recently, see [1], and can by far not be considered fully explored today.

b) Generalizability and out-of-sample performance: The quality of ML-learned functions is quite usually tested via some version of cross-validation [2]: Split the available data into a training and one or more testing sets, train the function on the former and test "generalizability" utilizing the latter. A key problem with the often rather high-dimensional spaces of function arguments is that no explicit definition of the domain of the learned function is provided. Therefore, when a new input argument is to be used in an application, there seem to be no systematic qualifiers that would indicate whether this argument may or may not be used as an argument of the learned function with any confidence.

c) Inefficiency in terms of data needed: DNNs are "big data" techniques, and it turns out they do, in fact, need rather large data sets for training – with consequences for the computational expense of their training, [3].

d) Inefficiency of function representation: DNNs based on ReLU (rectified linear unit) activation functions are popular in AI generically and as building blocks of more complex function constructions. ReLU-DNNs are known to represent piecewise linear functions on polygones in the space of function arguments, and He et al. [4] show that ReLU-DNNs require on the order of $D\kappa^D N$ free parameters to represent a piecewise linear function on a simplicial grid with N nodes in D dimensions, with $\kappa \geq 2$. Comparing this, for large D, with the number of degrees of freedom needed for the same task by a standard finite element ansatz, i.e., with $(D+1)N$, we find another reason for the extensive computational costs of DNN training.

e) Inefficiency of optimization algorithms: Thus far, (variants of) stochastic gradient descent methods for the solution of the DNN parameter estimation problem seem to be essentially the only reliable option [5]. Yet, these come with at most first order convergence, with further consequences for the computational expense of DNN training in comparison with methods the structure of which allows employing second order convergent Newton-type techniques in solving their parameter estimation problem.

f) Condition / Sensitivity: A mounting number of examples in the literature show that ML-learned functions can be tricked by so-called "adversarial attacks" to yield clearly false or low-quality results by effectively exploiting the often bad conditioning of the function learning problem, see [6, 7, 8, 9] and references therein.

Alternatives: We are not bound to utilizing neural network technology for machine learning and in the context of artificial intellence, however. Alternatives that overcome many of the abovementioned drawbacks and limitations are being developed. In this context, I have become aware of the recent family of "Sparse Probabilistic Approximations" (SPA) [3, 10, 11, 13] by Illia Horenko and co-workers. These methods turn out to be (i) at least as – and in many cases much more – powerful than DNNs in terms of the quality of functions learned; they (ii) come with natural indicators for the domain of the learned functions; targeting *small data problems* by design, they are (iii) generically much less data hungry than DNNs (thereby avoiding concern c), see above); their parameter estimation can (iv) be cast into optimization problems that are amenable to far-reaching partially analytical and iterative methods yielding second-order convergence; versions are available that (v) address a wide range of ML tasks, such as clustering, classification, regression, and more. On problems it has been used for, such as cancer classification, data-based El Niño prediction, or financial decision making [12], the recent eSPA+[1] technique for data classification in [13] has out-performed DNNs and other machine learning techniques decisively in terms of the quality of results or computational efficiency or both. As regards computational efficiency, the method comes with complexity $T \cdot D \cdot K$, where, T, is the number of available data samples, D is dimension of the function argument (or feature) space, and K is akin to the number of allowed input data clusters one would impose when applying some clustering technique for dimension reduction. In fact, this complexity is that of the K-means clustering algorithm, which is highly efficient but all by itself insufficient to solve machine learning problems beyond data clustering.

As is the case for the entire SPA family of methods, eSPA+ comes with a clean mathematical structure in which each ingredient has a transparent role and interpretation. This is seen in the following example: The parameter estimation problem of eSPA+ for the El Niño prediction problem reads as

$$(\boldsymbol{S}, \boldsymbol{\Gamma}, \boldsymbol{W}, \boldsymbol{\Lambda})^{\circ} = \underset{\boldsymbol{S},\boldsymbol{\Gamma},\boldsymbol{W},\boldsymbol{\Lambda}}{\operatorname{argmin}} \mathcal{L}^{+}_{\mathrm{eSPA}} \left(\boldsymbol{S}, \boldsymbol{\Gamma}, \boldsymbol{W}, \boldsymbol{\Lambda} \,\middle|\, \boldsymbol{X}, \boldsymbol{\Pi}^{\Delta t} \right) . \tag{1}$$

Here $\boldsymbol{X} = (X_t)_{t=1}^{T} \in I\!R^{D \times T}$ is the set of D-dimensional function arguments in the available data set. In the example, each $X_t \in I\!R^{D}$ consists of $D = 200$ degrees of freedom characterizing tropical pacific ocean sea surface and equatorial deep ocean water temperatures. $\boldsymbol{\Pi}^{\Delta t} = (\Pi_t^{\Delta t})_{t=1}^{T} \in \{0,1\}^{T}$ is the set of observations stating whether at time $t + \Delta t$ an El Niño did ($\Pi_t^{\Delta t} = 1$) or did not ($\Pi_t^{\Delta t} = 0$) occur; $\boldsymbol{S} \in I\!R^{D \times K}$ is a matrix whose columns $S_k \in I\!R^{D}$ are K centers of data clusters

[1] eSPA+ = entropy-optimal scalable probabilistic approximation, with algorithmic efficiency enhancements

or "boxes" in the space of function arguments akin to cluster centers in the K-means scheme, and $\mathbf{\Gamma} \in [0,1]^{K\times T}$ is a columnwise probabilistic matrix such that, when the problem has been solved, $(\boldsymbol{S}^{\circ}\mathbf{\Gamma}^{\circ})_t \approx X_t$ provides a reduced approximate representation of the argument-space input data $\boldsymbol{X}$; $\boldsymbol{W} \in [0,1]^D$ is a probability distribution over the dimensions of the argument (or feature) space with small W_d indicating weak influence of the dth data dimension on the prediction outcome, and $\mathbf{\Lambda} \in [0,1]^K$ is the set of probabilities Λ_k of El Niño occuring a time of Δt down the road if a data point belongs to the kth cluster.

The functional to be minimized then reads as

$$\mathcal{L}^{*}_{\mathrm{eSPA}}\left(\boldsymbol{S},\mathbf{\Gamma},\boldsymbol{W},\mathbf{\Lambda} \mid \boldsymbol{X},\mathbf{\Pi}^{\Delta t}\right) = \underbrace{\frac{1}{T}\sum_{d=1}^{D} W_d \sum_{t=1}^{T}\left(X_{d,t}-(S\Gamma)_{d,t}\right)^2}_{D\text{-red.: state approximation error}} \tag{2}$$

$$- \underbrace{\varepsilon_W \sum_{d=1}^{D} W_d \log\left(\frac{1}{W_d}\right)}_{D\text{-red.: feature discrimination}} - \underbrace{\frac{\varepsilon_\Lambda}{T}\sum_{t=1}^{T}\Pi_t^{\Delta t}\left(\ln\left(\mathbf{\Lambda}\right)\mathbf{\Gamma}\right)_t}_{\text{supervision}} .$$

All terms in this functional have a clear interpretation: The first term on the right is a W-weighted Euclidian norm measuring the quality of approximating $\boldsymbol{X}$ by $(\boldsymbol{S}\mathbf{\Gamma})$. The role of this term is to enable an effective dimension reduction in that the key information in the space of arguments is stored in the K reference points (or cluster centers) S°_k, with $K \ll T$ when the approach is successful. The dimension-wise weighting by W_d of the components $(X_{d,t}-(S\Gamma)_{d,t})^2$ of the Euklidian distance enables a further effective dimension reduction in that dimensions (or features) that only minimally affect the El Niño prediction receive a lesser weight in the solution and therefore contribute only marginally to the functional's value when the problem is solved. To achieve a least-biased discrimination of features in this way, the Shannon entropy of the distribution $\boldsymbol{W}$ is subtracted from the functional as a penalty in the second term on the right. That is, we seek to maximize Shannon-entropy and thus find the broadest possible distribution $\boldsymbol{W}$ under the given conditions.

The third term on the right, which implements the supervision of the classification learning problem, stems from interpreting the data $\Pi_t^{\Delta t} \in \{0,1\}$ as probabilities for the occurance of El Niño some time Δt in the future, and then

$$-\sum_{t=1}^{T}\Pi_t^{\Delta t}\left(\ln\left(\Lambda\right)\Gamma\right)_t \approx \sum_{t=1}^{T}\Pi_t^{\Delta t}\ln\left(\frac{1}{(\Lambda\Gamma)_t}\right) \tag{3}$$

is an approximation to that part of the Kullback-Leibler (KL) divergence between the data $\Pi_t^{\Delta t}$ and their approximations $(\mathbf{\Lambda\Gamma})_t$ which depends explicitly on the unknowns $\mathbf{\Lambda}$ and $\mathbf{\Gamma}$. It turns out that this latter approximation provably generates an upper bound for the functional utilizing the original KL-divergence, and in this sense the approximation is robust.

Now, once we have $(\boldsymbol{S}, \boldsymbol{\Gamma}, \boldsymbol{W}, \boldsymbol{\Lambda})^\circ$ determined by solving the above minimization problem, and today's state of ocean temperature data X^* is observed, then a probabilistic El Niño forecast is obtained as follows: Find the best-possible approximation of X^* by a convex combination $\boldsymbol{S}\Gamma^*$ of the reference points S_k, with $\Gamma^* \in [0,1]^K$, $\sum_{k=1}^K \Gamma_k^* = 1$. This yields the pertinent probabilistic weights Γ_k^* and the eSPA+-predicted probability for El Niño occurance a time Δt from today becomes $\Lambda\Gamma^* = \sum_{k=1}^K \Lambda_k \Gamma_k^*$.

Besides the clear interpretability of this method (concern a) taken care of), there is also an exceedingly efficient algorithm for its training: The idea detailed in [13] is to iteratively solve for one of the unknowns in $(\boldsymbol{S}, \boldsymbol{\Gamma}, \boldsymbol{W}, \boldsymbol{\Lambda})$ while keeping the other three fixed. Each of these steps allows for either an analytical or a numerical solution that scales linearly in the complexity parameters (D, T, K), and the entire iteration procedure can be cast as a Newton-type method [14]. This yields very fast (second order) convergence, so that concern e) does not arise for the SPA-family of methods.

Note also, that the convex hull of the reference points S_k° serves as a natural and robust estimate of the domain of the learned function. Thus, concern b) would at least in part be taken care of as well.

As regards concern d), it is shown in [3, 10] that the variant of eSPA+ described in (1)–(3) produces piecewise linear solutions on simplices with corners defined by the references points S_k. Therefore, these solutions are classical linear finite element functions, and the number of degrees of freedom needed to represent them is $(D+1)K$. Hence, eSPA+ does not suffer from concern d) either.

Addressing the remaining robustness concern f) for the SPA-family of methods is work in progress at the time of this writing.

References

[1] https://de.wikipedia.org/wiki/Explainable_Artificial_Intelligence

[2] https://en.wikipedia.org/wiki/Cross-validation_(statistics)

[3] I. Horenko, *On a Scalable Entropic Breaching of the Overfitting Barrier for Small Data Problems in Machine Learning*, Neural Comput., **32** (2020), 1563–1579

[4] J. He, L. Li, J. Xu, Ch. Zheng, *ReLU deep neural networks and linear finite elements*, J. Comput. Math., **38**:3 (2020), 502–527

[5] St.J. Wright, B. Recht, *Opitimzation for data analysis*, Cambridge University Press (2022), https://doi.org/10.1017/9781009004282

[6] https://openai.com/research/attacking-machine-learning-with-adversarial-examples

[7] Z. Klawikowska, A. Mikołajczyk, M. Grochowski, *Explainable AI for Inspecting Adversarial Attacks on Deep Neural Networks*, in: L. Rutkowski et al. (Eds.), ICAISC 2020, LNAI 12415, pp. 134–146, 2020.

[8] B. Dickson, https://bdtechtalks.com/2020/12/16/machine-learning-adversarial-attacks-against-machine-learning-time-bomb/ (original)
https://www.kdnuggets.com/2021/01/machine-learning-adversarial-attacks.html (re-post)

[9] N. Akhtar, A. Mian, N. Kardan, M. SHAH, *Advances in Adversarial Attacks and Defenses in Computer Vision: A Survey*, IEEE Access, **9**, 155161–155196

[10] S. Gerber, L. Pospisil, M. Navandar, I. Horenko, *Low-cost scalable discretization, prediction, and feature selection for complex systems*, Sci. Adv., **6**:5 (2020), eaaw0961

[11] I. Horenko, *Scalable Entropy-Optimal Machine Learning Classification for Small Data Problems*, PNAS, **119**:9 (2022), e2119659119

[12] E. Vecchi, *Efficient and scalable solution techniques for small data learning problems*, Dissertation, Faculty of Informatics, Università della Svizzera Italiana, Lugano (2023)
[13] E. Vecchi, L. Pospisil, S. Albrecht, T.J. O'Kane, I. Horenko, *Scalable Entropy-Optimal Machine Learning Classification for Small Data Problems*, Neural Comput., **34**:5 (2022), 1220–1255
[14] I. Horenko, E. Vecchi, J. Kardoš, Andreas Wächter, Olaf Schenk, Terence O'Kane, Patrick Gagliardini, Susanne Gerber, *On cheap entropy-sparsified regression learning*, PNAS, **120**:1 (2023), e2214972120

Numerically stable generative modelling using diffusion maps with application to subgridscale parametrization

Georg Gottwald

We consider the problem of sampling from an unknown distribution which is accessible only via a sufficiently large number of training samples. Such settings have recently drawn considerable interest in the context of generative modelling. Opposed to typical score-based generative models we do not attempt to estimate the score function and the generator of the underlying stochastic process but instead estimate a semi-group which admits the unknown probability density function as an asymptotic state. In particular, we propose a generative model combining diffusion maps and Langevin dynamics. Diffusion maps are used to approximate the drift term from the available training samples, which is then implemented in a discrete-time Langevin sampler to generate new samples. By setting the kernel bandwidth to match the time step size used in the unadjusted Langevin algorithm, our method effectively circumvents any stability issues typically associated with the time-stepping of stiff stochastic differential equations. Moreover, we introduce a novel split-step scheme, ensuring that the generated samples remain within the convex hull of the training samples. Our framework can be naturally extended to generate conditional samples. We demonstrate the performance of our proposed scheme through experiments on synthetic datasets with increasing dimensions and on a stochastic subgrid-scale parametrization conditional sampling problem. This is joint work with Fengyi Li, Youssef Marzouk and Sebastian Reich.

Tutorial on eSPA-Implementations

Davide Bassetti, Michael Groom

In this session, we provided a practical demonstration of the efficient implementation of various methods, between those discussed in other talks. Specifically, we provided the code and described the interface to generate synthetic datasets, load external ones and apply to them eSPA+, with different configurations. Moreover, we showed how to obtain insight on the output model through the use of plotting functions and via an implementation of the MAD algorithm.

eSPA for time series

Davide Bassetti

Analysis of (time) ordered data is a very common application for machine learning, for example we can consider classification of noisy time series, when assumptions of the persistence of the underlying states are available. In this talk, we present a novel methodology for performing supervised classification of (time) ordered data, which we call Entropic Sparse Probabilistic Approximation with H1 regularization (eSPAH1). It is an extension of entropic learning methodologies, that allows to learn H1-smooth segmentation patterns and simultaneously learning the optimal discretization and Bayesian classification rules.

Moreover, we also introduce cluster affiliation based on likelihood rather than Euclidean distance (as in "regular" entropic methods), which grants the ability to distinguish between states with overlapping mean but different distribution.

We demonstrate how this technique can be used for computationally-scalable identification of persistant (metastable) regime affiliations and regime switches from high-dimensional non-stationary and noisy time series, i.e., when the size of the data statistics is small compared to its dimensionality and when the noise variance is larger than the variance of the signal. Thus, we retain the advantages of entropic learning methodologies, expanding them to a new domain. We furthermore demonstrate the performance on a comprehensive set of toy learning examples, comparing it to state of the art techniques.

One particular application of this framework that was showcased was the direct learning of a model from training data, without iterative procedure. This can be performed if there exist a bijective map between the affiliation matrix Γ and the label probability matrix Π. As an example, this modality of learning was applied on a simplified RNA sequencing toy dataset.

Mathematics of adversarial AI attacks

Illia Horenko

The ubiquitous and boosting rise of AI technologies like AlphaGo, AlphaFold and GPT-4 heralded a new historic era. Multiple areas of human activity currently are affected by these rapid and - in many aspects, alarming and potentially dangerous - developments. Recent studies [1, 2, 3]indicate a high level of vulnerability of advanced AI platforms like AlphaGo through the several orders of magnitude more simple AI tools that can be trained to find and use the critical vulnerabilities to the so-called adversarial attacks.

In this talk, the problem of finding an optimal adversarial attack was formulated mathematically as an optimisation problem of finding the Minimal Adversarial Path (MAP) and Minimimal Adversarial Distance (MAD) [4]. Simply-verifiable mathematical conditions for existence, uniqueness and explicit analytical computability of MAP and MAD for (locally) uniquely-invertible classifiers, for generalized linear models (GLM), and for entropic AI (EAI) are formulated and proven. Practical computation of MAP and MAD, their comparison and interpretations for

various classes of AI tools (for neuronal networks, boosted random forests, GLM and EAI) was demonstrated on the common synthetic benchmarks from mathematics (including the number series challenges of Vladimir Arnold), climate research (prediction of the El Nino climate phenomenon), economics and biomedical sciences (finding minimal risk-mitigating policies for the health insurance holders).

References

[1] X. Yuan, P. He, Q. Zhu, X. Li, *Adversarial Examples: Attacks and Defenses for Deep Learning*, IEEE Transactions on Neural Networks and Learning Systems **30:9** (2019), 2805–2824.

[2] X. Yuan, P. He, Q. Zhu, X. Li, *Adversarial Policies Beat Superhuman Go AIs*, International Conference on Machine Learning **30:9** (2023).

[3] J. Chen, M.I. Jordan, *HopSkipJumpAttack: A Query-Efficient Decision-Based Attack*, 2020 IEEE Symposium on Security and Privacy (SP) **32** (2019), 1277–1294.

[4] I. Horenko, *On existence, uniqueness and scalability of adversarial robustness measures for AI classifiers*, ArXiv **32** (2023), abs/2310.14421.

Approximation rates for neural networks on Sobolev and Besov spaces

Jonathan W. Siegel

Machine learning techniques and especially neural networks [1] have recently been widely applied to a variety of problems in scientific computing (see for instance [2, 3, 4, 5]). Despite the empirical success of many of these methods, developing a theoretical understanding of their properties is still an active research problem. In particular, the questions of whether neural network-based methods can provably solve partial differential equations (PDEs), and if so, what advantages they have over traditional numerical methods, do not yet have satisfactory answers.

Some partial progress on this problem has been made in recent work demonstrating that shallow neural networks, i.e. neural networks with a single hidden layer, can be provably used to solve elliptic PDEs [6, 7]. However, the resulting methods are not yet competitive with traditional finite element solvers. Other recent work suggests that neural networks do not actually outperform traditional methods, such as finite element methods, for the solution of a variety of different PDEs [8]. We remark that we will mostly discuss the problem of solving PDEs using neural networks, which is typically low dimensional, instead of the more difficult problem of operator learning [5, 9], which is inherently high dimensional.

The perspective that we will take on these problems is to study the approximation theory of neural networks. Specifically, we will address the problem of how many parameters a neural network requires to achieve a specified accuracy on a given class of target functions, and will use this perspective to compare them with traditional numerical methods. The class of target functions we consider are functions from Sobolev and Besov spaces, and we consider approximation in the L_p-norm.

We recall that the class of Sobolev function $W^s(L_q(\Omega))$ for a domain $\Omega \subset \mathbb{R}^d$, $1 \leq q \leq \infty$, and $s > 0$ an integer, is defined by the norm

$$\|f\|_{W^s(L_q(\Omega))} = \|f\|_{L_q(\Omega)} + \|f^{(s)}\|_{L_q(\Omega)}, \tag{1}$$

where $f^{(s)}$ denotes the weak derivatives of f of order s (see [10], Chapter 5). The results we describe also apply to more general function spaces, such as fractional Sobolev spaces [11] and Besov spaces [12]. These function spaces play a fundamental role in analysis, approximation theory, and PDE theory [10, 13, 14]. For this reason, determining how efficiently neural networks can approximate functions from these spaces is an important theoretical problem for understanding the application of neural networks to scientific computing problems.

In order to approximate functions from the Sobolev class $W^s(L_q(\Omega))$ with error in the $L_p(\Omega)$ norm, with $1 \leq p, q \leq \infty$ and p and q not necessarily equal, it is necessary that the Sobolev embedding $W^s(L_q(\Omega)) \subset L_p(\Omega)$ hold. If Ω is a bounded domain in $\mathbb{R}^d$, this is guaranteed by the Sobolev embedding condition

$$\frac{1}{q} - \frac{1}{p} < \frac{s}{d}, \tag{2}$$

and in this case the embedding is guaranteed to be compact. At the embedding endpoint, i.e. when equality holds in (2), there may or may not be an embedding depending upon the values of s, p, q, d, but it will never be compact. We remark that similar embedding results hold also for Besov spaces [15].

Given that the Sobolev embedding condition (2) holds, we are interested in how efficiently the class $W^s(L_q(\Omega))$ can be approximated in $L_p(\Omega)$ by various methods of approximation. For linear methods of approximation, i.e. approximating via a linear map of rank n, the best error which can be achieved is given by the linear n-widths

$$\delta_n(W^s(L_q(\Omega)))_{L_p(\Omega)} \sim \begin{cases} n^{-\frac{s}{d}+\max\left(\frac{1}{2}-\frac{1}{p},\frac{1}{q}-\frac{1}{2}\right)} & 1 \leq q \leq 2 \leq p \leq \infty \\ n^{-\frac{s}{d}+\left(\frac{1}{q}-\frac{1}{p}\right)_+} & \text{otherwise.} \end{cases} \tag{3}$$

We refer to [16, 17] for the proof of this result and the theory of widths more generally. This implies that the rate of approximation by linear methods deteriorates when $q < p$. Using classical non-linear, i.e. adaptive, methods such as variable knot splines or non-linear wavelet expansions, one can recover an approximation rate of $n^{-s/d}$ even when $q < p$ (see [18]). For this reason the regime where $q < p$ is typically called the non-linear regime of approximation.

For approximation by deep neural networks with the ReLU activation function, we have been able to determine the optimal rates of approximation using n parameters [19], which are given by

$$E_n^{NN}(f)_{L_p(\Omega)} \leq C\|f\|_{W^s(L_q(\Omega))} n^{-\frac{2s}{d}} \tag{4}$$

whenever the embedding condition (2) is satisfied. This generalizes prior work [20, 21, 22] which only considered the case $q = \infty$, and in particular gives the optimal rate even in the non-linear regime where $q < p$. Remarkably, the rate obtained by deep neural networks significantly improves upon the rate that classical (even

non-linear) methods can achieve. However, this improved rate comes at the cost of parameters which cannot be encoded using a fixed number of bits. For this reason, these improved rates cannot really be obtained numerically. This can be made precise through the concept of metric entropy (see [17]), which shows that adaptive classical methods are optimal for classical smoothness spaces such as Sobolev and Besov spaces.

Interestingly, the corresponding problem for shallow ReLU^k neural networks is still open and appears to be much more difficult. Some partial results have been obtained [23, 24, 25], but determining the correct approximation rates for all s, p and q for which the compact embedding condition (2) is still an active research problem.

References

[1] Y. LeCun, Y. Bengio, and G, Hinton, *Deep learning*, Nature **521**, no. 7553 (2015), 436–444.

[2] M. Raissi, P. Perdikaris, and G. E. Karniadakis, *Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations*, Journal of Computational Physics **378** (2019), 686–707.

[3] E. Weinan and B. Yu, *The deep Ritz method: a deep learning-based numerical algorithm for solving variational problems*, Communications in Mathematics and Statistics **6.1** (2018), 1–12.

[4] G. Carleo, and M. Troyer, *Solving the quantum many-body problem with artificial neural networks*, Science **355.6325** (2017), 602–606.

[5] Z. Li, N. Borislavov Kovachki, K. Azizzadenesheli, K. Bhattacharya, A. Stuart, and A. Anandkumar, *Fourier Neural Operator for Parametric Partial Differential Equations*, International Conference on Learning Representations (2020).

[6] J. Xu, *Finite neuron method and convergence analysis*, Communications in Computational Physics **28**, no. 5 (2020), 1707–1745.

[7] J. W. Siegel, Q. Hong, X. Jin, W. Hao, and J. Xu, *Greedy training algorithms for neural networks and applications to PDEs*, Journal of Computational Physics **484** (2023), 112084.

[8] T. G. Grossmann, U. J. Komorowska, J. Latz, and C. Schönlieb. *Can physics-informed neural networks beat the finite element method?*, arXiv preprint arXiv:2302.04107 (2023).

[9] 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**, no. 3 (2021), 218–229.

[10] L. C. Evans, *Partial differential equations*, Vol. 19. American Mathematical Society, 2022.

[11] E. Di Nezza, G. Palatucci, and E. Valdinoci, *Hitchhiker's guide to the fractional Sobolev spaces*, Bulletin des sciences mathématiques **136**, no. 5 (2012), 521–573.

[12] R. A. DeVore, and R. C. Sharpley, *Besov spaces on domains in* $\mathbb{R}^d$, Transactions of the American Mathematical Society **335**, no. 2 (1993), 843–864.

[13] R. A. DeVore, and G. G. Lorentz, Constructive approximation. Vol. 303. Springer Science & Business Media, 1993.

[14] R. A. DeVore, and V. A. Popov, *Interpolation of Besov spaces*, Transactions of the American Mathematical Society **305**, no. 1 (1988), 397–414.

[15] O. V. Besov, V. P. Il'in, S. M. Nikol'skiĭ, Integral Representations of Functions and Imbedding Theorems, V. H. Winston & Sons, 1978.

[16] A. Pinkus, N-widths in Approximation Theory, Vol. 7. Springer Science & Business Media, 2012.

[17] G. G. Lorentz, M. von Golitschek, and Y. Makovoz, Constructive approximation: advanced problems, Vol. 304. Berlin: Springer, 1996.

[18] R. A. DeVore, *Nonlinear approximation*, Acta numerica **7** (1998): 51-150.
[19] J. W. Siegel, *Optimal approximation rates for deep ReLU neural networks on Sobolev and Besov spaces*, Journal of Machine Learning Research **24**, no. 357 (2023), 1–52.
[20] D. Yarotsky, *Optimal approximation of continuous functions by very deep ReLU networks*, In Conference on learning theory, pp. 639-649. PMLR, 2018.
[21] Z. Shen, H. Yang, and S. Zhang, *Optimal approximation rate of ReLU networks in terms of width and depth*, Journal de Mathématiques Pures et Appliquées **157** (2022), 101–135.
[22] J. Lu, Z. Shen, H. Yang, and S. Zhang, *Deep network approximation for smooth functions*, SIAM Journal on Mathematical Analysis **53**, no. 5 (2021), 5465–5506.
[23] P. P. Petrushev, *Approximation by ridge functions and neural networks*, SIAM Journal on Mathematical Analysis **30**, no. 1 (1998), 155–189.
[24] F.Bach, *Breaking the curse of dimensionality with convex neural networks* Journal of Machine Learning Research **18**, no. 19 (2017), 1–53.
[25] T. Mao, and D. X. Zhou, *Rates of approximation by ReLU shallow neural networks* Journal of Complexity **79** (2023): 101784.

rSPA and numerical W-solvers for eSPA

Lukas Pospisil

This contribution summarizes the solution aspects of so-called Entropy-optimal Scalable Probabilistic Approximation (eSPA, [1]) problem given by

$$
\begin{aligned}
[S^*, \Gamma^*, \Lambda^*, w^*] \quad = \quad & \arg\min \sum_{k=1}^{K} \sum_{t=1}^{T} \sum_{d=1}^{D} w_d (X_{d,t} - S_{d,t})^2 \\
& -\varepsilon_{KL} \sum_{t=1}^{T} \sum_{m=1}^{M} \sum_{k=1}^{K} \Pi_{m,t} \Gamma_{k,t} \log \Lambda_{m,k} \\
& +\varepsilon_w \sum_{d=1}^{D} w_d \log w_d,
\end{aligned} \tag{1}
$$

where $X \in \mathbb{R}^{D,T}$ are given data, $\Pi \in \mathbb{R}^{M,T}$ is given left-stochastic matrix of data labeling probabilities, $\Gamma \in \mathbb{R}^{K,T}$ is left-stochastic matrix with probability affiliations, $S \in \mathbb{R}^{D,K}$ are probabilistic centroids, $w \in \mathbb{R}^D$ is stochastic vector of feature importance, and $\Lambda \in \mathbb{R}^{M,K}$ is a left stochastic matrix of conditional probabilities from Bayesian model $\Pi = \Lambda\Gamma$.

The problem is formulated as an extension of original SPA algorithm [2] by Entropic regularization for dealing with feature sparsification.

Although the problem (1) is non-convex and hard to be solved directly, it is easy to prove (even on the blackboard) that if one solves the problem using Subspace algorithm (when the problem is solved in one variable and all others are fixed), the objective function is monotonically decreasing and all subproblems enjoy the analytic solution. To be more specific

- Γ-problem can be solved using the analytic formula from K-means clustering algorithm,
- S-problem is unconstrained quadratic programming problem and the solution can be obtained from the 1-order necessary optimality conditions,

- Λ-problem has similar solution to well-known transitional probabilities problem in discrete Markov chains,
- w-problem leads to softmax function.

Using the idea of affiliation regularization [4] by H1-seminorm (to enforce the persistency of classification in time or space), one can extend the formulation by the regularization of affiliation with respect to time or space [5]. In this case, problem in variable Γ does not have analytic solution, but can be solved efficiently using the Spectral Projected Gradient method for Quadratic programming [6].

References

[1] E. Vecchi, L. Pospíšil, S. Albrecht, T.J. O'Kane, I. Horenko, *Scalable Entropy-Optimal Machine Learning Classification for Small Data Problems*, Neural Comput., **34**:5 (2022), 1220–1255

[2] S. Gerber, L. Pospíšil, M. Navandar, I. Horenko, *Low-cost scalable discretization, prediction, and feature selection for complex systems*, Sci. Adv., **6**:5 (2020), eaaw0961

[3] I. Horenko, *Scalable Entropy-Optimal Machine Learning Classification for Small Data Problems*, PNAS, **119**:9 (2022), e2119659119

[4] S. Gerber and I. Horenko, *Improving clustering by imposing network information*, Science Advances, **1**, (2015)

[5] I. Horenko, L. Pospíšil, E. Vecchi, A. Steffen, A. Gerber, B. Rehbock, A. Stroh, S. Gerber, *Low-Cost Probabilistic 3D Denoising with Applications for Ultra-Low-Radiation Computed Tomography*, Journal of Imaging, **6**:8 (2022)

[6] L. Pospíšil, P. Gagliardini, W. Sawyer, I. Horenko, *On a scalable nonparametric denoising of time series signals*, Communications in Applied Mathematics and Computational Science **13** (2018),107–138

On the comparative utility of entropic learning versus deep learning for long-range ENSO prediction

Michael Groom

This talk compares the ability of deep learning and entropic learning methods to predict the probability of the Niño3.4 index being above 0.4° (El Niño), below −0.4° (La Niña) or within both of these thresholds (neutral) at lead times of 3 up to 24 months. In particular, the performance, robustness, interpretability and training cost of entropic learning methods, represented by the entropy-optimal Scalable Probabilistic Approximation (eSPA) algorithm, are compared with deep learning methods, represented by a Long Short-Term Memory (LSTM) classifier, trained on the same dataset. Using only data derived from sea surface temperature observations over the period 1958-2018 and a correspondingly surface-forced resimulated ocean model, the problem becomes a canonical small-data challenge, where the number of examples for learning is of similar size to the number of features. Relative to the LSTM model, eSPA exhibits substantially better out-of-sample performance in terms of area under the ROC curve (AUC) for all lead times at $\sim 0.02\%$ of the computational cost. Comparisons of AUC with other state-of-the-art deep learning models presented in the literature show that eSPA appears to also be more accurate than these models across all three classes. Examining the

feature importance for eSPA shows that it induces a strong sparsification of the feature space, while looking at the statistical significance of each of the obtained clusters shows that, in general, incorrect predictions on the test set are made from the least significant clusters. Composite images are generated for each of the cluster centroids from each trained eSPA model at each lead time. At shorter lead times the composite images for the most significant clusters correspond to patterns representing mature or emerging/declining El Niño or La Niña states, while at longer lead times they correspond to precursor states consisting of extra-tropical anomalies. Furthermore, an increased diversity is observed in the precursor patterns with the highest predictive skill as lead time increases. Finally, modifications to the baseline dataset are explored, showing that improvements can be made in the parsimony of the trained eSPA model without sacrificing predictive power.

An iterative method for the solution of Laplace-like equations in high and very high space dimensions

Harry Yserentant

The talk dealt with the equation $-\Delta u + \mu u = f$ on high-dimensional spaces $\mathbb{R}^m$, where the right-hand side $f(x) = F(Tx)$ is composed of a separable function F with an integrable Fourier transform on a space of a dimension $n > m$ and a linear mapping given by a matrix T of full rank and $\mu \geq 0$ is a constant. For example, the right-hand side can explicitly depend on some or all differences $x_i - x_j$ of the components of x. We have shown that the solution of this equation can be expanded into sums of functions of the same structure and have developed in this framework an equally simple and fast iterative method for its computation. The method is based on the observation that in almost all cases and for large problem classes the expression $\|T^t y\|^2$ deviates on the unit sphere $\|y\| = 1$ the less from its mean value the higher the dimension m is, a concentration of measure effect. The higher the dimension m, the faster the iteration converges.

References

[1] H. Yserentant, *An iterative method for the solution of Laplace-like equations in high and very high space dimensions*, Numer. Math **156** (2024), 777–811; also available as arXiv:2403.00682 [math.NA]

Understanding and computing with physical patterns

Karin Everschor-Sitte

To overcome the limitations of our modern computing technology, novel computational paradigms in combination with suitable hardware solutions are required. Physical reservoir computing [1, 2] exploits natural dynamical systems to efficiently transform complex tasks into simple ones. Reservoir computing is a universal computational paradigm that, in contrast to neural networks, requires fewer computational resources while considerably reducing computing time by several orders

of magnitude [3, 4]. In physical reservoir computing, the role of the reservoir is assumed by a physical material, see Figure 1.

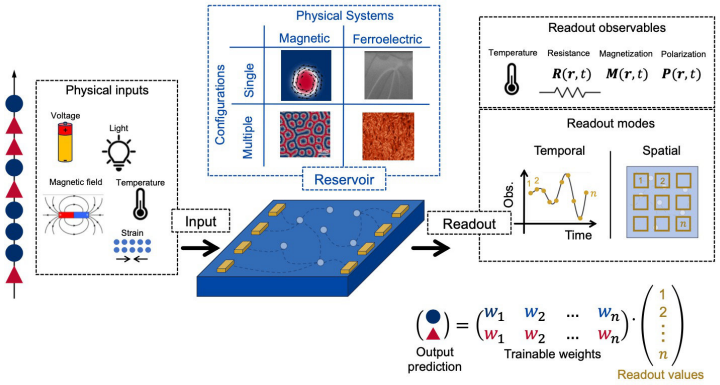

FIGURE 1. Sketch of the functional principle of physical reservoir computing. The physical reservoir (center) is stimulated by suitable physical input (left). The complex, non-linear response of the reservoir is then measured (right). The measurement can be done in different modes as indicated in the figure. The measured values serve as input data for training a linear classifier (indicated by the equation below). Fig. from Ref [5].

We showed that patterns in magnetic materials are a competitive ultra-low power and high-performing physical realization of the reservoir [6, 7, 8]. These magnetic patterns contain so-called skyrmions. Skyrmions are topologically stable whirls that are realized in different areas of physics and were initially discovered by Tony Skyrme in particle physics in the 1960's, see for example [9, 10]. Skyrmions, which occur in magnetic systems, were first observed experimentally in 2009 [11]. Since then, the field of magnetic skyrmions has developed into a very active area of research (see for example Ref. [12]), with the aim of exploiting the topological properties of the magnetic whirl-like particles for spintronics applications. For example, the peculiar twist of the magnetization in skyrmions leads to a very efficient coupling to electric currents and allows for "banana kicks" analogous to those in soccer (see Ref. [13] and references therein). More recently, magnetic skyrmions have been of strong interest for unconventional computing schemes such as stochastic computing, quantum computing and reservoir computing [14, 18, 15]. Using such magnetic patterns comprising skyrmions we have performed the benchmark test of classifying spoken digits [17]. Here, a model employing a nonlinear filtering technique has been developed, achieving a 97.4% accuracy in recognizing individual digits regardless of the speaker, and an even higher accuracy of 98.5% among female speakers. This represents the highest accuracy recorded for in-materio

reservoir computing. Our results are on par with state-of-the-art non-in-materio reservoir systems.

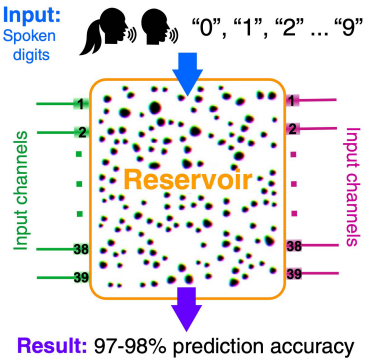

FIGURE 2. Sketch of spoken digit classification using a multi-channel skyrmion fabric reservoir computer. The dark points represent skyrmion cores and the cyclic colormap around them represents the orientation of in-plane magnetization. Graphical abstract from Ref. [16].

To generally characterize how good a physical reservoir is, we developed efficient task-agnostic metrics benchmarking spatially resolved the reservoir's key features – non-linearity, complexity, and fading memory [17]. Experimental realizations of skyrmion reservoirs further highlight the potential for energy-efficient high-performance skyrmion-based RC, see Ref [18] for an overview. Suitably combining reservoirs allows for improving the reservoir's properties and solving more complex tasks [19]. While on the one hand, we can employ materials to compute efficiently, on the other hand, we can use advanced inference methods to understand and improve materials. To improve materials for applications in general, knowledge of their inhomogeneities and defects is beneficial. We developed two computationally scalable data measures (the latent dimension and the latent entropy) taking into account latent temporal relations between processes of interest that provide deeper insight into experimentally observed noisy video data [20].

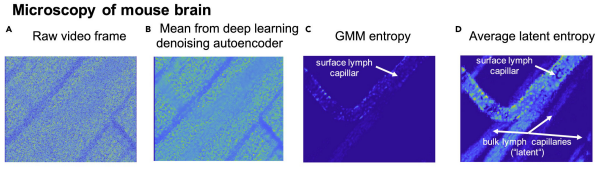

FIGURE 3. Analysis of the microscopy video of lymph flow in a mouse brain from Ref. [21].

We demonstrate their effectiveness by using the latent entropy and latent dimension on various examples in the natural sciences, uncovering previously unseen phenomena like a gradient in magnetic measurements and a hidden network of glymphatic channels in microscopy data from the mouse brain [21]. What makes these techniques unique is their independence from the typical restricting assumptions such as identical and independent distribution (i.i.d.), as well as, Gaussianity of the data found in many machine learning models.

References

[1] W. Maass, T. Natschlger, & H. Markram, *Real-time computing without stable states: A new framework for neural computation based on perturbations*, Neural Computation **14** (2002), 2531–2560

[2] G. Tanaka, T. Yamane, J. Héroux, R. Nakane, N. Kanazawa, S. Takeda, H. Numata, D. Nakano & A. Hirose, *Recent advances in physical reservoir computing: A review*, Neural Networks **115** (2019), 100–123

[3] H. Jaeger, M. Lukoševičius, D. Popovici, & U. Siewert, *Optimization and applications of echo state networks with leaky-integrator neurons*, Neural Networks **20** (2007), 335-352

[4] L. Grigoryeva, & J.Ortega, *Echo state networks are universal*, Neural Networks. **108** (2018), 495–508

[5] K. Everschor-Sitte, A. Majumdar, K. Wolk & D. Meier, *Novel implementations for reservoir computing-from spin to charge*, ArXiv Preprint ArXiv:2311.11929. (2023)

[6] D. Prychynenko, M. Sitte, K. Litzius, B. Krüger, G. Bourianoff, M. Kläui, J. Sinova & K. Everschor-Sitte, *Magnetic skyrmion as a nonlinear resistive element: a potential building block for reservoir computing*, Physical Review Applied **9** (2018), 014034

[7] G. Bourianoff, D. Pinna, M. Sitte & K. Everschor-Sitte, *Potential implementation of reservoir computing models based on magnetic skyrmions*, Aip Advances **8** (2018)

[8] D. Pinna, G. Bourianoff & K. Everschor-Sitte, *Reservoir computing with random skyrmion textures*, Physical Review Applied **14** (2020), 054020

[9] T. Skyrme, *A non-linear field theory*, Proceedings Of The Royal Society Of London. Series A. Mathematical And Physical Sciences **260** (1961), 127-138

[10] I. Aitchison *Tony Skyrme and the Origins of Skyrmions*, *ArXiv Preprint ArXiv:2001.09944.* (2019)

[11] S. Mühlbauer, B. Binz, F. Jonietz, C. Pfleiderer, A. Rosch, A. Neubauer, R. Georgii & P. Böni, *Skyrmion lattice in a chiral magnet*, Science **323** (2009), 915-919

[12] K. Everschor-Sitte, J. Masell, R. Reeve & M. Kläui, *Perspective: Magnetic skyrmions–Overview of recent progress in an active research field*, Journal Of Applied Physics **124** (2018)

[13] K. Everschor-Sitte, & M. Sitte, *Real-space Berry phases: Skyrmion soccer*, Journal Of Applied Physics **115** (2014)

[14] J. Grollier, D. Querlioz, K. Camsari, K. Everschor-Sitte, S. Fukami & M. Stiles, *Neuromorphic spintronics*, Nature Electronics **3** (2020), 360-370

[15] G. Finocchio, J. Incorvia, J. Friedman, Q. Yang, A. Giordano, J. Grollier, H. Yang, F. Ciubotaru, A. Chumak, A. Naeemi & Others *Roadmap for unconventional computing with nanotechnology*, Nano Futures (2023)

[16] R. Msiska, J. Love, J. Mulkers, J. Leliaert & K. Everschor-Sitte, *Audio classification with skyrmion reservoirs*, Advanced Intelligent Systems **5** (2023), 2200388

[17] J. Love, R. Msiska, J. Mulkers, G. Bourianoff, J. Leliaert & K. Everschor-Sitte, *Spatial analysis of physical reservoir computers*, Physical Review Applied **20** (2023), 044057

[18] O. Lee, R. Msiska, M. Brems, M. Kläui, H. Kurebayashi & K. Everschor-Sitte, *Perspective on unconventional computing using magnetic skyrmions*, Applied Physics Letters **122** (2023)

[19] K. Stenning, J. Gartside, L. Manneschi, C. Cheung, T. Chen, A. Vanstone, J. Love, H. Holder, F. Caravelli, K. Everschor-Sitte, & *Others Neuromorphic few-shot learning: generalization in multilayer physical neural networks*, ArXiv Preprint ArXiv:2211.06373. (2022)

[20] I. Horenko, D. Rodrigues, T. O'Kane & K. Everschor-Sitte, *Scalable computational measures for entropic detection of latent relations and their applications to magnetic imaging*, Communications In Applied Mathematics And Computational Science. **16** (2021), 267-297

[21] D. Rodrigues, K. Everschor-Sitte, S. Gerber & I. Horenko, *A deeper look into natural sciences with physics-based and data-driven measures*, Iscience **24** (2021)

Bayesian structure learning for climate

Terence O'Kane

Climate general circulation models are the main tools used for determining climate change risks associated with anthropogenic forcing. A severe limitation on our ability to determine such risks arise due to significant biases in representing the temporal behavior and seasonal phase locking of key climate teleconnections such as the El Nino Southern Oscillation (ENSO). Climate model biases and performance is typically assessed against observational products via systematic comparison of individual metrics, usually focused on the mean climate, over the recent historical period. These relationships are typically studied by regressing the postulated response on lagged values of the driver and vice versa. If substantial autocorrelations exist in the data, this approach can lead to the detection of spurious relationships. Where autoregressive models are employed, large uncertainties can arise over model selection. We demonstrate how Bayesian structure learning can enable a systematic probabilistic framework for process-based model evaluation of both the temporal behaviour of individual climate modes but also to identify and assess the causal teleconnections between those modes. Specifically, time homogeneous Dynamic Bayesian Network (DBN) models are constructed from observed (reanalyzed) data and climate model simulations from the leading Meteorological Centres as time series of empirical climate teleconnection indices. Reversible jump Markov Chain Monte Carlo (RJMCMC) is used to provide uncertainty quantification for selecting the respective network structures. The incorporation of confidence measures in structural features provided by the Bayesian approach is key to yielding informative measures of the differences between products if network-based approaches are to be used for model evaluation, particularly as point estimates alone may understate the relevant uncertainties. We show that network structures can be fitted simultaneously and feasibly across a representative sample of climate model simulations affording uncertainty estimation of the robustness of differences across models and observations and to robustly identify model biases between teleconnections in the climate simulated over the past six decades. Differences in the high confidence posterior probabilities assigned to edges of the respective directed acyclic graphs (DAG) provides a quantitative summary of departures in the CMIP5 models from reanalyses as quantified using Wasserstein distance and Kullback-Leibler divergences. In general terms the climate model simulations are

in better agreement with reanalyses where tropical processes dominate, and autocorrelation time scales are long. Seasonal effects are shown to be important when examining tropical-extratropical interactions with the greatest discrepancies and largest uncertainties present for the Southern Hemisphere teleconnections.

References

[1] T. J. O'Kane, D. Harries, M. A. Collier, *Bayesian structure learning for climate model evaluation*, ESS Open Archive, (2024), https://doi.org/10.22541/essor.169603507.70356103/v1

[2] D. Harries, T. J. O'Kane *Dynamic Bayesian networks for evaluation of Granger causal relationships in climate reanalyses*, Journal of Advances in Modeling Earth Systems **13** (2021), e2020MS002442. https://doi.org/10.1029/2020MS002442

Unresolved scales in weather and climate models: learning uncertainty from data

Nikki Vercauteren

Limited computer resources lead to a simplified representation of unresolved small-scale processes in weather and climate models, through parameterisation schemes. These approximate models of the unresolved scales express unresolved processes using the resolved variables of the prediction model as input and are often deterministic, but should sometimes be stochastic.

A systematic data-driven approach can help quantifying the uncertainty of parameterisations and inform us on how and when to incorporate uncertainty in the modelling, through stochastic parameterisation schemes. To enable such a systematic data-driven approach, methods from entropy-based learning and uncertainty quantification were combined in a model-based clustering framework, where the model is a stochastic differential equation with piecewise constant parameters. As a result, stochastic parameterisation can be learned from observations. The method is able to retrieve a hidden functional relationship between the parameters of a stochastic model and the resolved variables. As a result, a coarse-grained model is obtained, where the unresolved scales are expressed as stochastic differential equations whose parameters are continuous functions of the resolved variables.

Among the parameterised processes in weather and climate models, turbulent fluxes exert a critical impact on the exchange of heat, water and carbon between the land and the atmosphere. Turbulence theory was, however, developed for homogeneous and flat terrain, with stationary conditions. The theory fails in unsteady flow contexts or with heterogeneous landscapes, leading to important uncertainty in the parameterization of turbulence. Using field measurements of turbulence, the stochastic modelling framework is able to uncover a stochastic parameterisation that represent unsteady mixing in difficult conditions. Such methodology will be explored for further derivation of stochastic parameterisations, and should help to quantify uncertainties in climate projections related to uncertainties of the unresolved scales dynamics.

Machine learning in epigenetics

Nicolò Alagna

Epigenetics is the science that study stable and effective changes in the cell functionality that are not linked to the native DNA sequence, which links the gap between genetics and environmental influences on gene expression, offering insights into how the same genetic code can produce different phenotypes under various environmental contexts. One example is Chromatin remodelling and protein expression control, which control, part of the stem cells differentiation. Another aspect of epigenetics is to explore the chemical modifications of DNA and histone proteins which influence gene activity. Some of these modifications include methylation, acetylation, phosphorylation, and ubiquitylation, which can be inherited through cell division and potentially across generations, thereby affecting how genes are expressed. This field has profound implications for understanding the biological mechanisms underlying development, aging, and disease. Recent advancements highlight its role in the regulation of gene expression in response to environmental changes, the maintenance of cellular identity, and the reprogramming of cells. Furthermore, epigenetic mechanisms are implicated in numerous diseases, including cancer, neurological disorders, and autoimmune diseases, by governing genes that control cell cycle, apoptosis, and metabolism.

In our work, we focus on the analysis and identification of these modifications at both the DNA and RNA levels. In details, in our research we use deep neural network technology to track and identity possible DNA/RNA modifications with single base accuracy. Starting with epigenetic in DNA, DNA/RNA-hybrid systems were investigated, which are responsible for aging effects on the telomer. During cell splicing, Single based of RNA are included in the DNA sequence by coding errors due to the similarity of the DNA and RNA bases. This affects specific enzymes during the telomer formation, reducing their length and affecting life span of the system. It is possible to track small changes using deep learning? To answer this question, a neural network was trained to identify and distinguish RNA and DNA bases based on their experimental signal from Oxford Nanopore Technologies (ONT), which produce an electric signal depending on chemical structure of the base that it is reading. The deep learning results show the possibility to track the RNA bases in DNA strands with the accuracy of the single base. Moreover, it was able to establish what kind of RNA base was wrongly replaced in the DNA strand, classifying the four possible cases over the DNA strand. Similarly to the DNA-RNA hybrid, Neural networks were used to analyse RNA modifications in RNA strands, which present a more substantial chemical differences compared to the unmodified base. However, RNA modifications are much more in number (about 200) compared to the possible combination of DNA-RNA hybrid, making the problem to scale in terms of complexity. For the RNA modification analysis, we focused on four modifications (pseudouridine, M6A methylation, Inosine and Gm methylation) and see if the neural network can track modifications position and understands significant signal patters that are unique for modifications classifications. Also in this case, the neural network was able to distinguish and

classify the four different modifications, if they are present in the strand or not, and pointing the correct position in the RNA sequence reference.

These results underline the potential of deep learning in support our understanding of epigenetic mechanisms at the molecular level, especially understanding the crucial role and interplay between modifications and diseases. although this is a small step in the field of epigenetics, the ability of neural networks to localize and identify these modifications opens the question: can we leverage machine learning and deep learning to predict disease progression or response to therapy based on epigenetic signatures? What is the next step in the field of bioinformatics to explore and understand epigenetics in complex life systems?

Reporter: Davide Bassetti

Participants

Dr. Nicolò Alagna
Institute of Human Genetics, University Medical Center Mainz
Anselm-Franz-von-Bentzel-Weg 3
55128 Mainz
GERMANY

Dr. Davide Bassetti
Fachbereich Mathematik
T.U. Kaiserslautern
Postfach 3049
67618 Kaiserslautern
GERMANY

Cosima Caliendo
Institut für Mathematik
Johannes-Gutenberg Universität Mainz
Staudingerweg 9
55128 Mainz
GERMANY

Prof. Dr. Karin Everschor-Sitte
Fakultät für Physik
Universität Duisburg-Essen
Lotharstraße 1
47057 Duisburg
GERMANY

Prof. Dr. Susanne Gerber
Institute of Human Genetics, University Medical Center of the Johannes Gutenberg Universität Mainz
Anselm-Franz-von-Bentzel-Weg 3
55128 Mainz
GERMANY

Prof. Dr. Georg A. Gottwald
School of Mathematics and Statistics
The University of Sydney
Sydney 2206
AUSTRALIA

Dr. Michael Groom
CSIRO Environment
GPO Box 1538
Hobart, Tasmania 7001
AUSTRALIA

Dr. Charlotte Hewel
Institut für Humangenetik
Johannes-Gutenberg Universität Mainz
Anselm-Franz-von-Bentzel-Weg 3
55128 Mainz
GERMANY

Prof. Dr. Illia Horenko
RPTU Kaiserslautern-Landau
Fachbereich Mathematik
Postfach 3049
67653 Kaiserslautern
GERMANY

Prof. Dr. Rupert Klein
Fachbereich Mathematik und Informatik
Freie Universität Berlin
Arnimallee 6
14195 Berlin
GERMANY

Prof. Dr. Terence J. O'Kane
CSIRO Environment
GPO Box 1538
Hobart, Tasmania 7001
AUSTRALIA

Dr. Lukas Pospisil
Department of Mathematics
Faculty of Civil Engineering
VSB – Technical University of Ostrava
17. Listopadu 15/2172
708 33 Ostrava
CZECH REPUBLIC

Tim Prokosch
Fachbereich Mathematik
T.U. Kaiserslautern
Postfach 3049
67618 Kaiserslautern
GERMANY

Prof. Dr. Jonathan W. Siegel
Department of Mathematics
Texas A & M University
College Station, TX 77843-3368
UNITED STATES

Prof. Dr. Nikki Vercauteren
Institute of Geophysics and Meteorology
University of Cologne
50969 Köln
GERMANY

Prof. Dr. Harry Yserentant
Institut für Mathematik
Technische Universität Berlin
Straße des 17. Juni 136
10623 Berlin
GERMANY