Exploring the Universe with Artificial Intelligence (UniversAI)
Proceedings IAU Symposium No. 397, 2026
I. Liodakis & V. Efthymiou, eds.
doi:10.1017/S1743921325105048

Information Field Theory – Concepts, Applications, and AI-Perspective

Torsten Enßlin

Max Planck Institute for Astrophysics, Karl-Schwarzschild-Str. 1, 85748 Garching, Germany
Deutsches Zentrum für Astrophysik, Postplatz 1, 02826 Görlitz, Germany
email: ensslin@mpa-garching.mpg.de

Abstract. Information field theory (IFT) is the application of probabilistic reasoning to fields. Physical fields are mathematical functions over continuous spaces that exhibit certain properties of regularity, such as limited variance and finite gradients. Inferring a field from an observational dataset should exploit these regularities. However, the finite number of constraints that the data provides is insufficient to determine the infinite number of degrees of freedom of a field. IFT enables us to derive optimal field inference algorithms that explicitly exploit domain knowledge. These algorithms can be implemented via Numerical Information Field Theory (NIFTy). In NIFTy, neural operator forward models can be written and inverted probabilistically. NIFTy thereby infers fields and their remaining uncertainties. This is achieved using novel variational inference schemes that scale quasi-linearly, even for ultra-high dimensional problems. This paper introduces the basic concepts of IFT and NIFTy, highlights a few of their astrophysical applications, and discusses their artificial intelligence (AI) perspective. Finally, UBIK (the Universal Bayesian Imaging Kit), an emerging customisation of NIFTy for a suite of astrophysical telescopes, is presented as a central tool to the topic of the UniversAI conference.

Keywords. information field theory, artificial intelligence, signal reconstruction, inference

1. Introduction

In astrophysics, one wants to deduce from observational data the state of the physical fields that permeate the Universe, like the matter density, the radiation field, the gravitation field, and the like. From an artificial intelligence (AI) perspective, this is an ill posed inverse problem, as fields have more unknown parameters – actually an infinite number, as at any location of continuous space fields can and have different values – than data constraints, as any dataset is finite in size. The contemporary standard AI approach would be to use a training data set, pairs of discretized field realizations and corresponding data, in order to train and validate a neural network to become a field reconstruction machine. This approach has a number of problems, none severe enough to stop practitioners to go that route, but all sufficient worrisome to to be mentioned:

• The ground truth in astrophysics is usually not accessible. All we have is observational data, data products derived from these, and our theoretical understanding of the cosmos. The observational images we have from the universe are not a ground truth. Training networks on those images will imprint the artifacts caused by imperfections of the measurement and imaging processes.

• Synthetic skies can be obtained from numerical simulations, and mock data can be generated with those. These, however, depend heavily on the assumptions and approximations made and selection of included effects in the simulations. The risk is that a

© The Author(s), 2026. Published by Cambridge University Press on behalf of The International Astronomical Union. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

https://doi.org/10.1017/S1743921325105048 Published online by Cambridge University Press

neural network trained on those generalizes properties of the simulations that do not conform to reality.

• Many of the rules a network learns with large effort during training are trivial for humans. For example that the sky emission is a strictly positive quantity provides one bit of information per image pixel for free, as the negative values are excluded for each pixel. Letting a network to find this out by itself for a Mega-pixel image reconstruction method basically wastes one million bits of freely available information. Not using the available mathematical description of how a measurement sensed a cosmic field might waste even more. Furthermore, it might make it impossible to give accurately calibrated results since how should the machine tell apart a variation in instrumental sensitivity from a sky brightness variation?

• Typical neural networks get trained for fixed input and output spaces. But the variety of instruments, observational modes, and scientific questions requires methods and results that can be flexibly reused. The answer should be a representation of a physical field, from which as much as possible the measurement modality is removed, so that it can be used to answer various physical questions.

These problems are well known in AI research, and ways to handle those are increasingly used. One way is the usage of Bayesian reasoning, the extension of logic to uncertainty [1], in the way AI systems are set up. Another is to adapt appropriately abstracted representations. For example the generalization of traditional neural networks, which act on finite dimensional vectors, to neural operators, which operate on fields, is definitively going into the right direction for field inference. The reconstructed field should become independent of the used computational grid as soon the resolution of that is high enough and neural operators are a way to achieve this. Furthermore, the inclusion of domain knowledge into the construction of inference schemes is more often seen also at this conference. Physics informed networks are nice, they encourage solutions that orient themselves along physical laws via physics-violation loss terms in their objective functions [2, 3]. However, better would be solutions that can not violate these laws at all by construction. Finally, it is more and more realized that generative models are not only a good way to generate samples from conditional probability distributions (all what Bayes theorem is about), but also provide us with a language that makes the inclusion of effects straight forward: Isn't physics all about building a forward model for the universe?

Thus, current trends in AI research on field inference go into the direction to separate data and field representations (by the usage of neural operators), to use Bayesian logic, to quantify remaining uncertainties, to impose physical laws, and to use generative models. A mathematical framework for field inference that already does all of this is information field theory (IFT, [4, 5]). IFT allows one to derive optimal field reconstruction algorithms.

The algorithms derived with IFT can be implemented with the software package *Numerical Information Field Theory* (NIFTy, [6–8]). As explained in the following, these algorithms can be understood in terms of AI terminology as Variational Inference (VI) of latent spaces of neural operator based generative networks with a physics-inspired architecture. Thus, IFT can be regarded as a specific AI methodology that is designed for knowledge guided inference of physical fields [9].

These proceedings are organized as follows. First we discuss the mathematical preliminaries of inferring a field from data, leading us to IFT. Given this, ways to solve IFT problems on a computer using the NIFTy will be discussed, in particular the ideas behind novel VI methods for ultra-high dimensional problems. An outlook on the emerging *Universal Bayesian Imaging Kit* (UBIK, [10]) is given, a unification of image and signal reconstruction for all kinds of measurement devices. Finally, we conclude with some thoughts about the convergence of IFT and other AI methodologies.

https://doi.org/10.1017/S1743921325105048 Published online by Cambridge University Press

2. Theory of Field Inference

Any physical field $\varphi(x)$ depends on spatial and temporal coordinates $x = (\vec{x}, t)$ in a somehow smooth manner. Smoothness means that field values at nearby locations are related. If they were not, very often physical laws trigger an instantaneous dynamic that connects or mixes the field values of nearby locations. If one knows the field at x, the value at x' is similar. How similar might be expressed by the usage of a two point correlation function

$$C_\varphi(x, x') = \langle \varphi(x)\, \varphi^*(x') \rangle_{(\varphi)} \,, \tag{1}$$

where we introduced averaging about a field distribution function $\mathscr{P}(\varphi)$ and denote complex conjugation by an asterix. If only this field correlation structure is know about the field, the maximum entropy principle [11, 12] encourages to assign a Gaussian probability description with this covariance as a summary of the knowledge on the field,

$$\mathscr{P}(\varphi|C_\varphi) \,=\, \mathscr{G}(\varphi, C_\varphi) \;\equiv\; \mathscr{N}(\varphi|0, C_\varphi) \tag{2}$$

$$:= \frac{1}{\sqrt{|2\pi C_\varphi|}} \exp\left[-\frac{1}{2}\varphi^\dagger C_\varphi^{-1} \varphi\right]. \tag{3}$$

Here, the notion of a scalar product between two fields ϕ and ψ is introduced by defining $\phi^\dagger \psi = \int dx\, \phi^*(x)\, \psi(x)$.

Sometimes other constraints are known, for example the sky brightness field $s(x)$ has to be strictly positive. This can be modeled via suitable point wise transformations, e.g. by assigning

$$s(x) = s_0 \, \exp(\varphi(x)), \tag{4}$$

or the like. The exponentiation chosen here brings the often wanted property that s can easily vary over several orders of magnitude by a relative moderate variation of φ.

From now on, the field of interest should be denoted by s, where s stands for signal, the quantity to be inferred. Be I the collection of all the background knowledge we have on s (e.g. its correlations structure, whether it is positive, etc.), then $\mathscr{P}(s|I)$ is a knowledge state over the space of field configurations.

In order to learn more about a field its has to be probed by a measurement device which delivers data d. The data are informative on the field if its likelihood $\mathscr{P}(d|s, I)$ differs for different field configurations, $\mathscr{P}(d|s, I) \neq \mathscr{P}(d|I)$. The data might be considered to be split into a signal response $R(s) := \langle d \rangle_{(d|s,I)}$ and noise $n := d - R(s)$, so that $d = R(s) + n$, but this is not essential here. Essential is that data and its likelihood allow us to update the prior field knowledge state $\mathscr{P}(s|I)$ to the Bayesian posterior state (after the measurement)

$$\mathscr{P}(s|d, I) = \frac{\mathscr{P}(d, s|I)}{\mathscr{P}(d|I)} = \frac{\mathscr{P}(d|s, I)\, \mathscr{P}(s|I)}{\mathscr{P}(d|I)} \tag{5}$$

with $\mathscr{P}(d|I) = \int \mathscr{D}s\, \mathscr{P}(d, s|I)$ the so called evidence, which ensures proper normalization.

From this posterior moments of interests can be calculated, for example the posterior mean field $m = \langle s \rangle_{(s|d,I)}$ or the posterior uncertainty dispersion $D = \langle (s - m)(s - m)^\dagger \rangle_{(s|d,I)}$ and the like.

Quantum and statistical field theory have provided a great number of mathematical tools to calculate such conditional field expectation values like Feynman diagrams, renormalization techniques and others. In order to connect to these, we rephrase Bayesian probabilities in a mathematical language that is closer to those field theories. For this we introduce the information Hamiltonian $\mathscr{H}(\cdot|\cdot) := -\ln \mathscr{P}(\cdot|\cdot)$ in analogy to the

https://doi.org/10.1017/S1743921325105048 Published online by Cambridge University Press

Hamiltonian in statistical physics and the partition functions $\mathscr{Z} = \mathscr{P}(d|I)$, which turns out to be the evidence of Bayesian inference.

With these Bayes theorem can we written as

$$\mathscr{P}(s|d,I) = \frac{e^{-\mathscr{H}(d,s|I)}}{\mathscr{Z}}, \tag{6}$$

a form that now permits us to apply field theoretical methods. Interestingly, this form brings it also closer to the AI language, as the joint information Hamiltonian of data and signal is the sum

$$\mathscr{H}(d,s|I) = \mathscr{H}(d|s,I) + \mathscr{H}(s|I) \tag{7}$$

of what would be called in AI a data fidelity term $\mathscr{H}(d|s,I)$ and a regularization term $\mathscr{H}(s|I)$.

For example, in the above case of a log-normal signal $s = s_0 \exp\varphi$, measured by a device with a linear response and Gaussian noise, so that $\mathscr{P}(d|s,I) = \mathscr{G}(d - R\,s, N)$, with R the response operator and N the noise covariance, the joint information Hamiltonian for the log-signal field φ reads

$$\begin{aligned}\mathscr{H}(d,\varphi|I) = &\frac{1}{2}\left(d - R\,s_0 e^{\varphi}\right)^{\dagger} N^{-1}\left(d - R\,s_0 e^{\varphi}\right) + \\ &\frac{1}{2}\varphi^{\dagger} C_{\varphi}^{-1}\varphi + \mathscr{H}_0.\end{aligned} \tag{8}$$

Here $\mathscr{H}_0$ collects all terms that are independent of φ and the exponential function is applied at each location x locally, $(e^{\varphi})(x) = e^{\varphi(x)}$.

For fixed data, the minimum of $\mathscr{H}(d,\varphi|I)$ w.r.t. s would correspond to the maximum a posteriori (MAP) estimator of φ. In the specific example above this would be given by the solution to the equation

$$0 = \frac{\partial \mathscr{H}(d,\varphi|I)}{\partial \varphi} = -s_0 e^{\varphi} R^{\dagger} N^{-1}\left(d - R\,s_0 e^{\varphi}\right) + C_{\varphi}^{-1}\varphi. \tag{9}$$

In general, this equation needs to be solved numerically, for example by following the negative gradient, $-\frac{\partial \mathscr{H}(d,\varphi|I)}{\partial \varphi}$, until a minimum is reached. Although this is well possible, a different approach using generative models and variational inference (VI) is recommended for such problems for three reasons.

First, the MAP solution $\varphi_{\mathrm{MAP}} = \mathrm{argmax}_{\varphi}\mathscr{H}(d,\varphi|I)$ is often not the best from a minimal reconstruction error perspective, where one wants to minimize the expectation value of the square error $|\varphi - \psi|^2 := \int dx\,(\varphi(x) - \psi(x))^2$ of a reconstruction ψ. For this, not the mode (or maximum) of the posterior is wanted, but the mean $\overline{\varphi} = \langle\varphi\rangle_{(s|d,I)}$.

Second, the mode φ_{MAP} does (in general) not give the mode of the signal field, $s_{\mathrm{MAP}} \neq s_0 \exp\varphi_{\mathrm{MAP}}$, due to the non-linearity of the relation of s and φ. The same holds for the means, $\overline{s} \neq s_0 \exp\overline{\varphi}$.

And third, the MAP solution does not provide uncertainty quantification, as only a point estimate for the field is returned.

What is desired is a set of points that are drawn from the posterior, as from such, any moment of the fields s and φ can be estimated. VI provides such samples to a good approximation. To understand the usage of VI in NIFTy, we need to formulate field inference in terms of standardized generative models.

3. Generative Field Models

Generative models are widely used in AI. A standardized generative model f_{σ} transforms some unstructured white noise vector $\xi \hookleftarrow \mathscr{G}(\xi, \mathbb{1})$ into a more structured signal $s = f_{\sigma}(\xi)$. Here, σ denotes the collection of parameters of the model, for example the

https://doi.org/10.1017/S1743921325105048 Published online by Cambridge University Press

weights of a neural network that are trained to create a signal distribution following that of some training data. Generative models allow us to easily draw samples from the prior $\mathcal{P}(s|I) \equiv \mathcal{G}(\xi, \mathbb{1})\, |\partial f_\sigma(\xi)/\partial\xi|_{\xi=f_\sigma^{-1}(s)}$ via $\xi \hookleftarrow \mathcal{G}(\xi, \mathbb{1})$ and $s = f_\sigma(\xi)$.

In IFT, such a training of generative models is often not necessary as the parameters are specified by the prior knowledge. If, for example, the field φ has a Gaussian statistics with a priori a translation invariant correlation structure $C_\varphi(x, x') = C(x - x')$, then its power spectrum $P_\varphi(k) := \int dr\, C(r)\, e^{ikr} =: (\mathcal{F}\, C)(k)$, with $\mathcal{F}$ the Fourier operator, allows one to write down a generative model for $s = s_0 \exp\varphi$, namely

$$s = f_\sigma(\xi) = \widehat{s_0} \exp\left(\mathcal{F}^{-1} \widehat{P_\varphi^{1/2}} \xi\right). \tag{10}$$

Here, $f_\sigma = \widehat{s_0} \circ \exp \circ \mathcal{F}^{-1} \circ \widehat{P_\varphi^{1/2}}$ is a neural operator chain, $\sigma = (P_\varphi, s_0)$ is the collection of all parameters of the operator network, and $\widehat{a} = \mathrm{diag}(a)$ denotes a diagonal operator with diagonal a. The signal field s can have a non-Gaussian statistic, in this specific case a log-normal one. In case the power spectrum is not know a priori, it can be modeled as well as a log-normal field with its own standardized variables ξ_{P_φ}. This rewriting of fields (and their statistics) in terms of a standardized generative models does not only facilitate the generation of prior samples. It also helps to unify inference, as every standardized model has the same prior, $\mathcal{P}(\xi|I) = \mathcal{G}(\xi, \mathbb{1})$. The corresponding joint information Hamiltonian for ξ,

$$\begin{aligned} &\mathcal{H}(d, \xi|I) \\ &= \frac{1}{2}\left(d - R\, s_0 e^{\mathcal{F}^{-1}\widehat{P_\varphi^{1/2}}\xi}\right)^\dagger N^{-1} \left(d - R\, s_0 e^{\mathcal{F}^{-1}\widehat{P_\varphi^{1/2}}\xi}\right) \\ &+ \frac{1}{2}\xi^\dagger\xi + \mathcal{H}_0, \end{aligned} \tag{11}$$

has now absorbed all model complexity into the data fidelity term. The regularization term became a simple $\mathcal{L}^2$ norm on the latent field excitation parameters ξ.

4. Variational Inference

Minimizing the standardized information Hamiltonian $\mathcal{H}(d, \xi|I) = \mathcal{H}(d|\xi, I) + \mathcal{H}(\xi|I) = \mathcal{H}(d|s = f_\sigma(\xi), I) + \frac{1}{2}\xi^\dagger\xi + \frac{1}{2}\ln|2\pi\mathbb{1}|$ provides the MAP solution for ξ, which is not necessarily the MAP solution for $s = f_\sigma(\xi)$ due to non-linearities in the model. More problematic than this is that the MAP solution is often not a good estimator for deep and high dimensional models, like those where a field and its power spectrum are to be inferred simultaneously [13]. The size of the phase space volume is ignored by MAP, but such volume factors are essential for calculating proper expectation values.

Variational Inference (VI) improves on this. In VI a simpler parametric probability distribution $\mathcal{Q}(\xi|p)$ with parameters p is adapted to the posterior $\mathcal{P}(\xi|d, I)$ by minimizing their relative entropy or Kullback-Leibler divergence

$$\mathcal{D}_{\mathrm{KL}}(\mathcal{Q}||\mathcal{P}) = \int d\xi\, \mathcal{Q}(\xi|p)\, \ln\frac{\mathcal{Q}(\xi|p)}{\mathcal{P}(\xi|d, I)} \tag{12}$$

w.r.t. p. A convenient choice for $\mathcal{Q}$ is that of a Gaussian with mean $\overline{\xi}$ and dispersion D, $\mathcal{Q}(\xi|\overline{\xi}, D) = \mathcal{G}(\xi - \overline{\xi}, D)$. Optimizing for $\overline{\xi}$ is feasible, as this is as large as any latent space vector. However, optimizing w.r.t. to the entries of D is beyond current computer capabilities for most interesting field inference problems due to the quadratic scaling of their number with the number of signal representation pixels. Fortunately, it turns out that a good approximation can be build for D as a function of $\overline{\xi}$ that can be stored and

https://doi.org/10.1017/S1743921325105048 Published online by Cambridge University Press

computed (in its action and to a good accuracy) in quasi linear time,

$$D(\bar{\xi}) \approx \left[\mathbb{1} + \mathcal{M}_{d|\bar{\xi}}\right]^{-1}, \tag{13}$$

where

$$\mathcal{M}_{d|\xi} := \left\langle \frac{\partial \mathcal{H}(d,\xi|I)}{\partial \xi} \frac{\partial \mathcal{H}(d,\xi|I)}{\partial \xi}^{\dagger} \right\rangle_{(d|\xi)} \tag{14}$$

is the Fisher information matrix. Note that this matrix does not need to be instantiated or be calculated explicitly. All what is needed is that the data average is calculated analytically for the relevant noise statistics and that the action of the resulting matrix applied to a field can be computed numerically. Thanks to automatic differentiation in NIFTy and its underlying algebra package JAX [14], this happens in the background. The integral in the Kullback-Leibler divergence is approximated using samples drawn from $\mathcal{Q}(\xi|\bar{\xi}, D(\bar{\xi}))$.

With this approximation, only the latent space posterior mean $\bar{\xi}$ needs to be optimized for. An algorithm that does this is *Metric Gaussian Variational Inference* (MGVI, [15]). It has enabled the reconstruction of relative complex astrophysical field models in the full electromagnetic spectrum, ranging from radio to gamma rays [16–19].

An extension of MGVI is *geometric Varitational Inference* (geoVI, [20]). In geoVI a second latent space is constructed via a coordinate transformation $y = g_{\bar{\xi}}(\xi)$ that follows geodesics of the metric D_ξ in a Taylor approximation around an expansion point $\bar{\xi}$. In these new coordinates the posterior becomes approximatively normal, allowing to generate samples from it, therefore from $\xi = g_{\bar{\xi}}^{-1}(y)$, and ultimately from $s = f_\sigma(\xi)$. Thus, geoVI invokes a normalizing flow. This flow needs not to be trained, it is derived from the analytical structure of the used generative model. The resulting samples from drawing from a Gaussian in y-space and converting them into signal vectors via $s = f_\sigma(g_{\bar{\xi}}^{-1}(y))$ represent the posterior distributions more accurately than MGVI. This has proven to be in particular relevant for component separation problems, where there are large degeneracies between components that are often only solved by subtle forces of the prior (meaning gradients of $\mathcal{H}(s|I)$) [21].

The MGVI and geoVI methods take as an input a standardized generative operator model, which can be written in NIFTy, and the observational data that should be interpreted in the light of the model. They then construct the Fisher information matrix and the field uncertainty covariance D as implicit (never instantiated) operators, optimize for the latent space point $\bar{\xi}$ using stochastic samples to perform the involved integrals over phase space, and return the optimal $\bar{\xi}$ as well as a number of approximate posterior samples $(\xi_i)_{i=1}^n$ scattered around it. From those, any desired moment $h(s)$ of the field s can be calculated via a sample average

$$\langle h(s) \rangle_{(s|d,I)} \approx \frac{1}{n} \sum_{i=1}^{n} h(f_\sigma(\xi_i)). \tag{15}$$

We remind that our field is given via the generative model $s = f_\sigma(\xi)$.

5. Astrophysical Applications

The quasi-linear scaling of MGVI and geoVI with problem size has allowed for tackling large reconstruction problems in astrophysics. To date, the largest one is a reconstruction of the three dimensional Galactic dust distribution around the sun, with $6 \cdot 10^8$ voxels being tomographically reconstructed from starlight reddening data and stellar

https://doi.org/10.1017/S1743921325105048 Published online by Cambridge University Press

positions [22]. Other areas of application are radio [16, 17], X-ray [23, 24], and gamma-ray [18, 19] astronomy.

The usage of IFT and NIFTy is not yet wide spread in astrophysics, as these requires some expert knowledge in IFT, its numerics, and the underlying mathematics. This seems to be a threshold for many users, despite the existence of freely available educational material [25]. In order to facilitate the usage of IFT based methods, the *Universal Bayesian Imaging* Kit (UBIK) software package is under development [10]. The idea behind UBIK is that in the IFT framework, the problem of reconstructing a field is modular with respect to the knowledge on the field (as encoded in the prior), the way the field is sensed (as encoded in the likelihood), and how the inference is performed (whether via MAP, MGVI, or geoVI). Any combination of these elements gives rise to an algorithm tailored to a specific measurement situation and scientific question. Furthermore, thanks to the additivity of information, see Eq. 7, the fusion of data sets from different instruments is well possible. For example if data sets d_1 and d_2 are independent measurements of the same signal, $\mathscr{P}(d_1, d_2|s, I) = \mathscr{P}(d_1|s, I)\,\mathscr{P}(d_2|s, I)$, the corresponding joint information Hamiltonian of data and signal,

$$\mathscr{H}(d, s|I) = \mathscr{H}(d_1|s, I) + \mathscr{H}(d_2|s, I) + \mathscr{H}(s|I), \tag{16}$$

just contains the sum of the corresponding data fidelity terms. Provided with a suite of instrument descriptions, UBIK should enable IFT based reconstructions for a large number of astrophysical instruments and multi-messenger physics with those. In its current version, it is able to handle data from the *Chandra*, the *eROSITA*, and from the *JWST* satellite. The inclusion of a larger suite of radio astronomical instrument descriptions is foreseen next.

6. Conclusion

Information field theory (IFT, [4, 5]) emerged from the need to have a mathematical theory of field inference that works in practice, for example on astrophysical data sets. On a first glance, IFT seems to be unrelated to most artificial intelligence (AI) developments, except maybe Gaussian process regression. However, the computational tools developed for IFT problems benefited from and extended methods that were designed for AI. Thus methods used in IFT applications converge towards those used in AI.

Current trends in AI towards more detector independent representation of information, e.g. via neural operators, towards a stricter adherence to probabilistic logic and physical laws (via physics informed networks), indicate that there is a convergent movement also on the AI side towards IFT principles. From these observations, it is apparent that IFT is an AI methodology that puts more emphasis on the known structure of the physical world than pure data driven AI methods [9]. But the degree of data versus knowledge driven can be freely chosen, and the inclusion of purely data driven networks into the IFT methodology has already been demonstrated successfully [26].

An area where IFT might help the understanding and development of AI methods is that the IFT concept of probability distributions over fields is a useful language to phrase the AI problem of how to train a network. The versatility of neural networks stems from their *Universal Approximation Theorem* [27]. Any reasonable function can be approximated by a network of sufficient width (or depth). This theorem is, however, also a curse: Given a finite set data, there is also an infinite number of functions that will pass through it perfectly. How and whether the right one is chosen during neural network training, and what this implies in terms of implied data generalization rules, is unclear. IFT, the information theory for fields, might provide a suitable language to address such questions.

https://doi.org/10.1017/S1743921325105048 Published online by Cambridge University Press

Acknowledgments

TE thanks the organizers of the UniversAI IAU conference in Athen 2025 for the kind invitation and Anton Nöbauer, Julian Rüstig, and Lena Tauber for comments on the manuscript. Funded by the European Union (ERC, mw-atlas, 101166905). Views and opinions expressed are however those of the author only and do not necessarily reflect those of the European Union or the European Research Council Executive Agency. Neither the European Union nor the granting authority can be held responsible for them.

References

[1] Richard T Cox. "Probability, frequency and reasonable expectation". In: *American journal of physics* 14.1 (1946), pp. 1–13.

[2] Maziar Raissi, Paris Perdikaris, and George Em Karniadakis. "Physics Informed Deep Learning (Part I): Data-driven Solutions of Nonlinear Partial Differential Equations". In: *arXiv e-prints*, arXiv:1711.10561 (Nov. 2017), arXiv:1711.10561. DOI: 10.48550/arXiv.1711.10561. arXiv: 1711.10561 [cs.AI].

[3] Maziar Raissi, Paris Perdikaris, and George Em Karniadakis. "Physics Informed Deep Learning (Part II): Data-driven Discovery of Nonlinear Partial Differential Equations". In: *arXiv e-prints*, arXiv:1711.10566 (Nov. 2017), arXiv:1711.10566. DOI: 10.48550/arXiv.1711.10566. arXiv: 1711.10566 [cs.AI].

[4] Torsten A. Enßlin, Mona Frommert, and Francisco S. Kitaura. "Information field theory for cosmological perturbation reconstruction and nonlinear signal analysis". In: *Phys. Rev. D* 80.10, 105005 (Nov. 2009), p. 105005. DOI: 10.1103/PhysRevD.80.105005. arXiv: 0806.3474 [astro-ph].

[5] Torsten A. Enßlin. "Information Theory for Fields". In: *Annalen der Physik* 531.3 (2019), p. 1800127. DOI: https://doi.org/10.1002/andp.201800127. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/andp.201800127. https://onlinelibrary.wiley.com/doi/abs/10.1002/andp.201800127.

[6] M. Selig et al. "NIFTY - Numerical Information Field Theory. A versatile PYTHON library for signal inference". In: *Astron. and Astroph.* 554, A26 (June 2013), A26. DOI: 10.1051/0004-6361/201321236. arXiv: 1301.4499 [astro-ph.IM].

[7] Gordian Edenhofer et al. "Re-Envisioning Numerical Information Field Theory (NIFTy.re): A Library for Gaussian Processes and Variational Inference". In: *The Journal of Open Source Software* 9.98, 6593 (June 2024), p. 6593. DOI: 10.21105/joss.06593. arXiv: 2402.16683 [astro-ph.IM].

[8] Philipp Arras et al. *NIFTy5: Numerical Information Field Theory v5*. Astrophysics Source Code Library, record ascl:1903.008. Mar. 2019.

[9] Torsten Enßlin. "Information field theory and artificial intelligence". In: *Entropy* 24.3 (2022), p. 374.

[10] Vincent Eberle et al. "J-UBIK: The JAX-accelerated Universal Bayesian Imaging Kit". In: *arXiv e-prints*, arXiv:2409.10381 (Sept. 2024), arXiv:2409.10381. DOI: 10.48550/arXiv.2409.10381. arXiv: 2409.10381 [astro-ph.IM].

[11] E. T. Jaynes. "Information Theory and Statistical Mechanics". In: *Phys. Rev.* 106 (4 May 1957), pp. 620–630. DOI: 10.1103/PhysRev.106.620. https://link.aps.org/doi/10.1103/PhysRev.106.620.

[12] E. T. Jaynes. "Information Theory and Statistical Mechanics. II". In: *Phys. Rev.* 108 (2 Oct. 1957), pp. 171–190. DOI: 10.1103/PhysRev.108.171. https://link.aps.org/doi/10.1103/PhysRev.108.171.

[13] Torsten A. Enßlin and Mona Frommert. "Reconstruction of signals with unknown spectra in information field theory with parameter uncertainty". In: *Phys. Rev. D* 83 (10 May 2011), p. 105014. DOI: 10.1103/PhysRevD.83.105014. https://link.aps.org/doi/10.1103/PhysRevD.83.105014.

[14] James Bradbury et al. *JAX: composable transformations of Python+NumPy programs*. Version 0.3.13. 2018. http://github.com/jax-ml/jax.

https://doi.org/10.1017/S1743921325105048 Published online by Cambridge University Press

[15] Jakob Knollmüller and Torsten A. Enßlin. "Metric Gaussian Variational Inference". In: *arXiv e-prints*, arXiv:1901.11033 (Jan. 2019), arXiv:1901.11033. DOI: 10.48550/arXiv.1901.11033. arXiv: 1901.11033 [stat.ML].

[16] Henrik Junklewitz et al. "RESOLVE: A new algorithm for aperture synthesis imaging of extended emission in radio astronomy". In: *Astronomy & Astrophysics* 586 (2016), A76.

[17] Jakob Roth et al. "fast-resolve: Fast Bayesian radio interferometric imaging". In: *Astron. and Astroph.* 690, A387 (Oct. 2024), A387. DOI: 10.1051/0004-6361/202451107. arXiv: 2406.09144 [astro-ph.IM].

[18] Marco Selig et al. "The denoised, deconvolved, and decomposed Fermi γ-ray sky. An application of the D^{3}PO algorithm". In: *Astron. and Astroph.* 581, A126 (Sept. 2015), A126. DOI: 10.1051/0004-6361/201425172. arXiv: 1410.4562 [astro-ph.HE].

[19] L. I. Scheel-Platz et al. "Multicomponent imaging of the Fermi gamma-ray sky in the spatio-spectral domain". In: *Astron. and Astroph.* 680, A2 (Dec. 2023), A2. DOI: 10.1051/0004-6361/202243819. arXiv: 2204.09360 [astro-ph.HE].

[20] Philipp Frank, Reimar Leike, and Torsten A. Enßlin. "Geometric Variational Inference". In: *Entropy* 23.7, 853 (July 2021), p. 853. DOI: 10.3390/e23070853. arXiv: 2105.10470 [stat.ME].

[21] Sebastian Hutschenreuter et al. "Disentangling the Faraday rotation sky". In: *Astron. and Astroph.* 690, A314 (Oct. 2024), A314. DOI: 10.1051/0004-6361/202346740. arXiv: 2304.12350 [astro-ph.GA].

[22] Gordian Edenhofer et al. "A parsec-scale Galactic 3D dust map out to 1.25 kpc from the Sun". In: *Astron. and Astroph.* 685, A82 (May 2024), A82. DOI: 10.1051/0004-6361/202347628. arXiv: 2308.01295 [astro-ph.GA].

[23] M. Westerkamp et al. "The first spatio-spectral Bayesian imaging of SN1006 in X-rays". In: *Astron. and Astroph.* 684, A155 (Apr. 2024), A155. DOI: 10.1051/0004-6361/202347750. arXiv: 2308.09176 [astro-ph.HE].

[24] Vincent Eberle et al. "Bayesian Multi-wavelength Imaging of the LMC SN1987A with SRG/eROSITA". In: *arXiv e-prints*, arXiv:2410.14599 (Oct. 2024), arXiv:2410.14599. DOI: 10.48550/arXiv.2410.14599. arXiv: 2410.14599 [astro-ph.IM].

[25] T. Enßlin. *Information Field Theory*. 2025. eprint: http://www.mpa-garching.mpg.de/ift/.

[26] Jakob Knollmüller and Torsten A. Enßlin. "Bayesian Reasoning with Trained Neural Networks". In: *Entropy* 23.6, 693 (May 2021), p. 693. DOI: 10.3390/e23060693. arXiv: 2001.11031 [cs.LG].

[27] Kurt Hornik, Maxwell Stinchcombe, and Halbert White. "Multilayer feedforward networks are universal approximators". In: *Neural networks* 2.5 (1989), pp. 359–366.

https://doi.org/10.1017/S1743921325105048 Published online by Cambridge University Press