Information field dynamics for simulation scheme construction

Torsten A. Enßlin[*]

Max Planck Institute for Astrophysics, Karl-Schwarzschildstr. 1, 85741 Garching, Germany

Information field dynamics (IFD) is introduced here as a framework to derive numerical schemes for the simulation of physical and other fields without assuming a particular subgrid structure as many schemes do. IFD constructs an ensemble of nonparametric subgrid field configurations from the combination of the data in computer memory, representing constraints on possible field configurations, and prior assumptions on the subgrid field statistics. Each of these field configurations can formally be evolved to a later moment since any differential operator of the dynamics can act on fields living in continuous space. However, these virtually evolved fields need again a representation by data in computer memory. The maximum entropy principle of information theory guides the construction of updated data sets via *entropic matching*, optimally representing these field configurations at the later time. The field dynamics thereby become represented by a finite set of evolution equations for the data that can be solved numerically. The subgrid dynamics is thereby treated within auxiliary analytic considerations. The resulting scheme acts solely on the data space. It should provide a more accurate description of the physical field dynamics than simulation schemes constructed *ad hoc*, due to the more rigorous accounting of subgrid physics and the space discretization process. Assimilation of measurement data into an IFD simulation is conceptually straightforward since measurement and simulation data can just be merged. The IFD approach is illustrated using the example of a coarsely discretized representation of a thermally excited classical Klein-Gordon field. This should pave the way towards the construction of schemes for more complex systems like turbulent hydrodynamics.

I. INTRODUCTION

A. Motivation

Computer simulations of fields play a major role in science, engineering, economics, and many other areas of modern life. Computer limitations require that the infinite number of degrees of freedom of a field are represented by a finite data set that fits into computer memory. For example, in hydrodynamics with mesh codes, the average density, pressure, and velocities of the fluid within grid cells form the data. The data make statements about the field properties, and the simulation scheme describes how the present data determine the future data. This dynamics is usually set up such that the continuum limit of an infinite number of infinitesimal dense grid points recovers the partial differential equations governing the physical field dynamics. However, there are many possible schemes to discretize the differential operators of the field equations. Which one gives good results already at finite resolution? Which one takes the influence of processes on subgrid scales best into account? To address these questions, a rigorous approach to construct accurate simulation schemes, information field dynamics (IFD), is presented here. IFD rests on information field theory (IFT), the theory of Bayesian inference on fields [1,2]. In the ideal case, IFD and IFT provide identical results, since both can be used to make statements about fields at later times given some initial data. However, in real-world applications of simulation schemes, compromises with respect to accuracy and computational complexity are often unavoidable. Thus, IFD can be regarded as a particular approximation scheme within IFT, which may or may not provide optimal results from an information-theoretical point of view.

The basic idea is that IFT turns the data in computer memory into an ensemble of field configurations which are consistent with the data and the knowledge on the subgrid physics and field statistics. The differential operators of the field dynamics can then formally operate on these field configurations without the usual discretization approximation. An unavoidable approximation finally happens when these time-evolved fields get recast into the finite data representation in computer memory. The information-theoretical guideline of the maximum entropy principle (MEP) is used in order to ensure maximal fidelity of this operation, which we call in the following *entropic matching*. The subgrid dynamics is thereby treated within an auxiliary analytic consideration. In the end, an IFD simulation scheme for the time evolution of a field is a pure data updating operation in computer memory and, therefore, an implementable algorithm. Although this algorithm does not explicitly deal with a field living in continuous space, it was, however, derived with the continuous-space version of the original problem being very present in the mathematical reasoning. The subgrid information, which IFT used to construct the virtual continuous-space field configurations, is encapsulated implicitly in the resulting IFD scheme. Therefore, IFD schemes act solely on the data in computer memory without using any explicit subgrid field representation.

When constructing a computational simulation scheme for field dynamics, whether using IFD or not, one is facing two bottlenecks: finite computer memory and finite computational time. This work deals only with the first issue and explains how to construct schemes which optimally use the data stored in computer memory. Optimizing with respect to only one objective, memory in this case, very often results in solutions

_____

[*]ensslin@mpa-garching.mpg.de

which are ineffective with respect to another aim, which here is computational simplicity. Thus, we do not expect the resulting IFD schemes necessarily to be the optimal solution for a concrete computational problem. Deriving practically usable schemes will often require additional approximations in order to reduce the computational complexity. The IFD framework can, however, help to clarify the nature of the approximations made and guide the design of simulation schemes.

The concrete problem of how to discretize a thermally excited Klein-Gordon (KG) field in position space will illustrate the usage of the theoretical IFD framework.

B. Previous work

Our main motivation is to aid the construction of simulation schemes, for example, in hydrodynamics, for which a very rich body of previous work exists. Appendix A discusses briefly the relevant concepts of partial differential equation discretization, subgrid modeling, and information-theoretical concepts in simulation schemes and their relation to IFD.

C. Structure of this work

In Sec. II we introduce the necessary concepts of IFT, MEP, and IFD. In Sec. III IFD is developed in detail on an abstract level, as well as for the illustrative example of a KG field. The fidelity of IFD and a typical *ad hoc* scheme for the KG field are compared numerically and against an exact solution in Sec. IV. Section V contains the conclusion and outlook.

II. CONCEPTS

A. Information field theory

The idea of this work is that the data stored in a computer is only a constraint on possible field configurations but does not fully determine a unique subgrid field configuration. Instead, the ensemble of possible field configurations is constructed using IFT. IFT blends the information in the data and any prior knowledge on the field behavior into a single probability density function (PDF) over the space of all field configurations.

IFT is information theory applied to fields, probabilistic reasoning for an infinite set of unknowns, the field values at all space positions. It provides field reconstructions from finite data. For this IFT needs data, a data model describing how the data are determined by the field, and a prior PDF summarizing the statistical knowledge on the field degrees of freedom prior to the data. How this works in our case will be shown in the following. A general introduction to IFT can be found in Ref. [2] and in the references therein.

IFT exploits mathematical methods from quantum and statistical field theory. The unknown field ϕ is regarded as a signal, a hidden message to be revealed from the data d. A prior PDF $\mathcal{P}(\phi)$ describes the knowledge about the signal field prior to the data, and a likelihood PDF $\mathcal{P}(d|\phi)$ describes the probability of the data given a specific signal field configuration. Bayes's theorem allows one to construct the posterior PDF,

$$\mathcal{P}(\phi|d) = \frac{\mathcal{P}(d|\phi)\,\mathcal{P}(\phi)}{\mathcal{P}(d)}, \tag{1}$$

which summarizes the *a posteriori* (after the data is taken) knowledge on the signal field. The connection to statistical field theory becomes apparent when one realizes that Bayes's theorem can also be written as

$$\mathcal{P}(\phi|d) = \frac{e^{-H(d,\phi)}}{Z(d)}, \tag{2}$$

with the information Hamiltonian

$$H(d,\phi) = -\log \mathcal{P}(d,\phi) = -\log \mathcal{P}(d|\phi) - \log \mathcal{P}(\phi) \tag{3}$$

and the partition function

$$Z(d) = \mathcal{P}(d) = \int \mathcal{D}\phi\, \mathcal{P}(d,\phi) = \int \mathcal{D}\phi\, e^{-H(d,\phi)}. \tag{4}$$

Here, $\int \mathcal{D}\phi$ denotes a phase-space integral over all possible field configurations of ϕ, a so-called path integral.

The information Hamiltonian combines prior and likelihood into a signal energy, which determines the signal posterior according to the usual Boltzmann statistics. This Hamiltonian therefore contains all available information on the signal field.

The simplest IFT case is that of a free theory. This emerges when three conditions are met:

(i) The *a priori* distribution of the field is a multivariate Gaussian,

$$\mathcal{P}(\phi) = \mathcal{G}(\phi,\Phi) = \frac{1}{\sqrt{|2\pi\,\Phi|}} \exp\left(-\frac{1}{2}\phi^\dagger \Phi^{-1} \phi\right), \tag{5}$$

with signal covariance $\Phi = \langle \phi\,\phi^\dagger \rangle_{(\phi)} = \int \mathcal{D}\phi\, \mathcal{P}(\phi)\, \phi\,\phi^\dagger$, its determinant $|\Phi| = \det\Phi$, and $\phi^\dagger \psi = \int dx\, \overline{\phi}_x \psi_x$ denoting the scalar product.

(ii) The data depend linearly on the signal field,

$$d = R\,\phi + n, \tag{6}$$

with a known response operator R.

(iii) The noise $n = d - R\phi$ is signal independent with Gaussian statistics,

$$\mathcal{P}(n|\phi) = \mathcal{G}(n,N), \tag{7}$$

where $N = \langle n\,n^\dagger \rangle_{(n|\phi)} = \int \mathcal{D}n\, \mathcal{P}(n|\phi)\, n\,n^\dagger$.

In this case, the likelihood $\mathcal{P}(d|\phi) = \mathcal{P}(n = d - R\phi|\phi) = \mathcal{G}(d - R\phi, N)$ and the prior $\mathcal{P}(\phi)$ contribute terms to the Hamiltonian that are at most quadratical in the signal. Thus, the Hamiltonian is also quadratical, which is the mark of a free theory. In this specific case, the information Hamiltonian states that the posterior field is also Gaussian, but with a shifted mean $m = \langle \phi \rangle_{(\phi|d)} = \int \mathcal{D}\phi\, \phi\, \mathcal{P}(\phi|d)$ and uncertainty variance D, which can be read off from

$$
\begin{aligned}
H(d,\phi) &\cong \frac{1}{2}(d - R\phi)^\dagger N^{-1}(d - R\phi) + \frac{1}{2}\phi^\dagger \Phi^{-1}\phi^\dagger \\
&\cong \frac{1}{2}[\phi^\dagger \underbrace{(\Phi^{-1} + R^\dagger N^{-1} R)}_{D^{-1}} \phi + \phi^\dagger \underbrace{R^\dagger N^{-1} d}_{j} + j^\dagger \phi] \\
&= \frac{1}{2}(\phi^\dagger D^{-1}\phi + \phi^\dagger j + j^\dagger \phi) \\
&\cong \frac{1}{2}(\phi - m)^\dagger D^{-1}(\phi - m),
\end{aligned}
\tag{8}
$$

with

$$m = D\,j = \underbrace{(\Phi^{-1} + R^\dagger N^{-1} R)^{-1} R^\dagger N^{-1}}_{W}\,d = W d. \quad (9)$$

Here and later "$\hat{=}$" means equality up to irrelevant constants.[1] In analogy to the quantum field theory, an information propagator $D = (\Phi^{-1} + R^\dagger N^{-1} R)^{-1}$ and an information source $j = R^\dagger N^{-1} d$ can be identified. The information source j is given by the data d, weighted by the inverse noise covariance N^{-1} and back-projected with the Hermitian adjoint response $R^\dagger$ into the signal space. The *a posteriori* mean field m_x at some location x of the signal space is constructed by transporting the information j_y sourced by the data at some location y to x with the help of the information propagator D_{xy}. This happens by applying this as a linear operator to the information source field $m_x = \int dy\, D_{xy}\, j_y$. The resulting posterior mean field depends linearly on the data, $m = W d$. The corresponding linear filter operation W is well known in signal reconstruction as the (generalized) Wiener filter [3]. The information propagator D is also identical to the *a posteriori* uncertainty variance,

$$D = \langle (\phi - m)(\phi - m)^\dagger \rangle_{(\phi|d)}, \quad (10)$$

also known under the term Wiener variance. To conclude, in free IFT, the posterior is Gaussian with Wiener mean and variance,

$$\mathcal{P}(\phi|d) = \mathcal{G}(\phi - m, D). \quad (11)$$

Although the field mean m is a continuous function in the signal space, a full field with an apparently infinite number of field values, it has, strictly speaking, only effectively a finite number of degrees of freedom due to its construction. Since the mean field is a deterministic function of the data, $m = m(d) = W d$, the phase space of possible mean fields can have at most as many dimensions as the data have degrees of freedom. This sets a limit to the maximal possible accuracy a simulation scheme can achieve with finite data representation of the field. However, in this work, we evolve not only the mean field but also the full distribution of plausible fields around this, as characterized by $\mathcal{P}(\phi|d)$.

It should be noted that there exist two equivalent formulations of the Wiener filter operator,

$$\begin{aligned} W &= (\Phi^{-1} + R^\dagger N^{-1} R)^{-1} R^\dagger N^{-1} \\ &= \Phi R^\dagger (R\,\Phi\,R^\dagger + N)^{-1}. \end{aligned} \quad (12)$$

The first one is called the signal space and the second one the data space representation, since the operator inversions happen

in signal and data space, respectively. They are fully equivalent as long as Φ and N are regular matrices.[2]

The data-space representation of the Wiener filter $W = \Phi R (R\,\Phi\,R^\dagger + N)^{-1}$ can cope with the here relevant case of negligible noise, $N \to 0$, leading to $W = \Phi R (R\,\Phi\,R^\dagger)^{-1}$. This is possible only if $\tilde{\Phi} = R\,\Phi\,R^\dagger$, the data-space image of the signal field covariance, is (pseudo-)invertible, which is very often the case. If not, the data contain redundancies that could be used to tailor the data space until $\tilde{\Phi}$ is invertible.

This noiseless limit might be a desirable assumption for dealing with the data of a numerical simulation, since one might define the data to represent a statement about the field like $d = R\,\phi$ exactly, without any uncertainty in data space. However, in the course of a field dynamical simulation, the knowledge of the exact field configuration ϕ might not be present at later times due to unavoidable discretization errors. Therefore, a mismatch of the data d in computer memory and the correct discretized statement $R\,\phi$ for the true field might develop and this can be regarded as noise $n = d - R\,\phi$. Furthermore, a full error propagation of initial value uncertainties in a simulation might be of interest in case the initial data resulted from a real measurement with instrumental noise. For these reasons, we will keep the noise term in the formalism.

The Wiener filter theory described so far gives us a sufficient IFT background for this initial work on IFD. It should be noted, however, that in the case of nonlinear relations between data and signal, or non-Gaussian signal or noise statistics, IFT becomes an interacting field theory, and the resulting operations on the data to calculate *a posteriori* mean and variance become nonlinear. Such operations can be constructed using diagrammatic perturbation series, resummation, and renormalization techniques [2,4] or by use of the construction and minimization of an effective action, the Gibbs free energy [5,6]. In many cases, the posterior is well approximated by a multivariate Gaussian, which we assume in the following.

B. Entropic matching

We assume now that an ensemble of field configurations for a time t has been constructed with IFT, those being consistent with the data $d = d_t$ and any background information at that time. It has to be specified now how those evolve and how this can be represented by an updated data set $d' = d_{t'}$ at a later time t'.

Each of the possible field configurations is assumed to evolve for a short period according to the exact physical field dynamics. In order to recast this evolved ensemble of field configurations back into the data representation of the computational scheme, an updated data set has to be constructed. The field ensemble implied by the updated data should resemble the evolved field ensemble of the original

---

[1]This is, of course, a context-dependent convention, since it depends on what is regarded to be relevant. In the context of this work, any field-dependent quantity is relevant. Field-independent normalization constants of PDFs are not. The sign "$\hat{=}$" is here used as the logarithmic partner of the sign "$\propto$", since normalization constants become constant additive terms after taking the logarithms. Later, we will also regard terms of higher order in the time step δt as irrelevant, since they can be made to vanish by taking the limit $\delta t \to 0$.

---

[2]The equivalence of the two Wiener filter representations is easily verified via the following equivalence transformations:

$$\begin{aligned} &(\Phi^{-1} + R^\dagger N^{-1} R)^{-1} R^\dagger N^{-1} = \Phi R^\dagger (R\,\Phi\,R^\dagger + N)^{-1} \\ \Leftrightarrow\ &R^\dagger N^{-1}(R\,\Phi\,R^\dagger + N) = (\Phi^{-1} + R^\dagger N^{-1} R)\Phi R^\dagger \\ \Leftrightarrow\ &R^\dagger N^{-1} R\,\Phi\,R^\dagger + R^\dagger = R^\dagger + R^\dagger N^{-1} R\,\Phi\,R^\dagger. \end{aligned}$$

data as close as possible. We will use entropic matching for this, the usage of the MEP without any additional constrains. The MEP is the principle of our choice since it derives from very generic and desirable first principles on how to update a probability without introducing spurious knowledge.

For the MEP, entropy is just regarded as an abstract quantity that can be used to rank various possible PDFs according to how well they are suited to represent a knowledge state. A large entropy resembles an uninformed or ignorance state. MEP aims, therefore, for the least informed state that is still consistent with all known constraints. This should be the state with the least spurious assumptions.

A number of intuitively obvious requirements on the internal logic of such a ranking fully determines the functional form of this entropy [7–10]. These requirements are that local information should have only local effects, that the ranking should be independent of the coordinate system used, and that independent systems lead to separable PDFs. These requirements are further detailed in Appendix B. The only function on the space of PDFs that is consistent with these principles is the entropy

$$\mathcal{S}(\mathcal{P}|\mathcal{Q}) = -\int \mathcal{D}\phi\, \mathcal{P}(\phi)\, \log\left[\frac{\mathcal{P}(\phi)}{\mathcal{Q}(\phi)}\right], \qquad (13)$$

where $\mathcal{P}(\phi)$ denotes a PDF for some field ϕ to be ranked for its ignorance and $\mathcal{Q}(\phi)$ an *a priori* ignorance state. This entropy is the relative entropy of information theory, the Kullback-Leibler divergence of $\mathcal{P}$ to $\mathcal{Q}$ [10]. It is in general also equivalent (up to some constant) to the Gibbs energy of thermodynamics [5] and to the Boltzmann-Shannon entropy in case the ignorance knowledge state $\mathcal{Q}$ does not favor any region of physical phase space, i.e., $\mathcal{Q}(\phi) = $ const.

Since the information entropy is equivalent to the Kullback-Leibler distance of information theory, it can also be used to match one PDF optimally to another one. This entropic matching will be needed in this work in order to find the data constrained representation of the field PDF at a later instant that best matches the time-evolved PDF of an earlier instant. In case $\mathcal{P}(\phi)$ can be changed at any phase-space point ϕ, maximizing $\mathcal{S}(\mathcal{P}|\mathcal{Q})$ will reproduce the ignorance prior $\mathcal{P} \to \mathcal{Q}$. If there are, however, constraints limiting the flexibility of $\mathcal{P}(\phi)$ to adapt to $\mathcal{Q}(\phi)$, the MEP solution will differ. Such constraints can be imposed with the help of Lagrange multipliers, respective thermodynamical potentials, which can be used to imprint certain expectation values onto $\mathcal{P}$ as shown in Appendix B. In this work, constraints arise due to the fact that the degrees of freedom to represent functions and PDFs in computers are limited by the size of the computer memory. To be concrete, we write $\phi' = \phi_{t'}$ and assume for definiteness only that the short time step $\delta t = t' - t$ permits a deterministic and invertible functional relation between ϕ' and the earlier $\phi = \phi_t$, so $\mathcal{P}(\phi'|\phi) = \delta[\phi' - \phi'(\phi)]$ as well as $\mathcal{P}(\phi|\phi') = \delta[\phi - \phi(\phi')]$.[3]

Here and later, we assume further that the target knowledge state $\mathcal{Q}$ in our case is given by the Gaussian signal field posterior $\mathcal{P}(\phi|d,t) = \mathcal{G}(\phi - m, D)$ at time t as specified by the data $d = d_t$, and the background knowledge at this time, however, evolved according to the dynamical laws to a later time t', so

$$\mathcal{Q}(\phi') = \mathcal{P}(\phi'|d) = \int \mathcal{D}\phi\, \mathcal{P}(\phi'|\phi)\, \mathcal{P}(\phi|d)$$

$$= \mathcal{G}(\phi(\phi') - m, D)\left|\frac{\partial\phi}{\partial\phi'}\right|. \qquad (14)$$

The state $\mathcal{P}'$ we want to match to this using the MEP is one that can be represented by a new set of data $d' = d_{t'}$ at this later time via the IFT posterior $\mathcal{P}'(\phi') = \mathcal{P}(\phi'|d') = \mathcal{G}(\phi' - m', D')$. Since the data degrees of freedom are finite, the PDF implied by this new data [via $m' = W'd'$ and $D' = (\Phi'^{-1} + R'^\dagger N'^{-1} R')^{-1}$] will be of a parametric form, with the new data being the parameters. However, the evolved PDF will in general have a different functional form. Therefore, a matching between the PDFs $\mathcal{P}'(\phi'|d')$ and $\mathcal{Q}(\phi')$ is needed and using the MEP for this ensures that the least amount of spurious information is introduced in this unavoidable approximative step.

C. Simulation schemes construction

The IFD methodology to discretize the dynamics of a field can be summarized as followings:

(A) Field dynamics. The field dynamics equations have to be specified. The KG equation, which can be derived from a suitable Hamiltonian, will serve as an example in this work.

(B) Prior knowledge. The ignorance knowledge state in the case where there is an absence of data has to be specified. In our example the field will be assumed to be initially excited by contact with a thermal bath of known temperature. The Hamiltonian determining the field dynamics will, therefore, also determine the background knowledge on the initial state in our example.

(C) Data constraints. The relation of data and the ensemble of field configurations being consistent with data and background knowledge has to be established using IFT. Assimilation of external measurement data into the simulation scheme is naturally done during this step.

(D) Field evolution. The evolution of the field ensemble over a short time interval has to be described. This involves either the evolution of the mean and spread of the ensemble or, as we will use here, the analytical description of the evolution of all possible field configurations.

(E) Prior update. The background knowledge for the later time has to be constructed. In the chosen example, energy and phase-space conservation of the Hamiltonian dynamics guarantee that the same thermal ignorance state also holds at later times.

(F) Data update. The relation of data and field ensemble has to be invoked again to construct the data of the later time using entropic matching based on the MEP. Thereby a transformation rule is constructed that describes how the initial data determine the later data. This transformation forms the desired numerical simulation scheme. It has incorporated the physics of the subgrid degrees of freedom into operations solely in data space.

---

[3]Stochastic terms could easily be incorporated into the dynamics, e.g., by setting $\mathcal{P}(\phi'|\phi) = \mathcal{G}(\phi' - \phi'(\phi), \delta t\, \Xi)$ in case of additive Gaussian and temporally white noise ξ_t with covariance $\langle \xi_t \xi_{t'}^\dagger \rangle_{(\xi)} = \delta(t - t')\, \Xi$. This is a straightforward extension of the scheme presented here [11].

An IFD simulation scheme resulting from this recipe acts only on the data space. Any subgrid dynamics is encapsulated implicitly. This is ensured by the auxiliary analytic considerations that construct the ensemble of possible field configurations, evolve them analytically in time, and map them back onto the data representation using entropic matching.

III. INFORMATION FIELD DYNAMICS

The IFD program outlined above shall now be discussed in detail and by following the recipe of Sec. II C step by step. The discussion will deal only with linear dynamics and Gaussian knowledge states. Many interesting problems involve nonlinear dynamics and, consequently, should lead to non-Gaussian knowledge states. However, the construction of a nonlinear IFD theory will have its foundation in linear theory, which, therefore, needs to be developed first.

▶ In order to illustrate the IFD methodology, the problem of how to discretize the dynamics of a thermally excited Klein-Gordon (KG) field in one-dimensional position space is chosen as an example. Since exact solutions of the field dynamics can easily be given in Fourier-space representation, an exact, subgrid field model exists in this case to which numerical solutions using IFD and other discretization schemes can be compared. Passages dealing specifically with this example are marked as this paragraph and might be skimmed over on a first reading. ◀

A. Field dynamics

The linear dynamics of a field ϕ can in general be written as

$$\partial_t \phi = c + L \phi, \tag{15}$$

where L is a linear operator acting on the field vector of a time instance, thereby determining the field's time derivative. L can be a differential operator, it can include integro-differential operations, and it can depend on time. A dependence on earlier field values is excluded from L, which is therefore assumed here to be local in time. The field-independent, but potentially time- and position-dependent, additive term c is a source term of the field.

Nonlinear dynamics of the form

$$\partial_t \chi = F(\chi) \tag{16}$$

can often be cast approximatively into the form (15) via a Fréchet-Taylor expansion around a sufficiently good and known approximation ψ for $\chi = \psi + \phi$,

$$\partial_t \phi = \underbrace{F(\psi) - \partial_t \psi}_{c} + \underbrace{\partial_\psi F(\psi)}_{L} \phi + O(\phi^2). \tag{17}$$

One obvious choice of such an approximation would be to use a static function $\psi_t = \chi_{t_0}$ for some short period $[t_0, t_1]$ and afterwards $\psi_t = \chi_{t_1}$ for the next such period, always ensuring ϕ to be small and second-order effects to be negligible.

Stochastic terms in the evolution equations can also be included into the formalism; however, here we refrain from such complications and assume fully deterministic dynamics. If higher time derivatives are part of the linear or linearized evolution equation, these can be included as further components of ϕ.

▶ For example, the one-dimensional KG equation for a real scalar field with mass μ,

$$\partial_t^2 \varphi = (\partial_x^2 - \mu^2)\varphi, \tag{18}$$

which will serve as a concrete example in this work, can be cast into the form (15) by setting $\phi = (\varphi^\dagger, \pi^\dagger)^\dagger$ and

$$\partial_t \begin{pmatrix} \varphi \\ \pi \end{pmatrix} = L\, \phi = \begin{bmatrix} 0 & 1 \\ (\partial_x^2 - \mu^2) & 0 \end{bmatrix} \begin{pmatrix} \varphi \\ \pi \end{pmatrix} = \begin{bmatrix} \pi \\ (\partial_x^2 - \mu^2)\varphi \end{bmatrix}. \tag{19}$$

Here, $\pi = \partial_t \varphi$ is the canonical momentum field of the KG field φ, which can be discriminated by context from the number π. The dagger denotes transposing and complex conjugation of functional vectors so $\varphi^\dagger j = \int dx\, \bar{\varphi}_x j_x = \int dk\, \bar{\varphi}_k j_k / (2\pi)$ in real and Fourier space, respectively. The scalar product of two component fields $\phi = (\phi^{(\varphi)\dagger}, \phi^{(\pi)\dagger})^\dagger$ and $\psi = (\psi^{(\varphi)\dagger}, \psi^{(\pi)\dagger})^\dagger$ is

$$\phi^\dagger \psi = \int dx\, \left(\overline{\phi_x^{(\varphi)}} \psi_x^{(\varphi)} + \overline{\phi_x^{(\pi)}} \psi_x^{(\pi)} \right),$$
$$= \int \frac{dk}{2\pi} \left(\overline{\phi_k^{(\varphi)}} \psi_k^{(\varphi)} + \overline{\phi_k^{(\pi)}} \psi_k^{(\pi)} \right), \tag{20}$$

in real and Fourier space, respectively.

The KG field dynamics can be derived from the quadratic Hamiltonian of the dynamical system,

$$\mathcal{H}(\phi) = \frac{1}{2} \phi^\dagger E\, \phi = \int dx\, \frac{1}{2} \left[\pi_x^2 + (\partial_x \varphi_x)^2 + \mu^2 \varphi_x^2 \right]$$
$$= \int \frac{dk}{4\pi} [|\pi_k|^2 + (\mu^2 + k^2)|\varphi_k|^2] \tag{21}$$

in abstract, position space, and Fourier-space notation, respectively. Here and in the following, x and y are coordinates in position space, k and q coordinates in continuous or discrete Fourier space, t is a time coordinate, and coordinate labels determine in which functional basis a component of a field is to be read out. The kernel E of the Hamiltonian reads, in the Fourier basis,

$$E_{kq} = 2\pi \delta(k - q) \begin{pmatrix} \mu^2 + k^2 & 0 \\ 0 & 1 \end{pmatrix}. \tag{22}$$

This determines the KG dynamics via

$$\partial_t \phi = S\, \partial_\phi \mathcal{H}(\phi) = S\, E\, \phi, \tag{23}$$

with the symplectic matrix

$$S = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}. \tag{24}$$

Therefore, the linear time evolution operator is $L = S\, E$ and the temporal source is $c = 0$ in our example.

The Fourier-space representation of the KG dynamics, $(\partial_t^2 + k^2 + \mu^2)\varphi_k = 0$, has the solution

$$\varphi_k = a_k e^{i\omega t} + \overline{a_{-k}} e^{-i\omega t}$$
$$\pi_k = i\omega(a_k e^{i\omega t} - \overline{a_{-k}} e^{-i\omega t}) \tag{25}$$

with $\omega = \sqrt{k^2 + \mu^2}$, $i = \sqrt{-1}$, and $a_k \in \mathbb{C}$. With respect to the remaining degrees of freedom, the complex amplitudes a_k,

the Hamiltonian becomes

$$\mathcal{H}(a) = \int_0^\infty \frac{dk}{\pi} |a_k|^2 (k^2 + \mu^2), \quad (26)$$

which implies that these variables are stationary, $\partial_t a_k = 0$. Therefore, an exact high-resolution solution can be specified for the KG example for all times. This will be compared to approximative low-resolution solutions provided by simulation schemes derived from IFD and by the usual discretization of differential operators as described in Appendix A 1. ◄

B. Prior knowledge

The signal field prior $\mathcal{P}(\phi)$ has to be specified. The prior should summarize the data-independent knowledge on the field configuration at current time t. For practical reasons, one will typically approximate it by a Gaussian,

$$\mathcal{P}(\phi) = \mathcal{G}(\phi - \psi, \Phi), \quad (27)$$

with a properly chosen mean field $\psi = \langle \phi \rangle_{(\phi)}$ and prior uncertainty variance $\Phi = \langle (\phi - \psi)(\phi - \psi)^\dagger \rangle_{(\phi)}$. Such an approximation is often possible, since even non-Gaussian knowledge states are typically sufficiently well approximated by Gaussians. Any sophisticated treatment of the otherwise resulting nonlinear, interacting IFT is beyond the scope of this paper.

The Gaussian prior can also be justified from a pure information-theoretical point of view. In case only the prior mean ψ and variance Φ are known from physical considerations, the MEP distribution of the field ϕ representing exactly this knowledge is given by the Gaussian (27) with this mean and variance, as shown in Appendix B.

Any known mean field ψ can easily be absorbed by the redefinitions $\phi \to \phi' = \phi - \psi$ and $c \to c' = c + L\psi$. This, however, might create a c term even if none existed initially in the dynamical equation. Therefore, we keep the possibility of a prior mean in the formalism, but note that there is some freedom to trade a prior mean ψ against a field-independent c term and vice versa.

► For our illustrative example of a KG field, we assume that the field was initially in contact and equilibrium with a thermal reservoir at temperature β^{-1} and became decoupled from it at some time $t_0 = 0$. The initial probability function of the field is, therefore, thermal,

$$\mathcal{P}(\phi|\beta) = \frac{1}{Z_\beta} e^{-\beta \mathcal{H}(\phi)} = \prod_k \frac{1}{z_k} e^{-2\beta |a_k|^2 (k^2 + \mu^2)}. \quad (28)$$

It separates into independently excited modes, which do not exchange energy at later times because the amplitudes are stationary. Thus, an initially established thermal state stays thermal and at the same temperature for all times. The partition function is given by a complex Gaussian integral for each mode and is

$$Z_\beta \equiv \int \mathcal{D}\phi \, e^{-\beta \mathcal{H}(\phi)} = \prod_k \underbrace{\frac{\pi}{2\beta (k^2 + \mu^2)}}_{z_k}, \quad (29)$$

where the product goes over all accessible positive wave vectors.

Since the energy Hamiltonian $\mathcal{H}(\phi) = \frac{1}{2}\phi^\dagger E \phi$ is quadratic in ϕ, the prior information Hamiltonian $H(\phi|\beta) = \beta \mathcal{H}(\phi) = \frac{\beta}{2}\phi^\dagger E \phi$ is quadratic as well. The prior is simply a Gaussian $\mathcal{P}(\phi|\beta) = \mathcal{G}(\phi, \Phi)$ with zero mean $\psi = 0$ and covariance $\Phi = (\beta E)^{-1}$. In Fourier space this reads

$$\Phi_{kq} = \frac{2\pi}{\beta} \delta(k - q) \begin{bmatrix} (\mu^2 + k^2)^{-1} & 0 \\ 0 & 1 \end{bmatrix} \quad (30)$$

and in position space it is

$$\Phi_{xy} = \frac{1}{\beta} \begin{bmatrix} \frac{1}{2\mu} e^{-\mu |x-y|} & 0 \\ 0 & \delta(x - y) \end{bmatrix}. \quad (31)$$

A KG field realization drawn from (28) for $\beta = 1$ and $\mu = 1$ is displayed in Fig. 1. There the different spatial correlation structures of the field values with $\langle \varphi_x \varphi_y \rangle_{(\phi)} = (2\mu\beta)^{-1} e^{-\mu |x-y|}$ and field momenta with $\langle \pi_x \pi_y \rangle_{(\phi)} = \beta^{-1}\delta(x - y)$, as given by (31), can be seen. ◄
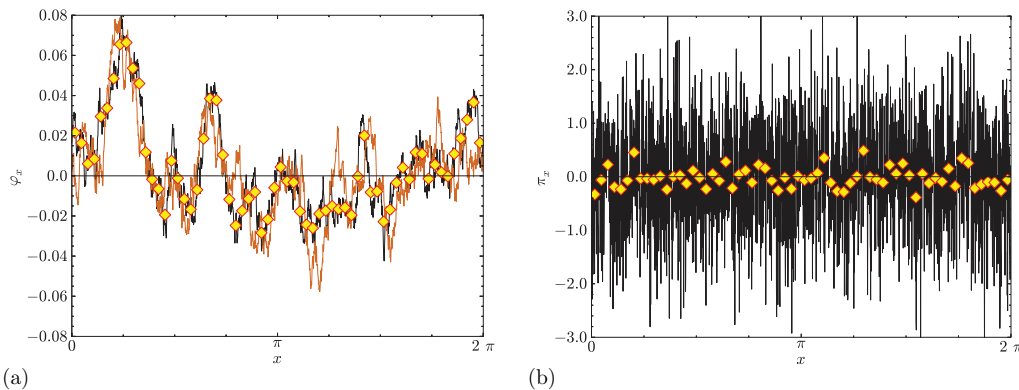

FIG. 1. (Color online) A realization of a thermally exited KG field φ_x (a) and its momentum distribution π_x (b) is shown for $\beta = 1$ and $\mu = 1$ at $t = 0$ with a resolution of 2048 pixels with black lines passing through the diamond symbols. The low-resolution data with $\mathcal{N} = 64$ data points describing the same fields are shown with yellow diamonds. The field configuration at $t = 0.1$ is also shown in panel (a) with a thin brown (gray) line. The KG field φ_x shows a correlated structure due to the suppression of small scale power by the gradient term in the Hamiltonian, whereas its momentum field π_x is just white noise. The loss of small-scale structure information in the low-resolution sampling is especially apparent for the momentum data.

C. Data constraints

In addition to the relatively vague prior knowledge, the field is constrained by the finite-dimensional data vector $d = (d_i)_i$ in computer memory. The data are assumed to represent linear statements on the field of the form $d = Rs + n$, cf. Eq. (6). In typical numerical simulation schemes, the response operator might just express an averaging of the field within some environment Ω_i of a grid point $x_i \in \Omega_i$, i.e.,

$$R_{ix} = \frac{1}{|\Omega_i|} \theta(x \in \Omega_i), \qquad (32)$$

where the logical θ function

$$\theta(x \in \Omega_i) = \mathcal{P}(x \in \Omega_i | x, \Omega_i) = \begin{cases} 1 & x \in \Omega_i \\ 0 & x \notin \Omega_i \end{cases} \qquad (33)$$

is 1 if the condition in its argument is true, otherwise it is zero. In schemes based on grid cells or space tessellations, the grid point volumes are disjoint, $\Omega_i \cap \Omega_j = \emptyset$ for $i \neq j$. In case a conserved quantity should be conserved as accurately as possible, the total amount of the quantity within the cells of a space tessellation as well as the currents of the quantity through the surfaces of the tessellation cells might be used as data. In smoothed particle hydrodynamics, the volumes overlap and are usually also structured by radially declining kernel functions that have evolving locations and sizes.

For the moment, we only have to deal with the data at one instant and need only to know that it depends linearly on the underlying field by a known relation of the form $d = R\phi + n$. This relation might or might not be the same at the next instant, depending on the design choices for $R = R_t$ (stationary grid or Lagrangian moving mesh). R_t could even be determined by the IFD formalism itself by requiring minimal information loss of the scheme, as we will do later for the KG field example in Sec. III F.

The simulation data vector d can even be extended also to contain measurement data on the system to be simulated (e.g., the weather) obtained for the current simulation time. If this auxiliary data $\mathfrak{d}$ resulted from a linear measurement $\mathfrak{d} = \mathfrak{R}\phi + \mathfrak{n}$ with response $\mathfrak{R}$ and Gaussian noise $\mathfrak{n}$ with covariance $\mathfrak{N}$, only the replacements

$$d \to \begin{pmatrix} d \\ \mathfrak{d} \end{pmatrix}, \quad R \to \begin{pmatrix} R \\ \mathfrak{R} \end{pmatrix}, \quad \text{and} \quad N \to \begin{pmatrix} N & 0 \\ 0 & \mathfrak{N} \end{pmatrix} \qquad (34)$$

are needed.[4] This way, the measurement information is assimilated into the simulation scheme and can be evolved into the future (or into the past, if the simulation is backward in time). The added data could become simulation degrees of freedom, or they could be discarded at the next simulation time

---

[4]The block diagonal structure of the extended noise covariance matrix assumes that the measurement error and the simulation error are uncorrelated. This assumption would be improper in case repeated measurements with the same incorrectly calibrated instrument are assimilated into the simulation. In that case, correlations among the simulation and measurement data errors could exist since the correlated measurement errors are partly imprinted onto the simulation data.

step after their information was transferred to the simulation data via the entropic matching operation. The former option would certainly conserve more information, and the latter is similar to what is done in particle filter methods as described in Appendix A 3.

The ensemble of field configurations constrained by the data via (6) and by the prior via (27) is then

$$\mathcal{P}(\phi | d) = \mathcal{G}(\phi - m, D), \qquad (35)$$

where

$$D = (\Phi^{-1} + R^\dagger N^{-1} R)^{-1}$$

and

$$m = \psi + W(d - R\psi) = D(R^\dagger N^{-1} d + \Phi^{-1}\psi). \qquad (36)$$

The mean is shifted here with respect to (9) due to the nonvanishing prior mean ψ.

In the case where the external data $\mathfrak{d}$ are to be assimilated into the simulation, applying replacements of (34) to (36) and expanding this yields $D = (\Phi^{-1} + R^\dagger N^{-1} R + \mathfrak{R}^\dagger \mathfrak{N}^{-1} \mathfrak{R})^{-1}$ and $d = D(R^\dagger N^{-1} d + \mathfrak{R}^\dagger \mathfrak{N}^{-1} \mathfrak{d} + \Phi^{-1}\psi)$. Thus, data assimilation is very naturally done in IFD since simulation and measurement data shape the field posterior $\mathcal{P}(\phi | d) = \mathcal{G}(\phi - m, D)$ in a similar way.

▶ In our example of the KG field we want to deal with the simplest possible data as given by (6) and (32) that lives on a regular grid, with equidistant space filling and disjoint pixel volumes $\Omega_i = [i\Delta, (i+1)\Delta)$, with $\Delta > 0$ being the grid spacings. Since on a computer one can deal only with finite domains, we assume periodic boundary conditions for the interval $\Omega = \cup_i \Omega_i = [0, 2\pi]$ and require that the number of grid points $\mathcal{N} = 2\pi/\Delta \in \mathbb{N}$. The Fourier transformed field is then

$$\phi_k = \int_0^{2\pi} dx \, e^{ikx} \phi_x, \qquad (37)$$

with

$$\phi_x = \sum_{k=-\infty}^{\infty} \frac{1}{2\pi} e^{-ikx} \phi_k. \qquad (38)$$

Here the following substitution with respect to the infinitely extended case have been made: $\int dx \to \int_0^{2\pi} dx$ and $\int \frac{dk}{2\pi} \to \sum_{k=-\infty}^{\infty} \frac{1}{2\pi}$, which are the appropriately weighted sums of the scalar products in position and Fourier space, respectively. Furthermore, we note that $\delta(k - q) \to \delta_{kq}$ in this case, so the unit operator is $\mathbb{1}_{kq} = 2\pi \delta_{kq}$ and the field covariance (30) reads

$$\Phi_{kq} = \frac{2\pi}{\beta} \delta_{kq} \begin{bmatrix} (\mu^2 + k^2)^{-1} & 0 \\ 0 & 1 \end{bmatrix}. \qquad (39)$$

Since the data space is finite, its Fourier space is also finite, where

$$d_k = \sum_{j=0}^{\mathcal{N}-1} \Delta \, e^{ikj\Delta} d_j, \qquad (40)$$

with

$$d_j = \sum_{k=0}^{\mathcal{N}-1} \frac{1}{2\pi} e^{-ikj\Delta} d_k, \qquad (41)$$

and $k \in \{0, \ldots \mathcal{N} - 1\}$. Higher or negative Fourier modes do not carry any additional information due to the Nyquist theorem.[5]

The Fourier transformed response,

$$R_{kq} = 2\pi \, \theta(q - k \in \mathcal{N} \, \mathbb{Z}) \frac{1 - e^{-iq\Delta}}{iq\Delta}, \tag{42}$$

$$= 2\pi \, \theta(q - k \in \mathcal{N} \, \mathbb{Z}) e^{-\frac{1}{2}iq\Delta} \mathrm{sinc}\left(\frac{1}{2}q\Delta\right), \tag{43}$$

is block diagonal in the reduced Fourier space of the data with $k \in \{0, \ldots \mathcal{N} - 1\}$. Note, however, that higher Fourier modes of the field ϕ_q with $q \in k + \mathcal{N} \, \mathbb{Z}$, which carry information on subgrid structure, imprint also onto the data and blend with the lower Fourier modes $k \in \{0, \ldots \mathcal{N} - 1\}$. Therefore, a unique reconstruction of the individual Fourier modes from the data alone is impossible even within the range $q \in \{0, \ldots \mathcal{N} - 1\}$.

The individual terms in (42) can easily be understood. The $\exp(-\frac{1}{2}iq\Delta)$ term stems from the fact that the centers of the pixel volumes are shifted by $\frac{1}{2}\Delta$ from the pixel positions $j\Delta$ used in the definition of the Fourier transformation. The sinc function is the Fourier-space transform of the pixel window. It encodes how well a given Fourier mode is represented in the data and, therefore, how well it is protected from noise and confusion with other modes imprinted onto the same data mode.

The data-space signal covariance, which is needed by the Wiener filter, is[6]

$$\tilde{\Phi}_{kq} = (R \, \Phi \, R^\dagger)_{kq} = \begin{pmatrix} \tilde{\Phi}_{kq}^{(\varphi)} & 0 \\ 0 & \tilde{\Phi}_{kq}^{(\pi)} \end{pmatrix}, \tag{44}$$

with

$$\tilde{\Phi}_{kq}^{(\varphi)} = \frac{\tilde{\Phi}_{kq}^{(\pi)}}{\mu^2} \begin{cases} 1 & k = 0 \\ \left[1 - \frac{2}{\mu\Delta} \frac{\sinh(\mu\Delta)\sin^2(\frac{1}{2}k\Delta)}{\cosh(\mu\Delta) - \cos(k\Delta)}\right] & k \neq 0 \end{cases},$$

$$\tilde{\Phi}_{kq}^{(\pi)} = \frac{2\pi \, \delta_{kq}}{\beta} \begin{cases} 1 & k = 0 \\ \frac{1 - \cos(k\Delta)}{2\sin^2(\frac{1}{2}k\Delta)} & k \neq 0 \end{cases}.$$

Since the field covariance and response are translationally invariant, we have every reason to believe that the noise statistics, which are fed only by approximation errors depending on

_____

[5]These conventions for the discrete Fourier transformation might appear a bit unusual, but they have the advantage that they match best the continuous-space Fourier convention used in physics. They permit us to use all derived Fourier-space equations for the KG field without changing normalization constants and with the intuitive identifications $dx \to \Delta$, $x \to j\Delta$, and $k \to k$.

[6]Here, we used the following identities:

$$\sum_{j \in \mathbb{Z}} \frac{1}{(a + j)^2} = \frac{\pi^2}{\sin^2(\pi a)}$$

and

$$\sum_{j \in \mathbb{Z}} \frac{1}{(a + j)^2[(a + j)^2 + b^2]}$$

$$= \frac{\pi}{b^3} \left[\frac{b\pi}{\sin^2(\pi a)} - \frac{\sinh(2\pi b)}{\cosh(2\pi b) - \cos(2\pi a)}\right].$$

these latter two quantities, will also be translationally invariant in data space. Therefore, its covariance will also be diagonal in discrete Fourier space:

$$N_{kq} = 2\pi \, \delta_{kq} \begin{pmatrix} \eta_k^{(\varphi)} & \eta_k^{(c)} \\ \eta_k^{(c)} & \eta_k^{(\pi)} \end{pmatrix}, \tag{45}$$

where $\eta^{(\varphi)}$, $\eta^{(\pi)}$, and $\eta^{(c)}$ are the noise spectra of the field value data, the field momenta data, and the cross-spectra of those, respectively. However, in Sec. III F we will show that the ideal IFD scheme stays noiseless if it was initially noiseless. Therefore, we can set $N \to 0$ for all times and use the η parameters to ensure consistency of all formulas. They will be set to zero at the end of the calculation if this is a permitted limit.

Taking the noiseless case as granted for the moment, the Wiener filter becomes

$$W_{kq} = (\Phi \, R^\dagger \tilde{\Phi}^{-1})_{kq}$$

$$= 2\pi\theta(q = k \bmod \mathcal{N}) \, e^{\frac{1}{2}ik\Delta} \mathrm{sinc}\left(\frac{1}{2}k\Delta\right) \frac{2\sin^2\left(\frac{1}{2}q\Delta\right)}{1 - \cos(q\Delta)}$$

$$\times \begin{pmatrix} \frac{\mu^2}{\mu^2 + k^2} \left[1 - \frac{2}{\mu\Delta} \frac{\sinh(\mu\Delta)\sin^2(\frac{1}{2}k\Delta)}{\cosh(\mu\Delta) - \cos(k\Delta)}\right]^{-1} & 0 \\ 0 & 1 \end{pmatrix}. \tag{46}$$

For a reconstructed signal image generated by this Wiener filter, any image Fourier mode $k \in \mathbb{Z}$ gets exited by its first Brillouin zone data-space mode $q = k \bmod \mathcal{N} \in \{0, \ldots \mathcal{N} - 1\}$. Thereby, all Fourier modes $k \in \mathbb{Z}$ of the mean field $m = W \, d$ get some nontrivial value if the corresponding data mode $k \bmod \mathcal{N}$ was nonzero. ◄

D. Field evolution

A Gaussian knowledge state $\mathcal{P}(\phi|t) = \mathcal{P}[\phi \, | \, d = d(t)] = \mathcal{G}(\phi - m, D)$ at some initial time t is represented by the data $d = d_t$, which determines the mean field via $m = W \, d$. The field uncertainty variance D is data independent in our example but not in general. The knowledge state $\mathcal{P}(\phi|t)$ has to be evolved to an infinitesimally later time $t' = t + \delta t$ via the evolution of the individual field configurations.

An individual field configuration $\phi = \phi_t$ at initial time t becomes $\phi' = \phi_{t'} \widehat{=} \phi_t + \delta t \, \dot{\phi}_t = \phi_t + \delta t \, (L \, \phi_t + c)$, where the time derivative is given by (15). Here, and in the following, we drop nonessential terms of $\mathcal{O}(\delta t^2)$, as indicated by "$\widehat{=}$". The time-evolved knowledge state therefore becomes

$$\mathcal{P}(\phi'|\, d) = \mathcal{P}(\phi|\, d) \left|\frac{\partial\phi}{\partial\phi'}\right| \tag{47}$$

by conservation of probability density. We need to calculate the Jacobian up to linear order in δt. This is most simply done from the inverse Jacobian,

$$\left|\frac{\partial\phi'}{\partial\phi}\right| = |\mathbb{1} + \delta t \, L| = \exp\log|\mathbb{1} + \delta t \, L|$$

$$\widehat{=} \exp\mathrm{Tr}(\delta t \, L) \widehat{=} 1 + \delta t \, \mathrm{Tr}(L). \tag{48}$$

In the case of linear Hamiltonian dynamics $\partial_t \phi = S \partial_\phi \mathcal{H}(\phi)$, with dynamical Hamiltonian of the from $\mathcal{H}(\phi) = \frac{1}{2}\phi^\dagger E \phi + b^\dagger \phi$ and E being block diagonal in the field value φ and field momentum π eigenspaces, we have $L = S E$ and $c = S b$. The Jacobian is then unity, since

$$\mathrm{Tr}(L) = \mathrm{Tr}(S\,E) = \mathrm{Tr}\left[\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}\begin{pmatrix} E^{(\phi)} & 0 \\ 0 & E^{(\pi)} \end{pmatrix}\right]$$
$$= \mathrm{Tr}\begin{pmatrix} 0 & -E^{(\pi)} \\ E^{(\phi)} & 0 \end{pmatrix} = 0. \tag{49}$$

This is not surprising, since it is well known that symplectic Hamiltonian systems conserve the phase-space density, so the unity of the Jacobian is also valid for noninfinitesimal time steps δt in such cases.

In general, for non-Hamiltonian systems, the Jacobian can differ from 1. It can be larger for systems with dynamical attractors or with dissipation (Navier-Stokes equations) and it can be smaller for systems with diverging phase-space flows, like chaotic inflation in cosmology or driven hydrodynamical turbulence (without significant dissipation).

The evolved knowledge state, or the knowledge state on the evolved field, is, therefore,

$$\mathcal{P}(\phi'\,|\,d) \,\widehat{=}\, \mathcal{P}(\phi = \phi' - \delta t\,\dot{\phi}\,|\,d)\,|\partial\phi/\partial\phi'|$$
$$\widehat{=}\, \mathcal{G}[\phi' - \delta t\,(L\phi' + c) - m,\, D]\,[1 - \delta t\,\mathrm{Tr}\,(L)]$$
$$\widehat{=}\, \mathcal{G}(\phi' - m^*,\, D^*), \tag{50}$$

with[7]

$$m^* \,\widehat{=}\, m + \delta t(c + L\,m)$$
$$\widehat{=}\, (1 + \delta t\,L)\,[\psi + W\,(d - R\,\psi)] + \delta t\,c,$$
$$D^* \,\widehat{=}\, D + \delta t\,(L\,D + D\,L^\dagger).$$
$$D^{*-1} \,\widehat{=}\, D^{-1} - \delta t\,(D^{-1}L + L^\dagger D^{-1}). \tag{51}$$

▶ In the case of our KG field, we have $\mathrm{Tr}(L) = 0$ due to the symplectic dynamics with $L = S E$ and $c = 0$, as well as $m^* \,\widehat{=}\, m + \delta t\,S E\,m$. Furthermore, using $L = S E$, $S^\dagger = -S$, $D^{-1} = \Phi^{-1} + R^\dagger N^{-1}R$, and $\Phi^{-1} = \beta E$, we get $D^{*-1} \,\widehat{=}\, D^{-1} - \delta t\,(R^\dagger N^{-1} R\,S E - E\,S R^\dagger N^{-1} R)$. ◀

————

[7]The key to understand this result is a short rearrangement in the exponent of the Gaussian,

$$[(1 - \delta t\,L)\phi' - m - \delta t\,c]^\dagger D^{-1}[(1 - \delta t\,L)\phi' - m - \delta t\,c]$$
$$= \{\phi' - \underbrace{[(1 - \delta t\,L)^{-1}m + \delta t\,c]}_{m^*}\}^\dagger$$
$$\times \underbrace{(1 - \delta t\,L)^\dagger D^{-1}(1 - \delta t\,L)}_{D^{*-1}}(\phi' - m'),$$

the δt expansion of the new mean field,

$$m^* \,\widehat{=}\, (1 + \delta t\,L)\,m + \delta t\,c,$$

that of the new uncertainty variance,

$$D^* \,\widehat{=}\, (1 + \delta t\,L)\,D\,(1 + \delta t\,L)^\dagger$$
$$\widehat{=}\, D + \delta t\,(L\,D + D\,L^\dagger),$$

and its determinant,

$$|D^*| \,\widehat{=}\, [1 + 2\delta t\,\mathrm{Tr}(L)]\,|D|.$$

The evolved mean field still can be regarded to be parametrized by the data; however, this is in a different way, $m^* = (1 + \delta t\,S E)\,W\,d$. It is not clear in general whether a new data set d' can be found that expresses this new mean field via the original parametrization $m' = W\,d'$ (or with the appropriate W', in case that D' also changed). This is because the functional forms of the two parametrizations differ since W and $L = S E$ operate on completely different vector spaces, the discrete data space and the continuous field space, respectively.

Therefore, entropic matching will be used to choose a d' that determines $\mathcal{P}'(\phi'\,|\,d')$ such that it captures most of the information content of $\mathcal{P}(\phi'\,|\,d)$. ◀

E. Prior update

The field prior for time t' has to be updated since the subgrid statistics might have changed. For example, some of the energy contained in subgrid modes might dissipate, leading to a different $\mathcal{P}(\phi') = \mathcal{G}(\phi' - \psi',\Phi')$ as parametrized via the updated prior mean ψ' and variance Φ'.

▶ In the case of our KG field, energy conservation of the dynamics leads to an unchanged prior for the evolved field $\mathcal{P}(\phi') = \mathcal{G}(\phi',\Phi)$, still with $\Phi = (\beta\,E)^{-1}$. ◀

F. Data update

The new data have to be determined from their relation to the updated field. Again, we assume the new data to depend linearly on the evolved field,

$$d' = R'\phi' + n'.$$

Note that we could chose a different pixilation at t', leading to a different response R', propagator D', and Wiener filter W'. This is needed, e.g., in case a simulation with moving or adaptive mesh is to be developed. It can even be considered that the response operator determination becomes a part of the entropic matching step, leading to an information optimal moving mesh.

Furthermore, we have to allow for a changed noise level, with new covariance N', since the meaning of the data values could have changed with changed pixilation and since we might have to allow for additional uncertainty in order to capture any mismatch between the new parametrized posterior and the evolved field posterior.

According to (35) and (36), the relation of new posterior and new data is

$$\mathcal{P}(\phi'|d') = \mathcal{G}(\phi' - m',\, D'), \tag{52}$$

where

$$D' = (\Phi'^{-1} + R'^\dagger N'^{-1} R')^{-1},$$
$$m' = \psi' + W'\,(d' - R'\,\psi')$$
$$= D'\,(R'^\dagger N'^{-1} d' + \Phi'^{-1}\psi'), \quad \text{and}$$
$$W' = D'\,R'^\dagger N'^{-1} = \Phi'R'^\dagger(\underbrace{R'\,\Phi'\,R'^\dagger}_{\tilde{\Phi}'} + N')^{-1}. \tag{53}$$

Now, the new posterior $\mathcal{P}' = \mathcal{P}(\phi'|d')$ should match the evolved posterior $\mathcal{P} = \mathcal{P}(\phi'|d)$ as well as possible. According

to (13), the cross entropy of the former with the latter is

$$\mathcal{S}(\mathcal{P}'|\mathcal{P}) = -\tfrac{1}{2}\text{Tr}[(\delta m\,\delta m^\dagger + D')\,D^{*-1}$$
$$+ \mathbb{1} + \log(D'\,D^{*-1})] \tag{54}$$

with $\delta m = m' - m^*$.

Maximizing this entropy with respect to the new data d' yields

$$-\partial_{d'}\mathcal{S} = (\partial_{d'}m')^\dagger D^{*-1}\delta m$$
$$= W'^\dagger D^{*-1}[W'(d' - R'\,\psi') + \psi' - m^*] = 0 \tag{55}$$
$$\Rightarrow d' = R'\,\psi' + (W'^\dagger D^{*-1}W')^{-1}W'^\dagger D^{*-1}(m^* - \psi').$$

This is the general formula to update the data. It should be expanded up to linear order in all the relevant changes in response $R' = R + \delta R$, noise covariance $N' = N + \delta N$, and prior parameters $\Phi' = \Phi + \delta\Phi$ and $\psi' = \psi + \delta\psi$, as well as in time $t' = t + \delta t$. The resulting general formula is lengthy and not directly instructive,[8] therefore, we concentrate here more on special cases.

The update of the uncertainty variance is also obtained by maximizing the entropy with respect to the degrees of freedom of $D' = (\Phi' + R'^\dagger N'^{-1}R')^{-1}$. These could be the location of the new pixel positions, which influence R', an updated noise level, influencing N', or properties of the field prior expressed via Φ' and ψ'.

We combine these degrees of freedom into the single vector η, irrespective of whether they determine R', N', Φ', ψ', or combinations thereof. The entropic matching of the updated uncertainty variance $D' = D(\eta + \delta\eta) = D(\eta) + \sum_i \delta\eta_i\,\Gamma_i + \mathcal{O}(\delta\eta^2)$, with $\Gamma_i = \partial_{\eta_i}D(\eta)$ the linear changes due to changes in the degrees of freedom, is then given by

$$-\partial_\eta\mathcal{S} = \tfrac{1}{2}\text{Tr}[(\partial_\eta D')(D^{*-1} - D'^{-1})] = 0$$

$$\Rightarrow \delta\eta = C^{-1}b,$$

with

$$b_i = \text{Tr}[\Gamma_i\,(D^{*-1} - D^{-1})]$$

and

$$C_{ij} = \text{Tr}[\Gamma_i\,D^{-1}\,\Gamma_j\,D^{-1}]. \tag{56}$$

---

[8] A few useful identities, when dealing with (55) might be in order. A short calculation shows that, up to linear order in δt,

$$(W'^\dagger D^{*-1}W')^{-1}W'^\dagger$$
$$= (\tilde\Phi' + N')(R'\Phi'D^{*-1}\Phi'R'^\dagger)^{-1}R'\Phi'$$
$$\cong (\tilde\Phi' + N')[R'\Phi'(D^{-1} - \delta t\,(D^{-1}L + L^\dagger D^{-1}))\Phi'R'^\dagger]^{-1}R'\Phi'$$
$$\cong (\tilde\Phi' + N')[\tilde D + \delta t\,\tilde D\,R'\Phi'(D^{-1}L + L^\dagger D^{-1})\Phi'R'^\dagger\,\tilde D\,]\,R'\Phi',$$

with $\tilde D = (R'\Phi'D^{-1}\Phi'R'^\dagger)^{-1}$ and that

$$D^{*-1}(m^* - \psi')$$
$$\cong [D^{-1} - \delta t\,(D^{-1}L + L^\dagger D^{-1})]\,[(1 + \delta t\,L)\,.$$
$$\times (\psi + W\,(d - R\,\psi)] + \delta t\,c$$
$$\cong D^{-1}\psi + R^\dagger N^{-1}(d - R\psi)$$
$$+ \delta t[D^{-1}c - L^\dagger(D^{-1}\psi + R^\dagger N^{-1}(d - R\,\psi))].$$

From the first line it is already apparent that if D' is able to match D^* exactly, then it will do so. The detailed formula for updating response, noise, and prior can be complex, since operator inversions are involved. In general, approximations might be necessary here in order to proceed with a reasonable computational complexity.

The formulas (55) and (56) form the desired simulation scheme. The scheme deals optimally with time-dependent pixelation, non-Hamiltonian dynamics, subgrid processes, as well as with the accumulation of discretization errors. The price of this generality is a higher complexity of the detailed formula compared to many *ad hoc* schemes. These formulas have to be analyzed case by case to identify the optimal numerical implementation strategy. In order to show this in a simple example, we turn again to the KG field.

▶ Assuming that we have all freedom to chose R', N', and Φ' to match $D'^{-1} = \Phi'^{-1} + R'^\dagger N'^{-1}R'$ exactly with

$$D^{*-1} \cong \Phi^{-1} + R^\dagger N^{-1}R - \delta t\,(R^\dagger N^{-1}RSE - ESR^\dagger N^{-1}R)$$

as derived in Sec. III D, we would immediately use $\Phi' = \Phi$ and try to accommodate the change in variance in a changed response or noise. Thus, the unchanged signal covariance also results from the data update via the MEP. The considerations to update the prior in Sec. III E were therefore superfluous in this case. The updated prior mean ψ' could also be derived by maximizing the entropy with respect to it. It is not surprising that it turns out to be $\psi' = \psi = 0$.

Writing $R' = R + \delta R$ and $N' = N + \delta N$, we find

$$D'^{-1} \cong \Phi^{-1} + R^\dagger N^{-1}R + \delta R^\dagger N^{-1}R$$
$$+ R^\dagger N^{-1}\delta R - R^\dagger N^{-1}\delta N\,N^{-1}R. \tag{58}$$

Comparing the terms of the last two equations, we conclude that the best match is found by the identification

$$\delta R = -\delta t\,R\,S\,E, \quad \delta N = 0. \tag{59}$$

Thus, the noise should stay unchanged and can be assumed to be zero for all times it was zero initially, which we will assume in the following. The response of an optimal scheme should, however, evolve according to $\partial_t R_t = -R_t\,S\,E$. This can actually be solved analytically, providing

$$R_t = R\,T_{-t}, \tag{60}$$

with the time translation operator

$$(T_t)_{kq} = (e^{SEt})_{kq}$$
$$= \mathbb{1}_{kq}\left[\cos(\omega_k t)\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \sin(\omega_k t)\begin{pmatrix} 0 & \omega_k^{-1} \\ -\omega_k & 0 \end{pmatrix}\right]. \tag{61}$$

In case we insist on using the original response R for all later times, the change in the uncertainty variance D^* would have been needed to be captured by either Φ' or by N'. Neither is optimal for this, which is why the resulting schemes would lose information in the course of the simulation. As we will see in Sec. III G, our scheme with evolving response is lossless with respect to information.

For the data update from $d = d_t$ to $d' = d_{t'}$ at $t' = t + \delta t$ we need only to expand (55) to first order in δt.

In our ideal case with $N \to 0$, we have $W' = \Phi R_{t'}^\dagger$ $(R_{t'} \Phi R_{t'}^\dagger)^{-1} = \Phi R_{t'}^\dagger (R \Phi R^\dagger)^{-1} = \Phi R_{t'}^\dagger \tilde{\Phi}^{-1} = W - \Phi (R_t - R_{t'})^\dagger \tilde{\Phi}^{-1} \,\hat{=}\, W - \delta t\, \Phi\, L^\dagger R_t^\dagger \tilde{\Phi}^{-1}$, as a short calculation verifies. The data evolution is then

$$
\begin{aligned}
d' &= (W'^\dagger D^{*-1} W')^{-1} W'^\dagger D^{*-1} m^* \\
&\,\hat{=}\, (W'^\dagger D^{*-1} W')^{-1} W'^\dagger D^{*-1} (1 + \delta t\, L)\, W\, d \\
&\,\hat{=}\, (W'^\dagger D^{*-1} W')^{-1} \\
&\quad \times [W'^\dagger D^{*-1} W' + W'^\dagger D^{*-1} (W - W' + \delta t\, L\, W)]\, d \\
&\,\hat{=}\, d + (W'^\dagger D^{*-1} W')^{-1} W'^\dagger D^{*-1} \delta t \\
&\quad \times (\underbrace{\Phi\, L^\dagger + L\, \Phi}_{0})\, R_t^\dagger \tilde{\Phi}^{-1} d \\
&= d, \tag{62}
\end{aligned}
$$

since $\Phi L^\dagger = \beta^{-1} E^{-1} E\, S^\dagger = -S\, E\, \beta^{-1} E^{-1} = -L\, \Phi$. Thus, $\partial_t d_t = 0$, the data should not be changed, and the evolution is completely captured by the response evolution. This scheme is optimal from an IFT point of view, as we will see in the following. Note that the scheme is completely specified in the data space, since (62) does not require any subgrid calculations, as it does not require any calculations at all. It will be shown in the next section that the evolution of binned field values is also completely specified in data space and that the subgrid field configuration predictions require the usage of a finer grid only at the very end of the calculation.

This simple data (non-)evolution equation $\partial_t d_t = 0$ is a consequence of our KG example having a linear symplectic evolution, as determined by $\mathcal{H}(\phi) = \frac{1}{2} \phi^\dagger E\, \phi$ and a thermal prior distribution, as characterized by $\tilde{H}(\phi|\beta) = \beta \mathcal{H}(\phi)$, both depending on the same energy matrix E. In general, $\partial_t d_t \neq 0$ can be expected as soon as prior and dynamics are more orthogonal in their eigenvector sets. ◄

G. Information field-theoretical solution

► The KG problem is exactly solvable and the later time field can be obtained from applying a time translation operator, as given by (61), to an earlier time field. This operator depends only on the time difference, $\phi_{t'} = T_{t'-t} \phi_t$, and is even invertible, so the earlier field can be calculated from the later one. With this, the time invariance of the field covariance can easily be verified,

$$
\Phi_t = \langle \phi_t \phi_t^\dagger \rangle_{(\phi)} = T_t \Phi_{t=0} T_t^\dagger = \Phi_0 \equiv \Phi, \tag{63}
$$

where the last identity requires a few lines of straightforward matrix multiplications using (30) and (61).

Since we want to infer the future field ϕ_t from the initial data $d = d_{t=0}$, we have to specify how the initial data depends on the future field. This backward-in-time response is simply given by

$$
d = R\, \phi_0 = \underbrace{R\, T_{-t}}_{R_t} \phi_t \equiv R_t \phi_t. \tag{64}
$$

Since we now have the response of the initial data $d = d_0$ to the field ϕ_t as well its variance Φ_t at a later time, we can simply write down the Wiener filter mean field at time t, which is

$$
m_t = \langle \phi_t \rangle_{(\phi_t|d)} = W_t d = \Phi R_t^\dagger \tilde{\Phi}^{-1} d. \tag{65}
$$

Here we used the identity $R_t \Phi R_t^\dagger = R\, T_{-t} \Phi T_{-t}^\dagger R^\dagger = R\, \Phi R^\dagger = \tilde{\Phi}$ that follows from (63). Therefore, any future mean field can be calculated directly from the original data, which therefore does not need to be evolved in time. The response R_t and Wiener filter W_t operators connecting the field at time t to the static data $d = d_{t=0}$ are exactly the ones which were found for the ideal IFD scheme. Thus, IFD reproduces IFT if the parameters of the future instances are able to capture all details of the evolved PDF.[9] The subgrid representation of the evolved field as given by (65) only requires complex operations in data space, since $\tilde{\Phi}$ is fully specified there. Solely the back-projection into continuous signal space by $R_t^\dagger$ and the subsequent spectral weighting by $\tilde{\Phi}$ require subgrid operations.

One might therefore ask how the virtual data $\tilde{d}_t = R \phi_t$ of the original response R applied to later field configurations would evolve and if this requires a subgrid field resolution. This is of importance to us, since we want to compare the IFD/IFT scheme with *ad hoc* schemes, which do not need to have a notion of a subgrid structure. Since the future field is not precisely known, the correct data at later times cannot be specified. The best we can do is to calculate the *a posteriori* expectation value of this hypothetical future data. This ideal data at later time, $\check{d}_t \equiv \langle \tilde{d}_t \rangle = \langle R\phi_t \rangle_{(\phi_t|d)}$, is, therefore,

$$
\check{d}_t = \underbrace{R \Phi R_t^\dagger (R \Phi R^\dagger)^{-1}}_{\tilde{T}_t} d \equiv \tilde{T}_t d. \tag{66}
$$

Note that the time translation operator of the data $\tilde{T}_t$ is not unity in general; basically it is only $\tilde{T}_t = \mathbb{1}$ for $t = 0$, since one of the response operators contains a time translation of the field

$$
\begin{aligned}
(\tilde{T}_t)_{kq} &= (R \Phi T_{-t}^\dagger R^\dagger \tilde{\Phi}^{-1})_{kq} \\
&= \sum_{k' \in k + \mathcal{N}\mathbb{Z}} \frac{2\, [1 - \cos(k'\Delta)]}{k'^2 \Delta^2} \\
&\quad \times \begin{bmatrix} \omega_{k'}^{-2} \cos(\omega_{k'} t) & -\omega_{k'}^{-1} \sin(\omega_{k'} t) \\ \omega_{k'}^{-1} \sin(\omega_{k'} t) & \cos(\omega_{k'} t) \end{bmatrix} \tilde{\Phi}_{kq}^{-1}. \tag{67}
\end{aligned}
$$

Since this time evolution operator is fully determined in data space, and the subgrid mode dynamics is just captured by a sum in a prefactor to the computational expensive operator $\tilde{\Phi}_{kq}^{-1}$, we can conclude that a data-space-only scheme was derived. The time-evolving data $\check{d}_t = \langle \tilde{d}_t \rangle$ contain the same information as d, since the latter can be reconstructed from the former via $d = \tilde{T}_t^{-1} \check{d}_t$. We can derive an evolution equation for $\check{d}_t$ by simply taking the temporal derivative of (66),

$$
\partial_t \check{d}_t = (\partial_t \tilde{T}_t) d = (\partial_t \tilde{T}_t)\, \tilde{T}_t^{-1} \check{d}_t.
$$

It is obvious that this ideal evolution equation of the virtual data according to the original response R is not only more complicated than just having an evolving response R_t and

---

[9]The observation that an entropic matching approximation enforced in any instance of continuous time can result in the exact equation for a dynamical system was observed previously in an attempt to reconstruct quantum mechanics from statistics [12].
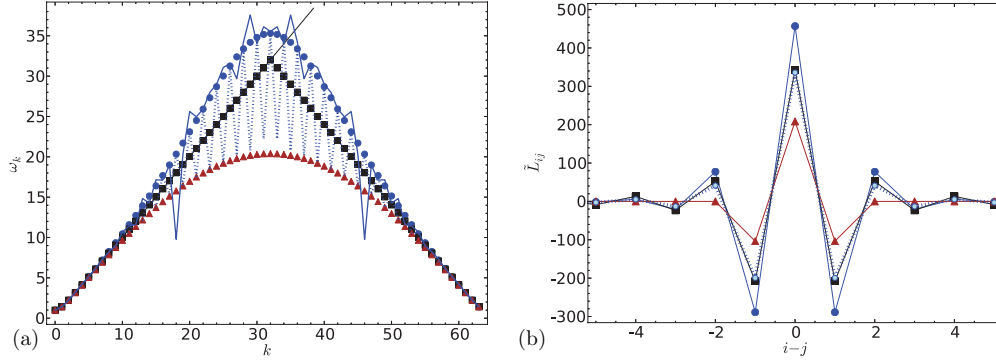
FIG. 2. (Color online) (a) Fourier–data-space dispersion relations $\tilde{\omega}_k$ of numerical schemes for the KG field simulation for the parameters $\mathcal{N} = 64$ and $\mu = 1$. The IFD scheme data mode frequencies $\check{\omega}_{k,t}$ are shown at initial time $t = 0$ as given by (69) (top, blue dots), an instance later at $t = 10^{-4}$ (top, blue solid line with kinks), and at time $t = \pi/2$ (strongly oscillating blue dotted line). At $t = \pi$, the IFD scheme dispersion relation looks similar to the initial one. The spectral scheme frequencies $\tilde{\omega}_k^{\text{spec}}$ as given by (71) (middle, black squares) follow the continuous-space field dispersion (thin, smooth, and black line). Finally, the finite-difference scheme $\tilde{\omega}_k^{\text{diff}}$ as given by (70) has the lowest frequencies (bottom, brick red triangles). (b) Data-space representation of the numerical scheme operator $\tilde{L}_{ij}$ as a function of the pixel number difference $i - j$ for small differences. The curves are given by the discrete Fourier transformations of $\tilde{\omega}_k^2$ for the IFD scheme at $t = 0$ (most extreme, blue dots and line) as well as for $t = \pi/2$ (smaller light blue dots and blue dotted line close to intermediate black line), the spectral scheme (intermediate values, black squares and line), and for the finite-difference scheme (most moderate values, brick red triangles and line). It should be noted that the IFD operator at $t = \pi/2$ also contains some power around positions $i - j = \pm \mathcal{N}/2 = \pm 32$ (not shown in this figure) as a consequence of the heavy oscillations of $\check{\omega}_{k,t}$ at this time that are visible in panel (a).

stationary data, it is also a differential equation with time-dependent coefficients. This might be surprising, since the dynamical equation of the underlying KG field is invariant under time translation. However, this time-translational symmetry is broken for our knowledge state on the field, for which the time $t = 0$ of the initial data set $d = R\phi_{t=0}$ is clearly singled out. The different Fourier data modes are mixtures of different field modes, which evolve with individual frequencies. Thus, the recovery of a similar mixture, $\tilde{d}_k = (R\,\phi_t)_k = \sum_{j\in\mathbb{Z}} 2\pi\, e^{-\frac{1}{2}ik\Delta} \text{sinc}(\frac{1}{2}k\Delta + \pi j)(T_t\phi)_{k+\mathcal{N}j}$, with the original phases in the response, works differently at different times, due to the changed phases of the individual modes. Therefore, the optimal IFD differential equation for data according to the original response becomes time dependent. Nevertheless, we would like to have something like a (now time-dependent) data mode frequency for a comparison with *ad hoc* simulation schemes. An observer of the data dynamics could estimate such a frequency in a pragmatic way by using $\partial_t^2 \check{d}_k + \check{\omega}_{k,t}^2 \check{d}_k = 0$ as an analog of $\partial_t^2 \varphi_k + \omega_k^2 \varphi_k = 0$ to define

$$\check{\omega}_{k,t}^2 = -\left(\partial_t^2 d_{k,t}^{(\varphi)}\right)\big/d_{k,t}^{(\varphi)}. \tag{68}$$

The resulting frequencies are best calculated numerically, since the involved formula (67) contains an infinite sum without a known closed form. For $t = 0$, however, a closed form can be derived,

$$\check{\omega}_{k,t=0}^2 = \mu^2 \left[1 - \frac{2}{\Delta\mu} \frac{\sinh(\mu\Delta)\sin\left(\frac{1}{2}k\Delta\right)^2}{\cosh(\mu\Delta) - \cos(k\Delta)}\right]^{-1}$$

$$= (k^2 + \mu^2)\left[1 + \frac{k^2\Delta^2}{12} + O(\Delta^4)\right], \tag{69}$$

that recovers the original continuous-space KG frequency $\omega_k = (k^2 + \mu^2)^{1/2}$ in the limit $\Delta \to 0$ but differs from it for finite grid spacings. The oscillation frequency of a data mode

is slightly higher than the directly corresponding continuous field mode, since the former also contains field modes from larger k, which have larger frequencies, due to the mode mixing of the response operator. The advanced revolution of the field modes at early times will be compensated later by a reduced oscillation speed. The initial and later time data dispersion relation is shown in Fig. 2 together with those of *ad hoc* schemes derived in Sec. IV. ◄

H. Summary of the derivation

► A brief summary of the essential steps of the IFD recipe applied to the KG problem might be instructive.

(A) Field dynamics. The KG equation was converted into a differential equation of first order in time, $\partial_t \phi = L\phi$, by the introduction of the momentum field $\pi_x = \dot{\varphi}_x$ as a second component of a two-component field $\phi = (\varphi, \pi)^\dagger$. The KG equation is linear as L is independent of ϕ. This simplified the derivation of an IFD scheme. If a nonlinear equation should be simulated, the equation has to be linearized around the current mean field at any simulation time step.

(B) Prior knowledge. The *a priori* KG field statistics was specified as a thermal distribution $\mathcal{P}(\phi) \propto \exp(-\beta\mathcal{H}(\phi))$. The fact that, in this case, the KG Hamiltonian $\mathcal{H}(\phi)$ determines both the dynamical operator L as well as the *a priori* statistics $\mathcal{P}(\phi)$ turns out to simplify the resulting scheme considerably. It is, however, not a general necessity for the applicability of IFD. The *a priori* distribution is a Gaussian since the Hamiltonian is quadratically in ϕ. If non-Gaussian priors are to be used it is recommended to find a Gaussian approximation since IFD is developed so far only for Gaussian priors.

(C) Data constraints. As a next step, the computer data space was introduced. The computer data d needs to be related to the field ϕ and this relation should be linear for

practical reasons and could be assumed to be noiseless for the KG example, $d = R\phi$. The initial discretization operator R was chosen here to perform a simple bin average. Therefore, the average field value in each bin is known if the data are available, but not the detailed field configuration within those. However, not all possible subgrid field configurations are equally plausible, since the prior gives them different weights. Combining prior and data information, the ensemble of plausible field configurations can be specified and characterized by its mean field $m = \langle\phi\rangle_{(\phi|d)}$ and uncertainty variance $D = \langle(\phi - m)(\phi - m)^\dagger\rangle_{(\phi|d)}$ determining a Gaussian *a posteriori* distribution $\mathcal{P}(\phi|d) = \mathcal{G}(\phi - m, D)$. This is a Gaussian thanks to the Gaussian prior and linear data model. The mean field and its variance are auxiliary mathematical objects used in the derivation of the simulation scheme that need not concrete representations in computer memory.

(D) Field evolution. The action of the time evolution operator on the posterior distribution had then to be worked out analytically. Since we insisted on linear or linearized operators, the time-evolved posterior is again a Gaussian, $\mathcal{P}(\phi'|d) = \mathcal{G}(\phi' - m^*, D^*)$, characterized by an updated mean m^* and uncertainty variance D^*, both again auxiliary mathematical objects.

(E) Prior update. The prior of the later time might differ and should be updated since it will be used again. However, due to energy and phase-space conservation of the KG dynamics, the KG prior is unchanged. This step could have been skipped, since the evolution of the prior can also be determined as part of the next step, the data update via entropic matching. However, this requires that the field dynamical equation captures all subgrid physics. If this not the case, the prior update step might permit implementation of subgrid processes not being present in the dynamical equation.

(F) Data update. Finally, an update formula for the later time data d' in computer memory was constructed. This was done by, first, specifying the mathematical relationship between any such data and the later time field *a posteriori* distribution, $\mathcal{P}(\phi'|d') = \mathcal{G}(\phi' - m', D')$, where $m' = D' R'^\dagger d'$ and R' and D' are response and propagator (or variance) at the later time. The time-evolved distribution $\mathcal{P}(\phi'|d)$ and the one determined by the new data $\mathcal{P}(\phi'|d')$ then were matched entropically. The parameters used to get an optimal match can be any of the later time, primed quantities. In the particular KG example it turned out to be most effective to vary d' and R' in the entropic matching since this way an information-lossless scheme could be obtained. This scheme maps the entire field evolution onto an evolving response operator R_t and stationary data. We showed that the resulting simulation scheme is indeed optimal by comparison to the exact information theoretically derived solution of the future field prediction problem. Since this particular KG simulation scheme does not modify the data, we asked how the binned field values (with stationary bin averaging) would evolve and derived their evolution equation. The time translation operator of this does also not require any explicit subgrid field representation but has encodes subgrid physics implicitly.

The derived simulation scheme can now be implemented on a computer. The resulting code performs only data-space operations and does not require any subgrid representation. The subgrid physics, the prior knowledge, and the details of

the measurement process (the data to fields relation) have all been included in the IFD scheme. ◄

IV. NUMERICAL VERIFICATION

A. Standard simulation schemes

► The IFD scheme for the KG field should now be compared to more standard simulation schemes for the KG equation as described in Appendix A 1.

The most common one is the finite-difference discretization of the differential operators by setting $\partial_x \varphi_x \approx (\varphi_{(i+1)\Delta} - \varphi_{i\Delta})/\Delta$ and $\partial_x^2 \varphi_x \approx (-\varphi_{(i+1)\Delta} + 2\varphi_{i\Delta} - \varphi_{(i-1)\Delta})/\Delta^2$. The KG equation discretized in this way, $\partial_t d = \tilde{L}^{\mathrm{diff}} d$ with $\tilde{L}_{ij}^{\mathrm{diff}} = \Delta^{-2}\delta_{i\,[j+1]_\mathcal{N}} - (2\Delta^{-2} + \mu^2)\delta_{ij} + \Delta^{-2}\delta_{i\,[j-1]_\mathcal{N}}$ and $[j]_\mathcal{N} = j \bmod \mathcal{N}$, becomes diagonal in Fourier space, just with the dispersion relation given by

$$\omega_k^2 \to \left(\tilde{\omega}_k^{\mathrm{diff}}\right)^2 = \mu^2 + 2\Delta^{-2}[1 - \cos(k\Delta)]. \qquad (70)$$

This and the IFD dispersion relation are shown in Fig. 2 in comparison to the one of the original KG field, $\omega^2 = \mu^2 + k^2$. Since the initial IFD frequencies are above, and the frequencies of the difference scheme are below the one of the KG field, it is also natural to consider the latter as another option. Thus, we also investigate a spectral simulation scheme with[10]

$$\left(\tilde{\omega}_k^{\mathrm{spec}}\right)^2 = \begin{cases} \mu^2 + k^2 & \text{for } k \in \{0, \ldots \mathcal{N}/2\} \\ \mu^2 + (\mathcal{N} - k)^2 & \text{for } k \in \{\mathcal{N}/2, \ldots \mathcal{N}\} \end{cases}. \qquad (71)$$

The Fourier-space data evolution equation can be solved analytically and has the solution

$$\begin{aligned} d_k^{(\varphi)} &= \tilde{a}_k e^{i\tilde{\omega}_k t} + \overline{\tilde{a}_{\mathcal{N}-k}} e^{-i\tilde{\omega}_k t}, \\ d_k^{(\pi)} &= i\tilde{\omega}_k(\tilde{a}_k e^{i\tilde{\omega}_k t} - \overline{\tilde{a}_{\mathcal{N}-k}} e^{-i\tilde{\omega}_k t}), \end{aligned} \qquad (72)$$

with the coefficients determined by the initial data,

$$\tilde{a}_k = \frac{d_{k\,t=0}^{(\varphi)}}{2} + \frac{d_{k\,t=0}^{(\pi)}}{2i\tilde{\omega}_k}. \qquad (73)$$

Thus, the most efficient simulation scheme for the KG field evolution schemes is to evolve the initial data according to these Fourier-space equations analytically and transform the field back to position space at the desired time.

The *ad hoc* simulation schemes are best implemented via (72) and (73), the corresponding data $\check{d}$ of the IFD scheme according to (66) and (67), whereas the full field including the subgrid modes can be followed via (25). ◄

B. Time evolution

► To see how well the different simulation schemes perform, we simulate a KG field by setting up its Fourier amplitudes $a_k \in \mathbb{C}$ up to $|k| = \mathcal{N}_\phi/2$ drawn from $\mathcal{P}(a_k) = \mathcal{G}\{a_k, 1/[4\beta(\mu^2 + k^2)]\}$ and $a_{\mathcal{N}_\phi - k} = \overline{a_k}$ for the "negative" modes, so (26), (28), and $\phi_x \in \mathbb{R}^2$ are satisfied. We use

---

[10]The distinctions of the cases is only necessary here, since we use $k \in \{0, \ldots \mathcal{N} - 1\}$ so the negative frequencies are represented by wave numbers in the second half of the range. If we would use $k \in \{-\mathcal{N}/2 + 1, \ldots \mathcal{N}/2\}$ as our first Brillouin zone, we would have $(\tilde{\omega}_k^{\mathrm{spec}})^2 = \mu^2 + k^2$.

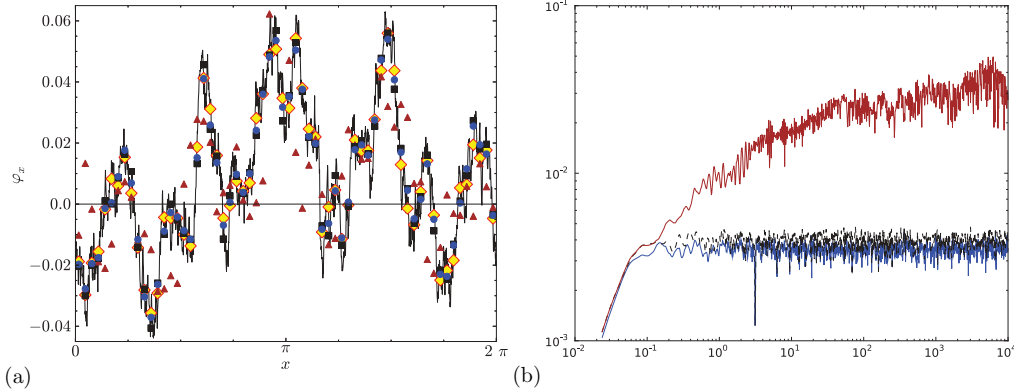FIG. 3. (Color online) (a) Evolved field (thin, black line) and data at $t = 10$ of the field also shown in Fig. 1 ($\beta = 1$, $\mu = 1$, $\mathcal{N} = 64$). The exact data $\tilde{d}_t = R\varphi_t$ are shown as yellow diamonds. The IFD data according to (66) and (67) (blue dots) follows the exact data closely. The data of the spectral scheme (black squares) is very close to the IFD data. The data of the difference scheme (brick red triangles) exhibit the poorest match to the correct data of the evolved field. The root-mean-square errors of the field data values $\sigma_d^{(\varphi)} = \sqrt{\sum_{i=0}^{\mathcal{N}-1}(\tilde{d}^{(\varphi)} - R\varphi)_i^2 / \mathcal{N}}$ of the three schemes are 0.003, 0.004, and 0.020 for the IFD, spectral, and difference schemes, respectively. (b) Temporal evolution of the data error $\sigma_d^{(\varphi)}(t)$ for the IFD (bottom solid blue line), spectral (dashed black line slightly above the former), and finite-difference (top brick red line) schemes. The dip in the IFD and spectral scheme error at $t = \pi$ is due to the nearly perfect alignment of the mode phases at this particular time.

$\mathcal{N}_\phi = 2048$, $\mu = 1$, and $\beta = 1$. A resulting field realization is displayed in Fig. 1. We time evolve all its Fourier modes according to (25). The initial and late time exact data are generated via $\tilde{d}_t = R\phi_t$ with the response given by (32) for $\mathcal{N} = 64$ data bins. This means that there are $\mathcal{N}_\phi/\mathcal{N} = 32$ independent field modes combined in a single datum, ensuring that there is substantial subgrid uncertainty, as is well observable in Fig. 1. For the spectral and difference schemes, the data are time evolved according to (72) and (73). For the IFD scheme, we use (66) and (67) to calculate corresponding late time data.

For time $t = 10$, the field is shown and the different data sets at this time are compared in Fig. 3. This time was chosen because the difference scheme already exhibits some significant but still moderate deviations from the correct solution. The IFD and spectral schemes are both relatively accurate. A difference between them exists but is hard to see by eye in this snapshot. However, a comparison of the spatially averaged errors of the two schemes reveals a significantly higher accuracy of the IFD scheme with respect to the spectral scheme at basically all times.

Although the IFD scheme has the highest fidelity, the spectral scheme is also very good for arbitrarily large times. The reason can easily be understood. Despite the fact that any data Fourier mode is a mixture of several field modes, the spectral scheme just follows the most dominant of these modes and treats the others as random noise. However, since the main mode is correctly captured, it can be followed for infinitely large intervals, and the ignored modes just contribute a fixed amount of uncertainty. The IFD scheme also assigns some power to these higher modes and follows their evolution. This is why it has a higher accuracy.

Optimally, one would have chosen an initial response that maps the first $\mathcal{N}$ Fourier modes of the field exactly into the data. Then these modes could have been followed with absolute precision, while one would have no information on the lower-amplitude higher Fourier modes. In this case, the IFD scheme

would have been identical to the spectral scheme, but it would not have served us well as a sufficiently complex example illustrating the inner workings of the IFD framework. ◄

V. CONCLUSIONS AND OUTLOOK

Information field dynamics serves as a framework to derive numerical simulation schemes. It rests on information field theory in order to construct continuous-space field configurations from the finite data in computer memory. It uses the maximum entropy principle to construct updated computer memory data so the ensemble of time-evolved continuous-space field configurations is matched by the ensemble implied by the updated data with minimal information loss.

The data-updating operations of an IFD simulation time step, as given by (55) and (56), are, in general, complex and might require the usage of linear algebra solvers. However, for numerical stability reasons, an implicit time step scheme might be adopted for a simulation anyway, and the linear algebra operations of the implicit and IFD schemes might be performed together.

As an illustrative example, we have derived the optimal IFD scheme for a thermally excited Klein-Gordon field. It could be shown that the resulting IFD scheme is identical to the one resulting from IFT. The scheme is much more accurate than a simplistic real-space discretization of the differential operator, and it is still significantly more accurate than a spectral scheme. In comparison to these two *ad hoc* schemes with stationary evolution equations for the data, the IFD scheme exhibits a time-dependent discretization of the differential equation. This is due to its ability to follow to some level the evolution of the subgrid scales without representing them explicitly in computer memory but capturing their influence implicitly in the data-update rules.

This initial work on IFD should be regarded as a proposal for how to incorporate information-theoretical considerations into the construction of simulation schemes. IFD permits us to

state and include explicitly background knowledge on subgrid behavior as well as external measurement data in a way that hopefully exploits and conserves as much of the available information as possible.

For technical reasons, one might compromise information-theoretical fidelity for reducing the numerical complexity. Also for this balance, the information-theoretical language introduced here should help to judge the choices. Finally, the language of IFD is already what is needed for data assimilation simulation schemes, as, for example, used in weather forecasts. The next goal of this research line is to develop IFD schemes for scientifically and technologically more relevant problems, like turbulent hydrodynamics. This, however, is left for future work.

ACKNOWLEDGMENTS

APPENDIX A: PREVIOUS WORK

1. Discretization of differential operators

Most of the dynamical systems in physics are described by partial differential equations. These contain differential operators acting on the dynamical fields. With the finite representation of the fields in computer memory, these operators need a discretized representation as well. A number of discretization schemes have been developed, including finite-difference methods, finite-volume methods, finite-element methods, spectral methods, smoothed particle hydrodynamics and others. Most of these schemes assume a distinct subgrid structure for the fields, in contrast to IFD.

Finite-difference methods [14] represent differentials by finite differences between the field values at the lattice grid points. These finite-difference operators are exact if the field is polynomial of the order of the operator. Thus, a finite-difference gradient operator implicitly assumes the field to be piecewise linear on subgrid scales, a Laplace operator the field to be quadratic and so forth. In Sec. IV we will show numerically that the IFD operator for the KG field evolution is superior to the finite-difference operator.

Finite-volume methods [15] are used when conserved quantities are simulated, such as, e.g., the fluid mass in hydrodynamics. The space is split into pixel volumes. The continuity equations for the conserved quantities can be turned into balance equations for the fluxes of the quantity through the boundaries of a pixel's volume. The simplest assumption for the subgrid field configuration is that it is constant within the pixels, with jumps at their boundaries. The resulting discontinuities have to be treated as separate Riemann problems at the boundaries in hydrodynamics. A conservative IFD scheme should also be possible if the stored data of the scheme are the amounts of the conserved quantity within pixel volumes and the fluxes between adjacent pixels.

Finite-element methods [16,17] also partition the space into subvolumes, the "elements." A set of basis functions for the field is defined, with a support covering only a small number of the elements or pixels. The field is represented as a linear combination of these basis functions and, therefore, with a tightly parametrized subgrid structure, e.g., being piecewise linear. The partial differential equations are only required to be solved weakly, in the Sobolev function space spanned by the chosen basis functions. This turns spatial differential operators into linear systems of equations, which then can be solved on a computer.

Spectral methods are also Sobolev space based, just with the basis functions being Fourier modes. We will compare the IFD scheme for the KG field to a spectral method and show that IFD provides a slightly more accurate simulation.

Smoothed particle hydrodynamics [18–20] discretizes the mass of the fluid and not the space. Smoothed particle hydrodynamics is one example of Lagrangian methods, in which the "grid" follows the flow. Each mass element has a dynamically evolving position and is thought to be distributed over some finite ball according to a radially declining and adaptively sized kernel function determining the subgrid field structure.

Moving mesh codes can be regarded as a compromise between Eulerian schemes with fixed lattices and Lagrangian schemes with a comoving but particle-based fluid discretization as smoothed particle hydrodynamics [21,22]. Moving mesh codes were recently improved by using Voronoi tessellation to create flexible volume cells around the moving grid points on which finite-volume methods can be used [23]. Thus, also the subgrid field representation is of a predetermined functional form.

In contrast to these approaches, IFD does not assume an *a priori* shape of the subgrid field structure. It considers all possible subgrid configurations consistent with the constraints given by the data and the field equations but weights them with *a priori* plausibilities. This requires knowledge on the subgrid dynamics.

2. Subgrid scale modeling

IFD, as proposed here, requires prior information on all modes of the dynamical field in order to constrain the unresolved degrees of freedom. The necessity to use information on subgrid scales in simulations was already realized for hydrodynamics. For this reason, the method of *large eddy simulations* was developed [24–26]. This resolves the largest scale of a flow by simulating a spatially filtered (convolved) dynamics in combination with subgrid scale models that try to summarize the effect of the unresolved scales on the global dynamics [27–30]. Usually stress tensors describe the subgrid scales. These are actually velocity fluctuation covariance matrices and, therefore, conceptually similar to the uncertainty dynamical field covariances in IFD. Large-eddy simulations have recently been combined with adaptive mesh refinement methods that increase the resolution at locations where small-scale dynamics is particularly important. This is especially important in astrophysical applications, where a large range of scales should be followed, as, for example, in galaxy clusters [31,32].

In *astrophysical hydrodynamics*, many additional processes on unresolved scales, like star formation and radiative feedback, are relevant yet cannot be followed in detail. In simulations of galaxies using smooth particle hydrodynamics, the interstellar medium is often described as a mixture of

interacting gas phases (e.g., molecular and ionized) forming a complex weather, with a single *effective equation of state* summarizing these phases [33]. However, the translation of subgrid physics into a concrete simulation scheme is usually done *ad hoc* without considering the resolution-dependent level of subgrid fluctuations.

In *oceanography*, it has been recognized that some information about subgrid eddy evolution is contained in the large-scale fluid motions due to the practical incompressibility of water and the resulting solenoidality of the flow patterns. Partial *reconstruction of the subgrid eddies* from a coarse resolution is therefore possible [34]. This has been used to construct accurate simulation schemes for advective tracers and for vorticity transport [35,36]. A *maximum entropy production principle* was introduced in this context in order to construct subgrid configurations that are numerically stable [35]. There maximum entropy was regarded merely as a numerical regularization trick, while in our work it plays an important role in ensuring optimal information flow between the simulation data at different time steps.

3. Data assimilation methods

Data assimilation methods are probably most similar in spirit to IFD. Data assimilation methods are used in weather forecast calculations to impose constraints from past measurements on numerical simulation of the atmosphere. A recent comparison of such methods can be found in Ref. [37]. The gold standard of the field is the full Bayesian posterior distribution of the dynamical system given all data. Typically, there are two broad classes of algorithms used to approximate this in a computationally affordable way: particle ensemble filters and variational methods.

Particle filters represent the knowledge and uncertainty on the system state as an ensemble of realizations, called the particles. These evolve individually according to the system dynamics to later times, when new measurements are available. Then, the particles are selected and/or reweighted according to their individual consistency with the new data. Resampling this distribution with a new set of particles (now with equal weights) closes the loop and prepares for the next simulation time step. A recent discussion of such methods can be found in Ref. [38].

Ensemble Kalman filters represent the system knowledge as well as an ensemble of realizations that can be propagated by the full nonlinear dynamics in time. The data assimilation step, however, is not done via reweighting or resampling but by Kalman filtering. Kalman filtering is basically Wiener filtering, which we introduced in Sec. II A, while using an empirically determined signal covariance matrix. This is computed from the ensemble, which is informed by the actual external measurement data.

Variational methods for data assimilation combine the action of a Lagrangian determining the dynamics and a loss function describing a penalty for any mismatch of the model prediction and the data [39]. From this combined Lagrangian, combining dynamics and data constraints, a variational equation aries that satisfies both the system dynamics and the data constraints. Variational methods treat information processing and field dynamics simultaneously, similarly to IFD.

A third approach to data assimilation has recently been proposed for the simulation of cosmic structure formation [40–42]. There the full posteriori of the cosmic matter field as determined by galaxy catalogs and the Gaussian initial condition statistics of cosmic structure formation is sampled via a *Hamiltonian sampling* method.

APPENDIX B: MAXIMUM ENTROPY PRINCIPLE

The MEP [7–10] is uniquely specified by the following three requirements on how probabilities should be ranked and updated with respect to new information. Entropy is defined to quantify how well a given PDF represents a knowledge state. Its functional form is determined by three requirements on the resulting probability updating scheme.

(a) Locality. Local information has local effects; information that affects only some part of the phase space should not modify the entropy and the implied MEP PDF in case this area is discarded.

(b) Coordinate invariance. The system of coordinates of the phase space does not carry information. Entropy should be invariant under coordinate transformation as well as the determined MEP PDF.

(c) Independence. Independent systems can be treated jointly or separately, yielding the same entropy in both cases. The joint MEP PDF therefore must be separable into a product of PDFs for the individual systems.

The unique (up to trivial rescaling) entropy functional on PDFs that is consistent with these requirements is given by (13), as shown in Refs. [7–10]. The usual way to use this entropy in order to specify the PDF $\mathcal{P}(\phi)$ is to maximize it subject to some constraints imposed on certain moments of the signal field statistics. An obvious one is the proper normalization $\langle 1 \rangle_{\mathcal{P}(\phi)} = 1$ of the PDF, but also a number of higher moments might be known *a priori*, and summarized in the form $\langle f_i(\phi) \rangle_{\mathcal{P}(\phi)} = a_i$. Here the functions could be simple moments like ϕ, $\phi\phi^\dagger$, etc., or more complicated functions thereof. These constraints on PDF moments are then incorporated into the entropy via Lagrange multiplier or thermodynamical potentials μ and $\lambda = (\lambda_i)_i$,

$$
\begin{aligned}
&\mathcal{S}(\mathcal{P}, \mu, \lambda | \mathcal{Q}) \\
&= \mathcal{S}(\mathcal{P} | \mathcal{Q}) - \langle \mu + \lambda^\dagger f(\phi) \rangle \\
&= -\int \mathcal{D}\phi \, \mathcal{P}(\phi) \left[\log\left(\frac{\mathcal{P}(\phi)}{\mathcal{Q}(\phi)} \right) + \mu + \lambda^\dagger f(\phi) \right]. \quad \text{(B1)}
\end{aligned}
$$

Maximizing this entropy with respect to all components of $\mathcal{P}(\phi)$ yields

$$
\mathcal{P}(\phi) = \frac{\mathcal{Q}(\phi)}{Z(\lambda)} e^{-\lambda^\dagger f(\phi)}, \quad \text{(B2)}
$$

where

$$
Z(\lambda) = \int \mathcal{D}\phi \, Q(\phi) \, e^{-\lambda^\dagger f(\phi)} \quad \text{(B3)}
$$

ensures proper normalization, and the Lagrange potentials λ have to be chosen to satisfy

$$
-\partial_\lambda \mathcal{S} = \partial_\lambda \log Z = \int \mathcal{D}\phi \, \mathcal{P}(\phi) \, f(\phi) = \langle f(\phi) \rangle_{\mathcal{P}(\phi)} = a.
$$

$$
\text{(B4)}
$$

In Sec. III B, it is claimed that the MEP distribution for ϕ with known mean ψ and covariance Φ is the Gaussian $\mathcal{G}(\phi - \psi, \Phi)$. This can now be verified by a short calculation. The entropy (B1) can be constrained by the knowledge of zero, first, and second moments of the field via the Lagrange-multiplier scalar μ, field λ, and matrix Λ, respectively,

$$
\begin{aligned}
&\mathcal{S}(\mathcal{P}, \mu, \lambda, \Lambda | \mathcal{Q}) \\
&= \mathcal{S}(\mathcal{P}|\mathcal{Q}) - \mu - \lambda^\dagger \langle \phi \rangle_{(\phi)} - \mathrm{Tr}(\Lambda \langle \phi \phi^\dagger \rangle_{(\phi)}) \\
&= - \int \mathcal{D}\phi \, \mathcal{P}(\phi) \left\{ \log\left[\frac{\mathcal{P}(\phi)}{\mathcal{Q}(\phi)} \right] + \mu + \lambda^\dagger \phi + \phi^\dagger \Lambda \, \phi \right\}.
\end{aligned}
\tag{B5}
$$

Minimizing this with respect to all components of $\mathcal{P}(\phi)$ for a flat prior-prior $\mathcal{Q}(\phi) = \mathrm{const}$ subject to the constraints

$$
-\partial_\mu \mathcal{S} = \langle 1 \rangle_{(\phi)} = 1, \tag{B6}
$$

$$
-\partial_\lambda \mathcal{S} = \langle \phi \rangle_{(\phi)} = \psi, \tag{B7}
$$

$$
-\partial_\Lambda \mathcal{S} = \langle \phi \, \phi^\dagger \rangle_{(\phi)} = \Phi + \psi \, \psi^\dagger, \tag{B8}
$$

to ensure proper PDF normalization, mean, and variance, respectively, yields $\mathcal{P}(\phi | \psi, \Phi) = \mathcal{G}(\phi - \psi, \Phi)$ as assumed in (27).

[1] J. C. Lemm, arXiv:physics/9912005.
[2] T. A. Enßlin, M. Frommert, and F. S. Kitaura, Phys. Rev. D **80**, 105005 (2009).
[3] N. Wiener, *Extrapolation, Interpolation, and Smoothing of Stationary Time Series* (Wiley, New York, 1949).
[4] T. A. Enßlin and M. Frommert, Phys. Rev. D **83**, 105014 (2011).
[5] T. A. Enßlin and C. Weig, Phys. Rev. E **82**, 051112 (2010).
[6] N. Oppermann, G. Robbers, and T. A. Enßlin, Phys. Rev. E **84**, 041118 (2011).
[7] E. T. Jaynes, Phys. Rev. **108**, 171 (1957).
[8] E. T. Jaynes, Phys. Rev. **106**, 620 (1957).
[9] E. T. Jaynes, Proc. IEEE **70**, 939 (1982).
[10] A. Caticha, AIP Conf. Proc. **1073**, 23 (2008).
[11] T. Ramalho, M. Selig, U. Gerland, and T. A. Enßlin, arXiv:1209.3700 [q-bio.QM].
[12] A. Caticha, *Proceeding of the 29th International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering* (Oxford, Mississippi, USA, 2009).
[13] W. Stein *et al.*, *Sage Mathematics Software (Version 4.7)*, the Sage development team (2011) [http://www.sagemath.org] .
[14] R. Courant, K. Friedrichs, and H. Lewy, Math. Annalen **100**, 32 (1928) [English translation, with commentaries by P. B. Lax, O. B. Widlund, S. V. Parter, IBM J. Res. Develop. 11 (1967)].
[15] S. K. Godunov, Math. Sbornik **47**, 271 (1959).
[16] W. Ritz, J. reine angew. Math. **135**, 1 (1908).
[17] B. Galerkin, Vestnik Inzh. **19**, 897 (1915).
[18] R. A. Gingold and J. J. Monaghan, Mon. Not. R. Astron. Soc. **181**, 375 (1977).
[19] L. B. Lucy, Astrophys. J. **82**, 1013 (1977).
[20] V. Springel and L. Hernquist, Mon. Not. R. Astron. Soc. **333**, 649 (2002).
[21] K.-H. A. Winkler, M. L. Norman, and M. J. Newman, Physica D **12**, 408 (1984).
[22] K.-H. A. Winkler, M. L. Norman, and D. Mihalas, J. Quantum Spec. Radiat. Transf. **31**, 473 (1984).
[23] V. Springel, Mon. Not. R. Astron. Soc. **401**, 791 (2010).
[24] J. Smagorinsky, Mon. Weather Rev. **91**, 99 (1963).
[25] J. W. Deardorff, J. Fluid Mech. **41**, 453 (1970).
[26] M. Lesieur and O. Metais, Annu. Rev. Fluid Mech. **28**, 45 (1996).
[27] V. M. Canuto, Astrophys. J. **428**, 729 (1994).
[28] V. M. Canuto, Astrolphys. J. **478**, 322 (1997).
[29] V. M. Canuto, Astrophys. J. **541**, L79 (2000).
[30] W. Schmidt and C. Federrath, Astron. Astrophys. **528**, A106 (2011).
[31] A. Maier, L. Iapichino, W. Schmidt, and J. C. Niemeyer, Astrophys. J. **707**, 40 (2009).
[32] F. Vazza, E. Roediger, and M. Brüggen, Astron. Astrophys. **544**, A103 (2012).
[33] V. Springel and L. Hernquist, Mon. Not. R. Astron. Soc. **339**, 289 (2003).
[34] A. F. Bennett, J. Atmos. Sci. **41**, 1881 (1984).
[35] F. Bouchet, arXiv:cond-mat/0305205.
[36] J. L. Sommer, F. d'Ovidio, and G. Madec, Ocean Model. **39**, 154 (2011).
[37] K. J. H. Law and A. M. Stuart, Mon. Weather Rev. **140**, 3757 (2012).
[38] P. J. van Leeuwen, Quart. J. R. Meteorol. Soc. **136**, 1991 (2010).
[39] A. Bennett, *Inverse Modeling of the Ocean and Atmosphere* (Cambridge University Press, Cambridge, UK, 2002).
[40] J. Jasche and B. D. Wandelt, arXiv:1203.3639 [astro-ph.CO].
[41] F.-S. Kitaura, Mon. Not. R. Astron. Soc. (in press), arXiv:1203.4184 [astro-ph.CO].
[42] F. Kitaura, J. Jasche, and R. B. Metcalf, Mon. Not. R. Astron. Soc. **403**, 589 (2010).