Information field theory for cosmological perturbation reconstruction and nonlinear signal analysis

Torsten A. Enßlin, Mona Frommert, and Francisco S. Kitaura

Max-Planck-Institut für Astrophysik, Karl-Schwarzschild-Straße 1, 85741 Garching, Germany

(Received 20 June 2008; published 9 November 2009)

We develop information field theory (IFT) as a means of Bayesian inference on spatially distributed signals, the information fields. A didactical approach is attempted. Starting from general considerations on the nature of measurements, signals, noise, and their relation to a physical reality, we derive the information Hamiltonian, the source field, propagator, and interaction terms. Free IFT reproduces the well-known Wiener-filter theory. Interacting IFT can be diagrammatically expanded, for which we provide the Feynman rules in position-, Fourier-, and spherical-harmonics space, and the Boltzmann-Shannon information measure. The theory should be applicable in many fields. However, here, two cosmological signal recovery problems are discussed in their IFT formulation. (1) Reconstruction of the cosmic large-scale structure matter distribution from discrete galaxy counts in incomplete galaxy surveys within a simple model of galaxy formation. We show that a Gaussian signal, which should resemble the initial density perturbations of the Universe, observed with a strongly nonlinear, incomplete and Poissonian-noise affected response, as the processes of structure and galaxy formation and observations provide, can be reconstructed thanks to the virtue of a response-renormalization flow equation. (2) We design a filter to detect local nonlinearities in the cosmic microwave background, which are predicted from some early-Universe inflationary scenarios, and expected due to measurement imperfections. This filter is the optimal Bayes' estimator up to linear order in the nonlinearity parameter and can be used even to construct sky maps of nonlinearities in the data.

DOI: 10.1103/PhysRevD.80.105005 PACS numbers: 89.70.−a, 11.10.−z, 95.75.−z, 98.80.−k

I. INTRODUCTION

A. Motivation

The optimal extraction and restoration of information from data on spatially distributed quantities like the cosmic large-scale structure (LSS) or the cosmic microwave background (CMB) temperature fluctuations in cosmology, but also on many other signals in physics and related fields, is essential for any quantitative, data-driven scientific inference. The problem of how to design such methods possesses many technical and even conceptual difficulties, which have led to a large number of recipes and methodologies.

Here, we address such problems from a strictly information theoretical point of view. We show, as others have done before, that information theory for distributed quantities leads to a statistical field theory, which we name information field theory (IFT). In contrast to the previous works, which mostly treat such problems on a classical field level, as will be detailed later, here, we take full advantage of the existing field theoretical apparatus to treat interacting and nonclassical fields. Thus, we show how to use diagrammatic perturbation theory and renormalization flows in order to construct optimal signal recovering algorithms and to calculate moments of their uncertainties. Nonclassicality manifests itself as quantum and statistical fluctuations in quantum and statistical field theory (QFT and SFT), and very similarly as uncertainty in IFT.

The information theoretical perspective on signal inference problems has technical advantages, since it permits one to design information-yield optimized algorithms and experimental setups. However, it also provides deeper insight into the mechanisms of knowledge accumulation, its underlying information flows, and its dependence on data models, prior knowledge and assumptions than pure empirical evaluations of *ad hoc* algorithms alone could provide.

We therefore hope that our work is of interest for two types of readers. The first are applied scientists, who are mainly interested in the practical aspect of IFT since they are facing a concrete inverse problem for a spatially distributed quantity, especially but not exclusively in cosmology. The second are more philosophical or theoretically inclined scientists, for whom IFT may serve as a framework to understand and classify many of the existing methods of signal extraction and reception. Since we expect that many interested readers are not very familiar with field theoretical formalisms, we introduce some of its basic mathematical concepts. Because of this anticipated nonuniform readership, not everything in this article might be of interest to everyone, and therefore we provide in the following a short overview on the structure and content of the article.

B. Overview of the work

The remainder of the Introduction contains a detailed discussion of the previous work on signal inference theory

as well as a very brief introduction into the here relevant works on the cosmic LSS and the CMB. The main part of this article falls into two categories: abstract IFT and its application. The concepts of IFT are introduced in Sec. II, where Bayesian methodology, the distinction of physical and information fields, the definition of signal response and noise, as well as the design of signal spaces are discussed. The basic IFT formalism including the free theory is introduced in Sec. III, which, according to our judgment, summarizes and unifies the previous knowledge on IFT before this paper. An impatient reader, only interested in applying IFT and not worrying about concepts, may start reading in Sec. III. From Sec. IV on, the new results of this work are presented, starting with the discussion of interacting information fields, their Hamiltonians and Feynman rules, and the Boltzmann-Shannon information measure. The normalizability of sensibly constructed IFTs is shown, as well as the classical information field equation. A step-by-step recipe of how to derive and implement an IFT algorithm is also provided.

Details of the notation can be found, if not defined in the main text, in Appendix A.

Applications of the theory are provided in the following two sections, which can be skipped by a reader interested only in the general theoretical framework. Although specific inference problems are addressed, they should serve as a blueprint for the tackling of similar problems. In Sec. V the problem of the reconstruction of the cosmic matter distribution from galaxy surveys is analyzed in terms of a Poissonain data model. In Sec. VI we derive an optimal estimator for non-Gaussianity in the CMB, and show how it can be generalized to map potential non-Gaussianities in the CMB sky. Our summary and outlook can be found in Sec. VII.

C. Previous works

The work presented here tries to unify information theory and statistical field theory in order to provide a conceptual framework in which optimal tools for cosmological signal analysis can be derived, as well as for inference problems in other disciplines. Below, we provide very brief introductions into each of the required fields[1] (information theory, image reconstruction, statistical field theory, cosmological large-scale structure, and cosmic microwave background), for the orientation of nonexpert readers. An expert in any of these fields might decide to skip the corresponding sections.

[1]This work has tremendously benefitted in a direct and indirect way from a large number of previous publications in those fields. We, the authors, have to apologize for being unable to give full credit to all relevant former works in those fields for only concentrating on a brief summary of the papers more or less directly influencing this work. This collection is obviously highly biased toward the cosmological literature due to our main scientific interests and expertise, and definitely incomplete.

1. Information theory and Bayesian inference

The fundament of information theory was laid by the work of Bayes [1] on probability theory, in which the celebrated Bayes theorem was presented. The theorem itself [see Eq. (7)] is a simple rule for conditional probabilities. It only unfolds its power for inference problems if used with belief or knowledge states, described by conditional probabilities.

The advent of modern information theory is probably best dated by the work of Shannon [2,3] on the concept of information measure, being the negative Boltzmann entropy, and the work of Jaynes, combining the language of statistical mechanics and Bayes probability theory and applying it to knowledge uncertainties [4–10]. The required numerical evaluation of Bayesian probability integrals suffered often from the curse of high dimensionality. The standard recipe against this, still in massive use today, is importance sampling via Markov-chain Monte Carlo methods (MCMC), following the ideas of Metropolis *et al.* [11], Hastings [12], and Geman and Geman [13], where the latter authors already had image reconstruction applications in mind. The Hamiltonian MCMC methods [14], in which the phase-space sampling is partly following Hamiltonian dynamics, are also of relevance here. There the Hamiltonian is introduced as the negative logarithm of the probability, as we do in this work.

With such tools, higher dimensional problems, as present in signal restoration, could and can be tackled, however, for the price of getting stochastic uncertainty into the computational results. For a recent review on image restoration MCMC techniques, see [15].

The applications and extensions of these pioneering works are too numerous to be listed here. Good monographs exist and the necessary references can be found there [16–21].

2. Image reconstruction in astronomy and elsewhere

The problem of image reconstruction from incomplete, noisy data is especially important in astronomy, where the experimental conditions are largely set by the nature of distant objects, weather conditions, etc., all mainly out of the control of the observer, as well as in other disciplines like medicine and geology, with similar limitations to arrange the object of observations for an optimal measurement. Some of the most prominent methods of image reconstruction, which are based on a Bayesian implementation of an assumed data model, are the Wiener filter [22], the Richardson-Lucy algorithm [23,24], and the maximum-entropy image restoration [25] (see also [26–37]).

The Wiener filter can be regarded to be a full Bayesian image inference method in case of Gaussian signal and noise statistics, as we will show in Sec. III B. It will be the working horse of the IFT formalism, since the Wiener filter represents the algorithm to construct the exact field theo-

retical expectation value given the data for an interaction-free information Hamiltonian. The filter can be decomposed into two essential information processing steps, first building the information source by response-over-noise weighting the data, and then propagating this information through the signal space, by applying the so-called Wiener variance.

The Richardson-Lucy algorithm is a maximum-likelihood method to reconstruct from Poissonian data and therefore is also of Bayesian origin. This method is usually to be regularized by hand, by truncation of the iterative calculations, against an overfitting instability due to the missing (or implicitly flat) signal prior. A Gaussian-prior based regularization was recently proposed by Kitaura and Enßlin [38], and the implementation of a variant of this is presented here in Sec. V D.

Maximum-entropy algorithms will not be the topic here, as well as not a number of other existing methods, which are partly within and partly outside the Bayesian framework. They may be found in existing reviews on this topic, e.g. [39,40].

3. Statistical and Bayesian field theory

The relation of signal reconstruction problems and field theory was discovered independently by several authors. In cosmology, a prominent work in this direction was Bertschinger [41], in which the path-integral approach was proposed to sample primordial density perturbations with a Gaussian statistics under the constraint of existing information on the large-scale structure. The work presented here can be regarded as a nonlinear, non-Gaussian extension of this. Many methods from statistics and from statistical mechanics were of course used even earlier, e.g. the usage of a moment generating function for cosmic density fields can already be found by Fry [42].

Simultaneously to Bertschinger's work, Bialek and Zee [43,44] argued that visual perception can be modeled as a field theory for the true image, being distorted by noise and other data transformations, which are summarized by a nuisance field. A probabilistic language was used, but no direct reference to information theory was made, since the optimal information reconstruction was not the aim, but a model for the human visual reception system. However, this work actually triggered our research.

Bialek *et al.* [45] applied a field theoretical approach to recover a probability distribution from data. Here, a Bayesian prior was used to regularize the solution, which was set up *ad hoc* to enforce smoothness of the reconstruction, obtained from the classical [or saddle point, or maximum *a posteriori* (MAP)] solution of the problem. However, an ''optimal'' value for the smoothness controlling parameter was derived from the data itself, a topic also addressed by Stoica *et al.* [46] and by a follow-up publication to ours [47]. Bialek *et al.* [45] also recognized, as we do, that an IFT can easily be nonlocal.

Finally, the work of Lemm and co-workers [48–55] established a tight connection between statistical field theory and Bayesian inference, and proposed the term Bayesian field theory (BFT) for this. However, we prefer the term information field theory since it puts the emphasis on the relevant object, the information, whereas BFT refers to a method, Bayesian inference. The term information field is rather self-explaining, whereas the meaning of a Bayesian field is not that obvious.

The applications considered by Lemm concentrate on the reconstruction of probability fields over parameter spaces and quantum mechanical potentials by means of the maximum *a posteriori* equation. The extensive book summarizing the essential insights of these papers, [48], clearly states the possibility of perturbative expansions of the field theory. However, this is not followed up by these authors probably for reasons of the computational complexity of the required algorithms. In contrast to many of the previous works on IFT, which deal with *ad hoc* priors, the publication by Lemm [56] is remarkable, since it provides explicit recipes of how to implement *a priori* information in various circumstances more rigorously.

The mathematical tools required to tackle IFT problems come from SFT and QFT, which have a vast literature. We have specially made use of the books of Binney *et al.* [57], Peskin and Schroeder [58], and Zee [59].

4. Cosmological large-scale structure

Our first IFT example in Sec. V is geared toward improving galaxy-survey based cosmography, the reconstruction of the large-scale structure matter distribution. We provide here a short overview on the relevant background and works.

The LSS of the matter distribution of the Universe is traced by the spatial distribution of galaxies, and therefore well observable. This structure is believed to have emerged from tiny, mostly Gaussian initial density fluctuations of a relative strength of 10^{-5} via a self-gravitational instability, partly counteracted by the expansion of the Universe. The initial density fluctuations are believed to be produced during an early inflationary epoch of the Universe, and to carry valuable information about the inflaton, the field which drove inflation, in their N-point correlation functions, to be extracted from the observational data.

The onset of the structure formation process is well described by linear perturbation theory and therefore to conserve Gaussianity, however, the later evolution, the structures on smaller scales, and especially the galaxy formation require nonlinear descriptions. The observational situation is complicated by the fact that the most important galaxy distance indicator, their redshift, is also sensitive to the galaxy peculiar velocity, which causes the observational data on the three-dimensional LSS to be partially degenerated. There are analytical methods to

describe these effects,[2] and also extensive work on N-body simulations of the structure formation, the latter probably providing us with the most detailed and accurate statistical data on the properties of the matter density field, e.g [75].

In recent years, it was recognized that the evolution of the cosmic density field and its statistical properties can be addressed with field theoretical methods by virtue of renormalization-flow equations. Detailed semianalytical calculations for the density field time propagator, the two- and three-point correlation functions are now possible due to this, which are expected to play an important role in future approaches to reconstruct the initial fluctuations from the observational data [76–90].

It was recognized early on that the primordial density fluctuations can in principle be reconstructed from galaxy observations [41]. This has led to a large development of various numerical techniques for an optimal reconstruction [91–131]. Many of them are based on a Bayesian approach, since they are implementations and extension of the Wiener filter. However, also other principles are used, like, e.g. the least action approach, or Voronoi tessellation techniques [132–138]. A discussion and classification of the various methods can be found in [38].

Especially the Wiener filter methods were extensively applied to galaxy-survey data[3] and permitted partly to extrapolate the matter distribution into the *zone of avoidance* behind the galactic disk and to close the data gap there, cf. [157–159], a topic we also address in Sec. V.

Another cosmological relevant information field to be extracted from galaxy catalogs is the LSS power spectrum [160–164]. This power is also measurable in the CMB, and for a long time the CMB provided the best spectrum normalization [165,166].

5. Cosmic microwave background

Since our second example deals with the CMB, we give a brief overview on it and related inference methods.

The CMB reveals the statistical properties of the matter field at a time when the Universe was about 1100 times smaller in linear size than it is today. The photon-baryon fluid, which decouples at that epoch into neutral hydrogen and freestreaming photons, has responded to the gravitational pull of the then already forming dark-matter structures. The photons from that epoch cooled due to the cosmic expansion since then into the CMB radiation we observe today, and carry information on the physical properties of the photon-baryon fluid of that time like density, temperature, and velocity. To very high accuracy, the spectrum of the photons from any direction is that of a blackbody, with a mean temperature of 2.7 K and fluctuations of the order of 10^{-5} K, imprinted by the primordial gravitational potentials at decoupling.

Therefore, mapping these temperature fluctuations permits one to precisely study many cosmological parameters simultaneously, like the amount of dark matter producing the gravitational potentials, the ratio of photons to baryons, balancing the pressure and weight of the fluid, and geometrical and dynamical parameters of space-time itself. The observations are technically challenging, and therefore require sophisticated algorithms to extract the tiny signal of temperature fluctuations against the instrument noise, but also to separate it from other astrophysical foreground emission with the best possible accuracy.

A number of such algorithms were developed [167–184], which in many cases implement the Wiener filter. Thus, the required numerical tools for an IFT treatment of CMB data are essentially available.

The expected temperature fluctuations spectrum can be calculated from a linear perturbative treatment of the Boltzmann equations of all dynamical active particle species at this epoch, and fast computational implementations exist permitting one to predict it for a given set of cosmological parameters. Well-known codes for this task are publicly available[4] and permit one to extract information on cosmological parameters from the measured CMB temperature fluctuation spectrum via comparison to their predictions for a given parameter set. It was recognized early on that this should happen in an information theoretically optimal way, and Bayesian methods were therefore adapted in that area well before other astrophysical disciplines [188–191].

The initial metric and density fluctuations, from which the CMB fluctuations and the LSS emerged, are believed to be initially seeded by quantum fluctuations of a hypothetical inflaton field, which should have driven an inflationary expansion phase in the very early Universe [192–197]. The inflaton-induced fluctuations have a very Gaussian probability distribution; however, some non-Gaussian features seem to be unavoidable in most scenarios and can serve as a fingerprint to discriminate among them [198–201]. Observational tests on such non-Gaussianities based on the three-point correlation function of the CMB data [202–206] were so far mostly negative, however not sensitive enough to seriously constrain the possible theoretical parameter space of inflationary scenarios; see e.g. [207,208]. Recently, there has been the claim of a detection of such non-Gaussianities by Yadav and Wandelt [209] and a confirmation of this with better data and improved algorithms is therefore highly desirable. In Sec. VI we make a proposal for improving the algorithmic side of this challenge. A recent review on the current status of CMB Gaussianity can be found in [210].

[2]Of special interest in this context may be [60], which already applies path integrals, [61–74], and the papers they refer to.

[3]Survey based reconstructions of the cosmic matter fields can be found in [139–156].

[4]For example, CMBFAST [185], CAMB [186], and CMBEASY [187].

II. CONCEPTS OF INFORMATION FIELD THEORY

A. Information on physical fields

In our attempts to infer the properties of our Universe from astronomical observations we are faced with the problem of how to interpret incomplete, imperfect and noisy data, draw our conclusions based on them, and quantify the uncertainties of our results. This is true for using galaxy surveys to map the cosmic LSS, for the interpretation of the CMB, as well as for many experiments in physical laboratories and compilations of geological, economical, sociological, and biological data about our planet. Information theory, which is based on probability theory and the Bayesian interpretation of missing knowledge as a probabilistic uncertainty, offers an ideal framework to handle such problems. It permits one to describe all relevant processes involved in the measurement probabilistically, provided a model for the Universe or the system under consideration is adopted.

The states of such a model, denoted by the state variable ψ, are identified with the possible physical realities. They can have probabilities $P(\psi)$ assigned to them, the so-called prior information. This prior contains our knowledge about the Universe as we model it before any other data are taken. For a given cosmological model, the prior may be the probability distribution of the different initial conditions of the Universe, which determine the subsequent evolution completely. Since our Universe is spatially extended, the state variable will in general contain one or several fields, which are functions over some coordinates x.

Also the measurement process is described by a data model which defines the so-called likelihood, the probability $P(d|\psi)$ to obtain a specific data set d given the physical condition ψ. In case the outcome d of the measurement is deterministic $P(d|\psi) = \delta(d - d[\psi])$, where $d[\psi]$ is the functional dependence of the data on the state. In any case, the probability distribution function of the data,

$$P(d) = \int \mathcal{D}\psi P(d|\psi)P(\psi), \tag{1}$$

is given in terms of a phase space or path integral over all possible realizations of ψ, to be defined more precisely later (Sec. II E 1).

A scientist is not actually interested in the total state of the Universe, but only in some specific aspects of it, which we call the signal $s = s[\psi]$. The signal is a very reduced description of the physical reality and can be any function of its state ψ, freely chosen according to the needs and interests of the scientist or the ability and capacity of the measurement and computational devices used. Since the signal does not contain the full physical state, any physical degree of freedom which is not present in the signal but influences the data will be received as probabilistic uncertainty, or shortly noise. The probability distribution function of the signal, its prior

$$P(s) = \int \mathcal{D}\psi \delta(s - s[\psi])P(\psi), \tag{2}$$

is related to that of the data via the joint probability

$$P(d, s) = \int \mathcal{D}\psi \delta(s - s[\psi])P(d|\psi)P(\psi), \tag{3}$$

from which the conditional signal likelihood

$$P(d|s) = P(d, s)/P(s), \tag{4}$$

and signal posterior

$$P(s|d) = P(d, s)/P(d), \tag{5}$$

can be derived.

Before the data are available, the phase space of interest is spanned by the direct product of all possible signals s and data d, and all regions with nonzero $P(d, s)$ are of potential relevance. Once the actual data d_{obs} have been taken, only a submanifold of this space, as fixed by the data, is of further relevance. The probability function over this subspace is proportional to $P(d = d_{\mathrm{obs}}, s)$, and needs just to be renormalized by dividing by

$$\begin{aligned}\int \mathcal{D}s P(d_{\mathrm{obs}}, s) &= \int \mathcal{D}s \int \mathcal{D}\psi \delta(s - s[\psi])P(d_{\mathrm{obs}}|\psi)P(\psi) \\ &= \int \mathcal{D}\psi P(d_{\mathrm{obs}}|\psi)P(\psi) = P(d_{\mathrm{obs}}),\end{aligned} \tag{6}$$

which is the unconditioned probability (or evidence) of that data. Thus, we find the resulting information of the data to be the posterior distribution $P(s|d_{\mathrm{obs}}) = P(d_{\mathrm{obs}}, s)/P(d_{\mathrm{obs}})$. This posterior is the fundamental mathematical object from which all our deductions have to be made. It is related via Bayes's theorem [1] to the usually better accessible signal likelihood,

$$P(s|d) = P(d|s)P(s)/P(d), \tag{7}$$

which follows from Eqs. (4) and (5).

The normalization term in Bayes's theorem, the evidence $P(d)$, is now also fully expressed in terms of the joint probability of data and signal,

$$P(d) = \int \mathcal{D}s P(d, s), \tag{8}$$

and the underlying physical field ψ basically becomes invisible at this stage in the formalism. The evidence plays a central role in Bayes inference, since it is the likelihood of all the assumed model parameters. Combining this parameter likelihood with parameter priors, one can start Bayesian inference on the model classes.

B. Signal response and noise

If signal and data depend on the same underlying physical properties, there may be correlations between the two, which can be expressed in terms of signal response R and

noise n of the data as

$$d = R[s] + n_s. \tag{9}$$

We have chosen two different ways of denoting the dependence of response and noise on the signal s, in order to highlight that the response should embrace most of the reaction of the data to the signal, whereas the noise should be as independent as possible. We ensure this by putting the linear correlation of the data with the signal fully into the response. The response is therefore the part of the data which correlates with the signal

$$R[s] \equiv \langle d \rangle_{(d|s)} \equiv \int \mathcal{D}d\, dP(d|s), \tag{10}$$

and the noise is just defined as the remaining part which does not:

$$n_s \equiv d - R[s] = d - \langle d \rangle_{(d|s)}. \tag{11}$$

Although the noise might depend on the signal, as it is well known, for example, for Poissonian processes, it is—per definition—linearly uncorrelated to it,

$$\langle n_s s^\dagger \rangle_{(d|s)} = (\langle d \rangle_{(d|s)} - R[s]) s^\dagger = 0 s^\dagger = 0, \tag{12}$$

whereas higher-order correlation might well exist and may be further exploited for their information content. The dagger denotes complex conjugation and transposing of a vector or matrix.

These definitions were chosen to be close to the usual language in signal processing and data analysis. They permit one to define signal response and noise for an arbitrary choice of the signal $s[\psi]$. No direct causal connection between signal and data is needed in order to have a nontrivial response, since both variables just need to exhibit some couplings to a common subaspect of ψ. The above definition of response and noise is however not unique, even for a fixed signal definition, since any data transformation $d' = T[d]$ can lead to different definitions, as seen from

$$R'[s] \equiv \langle d' \rangle_{(d|s)} = \langle T[d] \rangle_{(d|s)} \neq T[\langle d \rangle_{(d|s)}] = T[R[s]]. \tag{13}$$

Exceptions are some unique relations between signal and state, $P(\psi|s) = \delta(\psi - \psi[s])$, and maybe a few other very special cases. Thus, the concepts of signal response and therewith defined noise depend on the adopted coordinate system in the data space. This coordinate system can be changed via a data transformation T, and the transformed data may exhibit a better or worse response to the signal. Information theory aids in designing a suitable data transformation, so that the signal response is maximal, and the signal noise is minimal, permitting the signal to be best recovered. Thus, we may aim for an optimal T, which yields

$$T[d] = \langle s \rangle_{(s|d)}. \tag{14}$$

We define the posterior average of the signal, $m_d = \langle s \rangle_{(s|d)}$, to be the map of the signal given the data d and call T a *map-making algorithm* if it fulfills Eq. (14) at least approximately. As a criterion for this one may require that the signal response of a map-making algorithm,

$$R_T[s] \equiv \langle T[d] \rangle_{(d|s)}, \tag{15}$$

is positive definite with respect to signal variations as stated by

$$\frac{\delta R_T[s]}{\delta s} \geq 0. \tag{16}$$

This ensures that a map-making algorithm will respond with a non-negative correlation of the map to any signal feature, with respect to the noise ensemble. In general, T will be a nonlinear operation on the data, to be constructed from information theory if it should be optimal in the sense of Eq. (14). In any case, the fidelity of a signal reconstruction can be characterized by the quadratic signal uncertainty,

$$\sigma^2_{T,d} = \langle (s - T[d])(s - T[d])^\dagger \rangle_{(s|d)}, \tag{17}$$

averaged over typical realizations of signal and noise. Of special interest is the trace of this

$$\mathrm{Tr}\,(\sigma^2_{T,d}) = \int dx\, \langle |s_x - T_x[d]|^2 \rangle_{(s|d)}, \tag{18}$$

since it is the expectation value of the squared Lebesgue-L^2-space distance between a signal reconstruction and the underlying signal. Requesting a map-making algorithm to be optimal with respect to Eq. (18) implies $T[d] = \langle s \rangle_{(s|d)}$ and therefore to be optimal in an information theoretical sense according to Eq. (14).

The uncertainty $\sigma^2_{T,d}$ depends on d, since in Bayesian inference one averages over the posterior, which is conditional to the data. The frequentist uncertainty estimate, which is the expected uncertainty of any estimator before the data are obtained, is given by an average over the joint probability function:

$$\sigma^2_T = \langle (s - T[d])(s - T[d])^\dagger \rangle_{(d,s)}. \tag{19}$$

The latter is a good quantity to characterize the overall performance of an estimator, whereas $\mathrm{Tr}(\sigma^2_{T,d})$ is a more precise indicator of the actual estimator performance for a given data set. As we will see in our IFT applications, data dependence of the uncertainty is a common feature of nonlinear inference problems.

An illustrative example should be in order. Suppose our data are an exact copy of a physical field, $d = \psi$, our signal the square of the latter, $s = \psi^2$, and the physical field obeys an even statistics, $P(\psi) = P(-\psi)$. Then, the signal response is exactly zero, $R[s] = 0$, and the data contain only noise with respect to the chosen signal, $d = n_s$. Thus, we have chosen a bad representation of our data

to reveal the signal. If we, however, introduce the transformation $d' = T[d] = d^2$, we find a perfect response, $R'[s] = s$, and zero noise, $n'_s = 0$.

In this case, finding the optimal map-making algorithm was trivial, but in more complicated situations, it cannot be guessed that easily. Since the response and noise definitions depend on the signal definition, some thoughts should be given to how to choose the signal in a way that it can be well reconstructed.

C. Signal design

For practical reasons one will usually choose s according to a few guidelines, which should simplify the information induction process:

(1) The functional form of $s[\psi]$ should best be simple, steady, analytic, and if possible linear in ψ, permitting one to use the signal s to reason about the state of reality ψ.
(2) The degrees of freedom of s should be related to the ones of the data d in the sense that cross correlations exist which permit one to deduce properties of s from d. Signal degrees of freedom, which are insensitive to the data, will only be constrained by the prior and therefore just contain a large amount of uncertainty. This adds to the error budget and should be avoided as much as possible.
(3) The choice of $s[\psi]$ should also be led by mathematical convenience and practicality. In the examples presented in this work, simple signals are chosen which permit one to guess good approximations for signal likelihood $P(d|s)$ and prior $P(s)$ without the need to develop the full physical theory starting with $P(\psi)$.

To give a more specific example, we assume a cosmological model in which the reality is thought to be solely characterized by the primordial dark-matter density distribution $\psi(x)$, from which all observable cosmological phenomena like galaxies derive in a deterministic way. The coordinate x may refer to the comoving coordinates at some early epoch of the Universe. Although the LSS of the matter distribution at a later time may predominantly depend on the initial large-scale modes, and is reflected in the galaxy distribution, the actual positions of the individual galaxies also depend in a nontrivial way on the small-scale modes. Because of the discreteness of our observable, the galaxy positions, it may be impossible to reconstruct these small-scale modes. Therefore it could be sensible to define a signal $s[\psi] = F\psi$, with F being a linear low-pass filter, which suppresses all small-scale structures. This signal may be reconstructible with high precision, whereas any attempt to reconstruct ψ directly would be plagued by a larger error budget, since all the data-unconstrained small-scale modes represent uncertainties to a reconstruction of ψ, but not to one of s being defined as a low-pass filtered version of ψ.

D. Signal moment calculation

The information of some data d on a signal s defined over some set Ω, which in most applications will be a manifold like a subvolume of the R^n, or the sphere in case of a CMB signal, is completely contained in the posterior $P(s|d)$ of the signal given the data.[5] The expectation value of s at some location $x \in \Omega$, and higher correlation functions of s can all be obtained from the posterior by taking the appropriate average:

$$\langle s(x_1)\cdots s(x_n)\rangle_d \equiv \langle s(x_1)\cdots s(x_n)\rangle_{(s|d)} \equiv \int \mathcal{D}s\, s(x_1)\cdots s(x_n) P(s|d). \quad (20)$$

The problem is that often neither the expectation values nor even the posterior are easily calculated analytically, even for fairly simple data models. Fortunately, there is at least one class of data models for which the posterior and all its moments can be calculated exactly, namely, in case the posterior turns out to be a multivariate Gaussian in s. In this case analytical formulas for all moments of the signal are known and are in principle computable. Technically, one is still often facing a huge, but linear inverse problem. However, in the last decades a couple of computational high-performance map-making techniques were developed to tackle such problems either on the sphere, for CMB research, or in flat spaces with one, two, or three dimensions, for example, for the reconstruction of the cosmic LSS (detailed references are given in Sec. I C). The purpose of this work is to show how to expand other posterior distributions around the Gaussian ones in a perturbative manner, which then permits one to use the existing map-making codes for the computation of the resulting diagrammatic perturbation series. Since the diagrammatic perturbation series in the Feynman diagrams are well known and understood in QFT and SFT, the most economical way is to reformulate the information theoretical problem in a language which is as close as possible to the former two theories. Thereby, many of the results and concepts become directly available for signal inference problems. Moreover, it seems that expressing the optimal signal estimator in terms of Feynman diagrams immediately provides computationally efficient algorithms, since the diagrams encode the skeleton of the minimal necessary computational information flow.

E. Signal and data spaces

1. Discretization and continuous limit

Both, the signal and the data space may be continuous, however, in practice will most often be discrete since

[5] We are mostly dealing with scalar fields, however, multicomponent, vector or tensor fields can be treated analogously, and many of the equations just have to be reinterpreted for such fields and stay valid.

digital data processing only permits one to chose a discretized representation of the distributed information. The space in which the data and signal discretization happens can be chosen freely, and of course can be as well a Fourier, wavelet, or spherical-harmonics space. Even if we would like to analyze a continuous signal, the computationally required discretization will force an implicit redefinition of our actual signal to be the discretely sampled version of that continuous signal, and this discretization step should also be part of the data model, if it has the potential to significantly affect the analysis [211].

Although discretization implies some information loss it also has an advantage. We can just assume discretization and therefore read all scalar and tensor products as being the usual, componentwise ones, now just in high-, but finite-dimensional vector spaces.

To be concrete, let $\{x_i\} \subset \Omega$ be a discrete set of N_{pix} pixel positions, each of which has a volume size V_i attributed to it; then the scalar product of two discretized function vectors $f = (f_i)$, and $g = (g_i)$ sampled at these points via $f_i = f(x_i)$, and $g_i = g(x_i)$ could be defined by

$$g^\dagger f \equiv \sum_{i=1}^{N_{\mathrm{pix}}} V_i g_i^* f_i. \tag{21}$$

The asterisk denotes complex conjugation. This scalar product has the continuous limit

$$g^\dagger f \rightarrow \int_\Omega dx\, g^*(x) f(x). \tag{22}$$

In many cases the actual volume normalization in Eq. (21) does not matter for final results, since it usually cancels out, and therefore V_i is often dropped completely for equidistant sampling of signal and data spaces. The volume terms also disappear for a scalar product involving a function which is discretized via volume integration, $f_i = \int_{V_i} dx f(x)$, e.g. the number of counts within the cell i. Anyhow, higher-order tensor products are defined analogously.

The path integral of a functional $F[f] \equiv F(f_1, \ldots, f_{N_{\mathrm{pix}}})$ over all realizations of such a discretized field f is then just a high-dimensional volume integral, with as many dimensions as pixels:

$$\int \mathcal{D}f F[f] \equiv \left(\prod_{i=1}^{N_{\mathrm{pix}}} \int df_i\right) F(f_1, \ldots, f_{N_{\mathrm{pix}}}). \tag{23}$$

This definition of a finite-dimensional path integral is well normalized, since in case we want to integrate over a probability distribution over f, which is separable for all pixels, $P(f) = \prod_{i=1}^{N_{\mathrm{pix}}} P_i(f_i)$, as e.g. for white and Poissonian noise, we find

$$\langle 1 \rangle_{(f)} = \int \mathcal{D}f P(f) = \prod_{i=1}^{N_{\mathrm{pix}}} \underbrace{\int df_i P_i(f_i)}_{=1} = 1. \tag{24}$$

Although, in real data-analysis applications, it is practically never required to perform the continuous limit $N_{\mathrm{pix}} \rightarrow \infty$ with $V_i \rightarrow 0$ for all i, we stress that this limit can formally be taken and is well defined even for the path integral, as we argue in more detail in Sec. IV B. The basic argument is that suitable signals could and should be defined in such a way that path-integral divergences, which plague sometimes QFT, can easily be avoided by sensible signal design. Practically, the existence of a well-defined continuous limit of a well-posed IFT implies that two numerical implementations of a signal reconstruction problem, which differ in their space discretization on scales smaller than the structures of the signal, can be expected to provide identical results up to a small discretization difference, which vanishes with higher discretization resolution.

2. Parameter spaces

In many applications, the signal space is identified with the physical space or with the sphere of the sky. However, IFT can also be done over parameter spaces. In Sec. VI, a field theory over the sphere will implicitly define the knowledge state for an unknown parameter of that theory, which can be regarded again to define an information theory for that parameter. The latter is an IFT in case that the parameter has spatial variations.

However, there are also functions defined over a parameter space, $\Omega_{\mathrm{parameter}} = \{p\}$ for some parameter p, which one might want to obtain knowledge on from incomplete data. A very import one is the probability distribution of the parameter given the observational data, $P(p|d)$, which defines our parameter-knowledge state. This function may only be incompletely known and therefore requires an IFT approach for its reconstruction and interpolation. Such incomplete knowledge on the function could be due to incomplete numerical sampling of its function values because of large computational costs and the huge volumes of multidimensional parameter spaces. Or, there might be another unknown nuisance parameter q in the problem, which induces an uncertainty in $P_{(p|d)} = P(p|d)$ and therefore an IFT over all possible realizations of this knowledge state field function via

$$P[P_{(p|d)}] = \int \mathcal{D}P_{(p|d)} \delta\left[P_{(p|d)} - \int dq P(p, q|d)\right]. \tag{25}$$

In case that q is a field, the marginalization integral in the delta functional also becomes a path integral. Probabilistic decision theory, based on knowledge states as expressed by probability functions on parameters, has to deal with such complications. For inference directly on p, and not on the knowledge state $P_{(p|d)}$, the marginalized probability

$$P(p|d) = \int dq P(p, q|d) \tag{26}$$

contains all relevant information, and that will be sufficient for most inference applications, and especially for the ones in this work.

III. BASIC FORMALISM

A. Information Hamiltonian

We argued that the posterior $P(s|d)$ contains all available information on the signal. Although the posterior might not be easily accessible mathematically, we assume in the following that the prior $P(s)$ of the signal before the data are taken as well as the likelihood of the data given a signal $P(d|s)$ are known or at least can be Taylor-Fréchet expanded around some reference field configuration t. Then Bayes's theorem permits one to express the posterior as

$$P(s|d) = \frac{P(d, s)}{P(d)} = \frac{P(d|s)P(s)}{P(d)} \equiv \frac{1}{Z} e^{-H[s]}. \tag{27}$$

Here, the Hamiltonian

$$H[s] \equiv H_d[s] \equiv -\log[P(d, s)] = -\log[P(d|s)P(s)], \tag{28}$$

the evidence of the data

$$P(d) \equiv \int \mathcal{D}s\, P(d|s)P(s) = \int \mathcal{D}s\, e^{-H[s]} \equiv Z, \tag{29}$$

and the partition function $Z \equiv Z_d$ were introduced. It is extremely convenient to include a moment generating function into the definition of the partition function

$$Z_d[J] \equiv \int \mathcal{D}s\, e^{-H[s]+J^\dagger s}. \tag{30}$$

This means $P(d) = Z = Z[0]$, but also permits one to calculate any moment of the signal field via Fréchet differentiation of Eq. (30),

$$\langle s(x_1) \cdots s(x_n) \rangle_d = \frac{1}{Z} \frac{\delta^n Z_d[J]}{\delta J(x_1) \cdots \delta J(x_n)} \bigg|_{J=0}. \tag{31}$$

Of special importance are the so-called connected correlation functions or cumulants

$$\langle s(x_1) \cdots s(x_n) \rangle_d^c \equiv \frac{\delta^n \log Z_d[J]}{\delta J(x_1) \cdots \delta J(x_n)} \bigg|_{J=0}, \tag{32}$$

which are corrected for the contribution of lower moments to a correlator of order n. For example, the connected mean and dispersion are expressed in terms of their unconnected counterparts as

$$\begin{aligned} \langle s(x) \rangle_d^c &= \langle s(x) \rangle_d, \\ \langle s(x)s(y) \rangle_d^c &= \langle s(x)s(y) \rangle_d - \langle s(x) \rangle_d \langle s(y) \rangle_d, \end{aligned} \tag{33}$$

where the last term represents such a correction. For Gaussian random fields all higher-order connected correlators vanish:

$$\langle s(x_1) \cdots s(x_n) \rangle_d^c = 0 \tag{34}$$

for $n > 2$. For non-Gaussian random fields, they are in general nonzero, and for later usage we provide the connected three- and four-point functions,

$$\begin{aligned} \langle s_x s_y s_z \rangle_d^c &= \langle (s_x - \bar{s}_x)(s_y - \bar{s}_y)(s_z - \bar{s}_z) \rangle_d, \\ \langle s_x s_y s_z s_u \rangle_d^c &= \langle (s_x - \bar{s}_x)(s_y - \bar{s}_y)(s_z - \bar{s}_z)(s_u - \bar{s}_u) \rangle_d \\ &\quad - \langle s_x s_y \rangle_d^c \langle s_z s_u \rangle_d^c - \langle s_x s_z \rangle_d^c \langle s_y s_u \rangle_d^c \\ &\quad - \langle s_x s_u \rangle_d^c \langle s_y s_z \rangle_d^c, \end{aligned} \tag{35}$$

where we used $s_x = s(x)$ and defined $\bar{s}_x = \langle s(x) \rangle_d$.

The assumption that the Hamiltonian can be Taylor-Fréchet expanded in the signal field permits one to write

$$H[s] = \frac{1}{2} s^\dagger D^{-1} s - j^\dagger s + H_0 + \sum_{n=3}^{\infty} \frac{1}{n!} \Lambda^{(n)}_{x_1 \ldots x_n} s_{x_1} \cdots s_{x_n}. \tag{36}$$

Repeated coordinates are thought to be integrated over. The first three Taylor coefficients have special roles. The constant H_0 is fixed by the normalization condition of the joint probability density of signal and data. If $H'_d[s]$ denotes some unnormalized Hamiltonian, its normalization constant is given by

$$H_0 = \log \int \mathcal{D}s \int \mathcal{D}d\, e^{-H'_d[s]}. \tag{37}$$

Often H_0 is irrelevant unless different models or hyperparameters are to be compared.

We call the linear coefficient j information source. This term is usually directly and linearly related to the data. The quadratic coefficient, D^{-1}, defines the information propagator $D(x, y)$, which propagates information on the signal at y to location x, and thereby permits one, e.g., to partially reconstruct the signal at locations where no data were taken. Finally, the anharmonic tensors $\Lambda^{(n)}$ create interactions between the modes of the free, harmonic theory. Since this free theory will be the basis for the full interaction theory, we first investigate the case $\Lambda^{(n)} = 0$.

B. Free theory

1. Gaussian data model

For our simplest data model we assume a Gaussian signal with prior

$$P(s) = \mathcal{G}(s, S) \equiv \frac{1}{|2\pi S|^{1/2}} \exp\left(-\frac{1}{2} s^\dagger S^{-1} s\right), \tag{38}$$

where $S = \langle ss^\dagger \rangle$ is the signal covariance. The signal is assumed here to be processed by nature and our measurement device according to a linear data model

$$d = Rs + n. \tag{39}$$

Here, the response $R[s] = Rs$ is linear in and the noise $n_s = n$ is independent of the signal s. The linear response matrix R of our instrument can contain window and selection functions, blurring effects, and even a Fourier transformation of the signal space, if our instrument is an interferometer. Typically, the data space is discrete, whereas the signal space may be continuous. In that case the ith data point is given by

$$d_i = \int dx\, R_i(x) s(x) + n_i. \tag{40}$$

We assume, for the moment, but not in general, the noise to be signal independent and Gaussian, and therefore distributed as

$$P(n|s) = \mathcal{G}(n, N), \tag{41}$$

where $N = \langle nn^\dagger \rangle$ is the noise covariance matrix. Since the noise is just the difference of the data to the signal response, $n = d - Rs$, the likelihood of the data is given by

$$P(d|s) = P(n = d - Rs|s) = \mathcal{G}(d - Rs, N), \tag{42}$$

and thus the Hamiltonian of the Gaussian theory is

$$\begin{aligned} H_{\mathcal{G}}[s] &= -\log[P(d|s)P(s)] \\ &= -\log[\mathcal{G}(d - Rs, N)\mathcal{G}(s, S)] \\ &= \tfrac{1}{2} s^\dagger D^{-1} s - j^\dagger s + H_0^{\mathcal{G}}. \end{aligned} \tag{43}$$

Here

$$D = [S^{-1} + R^\dagger N^{-1} R]^{-1} \tag{44}$$

is the propagator of the free theory. The information source,

$$j = R^\dagger N^{-1} d, \tag{45}$$

depends linearly on the data in a response-over-noise weighted fashion and reads

$$j(x) = \sum_{ij} R_i^*(x) N_{ij}^{-1} d_j \tag{46}$$

in case of discrete data but continuous signal spaces. Finally,

$$H_0^{\mathcal{G}} = \tfrac{1}{2} d^\dagger N^{-1} d + \tfrac{1}{2} \log(|2\pi S||2\pi N|) \tag{47}$$

has absorbed all s-independent normalization constants.

The partition function of the free field theory,

$$\begin{aligned} Z_{\mathcal{G}}[J] &= \int \mathcal{D}s\, e^{-H_{\mathcal{G}}[s] + J^\dagger s} \\ &= \int \mathcal{D}s \exp\left\{ -\frac{1}{2} s^\dagger D^{-1} s + (J + j)^\dagger s - H_0^{\mathcal{G}} \right\}, \end{aligned} \tag{48}$$

is a Gaussian path integral, which can be calculated exactly, yielding

$$Z_{\mathcal{G}}[J] = \sqrt{|2\pi D|}\, \exp\{+\tfrac{1}{2}(J + j)^\dagger D (J + j) - H_0^{\mathcal{G}}\}. \tag{49}$$

The explicit partition function permits one to calculate via Eq. (32) the expectation of the signal given the data, in the following called the map m_d generated by the data d:

$$\begin{aligned} m_d = \langle s \rangle_d &= \frac{\delta \log Z_G}{\delta J} \bigg|_{J=0} = Dj \\ &= \underbrace{[S^{-1} + R^\dagger N^{-1} R]^{-1} R^\dagger N^{-1}}_{F_{\mathrm{WF}}} d. \end{aligned} \tag{50}$$

The last expression shows that the map is given by the data after applying a generalized Wiener filter, $m_d = F_{\mathrm{WF}} d$. The propagator $D(x, y)$ describes how the information on the density field contained in the data at location x propagates to position y: $m(y) = \int dx D(y, x) j(x)$.

The connected autocorrelation of the signal given the data,

$$\langle ss^\dagger \rangle_d^c = D = [S^{-1} + R^\dagger N^{-1} R]^{-1}, \tag{51}$$

is the propagator itself. All higher connected correlation functions are zero. Therefore, the signal given the data is a Gaussian random field around the mean m_d and with a variance of the residual error

$$r = s - m_d \tag{52}$$

provided by the propagator itself, as a straightforward calculation shows

$$\langle rr^\dagger \rangle_d = \langle ss^\dagger \rangle_d - \langle s \rangle_d \langle s^\dagger \rangle_d = \langle ss^\dagger \rangle_d^c = D. \tag{53}$$

Thus, the posterior should be simply a Gaussian given by

$$P(s|d) = \mathcal{G}(s - m_d, D). \tag{54}$$

As a test for the latter equation, we calculate the evidence of the free theory via

$$\begin{aligned} P(d) &= \frac{P(d|s)P(s)}{P(s|d)} = \frac{\mathcal{G}(d - Rs, N)\mathcal{G}(s, S)}{\mathcal{G}(s - Dj, D)} \\ &= \left(\frac{|D|/|S|}{|2\pi N|} \right)^{1/2} \exp\left\{ \frac{1}{2} (j^\dagger D j - d^\dagger N^{-1} d) \right\}, \end{aligned} \tag{55}$$

which is indeed independent of s and also identical to $Z_{\mathcal{G}}[0]$, as it should be.

2. Free classical theory

The Hamiltonian permits one to ask for *classical* equations derived from an extremal principle. This is justified, on the one hand, as being just the result of a the saddle point approximation of the exponential in the partition function. On the other hand, the extrema principle is equivalent to the MAP estimator, which is quite commonly used for the construction of signal filters. An exhaustive introduction into and discussion of the MAP approximation to Gaussian and non-Gaussian signal fields is provided by Lemm [48].

The classical theory is expected to capture essential features of the field theory. However, if the field fluctuations are able to probe phase-space regions away from the maximum in which the Hamiltonian (or posterior) has a more complex structure, deviations between classical and field theory should become apparent.

Extremizing the Hamiltonian of the free theory [Eq. (43)]

$$\left.\frac{\delta H_{\mathcal{G}}}{\delta s}\right|_{s=m} = D^{-1}m - j \equiv 0, \tag{56}$$

we get the classical mapping equation $m = Dj$, which is identical to the field theoretical result [Eq. (50)].

It is also possible to measure the sharpness of the maximum of the posterior by calculating the Hessian curvature matrix

$$\mathcal{H}[m] = \left.\frac{\delta^2 H[s]}{\delta s^2}\right|_{s=m} = D^{-1}. \tag{57}$$

In the Gaussian approximation of the maximum of the posterior, the inverse of the Hessian is identical to the covariance of the residual

$$\langle r r^\dagger \rangle = \mathcal{H}^{-1}[m] = D, \tag{58}$$

which for the pure Gaussian model is of course identical to the exact result, as given by the field theory [Eq. (53)].

IV. INTERACTING INFORMATION FIELDS

A. Interaction Hamiltonian

1. General form

All results of the free theory presented so far are well known within the field of signal reconstruction. IFT reproduces them elegantly and is therefore of pedagogical value. However, the new results presented in the rest of this paper arise as soon as one leaves the free theory. Non-Gaussian signal or noise, a nonlinear response, or a signal dependent noise create anharmonic terms in the Hamiltonian. These describe interactions between the eigenmodes of the free Hamiltonian.

We assume the Hamiltonian can be Taylor expanded in the signal fields, which permits one to write

$$H[s] = \underbrace{\frac{1}{2}s^\dagger D^{-1} s - j^\dagger s + H_0^{\mathcal{G}}}_{H_{\mathcal{G}}[s]} + \underbrace{\sum_{n=0}^{\infty}\frac{1}{n!}\Lambda^{(n)}_{x_1\cdots x_n} s_{x_1}\cdots s_{x_n}}_{H_{\mathrm{int}}[s]}. \tag{59}$$

Repeated coordinates are thought to be integrated over. In contrast to Eq. (36) we have now included perturbations which are constant, linear, and quadratic in the signal field, because we are summing from $n = 0$. This permits one to treat certain nonideal effects perturbatively. For example if a mostly position-independent propagator gets a small position-dependent contamination, it might be more convenient to treat the latter perturbatively and not to include it into the propagator used in the calculation. Note further that all coefficients can be assumed to be symmetric with respect to their coordinate indices.[6]

Often, it is more convenient to work with a shifted field $\phi = s - t$, where some (e.g. background) field t is removed from s. The Hamiltonian of ϕ reads

$$\begin{aligned} H'[\phi] &= \underbrace{\frac{1}{2}\phi^\dagger D^{-1}\phi - j'^\dagger\phi + H_0'}_{H'_G[\phi]} \\ &\quad + \underbrace{\sum_{n=0}^{\infty}\frac{1}{n!}\Lambda'^{(n)}_{x_1\cdots x_n}\phi_{x_1}\cdots\phi_{x_n}}_{H'_{\mathrm{int}}[\phi]}, \\ \text{with}\quad H_0' &= H_0^{\mathcal{G}} - j^\dagger t + \frac{1}{2}t^\dagger D^{-1} t, \\ j' &= j - D^{-1}t, \quad \text{and} \\ \Lambda'^{(m)}_{x_1\cdots x_m} &= \sum_{n=0}^{\infty}\frac{1}{n!}\Lambda^{(m+n)}_{x_1\cdots x_{m+n}} t_{x_1}\cdots t_{x_n}. \end{aligned} \tag{60}$$

2. Feynman rules

Since all the information on any correlation functions of the fields is contained in the partition sum and can be extracted from it, only the latter needs to be calculated:

[6]This means $D_{xy} = D_{yx}$ and $\Lambda^{(n)}_{x_{\pi(1)}\cdots x_{\pi(n)}} = \Lambda^{(n)}_{x_1\cdots x_n}$ with π any permutation of $\{1,\ldots,n\}$, since even nonsymmetric coefficients would automatically be symmetrized by the integration over all repeated coordinates. Therefore, we assume in the following that such a symmetrization operation has been already done, or we impose it by hand before we continue with any perturbative calculation by applying

$$\Lambda^{(n)}_{x_1\cdots x_n} \mapsto \frac{1}{n!}\sum_{\pi\in\mathcal{P}_n}\Lambda^{(n)}_{x_{\pi(1)}\cdots x_{\pi(n)}}.$$

This clearly leaves any symmetric tensor invariant if $\mathcal{P}_n$ is the space of all permutations of $\{1,\ldots,n\}$.

$$
\begin{aligned}
Z[J] &= \int \mathcal{D}s\, e^{-H[s]+J^\dagger s} \\
&= \int \mathcal{D}s\, \exp\Big[-\sum_{n=0}^{\infty}\frac{1}{n!}\Lambda^{(n)}_{x_1\cdots x_n} s_{x_1}\cdots s_{x_n}\Big] e^{-H_G[s]+J^\dagger s} \\
&= \exp\Big[-\sum_{n=0}^{\infty}\frac{1}{n!}\Lambda^{(n)}_{x_1\cdots x_n}\frac{\delta}{\delta J_{x_1}}\cdots\frac{\delta}{\delta J_{x_n}}\Big] \\
&\quad\times\int \mathcal{D}s\, e^{-H_G[s]+J^\dagger s} \\
&= \exp\Big[-H_{\mathrm{int}}\Big[\frac{\delta}{\delta J}\Big]\Big] Z_G[J]. \qquad (61)
\end{aligned}
$$

There exist well-known diagrammatic expansion techniques for such expressions [57]. The expansion terms of the logarithm of the partition sum, from which any connected moments can be calculated, are represented by all possible connected diagrams build out of lines (———), vertices (with a number of legs connecting to lines, like ,  , ,  , ...) and without any external line ends (any line ends in a vertex). These diagrams are interpreted according to the following Feynman rules:

(1) Open ends of lines in diagrams correspond to external coordinates and are labeled by such. Since the partition sum, in particular, does not depend on any external coordinate, it is calculated only from summing up closed diagrams. However, the field expectation value $m(x) = \langle s(x)\rangle_{(s|d)} = d\log Z[J]/dJ(x)|_{J=0}$ and higher-order correlation functions depend on coordinates and therefore are calculated from diagrams with one or more open ends, respectively.

(2) A line with coordinates x' and y' at its end represents the propagator $D_{x'y'}$ connecting these locations.

(3) Vertices with one leg get an individual internal, integrated coordinate x' and represent the term $j_{x'} + J_{x'} - \Lambda^{(1)}_{x'}$.

(4) Vertices with n legs represent the term $-\Lambda^{(n)}_{x'_1\cdots x'_n}$, where each individual leg is labeled by one of the internal coordinates $x'_1, \ldots, x'_n$. This more complex vertex structure, as compared to QFT, is a consequence of nonlocality in IFT.

(5) All internal (and therefore repeatedly occurring) coordinates are integrated over, whereas external coordinates are not.

(6) Every diagram is divided by its symmetry factor, the number of permutations of vertex legs leaving the topology invariant, as described in any book on field theory [57].

The nth cumulative moment of s is generated by taking the nth derivative of $\log Z[J]$ with respect to J, and then setting $J = 0$. This correspond to removing n end vertices from all diagrams. For example, the first four diagrams contributing to a map ($m = \langle s\rangle_{(s|d)}$) are

$$
\begin{aligned}
\text{  } &= D\,j = D_{xy}\,j_y \\
&\equiv \int dy\, D(x,y)\, j(y), \\
\text{  } &= -\frac{1}{2}\,D\,\Lambda^{(3)}[\cdot, D] = -\frac{1}{2} D_{xy}\,\Lambda^{(3)}_{yzu}\,D_{zu} \\
&\equiv -\frac{1}{2}\int dy\, D_{xy}\int dz\int du\,\Lambda^{(3)}_{xyu}\,D_{zu}, \\
\text{  } &= -\frac{1}{2}\,D\,\Lambda^{(3)}\,[\cdot, D\,j, D\,j] \\
&= -\frac{1}{2} D_{xy}\,\Lambda^{(3)}_{yuz}\,D_{zz'}\,j_{z'}\,D_{uu'}\,j_{u'} \\
&\equiv -\frac{1}{2}\int dy\, D_{xy}\int dz\int du\,\Lambda^{(3)}_{yzu} \\
&\quad\times\int dz'\, D_{zz'}\,j_{z'}\int du'\, D_{uu'}\,j_{u'},\ \text{and} \\
\text{  } &= -\frac{1}{2}\,D\,\Lambda^{(4)}[\cdot, D,\, D\,j] \\
&= -\frac{1}{2} D_{xy}\,\Lambda^{(4)}_{yzuv}\,D_{zu}\,D_{vv'}\,j_{v'} \\
&\equiv -\frac{1}{2}\int dy\, D_{xy}\int dz\int du\int dv\,\Lambda^{(4)}_{yzuv}\,D_{zu} \\
&\quad\times\int dv'\, D_{vv'}\,j_{v'}.
\end{aligned} \qquad (62)
$$

Here we have assumed that any first and second order perturbation was absorbed into the data source and the propagator, thus $\Lambda^{(1)} = \Lambda^{(2)} = 0$. Repeated indices are assumed to be integrated (or summed) over.

3. Local interactions and Fourier-space rules

In case of purely local interactions

$$
\Lambda^{(n)}_{x_1\cdots x_n} = \lambda_n(x_1)\delta(x_1 - x_2)\cdots\delta(x_1 - x_n), \qquad (63)
$$

the interaction Hamiltonian reads

$$
H_{\mathrm{int}} = \sum_{m=0}^{\infty}\frac{1}{m!}\lambda^\dagger_m s^m, \qquad (64)
$$

and the expressions of the Feynman diagrams simplify considerably. The fourth Feynman rule can be replaced by

(4) Vertices with n lines connected to it are associated with a single internal coordinate x' and represent the term $-\lambda_n(x')$.

For example, the last loop diagram in Eq. (62) becomes

$$
\text{  } = -\frac{1}{2}\int dy\, D_{xy}\,\lambda_4(y)\,D_{yy}\int dz\, D_{yz}\,j_z. \qquad (65)
$$

In case of local interactions, it can be helpful to do the calculations in Fourier space, for which the Feynman rules can be obtained by inserting a real-space identity operator $1 = F^\dagger F$ in between any scalar product and assigning the inverse Fourier transformation $F^\dagger$ to the left and the forward transform F to the right term, e.g.

$$Dj = F^\dagger \underbrace{FDF^\dagger}_{D'} \underbrace{Fj}_{j'} = F^\dagger D' j'.$$

This yields

(1) An open end of a line has an external momentum coordinate k, and gets an $\int dk e^{-ikx}/(2\pi)^n$ applied to it, if real-space functions are to be evaluated.
(2) A line connecting momentum k with momentum k' corresponds to a directed propagator between these momenta: $D_{kk'} = D(k, k')$.
(3) A data-source vertex is $(j + J - \lambda_1)(k'')$, where k'' is the momentum at the data end of the line.
(4) A vertex with $m > 1$ lines with momentum labels $k_1, \ldots, k_m$ is $-\lambda_m(k_0)(2\pi)^n \delta(\sum_{i=0}^m k_i)$.
(5) An internal end of a line has an internal (integrated) momentum coordinate k'. Integration means a term $\int dk'/(2\pi)^n$ in front of the expression.
(6) The expression gets divided by the symmetry factor of its diagram.

Here, $j(k) = (Fj)(k) = \int dx j(x) e^{ikx}$, $D(k,k') = (FDF^\dagger) \times (k,k') = \int dx \int dx' D(x,x') e^{i(kx - k'x')}$, etc. are the Fourier transformed information source, propagator, etc., respectively

Note that momentum directions have to be taken into account. The momenta that go into a vertex, data source, or open end get a positive sign in the delta function of momentum conservation, and the ones that go out of a vertex get a minus sign.

4. Simplistic interaction Hamiltonians

In order to have a toy case, which permits analytic calculations, we introduce a simplistic Hamiltonian by requiring the data model to be translational invariant and all interaction terms to be local. This is the case whenever the signal and noise covariances are fully characterized by power spectra over the same spatial space,

$$S(k, q) = (2\pi)^n \delta(k - q) P_S(k), \tag{66}$$

$$N(k, q) = (2\pi)^n \delta(k - q) P_N(k), \tag{67}$$

with $P_s(k) = \langle |s(k)|^2 \rangle / V$, and $P_n(k) = \langle |n(k)|^2 \rangle / V$, where V is the volume of the system. We assume further that the signal processing can be completely described by a convolution with an instrumental beam,

$$d(x) = \int dy\, R(x - y) s(y) + n(x), \tag{68}$$

where the response-convolution kernel has a Fourier power spectrum $P_R(k) = |R(k)|^2$ (no factor $1/V$). In this case D can be fully described by a power spectrum:

$$D(k, q) = (2\pi)^n \delta(k - q) P_D(k), \tag{69}$$

with $P_D(k) = (P_S^{-1}(k) + P_R(k) P_N^{-1}(k))^{-1}$.

The locality of the interaction terms requires $\lambda_m = \text{const}$ beside translational invariance and therefore the interaction Hamiltonian reads

$$H_{\text{int}}[s] = \sum_{m=1}^{\infty} \frac{\lambda_m}{m!} \int dx\, s^m(x) = \sum_{m=1}^{\infty} \frac{\lambda_m}{m!} \left(\prod_{i=1}^m \int \frac{dk_i}{(2\pi)^n} s_{k_i} \right) (2\pi)^n \delta\left(\sum_{j=1}^m k_j \right). \tag{70}$$

In that case, the Feynman rules simplify considerably. For the interaction Hamiltonian of Eq. (70), the Feynman rules are now

(1) unintegrated x-coordinate: $\exp(-ikx)$ (if real-space functions are to be evaluated),
(2) propagator: $P_D(k)$,
(3) data-source vertex: $(j + J - \lambda_1)(k)$,
(4) vertex with $m > 1$ lines: $-\lambda_m$,
(5) imply momentum conservation at each vertex: $(2\pi)^n \delta(\sum_{i=1}^m k_i)$, and integrate over every internal momentum: $\int \frac{dk}{(2\pi)^n}$,
(6) and divide by the symmetry factor.

5. Feynman rules on the sphere

For CMB reconstruction and analysis, but presumably also for terrestrial applications, the Feynman rules on the sphere $\Omega = S^2$ are needed and therefore provided in Appendix B.

B. Normalizability of the theory

In contrast to QFT, IFT should be properly normalized and not necessarily require any renormalization procedure. The reason is that IFT is not a low-energy limit of some unknown high-energy theory, but can be set up as the full (high-energy) theory. The Hamiltonian is just the logarithm of the joint probability function of data and signal, $H_d[s] \equiv -\log[P(d, s)]$, and therefore well defined and properly normalized if the latter is. Only if *ad hoc* Hamiltonians are set up, or if approximations lead to ill-normalized theories, should normalization be an issue.

However, since we are trying a perturbative expansion of the theory, there is no guarantee that all individual terms are providing finite results. For example in QFT, simple loop diagrams are known to be divergent and require renormalization. In the following we investigate a simplistic, but representative case of IFT, which shows that such problems are generally not to be expected.

Let us adopt the simplistic situation described in Sec. IV A 4 and estimate a simple loop diagram for which we assume for notational convenience $\lambda_3 = -2(2\pi)^n \lambda'$ (with $\lambda' > 0$):

$$\begin{aligned} \text{—}\bigcirc &= -\frac{1}{2} D\, \lambda_3\, \widehat{D} \\ &= \lambda' \int dk \int dk'\, \delta(k + k' - k')\, P_D(k)\, P_D(k')\, e^{ikx} \\ &\leq \lambda'\, P_D(0) \int dk'\, P_S(k') = \lambda'\, V\, P_D(0)\, \langle s^2(x) \rangle, \end{aligned} \tag{71}$$

where V is the volume of the system. Here and in the following, $\hat{C}$ denotes the diagonal of the matrix C.

Thus, as long the signal field is of bounded variance, the loop diagram is convergent due to $P_D \leq P_S$ for all k. Even a signal of unbounded variance would not lead to a divergent loop diagram if $\int dk(P_N/P_R)(k)$ is finite, since we also have $P_D \leq P_N/P_R$. A bounded variance signal is very natural, especially in a cosmological setting.[7]

Finally, since a signal as an information field can be chosen freely, we can define it to be the filtered version of the physical field (e.g. dark-matter distribution or CMB fluctuations), so that only modes of sufficiently bound variance are present in it. Since we have the freedom to choose information fields, which are mathematically well behaved, we can therefore ensure the convergence of expressions.

Although this is not a general proof of normalizability of the theory, which is beyond the scope of this paper, it should provide confidence in the well behavedness of the formalism in sensible applications. The price to be paid for this well behavedness is the more complex structure of the propagator, which, in comparison to QFT, even in simplistic cases can be nonanalytical and require numerical evaluation.

C. Expansion around the classical solution

1. General case

The classical solution of the Hamiltonian in Eq. (59) is provided by its minimum,

$$\frac{\delta H}{\delta s_x} = D^{-1}_{xy} s_y - j_x + \sum_{m=1}^{\infty} \frac{1}{m!} \Lambda^{(m+1)}_{xx_1\cdots x_m} s_{x_1} \cdots s_{x_m} = 0. \tag{72}$$

This leads to the equation for the classical field

$$s^{cl}_y = D_{yx}\left(j_x - \sum_{m=1}^{\infty} \frac{1}{m!} \Lambda^{(m+1)}_{xx_1\cdots x_m} s^{cl}_{x_1} \cdots s^{cl}_{x_m}\right), \tag{73}$$

which one can try to solve iteratively.

2. Local interactions

For simplicity, we concentrate for a moment on the case of purely local interactions, for which the equation for the classical field $s_{\rm cl}$ is

$$s_{\rm cl} = D\left(j - \sum_{m=1}^{\infty} \frac{\lambda^{\dagger}_{m+1}}{m!} s^m_{\rm cl}\right). \tag{74}$$

Iterating this equation and rewriting the resulting terms as Feynman diagrams shows that the classical solution contains the tree diagrams. The loop diagrams can be added by investigation of the nonclassical uncertainty field $\phi = s - s_{\rm cl}$.

A nonclassical expansion of the information field around the classical field is possible by inserting $s = s_{\rm cl} + \phi$ into the Hamiltonian [Eq. (64)]. Reordering terms according to the powers of the field ϕ leads to its Hamiltonian

$$\begin{aligned} H'[\phi] &\equiv H[s_{\rm cl} + \phi] \\ &= H'_0 + \frac{1}{2}\phi^{\dagger} D'^{-1}\phi - j'^{\dagger}\phi + \sum_{m=3}^{\infty} \frac{1}{m!} \lambda'^{\dagger}_m \phi^m, \end{aligned}$$

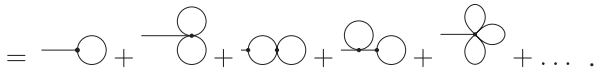
$$\text{with} \quad \lambda'_n \equiv \sum_{m=0}^{\infty} \frac{\lambda_{n+m}}{m!} s^m_{\rm cl},$$

$$H'_0 \equiv H[s_{\rm cl}] = H_0 + \frac{1}{2} s^{\dagger}_{\rm cl} D^{-1} s_{\rm cl} + \lambda'_0,$$

$$j' \equiv j - \lambda'_1 - D^{-1} s_{\rm cl}, \quad \text{and} \quad D' \equiv (D^{-1} + \hat{\lambda}'_2)^{-1}. \tag{75}$$

In case $s_{\rm cl}$ is exactly the classical solution, Eqs. (74) and (75) imply that $j' = 0$. Thus, there are no one-line internal vertices in any Feynman graphs of the ϕ theory, and only loop diagrams contribute uncertainty corrections[8] to any information theoretical estimator. For example, the uncertainty corrections to the classical map estimator are given by

$$\delta m = m_d - s_{\rm cl} = \langle \phi \rangle_d$$



(76)

However, in case $s_{\rm cl}$ is not (exactly) the classical solution, may this be due to a truncation error of an iteration scheme to solve for the classical field, or may $s_{\rm cl}$ be chosen for a completely different purpose, Eq. (75) provides the correct field theory for $\phi = s - s_{\rm cl}$ independent of the nature of $s_{\rm cl}$. In case of a truncation error, incorporating diagrams with data-source terms j' into any computation will permit one to correct the inaccuracy of $s_{\rm cl}$ in a systematic way.

D. Boltzmann-Shannon information

1. Helmholtz free energy

Information fields carry information on distributed physical quantities. The amount of signal information should be measurable in information units like bits and

[7]The cosmological signal of primary interest, the initial density fluctuations as revealed by the large-scale-structure and the CMB, is expected to exhibit a suppression of small-scale power due to the freestreaming of dark-matter particles before they became nonrelativistic. Also the CMB temperature fluctuations are damped on small scales, due to freestreaming of photons around the time of recombination.

[8]We propose the term *uncertainty corrections* in order to describe the influence of the spread of the probability distribution function around its maximum. The uncertainty corrections are the information field theoretical equivalent to quantum corrections in quantum field theories.

bytes. This is possible by adopting the Boltzmann-Shannon information measure of negative entropy. The entropy of a signal probability function measures the phase-space volume available for signal uncertainties, and therefore the constraintness of the remaining uncertainties. Thus we define

$$I_d \equiv \int \mathcal{D}s P(s|d) \log P(s|d) = -\int \mathcal{D}s \frac{1}{Z} e^{-H[s]}(H[s] + \log Z) = -\langle H[s]\rangle_d - \log Z \tag{77}$$

as the information measure. Introducing

$$Z_\beta[d, J] = \int \mathcal{D}s \exp\{-\beta(H[s] - J^\dagger s)\}, \quad \text{and}$$
$$F_\beta[d, J] = -\frac{1}{\beta}\log Z_\beta[d, J], \tag{78}$$

of which the latter is the Helmholtz free energy as a function of the inverse temperature β, we can write

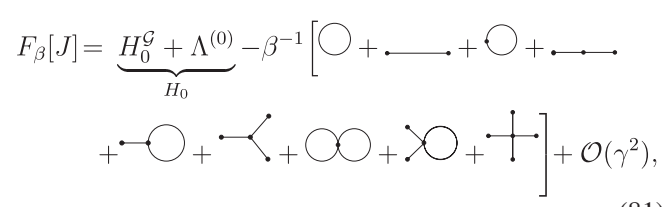
$$I_d = -\log Z_1[d, 0] - \langle H[s]\rangle_d = -\frac{\partial F_\beta[d, J]}{\partial \beta}\bigg|_{\beta=1, J=0}, \tag{79}$$

as can be verified by a direct calculation. The first expression for I_d in Eq. (79) is equivalent to the well-known thermodynamic relation $F = E - TS_B$ with the internal energy $E = \langle H[s]\rangle_d$, the Boltzmann entropy $S_B = -I_d$, and the temperature, which is set here to $T = 1$. The second expression actually holds even if the Hamiltonian is improperly normalized, e.g. H_0 can be chosen arbitrarily if $Z_\beta[d, J]$ is calculated consistently with this choice.

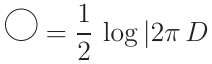
The Helmholtz free energy $F_\beta[J]$ is also the generator of all connected correlation functions of the signal $\langle s_{x_1} \cdots s_{x_n}\rangle^c_{(s|d)} = -\delta^n F_\beta[d, J]/\delta J_{x_1} \cdots \delta J_{x_n}|_{\beta=1, J=0}$. It can be calculated as follows:

$$F_\beta = -\frac{1}{\beta}\log\left(\frac{Z^{\mathcal{G}}_\beta[J]}{Z^{\mathcal{G}}_\beta[J]}\int \mathcal{D}s e^{-\beta H_{\text{int}}[s]} e^{-\beta(H_{\mathcal{G}}[s] - J^\dagger s)}\right) = -\frac{1}{\beta}\log Z^{\mathcal{G}}_\beta[J] - \frac{1}{\beta}\log\langle e^{-\beta H_{\text{int}}[s]}\rangle_{(s|J+j, \mathcal{G})}, \tag{80}$$

where the average in the last term is over the Gaussian probability function $P^{\mathcal{G}}_{J,\beta}[s] = \exp(-\beta(H_{\mathcal{G}}[s] - J^\dagger s))/Z^{\mathcal{G}}_\beta[J]$. This term can be calculated by using the well-known fact that the logarithm of the sum of all possible connected and unconnected diagrams with only internal coordinates (or without free ends), as generated by the exponential function of the interaction terms, is given by the sum of all connected diagrams [57]. For example, a free theory, perturbed by small, up-to-fourth-order interaction terms (all being proportional to some small parameter γ), has

$F_\beta[J] =$  $+ \mathcal{O}(\gamma^2)$, (81)

where an information source vertex reads $\beta(J + j - \Lambda^{(1)})$, an internal vertex with n lines $\beta\Lambda^{(n)}$, and the propagator $\beta^{-1}D$. Finally, we have defined



Thus, we have

$$\begin{aligned} F_\beta[J] = H_0 &- \frac{1}{2\beta}\mathrm{Tr}(\log(2\pi D\beta^{-1})) + \frac{1}{2\beta}\Lambda^{(2)}[D] \\ &+ \frac{1}{2}(J + j - \Lambda^{(1)})^\dagger(D + \Lambda^{(2)})(J + j - \Lambda^{(1)}) \\ &+ \frac{1}{2\beta}\Lambda^{(3)}[D, m_J] + \frac{1}{3!}\Lambda^{(3)}[m_J, m_J, m_J] \\ &+ \frac{1}{8\beta^2}\Lambda^{(4)}[D, D] + \frac{1}{4\beta}\Lambda^{(4)}[D, m_J, m_J] \\ &+ \frac{1}{4!}\Lambda^{(4)}[m_J, m_J, m_J, m_J] + \mathcal{O}(\gamma^2), \end{aligned} \tag{82}$$

where we introduced the zero-order map $m_J = D(J + j)$ for notational convenience. The power of β associated with the different diagrams in Eq. (81) is given by the number of vertices minus the number of propagators minus 1. Thus, all tree diagrams are of order β^0, the one-loop diagrams are of order β^{-1}, and the two-loop diagram of order β^{-2}, and only the latter two affect the information content:

$$\begin{aligned} I_d &= -\left[\frac{\varrho}{2} + \bigcirc + \bigcirc + \bigcirc\!\!-\!\!\bigcirc + \bigcirc\!\!\bigcirc + \bigcirc\!\!\times\!\!\bigcirc\right] + \mathcal{O}(\gamma^2) \\ &= \frac{1}{2}\left[-\mathrm{Tr}(1 + \log(2\pi D)) + \Lambda^{(2)}[D] + \Lambda^{(3)}[D, m_0] + \frac{1}{2}\Lambda^{(4)}[D \otimes (D + m_0 m_0^\dagger)]\right] + \mathcal{O}(\gamma^2), \end{aligned} \tag{83}$$

where $\varrho = \mathrm{Tr}(1)$, $\beta = 1$, $J = 0$, and thus $m_0 = Dj$.[9]

2. Free theory

To obtain the information content of the free theory, we can set $\gamma = 0$ in Eqs. (82) and (83) or use Eq. (49) with the

[9]Here, we introduced the symmetrized tensor product $A \otimes B$ of an n-rank tensor A and an m-rank tensor B, which has the property

$$(A \otimes B)_{x_1 \cdots x_{n+m}} = \frac{1}{m!}\sum_{\pi \in \mathcal{P}_{n+m}} A_{x_{\pi(1)} \cdots x_{\pi(n)}} B_{x_{\pi(n+1)} \cdots x_{\pi(n+m)}},$$

with $\mathcal{P}_l$ being the set of permutations of $\{1, \ldots, l\}$.

replacements $J \to \beta J$, $j \to \beta j$, $D \to \beta^{-1} D$, and $H_0 \to \beta H_0$. In both cases we find identically

$$F_\beta[J] = H_0^{\mathcal{G}} - \frac{1}{2}(J+j)^\dagger D(J+j) - \frac{1}{2\beta}\,\mathrm{Tr}\log\left(\frac{2\pi}{\beta}D\right),$$

$$\text{and} \quad I_d = -\frac{1}{2}\,\mathrm{Tr}(1 + \log(2\pi D)). \tag{84}$$

Very similarly, one can calculate the information prior to the data, which turns out to be

$$I_0 = -\tfrac{1}{2}\mathrm{Tr}(1 + \log(2\pi S)). \tag{85}$$

Thus, the data-induced information gain is

$$\begin{aligned}\Delta I_d &= I_d - I_0 = \tfrac{1}{2}\mathrm{Tr}(\log(SD^{-1})) \\ &= \tfrac{1}{2}\mathrm{Tr}(\log(1 + SR^\dagger N^{-1}R)).\end{aligned} \tag{86}$$

The information gain depends on the signal-response-to-noise ratio $Q \equiv RSR^\dagger N^{-1}$, also shortly denoted by the measurement fidelity or quality. The information increases linearly with Q as long as $Q \ll 1$, but levels off to a logarithmic increase for $Q \gg 1$.

We note that for the free theory only the information gain does not depend on the actual data realization.

E. IFT recipe

A typical IFT application will aim at calculating a model evidence $P(d)$, the expectation value of a signal given the data, the map $m(x) = \langle s(x)\rangle_{(s|d)}$ of the signal, or its variance $\sigma_s^2(x,y) = \langle (s(x) - m(x))(s(y) - m(y))\rangle_{(s|d)}$ as a measure of the signal uncertainty. The general recipe for such applications can be summarized as follows:

(i) Specify the signal s and its prior probability distribution $P(s)$. If the signal is derived from a physical field ψ, of which a prior statistic is known, the distribution of $s = s[\psi]$ is induced according to Eq. (2).

(ii) Specify the data model in terms of a likelihood $P(d|s)$ conditioned on s. Again, if the data are related to an underlying physical field ψ, the likelihood is given by Eq. (4).

(iii) Calculate the Hamiltonian $H_d[s] = -\log(P(d,s))$, where $P(d,s) = P(d|s)P(s)$ is the joint probability, and expand it in a Taylor-Fréchet series for all degrees of freedom of s. Identify the coefficients of the constant, linear, quadratic, and nth-order terms with the normalization H_0, information source j, inverse propagator D^{-1}, and nth-order interaction term $\Lambda^{(n)}$, respectively, as shown in Eq. (36) or (59).

(iv) Draw all diagrams, which contribute to the quantity of interest, consisting of vertices, lines, and open ends up to some order in complexity or some small ordering parameter. The log evidence is given by the sum of all connected diagrams without open ends, the expectation value of the signal by all connected diagrams with one open end, and the signal variance around this mean by all connected diagrams with two open ends.

(v) Read the diagrams as computational algorithms specified by the Feynman rules in Sec. IV, and implement them by using linear algebra packages or existing map-making codes for the information propagator and vertices. The required discretization is outlined in Sec. II E 1. Information on how to implement the required matrix inversions efficiently can be found in the literature given in Secs. I C 2, I C 4, and I C 5 and especially in [38].

(vi) If the resulting nonlinear data transformation (or filter) has the required accuracy, e.g. to be verified via Monte Carlo simulations using signal and data realizations drawn from the prior and likelihood, respectively, an IFT algorithm is established.

(vii) In case that too large interaction terms in the Hamiltonian prevent a finite number of diagrams to form a well-performing algorithm, a resummation of high order terms is due. This can be achieved by the saddle point approximation (classical solution, maximum *a posteriori* estimator), or even better by a detailed renormalization-flow analysis along the lines outlined in Sec. V F.

V. COSMIC LARGE-SCALE STRUCTURE VIA GALAXY SURVEYS

A. Poissonian data model and Hamiltonian

Many data sets suffer from Poisson noise, which is non-Gaussian and signal dependent, and therefore well suited to test IFT in the nonlinear regime. For example, the cosmological LSS is traced by galaxies, which may be assumed to be generated by a Poisson process. On large scales, the expectation value of the galaxy density follows that of the underlying (dark) matter distribution. The aim of cosmography is to recover the initial density field from the shot-noise contaminated galaxy data. Currently, large galaxy surveys are conducted in order to chart the cosmic matter distribution in three dimensions. Improving the galaxy based LSS reconstruction techniques and understanding their uncertainties better is therefore an imminent and important goal. Optimal techniques to reconstruct Poissonian-noise affected signals are also crucial for other problems, since e.g. imaging with photon detectors plays an important role in astronomy and other fields. Here, we outline how such problems can be treated, by discussing a specific data model motivated by the problem of large-scale-structure reconstruction from galaxies. For this problem we work out the optimal estimator and show its superiority numerically. A more general discussion of models of galaxy and structure formation and references to relevant works was given in Sec. I C 4.

In order to treat the Poissonian case in a convenient fashion, we subdivide the physical space into small cells with volumes ΔV, and assume that a cell located at x_i has

an expected number of observed galaxies

$$\mu_i \approx \kappa(1 + bs(x_i)), \tag{87}$$

with $\kappa = \bar{n}_g \Delta V$ being the cosmic average number of galaxies per cell and b being the bias of the galaxy overdensity with respect to the dark-matter overdensity s, still assumed to be a Gaussian random field [Eq. (38)]. However, this data model has two shortcomings. First, too negative fluctuations of the Gaussian random field with $s < -1$ lead to negative expectation values, for which the Poissonian statistics is not defined. Second, the mean density of observable galaxies κ and their bias parameter b are constant everywhere, whereas in reality both exhibit spatial variations.[10] Although being now spatially inhomogeneous, we assume κ and b to be known for the moment and to incorporate all the above observational effects.

To cure the above mentioned shortcomings we replace Eq. (87) by a nonlinear and nontranslational invariant model:

$$\mu_i = \kappa(x_i) \exp(b(x_i)s(x_i)), \tag{88}$$

where κ and b may depend on position in a known way, and the unknown Gaussian field s, the log-matter density, may exhibit unrestricted negative fluctuations. Note that μ is the signal response, by our definition in Eq. (10), since $\mu[s] = \langle d \rangle_{(d|s)}$. We call κ the *zero response*, since $\mu[0] = \kappa$. It should be stressed that the data model in Eq. (88) is just a convenient choice for illustration and proof-of-concept purposes, and is easily exchangeable with more realistic, and even nonlocal data models. However, this log-normal data model was originally proposed by Coles and Jones [212], investigated for constrained realizations by Sheth [107] and Vio *et al.* [213] and seems to reproduce the statistics of LSS simulations much better than the often used normal distribution of the overdensity [214].

Having chosen a Poissonian process to populate the Universe and our observational data with galaxies according to the underlying log-density field s, the likelihood is

$$\begin{aligned} P(d|s) &= \prod_i \frac{\mu_i^{d_i}}{d_i!} e^{-\mu_i} \\ &= \exp\left\{\sum_i [d_i \log \mu_i - \mu_i - \log(d_i!)]\right\}, \end{aligned} \tag{89}$$

where d_i is the actual number of galaxies observed in cell i. Since $P(s) = \mathcal{G}(s, S)$, the Hamiltonian is given by

$$\begin{aligned} H_d[s] &= -\log P(d, s) = -\log P(d|s) - \log P(s) \\ &= -d^\dagger bs + \kappa^\dagger \exp(bs) + H_0' + \frac{1}{2} s^\dagger S^{-1} s \\ &= \frac{1}{2} s^\dagger D^{-1} s - j^\dagger s + H_0 + \sum_{n=3}^{\infty} \frac{1}{n!} \lambda_n^\dagger s^n, \quad \text{with} \\ D^{-1} &= S^{-1} + \widehat{\kappa b^2}, \\ j &= b(d - \kappa), \\ H_0 &= \frac{1}{2} \log(|2\pi S|) + (\kappa + \log(d!))^\dagger 1 - d^\dagger \log \kappa, \quad \text{and} \\ \lambda_n &= \kappa b^n. \end{aligned} \tag{90}$$

The hat on a scalar field denotes that it should be read as a matrix, which is diagonal in position space (see Appendix A). A few remarks are in order. Comparing the propagator to the one of our Gaussian theory one can read off an inverse noise term $M = R^\dagger N^{-1} R = \widehat{\kappa b^2}$. Thus the effective (inversely response weighted) noise decreases with increasing mean galaxy number and bias, and seems to be infinite in regions without data ($\kappa = 0$) without causing any problem for the formalism.

The information source j increases with increasing response (bias) of the data (galaxies) to the signal (density fluctuations). However, it certainly vanishes for zero response ($b = 0$) or in case that the observed galaxy counts match the expected mean at a given location exactly. Finally, the interaction terms λ_n are local in position space, and vanish with decreasing b and κ. The latter parameter is under the control of the data analyst, since it is proportional to the volume of the individual pixel sizes, and therefore can be made arbitrarily small by choosing a more fine grained resolution in signal space. However, this would not change the convergence properties of the series since any interaction vertex has then to be summed over a correspondingly larger number of pixels within a coherence patch of the signal, which exactly compensates for the smaller coefficient.[11] The bias, in contrast, is set by nature and can be regarded as a power counting parameter, which provides naturally a numerical hierarchy among the higher-order vertices and diagrams for $b^2 S < 1$. Note that $j = \mathcal{O}(b)$.

B. Galaxy types and bias variations

Real galaxies can be cast into different classes, which all differ in terms of their luminosities, bias factors, and the frequencies with which they are found in the Universe.

[10] Such variations are due to the geometry of the observational survey sky coverage, due to a galaxy selection function which decreases with distance from the observer, and due to a changing composition of the galaxy population. The latter distance effects are caused by the cosmic evolution of galaxies and by the changing observational detectability of the different types with distance. We note that an observed sample of galaxies, which was selected deterministically or stochastically from a complete sample e.g. by their luminosity due to instrumental sensitivity, still possesses a Poissonian statistics, if the original distribution does.

[11] κ seems to control the stiffness of the later introduced response-renormalization flow equation and its values are therefore numerically relevant. A lower κ, due to a finer space pixelization, results in a less stiff and better behaved equation.

Although we are not going to investigate this complication in the following, it should be explained here how all the formulas in this section can easily be reinterpreted, in order to incorporate also the different classes of galaxies.

The galaxies can be characterized by a type variable $L \in \Omega_{\text{type}}$, which may be the intrinsic luminosity, the morphological galaxy type, or a multidimensional combination of all properties which determine the galaxy type's spatial distributions via a L-dependent bias b_L, and their detectability as encoded in κ_L. The data space is now spanned by $\Omega_{\text{data}} = \Omega_{\text{space}} \times \Omega_{\text{type}}$, and also μ, κ, and b can be regarded as functions over this space.

Performing the same algebra as in the previous section, just taking the larger data space into account, we get to exactly the same Hamiltonian, as in Eq. (90), if we interpret any term containing d, κ, and b to be summed or integrated over the type variable L. Thus, we read

$$\begin{aligned} j(x) &= (b(d-\kappa))(x) \equiv \int dL\, b_L(x)(d_L(x) - \kappa_L(x)), \\ D^{-1}_{xy} &= (S^{-1} + \widehat{\kappa b^2})_{xy} \equiv S^{-1}_{xy} + 1_{xy} \int dL\, \kappa_L(x) b^2_L(x), \\ \lambda_n(x) &= (\kappa b^n)(x) \equiv \int dL\, \kappa_L(x) b^n_L(x), \quad \text{and} \\ \mu[s](x) &= (\kappa e^{bs})(x) \equiv \int dL\, \kappa_L(x) e^{b_L(x)s(x)} \\ &= \int dL\, \mu_L[s](x), \end{aligned} \tag{91}$$

which all live in Ω_{space} solely, so that the computational complexity of the matter distribution reconstruction problem is not affected at all, and only a bit more bookkeeping is required in its setup.

A few observations should be in order. In case of all galaxies having the same bias factor, Eq. (91) is simply a marginalization of the type variable L, and any differentiation of the various galaxy types is not necessary. Since all known galaxy types seem to have $b \sim \mathcal{O}(1)$, such a marginalization seems to be justified, and explains why LSS reconstructions, which applied this simplification, are relatively successful, although the different galaxy masses, luminosities, and frequencies vary by orders of magnitude. As our numerical experiments below reveal, the data, and therefore the reconstructability of the density field, exhibit a sensitive dependence on the bias for s fluctuations with unity variance.[12] Such a variance is indeed observed on scales below $\sim$10 Mpc in the galaxy distribution, and therefore the galaxy type-dependent bias variation does indeed matter. Larger galaxies, which have larger biases, therefore provide per galaxy a slightly larger information source ($j \propto b$), less shot noise ($R^\dagger N^{-1} R \propto b^2$), and increasingly larger higher-order interaction terms ($\lambda_n \propto b^n$) in comparison to smaller galaxies. However, smaller galaxies are much more numerous by orders of magnitude, and therefore provide the largest total contribution to the information source, noise reduction, and most low-order interaction terms. Thus, the latter will dominate and therefore permit a reasonable accurate matter reconstruction from an inhomogeneous galaxy survey using a single bias value. Nevertheless, improvements of the bias treatment are possible by applying the recipes described here.

[12]This is found for our specific data model $\mu \propto \exp(bs)$; however, it should also apply for other models, which somehow have to keep $\mu \geq 0$ even for $bs < -1$

C. Nonlinear map making

The map, the expectation of our information field s given the data, is to the lowest order in interaction

$$\begin{aligned} m_1 &= \quad + \quad + \quad + \quad + \mathcal{O}(b^6) \\ &= D_{xy}\, j_y - \frac{1}{2} D_{xy}\, b^3_y\, \kappa_y\, D_{yy} - \frac{1}{2} D_{xy}\, b^3_y\, \kappa_y\, (D_{yz}\, j_z)^2 \\ &\quad - \frac{1}{2} D_{xy}\, b^4_y\, \kappa_y\, D_{yy}\, D_{yz}\, j_z + \mathcal{O}(b^6) \end{aligned} \tag{92}$$

or in compact notation

$$m_1 = m_0 - \tfrac{1}{2} D \widehat{b^3 \kappa} (\hat{D} + m_0^2 + \widehat{\hat{D} b m_0}) + \mathcal{O}(b^6). \tag{93}$$

It is apparent that the nonlinear map-making formula contains corrections to the linear map $m_0 = Dj$. The first two correction terms are always negative, reflecting the fact that our nonlinear data model has nonsymmetric fluctuations in the data with respect to the mean. The last correction term is oppositely directed to the linear map, thereby correcting for the curvature in the signal response.

A one-dimensional, numerical example is displayed in Fig. 1. There, the signal was realized to have a power spectrum $P_s(k) \propto (k^2 + q^2)^{-1}$, with a correlation length $q^{-1} = 0.04$. The normalization was chosen such that the autocorrelation function is $\langle s(x) s(x+r) \rangle_{(s)} = \exp(-|qr|)$ and therefore the signal dispersion is unity, $\langle s^2 \rangle_{(s)} = 1$. The data are generated by a Poissonian process from $\kappa_s = \kappa \exp(bs)$ with $b = 0.5$. All three displayed reconstructions exhibit less power than the original signal, as it is expected since the reconstruction is conservative, and therefore biased toward zero.

The nonlinear correction to the naive map m_0 should not be too large, otherwise higher-order diagrams have to be included. In the case displayed in Fig. 1, $b = 0.5$ ensured that the linear corrections were mostly going in the right direction. However, in case $b \approx 1$ there is no obvious ordering of the importance of the different interaction vertices, and numerical experiments reveal that the first-order corrections strongly overcorrect the linear map $m_0 = Dj$. In such a case interaction resummation techniques should be used to incorporate as many higher-order interaction terms as possible. One very powerful resummation is provided by the classical solution, as developed below,

which contains all tree diagrams simultaneously. This solution, also shown in Fig. 1, is very close to m_1 in this case.

D. Classical solution

The classical signal field or MAP solution is given by Eq. (74), which reads in this case

$$s_{\rm cl} = D\left(j - \sum_{m=2}^{\infty} \frac{b^{m+1}}{m!} \kappa s_{\rm cl}^m\right) = Db(d - \kappa(e^{bs_{\rm cl}} - bs_{\rm cl})) = Sb(d - \underbrace{\kappa e^{bs_{\rm cl}}}_{\kappa_{s_{\rm cl}}}). \tag{94}$$

The last expression motivates one to introduce the expected number of galaxies given the signal s:

$$\kappa_s = \kappa e^{bs}. \tag{95}$$

Also alternative forms of the MAP equation can be derived, for example, one, which is especially suitable for large j:

$$s_{\rm cl} = \frac{1}{b} \log\left[\frac{j - S^{-1}s_{\rm cl}}{\kappa b}\right] = \frac{1}{b} \log\left[\frac{d}{\kappa} - 1 - \frac{S^{-1}s_{\rm cl}}{\kappa b}\right]. \tag{96}$$

This may be solved iteratively, while ensuring that $s_{\rm cl}^{(i)} \leq Sj$ at all iterations i with equality only where $\kappa = 0$. This form of the classical field equation has some similarities to the naive inversion of the response formula, $\langle d \rangle_{(d|s)} = \kappa \exp(bs)$, which yields

$$s_{\rm naive} = \frac{1}{b} \log\left[\frac{d}{\kappa}\right], \tag{97}$$

a formula one can only dare to use in regimes of large d.

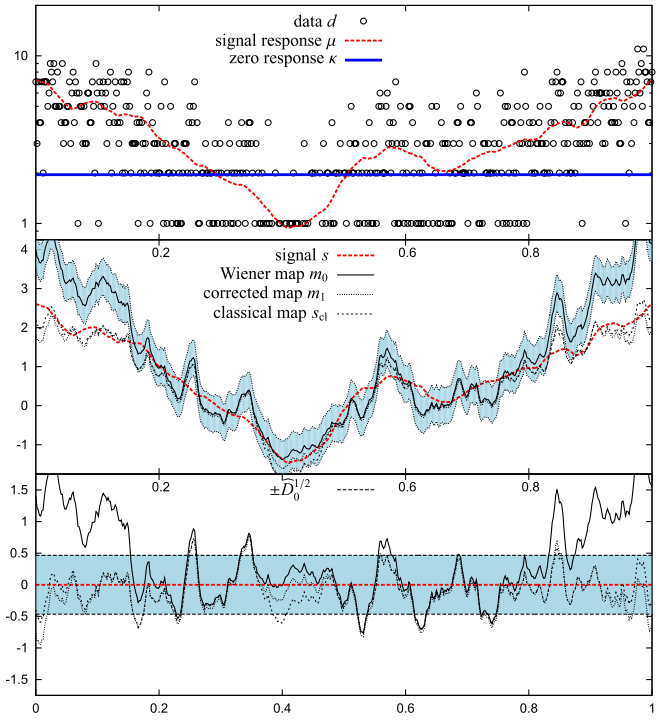


FIG. 1 (color online). Poissonian reconstruction of a signal with unit variance and correlation length $q^{-1} = 0.04$, observed with slightly nonlinear response ($b = 0.5$, resolution: 513 pixels per unit length, zero-signal galaxy density: 1000 galaxies per unit length). Top panel: data d, signal response μ, and zero response κ. Middle panel: signal s, linear Wiener-filter reconstruction $m_0 = Dj$, its 1σ error interval $m_0 \pm \hat{D}^{1/2}$, next order reconstruction m_1 according to Eq. (92), and classical solution $s_{\rm cl}$ according to Eq. (94). Although the linear Wiener filter is reconstructing well at most locations, the nonlinear response requires the perturbative corrections present in m_1 or the classical solution in regions of high signal strength. Bottom panel: The residuals, the deviations of m_0, m_1, $s_{\rm cl}$ from the signal, and the Wiener variance $\pm\hat{D}^{1/2}$.

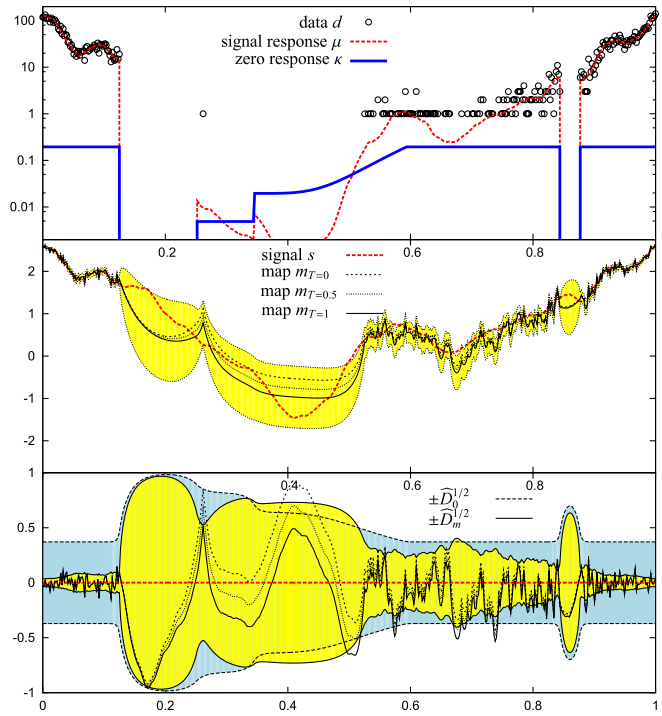


FIG. 2 (color online). Poissonian reconstruction of the same signal realization as in Fig. 1 (unit variance and correlation length $q^{-1} = 0.04$), observed now with a strongly nonlinear response ($b = 2.5$, resolution: 512 pixels per unit length, zero-signal galaxy density: 100 galaxies per unit length where the mask is 1) through a complicated mask. Top panel: data d, signal response μ, and zero response κ. Middle panel: signal s, classical solution $s_{\rm cl} = m_{T=0}$, intermediate solution $m_{T=0.5}$, and renormalization-based reconstruction $m_{T=1}$ with uncertainty interval $m_{T=1} \pm \hat{D}_{T=1}^{1/2}$, and mask $\kappa/(n_{\rm g}\Delta V)$. The linear Wiener-filter reconstruction m_0 as well as its next order corrected version m_1 are not displayed, since they are partly far outside of the displayed area. Bottom panel: Deviations of the three reconstructions from the signal, and the original and the renormalized uncertainty estimates $\pm\hat{D}_0^{1/2}$ and $\pm\hat{D}_{T=1}^{1/2}$, respectively. Note that in the regions with many observed galaxies, the high signal-to-noise ratio can be seen in the narrowness of $\hat{D}_{T=1}^{1/2}$, which is significantly smaller than the data unaffected $\hat{D}_0^{1/2}$ at these locations.

Since $s_{\rm naive}$ contains the full noise of the data, a suitable naive map may be given by $m_{\rm naive} = Ss_{\rm naive}$, after some fix for the locations without galaxy counts. The classical solution, however, is more conservative than this naive data inversion, in that there is a damping term, $S^{-1}s_{\rm cl}/(\kappa b)$, compensating a bit the influence of too large data points.

Those equations permit one to calculate the classical solution if suitable numerical regularization schemes are applied, since naive iterations can easily lead to numerical divergences in the nonlinear case.

One way of doing this is by turning the classical equation [Eq. (94)] into a dynamical system. Its initial conditions are given by a well-solvable linear or even trivial problem to which nonlinear complications are added successively during an interval of some pseudotime. The end point of this dynamics is then the required solution. The meaning of the pseudotime depends on the way it was set up. In any case, it can just be regarded as a mathematical trick to generate a differential equation, which might be easier to solve numerically than the original problem.

For example, a pseudotime τ can be introduced by setting $j(\tau) = \tau j$. Thus, the information source is successively injected into an initially trivial field state, $s_{\rm cl}(0) = 0$. This allows one to set up a differential equation for $s_{\rm cl}(\tau)$ by taking the time derivative of Eq. (94),

$$\dot{s}_{\rm cl} = D_{s_{\rm cl}} j \quad \text{with} \quad D_{s_{\rm cl}} = (S^{-1} + \kappa_{s_{\rm cl}} b^2)^{-1}, \tag{98}$$

which has to be solved for $s_{\rm cl}(1)$ starting from $s_{\rm cl}(0) = 0$. This equation is very appealing, since it looks like Wiener filtering an incoming information stream j and accumulating the filtered data, while simultaneously tuning the filter $D_{s_{\rm cl}(\tau)}$ to the accumulated knowledge on the signal $s_{\rm cl}(\tau)$ and thereby implied Poissonian-noise structure. Thus, it is a nice example system for continuous Bayesian learning and also illustrates how different data sets can successively be fused into a single knowledge basis.

Map-making algorithms with a higher fidelity are possible by not only investigating the maximum of the posterior, but by averaging the signal s over the full support of $P(s|d)$. Anyhow, we can assume that a good approximation $t \approx s_{\rm cl}$ to the classical solution can be achieved. Figures 1 and 2 display classical solutions for slightly and strongly nonlinear Poissonian inference problems. Especially the second example shows that the classical solution can be improved in regions of large uncertainty (see the region between $x = 0.2$ and 0.5 in Fig. 2, where apparently better estimators exist) for missing uncertainty-loop diagrams, which contain information about the non-Gaussian structure of the posterior $P(s|d)$ away from $s_{\rm cl}$.

E. Uncertainty-loop corrections

Now, we see how the missing uncertainty-loop corrections can be added to the classical solution. These corrections can be derived from the Hamiltonian of the uncertainty field $\phi = s - t$,

$$H_t[\phi] = \frac{1}{2}\phi^\dagger D_t^{-1}\phi - j_t^\dagger\phi + \kappa_t^\dagger g(b\phi) + H_{0,t},$$

where $D_t^{-1} = S^{-1} + b^2\hat{\kappa}_t$,

$$j_t = b(d - \kappa_t) - S^{-1}t,$$

$$g(x) = e^x - 1 - x - \frac{1}{2}x^2 = \sum_{m=3}^{\infty}\frac{x^m}{m!}, \tag{99}$$

and $H_{0,t}$ is a momentarily irrelevant normalization constant. Again, we have permitted for a nonzero j_t, since t might not be exactly the classical solution.

It is interesting to note that the interaction coefficients in this Hamiltonian, $\lambda_t^{(m)} = \kappa_t b^m$, all reflect the expected number of galaxies given the reference field t. Thus, the replacement $\kappa_0 \rightarrow \kappa_t$ would provide us with the shifted field Hamiltonian, as defined in Eq. (60), except for the term $-S^{-1}t$ in j_t. It turns out that this term is some sort of counterterm, which accumulates the effect of the nonlinear interactions.

We see that effective interaction terms arise when relevant parts of the solution are absorbed in the background field t. A similar approach is desirable for the loop diagrams. Instead of drawing and calculating all possible loop diagrams, we want to absorb several of them simultaneously into effective coefficients. For each vertex of the Poissonian Hamiltonian with m legs, there exist diagrams in any Feynman expansion, in which a number of n simple loops are added to this vertex. Such an n-loop enhanced m vertex is given by

$$= \frac{-1}{2^n\, n!}\,\lambda_t^{(m+2n)}\,\widehat{D}_t^{\,n} = \frac{-1}{2^n\, n!}\,\kappa_t\, b^{m+2n}\,\widehat{D}_t^{\,n}. \tag{100}$$

All these diagrams can be resummed into an effective interaction vertex, via

$$\lambda_t^{(m)} \rightarrow \lambda_t^{\prime(m)} = \kappa_t b^m \sum_{n=0}^{\infty}\frac{1}{2^n n!}b^{2n}\hat{D}^n = \kappa_t \exp\left(\frac{b^2}{2}\hat{D}\right)b^m$$
$$= \kappa_{t+b\hat{D}/2}\,b^m = \lambda_{t+b\hat{D}/2}^{(m)}. \tag{101}$$

Thus, this resummation is effectively equivalent to the replacement

$$\kappa_t \rightarrow \kappa_{t+b\hat{D}/2}, \tag{102}$$

which reflects the larger expected response to a reference field t due to the uncertainty fluctuations around it. Those fluctuations pick up the asymmetric shape of the exponential term in the Hamiltonian, where the larger response to positive fluctuations is not fully compensated by the lower response to negative fluctuations. One might wonder, if the simple replacement rule in Eq. (102) could supplement the classical solution with the missing uncertainty-loop corrections. Thus we ask, if the modified classical equation

$$m = bS(d - \kappa_{m+b\hat{D}/2}), \tag{103}$$

together with a self-consistently determined propagator

$$D^{-1} = S^{-1} + \widehat{b^2\kappa}_{m+b\hat{D}/2}, \tag{104}$$

could provide the mean field given the data. A more rigorous renormalization calculation will show that this is indeed the case, within some approximation.

The loop-corrected density and propagator permit one to construct estimators for the dark-matter density itself,

$$\varrho = \varrho_0 e^{cs}, \tag{105}$$

instead of its logarithm, s. Here c fixes the relation between s and ϱ, and ϱ_0 being the cosmic median dark-matter density. Translating our log-density map into the density results in the naive density estimator

$$m_\varrho^{\text{naive}} = \varrho_0 e^{cm}, \tag{106}$$

which is not optimal in the sense of minimal rms deviations. The proper estimator would be

$$m_\varrho = \langle \varrho_0 e^{cs} \rangle_{(s|d)} = \varrho_0 e^{cm + c^2\hat{D}/2}, \tag{107}$$

which contains uncertainty-loop corrections accounting for the shift of the mean under the nonlinear transformation between log density and density.

F. Response renormalization

Since we are dealing with a ϕ^∞-field theory, the zoo of loop diagrams is quite complex, and forms something like a *Feynman foam*. In order not to get stuck in the multitude of this foam, we urgently require a trick to keep either the maximal order of the diagrams low, or to limit the number of vertices per diagram, or both. We have basically two handles on any interaction term $\lambda_n = \kappa b^n$, the bias b and the zero response κ. We concentrate on the response, since it enters the Hamiltonian in a linear way and also the data can be regarded to be proportional to κ. Thus, the full Hamiltonian

$$H[s] = \tfrac{1}{2}s^\dagger S^{-1}s - bd^\dagger s + \kappa_0 e^{bs} \tag{108}$$

can be regarded to be proportional to the response, except for the prior term and also constant terms we immediately drop here and in the following.

Let us assume that prior to any data analysis we have an initial guess m_0 for the signal with some Gaussian uncertainty characterized by the covariance D_0. This can be expressed via a Hamiltonian of the form

$$H_0[s] = \tfrac{1}{2}(s - m_0)^\dagger D_0^{-1}(s - m_0), \tag{109}$$

which defines a probability density via $P_0(s) \propto \exp(-H_0[s])$. In case the prior should be our initial guess, we have $m_0 = 0$ and $D_0 = S$, but we need not restrict ourself to this case. Now, we want to anticipate step by step the information of the full problem and forget our initial guess with the same rate. This can be modeled by adopting an affine parameter τ, which measures how much we exposed ourselves to the full problem. For each τ, which we regard as a pseudotime, our knowledge state is described by a Hamiltonian H_τ. Increasing τ by some small amount ε should therefore lead to the next knowledge state characterized by

$$H_{\tau+\varepsilon} = H_\tau[s] + \varepsilon(H[s] - H_\tau[s]). \tag{110}$$

This equation just models an asymptotical approach to the correct Hamiltonian. If the initial guess was the prior, one sees that for infinitesimal steps ε the knowledge flow corresponds to tuning up all terms proportional to κ,

$$H_\tau[s] = \tfrac{1}{2}s^\dagger S^{-1}s + (1 - e^{-\tau})(-bd^\dagger s + \kappa_0 e^{bs}) \longrightarrow H[s].$$

This motivates the term *response renormalization* for this kind of continuous learning system, into which the information source as well as the interactions is fed with the same rate.

The trick for the renormalization procedure is to approximate the knowledge state at each moment τ to be of Gaussian shape and therefore the Hamiltonian to be free (quadratic in the signal). Thus we set

$$H_\tau[s] = \tfrac{1}{2}(s - m_\tau)^\dagger D_\tau^{-1}(s - m_\tau), \tag{111}$$

where m_τ and $D_\tau = (S^{-1} + M_\tau)^{-1}$ are the mean and dispersion of the field given the acquired knowledge at time τ, respectively.

These have to be updated when the next learning step is to be performed. The next Hamiltonian, before being again replaced by a free one, is

$$\begin{aligned} H_{\tau+\varepsilon}[\phi] &= \frac{1}{2}\phi^\dagger D_\tau^{-1}\phi + \varepsilon\Big((S^{-1}m_\tau - bd)^\dagger\phi \\ &\quad - \frac{1}{2}\phi^\dagger M_\tau\phi + \kappa_{m_\tau}e^{b\phi}\Big) \\ &= \frac{1}{2}\phi^\dagger D_\tau^{-1}\phi + \varepsilon\sum_{n=1}^{\infty}\frac{1}{n!}\lambda_n\phi^n, \end{aligned} \tag{112}$$

if expressed for the momentarily uncertainty field $\phi = s - m_\tau$. Here, the perturbative expansion coefficients are given by

$$\lambda_1 = \kappa_{m_\tau}b + S^{-1}m_\tau - bd,$$

$$\lambda_2 = \kappa_{m_\tau}b^2 - \hat{M}_\tau, \quad \text{and} \quad \lambda_n = \kappa_{m_\tau}b^n \quad \text{for } n > 2,$$

assuming for simplicity that M_τ is diagonal. This is a save restriction, since we will see that for $\tau \longrightarrow \infty$ this is the case asymptotically, even for a nondiagonal initial M_0. Thus we can require that our initial guess was also of this form.

In order to approximate this Hamiltonian by a free one, we have to calculate the shifted mean field and its connected two-point correlation function, the full propagator. To first order in ε only leaf diagrams with a single perturbative interaction vertex contribute to the perturbed expec-

tation value of ϕ:

$$\langle\phi\rangle^{(\tau+\varepsilon)}_{(s|d)} = \text{——•} + \text{—◯} + \text{◯◯} + \text{◯◯} + \ldots$$
$$= \varepsilon D_\tau \left[b\, d - S^{-1} m_\tau - b\, \kappa_{m_\tau} e^{b^2 \widehat{D}_\tau/2} \right]. \quad (113)$$

Note that only odd interaction terms shift the expectation value $m_{\tau+\varepsilon} = m_\tau + \langle\phi\rangle^{(\tau+\varepsilon)}_{(s|d)}$. The even ones do not exert any net forces in the vicinity of $\phi_\tau = 0$ since they represent a potential which is mirror symmetric about this point.

The renormalized propagator $D_{\tau+\varepsilon}$ is given by the connected two-point correlation function $\langle\phi\phi^\dagger\rangle^{(\tau+\varepsilon)}_{(s|d)}$, and this is up to linear order in ε

$$\langle\phi\phi^\dagger\rangle^{(\tau+\varepsilon)}_{(s|d)} = \text{——} + \text{—•—} + \text{◯—} + \text{◯◯} + \ldots$$
$$= D_\tau + \varepsilon D_\tau \left(M - b^2 \kappa_{m_\tau} e^{b^2 \widehat{D}_\tau/2} \right) D_\tau. \quad (114)$$

Rewriting this for an update of M_τ we find up to linear order in ε

$$M_{\tau+\varepsilon} = (1-\varepsilon) M_\tau - \varepsilon b^2 \kappa_{m_\tau} e^{b^2 \hat{D}_\tau/2}. \quad (115)$$

Taking the limit $\varepsilon \longrightarrow 0$ yields the integro-differential system

$$\dot{m} = D(bd - b\kappa_0 e^{bm + b^2 \hat{D}/2} - S^{-1} m),$$
$$\dot{M} = b^2 \kappa_0 e^{bm + b^2 \hat{D}/2} - M, \quad \text{and}$$
$$D = (S^{-1} + \hat{M})^{-1}. \quad (116)$$

This converges at a fixed point, which we previously guessed in Eqs. (103) and (104) for our uncertainty-loop enhanced classical equation.

The classical and the renormalization flow fix point equations can be unified:

$$m = bS(d - \kappa_{bm + Tb\hat{D}/2}),$$
$$D = (S^{-1} + \hat{\kappa}_{bm + Tb\hat{D}/2})^{-1}, \quad (117)$$

with $T = 0$ and $T = 1$ for the classical and renormalization result, respectively.

The parameter T is more than a pure convenience. If we would have introduced a temperature T at the beginning, via $P(d, s|T) = \exp(-H_d[s]/T)$, Eq. (117) would have been the result of the renormalization-flow calculation. And the classical limit naturally corresponds to the zero temperature regime, in which the field expectation value is not affected by any uncertainty fluctuations since the system is at its absolute energy minimum.

An example of such reconstructions can be seen in Fig. 2, and its uncertainty structures in Fig. 3. Here, the renormalization equation indeed seems to provide a better result compared to the classical one. However, a statistical comparison of the two reconstructions using 1000 realizations of the signal and data in Fig. 4 shows that there is at most a marginal difference. This may be surprising, since the classical and renormalization solutions are quite distinct, and the latter is always lower than the former. One might therefore ask if the two are bracketing the correct solution. And indeed, intermediate solutions constructed using $T = 1/2$ perform better than the ones for $T = 0$ and $T = 1$, as can be seen in Fig. 4.

If neither $T = 0$ nor $T = 1$ provide the optimal reconstruction, what would be the right choice? We have to remember that we replaced the probability density function at each step of the renormalization scheme by a Gaussian with the correct mean and dispersion. However, the real probability is not a Gaussian, and therefore our mean field estimator is not optimal. Reconstructions with different T probe the non-Gaussian probability structure with a different wide Gaussian kernel in phase space, and therefore result in a slightly different signal means due to the anharmonic nature of our Hamiltonian.

G. Uncertainty structure

The remaining uncertainties at the end of the renormalization flow can mainly be read of the renormalized propagator D, which we display in the top part of Fig. 3 in comparison to the original, unrenormalized one D_0. The renormalized propagator is a much better approximation to the uncertainty dispersion of the signal posterior distribution around the mean map than the original one. One can clearly see that the data imprinted a highly nonuniform structure into the uncertainty pattern visible in the renormalized propagator with small uncertainties where there were many galaxy counts. Also the density estimator in Eq. (107) benefits from the knowledge of the uncertainty

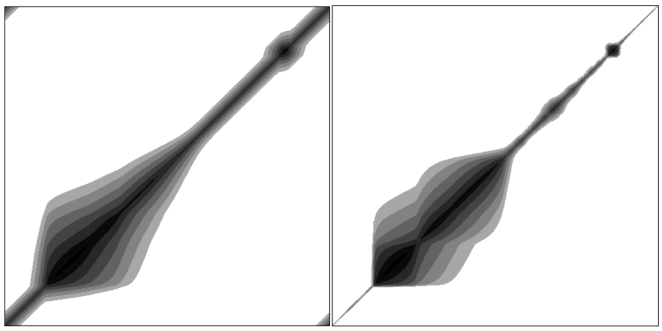


FIG. 3. The original propagator $D_0 = (S^{-1} + \widehat{\kappa_0 b^2})^{-1}$ (left panel) and the final of the renormalization flow D [Eq. (117), right panel] in logarithmic grey scaling for the data displayed in Fig. 2. The values of the diagonals show the local uncertainty variance (in Gaussian approximation) before ($\hat{D}_0$) and after ($\hat{D}$) the data are analyzed, respectively. The bottom left and top right corners exhibit nonvanishing propagator values due to the assumed periodic spatial coordinate, which puts these corners close to the two others on the matrix diagonal.

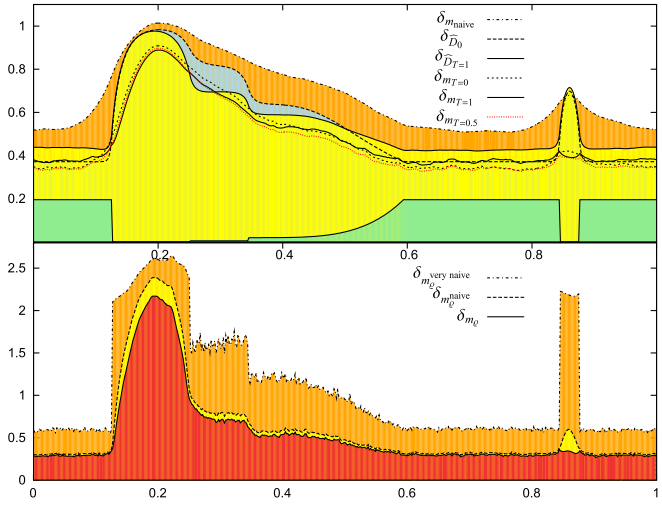


FIG. 4 (color online). Top panel: Statistical reconstruction error from 1000 signal and data realizations. Curves are, roughly in order from top (bad performance) to bottom (good performance): error $\delta_{m_{\text{naive}}} = \langle (s - m_{\text{naive}})^2 \rangle^{1/2}_{(d,s)}$ of the signal-covariance-convolved naive map $m_{\text{naive}} = S s_{\text{naive}}$ [see Eq. (97)], expected Wiener uncertainty $\delta_{\hat{D}_0} = \hat{D}_0^{1/2}$, averaged renormalized uncertainty $\delta_{\hat{D}_{T=1}} = \langle \hat{D}_{T=1} \rangle^{1/2}_{(d,s)}$, error of the classical map $\delta_{m_{T=0}} = \langle (s - m_{T=0})^2 \rangle^{1/2}_{(d,s)}$, error of the renormalized map $\delta_{m_{T=1}} = \langle (s - m_{T=1})^2 \rangle^{1/2}_{(d,s)}$, and error of the intermediate map $\delta_{m_{T=0.5}} = \langle (s - m_{T=0.5})^2 \rangle^{1/2}_{(d,s)}$. The lowest curve without a label is κ. Bottom panel: Error variance of estimators for the density, $\varrho = e^s$, namely, $\delta_{m_\varrho^{\text{very naive}}} = \langle (\varrho - e^{m_{\text{naive}}})^2 \rangle^{1/2}_{(d,s)}$, $\delta_{m_\varrho^{\text{naive}}} = \langle (\varrho - m_\varrho^{\text{naive}})^2 \rangle^{1/2}_{(d,s)}$, and $\delta_{m_\varrho} = \langle (\varrho - m_\varrho)^2 \rangle^{1/2}_{(d,s)}$ [see Eqs. (106) and (107)].

structure contained in the renormalized propagator, as the lower panel of Fig. 4 shows.

The propagators also visualize the effect any additional data would have at different locations. The height and width of the propagator values define, respectively, the strength of the response to and the distance of information propagation from an information source.

The structure of D_0 is imprinted by the prior and the mask. At D_0's widest locations the mask blocks any information source and the structure of the signal prior S becomes visible. At locations where the mask is transparent, the reconstruction response per information source is lower, as plenty of information can be expected there. Also the propagator width is smaller, since the individual informations do not need to be propagated that far, thanks to the richer information source density in such regions.

The structure of D_m has additionally imprinted the expected information source density structure given the reconstruction m. The strongly nonlinear signal response has led to regions with very high galaxy count rates, which have larger information densities, and therefore lower and narrower information propagators. This implies that any additional galaxy detection in the regions with high galaxy counts will have little impact on the updated map, whereas any additional detected galaxies in low density regions will more strongly change it. However, the number of additional galaxies per invested observing time will be larger in high density regions, which may compensate the lower information-per-galaxy ratio there. It is therefore interesting to look at the observational information content and how it depends on the actual data realization.

H. Information gain

In the case of a free theory, the amount of information depends on the experimental setup and on the prior, but is independent of the data obtained as we have shown in Sec. IV D 2. This changes in the case that one wants to harvest information in a situation described by a nonlinear IFT. There, the amount of information can strongly depend on the actual data.

This is well illustrated by our LSS reconstruction problem. A perturbative calculation of the nonlinear information gain is possible if either the bias factor or the signal amplitude, which both control the strength of the nonlinear interactions, are small compared to unity.[13]

The information gain, as given by Eq. (83), expanded to the first few orders in b

$$\Delta I_1 = \underbrace{\tfrac{1}{2}\mathrm{Tr}\log(1 + \widehat{S\kappa b^2})}_{\Delta I_0} + \tfrac{1}{2}(\kappa b^3 \hat{D}_0)^\dagger (m_0 + \tfrac{1}{2} b(\hat{D}_0 + m_0^2)) + \mathcal{O}(b^5), \quad (118)$$

clearly depends on the actual realization of the data. The different fluctuations in the Wiener map $m_0 = D_0 j$, with $D_0 = (S^{-1} + \widehat{b^2\kappa})^{-1}$ and $j = b(d - \kappa)$, imply positive and negative information density fluctuations.

To conveniently calculate the information gain of the observation in case of a large bias factor, we use the Gaussian approximation of the jointed probability function, as provided by the renormalization scheme. Because of the Gaussianity of this approximate solution, we can simply use the formula for the information gain of a free theory, as given by Eq. (86). This yields

$$\Delta I_d = \tfrac{1}{2}\mathrm{Tr}(\log(1 + \widehat{Sb^2\kappa\eta})), \quad (119)$$

with $\eta = e^{bm + (1/2)b^2\hat{D}_{T=1}}$ being proportional to the expected number density of galaxies in this region [see Eq. (107)]. It is also obvious that the information gain depends on the data. In regions with higher observed galaxy numbers, η is larger, and more information is expected to be harvested by further observations. This is

[13]The signal amplitude can, for example, be made small by defining the signal of interest to be the cosmic density field, smoothed on a sufficiently large scale (> 10 Mpc) so that $\langle s^2 \rangle_{(s)} < 1$.

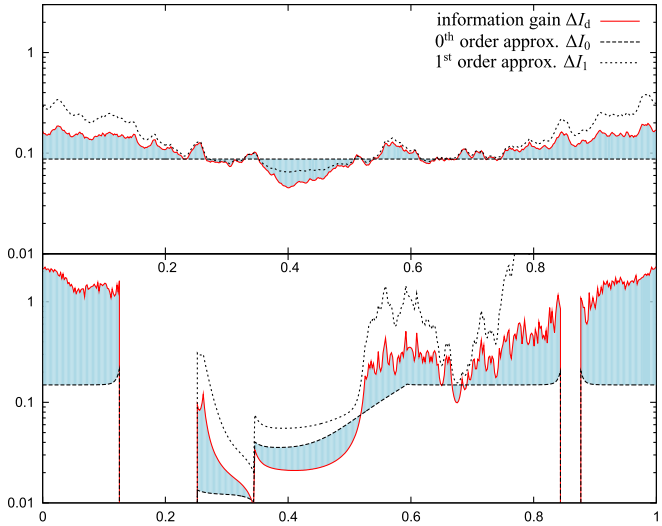


FIG. 5 (color online). Information gain density [the integrands of Eqs. (118) and (119)] for the two reconstruction examples presented, the only weakly nonlinear one (top panel, and Fig. 1) and the strongly nonlinear one (bottom panel, and Fig. 2). The renormalization result for $T = 1$ [Eq. (119)], and the zero- and first-order perturbative results [Eq. (118)] are shown. The information gain depends on the observational sensitivity as well as the actual data. The latter influence is stronger in the nonlinear regime and disappears in linear inference problems.

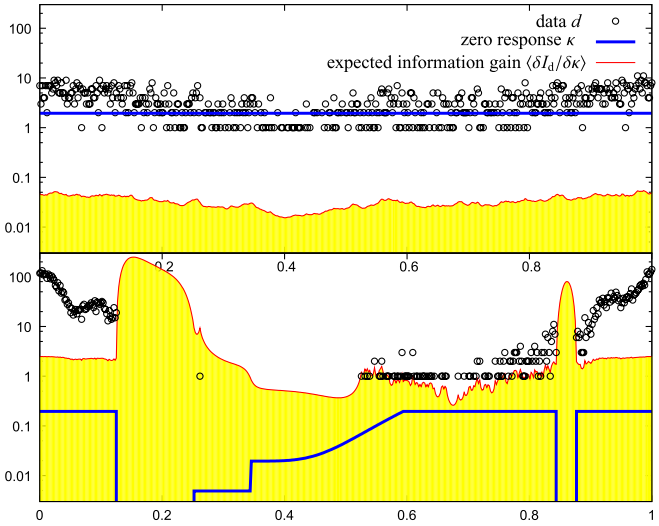


FIG. 6 (color online). Differential information gain density for the two reconstruction examples presented, the only weakly nonlinear one (top panel, and Fig. 1) and the strongly nonlinear one (bottom panel, and Fig. 2).

illustrated in Fig. 5, where the information gain density, the individual contributions to the trace in Eq. (119), as well as the first and all terms of Eq. (118) are shown for the cases displayed in Figs. 1 and 2. The approximate Eq. (118) seems to be adequate for $b \ll 1$, but not for our cases of $b = 0.5$ and 2.5.

The expected benefit of additional observations at location x can also be calculated by differentiating Eq. (119) with respect to $\kappa(x)$. Using Eqs. (88) and (117) we find

$$\left\langle \frac{\delta I_d}{\delta \kappa_0} \right\rangle_{(\text{new data}|d)} = \frac{1}{2} b^2 \eta \left(1 + \frac{1}{2} \widehat{\kappa b^2} \eta D^2 b^2 \right)^{-1} \hat{D}. \quad (120)$$

The expected information gain is especially large for observations at locations where the uncertainty $\hat{D}$ is large, where a large number density of galaxies ($\propto \eta$) can be expected, and where strong nonlinearities are present ($\propto b^2$). The inverse term caps the maximally available information gain at some level. For the two reconstruction examples given in Figs. 1 and 2 we display the expected information gain as a function of the observing position in Fig. 6.

It is apparent from the top panel, showing the case of uniform observation coverage, that additional observations are more advantageous at locations where already an increased matter density is identified. The bottom panel, showing the case of a very inhomogeneous observation of strongly nonlinear data, demonstrates that filling observational gaps should have the highest priority. But there again, regions where the extrapolated galaxy density seems to be larger should be preferred, as can be seen from the asymmetric shape of the expected information gain for observations in the gap around $x = 0.2$. In this example, the information harvest of high galaxy density regions can be so large that further observations of the already well-observed regions at the boundary of the domain seem to be more advantageous than improving the poorly observed regions around $x = 0.4$, where a low galaxy density is already apparent from the existing data.

Of course, in order to plan observations in a real case, the dependence of observational costs as a function of location x and already achieved zero response there, $\kappa(x)$, have to be folded into the considerations.

VI. NON-GAUSSIAN CMB FLUCTUATIONS VIA f_{nl} THEORY

A. Data model

As an IFT example on the sphere $\Omega = S^2$, involving two interacting uncertainty fields, we investigate the so-called f_{nl} theory of local non-Gaussianities in the CMB temperature fluctuations. This problem has currently a high scientific relevance due to the strongly increasing availability of high fidelity CMB measurements, which permit one to constrain the physical conditions at very early epochs of the Universe. The relevant references for this topic were provided in Sec. I C 5.

On top of the very uniform CMB sky with a mean temperature T_{CMB}, small temperature fluctuations on the level of $\delta T_{\text{obs}}^{\{\text{I,E,B}\}}/T_{\text{CMB}} \sim 10^{-\{5,6,7\}}$ are observed or expected in total intensity (Stokes I) and in polarization E and B modes, respectively. The weak B modes are mainly due to the lensing of E modes and some unknown level of

gravity waves. We will disregard them in the following. These CMB temperature fluctuations are believed and observed to follow mostly a Gaussian distribution. However, inflation predicts some level of non-Gaussianity. Some of the secondary anisotropies imprinted by the LSS of the Universe via CMB lensing, the integrated Sachs-Wolfe and the Rees-Sciama effects should also have imprinted non-Gaussian signatures [215,216]. The primordial, as well as some of the secondary CMB temperature fluctuations are a response to the gravitational potential initially seeded during inflation. Since we are interested in primordial fluctuations, we write

$$d \equiv \delta T_{\rm obs}^{\{\rm I,E\}}/T_{\rm CMB} = R\varphi + n, \tag{121}$$

where φ is the three-dimensional, primordial gravitational potential, and R is the response on it of a CMB instrument, observing the induced CMB temperature fluctuations in intensity and E-mode polarization. These are imprinted by a number of effects, like gravitational redshifting, the Doppler effect, and anisotropic Thomson scattering. In the case that the data of the instrument are foreground cleaned and deconvolved all-sky maps (assuming the data processing to be part of the instrument) the response, which translates the 3D gravitational field into temperature maps, is well known from CMB theory and can be calculated with publicly available codes like CMBFAST, CAMB, and CMBEASY (see Sec. I C 5). The precise form of the response does not matter for a development of the basic concept and can be inserted later.

Finally, the noise n subsumes all deviation of the measurement from the signal response due to instrumental and physical effects, which are not linearly correlated with the primordial gravitational potential, such as detector noise, remnants of foreground signals, but also primordial gravitational wave contributions to the CMB fluctuations.

The small level of non-Gaussianity expected in the CMB temperature fluctuations is a consequence of some non-Gaussianity in the primordial gravitational potential. Despite the lack of a generic non-Gaussian probability function, many of the inflationary non-Gaussianities seem to be well described by a local process, which taints an initially Gaussian random field, $\phi \hookleftarrow P(\phi) = \mathcal{G}(\phi, \Phi)$ (with the ϕ covariance $\Phi = \langle \phi\phi^\dagger \rangle_{(\phi)}$), with some level of non-Gaussianity. A well controllable realization of such a tarnishing operation is provided by a slightly nonlinear transformation of ϕ into the primordial gravitational potential φ via

$$\varphi(x) = \phi(x) + f_{\rm nl}(\phi^2(x) - \langle \phi^2(x) \rangle_{(\phi)}) \tag{122}$$

for any x. The parameter $f_{\rm nl}$ controls the level and nature of non-Gaussianity via its absolute value and sign, respectively. This means that our data model reads

$$d = R(\phi + f(\phi^2 - \hat{\Phi})) + n, \tag{123}$$

where we dropped the subscript of $f_{\rm nl}$. In the following we assume the noise n to be Gaussian with covariance $N = \langle nn^\dagger \rangle_{(n)}$ and define as usual $M = R^\dagger N^{-1} R$ for notational convenience.[14]

B. Spectrum, bispectrum, and trispectrum

The nonlinearity of the relation between the hidden Gaussian random field ϕ and the observable gravitational potential φ [Eq. (122)] imprints non-Gaussianity into the latter. In order to be able to extract the value of the non-Gaussianity parameter f from any data containing information on φ, we need to know its statistic at least up to the four-point function, the trispectrum, which we briefly derive with IFT methods.

To that end, it is convenient to define a φ-moment generating function $Z[J]$ and its logarithm

$$\begin{aligned} \log Z[J] &= \log \int \mathcal{D}\phi P(\phi) e^{J^\dagger \varphi(\phi)} \\ &= \frac{1}{2} J^\dagger (\Phi^{-1} - 2\widehat{fJ})^{-1} J - (fJ)^\dagger \hat{\Phi} \\ &\quad - \frac{1}{2} \mathrm{Tr}[\log(1 - 2\Phi \widehat{fJ})]. \end{aligned} \tag{124}$$

This permits one to calculate via J derivatives [see Eqs. (32)–(35)] the mean

$$\bar{\varphi} = \langle \varphi \rangle_{(\phi)} = 0, \tag{125}$$

the spectrum (or covariance)

$$\begin{aligned} C_{xy}^{(\varphi)} &= \langle \varphi_x \varphi_y \rangle_{(\phi)}^c = \langle (\varphi - \bar{\varphi})_x (\varphi - \bar{\varphi})_y \rangle_{(\phi)} \\ &= \Phi_{xy} + 2 f_x \Phi_{xy}^2 f_y, \end{aligned} \tag{126}$$

the bispectrum[15]

[14] Non-Gaussian noise components are in fact expected and would need to be included into the construction of an optimal $f_{\rm nl}$ reconstruction. However, currently we aim only at outlining the principles and we are furthermore not aware of a traditional $f_{\rm nl}$ estimator constructed while taking such noise into account. And finally, we show at the end how to identify some of such non-Gaussian noise sources by producing $f_{\rm nl}$ maps on the sphere, which can morphologically be compared to known foreground structures, like our Galaxy.

[15] Since the bispectrum contains most of the non-Gaussianity signature, we also provide its Fourier-space version, which is well known for the $f_{\rm nl}$ model [217]. The bispectrum for $f =$ const, expressed in terms of the φ covariance reads

$$B_{xyz}^{(\varphi)} = 2f[C_{xy}^{(\varphi)} C_{yz}^{(\varphi)} + C_{xz}^{(\varphi)} C_{zy}^{(\varphi)} + C_{yx}^{(\varphi)} C_{xz}^{(\varphi)}] + \mathcal{O}(f^3).$$

Fourier transforming this yields

$$\begin{aligned} B_{k_1 k_2 k_3}^{(\varphi)} &= 2f(2\pi)^3 \delta(k_1 + k_2 + k_3)[P(k_1)P(k_2) + P(k_2)P(k_3) \\ &\quad + P(k_3)P(k_1)] + \mathcal{O}(f^3), \end{aligned}$$

where $P(k)$ is the power spectrum of φ, which is identical to that of ϕ up to $\mathcal{O}(f^2)$.

$$B^{(\varphi)}_{xyz} = \langle(\varphi-\bar\varphi)_x(\varphi-\bar\varphi)_y(\varphi-\bar\varphi)_z\rangle_{(\phi)} = \langle\varphi_x\varphi_y\varphi_z\rangle^c_{(\phi)}$$
$$= 2[\Phi_{xy}f_y\Phi_{yz} + \Phi_{yz}f_z\Phi_{zx} + \Phi_{zx}f_x\Phi_{xy}] + 8\Phi_{xy}f_y\Phi_{yz}f_z\Phi_{zx}f_x, \quad (127)$$

and the trispectrum

$$T^{(\varphi)}_{xyzu} = \langle(\varphi-\bar\varphi)_x(\varphi-\bar\varphi)_y(\varphi-\bar\varphi)_z(\varphi-\bar\varphi)_u\rangle_{(\phi)}$$
$$= \Phi_{xy}\Phi_{zu} + \Phi_{xz}\Phi_{yu} + \Phi_{xu}\Phi_{yz} + \langle\varphi_x\varphi_y\varphi_z\varphi_u\rangle^c_{(\phi)}$$
$$= [\tfrac{1}{8}\Phi_{xy}\Phi_{zu} + 2\Phi_{xy}f_y\Phi_{yz}f_z\Phi_{zu} + \Phi_{xy}f_y\Phi_{yz}f_z\Phi_{zu}f_u\Phi_{ux}f_{x\cdot}] + 23 \text{ perm} \quad (128)$$

of the gravitational potential. Since we will investigate the possibility of a spatially varying non-Gaussianity parameter at the end of this section, we keep track of the spatial coordinate of f, but for the time being read $f_x = f$.

The spectrum, bispectrum, and trispectrum of our CMB measurement can easily be calculated from the gravitational spectrum and bispectrum, respectively:

$$C^{(d)} = RC^{(\varphi)}R^\dagger + N, \qquad B^{(d)}_{\hat n_1\hat n_2\hat n_3} = R_{\hat n_1 x}R_{\hat n_2 y}R_{\hat n_3 z}B^{(\varphi)}_{xyz},$$
$$T^{(d)}_{\hat n_1\hat n_2\hat n_3\hat n_4} = R_{\hat n_1 x}R_{\hat n_2 y}R_{\hat n_3 z}R_{\hat n_3 u}T^{(\varphi)}_{xyzu} + [(RC^{(\varphi)}R^\dagger + \tfrac{1}{8}N)_{\hat n_1\hat n_2}N_{\hat n_3\hat n_4} + 23 \text{ permutations}], \quad (129)$$

where $\hat n$ denotes the unit vector on the sphere, and we have made use of the assumption of the noise being Gaussian and independent of the signal. In case the noise itself has a bi- or trispectrum, or there is a signal dependent noise, e.g. due to an incorrect instrument calibration, then more terms have to be added to the expressions. The usually quoted formulas [204,217–219] can be obtained from Eq. (129) by applying spherical harmonic transformations.

C. CMB Hamiltonian

Although we are not interested in the auxiliary field ϕ, it is nevertheless very useful for its marginalization to define its Hamiltonian, which is

$$H_f[d,\phi] = -\log(\mathcal{G}(\phi,\Phi)\mathcal{G}(d-R(\phi+f(\phi^2-\hat\Phi)),N))$$
$$= \frac{1}{2}\phi^\dagger D^{-1}\phi + H_0 - j^\dagger\phi + \sum_{n=0}^{4}\frac{1}{n!}\Lambda^{(n)}[\phi,\ldots,\phi], \quad \text{with} \quad D^{-1} = \Phi^{-1} + R^\dagger N^{-1}R \equiv \Phi^{-1} + M,$$
$$j = R^\dagger N^{-1}d, \quad \Lambda^{(0)} = j^\dagger(f\hat\Phi) + \frac{1}{2}(f\hat\Phi)^\dagger M(f\hat\Phi), \quad \Lambda^{(1)} = -(f\hat\Phi)^\dagger M \text{ and } j' = j - \Lambda^{(1)\dagger},$$
$$\Lambda^{(2)} = -2\widehat{fj'}, \quad \Lambda^{(3)}_{xyz} = (M_{xy}f_y\delta_{yz} + 5 \text{ permutations}), \quad \Lambda^{(4)}_{xyzu} = \frac{1}{2}(f_x\delta_{xy}M_{yz}\delta_{zu}f_u + 23 \text{ permutations}), \quad (130)$$

and H_0 collects all terms independent of ϕ and f. The last two tensors should be read without the Einstein sum convention, but with all possible index permutations. Note, that this is a nonlocal theory for ϕ in case that either the noise covariance or the response matrix is nondiagonal, yielding a nonlocal M and therefore nonlocal interactions $\Lambda^{(3)}$ and $\Lambda^{(4)}$.

We should note that Babich [220] derived the now traditional $f_{\rm nl}$ estimator from a very similar starting point, the log probability for φ. The difference of the resulting estimators is not due to the slightly different approaches ($H_f[d,\varphi]$ versus $H_f[d,\phi]$), but because of the frequentist and Bayes statistics he and we use, respectively.

In case that the noise as well as the response is diagonal in position space, as it is often assumed for the instrument response of properly cleaned CMB maps, and is also approximately valid on large angular scales, where the Sachs-Wolfe effect dominates, we have $N_{xy} = \sigma_n^2(x)\delta(x-y)$, $R = -3$ [215] for the total intensity fluctuations, and thus $M_{xy} = 9\sigma_n^{-2}(x)\delta(x-y)$, if we restrict the signal space to the last-scattering surface, which we identify with S^2. This permits one to simplify the Hamiltonian to

$$H_f[d,\phi] = \frac{1}{2}\phi^\dagger D^{-1}\phi + H_0 - j^\dagger\phi + \sum_{n=0}^{4}\frac{1}{n!}\lambda_n^\dagger\phi^n,$$
$$\text{with} \quad D^{-1} = \Phi^{-1} + 9\widehat{\sigma_n^{-2}},$$
$$j' = j - \lambda_1 = 3(3\hat\Phi f - d)/\sigma_n^2,$$
$$\lambda_0 = 3(\hat\Phi/\sigma_n^2)^\dagger\left(\frac{3}{2}f^2\hat\Phi - fd\right), \qquad \lambda_2 = -2fj',$$
$$\lambda_3 = 54f/\sigma_n^2, \quad \text{and} \quad \lambda_4 = 108f^2/\sigma_n^2. \quad (131)$$

The numerical coefficients of the last two terms may look large, however, these coefficients stand in front of terms of typically $\phi^3 \sim 10^{-15}$, and $\phi^4 \sim 10^{-20}$, which ensures their well behavedness in any diagrammatic expansion series.

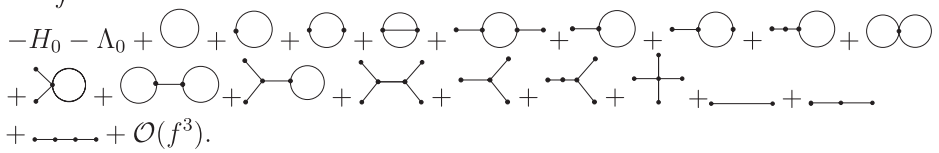
For later usage, we define the Wiener-filter reconstruction of the gravitational potential as $m_0 = Dj$.

D. $f_{\rm nl}$ evidence and map making

Since we are momentarily not interested in reconstructing the primordial fluctuations, but to extract knowledge on

$f_{\rm nl}$, we marginalize the former by calculating the log evidence $\log P(d|f)$ up to quadratic order in f:

$$\begin{aligned}\log Z_f[d] &= \log \int \mathcal{D}\phi\, P(d,\phi|f) \\ &= \log \int \mathcal{D}\phi\, e^{-H_f[d,\phi]} \\ &= -H_0 - \Lambda_0 + \text{(diagrams)} + \mathcal{O}(f^3).\end{aligned} \tag{132}$$



We have made use of the fact that the logarithm of the partition sum is provided by all connected diagrams, and that j' contains a term of the order $\mathcal{O}(f^0)$, $\Lambda^{(2)}$ and $\Lambda^{(3)}$ contain terms of the order $\mathcal{O}(f^1)$, and $\Lambda^{(4)}$ one of the order $\mathcal{O}(f^2)$, so that they can appear an unrestricted number of times, twice and once in diagrams of order up to $\mathcal{O}(f^2)$, respectively. Since only fourth-order interactions are involved, an implementation in spherical-harmonics space may be feasible using the only fourth-order C coefficients [Eq. (B3)], which can be calculated computer algebraically. Finally, we recall

$$\bigcirc = \frac{1}{2}\log|2\pi\, D| = \frac{1}{2}\mathrm{Tr}(\log(2\pi\, D)). \tag{133}$$

Although f is not known, the expressions in Eq. (132) proportional to f and f^2 can be calculated separately, permitting one to write down the Hamiltonian of f if a suitable prior $P(f)$ is chosen,

$$\begin{aligned}H_d[f] &\equiv -\log(P(d|f)P(f)) \\ &= \tilde{H}_0 + \tfrac{1}{2} f^\dagger \tilde{D}^{-1} f + \tilde{j}^\dagger f + \mathcal{O}(f^3),\end{aligned} \tag{134}$$

where we collected the linear and quadratic coefficients into $\tilde{j}$ and $\tilde{D}^{-1}$. It is obvious that the optimal f estimator to lowest order is therefore

$$m_f = \langle f \rangle_{(\phi,f|d)} = \tilde{D}\,\tilde{j}, \tag{135}$$

and its uncertainty variance is just

$$\langle (f - m_f)(f - m_f)^\dagger \rangle_{(\phi,f|d)} = \tilde{D}. \tag{136}$$

So far, we have assumed f to have a single universal value. However, we can also permit f to vary spatially, or on the sphere of the sky. In the latter case one would expand f as

$$f(x) = \sum_{l=0}^{l_{\max}} \sum_{m=-l}^{l} f_{lm} Y_{lm}(\hat{x}) \tag{137}$$

up to some finite $l_{\max}$. Then one would recalculate the partition sum, now separately for terms proportional to f_{lm} and $f_{lm}f_{l'm'}$, which are then sorted into the vector and matrix coefficients of $\tilde{j}$ and $\tilde{D}^{-1}$, respectively, and according to

$$\tilde{j}_{(lm)} = \left.\frac{dH_d[f]}{df_{lm}}\right|_{f=0}, \quad \text{and}$$
$$\tilde{D}^{-1}_{(lm)(l'm')} = \left.\frac{d^2 H_d[f]}{df_{lm} df_{l'm'}}\right|_{f=0}. \tag{138}$$

f-map making can then proceed as described above in spherical-harmonics space. Comparing the resulting map in angular space to known foreground sources, as our Galaxy, the level of non-Gaussian contamination due to their imperfect removal from the data may be assessed.

E. Comparison to traditional estimator

We conclude this chapter with a short comparison to traditional $f_{\rm nl}$ estimators. To our knowledge, the most developed estimator in the literature is based on the CMB bispectrum, which is the third order correlation functions of the data, e.g. [220,221], and references in Sec. IC 5. The IFT based filter presented here contains terms which are up to fourth order in the data, and therefore can be expected to be of higher accuracy since both methods are supposed to be optimal. Kogo and Komatsu [219] note that the CMB trispectrum should contain significant information on $f_{\rm nl}^2$, and may be superior to non-Gaussianity detection compared to the bispectrum on small angular scales. However, since the trispectrum is insensitive to the sign of $f_{\rm nl}$, its actual usage as a proxy is a bit more subtle. In the IFT estimator, any term proportional to $f_{\rm nl}^2$ enters the inverse of the propagator $\tilde{D}$, and therefore the trispectrum seems to unfold its $f_{\rm nl}$-estimation power mostly in combination with the bispectrum, which drives $\tilde{j}$.

Under which conditions does the traditional estimator emerge from the IFT one? There are three conceptual differences between the estimators, in that the IFT filter can handle inhomogeneous non-Gaussianity, correct for CMB sky and exposure chance coupling, and is unbiased with respect to the posterior.

The traditional estimator is usually written as

$$\varepsilon = \frac{1}{\mathcal{N}} \int dx\, A(x) B^2(x) = \frac{1}{\mathcal{N}} m_0^\dagger \Phi^{-1} m_0^2, \tag{139}$$

where $B = Dj = m_0$ is the Wiener-filter reconstruction of the gravitational potential, $A = \Phi^{-1}B$ is the same, just additionally filtered by the inverse power spectrum, and

$\mathcal{N}$ is a normalization constant [202]. This is fixed by the condition that the estimator should be unbiased with respect to all gravitational potential and noise realizations,

$$\begin{aligned}\mathcal{N} &= \langle m_0^\dagger \Phi^{-1} m_0^2 \rangle_{(d,\phi|f=1)} \\ &= B^{(\varphi)}_{xyz}|_{f=1}[(MD)_{xu}\Phi^{-1}_{uv}(DM)_{vy}(DM)_{vz}] \\ &= 2[\Phi_{xy}\Phi_{yz} + \Phi_{yz}\Phi_{zx} + \Phi_{zx}\Phi_{xy}] \\ &\quad \times [(MD)_{xu}\Phi^{-1}_{uv}(DM)_{vy}(DM)_{vz}]. \end{aligned} \tag{140}$$

The first difference between the estimators is obvious, in that the IFT estimator can handle a spatially varying $f(x)$. Therefore, we will only regard spatially constant nonlinearity parameters in the following. Since no CMB experiment is able to measure the monopole temperature fluctuation, the response to any spatially homogeneous signal is zero. This means, in the Fourier basis, that $R_{\hat{n},k=0} = 0$ and therefore $j_{k=0} = M_{k=0,k'} = 0$. Thus, we find for a universe with homogeneous statistics ($\hat{\Phi}_{k\neq 0} = 0$) that $\Lambda^{(0)} = \Lambda^{(1)} = 0$, $j' = j$, and $\Lambda^{(2)} = -2f\hat{j}$, which reduces the number of diagrams we have to calculate.

The IFT estimator is driven by the f-information source $\tilde{j}$, which is given by all diagrams which contain terms linear in f. There are four of them, yielding

$$\begin{aligned}\tilde{j} &= \frac{1}{f}\left[\right. \\ &= m_0^\dagger \Phi^{-1} m_0^2 + m_0^\dagger \left[\Phi^{-1}\widehat{D} - 2\,\widehat{MD}\right], \end{aligned} \tag{141}$$

where we used $M = D^{-1} - \Phi^{-1}$ in order to combine the two tree and the two loop diagrams into the first and second terms, respectively. The term resulting from the tree diagrams is actually identically to the unnormalized traditional estimator ε [Eq. (139)].

The terms resulting from the loop diagrams vanish for an homogeneous M, which a CMB experiment with uniform exposure and constant noise could produce. In case of an inhomogeneous M, which is the more realistic case, the loop term does not vanish and corrects for chance correlations between the CMB realization (as seen through j) and the noise and response structure of the experiment (as encoded in M and D). Creminelli *et al.* [222] already pointed out that such a linear correction term is necessary in case of an inhomogeneous sky coverage.

Anyhow, the second difference between the estimators is that the IFT based one applies a correction for chance correlations of CMB sky and sky exposure and the traditional one does not. This term is absent in the traditional estimator since the latter was constructed as the optimal estimator which is third order in the data. This excluded the loop term, which is linear in the data.

An inclusion of this term into the traditional estimator is straightforward and actually done by the more recent $f_{\rm nl}$ measurements [223]. The normalization constant $\mathcal{N}$ is unaffected by this, since the expectation value of the loop term averaged over all possible signal realization is zero.

This brings us to the third difference between the estimators, the different normalization. The traditional estimator is normalized by a data independent constant $\mathcal{N}$, where the IFT estimator is normalized by a data dependent term

$$\tilde{D}^{-1} = \frac{1}{\sigma_f^2} + \frac{2}{f^2}\left[\right. \tag{142}$$

where only the first three diagrams are data independent and σ_f is the variance of the prior, which we assume to be $P(f) = \mathcal{G}(f, \sigma_f^2)$. The detailed expressions for the different diagrams can be found in Appendix C. For both estimators, the traditional and the IFT one, the normalization is supposed to guarantee unbiasedness, however, with respect to different probability distributions.

The traditional estimator is unbiased in the frequentist sense, for an average over all signal f and data realizations. However, the IFT estimator is unbiased in the Bayesian sense, with respect to the posterior, the probability distribution of all signals given the data. Since the data are given, and not assumed to vary any more after the observation is performed, it can and should affect the normalization constant, which encodes the sensitivity of our non-Gaussianity estimation.

The reason for the IFT normalization constant (or f propagator) to be data dependent can be understood as follows. There are data realizations which are better suited to reveal the presence of non-Gaussianities than others, even if they have identical $\tilde{j}$. Such a dependence of the detectability of an effect on the concrete data realization is common in nonlinear Baysian inference, and was even more prominent in the example of the reconstruction of a log-normal density field in Sec. V.

VII. SUMMARY AND OUTLOOK

Starting with fundamental information theoretical considerations about the nature of measurements, signals, noise, and their relation to a physical reality given a model of the Universe or the system under consideration, we reformulated the inference problem in the language of IFT. IFT is actually a statistical field theory. The information field is identified with a spatially distributed signal, which can freely be chosen by the scientist according to needs and technical constraints. The mathematical apparatus of field theory permits one to deal with the ensemble of all possible field configurations given the data and prior information in a consistent way.

With this conceptual framework, we derived the Hamiltonian of the theory, showed that the free theory reproduces the well-known results of the Wiener-filter theory, and presented the Feynman rules for nonlinear, interacting Hamiltonians in general, and, in particular, cases. The latter are information fields over Fourier- and spherical harmonics spaces for inference problems in R^n and S^2, respectively. Our ''philosophical'' considerations permitted one to argue why the resulting IFTs are usually well normalized, but often nonlocal. Since the propagator of the theory is closely related to the Wiener filter, for which nowadays efficient numerical algorithms exist as image reconstruction and map-making codes, and the information source term is usually a noise weighted version of the data, the necessary computational tools are at hand to convert the diagrammatic expressions into well-performing algorithms.

Furthermore, we provided the Boltzmann-Shannon information measure of IFT based on the Helmholtz free energy, thereby highlighting the embedding of IFT in the framework of statistical mechanics.

As examples of the IFT recipe, two concrete IFT problems with cosmological motivation were discussed, which are also thought of as blueprints for other inference problems. The first was targeting at the problem of reconstructing the spatially continuous cosmic LSS matter distribution from discrete galaxy counts in incomplete galaxy surveys. The resulting algorithm can also be used for image reconstruction with low-number photon statistics, e.g. in low-dose x-ray imaging.

The second example was the design of an optimal method to measure or constrain any possible local nonlinearities in the CMB temperature fluctuations. This may serve as a blueprint for statistical monitoring of the linearity of a signal amplifier.

We conclude here with a short outlook on some problems that are accessible to the presented theory.

Many signal inference problems involve the reconstruction of fields without precisely known statistics. Some coefficients in the IFT Hamiltonians may only be phenomenological in nature, and therefore have to be derived from the same data used for the reconstruction itself. This more intricate interplay of parameter and information field can also be incorporated into the IFT framework, as we will show with a subsequent work.

For cosmological applications, along the lines started in this work, clearly more realistic data models need to be investigated. For example, to understand the response in galaxy formation to the underlying dark-matter distribution in terms of a realistic, statistical model, to be used in constructing the corresponding IFT Hamiltonian for a dark-matter information field, detailed higher-order correlation coefficients have to be distilled from numerical simulations or semianalytic descriptions. Also the CMB Hamiltonian may benefit from the inclusion of remnants from the CMB foreground subtraction process, permitting one to gather more solid evidence on fundamental parameters which are hidden in the CMB fluctuations, like the amplitude of non-Gaussianities.

Furthermore, there exist a number of more or less heuristic algorithms for inverse problems, which have proven to serve well under certain circumstances. Reverse engineering of their implicitly assumed priors and data models may permit one to understand better for which conditions they are best suited, as well as how to improve them in case these conditions are not exactly met.

Finally, we are very curious to see whether and how the presented framework may be suitable to inference problems in other scientific fields.

ACKNOWLEDGMENTS

It is a pleasure to thank the following people for helpful scientific discussions on various aspects of this work: Simon White on the dangers of perturbation theory, Benjamin Wandelt on the prospects of large-scale structure reconstruction, Jens Jasche on the pleasures and pains of signal processing, Jörg Rachen on the philosophy of science, and André Waelkens on the invariant, but vertiginous theory of isotropic tensors. We thank Cornelius Weig and Henrik Junklewitz for debates on the connection between IFT and QFT. We gratefully acknowledge helpful comments on the manuscript by Marcus Brüggen and Thomas Riller and by three very constructive referees.

APPENDIX A: NOTATION

We briefly summarize our notation of functions in position and Fourier space.

A real, but in principle also complex function $f(x)$ over the n-dimensional space is regarded as a vector f in a discrete and finite-dimensional, or continuous and infinite-dimensional Hilbert space. f will denote this vector, independently of the momentarily chosen function basis, be it the real space $f(x) = \langle x|f\rangle$ or the Fourier basis

$$f(k) = \langle k|f\rangle = \int dx\, f(x) e^{ik\cdot x}. \tag{A1}$$

Here, the volume integration usually is performed only over a finite domain with volume V. This leads to the convention for the origin of the delta function in k space,

$$\delta(0) = \frac{V}{(2\pi)^n}, \tag{A2}$$

and also to a Fourier-transformation operator $F = |k\rangle\langle x|$, with $F_{kx} = e^{ikx}$, and its inverse $F^\dagger = |x\rangle\langle k|$, with $F^\dagger_{xk} = e^{-ikx}$. The dagger is used to denote transposed and complex conjugated objects. We have $(F^\dagger F)_{xy} = 1_{xy}$ as well as $(FF^\dagger)_{kk'} = 1_{kk'}$ for the following definition of the scalar product of two functions f and g in real and Fourier space:

$$f^\dagger g = \langle f|g\rangle = \int dx\, f^*(x) g(x) = \int \frac{dk}{(2\pi)^n} f^*(k) g(k), \tag{A3}$$

where the asterisk denotes complex conjugation. The statistical power spectrum of f is denoted by $P_f(k) = \langle |f(k)|^2 \rangle_{(f)}/V$.

We also introduce for convenience the position-space componentwise product of two functions

$$(fg)(x) \equiv f(x) g(x), \tag{A4}$$

which also permits compact notations like

$$(\log f)(x) = \log(f(x)), \qquad (f/g)(x) = f(x)/g(x), \tag{A5}$$

and alike. The componentwise product should not be confused with the tensor product of two vectors $(fg^\dagger)(x, y) = f(x) g^*(y)$.

The diagonal components of a matrix M in position-space representation form a vector which we denote by

$$\hat{M} = \mathrm{diag}_x M, \quad \text{with} \quad \hat{M}_x = M_{xx}. \tag{A6}$$

Similarly, a diagonal matrix in position-space representation, whose diagonal components are given by a vector f, will be denoted by

$$\hat{f} = \mathrm{diag}_x f \quad \text{with} \quad \hat{f}_{xy} = f_x 1_{xy}. \tag{A7}$$

Thus, $\hat{\hat{M}} = M$ if and only if M is diagonal, and $\hat{\hat{f}} = f$ always.

In our notation a multivariate Gaussian reads

$$\mathcal{G}(s, S) = \frac{1}{|2\pi S|^{1/2}} \exp\left(-\frac{1}{2} s^\dagger S^{-1} s\right). \tag{A8}$$

Here, $S = \langle s s^\dagger \rangle_{(s)}$ denotes the covariance tensor of the Gaussian field s, which is drawn from $P(s) = \mathcal{G}(s, S)$. If s is statistically homogeneous, S is fully described by the power-spectrum $P_s(k)$:

$$\begin{aligned} S_{kk'} &= (2\pi)^n \delta(k - k') P_s(k), \\ S^{-1}_{kk'} &= (2\pi)^n \delta(k - k') (P_s(k))^{-1}. \end{aligned} \tag{A9}$$

The Fourier representation of the trace of a Fourier-diagonal operator,

$$\mathrm{Tr}(A) = \int dx\, A_{xx} = V \int \frac{dk}{(2\pi)^n} P_A(k), \tag{A10}$$

is very useful in combination with the following expression for the determinant of a Hermitian matrix:

$$\log|A| = \mathrm{Tr}(\log A). \tag{A11}$$

Furthermore, we usually suppress the dependency of probabilities on the underlying model I and its parameters θ in our notation, i.e., instead of $P(s|\theta, I)$ we just write $P(s)$ or $P(s|\theta)$ depending on our focus. Here $\theta = (S, N, R, \ldots)$ contains all the parameters of the model, which are assumed to be known within this work.

APPENDIX B: FEYNMAN RULES ON THE SPHERE

Here, we provide the Feynman rules on the sphere. The real-space rules are identical to those of flat spaces, with just the scalar product replaced by the integral over the sphere, etc. In case the problem at hand has an isotropic propagator, which only depends on the distance of two points on the sphere, but not on their location or orientation, the propagator is diagonal if expressed in spherical harmonics $Y_{lm}(x)$. Thanks to the orthogonality relation of spherical harmonics, we have for $x, y \in S^2$

$$(YY^\dagger)_{xy} = \sum_{lm} Y_{lm}(x) Y^*_{lm}(y) = \delta(x - y) = (1)_{xy}, \tag{B1}$$

and

$$\begin{aligned} (Y^\dagger Y)_{(l,m)(l',m')} &= \int dx\, Y^*_{lm}(x) Y_{l'm'}(x) = \delta_{ll'} \delta_{mm'} \\ &= (1)_{(l,m)(l',m')}. \end{aligned} \tag{B2}$$

Therefore, we can just insert real-space identity matrices $1 = YY^\dagger$ in between any expression in the real-space diagrammatic expression and assign $Y^\dagger$ to the right, and Y to the left term of it. This way we find the spherical-harmonics Feynman rules, which are very similar to the Fourier-space ones, in that they also require directed propagator lines for proper angular-momentum conservation. For a theory with only local interactions, these read

(1) An open end of a line has external (not summed) angular-momentum quantum numbers (l, m).
(2) A line connecting momentum (l, m) with momentum (l', m') corresponds to a propagator between these momenta: $D_{(l,m)(l',m')} = C_D(l) \delta_{ll'} \delta_{mm'}$, where $C_D(l)$ is the angular power spectrum of the propagator.
(3) A data-source vertex is $(j + J - \lambda_1)(l, m)$, where (l, m) is the angular momentum at the data end of the line.
(4) A vertex with quantum number (l_0, m_0) with n_{in} incoming and n_{out} outgoing lines $(n_{\mathrm{in}} + n_{\mathrm{out}} > 1)$ with momentum labels $(l_1, m_1) \cdots (l_{n_{\mathrm{in}}}, m_{n_{\mathrm{in}}})$ and $(l'_1, m'_1) \cdots (l'_{n_{\mathrm{out}}}, m'_{n_{\mathrm{out}}})$, respectively, is given by $-\lambda_m(l_0, m_0) C^{(l'_1,m'_1)\cdots(l'_{n_{\mathrm{out}}},m'_{n_{\mathrm{out}}})}_{(l_0,m_0)\cdots(l_{n_{\mathrm{in}}},m_{n_{\mathrm{in}}})}$, where C will be defined in Eq. (B3).
(5) An internal vertex has internal (summed) angular-momentum quantum numbers (l', m'). Summation means a term $\sum_{l'=0}^{\infty} \sum_{m=-l'}^{l'}$ in front of the expression.
(6) The expression gets divided by the symmetry factor of its diagram.

The interaction structure in spherical-harmonics space is complicated due to the nonorthogonality of powers and products of the spherical harmonic functions, compared to the Fourier-space case, where any power or product of Fourier-basis functions is again a single Fourier-basis function.

The spherical structure is encapsulated in the coefficients

$$C^{(l_1',m_1')\cdots(l_{n_{\rm out}}',m_{n_{\rm out}}')}_{(l_0,m_0)\cdots(l_{n_{\rm in}},m_{n_{\rm in}})} \equiv \int dx \left(\prod_{i=0}^{n_{\rm in}} Y_{l_i m_i}(x)\right)\left(\prod_{i=1}^{n_{\rm out}} Y^*_{l_i' m_i'}(x)\right), \tag{B3}$$

which can be expressed in terms of sums and products of Wigner coefficients, thanks to the relations $Y^*_{lm}(x) = Y_{l,-m}(x)$,

$$Y_{l_1 m_1}(x) Y_{l_2 m_2}(x) = \sum_{lm} \sqrt{\frac{(2l_1+1)(2l_2+1)(2l+1)}{4\pi}} \times \begin{pmatrix} l_1 & l_2 & l \\ m_1 & m_2 & m \end{pmatrix} Y_{lm}(x) \begin{pmatrix} l_1 & l_2 & l \\ 0 & 0 & 0 \end{pmatrix}, \tag{B4}$$

and the orthogonality relation in Eq. (B2), to be applied successively in this order. Because of this complication, it is probably most efficient to calculate propagation in spherical-harmonics space, but to change back to real space for any interaction vertex of high order.

APPENDIX C: $f_{\rm nl}$ PROPAGATOR

We provide in the following the individual terms of the $f_{\rm nl}$ propagator in Eq. (142). The individual diagrams are all $\mathcal{O}(f^2)$ and are given here for the case $f = 1$:

$$= -\mathrm{Tr}\left[D^2 M\right] - \frac{1}{2}\widehat{D}^\dagger M \widehat{D}, \tag{C1}$$

$$= \frac{1}{2}\left[2\widehat{DM} + \widehat{D}M\right]^\dagger D \left[2\widehat{MD} + M\widehat{D}\right], \tag{C2}$$

$$= \mathrm{Tr}\left[D^2 M D M\right] + 2 M_{xy} D_{yy'} M_{y'x'} D_{x'y} D_{xx'}, \tag{C3}$$

$$= j^\dagger D^2 j \tag{C4}$$

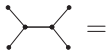
$$= -2\, m^\dagger M D^2 j - 4\,\mathrm{Tr}\left[\widehat{m} D \widehat{j} D M\right], \tag{C5}$$

$$= m^\dagger M D^2 M m + 4\,\mathrm{Tr}\left[\widehat{m} D \widehat{Mm} D M\right] + 2\,\mathrm{Tr}\left[\widehat{m} D (\widehat{m} M + M \widehat{m}) D M\right], \tag{C6}$$

$$= -2\left[2\widehat{DM} + \widehat{D}M\right]^\dagger D \widehat{j} m, \tag{C7}$$

$$= -m^{2\dagger} M \widehat{D} - 2\,\mathrm{Tr}\left[\widehat{m} M \widehat{m} D\right], \tag{C8}$$

$$= \left[2\widehat{DM} + \widehat{D}M\right]^\dagger D \left[2\widehat{m}Mm + Mm^2\right], \tag{C9}$$



$$= \frac{1}{2}\left[2\widehat{m}Mm + Mm^2\right]^\dagger D \left[2\widehat{m}Mm + Mm^2\right], \tag{C10}$$

$$= -2\,(mj)^\dagger D\,(M m^2 + 2\,\widehat{m} M m), \tag{C11}$$

$$= -\frac{1}{2} m^{2\dagger} M m^2, \tag{C12}$$

$$= 2\,(mj)^\dagger D\,(jm). \tag{C13}$$

We used here the conventions $m = Dj$ and $(D^2)_{xy} = (D_{xy})^2$ and remind one that $\Lambda^{(0)} = \Lambda^{(1)} = 0$, $j' = j$, $\Lambda^{(2)} = -2f\hat{j}$, $\Lambda^{(3)}_{xyz} = [M_{xy}\delta_{yz} + 5\ \text{perm}]$, and $\Lambda^{(4)}_{xyzu} = \frac{1}{2}[\delta_{xy} M_{yz} \delta_{zu} + 23\ \text{perm}]$.

[1] T. Bayes, Phil. Trans. R. Soc. London **53**, 370 (1763).
[2] C. E. Shannon, Bell Syst. Tech. J. **27**, 379 (1948).
[3] C. E. Shannon and W. Weaver, *The Mathematical Theory of Communication* (University of Illinois Press, Urbana, IL, 1949).
[4] E. T. Jaynes, Phys. Rev. **106**, 620 (1957).
[5] E. T. Jaynes, Phys. Rev. **108**, 171 (1957).
[6] E. T. Jaynes, *Statistical Physics 3, Brandeis Summer Institute 1962* (W. Benjamin, New York, 1963), p. 181.
[7] E. T. Jaynes, Am. J. Phys. **33**, 391 (1965).
[8] E. T. Jaynes, IEEE Trans. Syst. Sci. Cybern. **4**, 227 (1968).
[9] E. T. Jaynes, Proc. IEEE **70**, 939 (1982).
[10] E. T. Jaynes and R. Baierlein, Phys. Today **57**, 76 (2004).
[11] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. T. Teller, J. Chem. Phys. **21**, 1087 (1953).
[12] W. K. Hastings, Biometrika **57**, 97 (1970).
[13] S. Geman and D. Geman, IEEE Trans. Pattern Anal. Mach. Intell. **pami-6**, 721 (1984).
[14] S. Duan, A. Kennedy, B. Pendleton, and D. Roweth, Phys. Lett. B **195**, 216 (1987).
[15] K. P. N. Murthy, M. Janani, and B. Shenbga Priya, arXiv: cs/0504037.

[16] M. A. Tanner, *Tools for Statistical Inference* (Springer-Verlag, New York, 1996).
[17] R. M. Neal, Technical Report CRG-TR-93-1, Department of Computer Science, University of Toronto, 1993.
[18] C. P. Robert, *The Bayesian Choice* (Springer-Verlag, New York, 2001).
[19] A. Gelman, J. B. Carlin, H. S. Stern, and D. Rubin, *Bayesian Data Analysis* (Chapman & Hall/CRC, Boca Raton, FL, 2004).
[20] R. A. Aster, B. Brochers, and C. H. Thurber, *Parameter Estimation and Inverse Problems* (Elsevier Academic Press, London, 2005).
[21] R. Trotta, Contemp. Phys. **49**, 71 (2008).
[22] N. Wiener, *Extrapolation, Interpolation, and Smoothing of Stationary Time Series* (Wiley, New York, 1949).
[23] W. H. Richardson, J. Opt. Soc. Am. **62**, 55 (1972).
[24] L. B. Lucy, Astrophys. J. **79**, 745 (1974).
[25] B. R. Frieden, J. Opt. Soc. Am. **62**, 511 (1972).
[26] S. F. Gull and G. J. Daniell, Nature (London) **272**, 686 (1978).
[27] J. Skilling, A. W. Strong, and K. Bennett, Mon. Not. R. Astron. Soc. **187**, 145 (1979).
[28] R. K. Bryan and J. Skilling, Mon. Not. R. Astron. Soc. **191**, 69 (1980).
[29] S. F. Burch, S. F. Gull, and J. Skilling, Comp. Vis. Graph. Image Process. **23**, 113 (1983).
[30] S. F. Gull and J. Skilling, in *Proceedings of an International Symposium, Sydney, Australia, 1983*, edited by J. A. Roberts (Cambridge University Press, Cambridge, England, 1984), p. 267.
[31] S. Sibisi, J. Skilling, R. G. Brereton, E. D. Laue, and J. Staunton, Nature (London) **311**, 446 (1984).
[32] D. M. Titterington and J. Skilling, Nature (London) **312**, 381 (1984).
[33] J. Skilling and R. K. Bryan, Mon. Not. R. Astron. Soc. **211**, 111 (1984).
[34] R. K. Bryan and J. Skilling, J. Mod. Opt. **33**, 287 (1986).
[35] S. F. Gull, in *Maximum Entropy and Bayesian Methods*, edited by J. Skilling (Kluwer Academic Publishers, Dordtrecht, The Netherlands, 1989), pp. 53–71.
[36] S. F. Gull and J. Skilling, *The MEMSYS5 User's Manual* (Maximum Entropy Data Consultants Ltd., Royston, 1990).
[37] J. Skilling, in *Maximum Entropy and Bayesian Methods*, edited by G. J. Erickson, J. T. Rychert, and C. R. Smith (Springer, Berlin, 1998), p. 1.
[38] F. S. Kitaura and T. A. Enßlin, Mon. Not. R. Astron. Soc. **389**, 497 (2008).
[39] R. Narayan and R. Nityananda, Annu. Rev. Astron. Astrophys. **24**, 127 (1986).
[40] R. Molina, J. Nunez, F. J. Cortijo, and J. Mateos, IEEE Signal Processing Magazine **18**, 11 (2001).
[41] E. Bertschinger, Astrophys. J. **323**, L103 (1987).
[42] J. N. Fry, Astrophys. J. **289**, 10 (1985).
[43] W. Bialek and A. Zee, Phys. Rev. Lett. **58**, 741 (1987).
[44] W. Bialek and A. Zee, Phys. Rev. Lett. **61**, 1512 (1988).
[45] W. Bialek, C. G. Callan, and S. P. Strong, Phys. Rev. Lett. **77**, 4693 (1996).
[46] P. Stoica, E. G. Larsson, and J. Li, Astrophys. J. **120**, 2163 (2000).
[47] T. Enßlin and M. Frommert (unpublished).
[48] J. C. Lemm, arXiv:physics/9912005.
[49] J. C. Lemm and J. Uhlig, Few-Body Syst. **29**, 25 (2000).
[50] J. C. Lemm, J. Uhlig, and A. Weiguny, Phys. Rev. Lett. **84**, 2068 (2000).
[51] J. C. Lemm and J. Uhlig, Phys. Rev. Lett. **84**, 4517 (2000).
[52] J. C. Lemm, Phys. Lett. A **276**, 19 (2000).
[53] J. C. Lemm, in *Bayesian Inference and Maximum Entropy Methods in Science and Engineering*, edited by A. Mohammad-Djafari, AIP Conf. Proc. No. 568 (AIP, New York, 2001), pp. 425–436.
[54] J. C. Lemm, J. Uhlig, and A. Weiguny, Eur. Phys. J. B **20**, 349 (2001).
[55] J. C. Lemm, J. Uhlig, and A. Weiguny, Eur. Phys. J. B **46**, 41 (2005).
[56] J. C. Lemm, arXiv:cond-mat/9808039.
[57] J. Binney, N. Dowrick, A. Fisher, and M. Newman, *The Theory of Critical Phenomena* (Oxford University Press, Oxford, UK, 1992), ISBN0-19-851394-1.
[58] M. E. Peskin and D. V. Schroeder, *An Introduction to Quantum Field Theory* (Westview Press, Boulder, CO, 1995), ISBN-13 978-0-201-50397-5.
[59] A. Zee, *Quantum Field Theory in a Nutshell*, edited by A. Zee (Princeton University Press, Princeton, NJ, 2003), ISBN 0691010196.
[60] S. Matarrese, F. Lucchin, and S. A. Bonometto, Astrophys. J. **310**, L21 (1986).
[61] Y. B. Zel'dovich, Astron. Astrophys. **5**, 84 (1970).
[62] J. M. Bardeen, J. R. Bond, N. Kaiser, and A. S. Szalay, Astrophys. J. **304**, 15 (1986).
[63] P. J. E. Peebles, *The Large-Scale Structure of the Universe* Princeton University Press, Princeton, NJ, 1980), p. 435.
[64] N. Kaiser, Mon. Not. R. Astron. Soc. **227**, 1 (1987).
[65] P. J. E. Peebles, Astrophys. J. **362**, 1 (1990).
[66] F. Bernardeau, Astrophys. J. **390**, L61 (1992).
[67] S. Zaroubi and Y. Hoffman, Astrophys. J. **462**, 25 (1996).
[68] A. J. S. Hamilton, in *The Evolving Universe*, edited by D. Hamilton, Astrophysics and Space Science Library Vol. 231 (Kluwer Academic Publishers, Dordtrecht, The Netherlands, 1998), p. 185.
[69] F. Bernardeau, M. J. Chodorowski, E. L. Łokas, R. Stompor, and A. Kudlicki, Mon. Not. R. Astron. Soc. **309**, 543 (1999).
[70] E. Branchini *et al.*, Mon. Not. R. Astron. Soc. **308**, 1 (1999).
[71] A. Dekel and O. Lahav, Astrophys. J. **520**, 24 (1999).
[72] S. Zaroubi, arXiv:astro-ph/0206052.
[73] R. E. Smith, J. A. Peacock, A. Jenkins, S. D. M. White, C. S. Frenk, F. R. Pearce, P. A. Thomas, G. Efstathiou, and H. M. P. Couchman, Mon. Not. R. Astron. Soc. **341**, 1311 (2003).
[74] R. Scoccimarro, Phys. Rev. D **70**, 083007 (2004).
[75] V. Springel *et al.*, Nature (London) **435**, 629 (2005).
[76] P. Valageas, Astron. Astrophys. **421**, 23 (2004).
[77] P. Valageas, Astron. Astrophys. **476**, 31 (2007).
[78] P. Valageas, Astron. Astrophys. **484**, 79 (2008).
[79] M. Crocce and R. Scoccimarro, Phys. Rev. D **73**, 063519 (2006).
[80] M. Crocce and R. Scoccimarro, Phys. Rev. D **73**, 063520 (2006).
[81] P. McDonald, Phys. Rev. D **74**, 103512 (2006).
[82] P. McDonald, Phys. Rev. D **74**, 129901(E) (2006).

[83] P. McDonald, Phys. Rev. D **75**, 043514 (2007).
[84] D. Jeong and E. Komatsu, Astrophys. J. **651**, 619 (2006).
[85] D. Jeong and E. Komatsu, Astrophys. J. **691**, 569 (2009).
[86] S. Matarrese and M. Pietroni, J. Cosmol. Astropart. Phys. 06 (2007) 026.
[87] J. Gaite and A. Domínguez, J. Phys. A **40**, 6849 (2007).
[88] S. Matarrese and M. Pietroni, Mod. Phys. Lett. A **23**, 25 (2008).
[89] T. Matsubara, Phys. Rev. D **77**, 063530 (2008).
[90] M. Pietroni, J. Cosmol. Astropart. Phys. 10 (2008) 036.
[91] E. Bertschinger and A. Dekel, Astrophys. J. **336**, L5 (1989).
[92] E. Bertschinger and A. Dekel, in *Large-Scale Structures and Peculiar Motions in the Universe*, edited by D. W. Latham and L. A. N. da Costa, ASP. Conf. Ser. 15 (Astronomical Society of the Pacific, San Francisco, 1991), p. 67.
[93] P. J. E. Peebles, Astrophys. J. **344**, L53 (1989).
[94] A. Dekel, E. Bertschinger, and S. M. Faber, Astrophys. J. **364**, 349 (1990).
[95] N. Kaiser and A. Stebbins, in *Large-Scale Structures and Peculiar Motions in the Universe*, edited by D. W. Latham and L. A. N. da Costa ASP Conf. Ser. 15 (Astronomical Society of the Pacific, San Francisco, 1991), p. 111.
[96] Y. Hoffman and E. Ribak, Astrophys. J. **380**, L5 (1991).
[97] D. H. Weinberg, Mon. Not. R. Astron. Soc. **254**, 315 (1992).
[98] A. Nusser and A. Dekel, Astrophys. J. **391**, 443 (1992).
[99] G. B. Rybicki and W. H. Press, Astrophys. J. **398**, 169 (1992).
[100] M. Gramann, Astrophys. J. **405**, 449 (1993).
[101] G. Ganon and Y. Hoffman, Astrophys. J. **415**, L5 (1993).
[102] F. Bernardeau, Astron. Astrophys. **291**, 697 (1994).
[103] A. Nusser and M. Davis, Astrophys. J. **421**, L1 (1994).
[104] O. Lahav, in *Unveiling Large-Scale Structures Behind the Milky Way*, edited by C. Balkowski and R. C. Kraan-Korteweg, ASP Conf. Ser. 67 (Astronomical Society of the Pacific, San Francisco, 1994), p. 171.
[105] O. Lahav, K. B. Fisher, Y. Hoffman, C. A. Scharf, and S. Zaroubi, Astrophys. J. **423**, L93 (1994).
[106] K. B. Fisher, O. Lahav, Y. Hoffman, D. Lynden-Bell, and S. Zaroubi, Mon. Not. R. Astron. Soc. **272**, 885 (1995).
[107] R. K. Sheth, Mon. Not. R. Astron. Soc. **277**, 933 (1995).
[108] S. Zaroubi, Y. Hoffman, K. B. Fisher, and O. Lahav, Astrophys. J. **449**, 446 (1995).
[109] M. Tegmark and B. C. Bromley, Astrophys. J. **453**, 533 (1995).
[110] R. A. C. Croft and E. Gaztanaga, Mon. Not. R. Astron. Soc. **285**, 793 (1997).
[111] V. K. Narayanan and D. H. Weinberg, Astrophys. J. **508**, 440 (1998).
[112] U.-L. Pen, Astrophys. J. **504**, 601 (1998).
[113] U. Seljak, Astrophys. J. **503**, 492 (1998).
[114] U. Seljak, Astrophys. J. **506**, 64 (1998).
[115] V. Bistolas and Y. Hoffman, Astrophys. J. **492**, 439 (1998).
[116] A. Taylor and H. Valentine, Mon. Not. R. Astron. Soc. **306**, 491 (1999).
[117] V. K. Narayanan and R. A. C. Croft, Astrophys. J. **515**, 471 (1999).
[118] S. Zaroubi, Y. Hoffman, and A. Dekel, Astrophys. J. **520**, 413 (1999).
[119] D. M. Goldberg and D. N. Spergel, in *Cosmic Flows Workshop*, edited by S. Courteau and J. Willick, ASP Conf. Ser. 201 (Astronomical Society of the Pacific, San Francisco, 2000), p. 282.
[120] D. M. Goldberg and D. N. Spergel, Astrophys. J. **544**, 21 (2000).
[121] A. Kudlicki, M. Chodorowski, T. Plewa, and M. Różyczka, Mon. Not. R. Astron. Soc. **316**, 464 (2000).
[122] S. Basilakos and M. Plionis, Astrophys. J. **550**, 522 (2001).
[123] D. M. Goldberg, Astrophys. J. **552**, 413 (2001).
[124] U. Frisch, S. Matarrese, R. Mohayaee, and A. Sobolevski, Nature (London) **417**, 260 (2002).
[125] S. Zaroubi, Mon. Not. R. Astron. Soc. **331**, 901 (2002).
[126] Y. Brenier, U. Frisch, M. Hénon, G. Loeper, S. Matarrese, R. Mohayaee, and A. Sobolevskiĭ, Mon. Not. R. Astron. Soc. **346**, 501 (2003).
[127] R. Mohayaee, U. Frisch, S. Matarrese, and A. Sobolevskii, Astron. Astrophys. **406**, 393 (2003).
[128] R. Mohayaee, B. Tully, and U. Frisch, arXiv:astro-ph/0410063.
[129] C. S. Botzler, J. Snigula, R. Bender, and U. Hopp, Mon. Not. R. Astron. Soc. **349**, 425 (2004).
[130] R. Mohayaee and R. B. Tully, Astrophys. J. **635**, L113 (2005).
[131] R. Mohayaee, H. Mathis, S. Colombi, and J. Silk, Mon. Not. R. Astron. Soc. **365**, 939 (2006).
[132] V. Icke and R. van de Weygaert, Q. J. R. Astron. Soc. **32**, 85 (1991).
[133] S. Ikeuchi and E. L. Turner, Mon. Not. R. Astron. Soc. **250**, 519 (1991).
[134] F. Bernardeau and R. van de Weygaert, Mon. Not. R. Astron. Soc. **279**, 693 (1996).
[135] W. E. Schaap and R. van de Weygaert, Astron. Astrophys. **363**, L29 (2000).
[136] R. van de Weygaert and W. Schaap, in *Mining the Sky*, edited by A. J. Banday, S. Zaroubi, and M. Bartelmann (Springer, Berlin, 2001), p. 268.
[137] M. Ramella, W. Boschin, D. Fadda, and M. Nonino, Astron. Astrophys. **368**, 776 (2001).
[138] L. Zaninetti, Chin. J. Astron. Astrophys. **6**, 387 (2006).
[139] E. Bertschinger, A. Dekel, S. M. Faber, A. Dressler, and D. Burstein, Astrophys. J. **364**, 370 (1990).
[140] A. Yahil, M. A. Strauss, M. Davis, and J. P. Huchra, Astrophys. J. **372**, 380 (1991).
[141] K. B. Fisher, C. A. Scharf, and O. Lahav, Mon. Not. R. Astron. Soc. **266**, 219 (1994).
[142] E. J. Shaya, P. J. E. Peebles, and R. B. Tully, Astrophys. J. **454**, 15 (1995).
[143] E. Branchini, M. Plionis, and D. W. Sciama, Astrophys. J. **461**, L17 (1996).
[144] M. Webster, O. Lahav, and K. Fisher, Mon. Not. R. Astron. Soc. **287**, 425 (1997).
[145] C. Yess, S. F. Shandarin, and K. B. Fisher, Astrophys. J. **474**, 553 (1997).
[146] I. M. Schmoldt *et al.*, Astron. J. **118**, 1146 (1999).
[147] A. Nusser and M. Haehnelt, Mon. Not. R. Astron. Soc. **303**, 179 (1999).
[148] M. Tegmark and B. C. Bromley, Astrophys. J. **518**, L69 (1999).
[149] Y. Hoffman and S. Zaroubi, Astrophys. J. **535**, L5 (2000).

[150] D. M. Goldberg, Astrophys. J. **550**, 87 (2001).
[151] H. Mathis, G. Lemson, V. Springel, G. Kauffmann, S. D. M. White, A. Eldar, and A. Dekel, Mon. Not. R. Astron. Soc. **333**, 739 (2002).
[152] P. Erdoğdu *et al.*, Mon. Not. R. Astron. Soc. **352**, 939 (2004).
[153] M. S. Vogeley, F. Hoyle, R. R. Rojas, and D. M. Goldberg, in *IAU Colloq. 195: Outskirts of Galaxy Clusters: Intense Life in the Suburbs*, edited by A. Diaferio (Cambridge University Press, Cambridge, England, 2004), pp. 5–11.
[154] J. Huchra, T. Jarrett, M. Skrutskie, R. Cutri, S. Schneider, L. Macri, R. Steining, J. Mader, N. Martimbeau, and T. George, in *Nearby Large-Scale Structures and the Zone of Avoidance*, edited by A. P. Fairall and P. A. Woudt, ASP Conf. Ser. 329 (Astronomical Society of the Pacific, San Francisco, 2005), p. 135.
[155] W. J. Percival, Mon. Not. R. Astron. Soc. **356**, 1168 (2005).
[156] P. Erdoğdu *et al.*, Mon. Not. R. Astron. Soc. **373**, 45 (2006).
[157] Y. Hoffman, in *Unveiling Large-Scale Structures Behind the Milky Way*, edited by C. Balkowski and R. C. Kraan-Korteweg, ASP Conf. Ser. 67 (Astronomical Society of the Pacific, San Francisco, 1994), p. 185.
[158] S. Zaroubi, in *Mapping the Hidden Universe: The Universe Behind the Mily Way - The Universe in HI*, edited by R. C. Kraan-Korteweg, P. A. Henning, and H. Andernach, ASP Conf. Ser. 218 (Astronomical Society of the Pacific, San Francisco, 2000), p. 173.
[159] R. C. Kraan-Korteweg and O. Lahav, Astron. Astrophys. Rev. **10**, 211 (2000).
[160] J. A. Peacock and S. J. Dodds, Mon. Not. R. Astron. Soc. **267**, 1020 (1994).
[161] M. S. Vogeley and A. S. Szalay, Astrophys. J. **465**, 34 (1996).
[162] S. Zaroubi, I. Zehavi, A. Dekel, Y. Hoffman, and T. Kolatt, Astrophys. J. **486**, 21 (1997).
[163] M. Tegmark, Phys. Rev. Lett. **79**, 3806 (1997).
[164] D. J. Eisenstein and W. Hu, Astrophys. J. **511**, 5 (1999).
[165] G. Efstathiou, J. R. Bond, and S. D. M. White, Mon. Not. R. Astron. Soc. **258**, 1P (1992).
[166] E. F. Bunn, D. Scott, and M. White, Astrophys. J. **441**, L9 (1995).
[167] M. A. Janssen and S. Gulkis, in *The Infrared and Submillimetre Sky after COBE*, edited by M. Signore and C. Dupraz, NATO ASIC Proc. 359 (Kluwer Academic Publishers, Dordrecht, The Netherlands, 1992), pp. 391–408.
[168] E. F. Bunn, K. B. Fisher, Y. Hoffman, O. Lahav, J. Silk, and S. Zaroubi, Astrophys. J. **432**, L75 (1994).
[169] K. Maisinger, M. P. Hobson, and A. N. Lasenby, Mon. Not. R. Astron. Soc. **290**, 313 (1997).
[170] M. Tegmark, Phys. Rev. D **56**, 4514 (1997).
[171] M. Tegmark, Astrophys. J. **480**, L87 (1997).
[172] S. Dodelson, Astrophys. J. **482**, 577 (1997).
[173] M. P. Hobson, A. W. Jones, A. N. Lasenby, and F. R. Bouchet, Mon. Not. R. Astron. Soc. **300**, 1 (1998).
[174] P. Natoli, G. de Gasperis, C. Gheller, and N. Vittorio, Astron. Astrophys. **372**, 346 (2001).
[175] O. Doré, R. Teyssier, F. R. Bouchet, D. Vibert, and S. Prunet, Astron. Astrophys. **374**, 358 (2001).
[176] R. Stompor *et al.*, Phys. Rev. D **65**, 022003 (2001).
[177] B. D. Wandelt, D. L. Larson, and A. Lakshminarayanan, Phys. Rev. D **70**, 083511 (2004).
[178] H. K. Eriksen, I. J. O'Dwyer, J. B. Jewell, B. D. Wandelt, D. L. Larson, K. M. Górski, S. Levin, A. J. Banday, and P. B. Lilje, Astrophys. J. Suppl. Ser. **155**, 227 (2004).
[179] J. Jewell, S. Levin, and C. H. Anderson, Astrophys. J. **609**, 1 (2004).
[180] D. Yvon and F. Mayet, Astron. Astrophys. **436**, 729 (2005).
[181] E. Keihänen, H. Kurki-Suonio, and T. Poutanen, Mon. Not. R. Astron. Soc. **360**, 390 (2005).
[182] E. C. Sutton and B. D. Wandelt, Astrophys. J. Suppl. Ser. **162**, 401 (2006).
[183] D. L. Larson, H. K. Eriksen, B. D. Wandelt, K. M. Górski, G. Huey, J. B. Jewell, and I. J. O'Dwyer, Astrophys. J. **656**, 653 (2007).
[184] G. Hinshaw *et al.* (WMAP Collaboration), Astrophys. J. Suppl. Ser. **180**, 225 (2009).
[185] U. Seljak and M. Zaldarriaga, Astrophys. J. **469**, 437 (1996); also http://cmbfast.org, http://ascl.net/cmbfast.html.
[186] A. Lewis, A. Challinor, and A. Lasenby, Astrophys. J. **538**, 473 (2000); also http://camb.info/.
[187] M. Doran, J. Cosmol. Astropart. Phys. 10 (2005) 011; also http://www.cmbeasy.org/.
[188] E. F. Bunn and N. Sugiyama, Astrophys. J. **446**, 49 (1995).
[189] M. Tegmark, A. N. Taylor, and A. F. Heavens, Astrophys. J. **480**, 22 (1997).
[190] M. Tegmark, Phys. Rev. D **55**, 5895 (1997).
[191] M. R. Nolta *et al.* (WMAP Collaboration), Astrophys. J. Suppl. Ser. **180**, 296 (2009).
[192] A. H. Guth, Phys. Rev. D **23**, 347 (1981).
[193] A. D. Linde, Phys. Lett. **108B**, 389 (1982).
[194] A. Albrecht and P. J. Steinhardt, Phys. Rev. Lett. **48**, 1220 (1982).
[195] A. H. Guth and S.-Y. Pi, Phys. Rev. Lett. **49**, 1110 (1982).
[196] A. A. Starobinsky, Phys. Lett. **117B**, 175 (1982).
[197] J. M. Bardeen, P. J. Steinhardt, and M. S. Turner, Phys. Rev. D **28**, 679 (1983).
[198] W. Hu, Phys. Rev. D **64**, 083005 (2001).
[199] F. Bernardeau and J.-P. Uzan, Phys. Rev. D **66**, 103506 (2002).
[200] N. Bartolo, E. Komatsu, S. Matarrese, and A. Riotto, Phys. Rep. **402**, 103 (2004).
[201] D. Babich, P. Creminelli, and M. Zaldarriaga, J. Cosmol. Astropart. Phys. 08 (2004) 009.
[202] E. Komatsu, B. D. Wandelt, D. N. Spergel, A. J. Banday, and K. M. Górski, Astrophys. J. **566**, 19 (2002).
[203] D. Babich and M. Zaldarriaga, Phys. Rev. D **70**, 083005 (2004).
[204] E. Komatsu, D. N. Spergel, and B. D. Wandelt, Astrophys. J. **634**, 14 (2005).
[205] A. P. S. Yadav, E. Komatsu, and B. D. Wandelt, Astrophys. J. **664**, 680 (2007).
[206] A. P. S. Yadav, E. Komatsu, B. D. Wandelt, M. Liguori, F. K. Hansen, and S. Matarrese, Astrophys. J. **678**, 578 (2008).
[207] E. Komatsu *et al.*, Astrophys. J. Suppl. Ser. **148**, 119 (2003).
[208] A. Curto, J. F. Macias-Perez, E. Martinez-Gonzalez, R. B.

Barreiro, D. Santos, F. K. Hansen, M. Liguori, and S. Matarrese, arXiv:0804.0136.
[209] A. P. S. Yadav and B. D. Wandelt, Phys. Rev. Lett. **100**, 181301 (2008).
[210] E. Martinez-Gonzalez, arXiv:0805.4157.
[211] J. Jasche, F. S. Kitaura, and T. A. Ensslin, arXiv:0901.3043.
[212] P. Coles and B. Jones, Mon. Not. R. Astron. Soc. **248**, 1 (1991).
[213] R. Vio, P. Andreani, and W. Wamsteker, Publ. Astron. Soc. Pac. **113**, 1009 (2001).
[214] M. C. Neyrinck, I. Szapudi, and A. S. Szalay, Astrophys. J. **698**, L90 (2009).
[215] R. K. Sachs and A. M. Wolfe, Astrophys. J. **147**, 73 (1967).
[216] M. J. Rees and D. W. Sciama, Nature (London) **217**, 511 (1968).
[217] J. R. Fergusson and E. P. S. Shellard, Phys. Rev. D **80**, 043510 (2009).
[218] E. Komatsu and D. N. Spergel, Phys. Rev. D **63**, 063002 (2001).
[219] N. Kogo and E. Komatsu, Phys. Rev. D **73**, 083007 (2006).
[220] D. Babich, Phys. Rev. D **72**, 043003 (2005).
[221] A. F. Heavens, Mon. Not. R. Astron. Soc. **299**, 805 (1998).
[222] P. Creminelli, A. Nicolis, L. Senatore, M. Tegmark, and M. Zaldarriaga, J. Cosmol. Astropart. Phys. 05 (2006) 004.
[223] A. P. Yadav and B. D. Wandelt, Phys. Rev. D **71**, 123004 (2005).