Information field theory for cosmological perturbation reconstruction and non-linear signal analysis

Torsten A. Enßlin, Mona Frommert, and Francisco S. Kitaura
Max-Planck-Institut für Astrophysik, Karl-Schwarzschild-Str. 1, 85741 Garching, Germany
(Dated: June 20, 2008)

We develop *information field theory* (IFT) as a means of Bayesian, data based inference on spatially distributed signals, the information fields. A didactic approach is attempted in order to enable scientists not familiar with field theories to implement and apply inference algorithms derived within the IFT framework. 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. The theory should be applicable in many fields. However, here, two cosmological signal recovery problems are discussed in detail in their IFT-formulation. 1) Reconstruction of the cosmic large-scale structure matter distribution from discrete galaxy counts in incomplete galaxy surveys. It is demonstrated analytically and numerically that a Gaussian signal, which should resemble the initial density perturbations of the Universe, observed with a strongly non-linear, 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-renormalisation flow equation. Surprisingly, solving this equation numerically is much less expensive than solving the corresponding classical field equation, which means to calculate the maximum a posteriori estimator, despite the former's higher fidelity. 2) We design a filter to detect any possible local non-linearities in the cosmic microwave background (CMB), which are predicted from some Early-Universe inflationary scenarios, and expected due to measurement imperfections. This filter is optimal up to linear order in the non-linearity parameter, which is our signal, and can be used even to construct sky maps of non-linearities in the data. Since the filter uses up to 4$^{\text{th}}$ order data correlation functions, whereas current non-linearity filters rely on the bispectrum, which is 3$^{\text{rd}}$ order, its implementation may help to improve the detectability of this important messenger from the inflationary epoch. Finally, we provide the Boltzmann-Shannon information measure of IFT based on the Helmholtz free energy, thereby highlighting conceptual similarities of information and statistical field theory and outlining how to optimise observational strategies for maximal information retrieval.

I. INTRODUCTION

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 large-scale structure, for the interpretation of the cosmic microwave background (CMB), as well 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 probabilistic uncertainty, offers an ideal framework to handle such problems. It permits to describe all relevant processes involved in the measurement probabilistically, provided a model for the Universe is adopted. The possible physical realities defined here as states of such a model, denoted by the state variable ψ, 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 is 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 dataset d given the physical condition ψ. One could argue that given the complete state of the Universe ψ the outcome d of the measurement should be uniquely determined, since even in a quantum-physical Universe the Schrödinger equation is deterministic. In that case $P(d|\psi) = \delta(d - d[\psi])$, 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 path integral over all possible realisations of ψ, to be defined more precisely later (Sect. I 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 joined 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 is available, the phase-space of interest is spanned by the direct product of all possible signals s and data d, and all regions with non-zero $P(d, s)$ are of potential relevance. Once the actual data $d_{\rm obs}$ have been taken, only a sub-manifold of this space, as fixed by the data, is of further relevance. The probability function over this sub-space is proportional to $P(d = d_{\rm obs}, s)$, and needs just to be renormalised by dividing by

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

which is the unconditioned probability (or evidence) of that data. Thus, we find the resulting information of the data to be the posterior distribution $P(s|d_{\rm obs}) = P(d_{\rm obs}, s)/P(d_{\rm obs})$. 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 normalisation 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

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

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.

These definitions were chosen to be close to the usual language in signal processing and data analysis. They permit 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 non-trivial response, since both variables just need to exhibit some couplings to a common sub-aspect 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

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

except for linear transformations, $d' = T\, d$, 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 space can be changed via a data transformation T, and the transformed data may exhibit 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 $m_d = \langle s \rangle_{(s|d)}$ to be the map of the signal given the data d and call T a *map-making-algorithm* if it fulfils 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 non-linear 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 characterised by the quadratic signal uncertainty,

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

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

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

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 it to be optimal in an information theoretical sense according to Eq. 14.

An illustrative example should be in order. Suppose our data is 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 contains 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 can not 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 deduction process:

1. The functional form of $s[\psi]$ should best be simple, steady, analytic, and if possible linear in ψ, permitting to use the signal s to reason about the state of reality ψ with the help of the concept of a response function.

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 to deduce properties of s from d. Signal degrees of freedoms, which are insensitive to the data, will only be constrained by the prior and therefore just contain a large amount of uncertainty, which adds to the error budget, and should thus be avoided as far as possible. A good trade-off between signal response and noise on the one hand and signal uncertainty on the other hand should be considered. It should be noted, that dropping relatively unresponsive parts of the signal, and thereby increasing the data noise, typically leads to an *improved* signal recoverability, assuming an optimal information theoretical signal estimator is used.

3. The choice of $s[\psi]$ should also be lead by mathematical convenience and practicality. In the examples presented in this work, simple signals are chosen which permit 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 characterised by the 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 large-scale structure 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 non-trivial way on the small-scale modes. Due to 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 $s[\psi] = F\,\psi$, 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. Information field theory

Why this "philosophical" introduction to a mostly technical paper? The fact that the signal can be tailored to specific needs should also make us aware, that while interpreting data we are not reconstructing a physical reality, but only a condensed description of it, our chosen signal. We are gathering abstract information and this information can be about spatially distributed physical quantities in cosmological applications. Our signal is therefore an information field which should conform to the laws of information theory, or more specifically, to highlight the technical aspects of dealing with distributed quantities, to *information field theory* (IFT), which we want to introduce here.[1]

Physical fields and the information fields derived from them have a number of similarities, motivating the terminology chosen here, but also exhibit a number of distinct differences, explaining why we insist on not to identify an information field with its physical counterpart.

Physical fields form the physical reality, whereas information fields are only abstract and compressed descriptions of that physical reality. Physical reality should be understood here as the state of our physical model, which actually is a philosophical concept, and to some degree even subjective. Therefore, we can also state that information fields form a knowledge reality, which is obviously and explicitly subjective due to its dependence on the data available, and the data model used to construct the field. Physical fields may interact among themselves, whereas information fields are derived from the former but do not influence them.[2]

Nevertheless, also information fields interact among themselves, since two pieces of information can add up in a non-linear way and interaction terms in information Hamiltonians can have a similar structure to those of physical Hamiltonians. Physical signals travel at most with the speed of light, and any interaction of physical fields should be local due to physical causality. Information field theory, however, can contain non-local interactions of information fields of causally unconnected regions of space-time. This is for example possible in case a cosmological principle is assumed to be valid in the (model) Universe. This permits us to use conclusions derived from data received from one region of our Universe to understand properties of another part. If the data were received as light from two opposite cosmic directions, the two regions observed may never have been in causal contact during recent cosmic history.

Although our main motivation is the reconstruction of cosmic density and gravitational fields, and inference on their statistical properties, and our examples are geared towards such applications, it should be clear that the presented theory has more general applications. Even the presented concrete cosmological problems may be useful in other contexts as well with little modifications. For example, the large-scale structure reconstruction method based on galaxy counts of Sect. IV can be used for image reconstruction with low-number photon statistics, e.g in low-dose X-ray imaging. The CMB non-Gaussianity estimator of Sect. V may serve as a blueprint for statistical monitoring of the linearity of a signal amplifier.

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 sub-volume 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.[3] 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:

$$\begin{aligned} \langle s(x_1)\cdots s(x_n)\rangle_d \;&\equiv\; \langle s(x_1)\cdots s(x_n)\rangle_{(s|d)} \\ &\equiv\; \int \mathcal{D}s\, s(x_1)\cdots s(x_n)\, P(s|d). \quad (19) \end{aligned}$$

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 formulae 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 large-scale structure (detailed references are given in Sect. I F). The purpose of this

[1] We prefer the term *information field theory (IFT)* versus the proposal of *Bayesian field theory (BFT)* by Lemm [2] since the term IFT puts the emphasis on the relevant object, the information, whereas BFT refers to a method, Bayesian inference. The Bayesian methods are unavoidably needed while dealing with information, however, also other methods might be required. Furthermore, the term *information field* is rather self-explaining, whereas the meaning of a *Bayesian field* is not that obvious.

[2] At least no significant influence of a knowledge state described by an information field is expected in usual physical and especially cosmological settings. However, if IFT is to be applied to sociological systems, information on the state of distributed human motives might even change that state. This is well known e.g. in economics, where the information on a believe in an economical state of a company can have large influence on the same state due to the induced decisions of the companie's customers, especially if the company is a bank. However, such complications should not be the topic here.

[3] We are mostly dealing with scalar fields, however, multi-component, vector or tensor fields can be treated analogously, and many of the equations just have to be re-interpreted for such fields and stay valid.

work is to show how to expand other posterior distributions around the Gaussian ones in a perturbative manner, which then permits to use the existing map-making codes for the computation of the resulting diagrammatic perturbation series. Since the diagrammatic perturbation series in Feynman-diagrams are well known and understood in quantum- and statistical field theory (QFT & 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.

Formally, IFT can be regarded as a statistical field theory. However, there is a philosophical difference. In SFT any estimator is an expectation for a fluctuating field, and given a sufficient long observing period, the system will (usually) exhibit a time average identical to the estimator. In IFT, however, there exists presumably only a single realisation of the field. Its value is not known and therefore uncertainties exists, which are described probabilistically. Thus, an identical mathematical description of the uncertainties in IFT and the fluctuations in the SFT is used, but their meaning is conceptually different in that in one case it is missing knowledge on the system, and in the other case it is the description of a (rapidly) changing state. The missing knowledge may be supplemented by later observations, whereas the fluctuations of a thermodynamical system persist even after precise measurements as long the system is coupled to a heat bath.

Furthermore, adding constants to a QFT- or SFT-Hamiltonian does not change the predictions of the theory at all. In IFT, such constants are often meaningful in cases where parameters of the theory themselves have uncertainties and are to be estimated from the data or marginalised over. This is a clear difference to QFT, where, e.g., coupling constants and particle masses are assumed to be fixed (at a given energy), and not attributed with quantum-mechanical probability amplitudes which are part of the dynamics. Such parameter uncertainties can lead, as we will show in a follow-up paper, to non-local couplings in IFT. Inference based on data from one region may be used in conjunction with the cosmological principle to predict the properties of another corner of the Universe, which is the cause of such non-local information couplings. This is different to QFT, where only local coupling are possible due to the physical causality of space-time.

E. Signal and data spaces

1. *Discretisation and continuous limit*

Both, the signal and the data space may be continuous, however, in practice will most often be discrete since digital data processing only permits to chose a discretised representation of the distributed information. The space in which the data and signal discretisation 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 analyse a continuous signal, the computationally required discretisation will force an implicit redefinition of our actual signal to be the discretely sampled version of that continuous signal, and this discretisation step should also be part of the data model, if it has the potential to significantly affect the analysis.

Although discretisation implies some information loss it also has an advantage. Scientists, who are not so familiar with functional analysis and field theoretical calculations, can assume discretisation and therefore read all scalar and tensor products as being the usual, component-wise ones, now just in high-, but finite-dimensional vector spaces. This may lower the threshold for a broader usage of algorithms derived from IFT.

To be concrete, let $\{x_i\} \subset \Omega$ be a discrete set of $N_{\rm pix}$ pixel positions, each of which has a volume-size V_i attributed to it, then the scalar product of two discretised 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

$$g^\dagger f \equiv \sum_{i=1}^{N_{\rm pix}} V_i \, \bar{g}_i \, f_i. \tag{20}$$

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

$$g^\dagger f \longrightarrow \int dx \, \bar{g}(x) \, f(x). \tag{21}$$

In many cases the actual volume normalisation in Eq. 20 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 discretised 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 $F[f] \equiv F(f_1, \ldots, f_{N_{\rm pix}})$ over all realisations of such a discretised field f is then just a high-dimensional volume integral, with as many dimensions as pixels:

$$\int \mathcal{D}f \, F[f] \equiv \left(\prod_{i=1}^{N_{\rm pix}} \int df_i \right) F(f_1, \ldots, f_{N_{\rm pix}}). \tag{22}$$

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

$$\langle 1 \rangle_{(f)} = \int \mathcal{D}f\, P(f) = \prod_{i=1}^{N_{\rm pix}} \underbrace{\int df\, P_i(f)}_{=1} = 1. \tag{23}$$

The probabilities of the individual pixels may depend on the pixel volume, especially if the signal is an integral quantity like the number of galaxies within a pixel volume, or it might be independent of the volumes, e.g. if the signal is a physical field sampled only at the pixel positions.

Although it is practically never required in real data-analysis applications to perform the continuous limit $N_{\rm pix} \to \infty$ with $V_i \to 0$ 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. III 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 discretisation on scales smaller than the structures of the signal, can be expected to provide identical results up to a small discretisation difference, which vanishes with higher discretisation-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. V, 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, $\Omega_{\rm parameter} = \{p\}$ 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 require 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 multi-dimensional 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 realisations of this knowledge state field function via

$$P[P_{(p|d)}] = \int \mathcal{D}P_{(p|d)}\, \delta\left[P_{(p|d)} - \int dq\, P(p,q|d)\right]. \tag{24}$$

In case that q is a field, the marginalisation integral in the delta functional also becomes a path-integral. Probabilistic decision theory, based on knowledge state 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 marginalised probability

$$P(p|d) = \int dq\, P(p,q|d) \tag{25}$$

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

F. 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 (information theory, image reconstruction, statistical field theory, cosmological large-scale structure, and cosmic microwave background), for the orientation of non-expert readers. An expert in any of these fields might decide to skip the corresponding sections.

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 apologise 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 towards 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, 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 [3, 4] 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 [5, 6, 7, 8, 9, 10, 11]. 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, following the ideas of Metropolis et al. [12], Hastings [13], and Geman and Geman [14], where the latter authors already had image reconstruction applications in mind. With such tools, higher dimensional problems, as present in signal restauration, could and can be tackled, however, for the price of getting stochastic uncertainty into the computational results.

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 [15, 16, 17, 18, 19]. A review of Bayesian inference methods used for parameter estimation and model selection in cosmology is given by Trotta [20].

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 [21], the Richardson-Lucy algorithm [22, 23], and the maximum-entropy image restauration [24].[4]

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 Sect. II C. It will be the working horse of the IFT formalism, since the Wiener filter represents the algorithm to construct the exact field theoretical expectation value given the data for a non-linear 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 has usually to be regularised by hand, by truncation of the iterative calculations, against an over-fitting instability due to the missing (or implicitly flat) signal prior. A Gaussian-prior based regularisation was recently proposed by Kitaura and Enßlin [37], and the implementation of a variant of this is presented here in Sect. IV D.

The maximum entropy method can be regarded again as being fully Bayesian, however, for a slightly artificial data model of some ensemble of image signal packages or "photons" trying to occupy that state of the image-phase space, which has least information according to the Boltzmann-Shannon measure, but is still consistent with the data. The tradeoff between being data-conform and having maximal entropy is set by an ad-hoc control variable. For a classical review on maximum entropy methods see [38]. 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].

3. *Statistical and Bayesian field theory*

The relation of signal reconstruction problems and field theory was discovered independently by several authors. In cosmology, the most prominent work in this directions is probably by Bertschinger [40], 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 non-linear, non-Gaussian extension of this.

Simultaneously to Bertschinger's work, Bialek and Zee [41, 42] 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 summarised by a nuisance field. A probabilistic language was used, but no direct reference to information theory was made, since not the optimal information reconstruction was the aim, but a model for the human visual reception system. However, this work actually triggered our research.

Bialek et al. [43] applied a field theoretical approach to recover a probability distribution from data. Here, a Bayesian prior was used to regularise the solution, which was set up ad-hoc to enforce smoothness of the reconstruction, obtained from the classical (or saddlepoint, or maximum a posteriori) 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. [44] and by a follow up publication to ours. Bialek et al. [43] also recognised, as we do, that an IFT can easily be non-local.

Finally, the work of Lemm and coworkers[5] established a tight connection between statistical field theory and Bayesian inference is established, and proposed the term

---

[4] See also [24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36].

[5] [2, 45, 46, 47, 48, 49, 50, 51].

Bayesian field theory for this. The applications 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 summarising the essential insights of these papers, [2], 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 [52] is remarkable, since it provides explicit recipes of how to implement a priori information in various circumstances more rigorously.

The field theoretical mathematical tools required to tackle IFT problems come from statistical and quantum field theory, which have a vast literature. We have specially made use of the books of Binney et al. [53], Peskin and Schroeder [54], and Zee [55].

4. Cosmological large-scale structure

Our first IFT example in Sec. IV is geared towards 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 large-scale structure 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 intitial 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 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 non-linear 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 large-scale structure to be partially degenerated. There are analytical methods to describe these effects [6], 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.

In recent years, it was recognised that the evolution of the cosmic density field and its statistical properties can be addressed with field theoretical methods by virtue of renormalisation flow equations. Detailed semi-analytical 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[7] .

It was recognised early on, that the primordial density fluctuations can in principle be reconstructed from galaxy observations [40]. This has lead to a large development of various numerical techniques for an optimal reconstruction[8]. 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[9] . A discussion and classification of the various methods can be found in [37].

Especially the Wiener filter methods were extensively applied to galaxy survey data[10] and permitted partly to extrapolate the matter distribution into the *zone of avoidance* behind the galactic disk and to close the datagap there, c.f. [152, 153, 154], a topic we also address in Sect. IV.

Another cosmological relevant information field to be extracted from galaxy catalogues is the large-scale structure power spectrum[11]. This power is also measurable in the CMB, and for a long time the CMB provided the best spectrum normalisation [160, 161].

5. Cosmic Microwave Background

Since our second example deals with the CMB, we give a brief overview on it and on 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 photonbaryon fluid, which decouples at that epoch into neutral Hydrogen and free streaming photons, has responded to the gravitational pull of the then already forming dark matter structures. The photons from that epoch cooled

[6] Of special interest in this context may be [56], which already applies path-integrals, and [57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70] and the papers they refer to.

[7] Relevant references are [71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85].

[8] Some of the main references are: [86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, 123, 124, 125, 126].

[9] E.g. [127, 128, 129, 130, 131, 132, 133].

[10] Survey based reconstructions of the cosmic matter fields can be found in [134, 135, 136, 137, 138, 139, 140, 141, 142, 143, 144, 145, 146, 147, 148, 149, 150, 151].

[11] E.g. [155, 156, 157, 158, 159].

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 Kelvin and fluctuations of the order of 10^{-5} Kelvin, imprinted by the primordial gravitational potentials at decoupling.

Therefore, mapping these temperature fluctuations precisely permits to 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[12], 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 exists permitting to predict it for a given set of cosmological parameters. Well known codes for this task are publically available[13] and permit 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 recognised early on that this should happen in an information theoretically optimal way, and Bayesian methods were therefore adapted in that area well before in other astrophysical disciplines [e.g. 183, 184, 185, 186].

The initial metric and density fluctuations, from which the CMB fluctuations and the large-scale structure 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 [187, 188, 189, 190, 191, 192]. 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 [e.g. 193, 194, 195, 196]. Observational tests on such non-Gaussianities based on the three-point correlation function of the CMB data [e.g. 197, 198, 199, 200, 201] were so far mostly negative, however not sensitive enough to seriously constrain the possible theoretical parameter space of inflationary scenarios, see e.g. [202, 203]. Recently, there has been the claim of detection of such non-Gaussianities by Yadav and Wandelt [204], and a confirmation of this with better data and improved algorithms is therefore highly desirable. In Sect. V 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 [205].

G. Structure of the work

After the discussion of the basic concepts of IFT including the free theory in Sec. II, interacting information fields, their Hamiltonians and Feynman rules are introduced in Sec. III. The normalisability of sensibly constructed IFTs is shown, as well the classical information field equation is presented there. 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. In Sec. IV the problem of the reconstruction of the cosmic matter distribution from galaxy surveys is analysed in terms of a Poissonain data model. In Sec. V we derive an optimal estimator for non-Gaussianity in the CMB, and show how it can be generalised to map potential non-Gaussianities in the CMB sky. Finally, the Boltzmann-Shannon information measure is introduced in Sec. VI and used to outline how survey strategies could be optimised. Our summary and outlook can be found in Sec. VII.

II. BASICS CONCEPTS

A. Notation

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

A here usually 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

$$f(k) = \langle k|f\rangle = \int dx\, f(x)\, e^{i\, k\cdot x}. \tag{26}$$

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

$$\delta(0) = \frac{V}{(2\,\pi)^n}, \tag{27}$$

[12] E.g.by [162, 163, 164, 165, 166, 167, 168, 169, 170, 171, 172, 173, 174, 175, 176, 177, 178, 179].

[13] E.g. `cmbfast` (`http://cmbfast.org`, `http://ascl.net/cmbfast.html`, [180]), `camb` (`http://camb.info/`, [181]), and `cmbeasy` (`http://www.cmbeasy.org/`, [182]).

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

$$f^\dagger g = \langle f|g\rangle = \int dx\, \overline{f(x)}\, g(x) = \int \frac{dk}{(2\,\pi)^n}\, \overline{f(k)}\, g(k), \tag{28}$$

where the bar 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 component-wise product of two functions

$$(f\,g)(x) \equiv f(x)\,g(x), \tag{29}$$

which also permits compact notations like

$$(\log f)(x) = \log(f(x)),\ (f/g)(x) = f(x)/g(x), \tag{30}$$

and alike. The component-wise product should not be confused with the tensor product of two vectors $(f\,g^\dagger)(x,y) = f(x)\,\overline{g(y)}$.

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

$$\widehat{M} = \text{diag}_x M,\ \text{with}\ \widehat{M}_x = M_{xx}. \tag{31}$$

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

$$\widehat{f} = \text{diag}_x f\ \text{with}\ \widehat{f}_{xy} = f_x\, 1_{xy}. \tag{32}$$

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

In our notation a multivariate Gaussian reads:

$$G(s,S) = \frac{1}{|2\pi S|^{\frac{1}{2}}} \exp\left(-\frac{1}{2} s^\dagger S^{-1} s\right) \tag{33}$$

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

$$\begin{aligned} S_{k\,k'} &= (2\,\pi)^n\,\delta(k-k')\,P_s(k), \\ S^{-1}_{k\,k'} &= (2\,\pi)^n\,\delta(k-k')\,\left(P_s(k)\right)^{-1}. \end{aligned} \tag{34}$$

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

$$\text{Tr}(A) = \int dx\, A_{x\,x} = V \int \frac{dk}{(2\,\pi)^n}\, P_A(k), \tag{35}$$

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

$$\log|A| = \text{Tr}(\log A). \tag{36}$$

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 $\theta = (S, N, R, ...)$ contains all the parameters of the model, which are assumed to be known for the moment.

B. Information Hamiltonian

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 is 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 to express the posterior as

$$P(s|d) = \frac{P(d|s)\,P(s)}{P(d)} \equiv \frac{1}{Z}\, e^{-H[s]}\,. \tag{37}$$

Here, the Hamiltonian

$$H[s] \equiv H_d[s] \equiv -\log\left[P(d,s)\right] = -\log\left[P(d|s)\,P(s)\right], \tag{38}$$

the evidence of the data

$$P(d) \equiv \int \mathcal{D}s\; P(d|s)\,P(s) = \int \mathcal{D}s\; e^{-H[s]} \equiv Z, \tag{39}$$

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 sum

$$Z_d[J] \equiv \int \mathcal{D}s\; e^{-H[s] + J^\dagger s}. \tag{40}$$

This means $P(d) = Z = Z[0]$, but also permits to calculate any moment of the signal field via Fréchet-differentiation of Eq. 40

$$\langle s(x_1)\cdots s(x_n)\rangle_d = \frac{1}{Z}\, \frac{\delta^n\, Z_d[J]}{\delta J(x_1)\cdots\delta J(x_n)}\bigg|_{J=0}\,. \tag{41}$$

Of special importance are the so-called connected correlation functions

$$\langle s(x_1)\cdots s(x_n)\rangle_d^{\text{c}} \equiv \frac{\delta^n\, \log Z_d[J]}{\delta J(x_1)\cdots\delta J(x_n)}\bigg|_{J=0}\,, \tag{42}$$

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:

$$\begin{aligned} \langle s(x)\rangle_d^{\text{c}} &= \langle s(x)\rangle_d, \\ \langle s(x)\,s(y)\rangle_d^{\text{c}} &= \langle s(x)\,s(y)\rangle_d - \langle s(x)\rangle_d\,\langle s(y)\rangle_d, \end{aligned} \tag{43}$$

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

$$\langle s(x_1)\cdots s(x_n)\rangle_d^{\text{c}} = 0 \tag{44}$$

for $n > 2$.

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

$$H[s] = \frac{1}{2}\, s^\dagger D^{-1}\, s - j^\dagger s + H_0 + \sum_{n=3}^{\infty} \frac{1}{n!}\, \Lambda^{(n)}_{x_1 \dots x_n}\, s_{x_1} \cdots s_{x_n}. \quad (45)$$

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

$$H_0 = \log \int \mathcal{D}s \int \mathcal{D}d\; e^{-H'_d[s]}. \quad (46)$$

In many applications H_0 is irrelevant, as long as no comparison between different models or hyperparameter sets are needed, however otherwise, the normalisation constant H_0 is crucial.

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, e.g., to partially reconstruct the signal at locations where no data was 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 $\Lambda^{(n)} = 0$.

C. Free theory

1. *Gaussian data model*

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

$$P(s) = G(s, S) \equiv \frac{1}{|2\pi\, S|^{\frac{1}{2}}}\, \exp\left(-\frac{1}{2} s^\dagger S^{-1} s\right), \quad (47)$$

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

$$d = R\, s + n. \quad (48)$$

Here, the response $R[s] = R\, s$ 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 i-th data point is given by

$$d_i = \int dx\, R_i(x)\, s(x) + n_i. \quad (49)$$

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

$$P(n|s) = G(n, N), \quad (50)$$

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

$$P(d|s) = P(n = d - R\, s|s) = G(d - R\, s, N), \quad (51)$$

and thus the Hamiltonian of the Gaussian theory is

$$\begin{aligned} H_{\mathrm{G}}[s] &= -\log\left[P(d|s)\, P(s)\right] \\ &= -\log\left[G(d - R\, s, N)\, G(s, S)\right] \\ &= \frac{1}{2} s^\dagger D^{-1} s - j^\dagger s + H_0^{\mathrm{G}}\,. \end{aligned} \quad (52)$$

Here

$$D = \left[S^{-1} + R^\dagger N^{-1} R\right]^{-1} \quad (53)$$

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

$$j = R^\dagger N^{-1} d, \quad (54)$$

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

$$j(x) = \sum_{ij} \overline{R_i(x)} N^{-1}_{ij}\, d_j \quad (55)$$

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

$$H_0^{\mathrm{G}} = \frac{1}{2}\, d^\dagger\, N^{-1}\, d + \frac{1}{2}\, \log\left(|2\,\pi\, S|\, |2\,\pi\, N|\right) \quad (56)$$

has absorbed all s-independent normalisation constants.

The partition function of the free field theory,

$$\begin{aligned} Z_{\mathrm{G}}[J] &= \int \mathcal{D}s\, e^{-H_{\mathrm{G}}[s] + J^\dagger s} \quad (57) \\ &= \int \mathcal{D}s\, \exp\left\{-\frac{1}{2} s^\dagger D^{-1} s + (J + j)^\dagger s\, - H_0^{\mathrm{G}}\right\}, \end{aligned}$$

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

$$Z_{\mathrm{G}}[J] = \sqrt{|2\pi\, D|}\, \exp\left\{+\frac{1}{2}(J + j)^\dagger D (J + j) - H_0^{\mathrm{G}}\right\}. \quad (58)$$

The explicit partition function permits to calculate via Eq. 42 the expectation of the signal given the data, in the following called the map m_d generated by the data d:

$$\begin{aligned} m_d &= \langle s\rangle_d = \left.\frac{\delta \log Z_G}{\delta J}\right|_{J=0} = D\, j \\ &= \underbrace{\left[S^{-1} + R^\dagger N^{-1} R\right]^{-1} R^\dagger N^{-1}}_{F_{\mathrm{WF}}}\, d. \end{aligned} \quad (59)$$

The last expression shows that the map is given by the data after applying a generalised Wiener filter, $m_d = F_{\rm WF}\, d$. 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,

$$\langle s s^\dagger \rangle_d^c = D = \left[S^{-1} + R^\dagger N^{-1} R\right]^{-1}, \tag{60}$$

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

$$r = s - m_d \tag{61}$$

provided by the propagator itself, as a straightforward calculation shows:

$$\langle r r^\dagger \rangle_d = \langle s s^\dagger \rangle_d - \langle s \rangle_d \langle s^\dagger \rangle_d = \langle s s^\dagger \rangle_d^c = D. \tag{62}$$

Thus, the posterior should be simply a Gaussian given by

$$P(s|d) = G(s - m_d, D). \tag{63}$$

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

$$\begin{aligned} P(d) &= \frac{P(d|s)\, P(s)}{P(s|d)} = \frac{G(d - R\, s, N)\, G(s, S)}{G(s - D\, j, D)} \\ &= \left(\frac{|D|/|S|}{|2\pi\, N|}\right)^{\frac{1}{2}} \exp\left\{\frac{1}{2}(j^\dagger D\, j - d^\dagger N^{-1} d)\right\}, \end{aligned} \tag{64}$$

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

All of these results of the free theory are well-known within the field of signal reconstruction, and therefore demonstrate how elegantly the information field theoretical approach can be used to reproduce them.

2. *Free classical theory*

The Hamiltonian permits 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 maximum a posteriori (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 [2].

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.

Extremising the Hamiltonian of the free theory (Eq. 52)

$$\left.\frac{\delta H_G}{\delta s}\right|_{s=m} = D^{-1} m - j \equiv 0 \tag{65}$$

we get the classical mapping equation $m = D j$, which is identical to the field theoretical result (Eq. 59).

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

$$\mathcal{H}_{\rm G}[m] = \left.\frac{\delta^2 H[s]}{\delta s^2}\right|_{s=m} = D^{-1}. \tag{66}$$

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

$$\langle r\, r^\dagger \rangle = \mathcal{H}^{-1}[m] = D, \tag{67}$$

which for the pure Gaussian model is of course identical to the exact result, as given by the field theory (Eq. 62).

III. INTERACTING INFORMATION FIELDS

A. Interaction Hamiltonian

1. *General Form*

The assumption that the Hamiltonian can be Taylor expanded in the signal fields permits to write

$$H[s] = \underbrace{\frac{1}{2}\, s^\dagger D^{-1}\, s - j^\dagger s + H_0^{\rm G}}_{H_{\rm G}[s]} + \underbrace{\sum_{n=0}^{\infty} \frac{1}{n!}\, \Lambda^{(n)}_{x_1 \ldots x_n}\, s_{x_1} \cdots s_{x_n}}_{H_{\rm int}[s]}. \tag{68}$$

Repeated coordinates are thought to be integrated over. In contrast to Eq. 45 we have now included perturbations which are constant, linear and quadratic in the signal field, because we are summing from $n = 0$. This permits to treat certain non-ideal 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: $D_{x\,y} = D_{y\,x}$ and $\Lambda^{(n)}_{x_{\pi(1)} \cdots x_{\pi(n)}} = \Lambda^{(n)}_{x_1 \ldots x_n}$ with π any permutation of $\{1, \ldots, n\}$, since even non-symmetric coefficients would automatically be symmetrised by the integration over all repeated coordinates. Therefore, we assume in the following that such a symmetrisation operation has been already done, or we impose it by hand before we continue with any perturbative calculation by applying

$$\Lambda^{(n)}_{x_1 \ldots x_n} \longmapsto \frac{1}{n!} \sum_{\pi \in \mathcal{P}_n} \Lambda^{(n)}_{x_{\pi(1)} \ldots x_{\pi(n)}} \,. \tag{69}$$

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

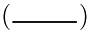
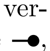
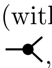
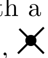
Often, it is more convenient to work with a shifted field $\phi = s - t$, where some (e.g. background) field t is removed from s. The Hamiltonian of ϕ reads

$$\begin{aligned}
H'[\phi] &= \underbrace{\frac{1}{2}\,\phi^\dagger D^{-1}\,\phi - j'^\dagger \phi + H'_0}_{H'_G[\phi]} \\
&\quad + \underbrace{\sum_{n=0}^{\infty} \frac{1}{n!}\,\Lambda'^{(n)}_{x_1 \ldots x_n}\,\phi_{x_1} \cdots \phi_{x_n}}_{H'_{\mathrm{int}}[\phi]}, \text{ with} \\
H'_0 &= H^{\mathrm{G}}_0 - j^\dagger t + \frac{1}{2}\,t^\dagger D^{-1} t, \\
j' &= j - D^{-1}\,t, \text{ and} \\
\Lambda'^{(m)}_{x_1 \ldots x_m} &= \sum_{n=0}^{\infty} \frac{1}{n!} \Lambda^{(m+n)}_{x_1 \ldots x_{m+n}}\,t_{x_1} \cdots t_{x_n}.
\end{aligned} \tag{70}$$

2. Feynman rules

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

$$\begin{aligned}
Z[J] &= \int \mathcal{D}s\, e^{-H[s] + J^\dagger s} \\
&= \int \mathcal{D}s\, \exp\left[-\sum_{n=0}^{\infty} \frac{1}{n!}\,\Lambda^{(n)}_{x_1 \ldots x_n}\, s_{x_1} \cdots s_{x_n} \right] e^{-H_G[s] + J^\dagger s} \\
&= \exp\left[-\sum_{n=0}^{\infty} \frac{1}{n!}\,\Lambda^{(n)}_{x_1 \ldots x_n}\,\frac{\delta}{\delta J_{x_1}} \cdots \frac{\delta}{\delta J_{x_n}} \right] \times \\
&\quad \int \mathcal{D}s\, e^{-H_G[s] + J^\dagger s} \\
&= \exp\left[-H_{\mathrm{int}}[\frac{\delta}{\delta J}] \right]\; Z_G[J].
\end{aligned} \tag{71}$$

There exist well known diagrammatic expansion techniques for such expressions. 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 $j_{x'} + J_{x'} - \Lambda^{(1)}_{x'}$.

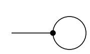
4. Vertices with n legs represent the term $-\Lambda^{(n)}_{x'_1 \ldots x'_n}$, where each individual leg is labeled by one of the internal coordinates $x'_1, \ldots, x'_n$. This more complex vertex-structure, as compared to QFT vertices, is a consequence of non-locality in IFT.

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

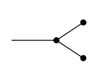
6. Every diagram is divided by its symmetry factor, the number of permutations of vertex legs leaving the topology invariant, as described in any book on field theory.

The n-th moment of s is generated by taking the n-th 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

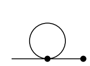
$$\begin{aligned}
&= D\,j = D_{xy}\,j_y \\
&\equiv \int dy\, D(x,y)\, j(y),
\end{aligned}$$



$$\begin{aligned}
&= -\frac{1}{2}\,D\,\Lambda^{(3)}[\cdot, D] = -\frac{1}{2} D_{xy}\,\Lambda^{(3)}_{yzu}\,D_{zu}, \\
&\equiv -\frac{1}{2} \int dy\, D_{xy} \int dz \int du\, \Lambda^{(3)}_{xyu}\, D_{zu}
\end{aligned}$$



$$\begin{aligned}
&= -\frac{1}{2}\,D\,\Lambda^{(3)}\,[\cdot, D\,j, D\,j] \\
&= -\frac{1}{2} D_{xy}\,\Lambda^{(3)}_{yuz}\,D_{zz'}\,j_{z'}\,D_{uu'}\,j_{u'} \\
&\equiv -\frac{1}{2} \int dy\, D_{xy} \int dz \int du\, \Lambda^{(3)}_{yzu} \times \\
&\quad \int dz'\, D_{zz'}\,j_{z'} \int du'\, D_{uu'}\,j_{u'},
\end{aligned} \tag{72}$$



$$\begin{aligned}
&= -\frac{1}{2}\,D\,\Lambda^{(4)}[\cdot, D,\, D\,j] \\
&= -\frac{1}{2} D_{xy}\,\Lambda^{(4)}_{yzuv}\,D_{zu}\,D_{vv'}\,j_{v'}
\end{aligned}$$

$$\equiv -\frac{1}{2}\int dy\, D_{xy} \int dz \int du \int dv\, \Lambda^{(4)}_{yzuv}\, D_{zu}\, \times$$
$$\int dv'\, D_{vv'}\, j_{v'},$$

where 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

$$\Lambda^{(n)}_{x_1 \ldots x_n} = \lambda_n(x_1)\, \delta(x_1 - x_2) \cdots \delta(x_1 - x_n) \tag{73}$$

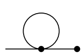
the interaction Hamiltonian reads

$$H_{\mathrm{int}} = \sum_{m=0}^{\infty} \frac{1}{m!}\, \lambda^{\dagger}_m s^m \tag{74}$$

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 72 becomes



$$= -\frac{1}{2}\int dy\, D_{xy}\, \lambda_4(y)\, D_{yy} \int dz\, D_{yz}\, j_z. \tag{75}$$

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 $F^{\dagger}$ to the left and F to the right term, e.g.

$$D\, j = F^{\dagger} \underbrace{F\, D\, F^{\dagger}}_{D'} \underbrace{F\, j}_{j'} = F^{\dagger}\, D'\, j'. \tag{76}$$

Note, that the Fourier space lines are directed since they carry momenta k:

1. An open end of a line has an external momentum coordinate k, and gets an $\int dk\, e^{-\,i\,k\,x}/(2\pi)^n$ 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 lines (m bigger than 1) with momentum lables k_1 ... 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, $j(k) = (F\, j)(k) = \int dx\, j(x) e^{i\,k\,x}$, $D(k, k') = (F\, D\, F^{\dagger})(k, k') = \int dx \int dx' D(x, x')\, e^{i\,(k\,x - k'\,x')}$, 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, 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 characterised by power spectra over the same spatial space,

$$S(k, q) \;=\; (2\pi)^n\, \delta(k - q)\, P_S(k), \tag{77}$$
$$N(k, q) \;=\; (2\pi)^n\, \delta(k - q)\, P_N(k), \tag{78}$$

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,

$$d(x) = \int dy\, R(x - y)\, s(y) + n(x), \tag{79}$$

where 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:

$$D(k, q) \;=\; (2\pi)^n\, \delta(k - q)\, P_D(k), \tag{80}$$

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

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

$$H_{\mathrm{int}}[s] \;=\; \sum_{m=1}^{\infty} \frac{\lambda_m}{m!} \int dx\, s^m(x) \tag{81}$$
$$\;=\; \sum_{m=1}^{\infty} \frac{\lambda_m}{m!} \int \frac{dk_1}{(2\pi)^n}\, s_{k_1} \cdots \int \frac{dk_m}{(2\pi)^n} s_{k_m}\, (2\pi)^n \delta(\sum_{i=1}^{m} k_i)$$

In that case, the Feynman rules simplify considerably. For the interaction Hamiltonian of equation 81, the Feynman rules are now

1. unintegrated x-coordinate: $\exp(-i\,k\,x)$ (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 lines (m bigger than 1): $-\lambda_m$,

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 here. Actually, 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 $x, y \in S^2$

$$(Y\,Y^\dagger)_{xy} = \sum_{lm} Y_{lm}(x)\,Y^*_{lm}(y) = \delta(x-y) = (1)_{xy} \quad (82)$$

and

$$\begin{aligned}(Y^\dagger Y)_{(l,m)(l',m')} &= \int dx\, Y^*_{lm}(x)\,Y_{l'm'}(x) \\ &= \delta_{ll'}\,\delta_{mm'} = (1)_{(l,m)(l',m')}. \end{aligned} \quad (83)$$

Therefore, we can just insert real-space identity matrices $1 = Y\,Y^\dagger$ in between any expression in 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 propagators-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 lables $(l_1, m_1)\ldots(l_{n_{\mathrm{in}}}, m_{n_{\mathrm{in}}})$ and $(l'_1, m'_1)\ldots(l'_{n_{\mathrm{out}}}, m'_{n_{\mathrm{out}}})$, respectively, is given by $-\lambda_m(l_0, m_0)\,C^{(l'_1,m'_1)\ldots(l'_{n_{\mathrm{out}}},m'_{n_{\mathrm{out}}})}_{(l_0,m_0)\ldots(l_{n_{\mathrm{in}}},m_{n_{\mathrm{in}}})}$, where C will be defined in Eq. 84.

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 non-orthogonality of powers and product 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

$$C^{(l'_1,m'_1)\ldots(l'_{n_{\mathrm{out}}},m'_{n_{\mathrm{out}}})}_{(l_0,m_0)\ldots(l_{n_{\mathrm{in}}},m_{n_{\mathrm{in}}})} \equiv \int dx \left(\prod_{i=0}^{n_{\mathrm{in}}} Y_{l_i m_i}(x)\right)\left(\prod_{i=1}^{n_{\mathrm{out}}} Y^*_{l'_i m'_i}(x)\right), \quad (84)$$

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

$$\begin{aligned} Y_{l_1 m_1}(x)\,Y_{l_2\,m_2}(x) \;&=\; \sum_{lm}\sqrt{\frac{(2\,l_1+1)\,(2\,l_2+1)\,(2\,l+1)}{4\,\pi}} \\ &\times \begin{pmatrix} l_1 & l_2 & l \\ m_1 & m_2 & m \end{pmatrix} Y_{lm}(x) \begin{pmatrix} l_1 & l_2 & l \\ 0 & 0 & 0 \end{pmatrix}, \end{aligned} \quad (85)$$

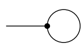
and the orthogonality relation in Eq. 83, to be applied successively in this order. Due to this complication, it is probably most efficient to calculate propagation not in spherical harmonics space, but to change back to real space for any interaction vertex of high order.

B. Normalisability of the theory

In contrast to quantum field theory and also to many applications in statistical field theory, IFT should be properly normalised and not necessarily require any renormalisation 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 joined probability function of data and signal, $H_d[s] \equiv -\log[P(d, s)]$, and therefore well defined and properly normalised if the latter is. Only if ad-hoc Hamiltonians are set up, or if approximations lead to ill-normalised theories, normalisation should be an issue.

However, since we are trying to do 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 renormalisation. 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 III A 4. and estimate a simple loop diagram for which we assume for notational convenience $\lambda_3 = -2\,(2\pi)^n\,\lambda'$ (with $\lambda' > 0$):



$$\begin{aligned} &= -\frac{1}{2}\,D\,\lambda_3\,\widehat{D} \\ &= \lambda' \int dk \int dk'\, \delta(k + k' - k')\, P_D(k)\, P_D(k')\, e^{ikx} \\ &\leq \lambda'\, P_D(0) \int dk'\, P_S(k') = \lambda'\, V\, P_D(0)\, \langle s^2(x) \rangle \end{aligned} \tag{86}$$

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

Thus, as long the signal field is of bounded variance, the loop diagram is convergent due to $P_D(k) \leq P_S(k)$. Even a signal of unbounded variance would not lead to a divergent loop diagram if $\int dk\, P_N(k)/P_R(k)$ is finite, since we also have $P_D(k) \leq P_N(k)/P_R(k)$. The latter scenario is not very natural, since the response might vanish for modes for which there is noise. However, a bounded variance signal is very natural, especially in a cosmological setting. 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 free-streaming of dark matter particles before they became non-relativistic. Also the CMB temperature fluctuations are damped on small scales, due to free streaming of photons around the time of recombination.

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 chose information fields, which are mathematically well behaved, we can therefore ensure convergence of expressions.

Although this is not a general proof of normalisability 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 payed for this well-behavedness is the more complex structure of the propagator, which, in comparison to QFT, even in simplistic cases can be non-analytical and require numerical evaluation.

C. Expansion around the classical solution

1. General case

The classical solution of the Hamiltonian in Eq. 68 is provided by its minimum,

$$\frac{\delta H}{\delta s_x} = D^{-1}_{x\,y} s_y - j_x + \sum_{m=1}^{\infty} \frac{1}{m!}\, \Lambda^{(m+1)}_{x_1 \dots x_m x}\, s_{x_1} \dots s_{x_m} = 0. \tag{87}$$

This leads to the equation for the classical field

$$s^{\mathrm{cl}}_y = D_{y\,x} \left(j_x - \sum_{m=1}^{\infty} \frac{1}{m!}\, \Lambda^{(m+1)}_{x_1 \dots x_m x}\, s^{cl}_{x_1} \dots s^{cl}_{x_m} \right), \tag{88}$$

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_{cl} is

$$s_{\mathrm{cl}} = D \left(j - \sum_{m=1}^{\infty} \frac{\lambda^{\dagger}_{m+1}}{m!}\, s^m_{\mathrm{cl}} \right). \tag{89}$$

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 non-classical uncertainty field $\phi = s - s_{\mathrm{cl}}$.

A non-classical expansion of the information field around the classical field is possible by inserting $s = s_{\mathrm{cl}} + \phi$ into the Hamiltonian (Eq. 74). Reordering terms according to the powers of the field ϕ leads to its Hamiltonian

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

with

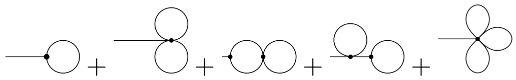
$$\begin{aligned} \lambda'_n &\equiv \sum_{m=0}^{\infty} \frac{\lambda_{n+m}}{m!}\, s^m_{\mathrm{cl}}, \qquad (90) \\ H'_0 &\equiv H[s_{\mathrm{cl}}] = H_0 + \frac{1}{2}\, s_{\mathrm{cl}}{}^{\dagger} D^{-1} s_{\mathrm{cl}} + \lambda'_0, \\ j' &\equiv j - \lambda'_1 - D^{-1}\, s_{\mathrm{cl}}, \ \ \text{and}\ D' \equiv (D^{-1} + \widehat{\lambda'_2})^{-1}\,. \end{aligned}$$

In case s_{cl} is exactly the classical solution, Eqs. 89 and 90 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[14]

[14] We propose the term *uncertainty-corrections* in order to describe

to any information theoretical estimator. For example, the uncertainty-corrections to the classical map estimator are given by

$$\delta m \;=\; m_d - s_{\rm cl} = \langle \phi \rangle_d \tag{91}$$

$=$  $+ \dots$

However, in case $s_{\rm cl}$ is not (exactly) the classical solution, may this 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, Eqs. 90-91 provide 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 to correct the inaccuracy of $s_{\rm cl}$ in a systematic way.

IV. COSMIC LARGE-SCALE STRUCTURE VIA GALAXY SURVEYS

A. Poissonian data model and Hamiltonian

Many datasets suffer from Poissonian-noise, which is non-Gaussian, and therefore well suited to test IFT in the non-linear regime. For example, the cosmological large-scale structure is traced by galaxies, which may be assumed to be roughly generated by a Poissonian process. On large-scales, the expectation value of the galaxy density follows that of the underlying (dark) matter distribution. The aim of cosmic cartography 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 large-scale-structure reconstruction techniques and understanding their uncertainties better is therefore an imminent and important goal. However, 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. A more general discussions of models of galaxy and structure fromation and references to relevant works was given in Sect. I F 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

$$\mu_i \approx \kappa\,(1 + b\,s(x_i)) \tag{92}$$

with $\kappa = \bar{n}_{\rm g}\,\Delta V$ 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. 47). 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. Such variations are due to the geometry of the observational survey sky coverage, due to a with distance from the observer decreasing galaxy selection function, and 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 from a complete sample e.g. by their luminosity due to instrumental sensitivity deterministically or stochastically, still possesses a Poissonian statistics, if the original distribution does. Although being spatially inhomogeneous, we assume κ and b to be known for the moment and to incorporate all above observational effects.

To cure the above mentioned shortcomings we replace Eq. 92 by a non-linear and non-translational invariant model:

$$\mu_i = \kappa(x_i)\, \exp(b(x_i)\, s(x_i)), \tag{93}$$

where κ and b may depend on position in a known way, and the unknown Gaussian field s 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. 93 is just a convenient choice for illustration and proof-of-concept purposes, and is easily exchangeable with more realistic, and even non-local data models. This data model was originally proposed by Coles and Jones [206] and investigated for constrained realisations by Sheth [120] and Vio et al. [207].

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

$$P(d|s) = \prod_i \frac{\mu_i^{d_i}}{d_i!} e^{-\mu_i} = \exp\left\{ \sum_i \left[d_i\, \log \mu_i - \mu_i - \log(d_i!) \right] \right\}, \tag{94}$$

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

$$H_d[s] \;=\; -\log P(d, s) = -\log P(d|s) - \log P(s)$$

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.

$$
\begin{aligned}
&= -d^\dagger b\, s + \kappa^\dagger \exp(b\, s) + H_0' + \frac{1}{2}\, s^\dagger S^{-1} s \\
&= \frac{1}{2}\, s^\dagger D^{-1} s - j^\dagger s + H_0 + \sum_{n=3}^{\infty} \frac{1}{n!}\, \lambda_n^\dagger\, s^n, \text{ with} \\
D^{-1} &= S^{-1} + \widehat{\kappa\, b^2}, \qquad (95)\\
j &= b\,(d-\kappa),\\
H_0 &= \frac{1}{2}\log(|2\pi\, S|) + (\kappa + \log(d!))^\dagger 1 - d^\dagger \log\kappa,\\
\text{and} & \\
\lambda_n &= \kappa\, b^n.
\end{aligned}
$$

A few remarks should be in order. Comparing the propagator to the one of our Gaussian theory one can read off an inverse noise term $M = R^\dagger N^{-1} R = \widehat{\kappa\, b^2}$, where $\widehat{c}$ denotes the diagonal matrix with c_x in the diagonal at location x. Thus the effective (inversely response weighted) noise decreases with increasing mean galaxy number and bias, and is infinite in regions without data ($\kappa = 0$).

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. Actually, the noise of our log-density signal is inversely proportional to the square root of the expected number density of galaxies, as it should be for a Poissonian process. 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. The bias, in contrast, is set by nature and can be regarded as a power counting parameter, which provides naturally a numerical hirarchy among the higher order vertices and diagrams for $b^2\langle s^2\rangle_{(s)} < 1$). 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. Although we are not going to investigate this complication in the following, it should be explained here how all the formulae in this section can easily be reinterpreted, in order to incorporate also the different classes of galaxies.

The galaxies can be characterised by a type-variable $L \in \Omega_{\text{type}}$, which may be the intrinsic luminosity, the morphological galaxy type, or a multi-dimensional combination of all properties which determine the galaxy types'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. 95, if we interpret any term containing d, κ and b to be summed or integrated over the type variable L. Thus we read

$$
\begin{aligned}
j(x) &= (b\,(d-\kappa))\,(x) \equiv \int dL\, b_L(x)\,(d_L(x) - \kappa_L(x)),\\
D^{-1}_{xy} &= \left(S^{-1} + \widehat{\kappa\, b^2}\right)_{xy} \equiv S^{-1}_{xy} + 1_{xy}\int dL\,\kappa_L(x) b_L^2(x),\\
\lambda_n(x) &= (\kappa\, b^n)\,(x) \equiv \int dL\,\kappa_L(x)\, b_L^n(x), \text{ and} \qquad (96)\\
\mu[s](x) &= \left(\kappa\, e^{b\,s}\right)(x) \equiv \int dL\,\kappa_L(x)\, e^{b_L(x)\,s(x)} = \int dL\,\mu_L[s](x),
\end{aligned}
$$

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

A few observations should be in order. In case of all galaxies having the same bias factor, Eq. 96 is simply a marginalisation of the type variable L, and any differentiation of the various galaxy types is not necessary. Since all known galaxies seem to have $b \sim \mathcal{O}(1)$, such a marginalisation seems to be justified, and explains why large-scale structure reconstructions, which applied this simplification, are relatively successful, although the different galaxy types masses, luminosities, and frequencies vary by orders of magnitude. However, as our numerical experiments below reveal, the data, and therefore the reconstructability of the density field, both exhibit a sensitive dependence on the bias for s-fluctuations with unity variance.[15] Such a variance is indeed observed on scales below ~ 10 Mpc in the galaxy distribution, and therefore the galaxy type-dependent bias variation does indeed matter to a non-negligible amount. 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 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 are possible by applying the recipes described here.

---

[15] This is found for our specific data model $\mu \propto \exp(b\,s)$, however, should also apply for more realistic models, which somehow have to keep $\mu \geq 0$ even for $b\,s < -1$

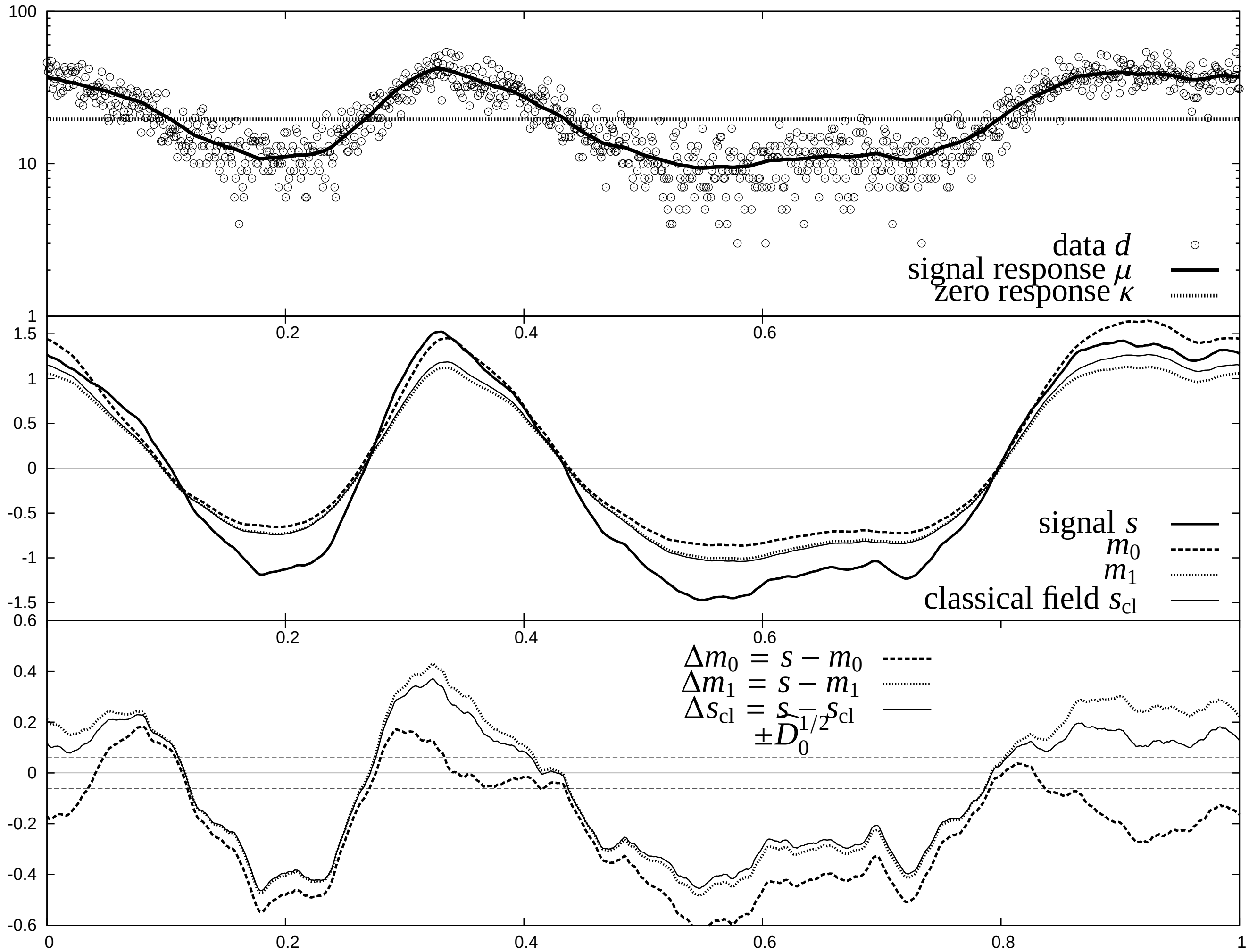


FIG. 1: Poissonian-reconstruction of a signal with unit variance and correlation length $q^{-1} = 0.05$, observed with slightly non-linear response ($b = 0.5$, resolution: 1025 pixels per unit length, zero-signal galaxy density: 2000 galaxies per unit length). **Top:** data d, signal response μ, and zero-response κ. **Middle:** signal s, linear Wiener-filter reconstruction $m_0 = D\,j$, next order reconstruction m_1 according to Eq. 97, and classical solution $s_{\rm cl}$ according to Eq. 99. **Bottom:** Deviations of m_0, m_1, $s_{\rm cl}$ from the signal, and the naive, Hessian-based error estimate $\widehat{D}^{1/2}$.

C. Non-linear map making

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

$$\begin{aligned} m_1 &= \text{———}\bullet + \text{—}\bullet\!\bigcirc + \text{—}\bullet\!\!< + \bigcirc\!\!\text{—}\bullet + \mathcal{O}(b^6) \\ &= D_{xy}\,j_y - \frac{1}{2}D_{xy}\,b_y^3\,\kappa_y\,D_{yy} - \frac{1}{2}D_{xy}\,b_y^3\,\kappa_y\,(D_{yz}\,j_z)^2 \\ &\quad -\frac{1}{2}D_{xy}\,b_y^4\,\kappa_y\,D_{yy}\,D_{yz}\,j_z + \mathcal{O}(b^6) \end{aligned} \tag{97}$$

or in compact notation

$$m_1 = D\left[j - \frac{1}{2}\,\widehat{b^3\,\kappa}\left(\widehat{D} + (D\,j)^2 + \widehat{D}\,b\,D\,j\right)\right] + \mathcal{O}(b^6). \tag{98}$$

It is apparent, that the non-linear map making formula contains corrections to the linear map $m_0 = D\,j$. The first two correction terms are always negative, reflecting the fact that our non-linear data model has non-symmetric fluctuations in the data with respect to the mean. Moderate positive fluctuations lead to large numbers of observed galaxies, due to the exponential dependence. Thus a large number of observed galaxies at a position leads to an overestimate of s in the linear map, which is corrected downwards by the first nonlinear terms. For negative field strength the response gets sublinear, so that a small number of observed galaxies (with respect to κ) leads to an underestimate of the magnitude of the (negative) fluctuation in the linear map. Thus also here a negative correction is required. The last correction term is oppositely directed to the linear map, thereby producing some non-linear damping of the effect of the information source and correcting for the curvature in the non-linear response of the data to the signal.

A one-dimensional, numerical example is displayed in Fig. 1. There, the signal was realised to have a power spectrum $P_s(k) \propto (k^2+q^2)^{-1}$, with a correlation length $q^{-1} = 0.05$. The normalisation was chosen such that the auto-correlation function is $\langle s(x)\, s(x+r)\rangle_{(s)} = \exp(-|q\, r|)$ and therefore the signal dispersion is unity, $\langle s^2\rangle_{(s)} = 1$. The data are generated by a Poissonian process from $\kappa_s = \kappa\, \exp(b\, s)$ 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 towards zero. Also shown in the bottom panel of the figure are the residuals in comparison to the square root of the diagonal part of the propagator $\widehat{D}$, as the Gaussian approximation of the uncertainty level. It is apparent, that the residuals have a non-Gaussian statistics, otherwise they should remain in 68% of the cases within the indicated uncertainty region, which they don't.

The non-linear 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 into 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 $m_0 = D\, j$ (not shown). In such a case interaction re-summation techniques should be used to incorporate as many higher order interaction terms as possible. One very powerful re-summation is provided by the classical solution, as developed below, which contains all tree-diagrams simultaneously. This solution, also show 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. 89, which reads in this case

$$\begin{aligned} s_{\mathrm{cl}} &= D\left(j - \sum_{m=2}^{\infty} \frac{b^{m+1}}{m!}\, \kappa\, s_{\mathrm{cl}}^{m}\right) \\ &= D\, b\left(d - \kappa\, \left(e^{b\, s_{\mathrm{cl}}} - b\, s_{\mathrm{cl}}\right)\right) \qquad (99) \\ &= S\, b\, (d - \underbrace{\kappa\, e^{b\, s_{\mathrm{cl}}}}_{\kappa_{s_{\mathrm{cl}}}}). \end{aligned}$$

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

$$\kappa_s = \kappa\, e^{b\, s}. \qquad (100)$$

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

$$s_{\mathrm{cl}} = \frac{1}{b}\, \log\left[\frac{j - S^{-1} s_{\mathrm{cl}}}{\kappa\, b}\right] = \frac{1}{b}\, \log\left[\frac{d}{\kappa} - 1 - \frac{S^{-1} s_{\mathrm{cl}}}{\kappa\, b}\right], \qquad (101)$$

which may be solved iteratively, while ensuring that $s_{\mathrm{cl}}^{(i)} < S\, j$ at all iterations i. This form of the classical field equation has some similarities to the naive inversion of the response formula, $\langle d\rangle_{(d|s)} = \kappa\, \exp(b\, s)$, which yields

$$s_{\mathrm{naive}} = \frac{1}{b}\, \log\left[\frac{d}{\kappa}\right], \qquad (102)$$

a formula one can only dare to use in regimes of large d. The classical solution is more conservative than this naive data inversion, in that there is a damping term, $S^{-1} s_{\mathrm{cl}}/(\kappa\, b)$, compensating a bit the influence of too large data points.

Those equations permit to calculate the classical solution if suitable numerical regularisation schemes are applied, since naive iterations can lead easily to numerical divergencies in the non-linear case. For example, a pseudo-time τ can be introduced by setting $j(\tau) = \tau\, j$. This allows to set up a differential equation for $s_{\mathrm{cl}}(\tau)$ by taking the time derivative of Eq. 99,

$$\dot{s}_{\mathrm{cl}} = D_{s_{\mathrm{cl}}}\, j \;\text{ with }\; D_{s_{\mathrm{cl}}} = \left(S^{-1} + \kappa_{s_{\mathrm{cl}}}\, b^2\right)^{-1}, \qquad (103)$$

which has to be solved for $s_{\mathrm{cl}}(1)$ starting from $s_{\mathrm{cl}}(0) = 0$. 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 to the accumulated knowledge on the signal and thereby implied Poissonian-noise structure. Thus, it is a nice example system for continous Bayesian learning and also illustrates how different datasets can successively be fused into a single knowledge basis.

Alternatively, we can recognize that j is proportional to the bias b, which is the essential signal response parameter controlling non-linearities, and therefore can serve us as an expansion parameter. Thus, we set $b \to \tau\, b$ in Eq. 99 and take the τ-derivative, yielding

$$\dot{s}_{\mathrm{cl}} = D_{(\tau\, s_{\mathrm{cl}})}\, b\, \left(d - \kappa\, e^{\tau\, b\, s_{\mathrm{cl}}}\, (1 + \tau\, b\, s_{\mathrm{cl}})\right), \qquad (104)$$

which is numerically different from Eq. 103 due to the with time growing effective bias factor $\tau\, b$ occurring in $\kappa_{\tau\, s_{\mathrm{cl}}}$ and $D_{\tau\, s_{\mathrm{cl}}(b)}$. The initially lower effective response has to be compensated later by an extra term not present in Eq. 103, which is proportional to $\tau\, b$, and therefore becoming important mostly at the later stages of the evolution.

Numerical implementations of Eqs. 103 and 104 by discretisation correspond to different approximation schemes, and therefore yield exactly identical results only in the continuous limit. Nevertheless, we can call all such implementations map-making algorithms, since we required a data transformation only to approximately reproduce the signal map given the data for this. It should become clear that higher-fidelity map-making algorithms are possible by not only investigating the maximum of the posterior, but by averaging the signal s over the full support of $P(s|d)$.

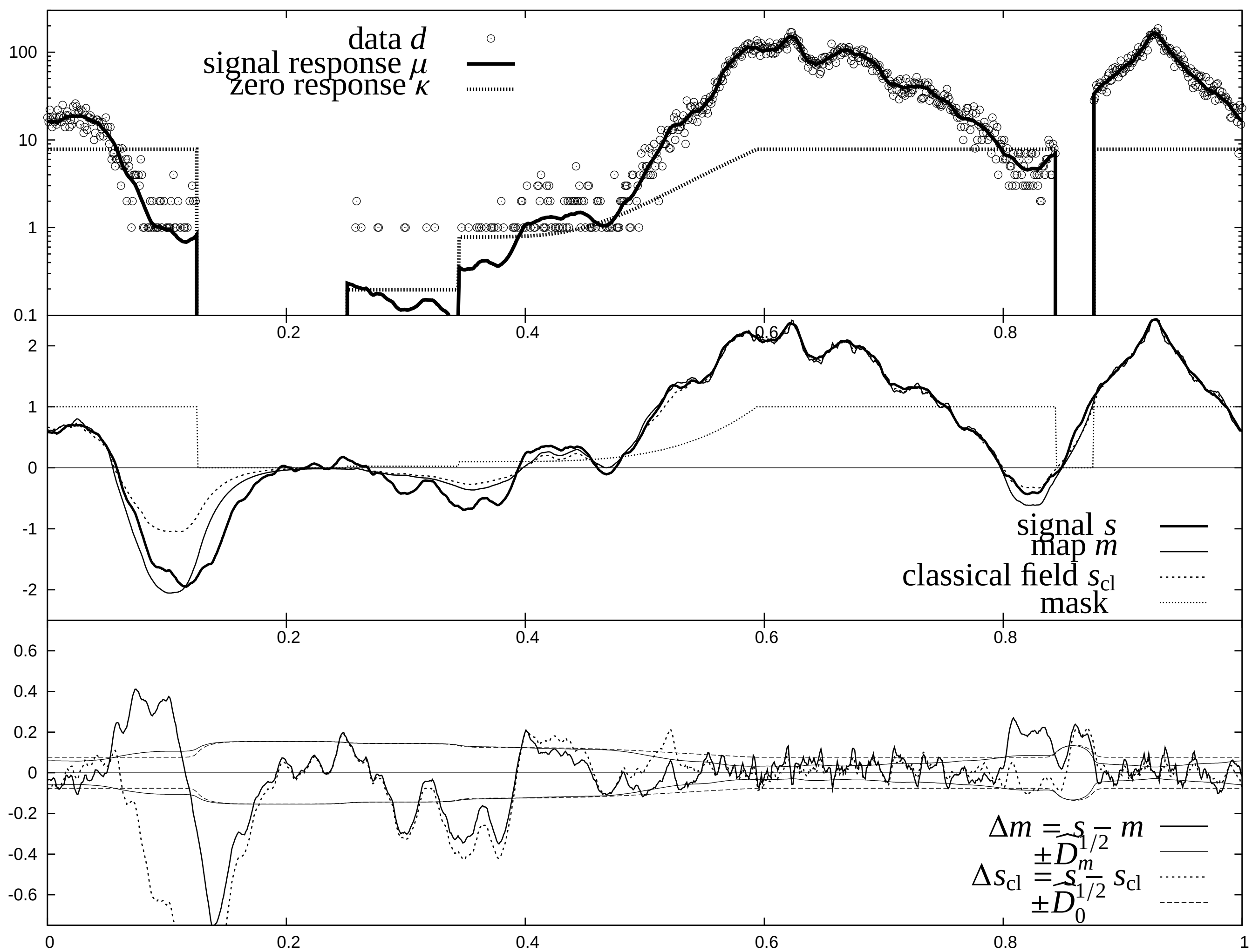


FIG. 2: Poissonian-reconstruction of a signal with unit variance and correlation length $q^{-1} = 0.03$, observed with strongly non-linear response ($b = 1.25$, resolution: 1025 pixels per unit length, zero-signal galaxy density: 8000 galaxies per unit length) through a complicated mask. **Top:** data d, signal response μ, and zero-response κ. **Middle:** signal s, renormalisation-based reconstruction m, classical solution $s_{\rm cl}$, 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 the displayed area. **Bottom:** Deviations of m and $s_{\rm cl}$ from the signal, and the naive, Hessian-based error estimates $\widehat{D}_0^{1/2}$ and $\widehat{D}_m^{1/2}$. Note, that in the regions with many observed galaxies, the high signal to noise ratio can be seen in the narrowness of $\widehat{D}_m^{1/2}$, which is significantly smaller than $\widehat{D}_0^{1/2}$ at these locations. Also the residual statistics seems to follow well a Gaussian statistics in these regions.

Anyhow, we can assume that a good approximation $t \approx s_{\rm cl}$ to the classical solution can be achieved. Figs. 1 and 2 display classical solutions for slightly and strongly non-linear Poissonian inference problems. Especially the second example shows that the classical solution is sometimes missing some significant contributions (see region around $x = 0.1$), due to the 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 $\phi = s - t$,

$$\begin{aligned}
H_t[\phi] &= \frac{1}{2}\,\phi^\dagger D_t^{-1}\phi - j_t^\dagger\phi + \kappa_t^\dagger g(b\,\phi) + H_{0,t}, \quad \text{where}\\
D_t^{-1} &= S^{-1} + \widehat{\kappa_t\,b^2},\\
j_t &= b\,(d-\kappa_t) - S^{-1}t,\\
g(x) &= e^x - 1 - x - \frac{1}{2}\,x^2 = \sum_{m=3}^{\infty}\frac{x^m}{m!},
\end{aligned} \tag{105}$$

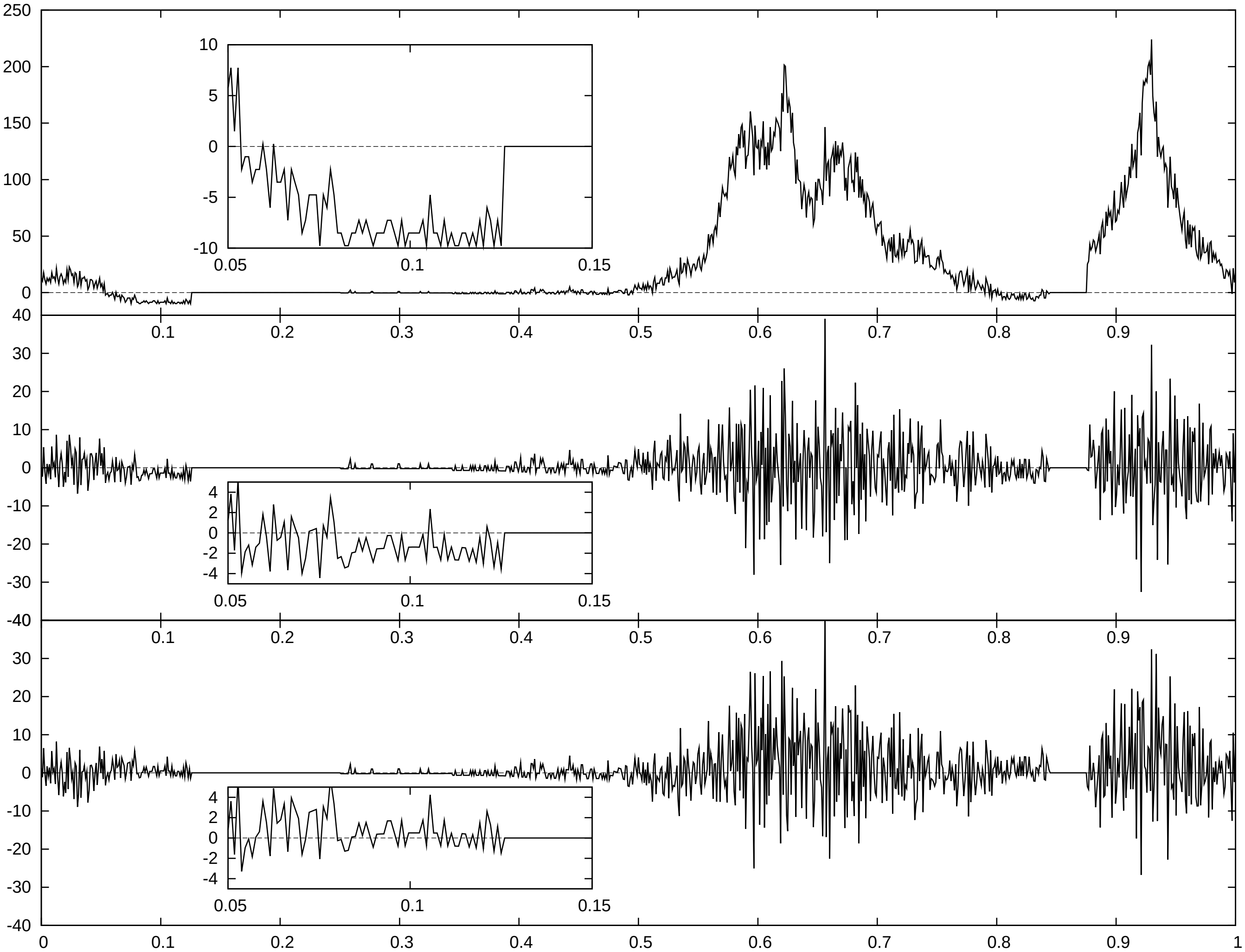


FIG. 3: Remaining information source before the reconstruction (top panel) after the classical reconstruction (middle panel) and after the exact field theoretical mapping (bottom panel). Displayed is the remaining information source $j'_t = b\,(d - \kappa_t)$ for the data of Fig. 2 and for $t = 0$, $t = s_{\mathrm{cl}}$, and $t = m$ in the top, middle and bottom panels, respectively. The inlets show an enlarged view on the region around $x = 0.1$, where the classical reconstruction was significantly worse than the field theoretical one, as can be seen from Fig. 2. This is due to the asymmetry of the Poissonian distribution with respect to its mean value, for which the classical map estimator is sub-optimal. This asymmetry is clearly visible in the region around $x = 0.1$ by the relatively rare, but stronger upwards spikes compared to downwards fluctuations. All the small scale fluctuations in the information source terms get erased by the smoothing operation of the information propagators, as shown in Fig. 4.

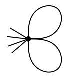
and $H_{0,t}$ is a momentarily irrelevant normalisation constant. Again, we have permitted for a non-zero j_t, since t might not be exactly the classical solution.

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 $\kappa_0 \to \kappa_t$ would provide us with the shifted field Hamiltonian, as defined in Eq. 70, expect for the term $-S^{-1}t$ in j_t. It turns out, that this term is some sort of counter-term, which accumulates the effect of the non-linear interactions.[16]

Anyhow, 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

[16] Dropping this term, and repeating the operation which have led to the classical solution iteratively, permits to reconstruct non-linearities better than classical. However, the price for this ad-hoc improvement is a data-overfitting instability, as numerical experiments show.

vertex of the Poissonian Hamiltonian with m legs, there exists diagrams in any Feynman-expansion, in which a number of n simple loops are added to this vertex. Such an n-loop enhanced vertex is given by



$$= \frac{-1}{2^n\, n!}\, \lambda_t^{(m+2n)}\, \widehat{D}^n = \frac{-1}{2^n\, n!}\, \kappa_t\, b^{m+2n}\, \widehat{D}^n. \quad (106)$$

All these diagrams can be re-summed into an effective interaction vertex, via

$$\begin{aligned} \lambda_t^{(m)} &\to \lambda'^{(m)}_t = \kappa_t\, b^m \sum_{n=0}^{\infty} \frac{1}{2^n\, n!}\, b^{2n}\, \widehat{D}^n \\ &= \kappa_t\, \exp\left(\frac{b^2}{2}\,\widehat{D}\right) b^m \\ &= \kappa_{t+\frac{b^2}{2}\widehat{D}}\; b^m = \lambda^{(m)}_{t+\frac{b^2}{2}\widehat{D}}. \end{aligned} \quad (107)$$

Thus, this re-summation is effectively equivalent to a shift of the reference field

$$t \to t' = t + b^2\, \widehat{D}/2. \quad (108)$$

One might consider to use this uncertainty-loop shifted field as the reference field t' in the Hamiltonian of Eq. 105, and to calculate with the corresponding propagator $D_{t'}$, information source term $j_{t'}$ and interaction coefficients $\lambda_{t'}^{(n)}$ the first few Feynman diagrams contributing to the map of $\phi' = s - t'$, like those diagrams in Eq. 97. Although the interaction coefficients are by no means smaller than the original ones, the ϕ'-field amplitudes should be much smaller than those of the original s-field, since most of the information is already condensed into t', and therefore one might hope that the higher-order diagrams should have smaller and smaller contributions.

The with this recipe derived total signal map, $m = t' + \langle\phi'\rangle$, contains many, however not all, contributing Feynman diagrams. The uncertainty corrections applied at the end indeed seem to shift the map closer to the signal, as numerical experiments for situations as displayed in Fig. 2 reveal, however, only by a small margin. Therefore, an inclusion of uncertainty corrections right from the beginning of the calculations is required. This can be done via renormlisation techniques, and the calculations above have paved the path for those.

F. Response renormalisation

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. Since we have basically only two handels on any interaction term $\lambda_n = \kappa\, b^n$, the bias b and the zero-response κ, we concentrate on the bias, since it has served us well in the classical case, and has the potential to suppress high order terms for small values of b. The bias is the main parameter controlling the signal response, and therefore we call our attempt *response renormalisation*, since the procedure developed below should be generic enough to serve as a blueprint for other data models with a non-linear response of the form $\mu = \kappa\, f(b\, s)$, with $f(x)$ being a strictly positive function of its argument. Again we replace $b \to \tau\, b$ in the theory, and use τ as our response renormalisation parameter.

For the response renormalisation we decompose the singal into a known, fixed background field t and an uncertainty field ϕ, both dependent on the renormaliation parameter τ, so that $s_\tau = t_\tau + \phi_\tau$. We start with $t_0 = 0$ at $\tau = 0$ and find that the uncertainties at that bias level are distributed according to the trivial prior-Hamiltonian

$$H_{t_0}^{(\tau=0)}[\phi_0] = \frac{1}{2}\, \phi_0^\dagger S^{-1} \phi_0, \quad (109)$$

where the subscript t_0 indicates the value of the background field around which the Hamiltonian is expanded. This Hamiltonian has a vanishing uncertainty field expectation value, $\langle\phi_0\rangle_{(s|d,\tau=0)} = 0$, indicating that the initial background field $t_0 = 0$ was well chosen.

We can therefore assume, that for some value of the bias we know a background solution t_τ, for which the corresponding uncertainty field $\phi_\tau = s_\tau - t_\tau$ has zero expectation value under its induced Hamiltonian $H_{t_\tau}^{(\tau)}[\phi_\tau]$, since this is at least valid for $\tau = 0$. Then the effective Hamiltonian for this field can be written as

$$H_{t_\tau}^{(\tau)}[\phi_\tau] = \frac{1}{2}\, \phi_\tau^\dagger D_\tau^{-1} \phi_\tau + V_{\rm int}^{(\tau)}[\phi_\tau], \quad (110)$$

where the effective propagator D_τ collecting all effective second order terms in ϕ, and the effective non-linear interaction potential $V_{\rm int}^{(\tau)}[\phi_\tau]$ have yet to be determined. The latter collects all effective terms which are non-quadratic in ϕ. However, their individual impacts on the uncertainty field expectation value must cancel each other, otherwise the expectation value would not be zero. For $\tau = 0$ we have $D_0 = S$ and $V_{\rm int}^{(0)}[\phi_0] = 0$.

If the renormalisation parameter is now increased by a small amount, $\tau \to \tau + \varepsilon$, we can use Eq. 105 to obtain a modified Hamiltonian for the uncertainty ϕ_τ. To first order in ε we find

$$\begin{aligned} &H_{t_\tau}^{(\tau+\varepsilon)}[\phi_\tau] = H_{t_\tau}^{(\tau)}[\phi_\tau] + \sum_{n=1}^{\infty} \frac{1}{n!}\, \lambda_n^{(\tau+\varepsilon)\,\dagger} \phi_\tau^n + \mathcal{O}(\varepsilon^2), \\ &\text{with } \lambda_1^{(\tau+\varepsilon)} = -\varepsilon\, b\,(d - \kappa_{t_\tau}) \\ &\text{and } \lambda_n^{(\tau+\varepsilon)} = \varepsilon\, \kappa_{t_\tau}\, n\, \tau^{n-1}\, b^n \text{ for } n > 1. \end{aligned} \quad (111)$$

The structure of this shifted Hamiltonian is that it decomposes into an old part $H_{t_\tau}^{(\tau)}[\phi_\tau]$ with zero field-expectation value, and a small ε-perturbation. Thus for very small perturbations, as provided by all the ε-terms, the old Hamiltonian can be regarded as being free of internal interactions and information sources.

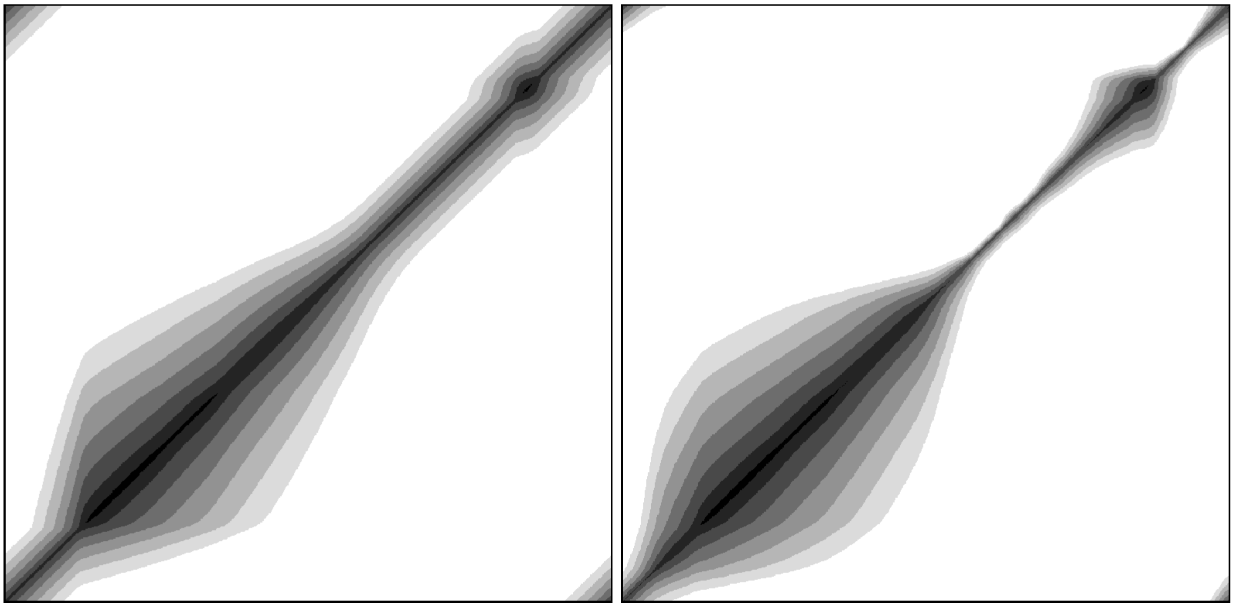
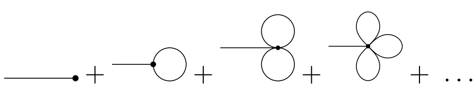


FIG. 4: IFT propagators $D_0 = (S^{-1} + \widehat{\kappa_0\, b^2})^{-1}$ (left) and $D_m = (S^{-1} + \widehat{\kappa_m\, b^2})^{-1}$ (right) in logarithmic grey scaling for the data displayed in Fig. 2. m is here the solution of the response renormalisation flow equation (Eq. 117). The values of the diagonals show the local uncertainty variance (in Gaussian approximation) before $(\widehat{D_0})$ and after $(\widehat{D_m})$ the data is analysed, respectively. The bottom left and top right corners exhibit non-vanishing propagator values due to the assumed periodic spatial coordinate, which puts these corners close to the two others on the matrix diagonal.

To first order in ε therefore only leaf diagrams with a single perturbative interaction vertex contribute to the perturbed expectation value of ϕ_τ:

$$\langle \phi_\tau \rangle_{(s|d,\tau+\varepsilon)} =$$



$$= \varepsilon\, D_\tau\, b \left[d - \kappa_{(\tau\, t_\tau)} \sum_{n=0}^{\infty} \frac{2\,n+1}{n!} \left(\frac{\tau^2\, b^2\, \widehat{D_\tau}}{2} \right)^n \right],$$

where the propagator to be used is that of the unperturbed Hamiltonian. Note, that only odd interaction terms shift the expectation value. 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 induced field expectation value can be absorbed into $t_{\tau+\varepsilon} \equiv t_\tau + \langle \phi_\tau \rangle_{(s|d,\tau+\varepsilon)}$, so that the then remaining uncertainty field $\phi_{\tau+\varepsilon} = s_{\tau+\varepsilon} - t_{\tau+\varepsilon}$ has zero expectation value. Its effective Hamiltonian is therefore of the form

$$H^{(\tau+\varepsilon)}_{t_\tau+\varepsilon}[\phi_{\tau+\varepsilon}] = \frac{1}{2}\, \phi^\dagger_{\tau+\varepsilon} D^{-1}_{\tau+\varepsilon} \phi_{\tau+\varepsilon} + V^{(\tau+\varepsilon)}_{\mathrm{int}}[\phi_{\tau+\varepsilon}], \quad (112)$$

and can again be regarded to be effectively free for small perturbations. The crucial step is to realise that

$$D_{\tau+\varepsilon} = D_\tau + \mathcal{O}(\varepsilon^2), \quad (113)$$

since all effects linear in ε were absorbed in the shift of the background field. Thus we conclude by iterating $N = \tau/\varepsilon$ times that

$$D_\tau = S + \mathcal{O}(\varepsilon\, \tau), \quad (114)$$

which has the well defined continuous limit $D_\tau = S$ for $\varepsilon \to 0$ since $0 \le \tau \le 1$.

Using the identity

$$\sum_{n=0}^{\infty} \frac{2\,n+1}{n!}\, x^n = e^x\,(1+2x), \quad (115)$$

we can write the background increment $dt_{\tau+\varepsilon} = t_{\tau+\varepsilon} - t_\tau$ as

$$dt_{\tau+\varepsilon} = \varepsilon\, D_\tau\, b \left[d - \kappa_{\tau\, t_\tau + \tau^2\, b\, \widehat{D_\tau}/2}\, (1 + \tau^2\, b^2\, \widehat{D_\tau}) \right] \quad (116)$$

and convert this into a differential equation for the background field by taking the limit $\varepsilon \to 0$:

$$\dot{t}_\tau = S\, b \left[d - \kappa\, e^{\tau\, b\, t_\tau + \tau^2\, b^2\, \widehat{S}/2}\, (1 + \tau^2\, b^2\, \widehat{S}) \right]. \quad (117)$$

This is the required response renormalisation flow equation, which has to be solved starting with $t_0 = 0$ at $\tau = 0$ until the required response bias is reached at $\tau = 1$, so that $m = t_1$.

It is instructive to compare it to the classical response flow equation, Eq. 104, which is reproduced here for an easier comparison:

$$\dot{s}_{\mathrm{cl}}(\tau) = D_{s_{\mathrm{cl}}(\tau)}\, b \left[d - \kappa\, e^{\tau\, b\, s_{\mathrm{cl}}(\tau)}\, (1 + \tau\, b\, s_{\mathrm{cl}}(\tau)) \right],$$

Although there is a structural similarity of the two equations, there are also significant differences.

The first and most prominent difference is that the propagator to be applied to some sort of effective information source flow (b times the terms in the square brackets) is the simple signal covariance matrix in case of the renormalisaton flow equation. This is a substantial computational simplification, since translation invariance of Cosmology enforces a diagonal covariance matrix in Fourier space. The classical flow equation requires the application of a much more numerically demanding propagator, which is the inverse of the sum of two terms, of which one is diagonal in Fourier space, S^{-1}, and the other in position space, $\widehat{\tau^2 \kappa_{s_{\rm cl}}}$. Thus, our effort to derive the exact information field expectation value is rewarded by a numerically much simpler and faster recipe.

The second difference between classical and field theoretical renormalisation flow equations is that the effective background field, for which the response is compared to the data in the renormalisation case, $t_\tau + \tau^2 \widehat{S}/2$, is actually shifted by loop corrections of the sort discussed in the previous subsection. These uncertainties (*or knowledge fluctuations*) probe the phase-space volume around the background field and feel the structure of the IFT-*potential* there. Since the potential is exponential ($\sim \exp(b\,\phi)$)and therewith asymmetric, positive uncertainty fluctuations have a larger impact than negative ones, leading to a shift in the effective potential.

And finally, the corrections provided by the last terms in both flow equations, which compensate for the initially low value of the effective bias $\tau\, b$ during the flow, are solely due to uncertainty loops in the renormalisation case, whereas they depend on the solution itself in the classical case.

However, the most important difference is probably the improved singal-recovery fidelity of the full IFT map making algorithm in comparison to the classical map, as Fig. 2 shows for many positions, and we also witnessed in additional numerical experiments, not shown here. The origin of this difference can be understood by inspection of Fig. 3, which shows the remaining information source after a background solution t has been adopted,

$$j'_t = b\,(d - \kappa_t), \tag{118}$$

for $t = 0$, $t = s_{\rm cl}$, and $t = m$. The remaining information source j'_t differs from the shifted information source j_t as provided by Eq. 105, by the omission of the counter term which compensates higher order interactions, $-S^{-1}t$. It is apparent in Fig. 3, that both, the classical and the field theoretical maps have removed most of the information present in j_0, however, $j_{s_{\rm cl}}$ still exhibits some exploitable net information around $x = 0.1$, whereas j_m seems to be free of any larger scale trend in this area. The better performance of the field theoretical mapping algorithm lies in the fact, that it attempts the correct expectation value estimation by integrating over the full posterior distribution $P(s|d)$, whereas the classical solution only searches the maximum of $P(s|d)$. And the latter is known to be a biased mean estimator for any asymmetrical probability density distribution like the Poissonian one we are dealing with, and especially in the limit of small number counts.

Since there are also instances, where the classical solution is closer to the underlying signal than the field theoretical solution, a more detailed statistical comparison between the two methods is required. This is provided in Fig. 5, where the mean variance of the residuals of the classical and correct solutions are shown, the $\widehat{\sigma^2_T}$ of Eq. 17, averaged over 1000 signal realisations of the inference problems of Fig. 2. Both methods exhibit a comparable ability to reconstruct missing data, and of course are unable to extrapolate into too large data-gaps. However, with increasing data availablility, the accuracy of the field theoretical method becomes clearly superior to the classical one, and this for lower computational costs.

It would be interesting to know something analytically about the remaining uncertainties at the end of the renormalisation flow. The theory, as developed here, was only suited to follow the mean evolution of the background field t_τ, which let to the map $m = t_1$. One could also ask for the evolution of the two point correlation function, however this is beyond the introductory scope of this work. However, some insight on the final uncertainty-covariance structure can be read off from the Hessian matrix of the bare (non-renormalised) Hamiltonian of the uncertainty field $\phi = s - m$. This is simply D_m, which we display in Fig. 4 in comparison to the zero information Hessian uncertainty-covariance matrix D_0. The height and width of the propagator values defines the strength of the response to, and the distance of information propagation from an information sources as can be seen in this fugure. 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 the latter can be expected there, due to the signal-response-to-noise weighting operation transforming the data into the source j. Also the individual informations do not need to be propagated that far, thanks to the rich information source density in such regions. It is interesting to see, how the propagator is structured at the intersections of transparent and opaque regions of the mask, such that information is propagated efficiently from the first to the second region, but any reverse information flow is largely suppressed. This can be read off the asymmetric vertical or horizontal profiles of the propagator at such locations. Finally, the structure of D_m has imprinted the expected information source density structure given the reconstruction m. The strongly non-linear signal response has lead to regions with very high galaxy count rates, which have larger information densities, and therefore lower and narrower information propagators.

Obvious extensions to the presented Poissonian reconstruction problems could involve treating more general or even non-local responses of the galaxy distribution to the dark-matter field, or incorporating knowledge beyond the Poissonian approximation of the galaxy formation statistics. However, these are left for future work, since our

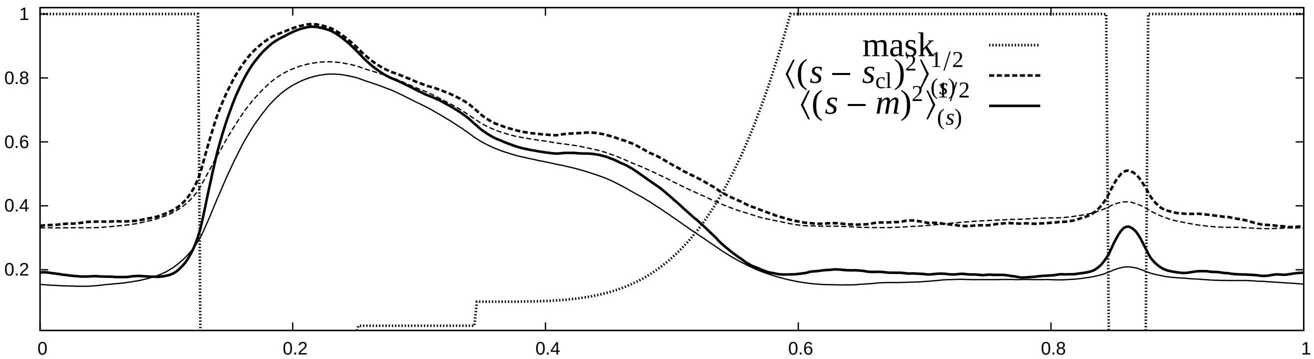


FIG. 5: Residual root-mean-square variance of the field theoretical ($\langle(s-m)^2\rangle^{1/2}_{(s)}$, solid) and the classical solution ($\langle(s-s_{\rm cl})^2\rangle^{1/2}_{(s)}$, dashed) from 1000 signal realisations. The data mask is also shown for comparison. The setup is very similar to the one of Fig. 2, just the resolution was set to 513 pixels per unit length, and the zero-signal galaxy density to 2000 galaxies per unit length, in order to speed up the computations. The thin lines show a similar average for a doubled correlation length $q^{-1} = 0.06$.

aim here was only to demonstrate the applicability of the theory to non-trivial problems and to outline the necessary calculations.

V. NON-GAUSSIAN CMB FLUCTUATIONS VIA $f_{\rm nl}$-THEORY

A. Data model

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

On top of the very uniform CMB sky with a mean temperature $T_{\rm CMB}$, small temperature fluctuations on the level of $\delta T^{\rm \{I,E,B\}}_{\rm obs}/T_{\rm CMB} \sim 10^{-\{5,6,7\}}$ are observed or expected in total Intensity (Stokes I) and in polarisation E- and B-modes, respectively. These CMB temperature fluctuations are believed and observed to follow mostly a Gaussian distribution. However, inflation predicts some level of non-Gaussianity, and also some of the secondary anisotropies imprinted by the large-scale structure of the Universe via CMB lensing, the Integrated Sachs-Wolfe and the Rees-Sciama effects should have imprinted non-Gaussian signatures due to the non-linear matter structures which have evolved in the modern Universe [208, 209]. 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

$$d \equiv \delta T^{\rm \{I,E,B\}}_{\rm obs}/T_{\rm CMB} = R\,\phi_{\rm primordial} + n, \tag{119}$$

where $\phi_{\rm primordial}$ is the 3-dimensional, primordial gravitational potential, and R is the response on it of an CMB-instrument, observing the induced CMB temperature fluctuations in intensity and polarisation. In case that the data of the instrument are forground-cleaned and deconvolved all-sky maps (assuming the data processing to be part of the instrument) the response, which translates the 3-d gravitational field into temperature maps, is well known from CMB-theory and can be calculated with publically available codes like `cmbfast, camb,` and `cmbeasy` (see Sect. I F 5), which use them internally. 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 subsummes 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 are 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 propability function, many of the inflationary non-Gaussianities seem to be well described by a local process, which taints an initially Gaussian random field, $\phi \hookleftarrow P(\phi) = G(\phi, \Phi)$ (with the ϕ-covariance $\Phi = \langle \phi\, \phi^\dagger \rangle_{(\phi)}$), with some level of non-Gaussianity. A well controllable realisation of such a tarnishing operation is provided by a slightly non-linear transformation of ϕ into the primordial gravitational potential via $\phi_{\rm primordial} = \phi(x) + f_{\rm nl}\,(\phi^2(x) - \langle \phi^2(x) \rangle_{(\phi)})$ for any $x \in R^3$. 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

$$d = R\,(\phi + f\,(\phi^2 - \widehat{\Phi})) + n, \tag{120}$$

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

B. CMB-Hamiltonian

The CMB-Hamiltonian is therefore given by

$$\begin{aligned}
H_f[d,\phi] &= -\log(G(\phi,\Phi)\, G(d-\phi+f\,(\phi^2-\widehat{\Phi}),N)) \\
&= \frac{1}{2}\phi^\dagger D^{-1}\phi + H_0 - j^\dagger\phi + \sum_{n=0}^{4}\frac{1}{n!}\,\Lambda^{(n)}[\phi,\ldots,\phi],
\end{aligned}$$

with

$$\begin{aligned}
D^{-1} &= \Phi^{-1} + R^\dagger N^{-1} R \equiv \Phi^{-1} + M, \\
j &= R^\dagger N^{-1} d, \\
\Lambda^{(0)} &= j^\dagger (f\,\widehat{\Phi}) + \frac{1}{2}\,(f\,\widehat{\Phi})^\dagger M\,(f\,\widehat{\Phi}), \\
\Lambda^{(1)} &= -R^\dagger N^{-1}(f\,\widehat{\Phi}) \text{ and } j' = j - \Lambda^{(1)}, \\
\Lambda^{(2)} &= -2\,\widehat{f j'}, \\
\Lambda^{(3)}_{xyz} &= (\delta_{xy}\, M_{yz}\, f_z + \text{permutations}), \\
\Lambda^{(4)}_{xyzu} &= \frac{1}{2}\,(f_x\,\delta_{xy}\, M_{yz}\,\delta_{zu}\, f_u + \text{permutations}),
\end{aligned} \tag{121}$$

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

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 also approximately valid on large angular scales, where the Sachs-Wolfe effect dominates, we have $N_{xy} = \sigma_n^2(x)\,\delta(x-y)$, $R = -3$ [209] 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 to simplify the Hamiltonian to

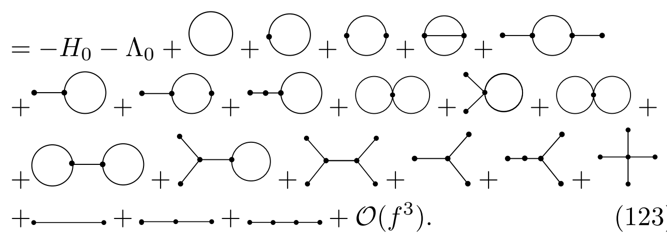
$$\begin{aligned}
H_f[d,\phi] &= \frac{1}{2}\phi^\dagger D^{-1}\phi + H_0 - j^\dagger\phi + \sum_{n=0}^{4}\frac{1}{n!}\,\lambda_n^\dagger\,\phi^n, \text{ with} \\
D^{-1} &= \Phi^{-1} + 9\,\widehat{\sigma_n^2},\ j' = j - \lambda_1 = -3\,(d + \widehat{\Phi}\,f)/\sigma_n^2, \\
\lambda_0 &= 3\,(\widehat{\Phi}/\sigma_n^2)^\dagger(\frac{3}{2}f^2\widehat{\Phi} - f\,d),\ \lambda_2 = -2\,f\,j', \\
\lambda_3 &= 54\,f/\sigma_n^2, \text{ and } \lambda_4 = 108\,f^2/\sigma_n^2.
\end{aligned} \tag{122}$$

The numerical coefficients of the last two terms may look large, and therefore question the applicability of perturbative methods, 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.

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

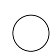
Since we are momentarily not interested in reconstructing the primordial fluctutations, but to extract knowledge on $f_{\rm nl}$, we marginalise the former by calculating the log-evidence $\log(P(d|f))$ up to quadratic order in f:

$$\begin{aligned}
\log Z_f[d] &= \log \int \mathcal{D}\phi\, P(\phi, f|d) \\
&= \log \int \mathcal{D}\phi\, e^{-H_f[d,\phi]}
\end{aligned}$$



(123)

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 4$^{\rm th}$ order interactions are involved, an implementation in spherical harmonics space may be feasible, using the only 4$^{\rm th}$ order C-coefficients (Eq. 84), which can be calculated with computer algebraic programs. Finally, we have defined



$$= \frac{1}{2}\,\log|2\pi\,D\,\beta^{-1}| = \frac{1}{2}\,\mathrm{Tr}(\log(2\pi\,D\,\beta^{-1})). \tag{124}$$

Although f is not known, the expressions in Eq. 123 proportional to f and f^2 can be calculated separately, permitting to write down the Hamiltonian of f if a suitable prior $P(f)$ is chosen,

$$H_d[f] \equiv -\log(P(d|f)\,P(f)) = \tilde{H}_0 + \frac{1}{2}\,f^\dagger\tilde{D}^{-1}f + \tilde{j}^\dagger f + \mathcal{O}(f^3), \tag{125}$$

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

$$m_f = \langle f\rangle_{(s,f|d)} = \tilde{D}\,\tilde{j}, \tag{126}$$

and its uncertainty variance is just

$$\langle (f - m_f)\,(f - m_f)^\dagger \rangle_{(s,f|d)} = \tilde{D}. \tag{127}$$

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

$$f(x) = \sum_{l=0}^{l_{\max}} \sum_{m=-l}^{l} f_{lm}\, Y_{lm}(x) \tag{128}$$

up to some finite $l_{\max}$. Then one would recalculate the partition sum, now separately for terms proporional 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

$$\begin{aligned} \tilde{j}_{(lm)} &= \left.\frac{dH_d[f]}{d_{lm} f}\right|_{f=0}, \text{ and} \\ \tilde{D}^{-1}_{(lm)\,(l'm')} &= \left.\frac{d^2 H_d[f]}{df_{lm}\, df_{l'm'}}\right|_{f=0}. \end{aligned} \tag{129}$$

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.

We conclude this chapter with a short comparison to existing f_{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. 199, and references in Sect. I F 5]. The estimator presented here is fourth order in the data, and therefore of higher accuracy since both estimators are supposed to be optimal within their order. It is not clear at the moment, if the presented new estimator is significantly more accurate, since the higher order corrections may be small due to the smallness of he perturbations. However, its implementation would not be of larger computational complexity than the existing f_{nl}-estimators, but would be rewarded by gaining a sensitive detector for non-Gaussianities structures imprinted by imperfectly subtracted foregrounds, or other data processing steps. This should motivate further research in this direction.

VI. BOLTZMANN-SHANNON INFORMATION

A. 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 bytes. This is possible by adopting the Boltzmann-Shannon information concept of information being negative entropy. The entropy of a signal probability function measures the phase-space volume available for signal uncertainties, and the information measure, as the negative entropy, expresses therefore the constraintness of the remaining uncertainties. It turns then out that further thermodynamical concepts, like the Helmholtz free energy, carry over to information theory, and both theories can be treated within an identical formalism.

The amount of available information on a signal is given by its negative entropy

$$\begin{aligned} I_d &= \int \mathcal{D}s\, P(s|d)\, \log P(s|d) \\ &= -\int \mathcal{D}s\, \frac{1}{Z}\, e^{-H[s]}\, (H[s] + \log Z) \\ &= -\langle H[s] \rangle_d - \log Z_d, \end{aligned} \tag{130}$$

the (negative of the) expectation value of the Hamiltonian plus logarithmic partition function. Introducing

$$Z_\beta[d,J] = \int \mathcal{D}s\, \exp\left\{-\beta\,(H[s] - J^\dagger s)\right\}, \text{ and} \tag{131}$$

$$F_\beta[d,J] = -\frac{1}{\beta}\, \log Z_\beta[d,J], \tag{132}$$

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

$$I_d = -\log Z_1[d,0] - \langle H[s] \rangle_d = -\left.\frac{\partial F_\beta[d,J]}{\partial \beta}\right|_{\beta=1,\,J=0}, \tag{133}$$

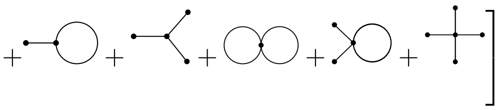
as can be verified by a direct calculation. The second expression for I_d in Eq. 133 actually holds even if the Hamiltonian is improperly normalised, e.g. H_0 can be chosen arbitrarily if $Z_\beta[d,J]$ is calculated consistently with this choice.

The Helmholtz free energy $F_\beta[J] = -\beta^{-1} \log Z_\beta[J]$ not only encodes the amount of signal information, $I_d = -\partial F_\beta[d,J]/\partial\beta|_{\beta=1,\,J=0}$, but is also the generator of all connected correlation functions of the signal $\langle s_{x_1} \cdots s_{x_n} \rangle^{\mathrm{c}}_{(s|d)} = -\delta^n F_\beta[d,J]/\delta J_{x_1} \cdots \delta J_{x_n}|_{\beta=1,\,J=0}$. The Helmholtz free energy can be calculated as follows:

$$\begin{aligned} F_\beta &= -\frac{1}{\beta}\, \log\left(\frac{Z_\beta^{\mathrm{G}}[J]}{Z_\beta^{\mathrm{G}}[J]} \int \mathcal{D}s\, e^{-\beta\, H_{\mathrm{int}}[s]}\, e^{-\beta\,(H_{\mathrm{G}}[s] - J^\dagger s)} \right) \\ &= -\frac{1}{\beta}\, \log Z_\beta^{\mathrm{G}}[J] - \frac{1}{\beta}\, \log \left\langle e^{-\beta\, H_{\mathrm{int}}[s]} \right\rangle_{(s|J+j,\mathrm{G})}, \end{aligned} \tag{134}$$

where the average in the last term is over the Gaussian probability function $P^{\mathrm{G}}_{J,\beta}[s] \propto \exp(-\beta\,(H_{\mathrm{G}}[s] - J^\dagger s))$. 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 [53]. For example, a free theory, perturbed by small, up-to-fourth-order interaction terms (all being proportional to some small parameter γ), has

$$F_\beta[J] = \underbrace{H_0^{\mathrm{G}} + \Lambda^{(0)}}_{H_0} - \beta^{-1} \Big[$$



$$\Big] + \mathcal{O}(\gamma^2), \quad (135)$$

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 recall

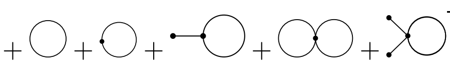
$$= \frac{1}{2}\,\log|2\pi\,D\,\beta^{-1}| = \frac{1}{2}\,\mathrm{Tr}(\log(2\pi\,D\,\beta^{-1})).$$

Thus, we have

$$\begin{aligned} F_\beta[J] \;=\; & H_0 - \frac{1}{2\,\beta}\,\mathrm{Tr}(\log(2\pi\,D\,\beta^{-1})) + \frac{1}{2\,\beta}\,\Lambda^{(2)}[D] \\ & + \frac{1}{2}\,(J+j-\Lambda^{(1)})^\dagger (D+\Lambda^{(2)})\,(J+j-\Lambda^{(1)}) \\ & + \frac{1}{2\,\beta}\Lambda^{(3)}[D, m_J] + \frac{1}{3!}\Lambda^{(3)}[m_J, m_J, m_J] \\ & + \frac{1}{8\,\beta^2}\Lambda^{(4)}[D, D] + \frac{1}{4\,\beta}\Lambda^{(4)}[D, m_J, m_J] \\ & + \frac{1}{4!}\Lambda^{(4)}[m_J, m_J, m_J, m_J] + \mathcal{O}(\gamma^2), \qquad (136) \end{aligned}$$

where we introduced the zero-order map $m_J = D\,(J+j)$ for notational convenice. The power of β associated with the different diagrams in Eq. 135 is given by the number of vertices minus the number of propagators minus one. 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:

$$I_d = -\Bigg[\frac{\varrho}{2} +$$



$$\Bigg] + \mathcal{O}(\gamma^2)$$

$$\begin{aligned} &= \frac{1}{2}\,\Big[-\mathrm{Tr}(1+\log(2\pi\,D)) + \Lambda^{(2)}[D] + \Lambda^{(3)}[D, m_0] \\ &\quad + \frac{1}{2}\Lambda^{(4)}[D \otimes (D + m_0\,m_0^\dagger)]\Big] + \mathcal{O}(\gamma^2), \qquad (137) \end{aligned}$$

where $\varrho = \mathrm{Tr}(1)$, $\beta = 1$, $J = 0$, and thus $m_0 = D\,j$. Here, we introduced the symmetrised tensor product $\otimes$, which has the property

$$A_{x_1 \dots x_n} \otimes B_{x_{n+1} \dots x_m} = \frac{1}{m!} \sum_{\pi \in \mathcal{P}_{n+m}} A_{x_{\pi(1)} \dots x_{\pi(n)}}\, B_{x_{\pi(n+1)} \dots x_{\pi(m)}}, \qquad (138)$$

with $\mathcal{P}_m$ being the set of permutations of $\{1, \dots, m\}$, while assuming $m > n$.

B. Free theory

To obtain the information content of the free theory, we can set $\gamma = 0$ in Eqs. 136 and 137 or use Eq. 58 with the 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

$$\begin{aligned} F_\beta[J] &= H_0^{\mathrm{G}} - \frac{1}{2}(J+j)^\dagger D\,(J+j) - \frac{1}{2\beta}\mathrm{Tr}\log\left(\frac{2\,\pi}{\beta}D\right), \text{and} \\ I_d &= -\frac{1}{2}\,\mathrm{Tr}\,(1 + \log(2\,\pi\,D))\,. \qquad (139) \end{aligned}$$

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

$$I_0 = -\frac{1}{2}\,\mathrm{Tr}\,(1 + \log(2\,\pi\,S))\,. \qquad (140)$$

Thus, the data-induced information gain is

$$\begin{aligned} \Delta I_d \;&=\; I_d - I_0 = \frac{1}{2}\,\mathrm{Tr}\left(\log\left(S\,D^{-1}\right)\right) \\ &=\; \frac{1}{2}\,\mathrm{Tr}\left(\log\left(1 + S\,R^\dagger N^{-1}\,R\right)\right). \qquad (141) \end{aligned}$$

The information gain depends on the signal-response-to-noise ratio $Q \equiv R\,S\,R^\dagger N^{-1}$, also shortly denoted by the measurement fidelity or quality. The information increases linearly with this ratio for it having a small value, but levels off to a logarithmic increase as soon the ratio exceeds one by large.

In the case of a 4^{th}-order interacting theory, where, however, the prior was Gaussian, the data induced information gain is still given by Eq. 141 plus the nonlinear terms on the rhs of Eq. 137.

C. Optimal interferometric information yield

We want to show how IFT can be used to optimise the scientific return of investment of observational instruments. In order to keep the discussion as transparent as possible, we concentrate on an idealised case. More realistic cases can be treated analogously, however, with a higher mathematical complexity, which is unwanted for the moment in order not to obscure our arguments.

An interferometer observing a Gaussian signal like the CMB may be described by a free theory, since signal and noise are Gaussian with high accuracy, the instrument signal response is linear, and non-Gaussian contaminating signals may be sufficiently reduced by a proper choice of observational wavelength and field on the sky. Assuming for the moment that the observational field is sufficiently small, so that the flat sky approximation holds, the instrument response is roughly independent of sky position well within the primary beam of the instrument, which can be large for a phased array. A situation in which the signal and noise statistics, as well as the instrumental response is position independent is described by our simplistic Hamiltonian[17] (Sect. III A 4). In this

[17] $S(k,q) = (2\pi)^n\,\delta(k-q)\,P_S(k)$, $N(k,q) = (2\pi)^n\,\delta(k-q)\,P_N(k)$, $P_s(k) = \langle|s(k)|^2\rangle/V$, $P_n(k) = \langle|n(k)|^2\rangle/V$. $d(x) = \int dy\,R(x-y)\,s(y) + n(x))$, $P_R(k) = |R(k)|^2$.

case we find

$$\Delta I = \frac{V}{2} \int \frac{dk}{(2\,\pi)^n} \, \log\left(1 + P_Q(k)\right), \tag{142}$$

where $P_Q(k) = P_S(k)\,P_R(k)/P_N(k)$ is the quality spectrum and V is the observed volume, or sky area in our example.[18] Note that the integrand is zero for k-space regions without response. If discrete samples were taken in Fourier space the above integral should be replaced by

$$V \int \frac{dk}{(2\,\pi)^n} \to \sum_{k_i}, \tag{143}$$

where k_i are the observed k-vectors, where $P_R(k) \neq 0$, which is also consistent with taking the trace in data-space.

We want to model the addition of further data in Fourier space, e.g. by adding baselines to an interferometric observation, which enhances the sensitivity for signals at a wavevector k. We therefore introduce the Fourier space exposure function $\varrho(k)$, which just tells us for how many observing (or time) units a measurement at a given k vector has been performed. The signal response adds up coherently and therefore quadratically with additional observations, whereas the noise adds up incoherently and thus only linearly. This means that the replacements $P_R(k) \to P_R(k)\,\varrho^2(k)$ and $P_N(k) \to P_N(k)\,\varrho(k)$ lead to a functional of ϱ describing the information gain:

$$\Delta I[\varrho] = \frac{V}{2} \int \frac{dk}{(2\,\pi)^n} \, \log\left(1 + P_Q(k)\,\varrho(k)\right) \equiv \int dk\, I_k. \tag{144}$$

Given that the differential information gain per observational unit $\varrho(k)$

$$\frac{\delta I}{\delta \varrho(k)} = \frac{V}{2\,(2\,\pi)^n} \, \frac{P_Q(k)}{1 + \varrho(k)\,P_Q(k)} \tag{145}$$

levels off for $\varrho > P_Q^{-1}$, one can ask what the optimal observation time is, from an information-economical point of view.

Assuming simply that an observation unit at k costs an amount of $c(k)$, and that an information unit on the signal at k has a value of $v(k)$, one finds by optimising the benefit $b_k = I_k(\varrho_k)\,v_k - \varrho_k\,c_k$ with respect to the observing time ϱ_k that the optimum is given by

$$\varrho_{\mathrm{opt}}(k) = \frac{V}{2\,(2\pi)^n} \, \frac{v_k}{c_k} - \left(\frac{P_N}{P_S\,P_R}\right)(k)\,. \tag{146}$$

This implies that – as far as our very simple information-economical model holds – observations should only be done at wavenumbers for which the costs stay below an economical limit provided by the quality times the value:

$$c_k < \frac{V\,P_Q(k)\,v_k}{2\,(2\pi)^n}. \tag{147}$$

The scientific value of some information at a given spatial wavenumber may be derived from its information content on yet unknown cosmological parameters. The price of such parameter information is actually set by indirect market mechanisms involving scientific funding agencies and the competition of scientific collaborations for fundamental discoveries. However, providing the recipes to assign concrete prices is beyond the scope of this paper. It should be added, that such a pricing without including the discovery potential of an observation risks to bias the scientific research direction towards being aligned to existing prejudices [210].

D. Poissonian information content

In 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. This changes in case that one wants to harvest information in a situation described by a non-linear IFT. There, the amount of information can strongly depend on the actual data.[19].

Here, we want to investigate how much information on an underlying Gaussian signal can be obtained in a non-linear case, and how it depends on the data. We choose the Poissonian Hamiltonian of Sect. IV as our reference case, and assume for our convenience that either the bias-factor or the signal amplitude, which both control the strength of the non-linear interactions, are small compared to unity. 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 $\langle s^2 \rangle_{(s)} < 1$.

Then the information gain expanded to the first few orders in b

$$\begin{aligned}
\Delta I \;=\; & \frac{1}{2}\,\mathrm{Tr} \log\left(1 + S\,\widehat{\kappa\, b^2}\right) \\
& + \frac{1}{2}\,\left(\kappa\, b^3 \widehat{D}\right)^{\dagger} \left(m_0 + \frac{1}{2}\, b\,(\widehat{D} + m_0^2)\right) + \mathcal{O}(b^5),
\end{aligned} \tag{148}$$

clearly depends on the actual realisation of the data. The different fluctuations in the naive map $m_0 = D\,j$, with $D = (S^{-1} + \widehat{b^2\,\kappa})^{-1}$ and $j = b\,(d - \kappa)$, contain positive or negative information. In general, the positive information fluctuations seem to dominate, since only in case

[18] The volume-factor enters with the replacement of the delta function in k-space $\delta(0) \to V/(2\pi)^n$.

[19] This may be confirmed by any astronomical observer, who faced the problem of how to convert a technically perfect, but unsuccessful search for astronomical objects into an acceptable scientific publication.

that $m_0(x) \in [m_-(x), m_+(x)]$, with

$$m_\pm(x) = \frac{1}{b_x}\left(-1 \pm \sqrt{1 - b_x^2\, D_{xx}}\right) \leq 0, \qquad (149)$$

the information fluctuation is negative, otherwise it is positive.

A bit more insight into the non-linear information content can be obtained by analysing the special case of a homogeneous response b, expected number of counts κ and a translation-invariant signal statistics. All these condition may approximately be met by a volume limited sample of observed galaxies in the universe, tracing the large-scale structure of the dark matter distribution. Note, that since a fundamental theory for translating our signal s, the logarithm of the cosmic matter density, into the expectation value of the observed number of galaxies at some location is lacking, at least the absolute normalisation for κ has to be deduced from the data itself.

If we know the spatial shape of $\kappa(x) = \kappa_0\, f(x)$, with $f(x)$ the known shape function ($f = \text{const}$ in our following example), and κ_0 the unknown normalisation to be derived observationally, we could trade the information of the zero Fourier mode for the κ_0-determination by requiring $\int dx\, j(x) = 0$, which yields

$$\kappa_0 = b^\dagger d / b^\dagger f, \qquad (150)$$

and is $\kappa_0 = \int dx\, d(x)/V$ in our translational invariant case.

By calibrating our zero-signal expectation value κ_0 we have removed any zero-Fourier-mode information in the signal field, and may have introduced a systematic error in our data model. We ignore the latter uncertainty for the moment, it will be addressed in a subsequent work, but model the missing zero-mode by removing its contribution completely from any Fourier space based information estimate. Having said this, we find

$$\begin{aligned}\Delta I \;=\; V \int_{k\neq 0} \frac{dk}{(2\pi)^n} \,\Big[\log\left(1 + \kappa\, b^2\, P_S(k)\right) \\ + \frac{\kappa\, b^4\, \widehat{D}}{4} \int_{k'\neq 0} \frac{dk'}{(2\pi)^n}\, m(k)\, m(k-k')\Big], \qquad (151)\end{aligned}$$

with $m(k) = P_D(k)\, j(k)$, $P_D(k) = (P_S(k)^{-1} + \kappa\, b^2)^{-1}$, $P_S(0) = P_D(0) = 0$, and

$$\widehat{D} = \int \frac{dk}{(2\pi)^n}\, P_D(k). \qquad (152)$$

For Gaussian signal and noise, one would expect $\langle m(k)\, m(k-k')\rangle_{(d,s)} = 0$ for $k' \neq 0$ and due to translational invariance, however, the Poissonian statistics should have introduced mode couplings, which can be exploited for better signal reconstruction and therefore contain information. In real space these non-linear information fluctuations are $\kappa\, b^4\, \widehat{D}\, m_x^2/4$ and therefore strictly positive.

For an ongoing survey, it might be advantageous to invest more observing times on regions with a higher information content, once they are identified. However, the future survey response function becomes then conditional on the data already obtained, a complication which is not yet fully understood how to be taken into account in a statistical data analysis. Therefore, we do not recommend such survey strategies for the moment, but like to encourage further theoretical research in this direction.

VII. SUMMARY AND OUTLOOK

A. Information Field Theory

The optimal extraction and restauration of information on spatially distributed quantities like the cosmic large-scale structure 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, methodologies and schools. Here, we addressed such problems from a strictly information theoretical point of view, with the aim of erasing any conceptual and practical ambiguity about the optimal method for a given problem.

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 *information field theory* (IFT). IFT is mathematically a statistical field theory (SFT), however, its probabilistic aspects are due to knowledge-uncertainties and not due to thermal fluctuations. 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 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 Wiener-filter theory, and presented the Feynman-rules for non-linear, 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 to argue why the resulting IFTs are usually well normalised, but often non-local. 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 for a large variety of applications in Cosmology and elsewhere.

B. 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 summarised as following:

- 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.

- 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.

- Calculate the Hamiltonian $H_d[s] = -\log(P(d,s))$, where $P(d,s) = P(d|s)\,P(s)$ is the joined 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 n^{th}-order terms with the normalisation H_0, information source j, inverse propagator D^{-1}, and n^{th}-order interaction term $\Lambda^{(n)}$, respectively, as shown in Eq. 45 or 68.

- 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.

- Read the diagrams as computational algorithms specified by the Feynman rules in Sect. III, and implement them by using linear algebra packages or existing map-making codes for the information propagator and vertices. The required discretisation is outlined in Sect. I E 1. Information on how to implement the required matrix inversions efficiently can be found in the literature given in Secs. I F 2, I F 4, and I F 5 and especially in [37].

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

- In case that too large interaction terms in the Hamiltonian prevent a finite number of diagrams to form a well performing algorithm, a re-summation 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 renormalisation-flow analysis along the lines outlined in Sect. IV F.

C. IFT Examples

As examples of the IFT recipe, two concrete IFT problems with cosmological motivation were discussed. The first was targeting at the problem of reconstructing the spatially continuous cosmic large-scale structure matter distribution from discrete galaxy counts in incomplete galaxy surveys. It was demonstrated analytically and numerically that a Gaussian signal, which should resemble the initial density perturbations of the Universe, observed with a strongly non-linear, incomplete and Poissonian-noise affected response, as the processes of structure and galaxy formation and observations may provide, can well be reconstructed thanks to the virtue of a response-renormalisation flow equation. Surprisingly, it turned out that the exact information field expectation value can be calculated with much lower numerical costs as compared to the classical maximum a posteriori estimator, despite its higher fidelity.

The second example was the design of an optimal method to measure or constrain any possible local non-linearities in the CMB temperature fluctuations, as it is predicted from some Early-Universe inflationary scenarios, and is also expected due to imperfections in the removal of CMB-foregrounds during the experimental data processing.

Finally, we provided the Boltzmann-Shannon information measure of IFT based on the Helmholtz free energy, and thereby highlighting conceptual similarities of IFT and SFT, and demonstrated, how this can be used to optimise the information yield of observational and experimental setups.

D. Outlook

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 with not precisely known statistics. Some the 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 semi-analytic descriptions. Also the CMB Hamiltonian may benefit from the inclusion of remnants from the CMB foreground subtraction process, permitting to gather more solid evidence on fundamental parameters which are hidden in the CMB fluctuations, like the amplitude of non-Gaussianities.

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

Acknowledgements

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 gratefully acknowledge helpful comments on the manuscript by Marcus Brüggen and Thomas Riller.

[1] T. Bayes, Phil. Trans. Roy. Soc. **53**, 370 (1763).
[2] J. C. Lemm, ArXiv Physics e-prints (1999), physics/9912005.
[3] C. E. Shannon, Bell System Technical Journal **27**, 379 (1948).
[4] C. E. Shannon and W. Weaver, *The mathematical theory of communication* (Urbana: University of Illinois Press, 1949, 1949).
[5] E. T. Jaynes, Physical Review **106**, 620 (1957).
[6] E. T. Jaynes, Physical Review **108**, 171 (1957).
[7] E. T. Jaynes, in *Statistical Physics 3* (1963), p. 181.
[8] E. T. Jaynes, American Journal of Physics **33**, 391 (1965).
[9] E. T. Jaynes, IEEE Trans. on Systems Science and Cybernetics **SSC-4**, 227 (1968).
[10] E. T. Jaynes, in *Proc. IEEE, Volume 70, p. 939-952* (1982), pp. 939–952.
[11] E. T. Jaynes and R. Baierlein, Physics Today **57**, 76 (2004).
[12] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. T. Teller, Journal of Chemical Physics **21**, 1087 (1953).
[13] W. K. Hastings, Biometrika **57**, 97 (1970).
[14] S. Geman and D. Geman, IEEE Transactions on Pattern Analysis and Machine Intelligence **6**, 721 (1984).
[15] R. A. Aster, B. Brochers, and C. H. Thurber, *Parameter estimation and inverse problems* (Elsevier Academic Press, London, 2005).
[16] A. Gelman, J. B. Carlin, H. S. Stern, and D. Rubin, *Bayesian data analysis* (Chapman & Hall/CRC, Boca Raton, Florida, 2004).
[17] R. M. Neal, in *Technical Report CRG-TR-93-1* (Dept. of Computer Science, University of Toronto, 1993).
[18] C. P. Robert, *The Bayesian choice* (Springer-Verlag, New York, 2001).
[19] M. A. Tanner, *Tools for statistical inference* (Springer-Verlag, New York, 1996).
[20] R. Trotta, ArXiv e-prints **0803.4089** (2008), 0803.4089.
[21] N. Wiener, *Extrapolation, Interpolation, and Smoothing of Stationary Time Series* (New York: Wiley, 1949).
[22] L. B. Lucy, AJ **79**, 745 (1974).
[23] W. H. Richardson, Journal of the Optical Society of America (1917-1983) **62**, 55 (1972).
[24] B. R. Frieden, Journal of the Optical Society of America (1917-1983) **62**, 511 (1972).
[25] R. K. Bryan and J. Skilling, MNRAS **191**, 69 (1980).
[26] R. K. Bryan and J. Skilling, Journal of Modern Optics **33**, 287 (1986).
[27] S. F. Burch, S. F. Gull, and J. Skilling, Computer Vision Graphics and Image Processing **23**, 113 (1983).
[28] S. F. Gull, in *Maximum Entropy and Bayesian Methods*, edited by J. Skilling (Kluwer Academic Publishers, Dordtrecht, 1989), pp. 53–71.
[29] S. F. Gull and G. J. Daniell, Nature (London) **272**, 686 (1978).
[30] S. F. Gull and J. Skilling, in *Indirect Imaging. Measurement and Processing for Indirect Imaging. Proceedings of an International Symposium held in Sydney, Australia, August 30-September 2, 1983. Editor, J.A. Roberts, Publisher, Cambridge University Press, Cambridge, England, New York, NY, 1984. LC # QB51.3.E43 I53 1984. ISBN # 0-521-26282-8. P.267, 1983* (1983), p. 267.
[31] S. F. Gull and J. Skilling, *The MEMSYS5 User's Manual* (Maximum Entropy Data Consultants Ltd, Royston, 1990).
[32] S. Sibisi, J. Skilling, R. G. Brereton, E. D. Laue, and J. Staunton, Nature (London) **311**, 446 (1984).
[33] J. Skilling, in *Maximum Entropy and Bayesian Methods*, edited by G. J. Erickson, J. T. Rychert, and C. R. Smith (1998), p. 1.
[34] J. Skilling and R. K. Bryan, MNRAS **211**, 111 (1984).
[35] J. Skilling, A. W. Strong, and K. Bennett, MNRAS **187**, 145 (1979).
[36] D. M. Titterington and J. Skilling, Nature (London) **312**, 381 (1984).
[37] F. S. Kitaura and T. A. Enßlin, ArXiv e-prints **0705.0429** (2007), 0705.0429.
[38] R. Narayan and R. Nityananda, ARAA **24**, 127 (1986).
[39] R. Molina, J. Nunez, F. J. Cortijo, and J. Mateos, Signal Processing Magazine, IEEE **18**, 11 (2001).
[40] E. Bertschinger, ApJL **323**, L103 (1987).
[41] W. Bialek and A. Zee, Physical Review Letters **58**, 741 (1987).
[42] W. Bialek and A. Zee, Physical Review Letters **61**, 1512 (1988).

[43] W. Bialek, C. G. Callan, and S. P. Strong, Physical Review Letters **77**, 4693 (1996), arXiv:cond-mat/9607180.
[44] P. Stoica, E. G. Larsson, and J. Li, AJ **120**, 2163 (2000).
[45] J. C. Lemm, Physics Letters A **276**, 19 (2000).
[46] J. C. Lemm, in *Bayesian Inference and Maximum Entropy Methods in Science and Engineering*, edited by A. Mohammad-Djafari (2001), vol. 568 of *American Institute of Physics Conference Series*, pp. 425–436.
[47] J. C. Lemm and J. Uhlig, Few-Body Systems **29**, 25 (2000), arXiv:quant-ph/0006027.
[48] J. C. Lemm and J. Uhlig, Physical Review Letters **84**, 4517 (2000), arXiv:nucl-th/9908056.
[49] J. C. Lemm, J. Uhlig, and A. Weiguny, Physical Review Letters **84**, 2068 (2000), arXiv:cond-mat/9907013.
[50] J. C. Lemm, J. Uhlig, and A. Weiguny, European Physical Journal B **20**, 349 (2001), arXiv:quant-ph/0005122.
[51] J. C. Lemm, J. Uhlig, and A. Weiguny, European Physical Journal B **46**, 41 (2005).
[52] J. C. Lemm, ArXiv Condensed Matter e-prints (1998), cond-mat/9808039.
[53] J. Binney, N. Dowrick, A. Fisher, and M. Newman, *The theory of critical phenomena* (Oxford University Press, Oxford, UK: ISBN0-19-851394-1, 1992).
[54] M. E. Peskin and D. V. Schroeder, *An Introduction to Quantum Field Theory* (Westview Press Boulder, Colorado: 1995, ISBN-13 978-0-201-50397-5., 1995).
[55] A. Zee, *Quantum field theory in a nutshell* (Quantum field theory in a nutshell, by A. Zee. Princeton, NJ: Princeton University Press, 2003, ISBN 0691010196., 2003).
[56] S. Matarrese, F. Lucchin, and S. A. Bonometto, ApJL **310**, L21 (1986).
[57] J. M. Bardeen, J. R. Bond, N. Kaiser, and A. S. Szalay, Astrophys. J. **304**, 15 (1986).
[58] F. Bernardeau, ApJL **390**, L61 (1992).
[59] F. Bernardeau, M. J. Chodorowski, E. L. Łokas, R. Stompor, and A. Kudlicki, MNRAS **309**, 543 (1999), astro-ph/9901057.
[60] E. Branchini, L. Teodoro, C. S. Frenk, I. Schmoldt, G. Efstathiou, S. D. M. White, W. Saunders, W. Sutherland, M. Rowan-Robinson, O. Keeble, et al., MNRAS **308**, 1 (1999), astro-ph/9901366.
[61] A. Dekel and O. Lahav, Astrophys. J. **520**, 24 (1999), astro-ph/9806193.
[62] A. J. S. Hamilton, in *The Evolving Universe*, edited by D. Hamilton (Kluwer Academic Publishers, Dordtrecht, 1998), vol. 231 of *Astrophysics and Space Science Library*, p. 185.
[63] N. Kaiser, MNRAS **227**, 1 (1987).
[64] P. J. E. Peebles, *The large-scale structure of the universe* (Research supported by the National Science Foundation. Princeton, N.J., Princeton University Press, 1980. 435 p., 1980).
[65] P. J. E. Peebles, Astrophys. J. **362**, 1 (1990).
[66] R. Scoccimarro, Phys. Rev. D **70**, 083007 (2004), astro-ph/0407214.
[67] 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, MNRAS **341**, 1311 (2003), arXiv:astro-ph/0207664.
[68] S. Zaroubi, ArXiv Astrophysics e-prints (2002), astro-ph/0206052.
[69] S. Zaroubi and Y. Hoffman, Astrophys. J. **462**, 25 (1996).
[70] Y. B. Zel'dovich, A&A **5**, 84 (1970).
[71] M. Crocce and R. Scoccimarro, Physical Review D **73**, 063520 (2006), arXiv:astro-ph/0509419.
[72] M. Crocce and R. Scoccimarro, Physical Review D **73**, 063519 (2006), arXiv:astro-ph/0509418.
[73] J. Gaite and A. Domínguez, Journal of Physics A Mathematical General **40**, 6849 (2007), arXiv:astro-ph/0610886.
[74] D. Jeong and E. Komatsu, Astrophys. J. **651**, 619 (2006), arXiv:astro-ph/0604075.
[75] D. Jeong and E. Komatsu, ArXiv e-prints **0805.2632** (2008), 0805.2632.
[76] S. Matarrese and M. Pietroni, Journal of Cosmology and Astro-Particle Physics **6**, 26 (2007), arXiv:astro-ph/0703563.
[77] S. Matarrese and M. Pietroni, Modern Physics Letters A **23**, 25 (2008), arXiv:astro-ph/0702653.
[78] T. Matsubara, Phys. Rev. D **77**, 063530 (2008), arXiv:0711.2521.
[79] P. McDonald, Phys. Rev. D **74**, 103512 (2006), arXiv:astro-ph/0609413.
[80] P. McDonald, Phys. Rev. D **74**, 129901(E) (2006).
[81] P. McDonald, Phys. Rev. D **75**, 043514 (2007), arXiv:astro-ph/0606028.
[82] M. Pietroni, ArXiv e-prints **0806.0971** (2008), 0806.0971.
[83] P. Valageas, A&A **421**, 23 (2004), arXiv:astro-ph/0307008.
[84] P. Valageas, A&A **476**, 31 (2007), arXiv:0706.2593.
[85] P. Valageas, A&A **484**, 79 (2008), arXiv:0711.3407.
[86] S. Basilakos and M. Plionis, Astrophys. J. **550**, 522 (2001), astro-ph/0011265.
[87] F. Bernardeau, A&A **291**, 697 (1994), astro-ph/9403020.
[88] E. Bertschinger and A. Dekel, ApJL **336**, L5 (1989).
[89] E. Bertschinger and A. Dekel, in *ASP Conf. Ser. 15: Large-Scale Structures and Peculiar Motions in the Universe*, edited by D. W. Latham and L. A. N. da Costa (1991), p. 67.
[90] V. Bistolas and Y. Hoffman, Astrophys. J. **492**, 439 (1998), astro-ph/9707243.
[91] C. S. Botzler, J. Snigula, R. Bender, and U. Hopp, MNRAS **349**, 425 (2004), arXiv:astro-ph/0312018.
[92] Y. Brenier, U. Frisch, M. Hénon, G. Loeper, S. Matarrese, R. Mohayaee, and A. Sobolevskiĭ, MNRAS **346**, 501 (2003), astro-ph/0304214.
[93] R. A. C. Croft and E. Gaztanaga, MNRAS **285**, 793 (1997), astro-ph/9602100.
[94] A. Dekel, E. Bertschinger, and S. M. Faber, Astrophys. J. **364**, 349 (1990).
[95] K. B. Fisher, O. Lahav, Y. Hoffman, D. Lynden-Bell, and S. Zaroubi, MNRAS **272**, 885 (1995), astro-ph/9406009.
[96] U. Frisch, S. Matarrese, R. Mohayaee, and A. Sobolevski, Nature (London) **417**, 260 (2002), arXiv:astro-ph/0109483.
[97] G. Ganon and Y. Hoffman, ApJL **415**, L5 (1993).
[98] D. M. Goldberg, Astrophys. J. **552**, 413 (2001), astro-ph/0008266.
[99] D. M. Goldberg and D. N. Spergel, Astrophys. J. **544**, 21 (2000), astro-ph/9912408.
[100] D. M. Goldberg and D. N. Spergel, in *ASP Conf. Ser. 201: Cosmic Flows Workshop*, edited by S. Courteau

and J. Willick (2000), p. 282.
[101] M. Gramann, Astrophys. J. **405**, 449 (1993).
[102] Y. Hoffman and E. Ribak, ApJL **380**, L5 (1991).
[103] N. Kaiser and A. Stebbins, in *ASP Conf. Ser. 15: Large-Scale Structures and Peculiar Motions in the Universe*, edited by D. W. Latham and L. A. N. da Costa (1991), p. 111.
[104] A. Kudlicki, M. Chodorowski, T. Plewa, and M. Różyczka, MNRAS **316**, 464 (2000), astro-ph/9910018.
[105] O. Lahav, in *ASP Conf. Ser. 67: Unveiling Large-Scale Structures Behind the Milky Way*, edited by C. Balkowski and R. C. Kraan-Korteweg (1994), p. 171.
[106] O. Lahav, K. B. Fisher, Y. Hoffman, C. A. Scharf, and S. Zaroubi, ApJL **423**, L93 (1994), astro-ph/9311059.
[107] R. Mohayaee, U. Frisch, S. Matarrese, and A. Sobolevskii, A&A **406**, 393 (2003), arXiv:astro-ph/0301641.
[108] R. Mohayaee, H. Mathis, S. Colombi, and J. Silk, MNRAS **365**, 939 (2006), astro-ph/0501217.
[109] R. Mohayaee, B. Tully, and U. Frisch, ArXiv Astrophysics e-prints (2004), astro-ph/0410063.
[110] R. Mohayaee and R. B. Tully, ApJL **635**, L113 (2005), astro-ph/0509313.
[111] V. K. Narayanan and R. A. C. Croft, Astrophys. J. **515**, 471 (1999), astro-ph/9806255.
[112] V. K. Narayanan and D. H. Weinberg, Astrophys. J. **508**, 440 (1998), astro-ph/9806238.
[113] A. Nusser and M. Davis, ApJL **421**, L1 (1994), astro-ph/9309009.
[114] A. Nusser and A. Dekel, Astrophys. J. **391**, 443 (1992).
[115] P. J. E. Peebles, ApJL **344**, L53 (1989).
[116] U.-L. Pen, Astrophys. J. **504**, 601 (1998), astro-ph/9711180.
[117] G. B. Rybicki and W. H. Press, Astrophys. J. **398**, 169 (1992).
[118] U. Seljak, Astrophys. J. **503**, 492 (1998), astro-ph/9710269.
[119] U. Seljak, Astrophys. J. **506**, 64 (1998), astro-ph/9711124.
[120] R. K. Sheth, MNRAS **277**, 933 (1995), astro-ph/9511096.
[121] A. Taylor and H. Valentine, MNRAS **306**, 491 (1999), astro-ph/9901171.
[122] M. Tegmark and B. C. Bromley, Astrophys. J. **453**, 533 (1995), astro-ph/9409038.
[123] D. H. Weinberg, MNRAS **254**, 315 (1992).
[124] S. Zaroubi, MNRAS **331**, 901 (2002), astro-ph/0010561.
[125] S. Zaroubi, Y. Hoffman, and A. Dekel, Astrophys. J. **520**, 413 (1999), astro-ph/9810279.
[126] S. Zaroubi, Y. Hoffman, K. B. Fisher, and O. Lahav, Astrophys. J. **449**, 446 (1995), astro-ph/9410080.
[127] F. Bernardeau and R. van de Weygaert, MNRAS **279**, 693 (1996).
[128] V. Icke and R. van de Weygaert, qras **32**, 85 (1991).
[129] S. Ikeuchi and E. L. Turner, MNRAS **250**, 519 (1991).
[130] M. Ramella, W. Boschin, D. Fadda, and M. Nonino, A&A **368**, 776 (2001), arXiv:astro-ph/0101411.
[131] W. E. Schaap and R. van de Weygaert, A&A **363**, L29 (2000), astro-ph/0011007.
[132] R. van de Weygaert and W. Schaap, in *Mining the Sky*, edited by A. J. Banday, S. Zaroubi, and M. Bartelmann (2001), p. 268.
[133] L. Zaninetti, Chinese Journal of Astronomy and Astrophysics **6**, 387 (2006), arXiv:astro-ph/0602431.
[134] E. Bertschinger, A. Dekel, S. M. Faber, A. Dressler, and D. Burstein, Astrophys. J. **364**, 370 (1990).
[135] E. Branchini, M. Plionis, and D. W. Sciama, ApJL **461**, L17 (1996), astro-ph/9512055.
[136] P. Erdoğdu, O. Lahav, J. Huchra, and et al., MNRAS **373**, 45 (2006), astro-ph/0610005.
[137] P. Erdoğdu, O. Lahav, S. Zaroubi, and et al., MNRAS **352**, 939 (2004), astro-ph/0312546.
[138] K. B. Fisher, C. A. Scharf, and O. Lahav, MNRAS **266**, 219 (1994), astro-ph/9309027.
[139] D. M. Goldberg, Astrophys. J. **550**, 87 (2001), astro-ph/0009046.
[140] Y. Hoffman and S. Zaroubi, ApJL **535**, L5 (2000), astro-ph/0003306.
[141] J. Huchra, T. Jarrett, M. Skrutskie, R. Cutri, S. Schneider, L. Macri, R. Steining, J. Mader, N. Martimbeau, and T. George, in *ASP Conf. Ser. 329: Nearby Large-Scale Structures and the Zone of Avoidance*, edited by A. P. Fairall and P. A. Woudt (2005), p. 135.
[142] H. Mathis, G. Lemson, V. Springel, G. Kauffmann, S. D. M. White, A. Eldar, and A. Dekel, MNRAS **333**, 739 (2002), astro-ph/0111099.
[143] A. Nusser and M. Haehnelt, MNRAS **303**, 179 (1999), astro-ph/9806109.
[144] W. J. Percival, MNRAS **356**, 1168 (2005), astro-ph/0410631.
[145] I. M. Schmoldt, V. Saar, P. Saha, E. Branchini, G. P. Efstathiou, C. S. Frenk, O. Keeble, S. Maddox, R. McMahon, S. Oliver, et al., Astrophys. J. **118**, 1146 (1999), astro-ph/9906035.
[146] E. J. Shaya, P. J. E. Peebles, and R. B. Tully, Astrophys. J. **454**, 15 (1995), astro-ph/9506144.
[147] M. Tegmark and B. C. Bromley, The Astrophysical Journal **518**, L69 (1999), URL `http://www.citebase.org/abstract?id=oai:arXiv.org:astro-ph`
[148] 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 (2004), pp. 5–11.
[149] M. Webster, O. Lahav, and K. Fisher, MNRAS **287**, 425 (1997), astro-ph/9608021.
[150] A. Yahil, M. A. Strauss, M. Davis, and J. P. Huchra, Astrophys. J. **372**, 380 (1991).
[151] C. Yess, S. F. Shandarin, and K. B. Fisher, Astrophys. J. **474**, 553 (1997), astro-ph/9605041.
[152] Y. Hoffman, in *ASP Conf. Ser. 67: Unveiling Large-Scale Structures Behind the Milky Way*, edited by C. Balkowski and R. C. Kraan-Korteweg (1994), p. 185.
[153] R. C. Kraan-Korteweg and O. Lahav, AAPR **10**, 211 (2000), astro-ph/0005501.
[154] S. Zaroubi, in *ASP Conf. Ser. 218: 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 (2000), p. 173.
[155] D. J. Eisenstein and W. Hu, Astrophys. J. **511**, 5 (1999), astro-ph/9710252.
[156] J. A. Peacock and S. J. Dodds, MNRAS **267**, 1020 (1994), astro-ph/9311057.
[157] M. Tegmark, Physical Review Letters **79**, 3806 (1997), astro-ph/9706198.
[158] M. S. Vogeley and A. S. Szalay, Astrophys. J. **465**, 34 (1996), astro-ph/9601185.
[159] S. Zaroubi, I. Zehavi, A. Dekel, Y. Hoffman, and T. Ko-

latt, Astrophys. J. **486**, 21 (1997), astro-ph/9610226.
[160] E. F. Bunn, D. Scott, and M. White, ApJL **441**, L9 (1995), astro-ph/9409003.
[161] G. Efstathiou, J. R. Bond, and S. D. M. White, MNRAS **258**, 1P (1992).
[162] E. F. Bunn, K. B. Fisher, Y. Hoffman, O. Lahav, J. Silk, and S. Zaroubi, ApJL **432**, L75 (1994), astro-ph/9404007.
[163] S. Dodelson, Astrophys. J. **482**, 577 (1997), astro-ph/9512021.
[164] O. Doré, R. Teyssier, F. R. Bouchet, D. Vibert, and S. Prunet, A&A **374**, 358 (2001), astro-ph/0101112.
[165] 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, ApJS **155**, 227 (2004), astro-ph/0407028.
[166] G. Hinshaw et al. (WMAP), arXiv **0803.0732** (2008), 0803.0732.
[167] M. P. Hobson, A. W. Jones, A. N. Lasenby, and F. R. Bouchet, MNRAS **300**, 1 (1998), astro-ph/9806387.
[168] M. A. Janssen and S. Gulkis, in *NATO ASIC Proc. 359: The Infrared and Submillimetre Sky after COBE*, edited by M. Signore and C. Dupraz (Kluwer Academic Publishers, Dordtrecht, 1992), pp. 391–408.
[169] J. Jewell, S. Levin, and C. H. Anderson, Astrophys. J. **609**, 1 (2004), astro-ph/0209560.
[170] E. Keihänen, H. Kurki-Suonio, and T. Poutanen, MNRAS **360**, 390 (2005), astro-ph/0412517.
[171] 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), astro-ph/0608007.
[172] K. Maisinger, M. P. Hobson, and A. N. Lasenby, MNRAS **290**, 313 (1997).
[173] P. Natoli, G. de Gasperis, C. Gheller, and N. Vittorio, A&A **372**, 346 (2001), astro-ph/0101252.
[174] R. Stompor, A. Balbi, J. D. Borrill, P. G. Ferreira, S. Hanany, A. H. Jaffe, A. T. Lee, S. Oh, B. Rabii, P. L. Richards, G.F. Smoot, C.D. Winan, J.-H. P. Wu, Phys. Rev. D **65**, 022003 (2001), astro-ph/0106451.
[175] E. C. Sutton and B. D. Wandelt, ApJS **162**, 401 (2006).
[176] M. Tegmark, Phys. Rev. D **56**, 4514 (1997), astro-ph/9705188.
[177] M. Tegmark, ApJL **480**, L87 (1997), astro-ph/9611130.
[178] B. D. Wandelt, D. L. Larson, and A. Lakshminarayanan, Phys. Rev. D **70**, 083511 (2004), astro-ph/0310080.
[179] D. Yvon and F. Mayet, A&A **436**, 729 (2005), astro-ph/0401505.
[180] U. Seljak and M. Zaldarriaga, Astrophys. J. **469**, 437 (1996), arXiv:astro-ph/9603033.
[181] A. Lewis, A. Challinor, and A. Lasenby, Astrophys. J. **538**, 473 (2000), astro-ph/9911177.
[182] M. Doran, Journal of Cosmology and Astro-Particle Physics **10**, 11 (2005), arXiv:astro-ph/0302138.
[183] E. F. Bunn and N. Sugiyama, Astrophys. J. **446**, 49 (1995), astro-ph/9407069.
[184] M. R. Nolta et al. (WMAP), arXiv **0803.0593** (2008), 0803.0593.
[185] M. Tegmark, Phys. Rev. D **55**, 5895 (1997), astro-ph/9611174.
[186] M. Tegmark, A. N. Taylor, and A. F. Heavens, Astrophys. J. **480**, 22 (1997), astro-ph/9603021.
[187] A. Albrecht and P. J. Steinhardt, Physical Review Letters **48**, 1220 (1982).
[188] J. M. Bardeen, P. J. Steinhardt, and M. S. Turner, Phys. Rev. D **28**, 679 (1983).
[189] A. H. Guth, Phys. Rev. D **23**, 347 (1981).
[190] A. H. Guth and S.-Y. Pi, Physical Review Letters **49**, 1110 (1982).
[191] A. D. Linde, Physics Letters B **108**, 389 (1982).
[192] A. A. Starobinsky, Physics Letters B **117**, 175 (1982).
[193] D. Babich, P. Creminelli, and M. Zaldarriaga, Journal of Cosmology and Astro-Particle Physics **8**, 9 (2004), arXiv:astro-ph/0405356.
[194] N. Bartolo, E. Komatsu, S. Matarrese, and A. Riotto, Phys. Rep. **402**, 103 (2004), astro-ph/0406398.
[195] F. Bernardeau and J.-P. Uzan, Phys. Rev. D **66**, 103506 (2002), hep-ph/0207295.
[196] W. Hu, Phys. Rev. D **64**, 083005 (2001), astro-ph/0105117.
[197] D. Babich and M. Zaldarriaga, Phys. Rev. D **70**, 083005 (2004), arXiv:astro-ph/0408455.
[198] E. Komatsu, D. N. Spergel, and B. D. Wandelt, Astrophys. J. **634**, 14 (2005), arXiv:astro-ph/0305189.
[199] E. Komatsu, B. D. Wandelt, D. N. Spergel, A. J. Banday, and K. M. Górski, Astrophys. J. **566**, 19 (2002), arXiv:astro-ph/0107605.
[200] A. P. S. Yadav, E. Komatsu, and B. D. Wandelt, Astrophys. J. **664**, 680 (2007), arXiv:astro-ph/0701921.
[201] A. P. S. Yadav, E. Komatsu, B. D. Wandelt, M. Liguori, F. K. Hansen, and S. Matarrese, Astrophys. J. **678**, 578 (2008), arXiv:0711.4933.
[202] A. Curto, J. F. Macias-Perez, E. Martinez-Gonzalez, R. B. Barreiro, D. Santos, F. K. Hansen, M. Liguori, and S. Matarrese, ArXiv e-prints **0804.0136** (2008), 0804.0136.
[203] E. Komatsu, A. Kogut, M. R. Nolta, C. L. Bennett, M. Halpern, G. Hinshaw, N. Jarosik, M. Limon, S. S. Meyer, L. Page, et al., ApJS **148**, 119 (2003), astro-ph/0302223.
[204] A. P. S. Yadav and B. D. Wandelt, Physical Review Letters **100**, 181301 (2008).
[205] E. Martinez-Gonzalez, ArXiv e-prints **0805.4157** (2008), 0805.4157.
[206] P. Coles and B. Jones, MNRAS **248**, 1 (1991).
[207] R. Vio, P. Andreani, and W. Wamsteker, PASP **113**, 1009 (2001), arXiv:astro-ph/0105107.
[208] M. J. Rees and D. W. Sciama, Nature (London) **217**, 511 (1968).
[209] R. K. Sachs and A. M. Wolfe, Astrophys. J. **147**, 73 (1967).
[210] S. D. M. White, Reports of Progress in Physics **70**, 883 (2007), arXiv:0704.2291.