Reconstruction of signals with unknown spectra in information field theory with parameter uncertainty

Torsten A. Enßlin and Mona Frommert
Max-Planck-Institut für Astrophysik, Karl-Schwarzschild-Str. 1, 85741 Garching, Germany
(Dated: February 15, 2010)

The optimal reconstruction of cosmic metric perturbations and other signals requires knowledge of their power spectra and other parameters. If these are not known a priori, they have to be measured simultaneously from the same data used for the signal reconstruction. We formulate the general problem of signal inference in the presence of unknown parameters within the framework of information field theory. To solve this, we develop a generic parameter uncertainty renormalized estimation (PURE) technique. As a concrete application, we address the problem of reconstructing Gaussian signals with unknown power-spectrum with five different approaches: (i) separate maximum-a-posteriori (MAP) power spectrum measurement and subsequent reconstruction, (ii) MAP reconstruction with marginalized power-spectrum, (iii) maximizing the joint posterior of signal and spectrum, (iv) guessing the spectrum from the variance in the Wiener filter map, and (v) renormalization flow analysis of the field theoretical problem providing the PURE filter. In all cases, the reconstruction can be described or approximated as Wiener filter operations with assumed signal spectra derived from the data according to the same recipe, but with differing coefficients. All of these filters, except the renormalized one, exhibit a perception threshold in case of a Jeffreys prior for the unknown spectrum. Data modes, with variance below this threshold do not affect the signal reconstruction at all. Filter (iv) seems to be similar to the so called Karhunen-Loève and Feldman-Kaiser-Peacock estimators for galaxy power spectra used in cosmology, which therefore should also exhibit a marginal perception threshold if correctly implemented. We present statistical performance tests and show that the PURE filter is superior to the others, especially if the post-Wiener filter corrections are included or in case an additional scale-independent spectral smoothness prior can be adopted.

I. INTRODUCTION

A. The generic sensing problem

Reception of a signal is strongly aided by prior knowledge of the signals properties. This is especially true in low signal to noise (S/N) situations, in which proper knowledge can make the difference between recognition of a signal and blindness. Our human senses like vision and hearing are strongly enhanced by our knowledge on the possible signals present in the data-stream entering the human brain. The very same is true for signal reception by artificial sensor systems, since signal knowledge permits us to construct optimal filters, suppressing the noise as far as possible while focusing on the data modes with stronger S/N. If a sufficient training data is available, or theoretical reasoning permits us to predict signal properties, optimal filter design is possible and relatively straightforward.

However, there are situations, where such knowledge is not available, or is to be excluded on purpose from the analysis, in order to have a prejudice-free signal reconstruction. In such a situation the required parameters have to be measured simultaneously from the same data which is used for the signal reconstruction. Due to the interdependence of reconstructed signal and parameters, the problem becomes non-trivial and in general non-linear, even if the original inference problem was linear for fixed parameter values.

Let us provide a concrete example in cosmology. The cosmic matter distribution and its imprinted metric fluctuations on large scales can be well approximated to be a Gaussian random field obeying statistical isotropy and homogeneity. Knowledge of the power spectrum of these fields permits us to construct optimal and linear reconstruction filters for data of any linear tracers like the cosmic microwave background, the galaxy distribution (approximatively), or the gravitational lensing signature. For a set of cosmological parameters (e.g. Hubble constant, cosmic matter content, ...) these power spectra are known and can be used. However, the cosmological parameters themselves are not precisely known, and our best knowledge might come from the data-set we are analyzing. Furthermore, if we want to be open for non-standard cosmological scenarios, we might not want to put any prior assumption on the functional form of the power spectrum into our signal reconstruction problem.

Therefore, we need signal reconstruction methods, which are capable of dealing with uncertainties in the parameters of the problem. Such methods would be very useful in many situations, where prior knowledge on signal properties are absent or should be avoided. Some loss in fidelity compared to the case where these parameters are known can be expected, however, such methods can be expected to be flexible and robust due to their generic nature and self-tuning abilities.

For the problem of the reconstruction of the cosmic large-scale structure, the key parameter is the cosmic matter power-spectrum. It is known in the field of signal detection, that a statistical verification of the presence of a signal due to an increase in the data variance is possible.
well before the signal can be reconstructed itself. Thus, a measurement of the signal power-spectrum is already possible while the S/N-ratio is too low for map-making, and is therefore immediately available for filter optimization as soon as the critical S/N-ratio is achieved.

B. Derived filters

The signal reconstruction filters derived in this work can all be regarded as or approximated to an application of a data-dependent Wiener filter operator onto the data, which results in a non-linear transformation of the data. The Wiener filter construction requires the knowledge of the signal covariance, or spectrum, the instrument response and the noise covariance. The signal covariance has to be extracted from the data itself, and therefore introduces a data dependence into the filter. The five filters presented in this work differ in the way the assumed covariance is constructed, due to the different philosophies:

1. MAP spectrum filter: The maximum a posteriori (MAP) of the spectrum given the data should be a reasonable guess for the signal spectrum assumed in the Wiener reconstruction.

2. Classical map: The inference problem should be marginalized over all possible power spectra. In doing so, and deriving the classical filter equation by extremizing the resulting effective posterior, a data-dependent Wiener filter is derived, in which an effective spectrum emerges. This spectrum differs in general from the MAP-spectrum.

3. Joint MAP filter: Instead of marginalizing the joint posterior of signal and spectrum and then extremizing it with respect to one of those, we can maximize it with respect to both, leading to the joint MAP filter.

4. Critical filter: Symmetry properties of the three classical filters described above indicate that a fourth one could be interesting. This filter also results if one requires the covariance of the Wiener filter map to exhibit exactly its expected variance. It turns out that this filter marks the demarcation line between filter with perception threshold, as the above, and filter without the next one.

5. PURE filter: Extremizing the posterior does not provide the optimal reconstruction algorithm in the sense of minimizing the reconstruction error variance. This can only be done by a full field theoretical treatment which incorporates spectrum-uncertainty effects correcting for imbalances of the induced errors due to over- and underestimations of the signal spectrum. Here, we incorporate such a correction by virtue of an uncertainty-renormalization calculation. The resulting parameter uncertainty renormalized estimation (PURE) filter appears only to be a Wiener filter in case only an infinitesimal amount of uncertainty added. The renormalized-optimal spectrum as a fixed point of this uncertainty adding operation is different from the spectra of the other filter. In case a finite amount of uncertainty is added, the PURE filter contains corrections terms which can not be described as Wiener filtering.

C. Previous works

The PURE approach is derived within information field theory (IFT). This deals with the information of data on spatially distributed quantities, and is a statistical field theory. The connection of inference problems and statistical field theories was discovered independently by several authors in cosmology [1, 5, 9], statistical field theory [2–4, 15–20], and quantum mechanics [15–20]. An introduction into IFT can be found in [5].

The uncomfortable dependence of information theoretical methods on signal prior information have lead several authors to think about methods to extract this information at least partly from the data. For example a smoothness prior for the signal can be used, where an “optimal” value for the smoothness controlling parameter derives from the data itself [2]. The optimal smoothness constraint for an Gaussian signal is provided by its covariance, as known from Wiener filter theory [31]. A natural proposal is therefore to measure the power spectrum (or any characteristics of the signal covariance) from the data and to use this for Wiener filtering or other signal reconstruction methods [21–24, 28]. Data gaps complicate the power spectrum measurement step, but extensions of such methods to this case exist [27]. However, a more theoretical understanding of the inference problem and the assumptions implicitly made by these methods would be beneficial to answer several questions. How should the spectrum be measured optimally? How can spectral prior information be incorporated into the filter? And is the best spectral estimator really the best choice for the spectrum assumed in the ‘Wiener filter’?

Only Bayesian approaches, which are explicitly dealing with all relevant prior information, can answer these questions accurately. For example, it is possible to use the MAP approach to the problem of Wiener filtering if the overall amplitude of the signal covariance is unknown, even on a logarithmic scale [15]. For a white signal, where all pixels are statistically independent, this can be generalized to the case that all pixels amplitudes are drawn from a scale-free distribution function [20].

In precision cosmology, the problem of inferring the image and its power spectrum simultaneously is very prominent in cosmic microwave studies and cosmography of the large scale structure. It has been addressed rigorously via the Gibbs sampling scheme [6, 10, 11, 50]. Since
this approach samples the full joint posterior of maps and spectra, it provides the full solution to the problem. However, the computational costs of Gibbs sampling are high. Also obtaining analytical insights into the general behavior of the scheme is not trivial. Computationally cheaper and analytically simpler methods are therefore interesting and should be provided by this work.

D. Structure of the work

We introduce IFT with parameter uncertainties in Sect. II. The generic PURE approach is described in Sect. III where for any case with fourth order interaction terms the generic uncertainty renormalization flow equation is provided. In Sect. IV the problem of signal spectrum uncertainty is introduced, and the five mentioned filters are derived. The perception thresholds of the filters are investigated in Sect. V and their fidelity in Sect. VI, where also a PURE filter with spectral smoothness prior is presented. Finally, we conclude in Sect. VII.

II. INFORMATION FIELD THEORY WITH PARAMETER UNCERTAINTY

A. Information field theory

We briefly introduce the concepts of IFT and extend them to the case of parameter uncertainties. More details on terminology and notation of the framework can be found in [5]. An information field is simply a spatially extended signal, where a signal \( s(x) \) is any quantity a scientist might be interested in measuring. We treat the signal \( s(x) = s_x \), a function of a spatial coordinate \( x \), as an abstract vector in Hilbert space with the scalar product \( j^* s = \int dx j(x) s(x) \). In the simplest case of the theory, signal and noise are independent Gaussian random variables, and the data depend linearly on them. This so-called free theory can be treated analytically and is our starting point.

B. Free theory from a Gaussian data model

We assume that the signal we want to reconstruct is a Gaussian random field, with a probability distribution prior to any measurement described as

\[
P(s|p) = \mathcal{G}(s, S_p) \equiv \frac{1}{2\pi S_p} \exp \left( -\frac{1}{2} s^\dagger S_p^{-1} s \right),
\]

where \( S_p = \langle ss^\dagger \rangle_{(s|p)} \) is the signal covariance given the parameter \( p \), which itself might be a vector or even a field over some space. The subscript \( (s|p) \) on the brackets of the expectation value indicate that the average should be done over the probability distribution \( P(s|p) \).

We further assume that the signal is processed by a linear measurement device with response matrix \( R \) and additive noise \( n \) according to:

\[
d = Rs + n.
\]

In general, response and noise can also depend on unknown parameters and the general theory developed here can also be applied to that case. To focus the discussion, we only consider here the concrete example of a parameter dependent signal covariance, and assume the response and noise statistics to be known. We assume the noise to be signal-independent and Gaussian, and thus

\[
P(n|s, p) = \mathcal{G}(n, N),
\]

where \( N = \langle nn^\dagger \rangle_{(n)} \) is the noise covariance matrix. Since the noise is just the difference of the data to the signal-response, \( n = d - Rs \), the likelihood of the data is

\[
P(d|s, p) = P(n = d - Rs|s, p) = \mathcal{G}(d - Rs, N).
\]

The information Hamiltonian as defined in [5] is the negative logarithm of the joint probability function of signal and data for given and fixed parameters:

\[
H_p[s] = -\log P(d, s|p) = -\log \left[ P(d|s, p) P(s|p) \right].
\]

Thus the Hamiltonian of the Gaussian theory,

\[
H_p^G = \frac{1}{2} s^\dagger D_p^{-1} s - j^\dagger s + H_{0,p}^G,
\]

is only quadratic in the signal, and therefore corresponds to a free field theory. Here

\[
D_p = \left[ S_p^{-1} + M \right]^{-1}, \quad \text{with} \quad M = R^\dagger N^{-1} R,
\]

is the information propagator, which depends on the unknown spectral parameters. The information source,

\[
j = R^\dagger N^{-1} d,
\]

depends linearly on the data in a response-over-noise weighted fashion. Finally,

\[
H_{0,p}^G = \frac{1}{2} d^\dagger N^{-1} d + \frac{1}{2} \log \left( \frac{1}{2\pi} \right) + \frac{1}{2} \log \left( \frac{1}{2\pi N} \right)
\]

absorbs all \( s \)-independent normalization constants. It can not be ignored here, since it depends on \( p \).

The key quantity, from which all relevant moments of the signal can be estimated, is the partition function,

\[
Z_p[J] = \int Ds e^{-H_p[s] + J^\dagger s}.
\]

For the free field theory the partition function is

\[
Z_p^G[J] = \sqrt{\frac{2\pi}{|D_p|}} \exp \left\{ \frac{1}{2} (J + j)^\dagger D_p (J + j) - H_{0,p}^G \right\}.
\]

\[1,2,3,4,5\]
This explicit formula permits us to calculate the expectation of the signal given the data (and the parameter), in the following called the map $m_p$:

$$m_p = \langle s \rangle_{(s,d,p)} = \frac{\delta \log Z_p^d}{\delta J} \bigg|_{j=0} = D_p j$$

$$= \left[ S_p^{-1} + R^t N^{-1} R \right]^{-1} R^t N^{-1} d. \quad (12)$$

The last expression shows that the map is given by the data after applying a generalized Wiener filter, $m_p = F_p d$, which depends on the parameter $p$ of the signal covariance.

Similarly, the quadratic uncertainty of the signal map can be worked out. It turns out that for a free theory it is the propagator itself

$$\langle (s - m_p)(s - m_p)^\dagger \rangle_{(s,d,p)} = \langle s s^\dagger \rangle_{(s,d,p)} - m_p m_p^\dagger = D_p$$

The first identity follows from

$$\langle s m_p^\dagger \rangle_{(s,d,p)} = \langle s^\dagger m_p \rangle_{(s,d,p)} = m_p m_p^\dagger$$

due to the fact, that the reconstructed map $m_p$ is solely determined by the data, and therefore given in this average. The second identity holds due to the identity of the connected correlation function and the propagator,

$$\langle s s^\dagger \rangle_{(s,d,p)} = \delta^2 \log Z_p^d / \delta Z_p^d \bigg|_{j=0} = D_p.$$  

### C. Classical field theory

In case of the free theory, the map, Eq. 12 would also be obtained from a classical treatment of the Hamiltonian by extremizing it:

$$\frac{\delta H_p[s]}{\delta s} = 0. \quad (14)$$

For a Hamiltonian with interaction terms the classical field (in field-theoretical language) or MAP estimator (in signal processing language) is a useful approximation to the correct expectation value. The inverse Hessian in the signal Hilbert space around this map,

$$\left( \frac{\delta^2 H_p[s]}{\delta s \delta s^\dagger} \right)^{-1}, \quad (15)$$

characterizes the uncertainty. For the free theory, this is the propagator, as given by Eq. 13.

The identity of fully field theoretical and classical results holds only for the case of a free theory. However, the latter is often an acceptable approximation to the former, while much easier to derive. Therefore, we will make also use of the classical approximation in the following.

### D. Parameter uncertainty and posterior

In many applications, there are parameters specifying the likelihood and prior, and thereby the coefficients of the Hamiltonian, which are not precisely known. These parameters, in the following denoted by the abstract vector $p$, are either to be determined from the data, to be marginalized over, or to be simultaneously determined with the signal.

In such a case we have to construct the joint posterior of the signal and the parameter given the data. This is given according to Bayes’ theorem as

$$P(s,p|d) = \frac{P(d,s,p)}{P(d)} = P(s|d,p) \frac{P(d|p)}{P(p)}, \quad (16)$$

where we had to introduce the parameter prior $P(p)$.

The last expression contains a Bayes factor $P(d|p)/P(d)$, the ratio of the evidence of data for a specific parameter set to that of the model at all. Thus, the joint posterior is weighted towards model-parameters for which the data provide larger evidence in addition to any prior-weighting.

The definition of the Hamiltonian for fixed-parameters as $H_p[s] = - \log P(d,s|p)$ permits us to construct the joint partition function

$$Z[J,K] = \int dp \int d s P(s,p|d) P(d) e^{J s + K s^\dagger}$$

$$= \int dp P(p) e^{K s^\dagger} \int d s P(d|s,p) P(s|p) e^{J s},$$

which is built upon $Z_p[J]$, the partition function of the theory for given parameters $p$. The information field estimators, marginalized for the unknown parameters, is then simply given by

$$\langle s \rangle_d = \frac{\delta \log Z[J,K]}{\delta J} \bigg|_{J,K=0} = \frac{1}{Z} \int dp P(p) \frac{\delta Z_p[J]}{\delta J} \bigg|_{J=0} \quad (18)$$

$$= \int dp P(p) \frac{Z_p}{Z} \langle s \rangle_{(s,d,p)}.$$  

The effective signal estimator (marginalized for unknown parameters) is not just the signal estimator multiplied by the parameter prior $P(p)$, but is additionally weighted by a parameter likelihood factor $P(d|p) = Z_p$, so that the parameter-dependent signal means are averaged over the parameter posterior $P(p|d)$. Therefore, parameter values which are especially compatible with the data get automatically a larger weight for the signal estimator, as recognized before.

### E. Effective marginalized Hamiltonian

If a parameter-dependent Hamiltonian $H_p[s] = - \log P(d,s|p)$ describes the conditional probability of
the signal and data given the parameters, an effective, parameter-marginalized Hamiltonian $H[s]$ is defined by

$$e^{-H[s]} = \int dp \, P(d, s, p) P(p) = \int dp e^{-H_p[s] - E_p},$$

(19)

with $E_p = -\log P(p)$ the parameter-prior-energy. It is crucial, that $H_p[s]$ obeys the correct normalization condition, $\int Dd \int Ds \exp(-H_p[s]) = 1$, otherwise a hidden prior on $p$ may enter the calculation.

In many cases, an analytical calculation of the effective Hamiltonian will be out of reach. Since the perturbative field theoretical treatment requires a polynomial representation anyway, it is often easier to obtain the coefficients of the effective Hamiltonian separately by Taylor-Frechét expansion around a reference field configuration $t$, so that $s = t + \phi$. The Hamiltonian for $\phi$ is then

$$H[\phi] = H_0 - j^\dagger \phi + \frac{1}{2} \phi^\dagger D^{-1} \phi + \sum_{n=3}^{\infty} \frac{1}{n!} \Lambda^{(n)}_{x_1 \ldots x_n} \phi_{x_1} \ldots \phi_{x_n},$$

with $H_0$ being the effective Hamiltonian around a reference field configuration $t$.

III. UNCERTAINTY RENORMALIZATION

A. General remarks

The expansion of the effective Hamiltonian around a reference field is expected to work best when the parameter prior is well localized around a specific value. The effective Hamiltonian will then be close to the original, parameter-dependent one for this parameter value. In case the original theory was free, the effective theory will have only small interaction terms. Diagrammatic expansions can then be conducted and truncated at low order.

Unfortunately, in many practical applications, the uncertainties of the parameters are substantial, and not described by a well localized prior. In this case it might be possible to construct the effective Hamiltonian by repeatedly adding smaller portions of parameter-uncertainty, with each uncertainty dose so small that the resulting Hamiltonian has only weak interactions, which can be re-absorbed into renormalized, effective propagator and data source terms. The accumulated uncertainty can thereby become large and equal to the required amount of entropy for the unknown parameter of the theory. In the following we will explain the basics of this uncertainty renormalization flow.

B. Parameter uncertainty renormalization

A broad prior for a parameter $p$ may be decomposed into a number $N$ of narrow and mutually independent priors for some auxiliary variables $\tau_j$ (with $j \in \{1, \ldots, N\}$):

$$P(p) = \left( \prod_{j=1}^{N} \int d\tau_j P(\tau_j) \right) \delta(p - \sum_{j=0}^{N} \tau_j).$$

(22)

We have chosen here the parameter to be the sum of the auxiliary variables for definiteness and simplicity, but other relations can be worked out in a similar way or be mapped onto this case. Also the mutual independence of the auxiliary variable is mostly a technical convenience and not a strict requirement. Note, that we have included a starting parameter value of $\tau_0$ into the sum. Since it
would be convenient to identify this with the prior expectation value \( \langle p \rangle_p \) throughout the full renormalization procedure we require

\[
\langle \tau_j \rangle_j = \delta_{j0} p_0,
\]

with \( p_0 = \langle p \rangle_p \). We further introduce the \( l \)-th parameter residual as \( r_l = \tau_0 + \sum_{j=l+1}^N \tau_j \), so that \( r_0 = p \) and \( r_N = \tau_0 = p_0 \). The effective Hamiltonian can now be expressed as

\[
e^{-H[s]} = \int dp \, P(p) \, e^{-H_p[s]}
\]

\[
= \int dp \left( \prod_{j=1}^N \int \tau_j \, P(\tau_j) \right) \delta(p - \sum_{j=0}^N \tau_j) \, e^{-H_p[s]}
\]

\[
= \prod_{j=1}^N \int \tau_j \, P(\tau_j) \, e^{-H_{r\tau}[s]}
\]

\[
= \prod_{j=2}^N \int \tau_j \, P(\tau_j) \int \tau_1 \, P(\tau_1) \, e^{-H_{r\tau+r}[s]} = e^{-H^{(1)}_{r\tau}[s]}
\]

\[
= \prod_{j=3}^N \int \tau_j \, P(\tau_j) \int \tau_2 \, P(\tau_2) \, e^{-H^{(2)}_{r\tau+r}[s]} = e^{-H^{(2)}_{r\tau}[s]}
\]

\[
= \cdots
\]

\[
= \int \tau N \, P(\tau N) \, e^{-H^{(N)}_{r\tau+r}[s]} = e^{-H^{(N)}_{r\tau}[s]},
\]

where \( H^{(N)}_{r\tau}[s] = H[s] \). This means that a series of effective Hamiltonians with increasing accumulated parameter uncertainty is defined, and an uncertainty adding operator:

\[
H^{(n)}_{r\tau} \rightarrow H^{(n+1)}_{r\tau} \equiv - \log \int d\tau_n \, P(\tau_n) \, e^{-H^{(n)}_{r\tau},r\tau}[s].
\]

Note that \( H^{(0)}_{r\tau} = H_{r\tau} = H \). To each Hamiltonian a time-like variable \( t \) can be assigned, which measures the amount of uncertainty accumulated so far. A suitable variable is the accumulated uncertainty dispersion,

\[
t_n = \sum_{j=1}^n \sigma_{\tau_j}^2 = \sum_{j=1}^n \left( \langle \tau_j^2 \rangle_j - \langle \tau_j \rangle_j^2 \right) = \sum_{j=1}^n \langle \tau_j^2 \rangle_j - \langle \tau_j \rangle_j^2,
\]

where we used Eq. \[23\] In case all auxiliary variables have the same prior, we find \( t_n = n t_1 = n \frac{\langle (p - p_0)^2 \rangle_p}{p_0} \).

At each time-step a renormalization of the Hamiltonian can be done, in which it is cast back into the structure it had before, just with modified coefficients (propagator, source and interaction terms). A renormalization flow can further be established by letting the individual time-steps of size \( t_1 \) become infinitesimally small, however, their number \( N \) infinitely large, while keeping the total added uncertainty constant, \( t = N t_1 \). The result are the renormalization flow equations for the coefficients of the Hamiltonian. The actual form of these equations depends on the Hamiltonian and is much simpler if the Hamiltonian has less interactions.

Therefore, even a free Hamiltonian as in Eq. \[6\] should be further simplified by suppressing the linear term \( j^p_0 \) as far as possible. This is done for the value of \( p = p_0 \), which is our starting point in parameter space, by changing to a new field variable \( \phi = s - m_0 \) with \( m_0 \equiv D_0 j_0, D_0 \equiv D_{p_0} \), and \( j_0 \equiv j_0 \). The Hamiltonian reads now

\[
H_p'[\phi] = \frac{1}{2} \phi^\dagger D^{-1}_p \phi - j^p_0 \phi + H_{0,p} \equiv \frac{1}{2} \phi^\dagger D^{-1}_p \phi + H_{0,p},
\]

with \( \phi = s - m_0 \),

\[
j^p_0 = j_p - D^{-1}_p m_0 = j_p - D^{-1}_p D_0 j_0, \quad \text{and} \quad (27)
\]

\[
H_{0,p} = H_{0,p} + \frac{1}{2} m_0 D^{-1}_p m_0 - j^p_0 m_0
\]

and is especially simple for \( p = p_0 \), since then \( j^0_0 = 0 \). Now, the effective Hamiltonian is calculated and expanded according to the recipe in Sect. \[11\] for a parameter prior well localized on \( p = p_0 \). The localization of the prior is typically characterized by a small parameter \( \delta t = \sigma^2 \), which also appears as a pre-factor of the various coefficients of the effective Hamiltonian.

### C. 4th order interactions

In order to perform the renormalization step, let us be a bit more specific about the effective Hamiltonian for definiteness. By virtue of our foresight on the calculations in Sect. \[IV\] we assume that up to linear order in \( \delta t \) the effective Hamiltonian is given by

\[
H_p'[\phi] = H_0' + \Lambda^{(1)} \phi + \frac{1}{2} \phi^\dagger \left( D_0^{-1} + \Lambda^{(2)} \right) \phi
\]

\[
+ \frac{1}{3!} \Lambda^{(3)} [\phi, \phi, \phi] + \frac{1}{4!} \Lambda^{(4)} [\phi, \phi, \phi, \phi] + O(\delta t^2),
\]

with \( \Lambda^{(1)}, \Lambda^{(2)}, \Lambda^{(3)}, \Lambda^{(4)} \) being of order \( O(\delta t) \), since our originally free Hamiltonian, Eq. \[27\] should be recovered in the limit of vanishing parameter uncertainty, \( \delta t \rightarrow 0 \). All higher order interaction terms are of higher order in \( \delta t \) and therefore ignored in the following. For our later convenience we introduce

\[
\lambda^{(n)} = \lim_{\delta t \rightarrow 0} \frac{\Lambda^{(n)} \delta t}{\delta t}.
\]

Now, we can renormalize by absorbing all diagrams of order \( O(\delta t) \) into renormalized propagator and source terms. Since \( f' \) in Eq. \[27\] is already of order \( \delta t \) and only the three- and four-leg vertices have contributions of order \( O(\delta t) \), only uncertainty loop corrections have to be
The inverse renormalized propagator up to first order is

\[ \delta t D^{-1} = \lambda^{(2)} + \frac{1}{2} \lambda^{(4)} [\cdot, D_t] , \]  

which takes the dominant uncertainty-loop correction into account. We dropped the subscript at \( D_0 \) and use the Feynman rules provided in [5]. The renormalized propagator up to linear order in \( \delta t \) is

\[
D_\# = D - D \Lambda^{(2)} D - \frac{1}{2} D \Lambda^{(4)} [\cdot, D, \cdot] D 
\]

\[ = D_{xy} - D_{zz} \Lambda^{(2)} D_{zz} D_{yy} - \frac{1}{2} D_{xx} \Lambda^{(4)} D_{zz} D_{yy} . \]

The inverse renormalized propagator up to first order is

\[
D^{-1}_\# = D^{-1} + \Lambda^{(2)} + \frac{1}{2} \Lambda^{(4)} [\cdot, D_t, \cdot] .
\]

These coefficients now define a renormalized effective Hamiltonian, \( H_\# [\phi] = \frac{1}{2} \phi^\dagger D^{-1}_\# \phi - j_\# \phi + H_{\#0} \), which belongs to a free theory, and is similar to \( H'[\phi] \), in that it has the same mean and uncertainty dispersion by construction. Higher order uncertainty correlations differ certainly, due to the approximation of the renormalization step. In contrast to the original Hamiltonian \( H_{\rho_0} [s] \) in Eq. 6, which was also free, the renormalized Hamiltonian has some amount of parameter uncertainty corrections included.

Now, the original field \( s = m_0 + \phi \) can be restored, leading to a free Hamiltonian with

\[ D_{t+\delta t} = D_\# , \]  

\[ j_{t+\delta t} = j_\# + D^{-1}_\# m_t , \]

where the subscript \( t + \delta t \) indicates that the parameter uncertainty is increased by \( \delta t \) from its original value of \( t \). Since \( \delta t \) can be made arbitrarily small, a system of differential equations can be derived,

\[
\frac{dD_t}{dt} = \lim_{\delta t \to 0} \frac{D_{t+\delta t} - D_t}{\delta t} = -D_t \lambda^{(2)} D_t - \frac{1}{2} D_t \lambda^{(4)} [\cdot, D_t, \cdot] D_t , \]

\[ \frac{dj_t}{dt} = \lim_{\delta t \to 0} \frac{j_{t+\delta t} - j_t}{\delta t} = -\lambda^{(1)} + \lambda^{(2)} D_t j_t - \frac{1}{2} \lambda^{(4)} [\cdot, D_t] + \frac{1}{2} \lambda^{(4)} [\cdot, D_t, j_t] , \]

which form the uncertainty renormalization flow equations. The pseudo-time \( t \) measures the accumulated dispersion of the resulting prior probability. These equations can be transformed into the more compact form

\[
\frac{dD_t}{dt} = \lambda^{(2)} + \frac{1}{2} \lambda^{(4)} [\cdot, D_t, \cdot] , \]

\[
\frac{dm_t}{dt} = -D_t \lambda^{(1)} - \frac{1}{2} D_t \lambda^{(3)} [\cdot, D_t] , \]

where \( m_t = D_t j_t \). The renormalization equations so far are evolution equations for operators. If they should become ordinary partial differential equations some sort of closure is required, in that the renormalized Hamiltonian has to be of its original structure so that it is always clear which terms are affected by the parameter uncertainty adding operation. Ideally, the change in the Hamiltonian can be mapped onto changes of effective parameter values.

The effective prior for the uncertain parameter can be expected to be a Gaussian, due to the central limit theorem of statistics,

\[ P(p) = \mathcal{G}(p - p_0, t) . \]

## IV. SIGNAL SPECTRUM UNCERTAINTY

### A. Spectrum parameterization

The signal covariance \( \langle S_p \rangle_{x,y} = \langle s_x s_y \rangle_{s[p]} \) may exhibit any dependence on the spatial coordinates as long as the matrix is symmetric and positive definite. In the cosmological relevant case of translationally and rotationally invariant signal statistics, the signal covariance is fully characterized by its power spectrum. This means, there is an orthonormal basis \( O \) of the signal Hilbert space which diagonalizes \( S_p \):

\[
(O S_p O^\dagger)_{k,q} \equiv \langle s_k s_q \rangle_{s[p]} D_{k,q} = 1_{k,q} P_{S_p} (k) ,
\]

with \( 1_{k,q} \) the identity in the transformed basis, \( P_{S_p} (k) \) the power-spectrum, and using Einstein sum convention. In case we are dealing with a signal over a \( d \)-dimensional Cartesian space, \( \langle s_k \rangle = \exp (i k \cdot x) \) is simply a Fourier transformation and the Fourier space identity is \( \delta(k - q) = (2 \pi)^d \delta(k - q) \), provided the scalar product in Fourier space is adopted as \( a^\dagger b = (2 \pi)^{-d} \int dk a(k) b(k) \). However, since the theory should also be applicable in curved spaces like the sphere, or even in spaces without translational invariance, we formulate it in an abstract way and just assume that the basis \( O \) diagonalizes the signal covariance, which is always possible.

In general, the signal covariance \( S_p \) may also exhibit any dependence on the unknown parameter \( p \) of the problem, as the power spectrum in cosmology is a complicated function of the cosmological parameters. However, in order not to depend on a specific model, we model the
power spectrum as being a linear combination of a number of positive basis functions \( f_i(k) \) with disjunct supports (the spectral bands) with respect to the basis \( O_{kx} \), so that
\[
P_{S_p}(k) = \sum_i p_i f_i(k) \quad (38)
\]
is positive for all \( k \) (all coefficients of \( p = (p_i)_i \) are positive and the spectral bands cover the full \( k \)-space domain). We define
\[
(S_i)_{xy} = \mathcal{O}_{kx} f_i(k) O_{ky} \quad (39)
\]
and therefore have
\[
S_p = \sum_i p_i S_i. \quad (40)
\]
Since we also need the inverse of the covariance matrix we further define
\[
g_i(k) = \begin{cases} 1/f_i(k) & f_i(k) > 0 \\ 0 & f_i(k) = 0 \end{cases} \quad (41)
\]
and the pseudo-inverse of the band-variances,
\[
S_i^{-1} = \mathcal{O}_{kx} g_i(k) O_{ky}, \quad (42)
\]
so that
\[
S_p^{-1} = \sum_i p_i^{-1} S_i^{-1}, \quad (43)
\]
is the inverse of \( S_p \), as one can easily verify.

B. Spectral prior and joint Hamiltonian

For definiteness, we assume that the individual signal-band amplitudes \( p_i \) have independent prior distributions,
\[
P(p) = \prod_i P(p_i), \quad (44)
\]
with the individual priors being power-laws with exponential low amplitude cutoff at \( q_i \):
\[
P(p_i) = \frac{1}{q_i} \Gamma(\alpha_i - 1) \left( \frac{p_i}{q_i} \right)^{-\alpha_i} \exp\left(-\frac{q_i}{p_i}\right). \quad (45)
\]
For \( \alpha_i > 1 \) this is an informative prior, where \( q_i/\alpha_i \) determines the preferred value. A non-informative prior would be given by Jeffreys prior with \( \alpha_i = 1 \) and \( q_i = 0. \)

The joint Hamiltonian is therefore
\[
H[s, p] = H_p[s] + E(p) \quad (46)
\]
with the parameter prior energy
\[
E(p) = \sum_i \left[ \frac{q_i}{p_i} + \alpha_i \log\left( \frac{p_i}{q_i} \right) + \log(q_i \Gamma(\alpha_i - 1)) \right]. \quad (47)
\]

\[\text{Since this would result in an improperly normalized prior, we understand this as } \alpha_i = 1 + \epsilon, q_i = \epsilon, \text{ and } \lim_{\epsilon \to 0} \text{ at the end of the calculation.}\]

---

Figure 1: Parameter \( \delta_i \) and \( \epsilon_i \) of the five different filters. The critical line between filter with and without perception threshold as given by Eq. [60] is also shown.

C. Generic filter formula

In the following, we derive five approximate filters. It turns out that they all can be cast into a single set of determining equations, with different coefficients. These filters can be expressed as Wiener filters for some specific spectrum \( S_{p^*} = \sum_i p_i^* S_i \), with different spectral parameters \( p^* \). The signal map and the spectrum assumed for its construction have to be calculated self-consistently from
\[
\begin{align*}
m_{p^*} &= D_{p^*} j, \quad \text{and} \\
p_i^* &= \frac{1}{\gamma_i + \delta_i} \left( q_i + \frac{1}{2} \text{Tr}[S_{p^*}^{-1} S_i S_{p^*}^{-1}] \right),
\end{align*}
\]
for example by simply iterating these two equations.

Here, the filter-specific parameters are \( \epsilon_i, \delta_i, \text{ and } \gamma_i = \alpha_i - 1 + g_i/2, \) where \( g_i = \text{Tr}[S_i^{-1} S_i] \) is the number of degrees of freedom of the \( i \)th spectral band. In order to simplify notation, we drop in the following the \( * \) from \( p^* \), assuming that the context makes it clear if we talk about the unknown parameter \( p \) or a parameter choice \( p^* \) for a specific filter.

In case a Jeffreys prior is adopted (\( q_i = 0 \) and \( \alpha_i = 1 \)) the trivial filter \( m(d) = 0 \) would be the preferred solution in all cases. However, since Jeffreys prior is an improper prior which is convenient to represents the class of very broad, but proper priors, we should not hesitate to remove the trivial filter solution by hand. Otherwise we would need to enter the discussion about an appropriate informative prior, which we like to avoid for simplicity since this can not be decided generically, but only for any concrete inference problems individually.

The parameters of the filters described in Sect. [13] and derived in the next few subsections are summarized in Fig. 1.
D. The critical filter

A good reference point seems to be the critical filter, which we derive first assuming for simplicity a Jeffreys prior \((\alpha_i = 1, q_i = 0)\), so that all spectral normalization are equally plausible on a logarithmic scale. Thus, the data alone should inform us about the spectrum. For this filter we require that the expectation value for the map variance in case we know \(p^* = p\) is a reliable proxy of the signal variance. From

\[
\langle m_p m_p \rangle_{d,s} = S_p - D_p, \tag{49}
\]

as can be derived from Eq. 13 or follows from a Maximum Likelihood calculation, we find the parameters \((\delta, \varepsilon) = (1, 0)\) by applying \(\text{Tr}[x S^{-1}_t]\), and by ignoring the average over \(P(d,s|p)\). The name critical filter should become clear in Sect. [V]. There, we show that at least in cases where the different spectral parameters are independent of each other, the different filters can be cast into two classes, such with and such without perception threshold. The critical filter marks the demarcation line between these phases.

E. Joint MAP filter

Extremizing the joint Hamiltonian, Eq. 40 with respect to \(p\) and \(s\) yields the joint MAP filter parameters \((\delta_i, \varepsilon_i) = (0, 1)\). We note, that if we extremize with respect to the log-spectral amplitudes \(\tau_i = \log p_i\), the parameters \((\delta_i, \varepsilon_i) = (0, 0)\) would have resulted due to the effect of the Jacobian of the prior transformation. This latter filter is identical to the classical one derived below in Sect. [IV.C].

F. MAP spectrum filter

Marginalizing the joint Hamiltonian Eq. 40 over the signal space provides the spectrum Hamiltonian

\[
H[p] = -\log(P(d,p)) - \log(P(d|p) P(p))
= \frac{1}{2} \log [1 + Q_p] - \frac{1}{2} d^t D_p S_p^{-1} H_0' + H_0
+ \sum_i \left( \frac{q_i}{p_i} + \alpha_i \log \left( \frac{p_i}{q_i} \right) \right), \tag{50}
\]

with \(D_p = (S_p^{-1} + M)^{-1}, Q_p = S_p M, \) and

\[
H_0' = \frac{1}{2} \log |N| + \frac{1}{2} d^t N d + \sum_i q_i \Gamma(\alpha_i - 1).
\]

Here we used Eq. 11 for \(P(d|p)\). A data-space view on this likelihood is given in Appendix A. Extremizing \(H[p]\) with respect to \(p_i\) and sorting for terms linear in it provides the MAP-spectrum parameter \((\delta_i, \varepsilon_i) = (1, 1)\). Again, if we extremize with respect to the parameters \(\tau_i = \log p_i\), we would get \((\delta_i, \varepsilon_i) = (1, 0)\), the parameters of the critical filter.

G. Classical map estimator

The effective, parameter marginalized signal Hamiltonian (Eq. 19) can be calculated analytically:

\[
H[s] = \frac{1}{2} s^t M s - j^*_t s + \sum_i \gamma_i \log \left( q_i + \frac{1}{2} s^t S_i^{-1} s \right)
+ H_0, \quad \text{with}
H_0 = \frac{1}{2} d^t N^{-1} d + \frac{1}{2} \log \left( \frac{1}{|2 \pi N|} \right)
- \log \left( \prod_i \frac{\Gamma(\gamma_i)}{\Gamma(\alpha_i - 1)} \frac{q_i^{-1}}{|2 \pi S_i|} \right). \tag{51}
\]

The classical mapping equation results from extremizing this Hamiltonian and is provided by Eq. 19 for \((\delta_i, \varepsilon_i) = (0, 1)\). In case of a single independent data and signal point, the effective Hamiltonian is an one dimensional function in signal space and is shown in Fig. [2].

H. Lognormal spectrum prior and parameter uncertainty renormalization

Now we want to calculate the PURE filter. In order to use the uncertainty renormalization method described
in Sect. [11] we need some additive auxiliary random variables, which each should have an unbiased distribution with zero mean according to Eq. [23]. For the moment, we concentrate on a single spectral parameter $p_i$ and change to the parameter variable $\tau_i = \log p_i$, which can be split up into additive auxiliary variables: $\tau_i = \sum_j \tau_{ij}$. For convenience we assume $\rho_{ij} = e^{\gamma_j}$ to be distributed according to Eq. [45] with properly chosen parameters $\alpha_{ij}$, and $q_{ij}$, as detailed in the Appendix [3]. There, it is shown that

$$P(\tau_i) \rightarrow G(\tau_i, t_i)$$  \hspace{1cm} (52)

for the limit of an infinite number of auxiliary parameters, with a finite total uncertainty dispersion of $t_i = (\tau_i^2(\tau_i) - \tau_i^2(\tau_i))$, as expected from the central limit theorem of statistics. The resulting statistics for $p_i = e^{\gamma_i}$ is therefore log-normal. If we take the limit $t_i \rightarrow \infty$ we obtain Jeffreys prior, which is flat on a logarithmic scale, and which conveniently permits us to compare the PURE filter to the others.

In the following we assume that all spectral coefficients receive uncertainty with the same infinitesimal rate, so that the prior distributions in Eq. [45] are all the same and narrowly centered on $p_i = 1$, which implies $\delta t_i = 1/(\alpha_i - 1) = \delta t$ and $q_i = \alpha_i - 3/2 = \delta t^{-1} - 1/2$ (see Appendix [3]).

Expanding the Hamiltonian in Eq. [51] around the reference map $m = D j$ recovers the original free Hamiltonian, shifted to $\phi = s - m$, and perturbed by some additional interaction terms $\Lambda(n) = \delta t \lambda(n) + O(\delta t^2)$ with

$$\lambda^{(1)} = \frac{1}{2} \left[ \left( \rho_i + 1 - p^{-1} m^i S^{-1} s \right) S^{-1} p^{-1} m, \right.$$ 
$$\lambda^{(2)} = \frac{1}{2} \left( \rho_i + 1 - p^{-1} m^i S^{-1} s \right) S^{-1} p^{-1} s - S^{-1} m m^i S^{-1} p^{-1}, \right.$$ 
$$\lambda^{(3)} = \lambda^{(4)} = \lambda^{(5)} = \lambda^{(6)} = 0 \quad \text{for } n > 4. \hspace{1cm} (53)$$

Here we have reinserted $p_i$ in order to have variables which capture the evolution of the renormalization flow dynamics. The renormalization flow equations are given by inserting the latter terms into two independent equations out of Eqs. [34]-[36]

$$\frac{dD^{-1}}{dt} = \sum_i \frac{1}{2} ((1 + q_i)p_i - \text{Tr}[B_i]) S^{-1} - S^{-1} B_i \right) p_i^{-2}$$
$$\frac{dj}{dt} = - \sum_i p_i^{-2}(m^i S^{-1} s) S^{-1} m, \quad \text{with} \hspace{1cm} (54)$$

$$B_i = (m m^i + D) S^{-1} \quad \text{and} \quad m = D j.$$ 

The system is not fully determined yet, since the evolution equations for the $p_i$'s are needed. Since $\frac{d}{dt} D^{-1} = \sum_i S_i^{-1} \frac{dp_i}{dt} p_i^{-1} + \frac{d}{dt} M$ contains the parameter evolution one has to specify how to split the evolution equation of the inverse propagator.

A natural way is to require all terms of the rhs of Eq. [54] which are parallel to the inverse signal covariance bands, to contribute to their evolution, and the ones which are orthogonal, to contribute to the evolution of $M$. The part of a matrix $A$ parallel to $S_i$ is obtained by the projector

$$P_i A \equiv \frac{1}{\delta t} \text{Tr}[A S_i] S_i^{-1}$$  \hspace{1cm} (55)

and the orthogonal part by $(1 - P_i) A$. Splitting the evolution equation this way yields

$$\frac{dp_i}{dt} = \beta_i p_i, \quad \text{or} \quad \frac{dp_i}{dt} = \beta_i, \quad \text{and} \hspace{1cm} (56)$$
$$\frac{dM}{dt} = \sum_i p_i^{-2} S_i^{-1} \left( \frac{1}{\rho_i} \text{Tr}[B_i] - B_i \right), \quad \text{with}$$

$$\beta_i = \left( \frac{1}{2} + \frac{1}{\rho_i} \right) \text{Tr}[B_i] p_i^{-1} - 1 + \frac{1}{\rho_i}. \hspace{2cm}$$

With this, the fastest evolution is assigned to the signal strength, whereas the inverse noise term evolves on much longer time-scales for $\rho_i \gg 1$. Actually, $M$ evolves only in directions orthogonal to all $S_i$, since

$$\frac{d}{dt}(P_i, M) = 0,$$  \hspace{1cm} (57)

meaning that the power within the spectral bands of $M$ gets only reshuffled, but is conserved. This implies that the evolution of $M$ interferes very little with the spectral evolution, since all changes to $M$ happen in directions which are projected out for $S_p$. The reverse is not true, since $M$ couples to the value of $p$. For an accurate reconstruction the evolution of $M$ needs to be followed, since it determines $D$ and thereby $m = D j$. However, we focus now only on the signal spectrum evolution and ignore the slow and perpendicular $M$ evolution.

The evolution equation for $p$ and $j$ have to be solved simultaneously. The resulting filter for $t \rightarrow \infty$ (which implies a Jeffreys prior) seems to be trivial, since $j \rightarrow 0$ and therefore $m_p \rightarrow 0$ in this limit. We remove this trivial solution by imposing $dj/dt = 0$ instead of Eq. [54]. This should be understood as looking for a stationary point of the $p$-evolution alone. Thus, we are asking for the unique spectrum, which taken as a sharp prior would remain unchanged if a small amount of spectral uncertainty is added. This fix point is given by $\beta_i = 0$ and therefore

$$p_i = \frac{1 + \frac{2}{\rho_i}}{\rho_i + 1} \text{Tr}[B_i]. \hspace{2cm} (58)$$

Although we have derived this filter only for Jeffreys prior, it is quite plausible to assume that the general spectrum formula, Eq. [10] with $(\delta_i, \varepsilon_i) = (1, -0.5/(1+2/\rho_i))$ also holds for $(\alpha_i, q_i) \neq (1, 0)$. We leave a formal proof of this for future work.
V. PERCEPTION THRESHOLD

A. Critical perception

In case of Jeffreys prior ($q_i = 0$, $\alpha_i = 1$, and $t = \infty$), the spectra used by some of our filters are only non-zero for spectral bands with a data variance above some threshold. Bands with less band power are fully suppressed. A perception threshold appears for filters within a certain critical line in the $\delta\varepsilon$-plane, which we calculate in the following.

Filter without perception threshold have to exhibit $p_i > 0$, even when the data has no power at all. Thus we investigate the extreme case $d = 0$ by inserting $m_p = 0$ into Eq. [48] and find after some algebra

\[
1 + 2 \frac{\varepsilon_i}{\theta_i} = \delta_i \frac{1}{\theta_i} \text{Tr} \left( (1 + Q_p)^{-1} I_i \right),
\]

with $I_i = S_i^{-1} S_i$ the unit matrix restricted to the $i$-th band. Since the marked expression on the rhs is one only for vanishing $p_i$, we find the critical line to be given by

\[
\delta_i^{\text{crit}} = 1 + 2 \frac{\varepsilon_i}{\theta_i}.
\]

Filters with $\delta_i > \delta_i^{\text{crit}}$ do not exhibit a perception threshold, since even for $d = 0$ all $p_i > 0$. Filters with $\delta_i < \delta_i^{\text{crit}}$ exhibit a perception threshold. We note that a non-Jeffreys prior with $\alpha_i > 1$ but still $q_i = 0$ can also be included into this classification scheme, by just adding $\alpha_i - 1$ to $\varepsilon_i$. Filters with $q_i > 0$ obviously do not exhibit a perception threshold, since even in the limit of vanishing data and vanishing propagator Eq. [48] has the positive solution $p_i = q_i/(\gamma_i + \varepsilon_i)$.

The point $(\delta_i, \varepsilon_i) = (1, 0)$ lies on top of the critical line, as can be seen in Fig. [1], and therefore the term critical filter seems to be appropriate for it. Although a general criterion for the position of the threshold in data space can easily be worked out, it is more instructive to investigate a simplified case.

B. Translation invariant data model

Here, we calculate the perception thresholds of our filters in the case of a translationally invariant data model. We assume the signal and noise to live in the same spatial space, and their covariances to be fully characterized by power spectra in Fourier space,

\[
S(k, q) = (2\pi)^n \delta(k - q) P_S(k),
\]

\[
N(k, q) = (2\pi)^n \delta(k - q) P_N(k),
\]

with $P_s(k) = |s(k)|^2/V$, and $P_N(k) = |n(k)|^2/V$, where $V$ is the observed volume. We define spectral bands with band spectra $P_{S_i}(k)$, so that $P_S(k) = \sum_i p_i P_{S_i}(k)$. We assume further that the signal processing can be completely described by a convolution with an instrumental beam,

\[
d(x) = \int dy R(x - y) s(y) + n(x),
\]

where the response-convolution kernel has a Fourier power spectrum $P_R(k) = |R(k)|^2$ (no factor $1/V$).

In this case $D$ can be fully described by a power spectrum,

\[
D(k, q) = (2\pi)^n \delta(k - q) P_{D}(k),
\]

with $P_D(k) = (P_{S_i}^{-1}(k) + P_R(k) P_N^{-1}(k))^{-1}$ and all spectral bands decouple.

C. Approximative treatment

The generic filter equations, Eqs. [48] now separate into independent equations for the individual $p_i$. Let us look first at the trace-terms in this equation, which now read

\[
\text{Tr} \left[ m_p m_p^T S_i^{-1} \right] = V \int \frac{dk}{(2\pi)^n} P_s(k) p_i^2 P_Q(k),
\]

\[
\text{Tr} \left[ D_p S_i^{-1} \right] = V \int \frac{dk}{(2\pi)^n} \left( 1 + p_i P_Q(k) \right).
\]

We define the data power $P_d(k) = |d(k)|^2/V$ and the i-band fidelity power $P_Q(k) = (P_S P_R/P_N)(k)$. We further use the approximation $V \int \frac{dk}{(2\pi)^n} f(k) \approx \int \frac{dk}{\pi} f(k)$, which assumes that $f(k)$, a combination of spectra, does not vary significantly over the narrow spectral band $i$. This permits us to write the characteristic equation, Eq. [48] which determines the filter band coefficients $p_i$ as an algebraic and dimensionless expression:

\[
x = \frac{1 + y}{y^2} \left[ (t y - u) ((1 + y) - \delta y) \right].
\]

Here, we have dropped the index $i$ and defined the noise-normalized data power $x = P_d(k_i)/P_N(k_i)$ and the measurement fidelity $y = p_i P_Q(k_i)$. The numerical coefficients are

\[
t = \frac{2}{\theta_i} (\gamma_i + \varepsilon_i) = 1 + \frac{2}{\theta_i} (\alpha_i - 1 + \varepsilon_i),
\]

\[
u = \frac{2}{\theta_i} q_i P_Q(k_i), \quad \text{and} \quad \delta = \delta_i.
\]

In case of Jeffreys prior, these simplify to $u = 0$ and $t = 1 + 2 \varepsilon_i/\theta_i$ and the characteristic equation has the following solutions

\[
y = 0, \quad \text{and}
\]

\[
y = \frac{x - x_0}{2t} \pm \sqrt{\left( \frac{x - x_0}{2t} \right)^2 - 1 + \frac{\delta}{t}}, \quad \text{with}
\]

\[
x_0 = 2t - \delta = 2 + 4 \varepsilon_i/\theta_i - \delta_i.
\]
Although there might be up to three simultaneous real solutions for a given $x$, always the largest value should be taken. This is non-zero only if

$$x \geq x_\text{th} = \begin{cases} 0, & x_0 < 1, \\ x_0 + 2\sqrt{l(l-1)}, & x_0 \geq 1. \end{cases} \quad (68)$$

The assumed dimensionless signal power $y$ is shown in Fig. 3 as a function of the dimensionless data variance $x$. Asymptotically, for $x \gg x_0$, we have a linear increase of assumed signal strength and data variance $y(x) = x - x_0$. The critical filter is special in that this relation holds exactly for the full region $x \geq x_\text{th} = x_0$. All of the MAP estimators in this work have $x_\text{th} > x_0$ and exhibit a jump from $y = 0$ to $y = \sqrt{1 - \delta/7}$ at $x = x_\text{th}$, followed by an approach to the linear asymptotic. The threshold approaches $x_\text{th} \to 1$ from above for $\delta \to \infty$ for the MAP spectrum filter, however, it is always $x_\text{th} = 4$ for the classical filter, independent of the spectral bin size $\delta_l$.

The PURE filter is the only one of our sample, which has no perception threshold since $y(x)$ is positive for all $x$. Even in case the data exhibits negligible variance $x \ll 1$, the filter still uses a non-negligible spectral amplitude, since $y(0) \approx 1/(\delta_l + 1)$. This might surprise, since the implied assumption of a significant signal variance is obviously not supported by the data. However, the renormalized filter aims for an optimal reconstruction, and not for an accurate power spectrum measurement, and letting some fraction of some data band with apparently low noise realization pass (remember $x \ll 1$) does not spoil this.

The combination of signal measurement and filtering can be regarded as a single response operator $R'_i$, with $R'_i = (m'_i)(d|x) = F'_i R_s = D M s$, which decomposes into separate pass-through factors for the individual bands, $R'_i = P_D(k_i) P_R(k_i) P_N^{-1}(k_i) = y/(1+y)$. This is also shown in Fig. 3.

D. Consequences for cosmological practice

The classical filter measures internally the power spectrum of a Wiener map, which is (iteratively) filtered with this very same spectrum (until convergence). Similar procedures are widely used in cosmology under the names Karhunen-Loève (KL, 12, 22, 29) and Feldman-Kaiser-Peacock (FKP, 7) estimators to measure power spectra of galaxy catalogues. These should therefore also exhibit a perception threshold for spectral modes with a data variance not largely exceeding the noise variance. Therefore, one would expect that cosmological spectra obtained by these estimator should exhibit modes with zero power. However, in applications of these scheme in the cosmological literature, the iterations of filtering and spectral measurements might not be repeated until convergence.2 Thus the knowledge system keeps some memory of the initial power spectrum choice, which can be regarded as a hidden prior regularizing the spectrum and preventing the perception threshold that a correctly implemented KL or FKP estimator would exhibit (see also 29 for a discussion on this).

VI. COMPARISON OF THE MAP MAKING ALGORITHMS

A. The test case

We want to examine the filter performances with an instructive test case. In case the spectral uncertainty is small, all filters in this work can be expected to pro-

---

2 Some random examples: Tegmark et al. 29, Percival et al. 23 as well as Feldman et al. 7 use a fixed and constant spectrum in the optimal data weighting step of the KL and FKP schemes, and do not iterate at all.
vide comparable results since they Wiener filter the data with basically the prior spectrum with small differences. Thus, in order to see the differences in performance more clearly, we again adopt Jeffreys prior for our spectral parameters ($\alpha_i = 1$ and $q_i = 0$, well, for numerical reasons $q_i = 0.01$). A spectrum, which naturally implements this distribution is the famous $1/f$-spectrum, which has equal power per decade in frequency space. To have a finite zero mode and signal variance, we adopt

$$P_S(k) = P_0 (1 + (k/k_0)^2)^{-\frac{1}{2}},$$

(69)

with $P_0 = 5$ and $k_0 = 2$. We further assume some white noise with $P_N(k) = \sigma_n^2 = 0.1$.

In case the response would be constant or a convolution, the spectral inference problems would be separable in Fourier space, as we have shown in the last section. In order to have a more complex problem, with coupling between the different unknown spectral parameters, we introduce a non-homogeneous observational signal response $R$ over the 257 pixel of our signal space, as displayed in Fig. 4 together with a test data set. We split the Fourier space in 64 disjunct spectral bands, with $p_i = 4$ for all but the lowest band, which has $p_0 = 5$, since it also contains the zero mode. Since we are dealing with a real-number signal in a discrete space, we have to take care of the negative frequency spectrum being identical to the positive ones, and therefore our bands are split into identical positive and negative parts, except the zero-band, which is continuous.

The signal reconstructions of the five filters are also shown in Fig. 4 and the used spectra in Fig. 5. The spectra are roughly ordered the way we expect them to be following our perception threshold analysis in Sect. V. However, there is the surprising modification that even the renormalized filter seems to suffer from a slight perception threshold, since many of the higher $k$-vector bands with lower signal to noise ratio are nearly free of power. A more informative prior for the power distribution would cure this, but this would limit the generality of our filter. So we should look for other yet unexploited prior information.

B. Spectral smoothness regularization

The $1/f$ signal spectrum adopted in our example is a member of the large class of smooth spectra, which do not exhibit spectral lines, jumps and edges. Spectral smoothness information can easily be incorporated into the framework. Since we do not want to specify a specific smoothness length scale, we require the double logarithmic derivative of the spectrum to be of limited variance. This can be done by introducing an additional prior energy for non-smoothness

$$E_{\text{reg}} = \frac{1}{2} \sigma_p^2 \int d\log k \left( \frac{d\log P_S(k)}{d\log k} \right)^2 \quad (70)$$

$$\approx \frac{1}{4} \sigma_p^2 \sum_i \frac{k_i + k_i-1}{k_i - k_{i-1}} (\tau_i - \tau_{i-1})^2 \equiv \frac{1}{2} \tau^T \tau.$$
Here we have (re-)introduced the logarithm of the power spectrum parameters \( \tau_i = \log p_i \), have discretized the integral and derivatives, and collected all coefficients in a matrix \( T \). The quadratic form in \( \tau \) in the last line shows that this is actually a log-normal prior contribution, which can be combined with the log-normal prior appearing in the renormalization calculation. Instead of repeating that calculation with now interdependent parameters, we just use our physical intuition to obtain the regularized characteristic equation for the filter spectrum parameters, we just use our physical intuition to obtain the regularized characteristic equation for the filter spectrum.

The unregularized evolution equation for \( \tau \), Eq. [56] can just be equipped with a regularizing force \(-dE_{\text{reg}}/d\tau\):

\[
\frac{d\tau}{dt} = \beta(\tau) - T \tau. \tag{71}
\]

The regularized Jeffreys prior case is then given by the fix point specified by \( \beta(\tau) = T \tau \) and reached asymptotically for \( t \to \infty \). The matrix \( T \) couples the neighboring bands together and thereby produces much smoother filter spectra without the gaps the other filter spectra exhibit, as can be seen in Fig. 5 where the regularized filter spectrum for \( \sigma_p = 2 \) is shown.

C. Full PURE filter

Spectral smoothness can not always be assumed, and therefore we should also think of other ways to improve the filter fidelity. One way is to be more precise in the PURE filter derivation. The largest approximation made was probably the neglect of the \( dj/dt \) term, which for infinite spectral uncertainty, \( t \to \infty \), leads to a trivial solution of \( m = 0 \). If we want to include this term, we can therefore only apply it for a finite amount of uncertainty, say up to \( t = 1 \). This implies that the initial starting point of the spectral renormalization flow would influence our result. In case of a concrete application, this might be very desirable, since there a good initial guess for the spectrum might be available.

In our more abstract discussion here, we want to avoid such choices, also in order to be sure not to have included too much spectral prior knowledge into the filter preventing a fair comparison to the others. Therefore we start the renormalisation flow including the \( dj/dt \) term with the fix point spectrum of the approximated PURE filter (without this term) and stop it at \( t = 1 \). This way we have both, independence of any prior spectrum and inclusion of non-Wiener corrections. The resulting filter seems to be partly cured from too generous predictions in regions without data while the results in better determined regions are practically unchanged, as can be seen in Fig. 6.

This can be understood in the following way. We have roughly \( dj/dt \propto -S^{-1}m \), since \( \langle m^2S^{-1}\rangle \approx g_i \) for most modes. If there is power at a poorly observed location in the map \( m \) on a level comparable to the well observed ones, \( j \) evolves in both regions with similar speed. However, the effect of this evolution to the map \( m = Dj \) is larger in regions with larger uncertainties, since \( D \) is larger there. Thus, any power spilled into observational gaps is removed faster than power in well observed regions. The full PURE filter seems to be aware of the lower certainty of the former.

D. Statistical comparison

A statistical assessment of the different filters is shown in Fig. 6. There it is apparent that the filters derived from MAP principles are worse than the PURE filter, with only the critical filter being comparable in performance. The underestimation of the power spectra due to
the perception threshold obviously reduces the fidelity of those filters.

The spectral smoothness regularized, renormalized filter clearly outperforms the unregularized ones, probably due to the lack of spectral gaps. Its performance is comparable to that of the Wiener filter using the correct signal power spectrum $P_S(k)$. The error variance for the latter filter is also displayed in Fig. 6 in comparison to its theoretical value given by the Wiener variance $D_{xx}$ (see Eq. 13). Finally, also the full PURE filter as described in the last section is shown. Its fidelity is comparable to the spectrally regularized one, without that any spectral smoothness assumptions had to be made. Of course, such assumptions could also be included into this filter.

VII. CONCLUSIONS

We showed how to deal with parameter uncertainties in information field theory by introducing an effective Hamiltonian over the joint space of the signal field and the parameters. In order to go beyond a classical, or Maximum a Posteriori treatment of the problem we presented an uncertainty renormalization scheme, in which the parameter uncertainty is successively fed into the knowledge system. The resulting parameter uncertainty renormalized estimation, PURE, can be used to tackle many signal inference problems including calibration uncertainties.

To demonstrate the advantage of PURE with a concrete example, we investigated the general problem of inferring a Gaussian signal with unknown spectrum from noisy data, which follows from a linear, but inhomogeneous data model. Following the parameter uncertainty renormalization and various classical approaches, four classical and one renormalized filter were derived. All filters can be regarded as Wiener filter operations with assumed signal spectra to be calculated from the data by a single recipe, Eq. 48, with just differences in two of its numerical coefficients.

The computational complexity of all those filters is therefore very similar and should not be a reason to prefer one over the other. Their signal fidelity, however, differs significantly. In case a non-informative Jeffreys prior is adopted for the spectral amplitudes, all classical filters suffer from a perception threshold. Spectral bands, which do not show more data power than the threshold, are completely filtered out. Three out of the four classical filters investigated have a perception threshold which requires data variance significantly above the noise level. The fourth one, the critical filter, lives on the critical line between filters with and without perception threshold in our space of filter parameters. The critical filter tries to match the correct spectrum on a logarithmic power scale. Its perception starts therefore for modes with a variance just above the noise level, as soon the data indicates some potential signal power.

The critical filter coresponds in general to the Karhunen-Loève method \cite{12, 22, 29}, and for an infinite window function to the FKP method \cite{7} frequently used in cosmology to estimate power spectra of galaxy catalogs. It seems that the perception threshold of this method is often ‘cured’ in applications by a truncation of the full iterative scheme. This implies the presence of a hidden spectrum prior in such estimates.

The PURE filter percepts also for spectral bands, which by chance exhibit less power than expected for the noise alone. This might appear as being too generous – the signal spectrum adopted by this filter is typically larger than the correct and therefore optimal one. However, the PURE filter exhibits the largest fidelity of our filter sample, even slightly better than that of the critical filter. The reason lies in the asymmetric fidelity loss for under- and overestimating the true signal spectrum. Spectrum underestimation is much worse than overestimation in terms of signal reconstruction accuracy. The renormalized filter knows about this and adds a safety margin to any spectral band. This margin is inversely proportional to the number of data degrees of freedom informing about the signal spectrum in this band. Thus, in the limit of a large number of data points determining the band spectrum the renormalized filter approaches the critical one, but always from the perception threshold free side.

Although the classical filter resulted from maximizing the exact effective, parameter marginalized Hamiltonian (Eq. 51), it performs much worse than the critical and PURE filters. Thus, this is an example where the MAP principle, or equivalently a tree-level IFT calculation, provides a poorly performing algorithm, and uncertainty loop corrections as explicitly included in the PURE filter are essential.

The PURE filter, as well as the others, can be further improved by adding any additional spectral information. One way is to use informative priors on the spectral behavior, which instantaneous cure the perception threshold problem. However, even in case no information on the location of the spectrum is available, information about its smoothness as a function of the Fourier space coordinate may be exploited. We show that the performance of the PURE filter with spectral smoothness prior approaches that of the optimal Wiener filter for known signal power spectrum.

Since the computational complexity of the renormalized filter is identical to the critical one already used in cosmology, there exists no reason not to use it for Wiener filtering of signals with unknown spectra. One only has to keep in mind that the internally used spectrum of the filter is not the best estimate of the signal spectrum, but an overestimate. The critical spectrum provides such an estimate, using the posterior maximum for the logarithm of the spectral amplitudes.

The full PURE filter, which contains non-Wiener filter corrections and requires the more expensive evaluation of the renormalization flow equation, performs best among all spectrally unreguarlized filters. Spectral smoothness
information can also be incorporated into it if available.

To conclude, the PURE scheme to construct optimized filters presented in this work is very general and should also be applicable to the problems of inference with uncertainties in the instrument response, the typical calibration problem, and for measurements without known noise level. A better understanding of the implications and assumptions of the commonly used process of self-calibration should be feasible, and possibly also improvements thereof. The pseudo-time parameter appearing in the renormalization flow, the amount of uncertainty or parameter dispersion per physical time is fixed, the equation permits to continuously update the unknown. The renormalization flow, the amount of uncertainty or parameter uncertainty renormalization equation offers a natural possibility to model this. Once the amount of uncertainty dispersion per physical time is fixed, the equation permits to continue to update the unknown parameters by combining past and novel information in an optimal, and controlled way. The PURE approach may thereby make contributions to the technologically important field of optimal control and time dependent instrument calibration.

Acknowledgments

We acknowledge helpful discussions with and comments on the manuscript by Jens Jasche, Henrik Junklewitz, Cornelius Weig, and Gerhard Börner.

Appendix A: Signal covariance likelihood

In order to find the posterior of the signal covariance, we have to calculate \( P(p) Z_p / Z \). In case that \( S, N, \) and \( R \) are all invertible operators, one can show that the evidence \( Z_p[0] \) for any parameter \( p \) of the free Hamiltonian, given by Eq. [11] is

\[
Z_p = P(d)p = G(d, RS_p R^\dagger + N). \tag{A1}
\]

This formula can also intuitively be read as the data likelihood given \( p \), since it compares the power in the data to their expected fluctuations level \( \langle dd^\dagger \rangle_{(d,dp)} = RS_p R^\dagger + N \). It can therefore be used for a Bayesian estimate of any model-parameter of the free theory, not only for spectral parameters as in this work.

Appendix B: Derivation of the Gaussian prior

Here we show how the different auxiliary variables \( \tau_{ij} \) combine into a normal distribution for \( \tau_i = \sum_j \tau_{ij} \), as was assumed in Sect. [V H]. We drop in the following the index \( i \), which labels the signal bands. Since we assume \( e^{-\tau} \) to be distributed according to Eq. [43] we have

\[
P(\tau_j) = \frac{\exp \left[ - (\alpha - 1) (\tau_j - \log q - q e^{-\tau_j}) \right]}{\Gamma[\alpha - 1]} \tag{B1}
\]

The non-bias condition, Eq. [23] translates into

\[
\langle \tau_i \rangle_{(\tau_j)} = \log q - \psi_0(\alpha - 1) = 0, \tag{B2}
\]

with \( \psi_0(z) \) being the Polygamma function. This condition fixes \( q(\alpha) = e^{\psi_0(\alpha - \frac{1}{2})} \), which for large values of \( \alpha \), and thereby for well localized auxiliary parameters, is asymptotically \( q = \alpha = \frac{1}{2} \). The dispersion of the auxiliary variables is

\[
\delta t = \langle \tau_j^2 \rangle_{(\tau_j)} = \psi_1(\alpha - 1), \tag{B3}
\]

which asymptotically is \( \delta t = 1/(\alpha - 1) \) for large \( \alpha \).

Now, we can work out the total prior resulting from the combination of \( N = t/\delta t \) auxiliary variables, where \( t \) is the uncertainty level of the prior, and \( \delta t \) that of the individual variables:

\[
P(\tau) = \left( \prod_{j=1}^{N} \int d\tau_j \, P(\tau_j) \right) \delta(\tau - \sum_{j=1}^{N} \tau_j)
\]

\[
= \int \frac{dk}{2\pi} \left( \prod_{j=1}^{N} \int d\tau_j \, P(\tau_j) \right) e^{-ik(\tau - \sum_{j=1}^{N} \tau_j)}
\]

\[
= \int \frac{dk}{2\pi} \left( \prod_{j=1}^{N} \int d\tau_j \, q^{ik} \exp \left[ - (\delta t^{-1} - ik) \tau_j - q e^{-\tau_j} \right] \right) N
\]

\[
\quad \times e^{-ik\tau}
\]

\[
= \int \frac{dk}{2\pi} \left( \frac{\Gamma[\delta t^{-1} - ik]}{\Gamma[\delta t^{-1}]} q^{ik} \right)^N e^{-ik\tau}
\]

\[
= \frac{N}{\Gamma[\delta t^{-1}]} \exp \left[ - i k \tau + N \log \left( \frac{\Gamma[\delta t^{-1} - ik]}{\Gamma[\delta t^{-1}]} \right) \right]
\]

\[
\quad - i k N \psi_0(\alpha - 1)
\]

\[
\quad \rightarrow G(\tau, t) \text{ for } \delta t \rightarrow 0, \tag{B4}
\]

as also expected from the central limit theorem of statistics. Thus, the resulting distribution for the \( p_i \) parameter is log-normal.
In case \( R \) is not invertible, one could perturb it with a small \( \epsilon \)-contribution, which makes it invertible, derive Eq. A1 and take the limit \( \epsilon \to 0 \), which is permitted by this equation. More formal proofs can be found in [8][14].