Reconstructing signals from noisy data with unknown signal and noise covariance

Niels Oppermann,[*] Georg Robbers, and Torsten A. Enßlin

Max Planck Institute for Astrophysics, Karl-Schwarzschild-Str. 1, D-85741 Garching, Germany

We derive a method to reconstruct Gaussian signals from linear measurements with Gaussian noise. This new algorithm is intended for applications in astrophysics and other sciences. The starting point of our considerations is the principle of minimum Gibbs free energy, which was previously used to derive a signal reconstruction algorithm handling uncertainties in the signal covariance. We extend this algorithm to simultaneously uncertain noise and signal covariances using the same principles in the derivation. The resulting equations are general enough to be applied in many different contexts. We demonstrate the performance of the algorithm by applying it to specific example situations and compare it to algorithms not allowing for uncertainties in the noise covariance. The results show that the method we suggest performs very well under a variety of circumstances and is indeed qualitatively superior to the other methods in cases where uncertainty in the noise covariance is present.

I. INTRODUCTION

The problem of signal inference consists of reconstructing a set of parameters or even a continuous field s from some data set d, which is influenced in some way by the signal:

$$d = f(s) + n. \tag{1}$$

Two problems will arise. First, the function f may not be invertible and, second, the noise term n will not be known. In the Bayesian framework, one uses prior information on the signal and the noise term to calculate a best estimate for the true signal realization or, ideally, the whole probability distribution for the signal given the prior information and the information contained in the data.

Symmetry considerations and knowledge about the underlying physics of the signal and the measurement process may restrict the class of priors that one has to consider. They might, however, still contain some free parameters that then become part of the inference problem. The case in which the signal covariance contains uncertain parameters was tackled in Ref. [1], producing a whole class of filters for this problem. The filter that we extend in this work was reproduced in Ref. [2], where the principle of minimum Gibbs free energy was introduced (cf. also Sec. III) and successfully applied in an astrophysical setting in Ref. [3].

Here we focus on the case where we can assume zero-mean Gaussian priors both for the signal and for the noise. The priors are therefore completely characterized by the respective covariance matrices. Our goal is to extend the study of Ref. [2] to the case in which both the signal covariance and the noise covariance contain parameters that are not known *a priori*. This is motivated mainly by applications from the field of astrophysics. The theory and resulting filter formulas, however, are of general applicability. Gaussian noise, e.g., is omnipresent in nearly every area of the natural sciences, and the situation in which its variance is not precisely known should be a rather common one.

Previous work dealing with the problem of unknown noise variance has mainly dealt with specific applications. One of

these applications is the field of image reconstruction. Here it is usually assumed that the measured picture is the sum of the underlying signal and a white Gaussian noise term. Often it is further assumed that the noise level, i.e., its variance, is the same in every image pixel. A comparison of different algorithms for noise estimation under these assumptions can be found, e.g., in Ref. [4]. An example for an algorithm allowing for inhomogeneous noise is presented in Ref. [5], where a wavelet transform of the image is applied and the lack of correlated noise is exploited. Most of these algorithms, however, are not derived by rigorous statistical calculations but rather by a combination of intuition and experience.

From a mathematical viewpoint, the problem of an unknown noise prior has received some attention in the theory of density deconvolution, which deals with the inference of the probability density for a signal from measurements with additive noise. Here the signal is usually assumed to consist of independent identically distributed variables. The case of Gaussian noise with unknown variance has been considered, e.g., by Refs. [6] and [7].

In this work we create a general setting with well-defined assumptions and a traceable derivation of a general filter formula within a Bayesian framework, not losing sight of its applicability. Our result can accommodate a host of different assumptions and models, such as correlated or uncorrelated noise. It allows for a distinction between the data space and the signal space, with possibly different numbers of degrees of freedom.

The remainder of the paper is organized as follows. In Sec. II we introduce our model for the measurement process and the notation that is required. The derivation of the filter formulas follows in Sec. III. We then demonstrate the usefulness of our filter by applying it in a set of mock observational situations in Sec. IV and discuss the implications in Sec. V.

II. SIGNAL MODEL AND NOTATION

We assume a linear measurement process where the data are a superposition of a linear signal response and a noise term:

$$d = Rs + n. \tag{2}$$

---

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

Here both the data and noise and the signal can be finite-dimensional vectors or continuous fields defined on some manifold. The response matrix R maps a field in the signal space to a field in the data space. In the continuous limit the matrix vector product becomes

$$(Rs)_i = \int dx\, R_{ix} s_x, \tag{3}$$

where the index denotes the value of a field at this position. In physical applications the data vector will always be discrete, since only a finite number of measurements can be taken, while the signal space might well be continuous. The result of any numerical signal reconstruction, however, will at best be a discretized version of the continuous field.

Further, we assume Gaussian prior statistics both for the signal and for the noise contribution, i.e., $s \hookleftarrow \mathcal{G}(s,S)$, $n \hookleftarrow \mathcal{G}(n,N)$, where

$$\mathcal{G}(a,A) = \frac{1}{\sqrt{|2\pi A|}} \exp\left(-\frac{1}{2} a^\dagger A^{-1} a\right) \tag{4}$$

denotes a multivariate Gaussian distribution in a with covariance matrix A. We use the dagger symbol to indicate a scalar product,

$$a^\dagger b = \int dx\, a_x^* b_x, \tag{5}$$

and the asterisk to denote complex conjugation. This corresponds to the notation introduced in Ref. [8].

The problem of signal reconstruction is to find an optimal estimate m for the signal realization that the measured data arose from. Optimality in an L^2-norm sense leads to

$$m = \langle s \rangle_{\mathcal{P}(s|d)} := \int \mathcal{D}s\, s\, \mathcal{P}(s|d), \tag{6}$$

i.e., the posterior mean. The integration is performed over all possible signal configurations. In the discrete case this becomes a product of one-dimensional integrals,

$$\int \mathcal{D}s = \int_{-\infty}^{+\infty} ds_1 \int_{-\infty}^{\infty} ds_2 \cdots, \tag{7}$$

where $s = (s_1, s_2, \ldots)$ is the vector of signal values at locations $1, 2, \ldots$. Ideally, we would also like to obtain some information on the posterior distribution $\mathcal{P}(s|d)$ other than its mean. If the signal and noise covariances are known, the posterior is a Gaussian $\mathcal{G}(s - m, D)$ with mean

$$m = Dj \tag{8}$$

and covariance D, where $j = R^\dagger N^{-1} d$ is called the *information source* and $D = (S^{-1} + R^\dagger N^{-1} R)^{-1}$ the *information propagator* (see Ref. [8]), and the dagger attached to a matrix denotes its Hermitian conjugate.

In this paper we are concerned with the case in which neither the signal covariance matrix S nor the noise covariance matrix N is known. We parametrize these matrices as sums of their eigenvalues $\tilde{p}_k$ and $\tilde{\eta}_j$ multiplied with the projectors onto the respective eigenspaces $\tilde{S}_k$ and $\tilde{N}_j$. The parameters can be rescaled by including some numerical values $\tilde{s}_k$ and $\tilde{n}_j$ in the

projection-like matrices, making the rescaled version of the parametrization

$$S = \sum_k p_k S_k, \tag{9}$$

$$N = \sum_j \eta_j N_j, \tag{10}$$

where

$$p_k = \frac{\tilde{p}_k}{\tilde{s}_k}, \quad \eta_j = \frac{\tilde{\eta}_j}{\tilde{n}_j} \tag{11}$$

and

$$S_k = \tilde{s}_k \tilde{S}_k, \quad N_j = \tilde{n}_j \tilde{N}_j. \tag{12}$$

Furthermore, we define the pseudo-inverse matrices $S_k^{-1} = \tilde{s}_k^{-1} \tilde{S}_k$ and $N_j^{-1} = \tilde{n}_j^{-1} \tilde{N}_j$, so that $S_k^{-1} S_k$ and $N_j^{-1} N_j$ are identity operators on the respective eigenspaces.

We assume here that the eigenspaces corresponding to the different eigenvalues are known *a priori*, e.g., from symmetry considerations. However, the formalism allows for eigenvalues of different eigenspaces becoming equal *a posteriori*.

Finally, we also need to define some priors for the parameters p_k and η_j. As was done in Refs. [1] and [2], we assume each parameter to be *a priori* independent from all the others and use inverse Gamma distributions, i.e., power laws with exponential cutoff, as priors for the individual parameters,

$$\mathcal{P}(p,\eta) = \mathcal{P}(p)\mathcal{P}(\eta), \tag{13}$$

$$\mathcal{P}(p) = \prod_k \frac{1}{q_k \Gamma(\alpha_k - 1)} \left(\frac{p_k}{q_k}\right)^{-\alpha_k} \exp\left(-\frac{q_k}{p_k}\right), \tag{14}$$

$$\mathcal{P}(\eta) = \prod_j \frac{1}{r_j \Gamma(\beta_j - 1)} \left(\frac{\eta_j}{r_j}\right)^{-\beta_j} \exp\left(-\frac{r_j}{\eta_j}\right). \tag{15}$$

The parameters α_k and β_j determine the steepness of the power law, and the parameters q_k and r_j give the position of the cutoff. In the limit $(\alpha_k, \beta_j) \to (1,1)$ and $(q_k, r_j) \to (0,0)$, this turns into the so-called Jeffreys prior, which is flat on a logarithmic scale and can therefore be characterized as noninformative.

III. DERIVATION OF THE FILTER FORMULAS

With the priors for s, n, p, and η, we can calculate the joint probability of the signal and the data by marginalizing over the parameters p and η,

$$\begin{aligned}
\mathcal{P}(s,d) &= \int \mathcal{D}p \int \mathcal{D}\eta\, \mathcal{P}(s,d|p,\eta)\mathcal{P}(p,\eta) \\
&= \int \mathcal{D}p \int \mathcal{D}\eta\, \mathcal{P}(d|s,p,\eta)\mathcal{P}(s|p)\mathcal{P}(p,\eta) \\
&= \int \mathcal{D}p \int \mathcal{D}\eta\, \mathcal{G}(d - Rs, N)\mathcal{G}(s,S)\mathcal{P}(p,\eta). \tag{16}
\end{aligned}$$

Solving the integrals yields

$$
\begin{aligned}
\mathcal{P}(s,d) &= \prod_k \frac{\Gamma(\gamma_k)q_k^{\alpha_k-1}}{\Gamma(\alpha_k-1)(2\pi)^{\rho_k/2}} \left(q_k + \frac{1}{2}s^\dagger S_k^{-1}s \right)^{-\gamma_k} \\
&\times \prod_j \frac{\Gamma(\delta_j)r_j^{\beta_j-1}}{\Gamma(\beta_j-1)(2\pi)^{\mu_j/2}} \\
&\times \left[r_j + \frac{1}{2}(d-Rs)^\dagger N_j^{-1}(d-Rs) \right]^{-\delta_j}, \quad (17)
\end{aligned}
$$

where $\rho_k = \mathrm{tr}(S_k^{-1}S_k)$, $\mu_j = \mathrm{tr}(N_j^{-1}N_j)$, $\gamma_k = \rho_k/2 + \alpha_k - 1$, and $\delta_j = \mu_j/2 + \beta_j - 1$. Note that the posterior is proportional to this joint likelihood for any given data set.

One could construct the *maximum a posteriori* estimator; however, this was shown in Ref. [1] to perform poorly due to a perception threshold; i.e., modes with too little power in the data are completely filtered out. A better estimate for the posterior mean of the signal can be constructed using the formalism of minimum Gibbs free energy, derived in Ref. [2], where thermodynamic quantities are introduced by identifying the posterior probability density with a canonical density function according to

$$
\mathcal{P}(s|d) = \frac{\mathcal{P}(s,d)}{\mathcal{P}(d)} = \frac{e^{-T^{-1}H(s,d)}}{Z(d)}. \quad (18)
$$

The Gibbs energy is then

$$
G = U - T S_B, \quad (19)
$$

where $U = \langle H \rangle_{\mathcal{P}(s|d)}$ is the internal energy, $S_B = \langle -\log\mathcal{P}(s|d) \rangle_{\mathcal{P}(s|d)}$ the Boltzmann entropy, and $H = -\log\mathcal{P}(s,d)$ is called the information Hamiltonian. The temperature T serves as a tuning parameter, shifting the weight between the internal energy term and the entropy term in the Gibbs free energy.

Approximating the posterior with a Gaussian with mean m and covariance D,

$$
\mathcal{P}(s|d) \approx \mathcal{G}(s-m,D), \quad (20)
$$

gives an approximate internal energy $\tilde{U}$, an approximate entropy $\tilde{S}_B$, and therefore an approximate Gibbs energy

$$
\begin{aligned}
\tilde{G}(m,D) &= \tilde{U}(m,D) - T\tilde{S}_B \\
&= \langle H(s,d) \rangle_{\mathcal{G}(s-m,D)} - \frac{T}{2}\mathrm{tr}[1 + \log(2\pi D)]. \quad (21)
\end{aligned}
$$

For $T = 1$, this approximate energy is, apart from an additive constant, identical to the nonsymmetric Kullback-Leibler distance [9] between the full posterior and the Gaussian approximation,

$$
\begin{aligned}
\tilde{G}(m,D) &= \langle H(s,d) + \log[\mathcal{G}(s-m,D)] \rangle_{\mathcal{G}(s-m,D)} \\
&= \int \mathcal{D}s\, \mathcal{G}(s-m,D) \log\left[\frac{\mathcal{G}(s-m,D)}{\mathcal{P}(s,d)} \right] \\
&= \int \mathcal{D}s\, \mathcal{G}(s-m,D) \log\left[\frac{\mathcal{G}(s-m,D)}{\mathcal{P}(s|d)} \right] \\
&\quad + \log[\mathcal{P}(d)] \\
&= d_{\mathrm{KL}}[\mathcal{G}(s-m,D),\mathcal{P}(s|d)] + \log[Z(d)], \quad (22)
\end{aligned}
$$

as was shown already in Ref. [2].

The approximate internal energy in our case, calculated from the joint probability of Eq. (17), is

$$
\begin{aligned}
\tilde{U}(m,D) \cong &\sum_k \gamma_k \underbrace{\left\langle \log\left(q_k + \frac{1}{2}s^\dagger S_k^{-1}s \right) \right\rangle_{\mathcal{G}(s-m,D)}}_{=:\mathcal{A}_k} \\
&+ \sum_j \delta_j \underbrace{\left\langle \log\left[r_j + \frac{1}{2}(d-Rs)^\dagger N_j^{-1}(d-Rs) \right] \right\rangle_{\mathcal{G}(s-m,D)}}_{=:\mathcal{B}_j},
\end{aligned}
$$

$$(23)$$

where we have dropped terms that are independent of m and D. The logarithms can be expanded in an asymptotic power series, giving

$$
\begin{aligned}
\mathcal{A}_k = &\log(\tilde{q}_k) \\
&- \sum_{i=1}^\infty \frac{(-1)^i}{i\tilde{q}_k^i} \underbrace{\left\langle \left(q_k + \frac{1}{2}s^\dagger S_k^{-1}s - \tilde{q}_k \right)^i \right\rangle_{\mathcal{G}(s-m,D)}}_{=:\tilde{A}_{ki}} \quad (24)
\end{aligned}
$$

and

$$
\begin{aligned}
\mathcal{B}_j = &\log(\tilde{r}_j) - \sum_{i=1}^\infty \frac{(-1)^i}{i\tilde{r}_j^i} \\
&\times \underbrace{\left\langle \left[r_j + \frac{1}{2}(d-Rs)^\dagger N_j^{-1}(d-Rs) - \tilde{r}_j \right]^i \right\rangle_{\mathcal{G}(s-m,D)}}_{=:\tilde{B}_{ji}},
\end{aligned}
$$

$$(25)$$

where we have chosen the linear dependencies to be captured by

$$
\begin{aligned}
\tilde{q}_k &= \left\langle q_k + \tfrac{1}{2}s^\dagger S_k^{-1}s \right\rangle_{\mathcal{G}(s-m,D)} \\
&= q_k + \tfrac{1}{2}\mathrm{tr}\left[(mm^\dagger + D)S_k^{-1} \right] \quad (26)
\end{aligned}
$$

and

$$
\begin{aligned}
\tilde{r}_j &= \left\langle r_j + \tfrac{1}{2}(d-Rs)^\dagger N_j^{-1}(d-Rs) \right\rangle_{\mathcal{G}(s-m,D)} \\
&= r_j + \tfrac{1}{2}\mathrm{tr}\left\{ [(d-Rm)(d-Rm)^\dagger + RDR^\dagger]N_j^{-1} \right\}, \quad (27)
\end{aligned}
$$

respectively.

Here, we restrict ourselves to the zeroth-order solution; i.e., we neglect all contributions from $\tilde{\mathcal{A}}$ and $\tilde{\mathcal{B}}$. Furthermore we set $T = 1$. The case with $T \neq 1$ is discussed up to second order in the Appendix.

Now we search for the optimal Gaussian approximation to the posterior by minimizing the approximate Gibbs energy, which is equivalent to minimizing the Kullback-Leibler distance between the two probability densities, according to Eq. (22). Taking the functional derivatives of Eq. (21) with

respect to m and D and equating them to zero yields the equations

$$m = Dj, \quad (28)$$

$$j = \sum_j \frac{\delta_j}{\tilde{r}_j} R^\dagger N_j^{-1} d, \quad (29)$$

$$D = \left(\sum_k \frac{\gamma_k}{\tilde{q}_k} S_k^{-1} + \sum_j \frac{\delta_j}{\tilde{r}_j} R^\dagger N_j^{-1} R \right)^{-1}. \quad (30)$$

By comparing these expressions to the Wiener filter formula, Eq. (8), we can read off the parameters $p_k = \frac{\tilde{q}_k}{\gamma_k}$ and $\eta_j = \frac{\tilde{r}_j}{\delta_j}$ for the signal and noise covariance matrix, respectively.

So altogether the equations that need to be solved simultaneously are

$$m = Dj, \quad (31a)$$

$$p_k = \frac{q_k + \frac{1}{2}\text{tr}[(mm^\dagger + D)S_k^{-1}]}{\frac{\rho_k}{2} + \alpha_k - 1}, \quad (31b)$$

$$\eta_j = \frac{r_j + \frac{1}{2}\text{tr}\{[(d - Rm)(d - Rm)^\dagger + RDR^\dagger]N_j^{-1}\}}{\frac{\mu_j}{2} + \beta_j - 1}. \quad (31c)$$

Thus, we find both the posterior mean and the posterior covariance for the signal. Note that the first two of these three equations were already found in Refs. [1] and [2], where the reconstruction of signals with unknown power spectra is discussed. The term *critical filter* was coined in Ref. [1] to refer to this filter since it belongs to a family of filters lying on a line in the parameter plane of Ref. [1] that separates the filters with a perception threshold from those without. The additional uncertainty in the noise covariance that we introduce here simply adds one more equation, leading to an *extended critical filter*.

IV. APPLICATION TO SIMULATED SIGNALS

Here we demonstrate the performance of our signal reconstruction algorithm under different circumstances. The computations were performed using the HEALPIX [12] and SAGE [13] software packages.

A. Setup

We consider two different scenarios. First, we consider a simple one-dimensional test case, where the signal is supposed to be a real field defined over some interval with periodic boundary conditions. We discretize the interval into 2048 pixels. For simplicity, we set the response matrix R to be the identity operator, so that our data set consists of 2048 individual points as well. We further assume statistical homogeneity for the signal field, leading to a covariance matrix that is diagonal in Fourier representation,

$$S_{kk'} = \langle s_k s_{k'}^* \rangle_{\mathcal{P}(s)} = \delta_{kk'} P_s(k), \quad (32)$$

with the power spectrum $P_s(k)$ on its diagonal. For this power spectrum we choose a simple power law

$$P_s(k) \propto (1 + k)^{-2} \quad (33)$$

and draw a random realization of the signal from it.

Motivated by astrophysical applications, we also consider a real signal field on the sphere,

$$s : \mathcal{S}^2 \longrightarrow \mathbb{R}. \quad (34)$$

Using again $R = \mathbb{1}$, the data and noise are also fields on the sphere,

$$d, n : \mathcal{S}^2 \longrightarrow \mathbb{R}. \quad (35)$$

In the numerical implementation, we use the HEALPIX [10] discretization scheme at a resolution of $N_{\text{side}} = 16$, which leads to 3072 pixels. Under the assumption of statistical homogeneity and isotropy, the signal covariance matrix in this case becomes diagonal in the basis given by the spherical harmonics components,

$$S_{(\ell m)(\ell' m')} = \langle s_{\ell m} s_{\ell' m'}^* \rangle_{\mathcal{P}(s)} = \delta_{\ell \ell'} \delta_{mm'} C_\ell, \quad (36)$$

where C_ℓ are the angular power spectrum components. We draw our signal realization again from a power-law spectrum,

$$C_\ell = (1 + \ell)^{-2}. \quad (37)$$

We assume the noise to be uncorrelated in the position basis, making the noise covariance matrix diagonal in this basis,

$$N_{\hat{n}\hat{n}'} = \langle n_{\hat{n}} n_{\hat{n}'} \rangle_{\mathcal{P}(n)} = \delta_{\hat{n}\hat{n}'} \sigma_{\hat{n}}^2, \quad (38)$$

where $\hat{n}$ and $\hat{n}'$ denote positions on the sphere or on the interval, respectively. Within this framework, we consider three cases for the noise statistics. In the first one, we use homogeneous noise with variance $\sigma_{\hat{n}}^2 = 1/4$ independent of $\hat{n}$. For the second case we divide the data space into three zones. In the left/southern third, we suppress the noise variance by a factor of nine, and in the right/northern third we enhance it by a factor of nine, while we leave it unchanged in the middle. Finally, in the third case we again assume homogeneous noise with variance 1/4, but we enhance the variance in five percent of the pixels, randomly selected, by a factor of 100. Both the signal and the three resulting data realizations are shown in Fig. 1 for the one-dimensional case and in Fig. 2 for the spherical case, along with the results of different reconstructions that we discuss next.

B. Reconstructions

We first apply the standard Wiener filter formula, the results of which are shown in the second row of Fig. 1 for the one-dimensional case and in the middle row of Fig. 2 for the spherical case. For this we assume the correct power spectrum to be known, but we assume homogeneous noise with variance 1/4 in all three cases. In the case where this assumption is correct, the reconstruction is known to be optimal, and this is confirmed by visual inspection of the outcome. In the cases with inhomogeneous noise, the Wiener filter fails to completely filter out the noise structures in the data in the regions where the noise is underestimated and therefore reproduces some of them in the reconstruction, as one would expect. This is true for the right (northern) third in the middle column of Fig. 1 (Fig. 2), as well as the noisy pixels in the right column. The opposite should happen in the left (southern) third in the middle column of Fig. 1 (Fig. 2), where the noise is overestimated. One would expect that structures in the data that
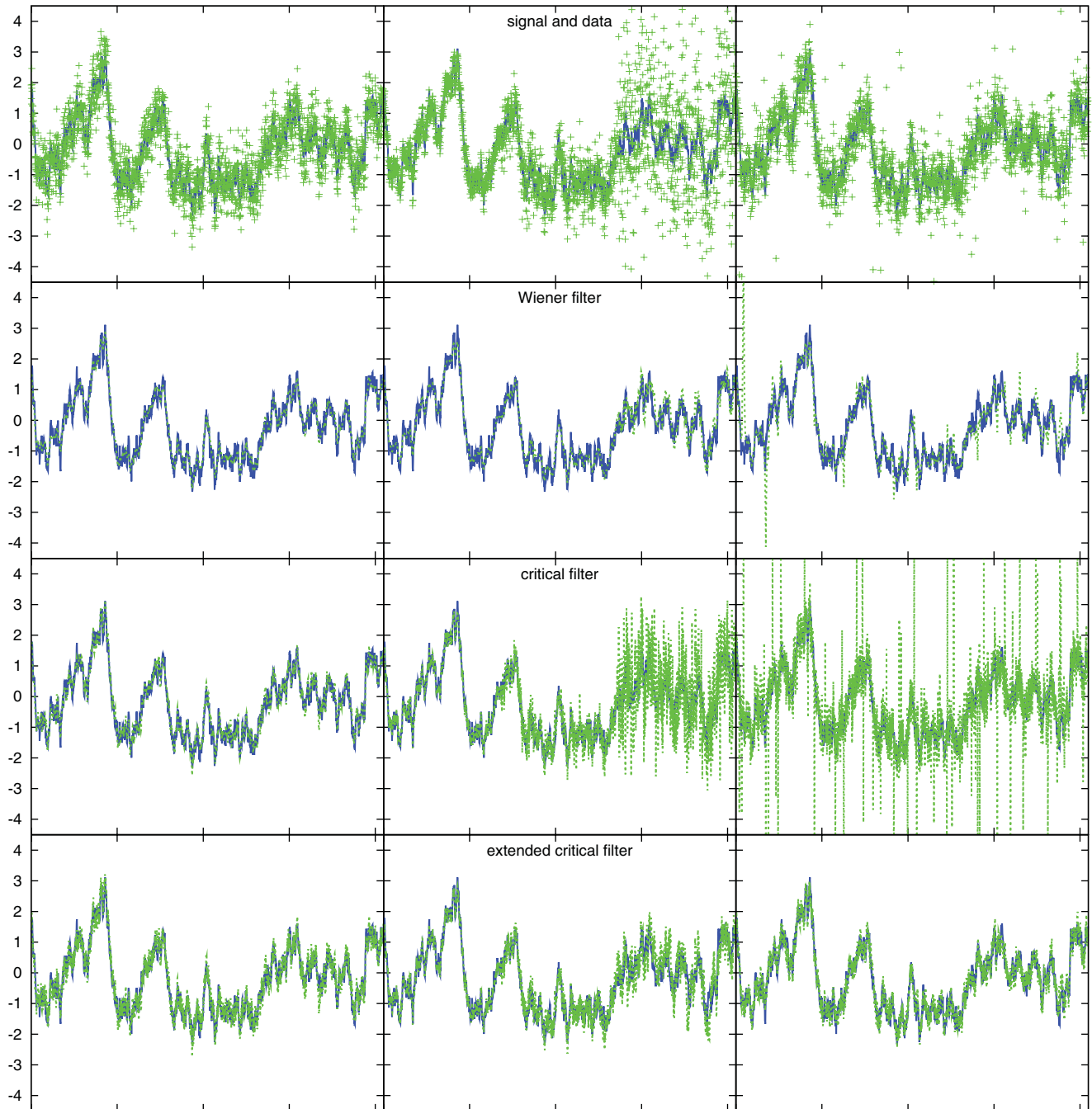
FIG. 1. (Color online) Comparison of different filter algorithms in the one-dimensional test case. Each column corresponds to a different setting. The signal, drawn from a power-law power spectrum, is the same in each case and depicted in each panel with a solid line. The left column contains homogeneous noise, while in the middle column, the noise is suppressed in the left third of the interval and enhanced in the right third, and in the right column the noise is enhanced in some individual pixels. The first row shows the signal realization along with the data. The second row shows the reconstruction using the Wiener filter formula, assuming the correct power spectrum and under the assumption of homogeneous noise; the third row shows the critical filter reconstruction, assuming the power spectrum to be unknown, but still assuming homogeneous noise. The last row shows the extended critical filter reconstruction in which both the signal power spectrum as well as the noise variance are assumed to be unknown. The respective reconstructions are depicted by a dashed line that lies on top of the solid one in many cases. In the two right panels of the first row, some of the data points lie outside the area that is shown.

are actually due to the signal get filtered out. This is actually happening, although it is barely visible in the resulting plots.

Next we assume that the power spectrum is not known *a priori*: i.e., we apply the critical filter. The resulting plots are shown in the third row of Fig. 1 for the one-dimensional case

and the fourth row of Fig. 2 for the two-dimensional case. In the one-dimensional case we define the S_k operators of Eq. (9) to be projections onto bins of width $\Delta_k = 2$ in Fourier space, effectively assuming that the two scales that enter the bin have the same power. This binned power is then represented by

FIG. 2. (Color online) Comparison of different filter algorithms in the spherical case. Each column corresponds to a different setting. The signal, drawn from a power-law power spectrum, is the same in each case. The left column contains homogeneous noise, while in the middle column, the noise is suppressed in the southern third of the sphere and enhanced in the northern third, and in the right column the noise is enhanced in some individual pixels. The first row shows the signal realization and the second row the data. The third row shows the reconstruction using the Wiener filter formula, assuming the correct power spectrum and under the assumption of homogeneous noise; the fourth row shows the critical filter reconstruction, assuming the power spectrum to be unknown, but still assuming homogeneous noise. The last row shows the extended critical filter reconstruction in which both the signal power spectrum as well as the noise variance are assumed to be unknown.

the parameter p_k in Eq. (9). This binning is necessary in one dimension since each individual Fourier component contains only two degrees of freedom [14]. In the spherical case, we can directly use the angular power spectrum components C_ℓ as parameters, and the projection-like operators S_k become actual projections onto the ℓ-th angular scale, which contains $2\ell + 1$ degrees of freedom. In both cases we assume Jeffreys prior for the unknown parameters. Then we simply iterate the

first two lines of Eq. (31), while keeping the assumption of homogeneous noise with variance 1/4.

In the cases where our assumptions about the noise are true, the resulting map is very close to the Wiener filter reconstruction, confirming the assessment of Refs. [1–3] that the critical filter can yield a very accurate reconstruction, even if the power spectrum is completely unknown. In the cases where we have made false assumptions about the noise,

however, we see the same problems that the Wiener filter reconstruction has, only much stronger pronounced. This is because the reconstructed power spectrum now actually accounts for the features in the data that are due to noise where this is underestimated. With this power spectrum, the map reconstruction tends to favor these features even more than when the correct power spectrum is used. This amplifying effect is again much more prominent where the noise was underestimated than where it was overestimated.

Finally, we account for the possibility that we might have misestimated the noise statistics by applying the full extended critical filter, derived in Sec. III. As projection-like matrices N_j we choose projections onto the jth pixel of the interval and sphere, respectively, multiplied with our original guess for the noise variance in that pixel, $\sigma_j^2 = 1/4$. In this way, the parameters η_j become correction factors for the noise variance of each data point. For the prior parameters we choose $\beta_j = 2$ and we adapt r_j such that $\langle \log(\eta_j) \rangle_{\mathcal{P}(\eta)} = 0$. After iterating the full set of Eqs. (31), we obtain the results shown in the bottom rows of Figs. 1 and 2. In the case with homogeneous noise, we still get a result that is similar to the Wiener filter one. This shows that we do not lose much by allowing for some uncertainty in the noise covariance. In the cases in which our original noise estimate was wrong, however, we obtain reconstructions of a much higher quality than from the critical filter. Obviously, our algorithm succeeds in uncovering the false error bars in our data set and correcting them. This works especially well in the case where only individual pixels have underestimated noise variance. This setting makes it especially easy for the algorithm to infer the signal statistics from all the other pixels and find the pixels in which the data points and the signal are inconsistent with one another. However, even in the case where one-third of the space is covered with underestimated noise, our algorithm still does a good job in reconstructing the original signal. In the spherical case, the extended critical filter performs even better than the Wiener filter. This is also true for the one-dimensional case in the scenario where the noise in individual pixels is enhanced. In the scenario with enhanced and suppressed noise in one-third of the one-dimensional interval, the Wiener filter performs better than the extended critical filter. It should be noted, however, that using the Wiener filter is not an option if the power spectrum of the signal is not known *a priori*.

Some further insight can be gained by looking at the reconstructed angular power spectra for the spherical case. These are shown in Fig. 3. In the case with homogeneous noise, the critical filter recovers the true power spectrum almost perfectly, while the extended critical filter misses some power on the smallest scales; i.e., some of the small-scale power in the data is falsely attributed to noise and therefore not represented in the signal power spectrum. This effect is small, however, and does not greatly influence the resulting map.

In the case in which the noise is highly inhomogeneous, being higher and lower in one-third of the data space each, the extended critical filter misses quite a lot of power on small scales. This results in the slightly oversmoothed map seen in Fig. 2. The critical filter, however, which operates under wrong assumptions for the noise statistics, overestimates the power on small scales significantly. This is in agreement with the very noisy reconstructed map.
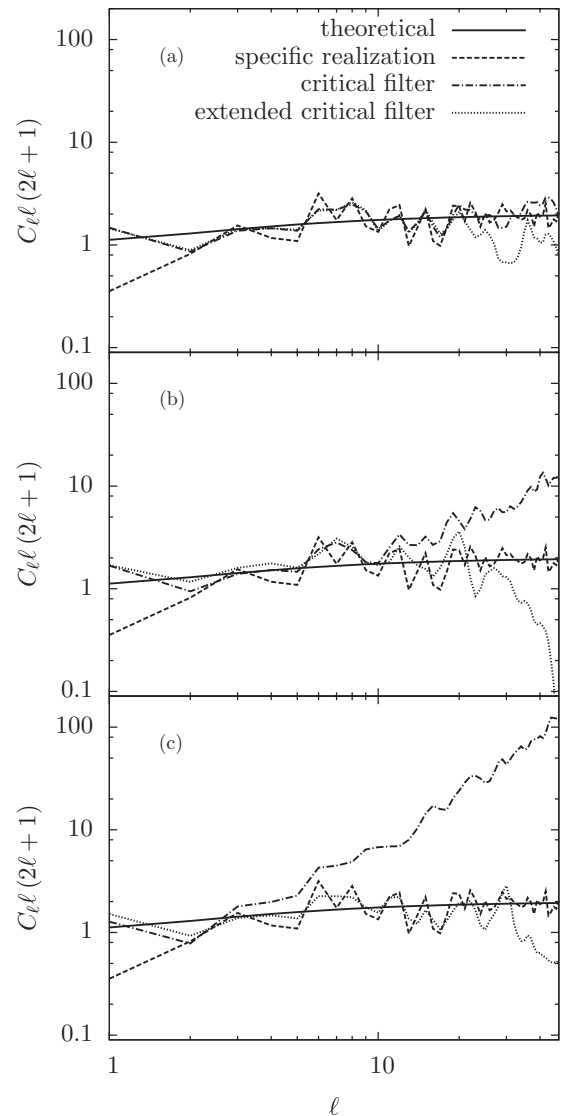

FIG. 3. Comparison of the different reconstructed angular power spectra for the spherical scenario. The solid line depicts the theoretical power spectrum, which is also used in the Wiener filter reconstructions. The dashed line corresponds to the power of the specific signal realization and the dash-dotted and dotted lines to the power spectrum reconstructed with the critical filter and the extended critical filter, respectively. Panel (a) shows the case with homogeneous noise, panel (b) the one in which the noise is enhanced and suppressed in one-third of the sphere each, and panel (c) the one in which the noise is enhanced in individual pixels.

It is in the third case, in which the noise is greatly enhanced in individual pixels, that the extended critical filter shows its full strength. While the critical filter attributes the power in the faulty pixels to the signal and therefore overestimates the signal power by orders of magnitude, the extended version accounts for the misestimated error bars and does not account for these pixels in the signal power spectrum. While the result is a power spectrum that is slightly underestimated on the smallest scales, this is again only a comparatively small error, still leading to a good reconstruction.

These findings are confirmed by Fig. 4, which shows the differences of the nine reconstructions and the signal
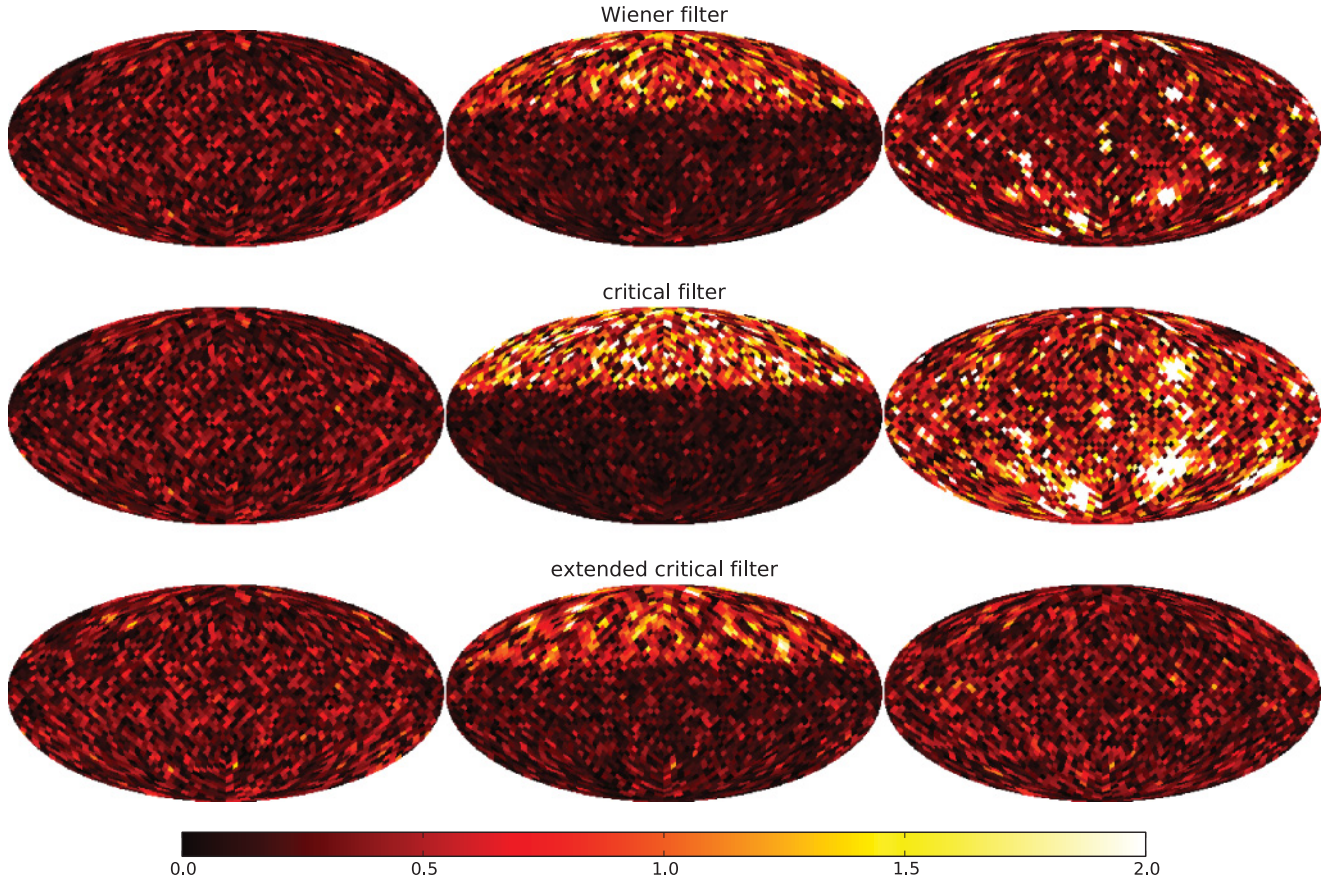
FIG. 4. (Color online) Absolute value of the pixelwise difference between the reconstructed maps and the signal realization for the spherical scenario. Each row shows the results for a different filter algorithm. As in Fig. 2, the left column shows the case with homogeneous noise, the middle column the one with enhanced noise in the northern third of the sphere and suppressed noise in the southern third, and the right column shows the case where the noise is enhanced in individual pixels. Note that the color bar differs from the one used in Fig. 2.

realization. Our extension of the critical filter clearly brings the strongest improvement in the case where the noise is enhanced in individual pixels, while also lowering the error in the case with an extended region of underestimated noise. The same can directly be seen for the one-dimensional case in Fig. 1.

Finally, we plot the standard deviation per pixel of the Gaussian approximation (20) to the posterior probability distribution, i.e., the square root of the diagonal of the covariance matrix in the pixel basis, $\sigma = \sqrt{\text{diag}(D)}$, in Fig. 5

for the extended critical filter. This can be interpreted as an estimate for the 1σ error bar of the reconstructed maps. The region with enhanced noise in the second scenario is clearly marked out by a higher uncertainty of the reconstruction due to the corrected entries of the noise covariance matrix entering the information propagator D. Note, however, that the full posterior is non-Gaussian and the 1σ range can therefore not necessarily be interpreted as a 68% confidence interval, especially since we are using a zeroth-order approximation to
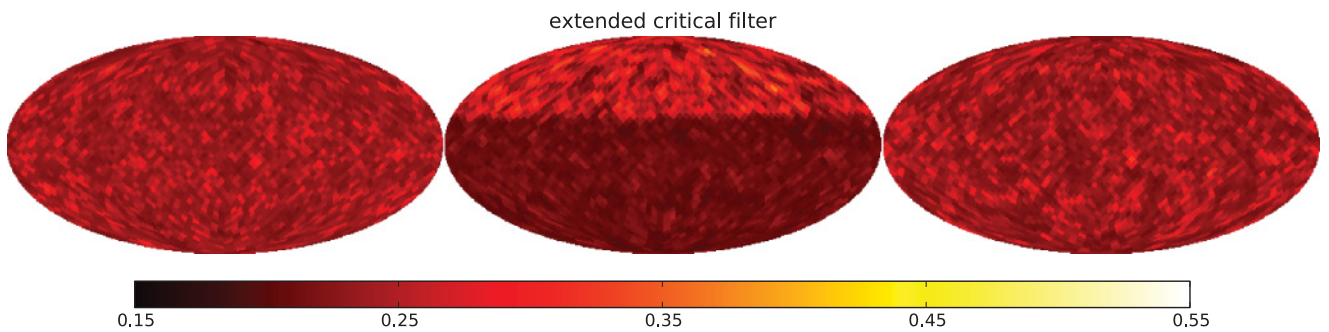

FIG. 5. (Color online) Pixelwise uncertainty of the extended critical filter reconstructions in the spherical scenario. The left panel shows the case with homogeneous noise, the middle panel the case with enhanced noise in the northern third and suppressed noise in the southern third, and the right panel the case with enhanced noise in individual pixels.
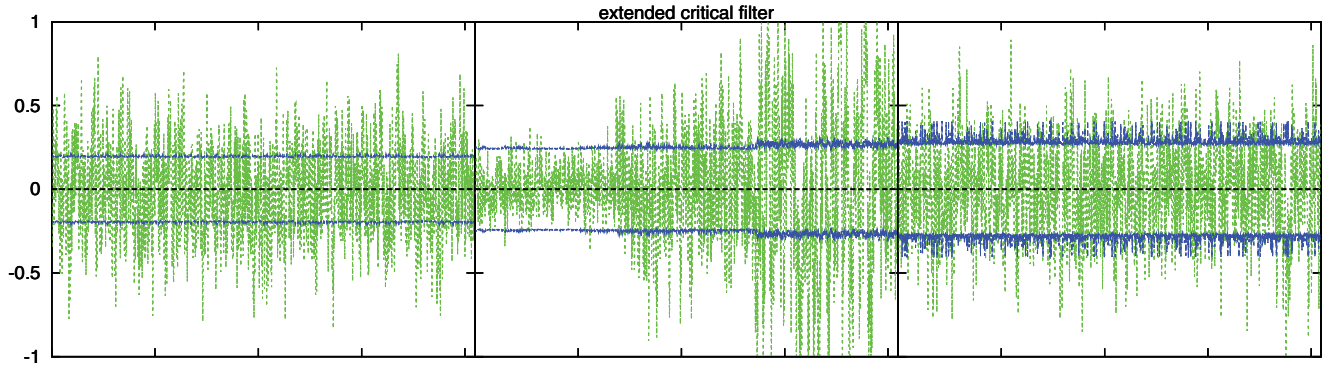
FIG. 6. (Color online) Pixelwise uncertainty of the extended critical filter reconstructions in the one-dimensional scenario. The left panel shows the case with homogeneous noise, the middle panel the case with enhanced noise in the right third and suppressed noise in the left third, and the right panel the case with enhanced noise in individual pixels. The dark curves represent $\pm\sigma$ and the light curve the difference between the reconstruction and the correct signal.

calculate the Gaussian approximation. In fact, in our spherical example only about 50% of the pixels of the reconstructions lie within 1σ of the correct signal in all three noise scenarios.

Figure 6 extends this study to the one-dimensional case. We plot the difference of the signal and the reconstruction result of the extended critical filter, along with lines depicting $\pm\sigma$, for the three different noise settings. In the case where the noise variance is constant within each third of the interval, the one-sigma curve exhibits a steplike behavior at the boundaries of the thirds, although this effect is relatively small. In the case of homogeneous noise, the 1σ curve is roughly constant while the individual noisy pixels in the last scenario are reflected in the 1σ curve by its large variations from pixel to pixel. The fraction of the pixels for which the reconstruction lies within 1σ of the correct signal in the one-dimensional case is 50% in the case with homogeneous noise, 63% in the case with enhanced and suppressed noise in a third of the interval each, and 69% in the case with enhanced noise in individual pixels.

V. DISCUSSION

Using the formalism of minimum Gibbs free energy we have extended the critical filter algorithm, developed in Refs. [1] and [2], to an algorithm that allows for uncertainties both in the signal covariance and in the noise covariance. We have demonstrated the performance of our algorithm, Eqs. (31), by applying it to a set of mock observations on the sphere, as well as in a simple one-dimensional test case.

These applications have shown that the extended critical filter performs outstandingly if only a few individual data points have a misestimated error bar. However, even in a case where large portions of the data are affected, the algorithm was shown to perform inarguably better than the critical filter, using a fixed—and faulty—assumption about the noise statistics. We have also compared the results to those obtained from a Wiener filter reconstruction, using the correct power spectrum, which is known to be optimal if the assumptions about the noise statistics are correct. This filter was demonstrated, however, to lead to reconstructed maps that are much further from the true signal than the results of the extended critical filter in some cases where the assumptions are not correct.

The choice of the two-sphere as the space on which our signal is defined was motivated by astrophysical applications, where we could think of the signal as an all-sky field or some quantity defined on the surface of a star or a planet. Applications in other fields of physics are abundant. However, it should be noted that there is nothing special about the sphere. We could equally well have chosen a more-dimensional Euclidean space, using the power spectrum defined in Fourier space instead of the angular power spectrum, as we have done in the one-dimensional scenario.

Furthermore, our choice of the identity operator as response matrix was made only on account of simplicity. It allowed us to represent the data in the same fashion as the signal. It should be clear, however, that the derived filter formulas, Eqs. (31), are valid for any response matrix, even a singular one. Applications of the critical filter with nontrivial response matrices were presented in Refs. [1] and [3], and such a response would not pose a problem for the extended version of the filter.

The problem of signal reconstruction with some uncertainty in the noise variance is certainly one of general interest. There are several ways in which uncertainty in the noise variance might arise. It may be due to questionable assumptions that enter in the calculation of the error bars of the data. Another possibility is that it arises from the definition of the signal itself. The quantity of interest may only be part of what has been measured in the first place in which case the rest of the data would be noise with essentially unknown variance. All these factors come together in the reconstruction problem considered in Ref. [3]. An extension of that reconstruction, using additional sets of data and the improved algorithm presented here, is planned [11].

It should be noted, however, that even with the extended critical filter, some knowledge about either the parameters in the signal covariance or the ones in the noise covariance is needed to arrive at a sensible reconstruction. Leaving them both completely free would lead to a degeneracy between signal and noise that cannot be resolved. Only by assigning an informative prior to at least one of the two sets of parameters is this degeneracy broken. Furthermore, the functional bases in which the signal and noise covariances are diagonal, i.e., their eingenspaces, need to differ to allow for a separation of noise from signal.

APPENDIX: HIGHER-ORDER SOLUTIONS

Here we briefly list the results for nontrivial temperatures up to second order, i.e., considering terms up to $i = 2$ in Eqs. (24) and (25). Since the first-order terms are zero for our choice of $\tilde{q}_k$ and $\tilde{r}_j$, we list only the resulting filter formulas for the zeroth- and second-order internal energy.

1. Zeroth order

The zeroth-order solution with arbitrary temperature is rather similar to the one with $T = 1$ presented in Sec. III. It is given by

$$m = D'j, \tag{A1}$$

$$j = \sum_j \frac{\delta_j}{\tilde{r}_j} R^\dagger N_j^{-1} d, \tag{A2}$$

$$D' = \left(\sum_k \frac{\gamma_k}{\tilde{q}_k} S_k^{-1} + \sum_j \frac{\delta_j}{\tilde{r}_j} R^\dagger N_j^{-1} R \right)^{-1} \tag{A3}$$

$$D = T D'. \tag{A4}$$

The mean m is completely unchanged. However, the covariance D of the Gaussian approximation is now T times the information propagator D'; i.e., the Gaussian approximation becomes wider at higher temperature.

2. Second order

The second-order solution is given by

$$m = D'j, \tag{A5}$$

$$j = \sum_j \frac{\delta_j}{\tilde{r}_j} Y_j R^\dagger N_j^{-1} d, \tag{A6}$$

$$D' = \left(\sum_k \frac{\gamma_k}{\tilde{q}_k} X_k + \sum_j \frac{\delta_j}{\tilde{r}_j} Y_j R^\dagger N_j^{-1} R \right)^{-1}, \tag{A7}$$

$$X_k = 1 + \frac{1}{\tilde{q}_k^2} \mathrm{tr} \left[\left(mm^\dagger + \frac{1}{2} D \right) S_k^{-1} D S_k^{-1} \right] - \frac{1}{\tilde{q}_k} S_k^{-1} D, \tag{A8}$$

$$Y_j = 1 + \frac{1}{\tilde{r}_j^2} \mathrm{tr} \left\{ \left[(d - Rm)(d - Rm)^\dagger + \frac{1}{2} R D R^\dagger \right] \right.$$
$$\left. \times N_j^{-1} R D R^\dagger N_j^{-1} \right\} - \frac{1}{\tilde{r}_j} R^\dagger N_j^{-1} R D, \tag{A9}$$

$$D = T \left\{ D'^{-1} - \sum_k \frac{\gamma_k}{\tilde{q}_k^2} S_k^{-1} (mm^\dagger) S_k^{-1} - \sum_j \frac{\delta_j}{\tilde{r}_j^2} R^\dagger N_j^{-1} \right.$$
$$\left. \times [(d - Rm)(d - Rm)^\dagger] N_j^{-1} R \right\}^{-1}. \tag{A10}$$

Again, the only effect of the temperature is to broaden the approximate Gaussian. However, in the second-order solution the operators X_k and Y_j appear, destroying the one-to-one correspondence between the terms in these expressions for D, D', and j and the Wiener filter formula, Eq. (8). Therefore, the values of the parameters p_k and η_j are not immediately determined by these equations. Note, however, that the goal was not to determine the signal and noise covariance matrices, but to find the optimal Gaussian approximation to the signal posterior, given by m and D.

[1] T. A. Enßlin and M. Frommert, Phys. Rev. D **83**, 105014 (2011).

[2] T. A. Enßlin and C. Weig, Phys. Rev. E **82**, 051112 (2010).

[3] N. Oppermann, H. Junklewitz, G. Robbers, and T. A. Enßlin, Astron. Astrophys. **530**, A89 (2011).

[4] S. I. Olsen, CVGIP: Graph. Models Image Proc. **55**, 319 (1993).

[5] J.-L. Starck and F. Murtagh, Publ. Astron. Soc. Pac. **110**, 193 (1998).

[6] V. I. Koltchinskii, Ann. Stat. **28**, 591 (2000).

[7] M. Schwarz and S. van Bellegem, Stat. Prob. Lett. **80**, 236 (2010).

[8] T. A. Enßlin, M. Frommert, and F. S. Kitaura, Phys. Rev. D **80**, 105005 (2009).

[9] S. Kullback and R. A. Leibler, Ann. Math. Stat. **22**, 79 (1951).

[10] The HEALPIX package is available from http://healpix.jpl.nasa.gov.

[11] N. Oppermann *et al.* (unpublished).

[12] K. M. Górski, E. Hivon, A. J. Banday, B. D. Wandelt, F. K. Hansen, M. Reinecke, and M. Bartelmann, Astrophys. J. **622**, 759 (2005).

[13] Sage, Sage mathematical software, version 4.3.3 (Feb. 2010), [http://www.sagemath.org].

[14] Note that since our signal is real, the Fourier components associated with negative k-values do not contain additional degrees of freedom.