Reconstructing signals from noisy data with unknown signal and noise covariance

N. Oppermann, G. Robbers and T.A. Enßlin

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

Abstract. The reconstruction of spatially or temporally distributed signals from a data set can benefit from using information on the signal's correlation structure. In this work, we present a method, developed within the novel discipline of *Information Field Theory*, that uses this correlation information, even though it is not a priori available. The method is furthermore robust against outliers in the data since it allows for errors in the error estimates themselves. The algorithm is derived in a fully Bayesian way making use of tools from statistical mechanics. We demonstrate its performance by applying it to an astrophysical map-making problem, the reconstruction of an all-sky image of the Galactic Faraday depth. We briefly discuss the underlying physical processes.

Keywords: statistics; information theory; statistical mechanics; interstellar medium: magnetic field
PACS: 02.50.Tt, 89.70.-a, 98.35.Eg, 98.38.Am

1. INTRODUCTION

The problems of data analysis and signal reconstruction permeate all branches of science. Whenever observations are made, i.e. data are taken, the question arises how to infer some underlying quantity that is ultimately of interest from the observed data. Of course, this is only possible if there is some relationship between the data d and the underlying signal s,

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

The relationship f, however, is not necessarily invertible and might contain some stochastic processes. Here, we will specifically deal with the situation in which the signal of interest is a field on some manifold. We therefore develop our reconstruction algorithm within the framework of *information field theory* [1] (see also the article by Enßlin in this volume). We further restrict ourselves to cases in which the data arise from the signal in a linear way, subject to additional noise,

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

We refer to the linear operator R as the response operator and n is the additive noise term. This linear data model is, however, general enough to incorporate a vast array of observational situations, such as point measurements, convolutions of the signal field, interferometric observations, etc.

The problem of inferring s from d is twofold. In cases where the response operator is simple (e.g. if we measure the signal directly, $R = \mathbb{1}$), we need to filter out the noise from the data. In order to do so, Bayesian probability theory tells us to make use of prior knowledge that we have about the noise, such as an estimate of its variance. If, on the other hand, the noise contribution to the data is negligible, we simply have to invert the response operator. This, however, is not always possible. In fact, if s is a field containing an infinite number of degrees of freedom, this is never possible, since any data set only ever contains a finite amount of information. So in addition to the prior knowledge about the noise, we need to make use of some prior knowledge about the signal as well, such as an energy that tends to be minimized by the types of signals under considereation. In a general situation, both complications apply.

In this work, we will focus on cases in which Gaussian priors can be used both for the noise n and the signal field s. These priors are excellent approximations to the truth in many cases. They can also be derived from the maximum entropy principle in cases where only the first and second moments of the signal and noise, i.e. their means and covariances, are known [e.g. 2, 3]. However, we very often encounter cases in which we don't know these moments, especially the covariances, with certainty. The question we address here is how to reconstruct a signal from a data set following the relation (2) using Gaussian priors whose covariances are themselves uncertain.

We outline the derivation of the *extended critical filter* algorithm [4] which is based on an a priori assumption of Gaussianity both for the noise and the signal field, as well as the existence of known symmetries in their correlation structures, in Sect. 2. We furthermore demonstrate the application of the algorithm to an astrophysical map making problem, the reconstruction of an all-sky image of the Galactic Faraday depth from point-like measurements of extragalactic Faraday depths [5], in Sect. 3. We conclude in Sect. 4.

2. THE ALGORITHM

2.1. Assumptions

The basic assumption behind the *extended critical filter* algorithm is that both the signal s and the noise n in Eq. (2) can be assumed to follow Gaussian prior statistics,

$$\mathscr{P}(\phi) = \mathscr{G}(\phi,\Phi) = \frac{1}{|2\pi\Phi|^{1/2}} \exp\left(-\frac{1}{2}\phi^{\dagger}\Phi^{-1}\phi\right), \tag{3}$$

where ϕ stands for either s or n, $\mathscr{G}(\phi,\Phi)$ denotes a zero-mean Gaussian in ϕ with covariance Φ, and the $\dagger$-symbol denotes the conjugate-transpose operator, so that the scalar product can be written as

$$\phi^{\dagger}\psi = \int \mathrm{d}x\, \bar{\phi}(x)\psi(x), \tag{4}$$

where the bar denotes complex conjugation and the integration is performed over the whole domain of the fields ϕ and ψ.

Both the signal and noise covariance matrices are generally unknown. We assume that each can be decomposed into a sum of unknown parameters and known matrices,

$$\Phi = \sum_{k=0}^{k_{\max}} p_k^{(\phi)} \Phi_k, \tag{5}$$

The same expansion can be written down for the inverse covariance matrices,

$$\Phi^{-1} = \sum_{k=0}^{k_{\max}} \left(p_k^{(\phi)}\right)^{-1} \Phi_k^{-1}, \tag{6}$$

where Φ_k^{-1} denotes the pseudo-inverse of Φ_k[1], so that $\Phi_k^{-1}\Phi_k = \mathrm{id}|_{\mathrm{support}(\Phi_k)} = \Phi_k\Phi_k^{-1}$ and $\Phi_j^{-1}\Phi_k = 0 = \Phi_k\Phi_j^{-1}$ for $j \neq k$. This turns the problem of unknown covariance matrices into a problem of – potentially few – unknown parameters $p_k^{(s)}$ and $p_k^{(n)}$.

For these unknown parameters, we assume a priori statistical independence and inverse gamma priors for each individual parameter,

$$\mathscr{P}(p_k^{(\phi)}) = \frac{1}{q_k^{(\phi)}\Gamma(\alpha_k^{(\phi)}-1)} \left(\frac{p_k^{(\phi)}}{q_k^{(\phi)}}\right)^{-\alpha_k^{(\phi)}} \exp\left(-\frac{q_k^{(\phi)}}{p_k^{(\phi)}}\right). \tag{7}$$

The parameter $\alpha_k^{(\phi)}$ determines the slope of a power-law drop-off for this prior probability distribution at large values of $p_k^{(\phi)}$, while the parameter $q_k^{(\phi)}$ determines the location of an exponential cut-off at low values of $p_k^{(\phi)}$. By varying these parameters $\alpha_k^{(\phi)}$ and $q_k^{(\phi)}$, different states of prior knowledge for the parameters $p_k^{(\phi)}$ can be modeled. In the limit of $q_k^{(\phi)} \to 0$ and $\alpha_k^{(\phi)} \to 1$ this inverse gamma prior turns into Jeffreys prior which is uninformative on a logarithmic scale.

2.2. Derivation

Using the assumptions made in Sect. 2.1, it is straightforward to calculate the joint probability distribution of signal s and data d by marginalizing over the unknown parameters $\left(p_k^{(s)}\right)_k$ of the signal covariance matrix and $\left(p_i^{(n)}\right)_i$ of the noise covariance matrix. The result is

$$\mathscr{P}(d,s) = \left(\prod_{k=0}^{k_{\max}} \frac{\Gamma(\gamma_k^{(s)})\left(q_k^{(s)}\right)^{\alpha_k^{(s)}-1}}{\Gamma(\alpha_k^{(s)})\,(2\pi)^{\rho_k^{(s)}/2}} \left(q_k^{(s)} + \frac{1}{2}s^\dagger S_k^{-1} s\right)^{-\gamma_k^{(s)}}\right)$$
$$\left(\prod_{i=0}^{i_{\max}} \frac{\Gamma(\gamma_i^{(n)})\left(q_i^{(n)}\right)^{\alpha_i^{(n)}-1}}{\Gamma(\alpha_i^{(n)})\,(2\pi)^{\rho_i^{(n)}/2}} \left(q_i^{(n)} + \frac{1}{2}(d-Rs)^\dagger N_i^{-1}(d-Rs)\right)^{-\gamma_i^{(n)}}\right). \tag{8}$$

[1] We use the pseudo-inverse since the matrices Φ_k might well be singular.

Here, $\gamma_k^{(\phi)} = \alpha_k^{(\phi)} - 1 + \rho_k^{(\phi)}/2$ and $\rho_k^{(\phi)} = \mathrm{tr}\left(\Phi_k^{-1}\Phi_k\right)$. The posterior for the signal, $\mathscr{P}(s|d) = \mathscr{P}(d,s)/\mathscr{P}(d)$, is proportional to this joint probability distribution.

This posterior describes our knowledge state about the signal after taking the data. It is, however, hard to interpret since it is obviously non-Gaussian and a calculation of its moments is in general not possible analytically. We therefore aim to approximate it with a Gaussian that captures most of the information contained in the full posterior while being easy to interpret and analytically manageable. In order to find the best Gaussian approximation, we use the formalism of Gibbs free energy minimization presented in [6]. This procedure leads to equations for the mean and covariance of. In these expressions, linear combinations of the matrices $(\Phi_k)_k$ appear. If one reidentifies the coefficients in these linear combinations with the coefficients $\left(p_k^{(\phi)}\right)_k$ of the decomposition of the covariance matrices, the expressions for the mean and covariance turn out to be exactly the same as for a Gaussian posterior in a case of fully known signal and noise covariances. The mean m and covariance D of the Gaussian and these parameters are then given by

$$m = DR^\dagger \left(\sum_{i=0}^{i_{\max}} \left(p_i^{(n)}\right)^{-1} N_i^{-1} \right) d, \tag{9a}$$

$$D = \left(\left(\sum_{k=0}^{k_{\max}} \left(p_k^{(s)}\right)^{-1} S_k^{-1} \right) + R^\dagger \left(\sum_{i=0}^{i_{\max}} \left(p_i^{(n)}\right)^{-1} N_i^{-1} \right) R \right)^{-1}, \tag{9b}$$

$$p_k^{(s)} = \frac{1}{\gamma_k^{(s)}} \left(q_k^{(s)} + \frac{1}{2}\mathrm{tr}\left(\left(mm^\dagger + D \right) S_k^{-1} \right) \right), \tag{9c}$$

$$p_i^{(n)} = \frac{1}{\gamma_i^{(n)}} \left(q_i^{(n)} + \frac{1}{2}\mathrm{tr}\left(\left((d - Rm)(d - Rm)^\dagger + RDR^\dagger \right) N_i^{-1} \right) \right). \tag{9d}$$

These equations can then be solved by simple iteration, yielding the Gaussian mean m as a reconstruction of the signal field and its covariance D as a measure of the reconstruction's uncertainty. We call this procedure the *extended critical filter*. While no formal proof for the convergence of this iteration exists, it was shown in [4] with mock observations, that the correct solution can be expected to be found.

2.3. Verification

In this section we present a simple one-dimensional mock example to demonstrate the performance of the *extended critical filter*. We draw a random realization of a Gaussian signal defined on the interval $[0,128]$ and assume that it has been observed in 91 locations with a response that simply gives back the signal values at these locations. Then we add a random noise contribution to these values. These noise contributions are drawn independently from a Gaussian with unit variance, except for 10% of the data points for which we have increased the noise variance by a factor 100. This setup is shown in the left panel of Fig. 1.

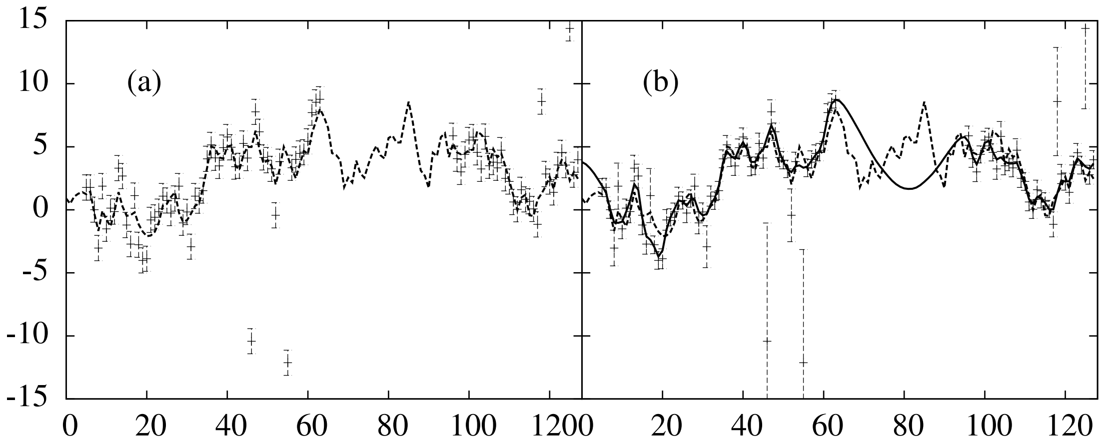

FIGURE 1. One-dimensional test case for the *extended critical filter*. Panel (a) shows the signal (dashed line) along with the data and their initially assumed error bars. Panel (b) shows the signal (dashed line), its reconstruction (solid line), as well as the data with their reconstructed error bars.

We then apply the *extended critical filter* algorithm to this problem in the following way: We inform the algorithm of the fact that the noise contributions are uncorrelated, i.e. the noise covariance matrix N is diagonal. We expand it according to Eq. (5) where the matrix N_k is a simple projection onto the k-th data point[2] and the prefactor $p_k^{(n)}$ is introduced to potentially increase (or decrease) the error variance of individual data points. Note that while the algorithm is taking into account the possibility of underestimated error bars, it is not a priori informed about the factor by which the error bars are underestimated and which of the data points are afflicted by this problem.

For the signal, we assume statistical homogeneity, so that its covariance matrix becomes diagonal in Fourier space. We are therefore able to decompose it according to Eq. (5), where the matrix S_k projects onto the k-th length scale, i.e. onto the Fourier components corresponding to a wavenumber k. The prefactor $p_k^{(s)}$ then corresponds to the k-th power spectrum component.

For the inverse gamma prior, Eq. (7), we choose parameters $q_k^{(s)} \to 0$ and $\alpha_k^{(s)} \to 1$ for the signal, modeling our complete lack of knowledge about the power spectrum on a logarithmic scale. For the noise we choose $\alpha_k^{(n)} = 2$ and adapt $q_k^{(n)}$ so that the prior expectation value of $\log\left(p_k^{(n)}\right)$ is 0. This corresponds to the expectation that our initial guess of the error variance will be roughly correct for most data points.

We then iterate Eqs. (9) until convergence, smoothing the power spectrum in each iteration step and enforcing a minimum logarithmic slope. The final estimate for the signal, m, as well as the reconstructed error bars of the data, $\left(p_k^{(n)}\right)^{1/2}$, are shown in the right panel of Fig. 1. This demonstrates that the *extended critical filter* is well suited to tackle problems like the one studied here. It has clearly identified the outliers in the data and increased their error bars accordingly. Furthermore, the reconstruction of the signal

[2] This includes a multiplication with the a priori estimate of the error variance, which is unity here.

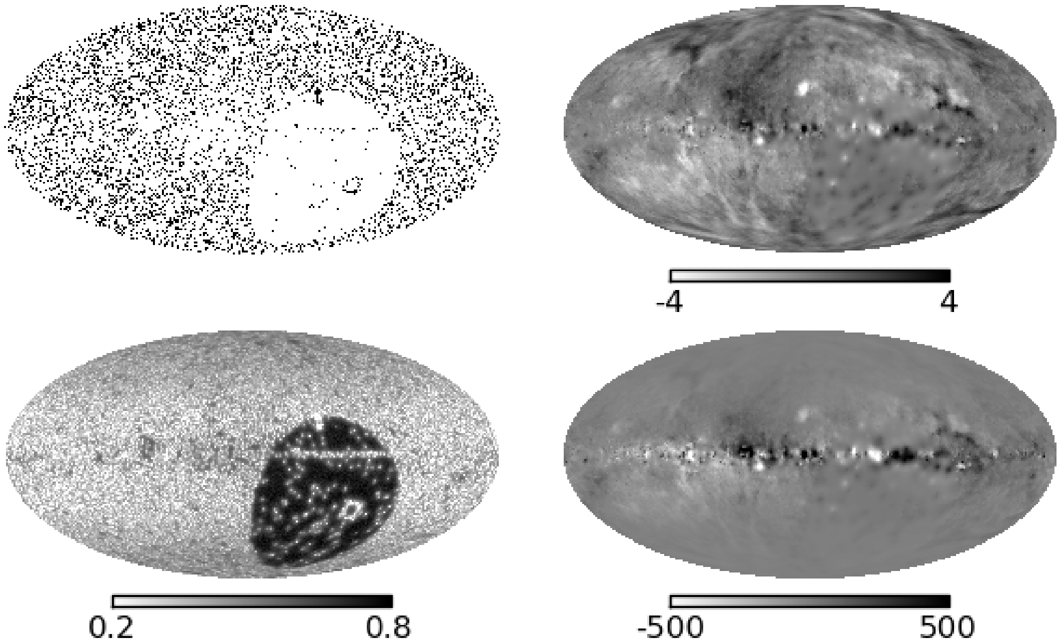

FIGURE 2. Application of the *extended critical filter* to the Galactic Faraday sky. The top left panel shows the distribution of the data points on the sky, the top right panel the reconstruction m of the signal field, the lower left panel a map of the uncertainty $\sqrt{\mathrm{diag}\,(D)}$ of this reconstruction, and the lower right panel the corresponding reconstruction of the Galactic Faraday depth ϕ in $\mathrm{rad}/\mathrm{m}^2$.

is rather close to the original signal in regions about which the data contain information and interpolates smoothly through the large region without data.

Further test case studies of the *extended critical filter* can be found in [4].

3. AN APPLICATION EXAMPLE

In this section we present the application of the *extended critical filter* to the problem of the reconstruction of an all-sky map of the Galactic Faraday depth. In radio astronomy, the Faraday depth corresponding to a physical distance r_0 from the observer is defined – up to proportionality constants – as the integral

$$\phi(r_0) \propto \int_{r_0}^{0} \mathrm{d}r\, n_{\mathrm{e}}(\vec{r}) B_r(\vec{r}) \tag{10}$$

from the distance r_0 to the observer over the density of thermal electrons n_{e} and the line-of-sight component of the magnetic field B_r. It manifests itself observationally in a wavelength-dependent rotation of the polarization plane of linearly polarized radiation. Therefore, the Faraday depth of sources of linearly polarized radiation can be estimated by measuring the position angle of the polarization plane of this radiation at different wavelengths.

The observed Faraday depths of 41 330 sources were assembeld in [5] and their positions are shown in the top left panel of Fig. 2. However, these observed values are not equal to the Faraday depth of the Milky Way, $\phi(r_{\mathrm{MW}})$, which corresponds to the integral in Eq. (10) with its lower boundary at the outer edge of the Galaxy r_{MW}, but rather contain an unknown contribution from outside the Milky Way. This extragalactic contribution is expected to be subdominant for most sources, however, it might be significant for some of them.

When reconstructing the Galactic Faraday depth, the extragalactic contributions should enter the error budget. This is one of several causes described in [5] that lead to uncertain error bars of the data. The approximation of uncorrelated noise contributions is still used and therefore the situation for the noise covariance is exactly the same as in the test case studied in Sect. 2.3, i.e. it can be split up into projections onto the data points with their initially assumed error variances and correction factors that can e.g. increase the error bars for data points with significant extragalactic contamination.

For the signal covariance, however, the assumption of statistical homogeneity made in Sect. 2.3 is no longer valid. The absolute value of the Galactic Faraday depth is expected to be larger on average for lines of sight through the Galactic plane than for lines of sight toward the Galactic poles. This is an effect that can already be seen in the raw data. The easiest way to circumvent this problem, however, is not to drop the statistical symmetry assumption, but to redefine the signal. Defining the ratio

$$s(l,b) = \frac{\phi(l,b)}{v(b)} \tag{11}$$

as the signal field, where $\phi(l,b)$ is the Galactic Faraday depth as a function of Galactic longitude l and latitude b and $v(b)$ is a root mean square profile of the variations as a function of Galactic latitude only which takes into account this largest-scale anisotropy, statistical homogeneity can still be assumed as a viable prior assumption. The signal covariance can therefore be split up in the same way as in Sect. 2.3 into projections onto length-scales and the power spectrum components.

With this signal definition, the response operator R consists on the one hand of the evaluation of the signal field at the positions on the sky where sources have been observed and on the other hand of the multiplication with the profile function $v(b)$, evaluated at the sources' latitudes. The calculation of the profile function $v(b)$ becomes an additional iteration step in the reconstruction procedure. In the first step, it is calculated directly from the data as their root mean square value within a latitude bin. Then a few reconstructions are performed using a given profile function and calculating a new profile function from the mean m and covariance D of the Gaussian approximation to the posterior according to

$$v = \sqrt{\int \mathscr{D}s\, s^2 \mathscr{G}(s-m,D)} = \sqrt{m^2 + \mathrm{diag}\,(D)}, \tag{12}$$

where again averages over bins of Galactic latitude are taken.

Figure 2 shows the results of this analysis. The top right panel shows the mean map m of the Gaussian posterior. The pixelwise uncertainty of this reconstruction is given by the diagonal of the covariance matrix D, i.e. the one-sigma uncertainty in pixel x is given by $D_{xx}^{1/2}$. This uncertainty map is shown in the lower left panel of Fig. 2. It can be clearly seen how the uncertainty of the reconstruction is lower near the positions of the data points. Finally, the best estimate of the physical Galactic Faraday depth is given by the product of the profile function v and the mean map m, shown in the lower right panel of Fig. 2. An uncertainty map for this field can be calculated analogously. More detailed results are presented in [5][3].

4. SUMMARY AND CONCLUSIONS

We have shown how Gaussian signal fields can be reconstructed from data that are linearly dependent on the signal and subject to additive Gaussian noise. The unknown covariances of signal and noise are parameterized and marginalized over, choosing appropriate priors for the unknown parameters. The resulting complex probability distribution is then matched with a Gaussian using the principle of minimum Gibbs free energy. This Gaussian is described by a mean and a covariance which are the main results of the filtering procedure. We have demonstrated how to apply this *extended critical filter* to the problem of the reconstruction of a statistically isotropic signal with unknown power spectrum from data with uncertain error bars.

This algorithm was applied to the problem of all-sky map-making of the Galactic Faraday depth, yielding a map of this quantity showing unprecedented detail, along with a measure of its uncertainty. This is, however, only one of many possible fields in which it is applicable. The generality of the assumptions leads to a high versatility of the filter; with appropriate models of the response operator and different decompositions of the signal and noise covariances, a wealth of differend scenarios can be handled.

REFERENCES

1. T. A. Enßlin, M. Frommert, and F. S. Kitaura, *Phys. Rev. D* **80**, 105005 (2009), `0806.3474`.
2. E. T. Jaynes, *Physical Review* **106**, 620–630 (1957).
3. E. T. Jaynes, *Physical Review* **108**, 171–190 (1957).
4. N. Oppermann, G. Robbers, and T. A. Enßlin, *Phys. Rev. E* **84**, 041118 (2011), `1107.2384`.
5. N. Oppermann, H. Junklewitz, G. Robbers, M. R. Bell, T. A. Enßlin, A. Bonafede, R. Braun, J. C. Brown, T. E. Clarke, I. J. Feain, B. M. Gaensler, A. Hammond, L. Harvey-Smith, G. Heald, M. Johnston-Hollitt, U. Klein, P. P. Kronberg, S. A. Mao, N. M. McClure-Griffiths, S. P. O'Sullivan, L. Pratley, T. Robishaw, S. Roy, D. H. F. M. Schnitzeler, C. Sotomayor-Beltran, J. Stevens, J. M. Stil, C. Sunstrum, A. Tanna, A. R. Taylor, and C. L. Van Eck, *A&A* **542**, A93 (2012), `1111.6186`.
6. T. A. Enßlin, and C. Weig, *Phys. Rev. E* **82**, 051112 (2010), `1004.2868`.

---

[3] The results are also available in digital form at `http://www.mpa-garching.mpg.de/ift/faraday/`