

# Improving stochastic estimates with inference methods: calculating matrix diagonals

Marco Selig, Niels Oppermann, and Torsten A. Enßlin

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

(Received 2011 August 02; Accepted 2012 February 23)

Estimating the diagonal entries of a matrix, that is not directly accessible but only available as a linear operator in the form of a computer routine, is a common necessity in many computational applications, especially in image reconstruction and statistical inference. Here, methods of statistical inference are used to improve the accuracy or the computational costs of matrix *probing* methods to estimate matrix diagonals. In particular, the generalized Wiener filter methodology, as developed within information field theory, is shown to significantly improve estimates based on only a few sampling probes, in cases in which some form of continuity of the solution can be assumed. The strength, length scale, and precise functional form of the exploited autocorrelation function of the matrix diagonal is determined from the probes themselves. The developed algorithm is successfully applied to mock and real world problems. These performance tests show that, in situations where a matrix diagonal has to be calculated from only a small number of computationally expensive probes, a speedup by a factor of 2 to 10 is possible with the proposed method.

PACS numbers: 02.50.Tt, 02.30.Zz, 89.70.Eg, 11.10.-z, 89.70.-a

Keywords: Inference methods – Inverse problems – Computational complexity – Field theory – Information and communication theory

## I. INTRODUCTION

Imagine the inside of a black box, which computes an output in a concealed way from a given input, needs to be investigated. This is a frequent computational task in image reconstruction problems and it might also be necessary in reverse engineering the functionality of a given, already compiled computer code. In this work, the black box acts as a linear operation on its high-dimensional input, which can be freely chosen in order to probe the box.

However, an internal property of the box that is of interest might be obfuscated and therefore needs to be inferred indirectly by analyzing suitable probes. In practice, the obfuscation could be just due to the complexity of the box-internal operations, as it happens in image reconstruction problems. These operations could be any combination of many possible linear operators, like scalar or componentwise multiplications, vector space rotations, Fourier transformations, linear equation solvers, etc. The output, however, is assumed to be fully available.

Mathematically, the linear operation performed by the black box can be represented by a matrix in case the input and output spaces are of finite dimension. In case these spaces have the same dimension, the matrix is a square matrix. In this paper, we are interested in the diagonal of a matrix for which input and output space can be identified with each other, e.g., in image reconstruction problems by component- or pixelwise correspondence. There, the matrix diagonal expresses how much a component of the input vector imprints onto the corresponding component of the output vector. This matrix diagonal, or any other component of the matrix representing the black box operator, can be probed by feeding a number of test vectors into the box and analyzing

the results.

The simplest, but most accurate, scheme is to probe any of the diagonal components sequentially, one by one. This can be done by choosing an input vector that contains only one nonzero component at the location of a certain pixel, and repeating this for all pixels. The matrix entries can then be read off from the output vectors. This scheme, however, is computationally very expensive if the dimension of the involved vector spaces is very high. In image reconstruction, one deals with the space of all possible images, which has the dimension of the number of image pixels. Probing these sequentially is nearly always computationally prohibitive.

A cheaper, but less accurate, scheme is to feed white noise vectors into the black box and to correlate the outputs to the inputs component- or pixelwise. Averaging these correlations over a sufficiently large set of input vectors yields an estimator for the desired matrix diagonal.

A first proposal for such a probing scheme can be found in the work by Hutchinson [1, and references therein]. There, the functionality and efficiency of probing for obtaining trace estimates has been proven. Bekas, Kokiopoulou and Saad [2] extended the probing to Hadamard vectors (rows of the Hadamard matrix) to improve the estimation of diagonals of banded matrices. Those methods have been oriented to applications in density functional theory. Similar problems were approached by Tang and Saad [3, and references therein] with a focus on nonstochastic estimators. The recent paper by Aune and Simpson [4] transfers the probing technique to the field of information theory, in particular to the calculation of log-likelihoods. Finally, in the extensive work of Rohde and Tsybakov [5], the noise corrupted observation of unknown matrix entries is investigated from a more mathematical point of view.

In these schemes, the number of required probes is much smaller than in the simple sequential probing scheme. However, this comes at the price of a lower accuracy of the result due to the stochastic nature of these schemes, which leads to sample variance or uncertainty in the result. The situation is completely analogous to a noise contaminated signal measurement. Higher accuracy can be achieved by calculating or measuring more samples, since the noise will average out in the long run. This, however, might become computationally too expensive at some point.

Noise suppression is a common theme also in the discipline of image and signal reconstruction. Some *a priori* knowledge on signal properties may exist, like the presence of spatial smoothness, or the knowledge of limited variance of certain image features (pixels, regions, Fourier modes, etc.). This knowledge can be exploited in order to reduce the impact of the noise on the reconstructed image beyond the simple noise averaging.

Here, we propose to use well-suited image reconstruction techniques, like generalizations of the Wiener filter, for an improved reconstruction of the matrix diagonals. We show that when determining the diagonals of covariance matrices, as needed in problems of image reconstruction, the usage of suitable filters can improve the accuracy of the inferred matrix diagonal significantly. The method proposed here yields more accurate results for a given amount of CPU time or, respectively, can reach a given level of accuracy earlier. In the examined applications, we found a performance increase by a factor of 2 to 10. This definitely enlarges the number of signal reconstruction problems which can be tackled, and might enable computations which were marginally prohibitive before.

The remainder of this paper is structured as follows. In Sec. II, we highlight the importance of obtaining estimates for matrix diagonals, in particular of uncertainty covariance matrices in the context of information field theory (IFT) for signal reconstruction. The stochastic probing approach is reviewed in Sec. III. In Sec. IV, we propose using filters derived within the framework of IFT to solve the problem of estimating matrix diagonals. Both methods are verified by simple mock examples as well as by a real-world problem in Sec. V. We conclude in Sec. VI.

## II. PROBLEM OF MATRIX DIAGONALS

Linear operators are fundamental in any area of computation and thereby often expressed in their matrix representation.

In the field of information theory, the covariance matrix of a quantity (which equals the inverse of the precision matrix) holds a key role. To stress this, let us

consider a multidimensional zero-mean Gaussian,

$$\mathcal{G}(\varphi, X) = \frac{1}{\sqrt{\det[2\pi X]}} \exp\left(-\frac{1}{2}\varphi^\top X^{-1}\varphi\right), \quad (1)$$

with the covariance matrix  $X = \langle \varphi\varphi^\top \rangle_{\mathcal{G}}$ , where  $\varphi$  is a random field defined over some pixelized vector space,  $\top$  denotes a transpose, and  $\langle \cdot \rangle_{\mathcal{G}}$  is the expectation value weighted by this Gaussian. (In this context “pixel” is to be understood as a discretized coordinate which elsewhere may be referred to as “grid point”, “bin” or “voxel”.)

A diagonal entry of the covariance matrix is the squared standard deviation  $\sigma_i$  assigned to pixel  $i$  and expresses the pixelwise uncertainty in  $\varphi$ ,

$$\sigma_i^2 = \langle \varphi_i^2 \rangle_{\mathcal{G}} = X_{ii}. \quad (2)$$

A sophisticated and effective reconstruction tool is the generic filter [6] that we will review in the following. We provide this review in order to further emphasize the importance and problem of obtaining matrix diagonals for stochastic inference. Furthermore, this filter forms also the basis of our proposed algorithm discussed in Sec. IV B.

### A. Generic filter

Signal inference focuses on the reconstruction of some signal  $s$  in order to explain a set of measurements  $d$ , both of which are connected by a forward data model,

$$d = R[s] + n, \quad (3)$$

where  $n$  is the noise and  $R$  a (not necessarily linear) response operator that maps from signal to data space.

The generalized Wiener filter, as derived, e.g., in Ref. [7] in a Bayesian framework, is for one thing based on a linear forward data model,

$$d = R s + n, \quad (4)$$

where the data  $d$  is a sum of the signal response  $R s$  and the noise  $n$ . In this scenario, the response is a linear operator that inherits all aspects of the signal measurement. The generalized Wiener filter arises in case one can assume a Gaussian distribution for the signal’s prior and the signal-independent noise in addition to the described forward data model,

$$P(s|S) = \mathcal{G}(s, S), \quad (5)$$

$$P(n|N) = \mathcal{G}(n, N), \quad (6)$$

where  $S$  and  $N$  stand for the signal and noise covariance matrix, respectively. The choice of zero-mean Gaussians shall solely simplify the notation at this point and does not present a restriction for the theory. In fact, choosing an appropriate nonzero mean is often reasonable, as we

demonstrate in Sec. IV B. In consequence of the forward model, the likelihood of the data given the signal and its response is modeled by the signal-independent noise distribution,<sup>1</sup>

$$\begin{aligned} P(d|s, R, N) &= P(n = d - R s | s, R, N) \\ &= \mathcal{G}(d - R s, N). \end{aligned} \quad (7)$$

The actual inverse problem of estimating the signal given the data leads to a posterior that, logically, is also a Gaussian,

$$P(s|d) = \mathcal{G}(s - m, D), \quad (8)$$

with the mean  $m$  and the covariance  $D$  that encodes the *a posteriori* signal uncertainty. The resulting filter formula, whose straightforward derivation is detailed in Refs. [6–8], reads

$$m = \underbrace{(S^{-1} + R^T N^{-1} R)^{-1}}_D \underbrace{(R^T N^{-1} d)}_j, \quad (9)$$

where the map  $m$  is the Bayesian estimator for the signal, i.e. its posterior mean,  $D$  is referred to as *information propagator* and  $j$  as *information source* in IFT [7] in close analogy to quantum field theory. Both the signal covariance  $S$  and the noise covariance  $N$  needed for this filter are here assumed to be known. A convenient description of these covariances is in terms of the power spectra, the spectra of the eigenvalues of these matrices, as well as the eigenbases in which the covariance matrices are diagonal.

In the following, we decompose the signal covariance as  $S = \sum_l C_l S_l$ , where the  $S_l$  are projection operators that project onto eigenspaces, the spectral bands (which correspond to the spherical harmonics of equal degree  $l$  in our examples in Sec. V). An analogous decomposition exists for the inverse  $S^{-1} = \sum_l C_l^{-1} S_l^{-1}$ , where  $S_l^{-1}$  is the pseudoinverse of  $S_l$ .

The signal’s power spectrum might be unknown *a priori*, whereas the eigenbasis can often be guessed from statistical symmetries (e.g., the spherical harmonics basis in case of a statistically isotropic distribution on the sphere). Thus the spectral coefficients  $C_l$  allow for a parametrization of the covariance matrix. In such applications without spectral knowledge, the generalized Wiener filter can be extended to a generic filter derived in Ref. [6]. For this purpose, first, a logarithmically flat prior is assumed for the unknown spectral coefficients. Second, the  $C_l$ -marginalized posterior for the signal is calculated. This posterior then allows for the reidentification of appropriate terms with the spectral coefficients.<sup>2</sup>

The resulting generic filter formulas are Eq. (9) complemented by a reconstruction rule for the power spectrum, i.e., for each spectral coefficient one calculates

$$C_l = \frac{1}{\varrho_l + 2\epsilon_l} \text{tr} [(mm^T + \delta_l D) S_l^{-1}], \quad (10)$$

where  $\varrho_l = \text{tr} [S_l^{-1} S_l]$  are the numbers of degrees of freedom for each spectral band. The parameters  $(\delta_l, \epsilon_l)$  characterize the different filter options: two specific forms are the *classical filter*, for which one chooses  $(\delta_l, \epsilon_l) = (0, 0)$ , and the *critical filter*, for which  $(\delta_l, \epsilon_l) = (1, 0)$ . The former can be derived from a “classical” maximum *a posteriori* approximation of the spectral uncertainty marginalized problem. The latter is called “critical” because it exhibits (in contrast to the classical filter) only a marginal perception threshold. For a filter with a perception threshold, the signal-to-noise ratio of a spectral mode has to exceed a certain threshold in the data before the filter recognizes it at all. There exists a critical line in the  $\delta$ - $\epsilon$ -plane separating filters that fully suppress bands with insufficient spectral power from filters that do not. The critical filter resides exactly on this line, while the classical filter is in the region with such a perception threshold.

All in all, Eqs. (9) and (10) provide an iterative scheme for the full inverse problem of signal reconstruction with unknown power spectrum, i.e., unknown correlation structure. The signal reconstruction benefits from the additional spectral information recovered from the data since it encodes internal structure of the signal. If, in addition, one lacks the *a priori* knowledge about the correlation structure of the noise, an analogous approach is conceivable; cf. [9].

The result of the inference is, first, a map  $m$  describing the posterior mean field of the signal in order to explain the given data  $d$  and, second, a covariance matrix  $D$  that encodes the underlying uncertainty correlation structure. The matrix diagonal of  $D$  is thereby of special importance, since it expresses the pixelwise variance, cf. Eq. (2), allowing one to make a statement about the uncertainty of the result in each pixel.

In order to apply the critical or other generic filters we may need to calculate the trace of  $D S_l^{-1}$  in Eq. (10) in each iteration, and we have to evaluate the diagonal of  $D$  in order to interpret the reliability of our results. This motivates our ambition to develop faster and more accurate matrix probing schemes.

Generic filters are applied, e.g., in Refs. [6, 8–11].

## B. Exact matrix diagonal

The diagonal of the uncertainty covariance  $D$  is a quantity of interest, but unfortunately not directly accessible in most cases. Its calculation involves complex matrix operations, such as matrix inversion; see Eq. (9). Often, the complete matrix is not known explicitly, only the matrix-vector multiplication is available as a black

<sup>1</sup> All noise contributions are marginalized over in an intermediate step,  $P(d|s, R, N) = \int \mathcal{D}n P(d|s, n, R) P(n|N)$ , where  $P(d|s, n, R) = \delta(d - R s - n)$  according to Eq. (4).

<sup>2</sup> Equation (10) (with  $\delta_l = 1$ ) can alternatively be derived as the maximization of a signal-marginalized posterior with respect to (the logarithm of) the spectral coefficients.

box in the form of a computer routine which reads in and returns a vector.

Calculating the diagonal of a matrix  $X$  of dimension  $r$  seems still possible using the canonical basis vectors  $e^{(k)}$  (with  $e_i^{(k)} = \delta_{ik} \forall i, k \in \{1, \dots, r\}$ ), since

$$\text{diag}[X] = \sum_k e^{(k)} * X e^{(k)}, \quad (11)$$

where  $*$  denotes a componentwise product in the way that  $(a * b)_i = a_i b_i \forall i \in \{1, \dots, r\}$ .

It is obvious that this “true” diagonal is too expensive computationally because one needs to evaluate the matrix-vector multiplication exactly  $r$  times looping through all canonical basis vectors where the dimension  $r$  of the problem can be very high ( $r \gg 1$ ). In addition, each of those products alone can be expensive because it may invoke numerical inversion techniques, e.g., a conjugate gradient scheme [12], which is the case in most of the examples in Sec. V.

### III. PROBING ESTIMATE

The question arises if one can choose another set of vectors instead of the full set of canonical basis vectors to speed up the computation. Independent and identically distributed random variables stored in a set of vectors  $\{\xi\}$  (with sample size  $|\{\xi\}| = A$ ) work out if they fulfill the property

$$\langle \xi_i \xi_j \rangle_{\{\xi\}} \xrightarrow{A \rightarrow \infty} \delta_{ij}. \quad (12)$$

The average  $\langle \cdot \rangle_{\{\xi\}}$  stands for the arithmetic mean over a set  $\{\xi\}$  and  $\delta_{ij}$  for the Kronecker delta.

Two of many possible options are (i) equally probable values of  $\pm 1$  for the components of  $\xi$  [1]<sup>3</sup> or (ii) zero-mean Gaussian random numbers with unit variance. Both were originally developed for trace estimation. We use (ii) in the following applications.

Regardless of the choice of the random vectors, the sample average

$$\langle \xi * X \xi \rangle_{\{\xi\}} \xrightarrow{A \rightarrow \infty} \text{diag}[X] \quad (13)$$

over an infinite set results in the “true” diagonal; see the Appendix A.

The average over a finite but sufficiently large set ( $A < r < \infty$ ) gives the probing estimate  $f$  of the matrix diagonal,

$$\text{diag}[X] \approx \langle \xi * X \xi \rangle_{\{\xi\}} = f. \quad (14)$$

Given this estimator, a trace estimate is obtained by summing up all elements of  $f$ , as

$$\text{tr}[X] \approx \langle \xi^\top X \xi \rangle_{\{\xi\}} = \sum_i f_i. \quad (15)$$

Since one wants to obtain an estimator in a finite period of time, one has to find an acceptable trade-off between the sample size  $A$  and the residual error, where the latter scales with  $1/\sqrt{A}$  according to the law of large numbers. Aiming for a certain precision, therefore, requires a particular amount of computation time.

The estimator given by Eq. (14) is absolutely generic and applicable to a variety of matrices. Recent applications of it can be found in Refs. [2, 4, 10].

## IV. BAYESIAN ESTIMATE

### A. Forward model

Instead of doing a simple averaging of the probes, we now want to develop a Bayesian estimate which exploits additional knowledge of the problem to infer the matrix diagonal from a smaller set of samples. For this purpose, we consider the sampling described by Eq. (14) as a linear forward model of a measurement process for the signal  $\tilde{s} = \text{diag}[X]$  we are interested in. (In order to avoid confusion, already introduced synonymous quantities that appear now in another context are marked with a tilde.)

For one sample,  $a \in \{1, \dots, A\}$ , the measurement equation takes the form

$$\begin{aligned} \tilde{d}^{(a)} &= \xi^{(a)} * X \xi^{(a)} \\ &= \text{diag} \left[ \underbrace{(\xi_1^{(a)})^2, \dots, (\xi_r^{(a)})^2}_{\tilde{R}^{(a)}} \right] \tilde{s} + \tilde{n}^{(a)}. \end{aligned} \quad (16)$$

For all samples, it is

$$\begin{aligned} \tilde{d} &= \left( \tilde{d}^{(1)}, \dots, \tilde{d}^{(A)} \right)^\top \\ &= \underbrace{\left( \tilde{R}^{(1)}, \dots, \tilde{R}^{(A)} \right)^\top}_{\tilde{R}} \tilde{s} + \underbrace{\left( \tilde{n}^{(1)}, \dots, \tilde{n}^{(A)} \right)^\top}_{\tilde{n}} \\ &= \tilde{R} \tilde{s} + \tilde{n}, \end{aligned} \quad (17)$$

where  $\tilde{d}$  represents the “measured” data,  $\tilde{R}$  the signal response, and  $\tilde{n}$  the noise. The contributions from all off-diagonal matrix elements are considered to be noise, i.e.,

$$\tilde{n}^{(a)} = \xi^{(a)} * (X - \text{diag}[X_{11}, \dots, X_{rr}]) \xi^{(a)}, \quad (18)$$

and they can be estimated using Eq. (17), once we have an estimator for the signal.

Note that, if one chooses the random variables  $\xi$  to be  $\pm 1$ , first, one does not have to draw normal variables as

<sup>3</sup> In Ref. [2], a much more sophisticated choice, based on Ref. [1], is presented.

originally pointed out by Ref. [1] and, second, all the response matrices  $\tilde{R}^{(a)}$  equal  $\mathbb{1}$  and hence do not need to be treated separately for the different samples. This speeds up the algorithm and reduces the memory requirements.

## B. Proposed algorithm

Our goal is to find an estimator for the matrix diagonal which is close to the minimum mean square error estimate, but still computationally affordable. This estimator has to account for our missing knowledge about the underlying correlation structure. Given these requirements, the generic filter formulas are potentially an appropriate choice. Therefore, our proposed algorithm is based on this filter.

We start by probing the matrix as described in Sec. III and as a result obtain a first estimator  $f$  for our signal, i.e., the matrix diagonal. This additional information changes our state of knowledge about the matrix diagonal in the way that the assumed prior in Eq. (5) is not adequate. Although  $f$  is a sufficient starting value for the iterative part of the scheme, it is not suitable for the construction of a prior, since, after only a few samples, several diagonal entries of the matrix may still be considerably over- or underestimated. By contrast,  $f$  provides already a sufficiently accurate estimator for the trace; see Eq. (15). For that reason, we can *a priori* expect the matrix diagonal  $\tilde{s}$  to be distributed around some  $\tilde{t}$  rather than around zero, where for all  $i \in \{1, \dots, r\}$  we set  $\tilde{t}_i = \sum_j f_j / r \approx \langle \text{tr}[X] \rangle_{\{\xi\}} / \dim[X]$ . Therefore, the prior of the matrix diagonal is chosen to have a nonzero mean  $\tilde{t}$ ,

$$P(\tilde{s}, \tilde{S}) = \mathcal{G}(\tilde{s} - \tilde{t}, \tilde{S}). \quad (19)$$

As a consequence, the filter formulas Eqs. (9) and (10) undergo a shift,

$$\tilde{m} = \tilde{D} \left( \tilde{R}^\top \tilde{N}^{-1} \tilde{d} + \tilde{S}^{-1} \tilde{t} \right), \quad (20)$$

$$\tilde{D} = \left( \tilde{S}^{-1} + \tilde{R}^\top \tilde{N}^{-1} \tilde{R} \right)^{-1}, \quad (21)$$

$$\tilde{C}_l = \frac{1}{\tilde{\varrho}_l + 2\tilde{\epsilon}_l} \text{tr} \left[ \left( (\tilde{m} - \tilde{t}) (\tilde{m} - \tilde{t})^\top + \tilde{\delta}_l \tilde{D} \right) \tilde{S}_l^{-1} \right]. \quad (22)$$

Furthermore, the noise covariance, i.e., its required inverse, is unknown *a priori* and needs to be estimated for our algorithm. If we use the data model described in Sec. IV A,  $\tilde{N}^{-1}$  can be approximated by the noise given the data and an estimator for the signal,

$$\tilde{n} = \tilde{d} - \tilde{R} \tilde{m}. \quad (23)$$

We simplify our calculation by using

$$\tilde{N}^{-1} = (\tilde{n} \tilde{n}^\top)^{-1} \approx (\text{diag}[\tilde{n} * \tilde{n}])^{-1}. \quad (24)$$

This is done in order to limit the computational effort, and it can be shown that this corresponds to the treatment of an unknown noise covariance presented in Ref.[9] by means of a classical filter.

Equations (20) to (22) are solved iteratively in the following scheme:

- (1) Start with  $\tilde{m}^{(\nu=0)} = f$  according to Eq. (14).
- (2) Compute  $\tilde{n}^{(\nu+1)}$  according to Eq. (23).
- (3) Compute  $\tilde{C}_l^{(\nu+1)}$  according to Eq. (22), while ignoring  $\tilde{t}$  and  $\tilde{D}$  for  $\nu = 0$ .
- (4) Compute  $\tilde{m}^{(\nu+1)}$  according to Eq. (20) using Eqs. (21) and (24).
- (5) Repeat steps (2) to (4) until convergence.

As an initial guess for the power spectrum in step (3), we use an overestimation. This accelerates the convergence process as can be seen in the extreme limits:  $\tilde{C}_l \rightarrow \infty : \tilde{m} \sim \tilde{R}^{-1} \tilde{d}$ , whereas  $\tilde{C}_l \rightarrow 0^+ : \tilde{m} \sim \tilde{t}$ , i.e., a strong overestimate still gives a nontrivial result for  $\tilde{m}$ , whereas a strong underestimate gives a nearly trivial one.

Following Sec. II A, we generally recommend the critical filter, since it does not exhibit a significant perception threshold. Nevertheless, in the presented examples, the correction term  $\text{tr}[\tilde{D} \tilde{S}_l^{-1}]$  contributes only marginally to the accuracy and therefore the classical filter, which does not require the calculation of this term, is applied in the following.

## V. VERIFICATION & APPLICATION

### A. Numerical experiments

To verify the proposed algorithm, we perform some numerical experiments that are posed on signals living on the sphere. The examples in this section are represented by all-sky HEALPIX [13, 14] maps with  $N_{\text{side}} = 8$ , resulting in  $r = 768$  pixels and in a maximal spectral index  $l_{\text{max}} = 23$  of the spherical harmonics basis in which we *a priori* assume our signal covariance to be diagonal due to a statistical isotropy of the signal.

The computations were performed using the free open-source mathematics software system SAGE [15] and the HEALPIX package.

#### 1. Trivial case

At first, we consider a trivial case where the matrix in question is given explicitly. To ensure that this matrix is a valid covariance matrix, i.e., it is positive and symmetric, we constructed it to be

$$X = \begin{pmatrix} X_{11} & -1 & & 0 \\ -1 & \ddots & \ddots & \\ & \ddots & \ddots & -1 \\ 0 & & -1 & X_{rr} \end{pmatrix}, \quad (25)$$

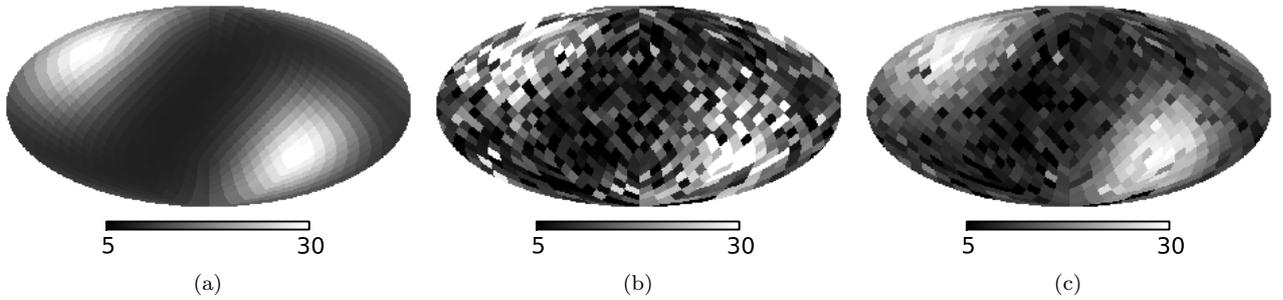


Figure 1: Trivial case: (a) the exact matrix diagonal, (b) the probing estimate, and (c) the Bayesian estimate started with four probes, both after around 0.3 s.

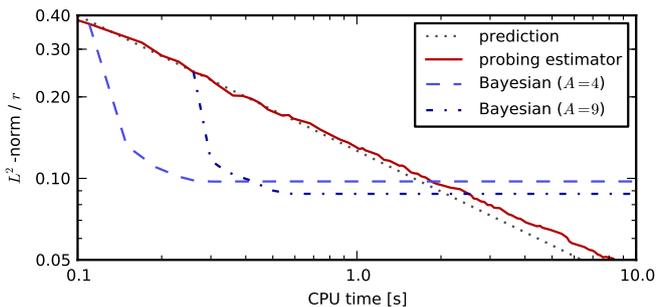


Figure 2: (Color online)  $L^2$ -norm of the error (divided by the number of pixels  $r$ ) as a function of CPU time for the trivial case: the evolution of the probing estimator (solid), its theoretical prediction  $\propto 1/\sqrt{A}$  (dotted) and the Bayesian estimators starting with four (dashed) and nine samples (dash-dotted) are shown.

where the diagonal entries need to fulfill  $X_{ii} \geq 2 \forall i \in \{1, \dots, r\}$  for positive definiteness and are drawn with a simple structure on the sphere; see Fig. 1.

The normalized  $L^2$ -norms of the residual error<sup>4</sup> serve as an accuracy measure and are shown as a function of CPU time in Fig. 2.

Although  $X$  as an operator could be implemented very efficiently, we use the much more expensive full matrix multiplication to have realistic computational costs like those of applying a more complex matrix. But this trivial case should only hold as a proof of concept; more sensible examples are discussed in the following.

As one can clearly see in Fig. 2, the probing estimator improves continuously with an increasing number of probes and shows an overall proportionality to  $1/\sqrt{A}$  as argued in Sec. III. However, the Bayesian estimator given a set of samples converges in only a couple of iterations to a result with an accuracy the pure probing will first

reach after investing a factor of a few more CPU time. For a fixed amount of computation time, the Bayesian estimator excels the probing estimator, as can be seen in Fig. 1. It is also evident that the Bayesian estimator must reach a lower limit in its progress because only a limited data set is provided containing finitely accurate information.

## 2. Realistic case

For a more realistic mock example, we consider a covariance matrix  $D = (S^{-1} + N^{-1})^{-1}$  similar to the one described by Eq. (9) where the signal covariance  $S$  is completely defined by a power spectrum,

$$C_l \propto (\max\{1, l\})^{-2}. \quad (26)$$

The noise covariance  $N$  is characterized by two effects, first, high noise in one of the twelve HEALPIX basis pixels representing a defect in the detector and, second, smoothly increased noise toward the poles imitating an observational effect.<sup>5</sup> The described noise covariance and the resulting propagator  $D$  are illustrated in Fig. 3, where one can see the conservation of the noise structure and the smoothing effect of the power spectrum.

The performance of both algorithms is shown in Fig. 4. Our algorithm performs qualitatively in the same way as in the trivial case, but the overall gain in accuracy or time is quantitatively lower. It is also noticeable that the relative advantage of the proposed method decreases with the number of used random vectors. Consequently, the matrix diagonal inference method pays off best in cases where a rough estimate using only a few probes is sufficient.

<sup>4</sup> Meaning mathematically  $\|\text{diag}[X] - f\|_2 / r$  or  $\|\text{diag}[X] - \tilde{m}\|_2 / r$ , respectively, to be exact.

<sup>5</sup> The noise variance is assigned to each pixel  $i$  according to  $N_{ii}^{-1} = (0.005 + 8(h_i(h_i - h_{\max}))^2 / h_{\max}^4)$ , where  $h_i$  is the HEALPIX ring number associated to the pixel  $i$ .

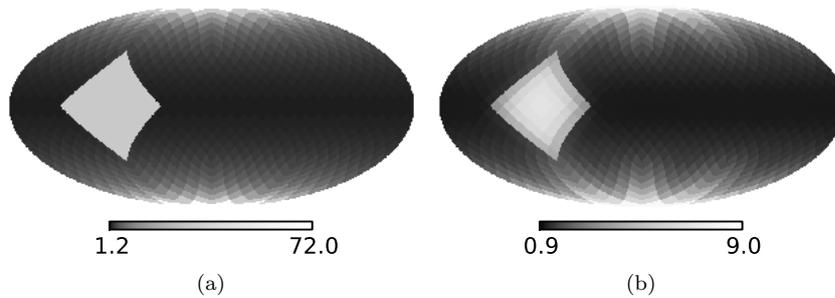


Figure 3: Realistic case: (a) the matrix diagonal of the mock noise covariance  $N$  and (b) the “true” diagonal of the propagator  $D = (S^{-1} + N^{-1})^{-1}$ .

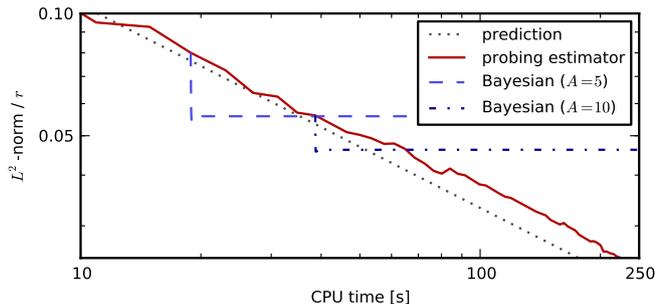


Figure 4: (Color online) Same as Fig. 2, only for the propagator for the realistic case.

## B. Faraday sky uncertainty

Next, we attempt to use our algorithm in a real physical application. We consider the inference problem discussed in Ref. [10]. In that work, an all-sky signal – the galactic Faraday depth – was reconstructed from measurements in 37 543 individual directions [16]. The data were modeled according to a linear measurement procedure, Eq. (4), with a response  $R$  that is nonzero only for directions in which measurements had been made and is larger within the galactic plane.

To reconstruct the field, the critical filter algorithm that was discussed in Sec. II A was used, yielding an estimate for the posterior mean  $m$  of the signal field  $s$ , as well as an estimate for the components of the angular power spectrum of this field,  $C_l$ . In addition, a map showing the uncertainty of the signal estimate  $m$ , given by  $\text{diag}[D]$ , is provided in Ref. [10]. This was calculated from the information propagator  $D$ , which takes on the form (21), by applying the probing estimator discussed in Sec. III.

Here, we show how the application of our Bayesian algorithm to this problem can improve the accuracy and speed up the calculation of this matrix diagonal. In order to be able to compare the results of the probing and Bayesian estimators to the correct matrix diagonal, we reduce the dimensionality of the problem to facilitate the

exact calculation of the diagonal via Eq. (11). We do this by reducing the resolution of the all-sky map with respect to the one presented in Ref. [10] to HEALPIX parameter  $N_{\text{side}} = 16$ , leading to 3 072 pixels, and truncating the reconstructed power spectrum at  $l_{\text{max}} = 47$ . Furthermore, we use a coarser version of the response matrix. In this way, a coarse-grained version of the propagator  $D$  is defined and we can calculate its diagonal exactly, as well as by using the probing estimate and our Bayesian extension.

Figure 5 shows the results of these calculations, where the matrix-vector multiplication was conducted for ten different random vectors in the case of the probing and Bayesian estimators. The sphere in this case corresponds to a map of the whole sky, where the disk of the Milky Way extends horizontally in the middle of the image. While still exhibiting a large amount of noise, both the probing and Bayesian results show roughly the right structure after only a few iterations. This structure is determined mainly by an oval region of high uncertainty, i.e. large diagonal entries, where no observations had been made and a dependence on galactic latitude due to the larger signal response within the galactic plane.

From Fig. 5 alone, it is hard to judge whether the Bayesian estimator leads to an improvement over the probing one. We therefore plot again the  $L^2$ -norm of the difference between the estimated matrix diagonal and the “true” one as a function of CPU time in Fig. 6. Shown is the curve for the pure probing estimator as well as two examples for Bayesian improvements, using two and ten random vectors, respectively. It is evident that for both cases the Bayesian method gives a boost in accuracy with only marginal time consumption. The absolute and relative improvement is larger if one uses fewer random vectors. This shows again that the main strength of the Bayesian method does not lie in the absolute accuracy that can be reached, but rather in the speedup it provides for obtaining an estimate for the matrix diagonal with intermediate accuracy.

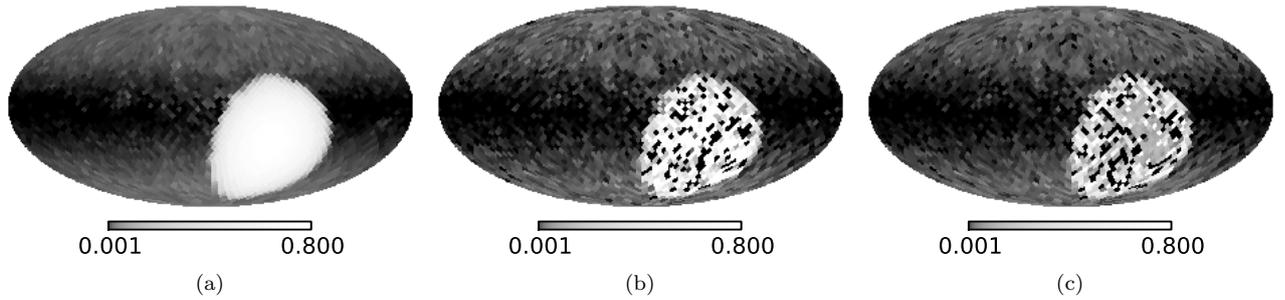


Figure 5: Diagonal of the propagator for the reconstruction of the galactic Faraday depth. Panel (a) shows the result of the exact calculation according to Eq. (11), panel (b) the probing result after ten iterations, and panel (c) the result of the Bayesian estimator, using ten random vectors as well.

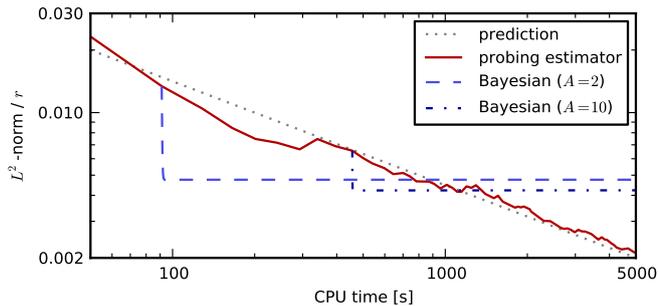


Figure 6: (Color online) Same as Fig. 2, only for the propagator for the reconstruction of the galactic Faraday depth.

## VI. CONCLUSIONS

This work aims at transferring successful signal inference methods to the realm of numerical computation. In particular, the problem of inferring internal properties of a computational black box is analyzed. To this end, the well known method of stochastic probing of matrix diagonals is augmented by a sophisticated signal reconstruction scheme. This inference algorithm interprets the probing of the matrix diagonal as a numerical measurement including unwanted uncertainties, or plainly noise.

As in signal reconstruction, additional knowledge on the matrix diagonal, i.e., our signal, can be exploited to improve the accuracy of the obtained result. We use known symmetries of the underlying continuous structure of the matrix diagonal, namely approximate statistical isotropy.

This improves the estimates acquired from probing, without increasing computational costs significantly. In order to achieve the same level of accuracy, the traditional probing method requires computational costs, which are a factor of 2 to 10 times larger than the proposed scheme, in the cases investigated here. This algo-

rithm is especially effective when matrix diagonals need to be calculated only roughly, since the relative gain in accuracy is larger if only a few probes are available.

This has been shown in numerical examples as well as for the uncertainty map appearing in the reconstruction of the galactic Faraday depth.

## VII. ACKNOWLEDGMENTS

We thank Mike Bell, Henrik Junklewitz, Georg Robbers and three anonymous referees for the insightful discussions and productive comments on the manuscript.

### Appendix A: Proof of the probing estimator

Here we prove that Eq. (13) is indeed implied by Eq. (12). Given a sufficiently large but not necessarily finite set  $\{\xi\}$  (with  $|\{\xi\}| = A$ ), the condition given by Eq. (12) becomes

$$\lim_{A \rightarrow \infty} \langle \xi_i \xi_j \rangle_{\{\xi\}} = \lim_{A \rightarrow \infty} \frac{1}{A} \sum_{a=1}^A \xi_i^{(a)} \xi_j^{(a)} = \delta_{ij}. \quad (\text{A1})$$

Inserting this equality in Eq. (13), without loss of generality restricted to the average's component  $i \in \{1, \dots, r\}$ ,

$$\begin{aligned} \left( \langle \xi * X \xi \rangle_{\{\xi\}} \right)_i &= \frac{1}{A} \sum_{a=1}^A \sum_{j=1}^r \xi_i^{(a)} X_{ij} \xi_j^{(a)} \\ &= \sum_{j=1}^r X_{ij} \underbrace{\frac{1}{A} \sum_{a=1}^A \xi_i^{(a)} \xi_j^{(a)}}_{\rightarrow \delta_{ij}} \rightarrow X_{ii}, \quad (\text{A2}) \end{aligned}$$

proves the statement.

[1] M. F. Hutchinson, Communications in Statistics - Simulation and Computation **18**, 1059 (1989).

[2] C. Bekas, E. Kokiopoulou, and Y. Saad, Applied Numer-

- ical Mathematics archive **57**, 1214 (2007).
- [3] J. M. Tang and Y. Saad, *Numerical Linear Algebra with Applications* (2011).
  - [4] E. Aune and D. P. Simpson, *ArXiv e-prints* (2011), 1105.5256.
  - [5] A. Rohde and A. B. Tsybakov, *ArXiv e-prints* (2009), 0912.5338.
  - [6] T. A. Enßlin and M. Frommert, *Phys. Rev. D* **83**, 105014 (2011), 1002.2928.
  - [7] T. A. Enßlin, M. Frommert, and F. S. Kitaura, *Phys. Rev. D* **80**, 105005 (2009), 0806.3474.
  - [8] T. A. Enßlin and C. Weig, *Phys. Rev. E* **82**, 051112 (2010), 1004.2868.
  - [9] N. Oppermann, G. Robbers, and T. A. Enßlin, *Phys. Rev. E* **84**, 041118 (2011), 1107.2384.
  - [10] N. Oppermann, H. Junklewitz, G. Robbers, and T. A. Enßlin, *A&A* **530**, A89+ (2011), 1008.1246.
  - [11] 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, et al., *ArXiv e-prints* (2011), 1111.6186.
  - [12] J. R. Shewchuk, Technical report, Carnegie Mellon University, Pittsburgh, PA (1994).
  - [13] K. M. Górski, E. Hivon, A. J. Banday, B. D. Wandelt, F. K. Hansen, M. Reinecke, and M. Bartelmann, *Astrophys. J.* **622**, 759 (2005), arXiv:astro-ph/0409513.
  - [14] HEALPIX homepage <http://healpix.jpl.nasa.gov/>.
  - [15] SAGE homepage <http://www.sagemath.org/>.
  - [16] A. R. Taylor, J. M. Stil, and C. Sunstrum, *Astrophys. J.* **702**, 1230 (2009).