Operator calculus for information field theory

Reimar H. Leike and Torsten A. Enßlin

Max-Planck-Institut für Astrophysik, Karl-Schwarzschildstrasse 1, 85748 Garching, Germany
and Ludwig-Maximilians-Universität München, Geschwister-Scholl-Platz 1, 80539 Munich, Germany

Signal inference problems with non-Gaussian posteriors can be hard to tackle. Through using the concept of Gibbs free energy these posteriors are rephrased as Gaussian posteriors for the price of computing various expectation values with respect to a Gaussian distribution. We present a way of translating these expectation values to a language of operators which is similar to that in quantum mechanics. This simplifies many calculations, for instance such as those involving log-normal priors. The operator calculus is illustrated by deriving a self-calibrating algorithm which is tested with mock data.

I. INTRODUCTION

Information field theory (IFT) [1] is a Bayesian formalism for solving field inference problems. Given a prior probability density and a data model IFT enables us to calculate posterior field expectation values when data have been measured. A common way to summarize the posterior is in terms of an estimate of the signal posterior mean and its variance. One way to obtain an approximation to those is provided by the Gibbs free energy method [2,3], also called "variational Bayes" or "mean field approximation." The minimum of the Gibbs free energy is an estimate of the posterior mean. The curvature of this minimum encodes the posterior covariance. The involved mathematical expressions contain Gaussian field averages over many functions which can become very difficult to evaluate, especially in the case of log-normal signal distributions, interactive Hamiltonians, or nonlinear responses.

We present an alternative way to calculate such Gaussian integrals; by translating them to operator calculations and using well-known formulas from differential geometry we are able to handle them efficiently. This is a general technique for calculating expectation values over Gaussian distributions which could have implications to other contexts as well. We chose to apply it to the Gibbs formalism because on the one hand Gibbs free energy inference is a very general tool for tackling inference problems and on the other hand because in this context the introduced operator formalism proves to be exceedingly useful.

In Sec. II we give a short review of Ref. [3], introducing the reader to the concept of Gibbs free energy inference and explaining its advantages and challenges. In Sec. III we introduce a typical problem set of image reconstruction as an example. In Sec. IV we translate expectation values over a Gaussian distribution into the language of operators. We then show how to leverage the power of our operator calculus with a certain set of algebraic tools in Sec. V. The algorithm that is derived using these algebraic tools is then implemented and tested for mock data. The results are discussed in Sec. VI. We conclude in Sec. VII. The derivation of the algorithm for the image reconstruction problem introduced earlier is in the Appendix.

II. GIBBS FREE ENERGY INFERENCE

To give a better understanding of the benefit we get from the operator calculus to be introduced in Sec. IV, we first give a brief introduction to the Gibbs formalism [3].

In signal reconstruction we try to infer a signal s when given the data

$$d = r(s,n) \tag{1}$$

for a given response operator r and measurement noise n. Here

$$s : X \to \mathbb{C},$$
$$x \mapsto s(x) = s_x \tag{2}$$

is a field over some measure space X. In order to infer the signal, we use the posterior probability density

$$P(s|d) = \frac{P(d|s)P(s)}{P(d)} = \frac{P(d|s)P(s)}{\int ds\, P(d|s)P(s)}. \tag{3}$$

For more complicated problems, this posterior probability density is often not accessible because integrations like that in the denominator might not be analytically solvable. In most situations we do however have access to the so called information Hamiltonian $H(s,d) = -\ln[P(s)] - \ln[P(d|s)]$ which contains all information available on the signal s. We can ignore additive constants in the Hamiltonian that depend only on d since they cancel when we reconstruct $P(s|d)$ from the Hamiltonian,

$$P(s|d) = \frac{e^{-H(s,d)}}{Z(d)}, \tag{4}$$

where

$$Z(d) = \int ds\, e^{-H(s,d)} \tag{5}$$

is the partition function. Thus we try to infer s using an estimator that only uses the Hamiltonian. A simple way is by minimizing $H(s,d)$, but this yields suboptimal results for asymmetric posterior distributions as well as unsatisfying error estimates for posterior distributions that deviate strongly from Gaussianity; see, e.g., [4]. There exist other estimators, but our formalism of operator calculus is most suited for the Gibbs free energy method and thus we will concentrate on this approach. In the Gibbs formalism we approximate the posterior by a Gaussian distribution

$$\tilde{P}(s|d) = \mathscr{G}(s-m, D) = \frac{e^{(1/2)(s-m)^\dagger D^{-1}(s-m)}}{|2\pi D|^{1/2}} \tag{6}$$

with mean m and covariance D that depend on the data in a way still to be found. Here we use the scalar product of fields

$$a^\dagger b = \int dx \, a_x^* b_x \qquad (7)$$

with $*$ denoting complex conjugation. For a Gaussian posterior, calculating the partition function is possible (see Sec. II E of Ref. [3]). This approach is also known in the literature under the names of "variational Bayes" and "mean field approximation" but we will call it Gibbs free energy because it was historically the earliest of the three concepts.

Almost every scientific result is given as a pair of estimate and standard deviation. Most people assume Gaussian statistics when seeing a result like that, consistent with its maximum entropy translation into a probability density function. Thus approximating the posterior by a Gaussian is basically common practice and information theoretically supported [5]. While doing so, we might as well try to invent the least amount of information through the approximation. In order to obtain m and D we therefore minimize the Kullback-Leibler divergence [6]:

$$\begin{aligned} d_{KL}(\tilde{P}, P) &= \int ds \, \tilde{P}(s|d) \ln\left(\frac{\tilde{P}(s|d)}{P(s|d)} \right) \\ &= \left\langle \ln\left(\frac{\tilde{P}(s|d)}{P(s|d)} \right) \right\rangle_{\mathscr{G}(s-m,D)} \\ &= \langle \ln(\tilde{P}(s|d)) \rangle_{\mathscr{G}(s-m,D)} \\ &\quad + \langle H(s|d) \rangle_{\mathscr{G}(s-m,D)}. \end{aligned} \qquad (8)$$

The Kullback-Leibler divergence is therefore up to the irrelevant constant $\ln[Z(d)]$ equal to the Gibbs free energy $G(m,D) = U - TS$ at temperature $T = 1$ with the Shannon entropy [7]

$$S = -\langle \ln(\tilde{P}(s|d)) \rangle_{\mathscr{G}(s-m,D)} \qquad (9)$$

and the internal energy

$$U = \langle H(s,d) \rangle_{\mathscr{G}(s-m,D)}. \qquad (10)$$

The posterior mean m is now given within our approximation by the minimum of the Gibbs free energy

$$m : \frac{\delta G}{\delta m} = 0. \qquad (11)$$

From a Kullback-Leibler divergence perspective, the posterior uncertainty dispersion is given by

$$D : \frac{\delta G}{\delta D} = 0. \qquad (12)$$

Derivatives with respect to operators can be difficult to handle. Fortunately the thermodynamical relation

$$D = \left(\frac{\delta^2 G}{\delta m \delta m^\dagger} \right)^{-1}, \qquad (13)$$

which holds at the minimum of the Gibbs free energy, requires only derivatives with respect to the mean field m.

Note that following the Gibbs free energy approach we only need to calculate expectation values over a Gaussian distribution and instead of the full posterior $P(s|d)$ it suffices

to know the Hamiltonian $H(s,d)$ of the joint probability of data and signal.

III. A SELF-CALIBRATING SYSTEM AS EXAMPLE

Suppose now we have a measurement scenario where a real field a contributes to the data via being exponentiated,

$$d = re^a + n, \qquad (14)$$

$$P(a) = \mathscr{G}(a,A). \qquad (15)$$

This corresponds to a linear data model with response operator r and a log-normal prior which is the natural prior for strictly positive signals that vary over orders of magnitude. For example the galaxy densities in the cosmos show roughly log-normal distributions as was supported empirically [8,9] and theoretically [10–15]. Suppose additionally that the response operator is unknown and we have an independent Gaussian prior for it,

$$P(r) = \mathscr{G}(r,R). \qquad (16)$$

Now we are actually dealing with a self-calibration problem, as the unknown instrument response r has to be inferred from the unknown signal observation. These are notoriously hard; see [16] or [17].

We define a joint signal vector

$$s = \begin{pmatrix} r \\ a \end{pmatrix}, \qquad (17)$$

$$P(s) = \mathscr{G}(s,S) = \mathscr{G}\left[s, \begin{pmatrix} R & 0 \\ 0 & A \end{pmatrix} \right] \qquad (18)$$

for all quantities we would like to infer.

In the simple case of Gaussian additive noise $P(n) = \mathscr{G}(n,N)$ we get as the Hamiltonian

$$\begin{aligned} H(d,s) &= H(s) + H(d|s) \\ &= \tfrac{1}{2} s^\dagger S^{-1} s \\ &\quad + \tfrac{1}{2} (d - re^a)^\dagger N^{-1} (d - re^a). \end{aligned} \qquad (19)$$

The Hamiltonian thus contains the interacting signal terms $d^\dagger N^{-1} re^a$ and $(re^a)^\dagger N^{-1} re^a$ for which the expectation value over the generic Gaussian distribution $\mathscr{G}(s-m,D)$ has to be taken to calculate the Gibbs free energy. Although this can be done by hand, calculations can get very tedious and require a lot of time. We will be able to handle them quite nicely with our operator formalism in the Appendix.

In the case of a nonlinear response or a signal depended noise model we get even more exponentials and potentially additional factors of polynomials in s.

Now that we have seen a typical problem set, let us proceed by introducing the tools to translate expectation values over Gaussian distributions to operator action.

IV. FORMULATING GAUSSIAN AVERAGES IN OPERATOR CALCULUS

In this section we are concerned with the task of calculating the expectation value

$$\langle f(s) \rangle_{\mathscr{G}(s-m,D)} \qquad (20)$$

for a Gaussian distribution in s with mean m and covariance D.

Let us start with the much more simple task of calculating $\langle s \rangle_{\mathscr{G}(s-m,D)}$. We let us guide by a calculation trick from statistical physics where a lot of expectation values are calculated by taking different derivatives of the partition sum and thus try to obtain s by taking the derivative of $\mathscr{G}(s-m,D)$:

$$\frac{\delta}{\delta m}\mathscr{G}(s-m,D) = D^{-1}(s-m)\mathscr{G}(s-m,D)$$

$$\Rightarrow \left(D\frac{\delta}{\delta m} + m\right)\mathscr{G}(s-m,D)$$

$$= s\,\mathscr{G}(s-m,D). \qquad (21)$$

Thus we have

$$\langle s \rangle_{\mathscr{G}(s-m,D)} = \left\langle D\frac{\delta}{\delta m} + m \right\rangle_{\mathscr{G}(s-m,D)}. \qquad (22)$$

Here the linear operator $D\frac{\delta}{\delta m} + m$ does not depend on s, so one may pull it out of the expectation value:

$$\langle s \rangle_{\mathscr{G}(s-m,D)} = \left(D\frac{\delta}{\delta m} + m\right)\langle 1 \rangle_{\mathscr{G}(s-m,D)} = m. \qquad (23)$$

This is not a surprising result. However, it is remarkable that this works for any moment of the Gaussian

$$\langle s^n \rangle_{\mathscr{G}(s-m,D)} = \left(D\frac{\delta}{\delta m} + m\right)^n 1. \qquad (24)$$

We call $\Phi := D\frac{\delta}{\delta m} + m$ the s operator. Let us look at the expectation value of an arbitrary analytical function f. By definition, an analytical function can be expanded locally in a point s_0 in a series $f(s) = \sum_{n=0}^{\infty} \Lambda_n (s-s_0)^n$ that has a positive convergence radius. We use a short notation for the Taylor-Fréchet expansion of the function f,

$$f(s) = \sum_{n=0}^{\infty} \Lambda_n (s-s_0)^n$$

$$= \sum_{n=0}^{\infty} \int dx_1 \ldots \int dx_n$$

$$\times \Lambda_n(x_1,\ldots,x_n)(s-s_0)_{x_1}\ldots(s-s_0)_{x_n} \qquad (25)$$

and calculate

$$\langle f(s) \rangle_{\mathscr{G}(s-m,D)} = \sum_{n=0}^{\infty} \Lambda_n \langle (s-s_0)^n \rangle_{\mathscr{G}(s-m,D)}$$

$$= \sum_{n=0}^{\infty} \Lambda_n \sum_{i=0}^{n} \binom{n}{i} \langle s^i(-s_0)^{n-i} \rangle_{\mathscr{G}(s-m,D)}$$

$$= \sum_{n=0}^{\infty} \Lambda_n \sum_{i=0}^{n} \binom{n}{i} \langle \Phi^i(-s_0)^{n-i} \rangle_{\mathscr{G}(s-m,D)}$$

$$= \sum_{n=0}^{\infty} \Lambda_n \langle (\Phi-s_0)^n \rangle_{\mathscr{G}(s-m,D)}$$

$$= \sum_{n=0}^{\infty} \Lambda_n (\Phi-s_0)^n 1 = f(\Phi)1. \qquad (26)$$

Thus instead of calculating the expectation value of $f(s)$ with respect to a Gaussian distribution we can let the operator $f(\Phi)$ act on 1.

When dealing with complex numbers we have to treat s and s^* separately and replace them with $\Phi' := 2D\frac{\delta}{\delta m^*} + m$ and $\Phi'^* := 2D\frac{\delta}{\delta m} + m^*$ respectively. These two operators commute $[\Phi',\Phi'^*] = 0$ and calculations thus follow a similar line for complex fields.

V. CALCULATING GAUSSIAN EXPECTATION VALUES ALGEBRAICALLY

In order to highlight the benefit of this reformulation of integrations to operator actions, we introduce the reader to certain useful algebraic tools and show how to apply them. The first step to all calculations is to separate

$$\Phi_x = \int dy\, D_{xy}\frac{\delta}{\delta m_y} + m_x = c^x + b^x. \qquad (27)$$

We call $b^x = m_x$ the creation operator and $c^x = \int dy\, D_{xy}\frac{\delta}{\delta m_y}$ the annihilation operator. Our goal is to get the annihilation operators to the right hand side because they cancel,

$$c^x 1 = \int dt\, D_{xt}\frac{\delta}{\delta m_t} 1 = 0. \qquad (28)$$

To achieve this we use the commutation relations of the creation and annihilation operators,

$$[b^x, b^y] = [c^x, c^y] = 0, \qquad (29)$$

$$[c^x, b^y] = D_{xy}. \qquad (30)$$

How exactly we bring the annihilation part to the right side differs for different classes of functions. For polynomials we can simply use distributivity of multiplication,

$$\Phi_x\Phi_y = (b^x + c^x)(b^y + c^y)$$

$$= b^x b^y + b^x c^y + c^x b^y + c^x c^y, \qquad (31)$$

and then apply the commutation relations to obtain

$$\Phi_x\Phi_y = b^x b^y + 2b^x c^y + [c^x, b^y] + c^x c^y$$

$$= b^x b^y + 2b^x c^y + D_{xy} + c^x c^y$$

$$\Rightarrow \Phi_x\Phi_y 1 = m_x m_y + D_{xy}. \qquad (32)$$

We can separate creation and annihilation parts for exponential functions by making use of the Baker-Campbell-Hausdorff (BCH) formula [18]

$$e^{b^x + c^y + (1/2)[b^x, c^y]} = e^{b^x} e^{c^y}. \qquad (33)$$

Thereby, we can omit further iterations of the commutator that appear in the full BCH formula because $[b^x, c^y] = -D_{xy}$ is central in the algebra of linear operators on functions of m, i.e., it commutes with c^y and b^x. Applying this yields

$$e^{\Phi_x} = e^{b^x + c^x}$$

$$= e^{-(1/2)[b^x, c^x]} e^{b^x} e^{c^x}$$

$$= e^{(1/2)D_{xx}} e^{b^x} e^{c^x}. \qquad (34)$$

Thus for certain functions $f(\Phi)$ we are able to separate the annihilation part and the creation part of Φ,

$$f(\Phi) = \sum_i f_i^b(b) f_i^c(c), \tag{35}$$

using algebraic tools. One major advantage of using that approach instead of calculating the expectation value directly is that now calculating the expectation value of the product of two functions $\langle f(s)g(s)\rangle_{\mathscr{G}(s-m,D)}$ simply amounts to calculating the commutator,

$$
\begin{aligned}
\langle f(s)g(s)\rangle_{\mathscr{G}(s-m,D)} &= f(\Phi)g(\Phi)1 \\
&= \sum_i f_i^b(b) f_i^c(c) \sum_j g_i^b(b) g_i^c(c)1 \\
&= \sum_{i,j} f_i^b(m) g_i^b(m) f_i^c(0) g_i^c(0)1 \\
&\quad + \sum_{i,j} f_i^b(b)\big[f_i^c(c), g_i^b(b)\big] g_i^b(0)1 \\
&= \langle f(s)\rangle_{\mathscr{G}(s-m,D)} \langle g(s)\rangle_{\mathscr{G}(s-m,D)} \\
&\quad + \sum_{i,j} f_i^b(b)\big[f_i^c(c), g_i^b(b)\big] g_i^c(0)1,
\end{aligned}
\tag{36}
$$

of the two involved functions.

We can calculate those commutators using algebraic tools. For example to exchange c^x and e^{b^y} we use the fact that $[c^x, \_]$ has the algebraic properties of a derivation, meaning it is linear and obeys the product rule

$$[c^x, ab] = [c^x, a]b + a[c^x, b]. \tag{37}$$

Thus

$$
\begin{aligned}
[c^x, e^{b^y}] &= \sum_{n=0}^{\infty} \frac{[c^x, (b^y)^n]}{n!} \\
&= \sum_{n=0}^{\infty} \frac{n(b^y)^{n-1}[c^x, b^y]}{n!} \\
&= e^{b^y}[c^x, b^y] \\
&= D_{xy} e^{b_y}.
\end{aligned}
\tag{38}
$$

We can calculate the commutator of two exponential functions using the BCH formula twice:

$$
\begin{aligned}
[e^{c^x}, e^{b^y}] &= e^{c^x} e^{b^y} - e^{b^y} e^{c^x} \\
&= e^{b^y + c^x - (1/2)[b^y, c^x]} - e^{b^y} e^{c^x} \\
&= e^{b^y} e^{c^x} e^{-[b^y, c^x]} - e^{b^y} e^{c^x} \\
&= e^{b^y} e^{c^x} (e^{D_{xy}} - 1).
\end{aligned}
\tag{39}
$$

If we just want to exchange the position of these exponentials the formula (39) simplifies to

$$e^{c^x} e^{b^y} = e^{b^y} e^{c^x} e^{D_{xy}}. \tag{40}$$

Having aggregated these tools, calculating the Gibbs free energy of the self-calibration problem introduced in Sec. III is straightforward. This calculation is done in the Appendix.

VI. A NUMERICAL EXAMPLE

To conclude this paper we implemented and verified the derived algorithm. For our implementation we use the slightly altered data model

$$d = (r + r_0)\mathbb{1}e^a + n, \tag{41}$$

where $\mathbb{1}$ denotes the identity matrix and r is just a scalar. This simplified model has the advantage of being easier to implement because it is less degenerate and its results are easier to visualize. The constant r_0 encodes that we usually have some rough idea about the typical response of our instrument. We assume a Gaussian noise distribution

$$P(n) = \mathscr{G}\big(n, \sigma_N^2 \delta_{ij}\big)$$

that has a scalar covariance σ_N^2 and also a Gaussian prior distribution for a and r as in Eqs. (15) and (16). The only difference is that the covariance matrix for r is now just a constant. We take the signal prior covariance A to be diagonal in Fourier space with known power spectrum. Analogously to the derivation in the Appendix we arrive at

$$
\begin{aligned}
G(m, D) =& -\frac{1}{2}\mathrm{tr}[1 + \ln(2\pi D)] \\
&+ \frac{1}{2}m^\dagger S^{-1} m + \frac{1}{2}\mathrm{tr}(S^{-1} D) \\
&- \int di \left(\frac{d^\dagger}{\sigma_N^2}\right)_i [m_r + r_0 + (D_{ra})_i](e^{m_a + (1/2)\widehat{D_{aa}}})_i \\
&+ \frac{1}{2}\int di\, (e^{2m_a + 2\widehat{D_{aa}}})_i \\
&\times \{D_{rr} + [m_r + r_0 + 2(D_{ra})_i]^2\}.
\end{aligned}
\tag{42}
$$

Taking the derivative by m we arrive at the gradient in signal direction,

$$
\begin{aligned}
\frac{\delta G(m, D)}{\delta(m_a)_i} =& A^{-1} m_a - \left(\frac{d^\dagger}{\sigma_N^2}\right)_i \\
&\times [m_r + r_0 + (D_{ra})_i](e^{m_a + (1/2)\widehat{D_{aa}}})_i \\
&+ (e^{2m_a + 2\widehat{D_{aa}}})_i \{D_{rr} + [m_r + r_0 + 2(D_{ra})_i]^2\}
\end{aligned}
\tag{43}
$$

and in the response factor direction

$$
\begin{aligned}
\frac{\delta G(m, D)}{\delta m_r} =& \frac{m_r}{R} - \int di \left(\frac{d^\dagger}{\sigma_N^2}\right)_i (e^{m_a + (1/2)\widehat{D_{aa}}})_i \\
&+ \int di (e^{2m_a + 2\widehat{D_{aa}}})_i [m_r + r_0 + 2(D_{ra})_i],
\end{aligned}
\tag{44}
$$

respectively. Taking the derivative again we arrive at the Hessian matrix

$$
\begin{aligned}
\frac{\delta^2 G(m, D)}{\delta(m_a)_i \delta(m_a^\dagger)_j} =& A_{ij}^{-1} - \delta_{ij}\left(\frac{d^\dagger}{\sigma_N^2}\right)_i [m_r + r_0 + (D_{ra})_i] \\
&\times (e^{m_a + (1/2)\widehat{D_{aa}}})_i + 2\delta_{ij}(e^{2m_a + 2\widehat{D_{aa}}})_i \\
&\times \{D_{rr} + [m_r + r_0 + 2(D_{ra})_i]^2\},
\end{aligned}
\tag{45}
$$

$$\frac{\delta^2 G(m,D)}{\delta m_r \delta (m_a)_i} = -\left(\frac{d^\dagger}{\sigma_N^2}\right)_i (e^{m_a+(1/2)\widehat{D_{aa}}})_i + 2(e^{2m_a+2\widehat{D_{aa}}})_i$$

$$\times [m_r + r_0 + 2(D_{ra})_i], \tag{46}$$

$$\frac{\delta^2 G(m,D)}{\delta^2 m_r} = R + \int di (e^{2m_a+2\widehat{D_{aa}}})_i. \tag{47}$$

Using these in a Newton scheme to find the minimum, we arrive at an algorithm that we implemented. In Fig. 1 we show mock data that were generated by sampling from the prior in comparison to the projected signal response $(r + r_0)e^a$ and the expected value of the signal response as was computed by the Gibbs reconstruction algorithm. Notice that the signal response varies over a few orders of magnitude due to its log-normal nature. One can also see that the signal has strong spatial correlations that were encoded into the prior. Figure 2 shows the reconstruction of the signal we get from our algorithm. These results only deviate slightly from the results we get when using a maximum *a posteriori* (MAP) estimator because the posterior is still relatively near to a Gaussian. The response factor $r + r_0$ was sampled to be 3.62, the corresponding Gibbs estimate is $m_r + r_0 = 3.11 \pm 0.41$, and the MAP estimate of it is $(m_{\text{MAP}})_r + r_0 = 3.15 \pm 0.46$. Thus both deviate about one σ from the actual value. In this case, the Gibbs result is comparable to the MAP estimator. However, this inference problem was chosen to demonstrate how the operator formalism works and not to highlight the differences in the performance of the Gibbs estimator with respect to that of the MAP estimator.

VII. CONCLUSION

With the help of the Gibbs free energy one can easily write down expressions for the posterior mean and covariance. Using the operator formalism introduced in this paper we formulated expectation values as operators acting on 1 which eliminates

FIG. 2. The signal reconstruction m_a (dashed line) with position dependent uncertainty range (shaded area). This uncertainty is given by the square root of the diagonal of the covariance matrix $\widehat{D_{aa}}$. Notice that due to the log-normal nature of the problem, higher signal values are reconstructed more accurately.

the need to calculate an integral over the Hamiltonian and the Gaussian distribution. This process of translating the expectation value to operators works generally, albeit possibly entailing algebraic complexity. For expectation values over products of exponential functions and polynomials which are typical for problems with log-normal statistics, we aggregated a collection of algebraic tools that enable us to nevertheless calculate them in a few lines of straightforward calculation. We demonstrated their usage by applying our operator calculus to a signal inference problem with log-normal prior and unknown but linear response operator for which we worked out all the occurring terms with regard to the posterior mean. The resulting algorithm was implemented and found to be working for mock data.

Future research might be directed towards finding analogies to the BCH formula for function classes other than the exponential function which will allow us to apply the operator formalism to an even broader range of problems.

APPENDIX: FACILITATING CALCULATIONS WITH OPERATORS

By making use of our operator formalism we are able to quickly calculate expectation values of products of exponentials and polynomials like those we encountered in Sec. III. We calculate the Gibbs free energy from the Hamiltonian we got at the end of Sec. III. Combining Eqs. (8) and (19)
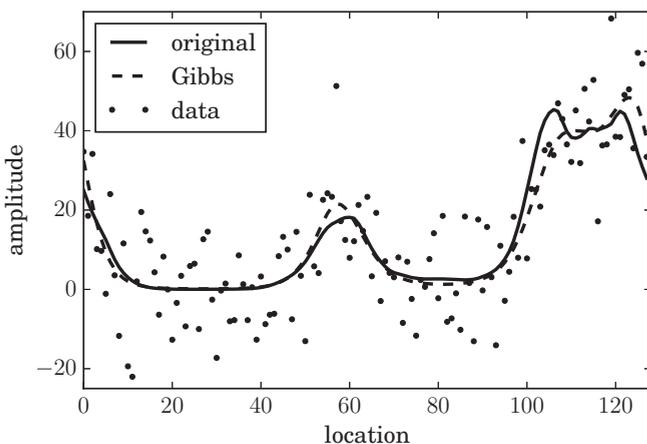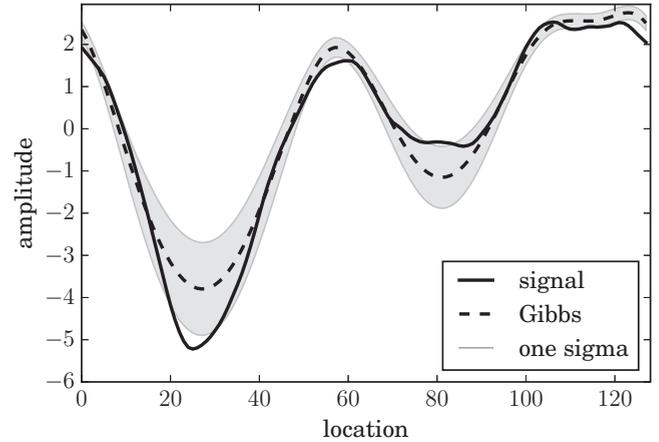

FIG. 1. Original signal response (solid line), expected value of the signal response given via the Gibbs estimate (dashed line) and data that were sampled from the prior (points). The expected value of the signal response was computed with the formula $(r_0 + m_r + D_{ra})e^{m_a+(1/2)\widehat{D_{aa}}}$.

yields

$$
\begin{aligned}
G(m,D) = & \, G\left[\begin{pmatrix} m_r \\ m_a \end{pmatrix}, \begin{pmatrix} D_{rr} & D_{ra} \\ D_{ar} & D_{aa} \end{pmatrix}\right] \\
\widehat{=} & \, \langle \ln(\tilde{P}(s|d)) \rangle_{\mathscr{G}(s-m,D)} + \left\langle \frac{1}{2}s^\dagger S^{-1}s + \frac{1}{2}(d-re^a)^\dagger N^{-1}(d-re^a) \right\rangle_{\mathscr{G}(s-m,D)} \\
\widehat{=} & \, \left\langle -\frac{1}{2}(s-m)^\dagger D^{-1}(s-m) + \ln(|2\pi D|^{-1/2}) \right\rangle_{\mathscr{G}(s-m,D)} + \left\langle \frac{1}{2}s^\dagger S^{-1}s \right\rangle_{\mathscr{G}(s-m,D)} - \langle d^\dagger N^{-1}re^a \rangle_{\mathscr{G}(s-m,D)} \\
& + \left\langle \frac{1}{2}(re^a)^\dagger N^{-1}re^a \right\rangle_{\mathscr{G}(s-m,D)}.
\end{aligned}
\tag{A1}
$$

Here "$\widehat{=}$" denotes equality up to irrelevant constants, which are constants that do not depend on m or D.

Following the formalism introduced in Sec. IV we replace

$$
a \leftrightarrow \Phi^a = D_{ax}\frac{\delta}{\delta m_x} + m_a,
\tag{A2}
$$

$$
r \leftrightarrow \Phi^r = D_{rx}\frac{\delta}{\delta m_x} + m_r,
\tag{A3}
$$

$$
s \leftrightarrow \Phi = D\frac{\delta}{\delta m} + m.
\tag{A4}
$$

We now evaluate the terms of Eq. (A1) one by one. The first two terms are simply second moments of a Gaussian distribution and thus the calculation can easily be done by hand. For illustration we use our formalism anyway and focus on the second term:

$$
\left\langle \frac{1}{2}s^\dagger S^{-1}s \right\rangle_{\mathscr{G}(s-m,D)} = \frac{1}{2}\int di\,dj\,\langle s_i S_{ij}^{-1}s_j \rangle_{\mathscr{G}(s-m,D)} = \frac{1}{2}\int di\,dj\,\Phi_i S_{ij}^{-1}\Phi_j 1.
\tag{A5}
$$

We separate

$$
\Phi_t^x = \left(D\frac{\delta}{\delta m_x}\right)_t + (m_x)_t = c_t^x + b_t^x
\tag{A6}
$$

with

$$
c_t^x = \left(D\frac{\delta}{\delta m}\right)_t = \int dv\, D_{tv}\frac{\delta}{(\delta m_x)_v}, \quad b_t^x = (m_x)_t,
\tag{A7}
$$

where x labels "a," "r," or no index and arrive at the commutation relations

$$
[b_i^x, b_j^y] = [c_i^x, c_j^y] = 0,
\tag{A8}
$$

$$
[c_i^x, b_j^y] = (D_{xy})_{ij}.
\tag{A9}
$$

Our goal is to get the annihilation operators to the right hand side because they cancel. Doing so one gets

$$
\begin{aligned}
\frac{1}{2}\int di\,dj\, S_{ij}^{-1}\Phi_i\Phi_j 1 = & \, \frac{1}{2}\int di\,dj\, S_{ij}^{-1}(c_i+b_i)(c_j+b_j)1 \\
= & \, \frac{1}{2}\int di\,dj\, S_{ij}^{-1}([c_i,b_j]+b_ib_j)1 = \frac{1}{2}m^\dagger S^{-1}m + \frac{1}{2}\mathrm{tr}(S^{-1}D).
\end{aligned}
\tag{A10}
$$

We proceed with the third term,

$$
\langle d^\dagger N^{-1}re^a \rangle_{\mathscr{G}(s-m,D)} = d^\dagger N^{-1}\Phi^r e^{\Phi^a}1.
\tag{A11}
$$

To simplify $\Phi^r e^{\Phi^a}1$ we apply the BCH formula:

$$
\begin{aligned}
(\Phi^r e^{\Phi^a}1)_j = & \int di\,(b_{ji}^r + c_{ji}^r)e^{c_i^a + b_i^a}1 \\
= & \int di\,(b_{ji}^r + c_{ji}^r)e^{b_i^a + (1/2)(D_{aa})_{ii}}e^{c_i^a}1.
\end{aligned}
\tag{A12}
$$

To exchange c^r_{ji} and $e^{b^a_i}$ we use the fact that $[c^r_{ji},\_]$ has the algebraic properties of a derivation, thus

$$\Phi^r e^{\Phi^a} 1 = \int di \left(b^r_{ji} + (D_{ra})_{(ji)i} \right) e^{b^a_i + (1/2)(\widehat{D_{aa}})_i} e^{c^a_i} 1$$

$$= \int di \left((m_r)_{ji} + (D_{ra})_{(ji)i} \right) (e^{m_a + (1/2)\widehat{D_{aa}}})_i. \tag{A13}$$

With $\widehat{D_{aa}}$ we denote the diagonal of the operator D_{aa}. We used that if we Taylor expand $e^{c^a_i}$ only the first term will contribute since all terms containing c cancel with the 1.

The last term of Eq. (A1) is

$$\left\langle \frac{1}{2}(re^a)^\dagger N^{-1} re^a \right\rangle. \tag{A14}$$

Translating this into operator language we arrive at

$$\left\langle \frac{1}{2}(re^a)^\dagger N^{-1} re^a \right\rangle = e^{\Phi^\dagger_a} \Phi^\dagger_r N^{-1} \Phi_r e^{\Phi_a} 1 = \int di\,dj\,dk\,dl (e^{\Phi_a})_l (\Phi_r)_{kl} (N^{-1})_{kj} (\Phi_r)_{ji} (e^{\Phi_a})_i 1. \tag{A15}$$

First we separate the exponentials with the BCH formula as we have done with the previous term and get

$$(N^{-1})_{kj}(e^{\Phi_a})_l (e^{\Phi_a})_i (\Phi_r)_{kl}(\Phi_r)_{ji} 1 = (N^{-1})_{kj} e^{b^a_l + (1/2)(D_{aa})_{ll}} e^{c^a_l} e^{b^a_i + (1/2)(D_{aa})_{ii}} e^{c^a_i}$$

$$\left(c^r_{kl} b^r_{ji} + b^r_{kl} b^r_{ji} \right) 1 = (N^{-1})_{kj} e^{b^a_l + (1/2)(D_{aa})_{ll}} e^{c^a_l} e^{b^a_i + (1/2)(D_{aa})_{ii}} e^{c^a_i}$$

$$\left((D_{rr})_{(kl)(ji)} + b^r_{kl} b^r_{ji} \right) 1 = (N^{-1})_{kj} e^{b^a_l + (1/2)(D_{aa})_{ll}} e^{b^a_i + (1/2)(D_{aa})_{ii}} e^{(D_{aa})_{li}} e^{c^a_l} e^{c^a_i}$$

$$\left((D_{rr})_{(kl)(ji)} + b^r_{kl} b^r_{ji} \right) 1 = (N^{-1})_{kj} e^{b^a_l + (1/2)(D_{aa})_{ll}} e^{b^a_i + (1/2)(D_{aa})_{ii}} e^{(D_{aa})_{li}} e^{c^a_l}$$

$$\left[(D_{rr})_{(kl)(ji)} + \left(b^r_{kl} + (D_{ra})_{(kl)i} \right) e^{c^a_i} b^r_{ji} \right] 1 = (N^{-1})_{kj} e^{b^a_l + (1/2)(D_{aa})_{ll}} e^{b^a_i + (1/2)(D_{aa})_{ii}} e^{(D_{aa})_{li}}$$

$$\times \left[(D_{rr})_{(kl)(ji)} + \left(b^r_{kl} + (D_{ra})_{(kl)l} + (D_{ra})_{(kl)i} \right)\left(b^r_{ji} + (D_{ra})_{(ji)l} + (D_{ra})_{(ji)i} \right) \right] 1$$

$$= (N^{-1})_{kj} (e^{m_a + (1/2)\widehat{D_{aa}}})_l (e^{m_a + (1/2)\widehat{D_{aa}}})_i (e^{D_{aa}})_{li}$$

$$\times \left[(D_{rr})_{(kl)(ji)} + ((m_r)_{kl} + (D_{ra})_{(kl)l} + (D_{ra})_{(kl)i})((m_r)_{ji} + (D_{ra})_{(ji)l} + (D_{ra})_{(ji)i}) \right]. \tag{A16}$$

For the Gibbs energy we therefore arrive at

$$G(m,D) = -\frac{1}{2}\mathrm{tr}[1 + \ln(2\pi D)] + \frac{1}{2}m^\dagger S^{-1} m + \frac{1}{2}\mathrm{tr}(S^{-1}D) - \int di\,dj\,(d^\dagger N^{-1})_j ((m_r)_{ji} + (D_{ra})_{(ji)i})(e^{m_a + (1/2)\widehat{D_{aa}}})_i$$

$$+ \frac{1}{2}\int di\,dj\,dk\,dl (e^{m_a + (1/2)\widehat{D_{aa}}})_l [(D_{rr})_{(kl)(ji)} + ((m_r)_{kl} + (D_{ra})_{(kl)l} + (D_{ra})_{(kl)i})((m_r)_{ji} + (D_{ra})_{(ji)l} + (D_{ra})_{(ji)i})]$$

$$\times (N^{-1})_{kj} e^{(D_{aa})_{li}} (e^{m_a + (1/2)\widehat{D_{aa}}})_i. \tag{A17}$$

We separately compute the derivative for the reconstructed signal,

$$\frac{\delta G(m,D)}{\delta(m_a)_i} = (A^{-1}m_a)_i - \int dj (d^\dagger N^{-1})_j ((m_r)_{ji} + (D_{ra})_{(ji)i})(e^{m_a + (1/2)\widehat{D_{aa}}})_i + \int dj\,dk\,dl (e^{m_a + (1/2)\widehat{D_{aa}}})_l [(D_{rr})_{(kl)(ji)} + ((m_r)_{kl}$$

$$+ (D_{ra})_{(kl)l} + (D_{ra})_{(kl)i})((m_r)_{ji} + (D_{ra})_{(ji)l} + (D_{ra})_{(ji)i})](N^{-1})_{kj} e^{(D_{aa})_{li}} (e^{m_a + (1/2)\widehat{D_{aa}}})_i, \tag{A18}$$

and response,

$$\frac{\delta G(m,D)}{\delta(m_r)_{ij}} = (R^{-1}m_r)_{ij} - (d^\dagger N^{-1})_j (e^{m_a + (1/2)\widehat{D_{aa}}})_i + \int dk\,dl (e^{m_a + (1/2)\widehat{D_{aa}}})_l ((m_r)_{kl} + (D_{ra})_{(kl)l} + (D_{ra})_{(kl)j})(N^{-1})_{ki}$$

$$\times e^{(D_{aa})_{lj}} (e^{m_a + (1/2)\widehat{D_{aa}}})_j. \tag{A19}$$

To finalize the derivation we take the second derivative of the Gibbs free energy which will give us an estimate for D^{-1}. Via the relationship

$$D^{-1} = \frac{\delta^2 G(m,D)}{\delta m \delta m^\dagger}$$

we compute

$$\frac{\delta^2 G(m,D)}{\delta(m_a)_i \delta(m_a^\dagger)_l} = A_{il}^{-1} - \int dj\,(d^\dagger N^{-1})_j((m_r)_{ji} + (D_{ra})_{(ji)i})(e^{m_a+(1/2)\widehat{D_{aa}}})_i \delta_{il} + \int dj\,dk\,dn\,(e^{m_a+(1/2)\widehat{D_{aa}}})_n[(D_{rr})_{(kn)(ji)}$$

$$+ ((m_r)_{kn} + (D_{ra})_{(kn)n} + (D_{ra})_{(kn)i})((m_r)_{ji} + (D_{ra})_{(ji)n} + (D_{ra})_{(ji)i})](N^{-1})_{kj} e^{(D_{aa})_{ni}}(e^{m_a+(1/2)\widehat{D_{aa}}})_i(\delta_{il} + \delta_{nl}), \tag{A20}$$

$$\frac{\delta^2 G(m,D)}{\delta(m_a)_k \delta(m_r^\dagger)_{ji}} = -(d^\dagger N^{-1})_j (e^{m_a+(1/2)\widehat{D_{aa}}})_i \delta_{ik} + \int dm\,dl\,dn(\delta_{kl} + \delta_{kj})(e^{m_a+(1/2)\widehat{D_{aa}}})_l$$

$$\times ((m_r)_{nl} + (D_{ra})_{(nl)l} + (D_{ra})_{(nl)j})(N^{-1})_{ni} e^{(D_{aa})_{lj}}(e^{m_a+(1/2)\widehat{D_{aa}}})_j, \tag{A21}$$

$$\frac{\delta^2 G(m,D)}{\delta(m_r)_{kl} \delta(m_r^\dagger)_{ji}} = R^{-1} + (e^{m_a+(1/2)\widehat{D_{aa}}})_l(N^{-1})_{ki} e^{(D_{aa})_{lj}}(e^{m_a+(1/2)\widehat{D_{aa}}})_j, \tag{A22}$$

where with δ_{xy} we denote the Dirac δ function. Now we arrive at a point where we have a fully operational reconstruction algorithm. By using a minimization technique like gradient descent one can simultaneously reconstruct the signal field and response operator for given data d.

[1] T. A. Enßlin, M. Frommert, and F. S. Kitaura, Information field theory for cosmological perturbation reconstruction and nonlinear signal analysis, Phys. Rev. D **80**, 105005 (2009).

[2] M. Opper and D. Saad, *Advanced Mean Field Methods: Theory and Practice*, Neural Information Processing Series (MIT Press, Cambridge, Massachusetts, 2001).

[3] T. A. Enßlin and C. Weig, Inference with minimal Gibbs free energy in information field theory, Phys. Rev. E **82**, 051112 (2010).

[4] T. A. Enßlin and M. Frommert, Reconstruction of signals with unknown spectra in information field theory with parameter uncertainty, Phys. Rev. D **83**, 105014 (2011).

[5] E. T. Jaynes, *Probability Theory*, edited by G. L. Bretthorst (Cambridge University Press, Cambridge, UK, 2003).

[6] S. Kullback and R. Leibler, On information and sufficiency, Ann. Math. Stat. **22**, 79 (1951).

[7] C. E. Shannon, A mathematical theory of communication, Bell Syst. Tech. J. **27**, 379 (1948).

[8] E. Hubble, The distribution of extra-galactic nebulae, Astrophys. J. **79**, 8 (1934).

[9] F. S. Kitaura, J. Jasche, C. Li, T. A. Enßlin, R. B. Metcalf, B. D. Wandelt, G. Lemson, and S. D. M. White, Cosmic cartography of the large-scale structure with Sloan Digital Sky Survey data release 6, Mon. Not. R. Astron. Soc. **400**, 183 (2009).

[10] D. Layzer, A new model for the distribution of galaxies in space, Astron. J. **61**, 383 (1956).

[11] P. Coles and B. Jones, A lognormal model for the cosmological mass distribution, Mon. Not. R. Astron. Soc. **248**, 1 (1991).

[12] R. K. Sheth, Constrained realizations and minimum variance reconstruction of non-Gaussian random fields, Mon. Not. R. Astron. Soc. **277**, 933 (1995).

[13] I. Kayo, A. Taruya, and Y. Suto, Probability distribution function of cosmological density fluctuations from a gaussian initial condition: Comparison of one-point and two-point lognormal model predictions with N-body simulations, Astrophys. J. **561**, 22 (2001).

[14] R. Vio, P. Andreani, and W. Wamsteker, Numerical simulation of non-gaussian random fields with prescribed correlation structure, Publ. Astron. Soc. Pac. **113**, 1009 (2001).

[15] M. C. Neyrinck, I. Szapudi, and A. S. Szalay, Rejuvenating the matter power spectrum: Restoring information with a logarithmic density mapping, Astrophys. J. Lett. **698**, L90 (2009).

[16] S. Dorn, T. A. Enßlin, M. Greiner, M. Selig, and V. Boehm, Signal inference with unknown response: Calibration-uncertainty renormalized estimator, Phys. Rev. E **91**, 013311 (2015).

[17] T. A. Enßlin, H. Junklewitz, L. Winderling, M. Greiner, and M. Selig, Improving self-calibration, Phys. Rev. E **90**, 043301 (2014).

[18] J. E. Campbell, On a law of combination of operators (second paper), Proc. London Math. Soc. **1**, 14 (1897).