Denoising, Deconvolving, and Decomposing Photon Observations

Derivation of the D³PO Algorithm

Marco Selig[1,2] and Torsten A. Enßlin[1,2]

[1] Max Planck Institut für Astrophysik (Karl-Schwarzschild-Straße 1, D-85748 Garching, Germany)
[2] Ludwig-Maximilians-Universität München (Geschwister-Scholl-Platz 1, D-80539 München, Germany)

ABSTRACT

The analysis of astronomical images is a non-trivial task. The D³PO algorithm addresses the inference problem of denoising, deconvolving, and decomposing photon observations. Its primary goal is the simultaneous but individual reconstruction of the diffuse and point-like photon flux given a single photon count image, where the fluxes are superimposed. In order to discriminate between these morphologically different signal components, a probabilistic algorithm is derived in the language of information field theory based on a hierarchical Bayesian parameter model. The signal inference exploits prior information on the spatial correlation structure of the diffuse component and the brightness distribution of the spatially uncorrelated point-like sources. A maximum *a posteriori* solution and a solution minimizing the Gibbs free energy of the inference problem using variational Bayesian methods are discussed. Since the derivation of the solution is not dependent on the underlying position space, the implementation of the D³PO algorithm uses the NIFTY package to ensure applicability to various spatial grids and at any resolution. The fidelity of the algorithm is validated by the analysis of simulated data, including a realistic high energy photon count image showing a 32×32 arcmin^2 observation with a spatial resolution of 0.1 arcmin. In all tests the D³PO algorithm successfully denoised, deconvolved, and decomposed the data into a diffuse and a point-like signal estimate for the respective photon flux components.

Key words. methods: data analysis – methods: numerical – methods: statistical – techniques: image processing – gamma-rays: general – X-rays: general

1. Introduction

An astronomical image might display multiple superimposed features, such as point sources, compact objects, diffuse emission, or background radiation. The raw photon count images delivered by high energy telescopes are far from perfect; they suffer from shot noise and distortions caused by instrumental effects. The analysis of such astronomical observations demands elaborate denoising, deconvolution, and decomposition strategies.

The data obtained by the detection of individual photons is subject to Poissonian shot noise which is more severe for low count rates. This causes difficulty for the discrimination of faint sources against noise, and makes their detection exceptionally challenging. Furthermore, uneven or incomplete survey coverage and complex instrumental response functions leave imprints in the photon data. As a result, the data set might exhibit gaps and artificial distortions rendering the clear recognition of different features a difficult task. Point-like sources are especially afflicted by the instrument's point spread function (PSF) that smooths them out in the observed image, and therefore can cause fainter ones to vanish completely in the background noise.

In addition to such noise and convolution effects, it is the superposition of the different objects that makes their separation ambiguous, if possible at all. In astrophysics, photon emitting objects are commonly divided into two morphological classes, diffuse sources and point sources. Diffuse sources span rather smoothly across large fractions of an image, and exhibit apparent internal correlations. Point sources, on the contrary, are local features that, if observed perfectly, would only appear in one pixel of the image. In this work, we will not distinguish between diffuse sources and background, both are diffuse contributions. Intermediate cases, which are sometimes classified as extended or compact sources, are also not considered here.

The question arises of how to reconstruct the original source contributions, the individual signals, that caused the observed photon data. This task is an ill-posed inverse problem without a unique solution. There are a number of heuristic and probabilistic approaches that address the problem of denoising, deconvolution, and decomposition in partial or simpler settings.

SEXTRACTOR (Bertin & Arnouts 1996) is one of the heuristic tools and the most prominent for identifying sources in astronomy. Its popularity is mostly based on its speed and easy operability. However, SEXTRACTOR produces a catalog of fitted sources rather than denoised and deconvolved signal estimates. CLEAN (Högbom 1974) is commonly used in radio astronomy and attempts a deconvolution assuming there are only contributions from point sources. Therefore, diffuse emission is not optimally reconstructed in the analysis of real observations using CLEAN.
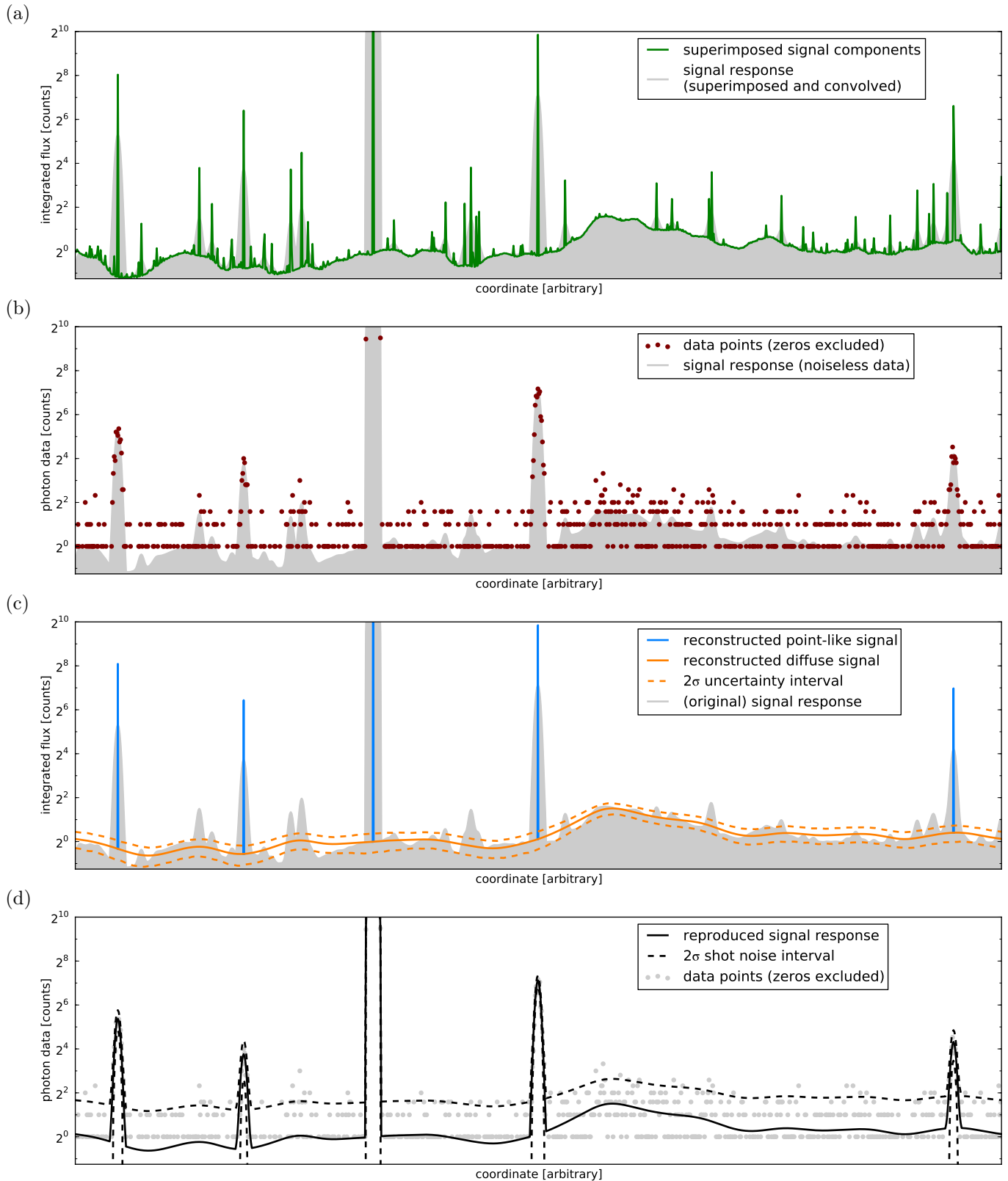
Fig. 1. Illustration of a 1D reconstruction scenario with 1024 pixels. Panel (a) shows the superimposed diffuse and point-like signal components (green solid line) and its observational response (gray contour). Panel (b) shows again the signal response representing noiseless data (gray contour) and the generated Poissonian data (red markers). Panel (c) shows the reconstruction of the point-like signal component (blue solid line), the diffuse one (orange solid line), its 2σ reconstruction uncertainty interval (orange dashed line), and again the original signal response (gray contour). The point-like signal comprises 1024 point-sources of which only five are not too faint. Panel (d) shows the reproduced signal response representing noiseless data (black solid line), its 2σ shot noise interval (black dashed line), and again the data (gray markers).

Multiscale extensions of CLEAN (Cornwell 2008; Rau & Cornwell 2011) improve on this, but are also not prefect (Junklewitz et al. 2013). Decomposition techniques for diffuse backgrounds, based on the analysis of angular power spectra have recently been proposed by Hensley et al. (2013).

Inference methods, in contrast, investigate the probabilistic relation between the data and the signals. Here, the signals of interest are the different source contributions. Probabilistic approaches allow a transparent incorporation of model and *a priori* assumptions, but often result in computationally heavier algorithms.

As an initial attempt, a maximum likelihood analysis was proposed by Valdes (1982). In later work, maximum entropy (Strong 2003) and minimum χ^2 methods (e.g., Bouchet et al. 2013) were applied to the INTEGRAL/SPI data reconstructing a single signal component, though. On the basis of sparse regularization a number of techniques exploiting waveforms (based on the work by Haar 1910, 1911) have proven successful in performing denoising and deconvolution tasks in different settings (González-Nuevo et al. 2006; Willett & Nowak 2007; Dupe et al. 2009; Figueiredo & Bioucas-Dias 2010; Dupé et al. 2011). For example, Schmitt et al. (2010, 2012) analyzed simulated (single and multi-channel) data from the Fermi γ-ray space telescope focusing on the removal of Poisson noise and deconvolution or background separation. Furthermore, a (generalized) morphological component analysis denoised, deconvolved and decomposed simulated radio data assuming Gaussian noise statistics (Bobin et al. 2007; Chapman et al. 2013).

Still in the regime of Gaussian noise, Giovannelli & Coulais (2005) derived a deconvolution algorithm for point and extended sources minimizing regularized least squares. They introduce an efficient convex regularization scheme at the price of *a priori* unmotivated fine tuning parameters. The fast algorithm PowellSnakes I/II by Carvalho et al. (2009, 2012) is capable of analyzing multi-frequency data sets and detecting point-like objects within diffuse emission regions. It relies on matched filters using PSF templates and Bayesian filters exploiting, among others, priors on source position, size, and number. PowellSnakes II has been successfully applied to the Planck data (Planck Collaboration et al. 2011).

The approach closest to ours is the background-source separation technique used to analyze the ROSAT data (Guglielmetti et al. 2009). This Bayesian model is based on a two-component mixture model that reconstructs extended sources and (diffuse) background concurrently. The latter is, however, described by a spline model with a small number of spline sampling points.

The strategy presented in this work aims at the simultaneous reconstruction of two signals, the diffuse and point-like photon flux. Both fluxes contribute equally to the observed photon counts, but their morphological imprints are very different. The proposed algorithm, derived in the framework of information field theory (IFT) (Enßlin et al. 2009; Enßlin 2013), therefore incorporates prior assumptions in form of a hierarchical parameter model. The fundamentally different morphologies of diffuse and point-like contributions reflected in different prior correlations and statistics. The exploitation of these different prior models is crucial to the signal decomposition. In this work, we exclusively consider Poissonian noise, in particular, but not exclusively, in the low count rate regime, where the signal-to-noise ratio becomes challengingly low. The D³PO algorithm presented here targets the simultaneous denoising, deconvolution, and decomposition of photon observations into two signals, the diffuse and point-like photon flux. This task includes the reconstruction of the harmonic power spectrum of the diffuse component from the data themselves. Moreover, the proposed algorithm provides *a posteriori* uncertainty information on both inferred signals.

The fluxes from diffuse and point-like sources contribute equally to the observed photon counts, but their morphological imprints are very different. The proposed algorithm, derived in the framework of information field theory (IFT) (Enßlin et al. 2009; Enßlin 2013, 2014), therefore incorporates prior assumptions in form of a hierarchical parameter model. The fundamentally different morphologies of diffuse and point-like contributions reflected in different prior correlations and statistics. The exploitation of these different prior models is key to the signal decomposition.

The diffuse and point-like signal are treated as two separate signal fields. A signal field represents an original signal appearing in nature; e.g., the physical photon flux distribution of one source component as a function of real space or sky position. In theory, a field has infinitely many degrees of freedom being defined on a continuous position space. In computational practice, however, a field needs of course to be defined on a finite grid. It is desirable that the signal field is reconstructed independently from the grid's resolution, except for potentially unresolvable features.[1] We note that the point-like signal field hosts one point source in every pixel, however, most of them might be invisibly faint. Hence, a complicated determination of the number of point sources, as many algorithms perform, is not required in our case.

The derivation of the algorithm makes use of a wide range of Bayesian methods that are discussed below in detail with regard to their implications and applicability. For now, let us consider an example to emphasize the range and performance of the D³PO algorithm. Figure 1 illustrates a reconstruction scenario in one dimension, where the coordinate could be an angle or position (or time, or energy) in order to represent a 1D sky (or a time series, or an energy spectrum). The numerical implementation uses the NIFTY[2] package (Selig et al. 2013). NIFTY permits an algorithm to be set up abstractly, independent of the finally chosen topology, dimension, or resolution of the underlying position space. In this way, a 1D prototype code can be used for development, and then just be applied in 2D, 3D, or even on the sphere.

The remainder of this paper is structured as follows. Sec. 2 discusses the inference on photon observations; i.e., the underlying model and prior assumptions. The D³PO algorithm solving this inference problem by denoising, decon-

---

[1] If the resolution of the reconstruction were increased gradually, the diffuse signal field might exhibit more and more small scale features until the information content of the given data is exhausted. From this point on, any further increase in resolution would not change the signal field reconstruction significantly. In a similar manner, the localization accuracy and number of detections of point sources might increase with the resolution until all relevant information of the data was captured. All higher resolution grids can then be regarded as acceptable representations of the continuous position space.

[2] NIFTY homepage `http://www.mpa-garching.mpg.de/ift/nifty/`

volution, and decomposition is derived in Sec. 3. In Sec. 4 the algorithm is demonstrated in a numerical application on simulated high energy photon data. We conclude in Sec. 5.

2. Inference on photon observations

2.1. Signal inference

Here, a signal is defined as an unknown quantity of interest that we want to learn about. The most important information source on a signal is the data obtained in an observation to measure the signal. Inferring a signal from an observational data set poses a fundamental problem because of the presence of noise in the data and the ambiguity that several possible signals could have produced the same data, even in the case of negligible noise.

For example, given some image data like photon counts, we want to infer the underlying photon flux distribution. This physical flux is a continuous scalar field that varies with respect to time, energy, and observational position. The measured photon count data, however, is restricted by its spatial and energy binning, as well as its limitations in energy range and observation time. Basically, all data sets are finite for practical reasons, and therefore cannot capture all of the infinitely many degrees of freedom of the underlying continuous signal field.

There is no exact solution to such signal inference problems, since there might be (infinitely) many signal field configurations that could lead to the same data. This is why a probabilistic data analysis, which does not pretend to calculate the correct field configuration but provides expectation values and uncertainties of the signal field, is appropriate for signal inference.

Given a data set $\boldsymbol{d}$, the *a posteriori* probability distribution $P(\boldsymbol{s}|\boldsymbol{d})$ judges how likely a potential signal $\boldsymbol{s}$ is. This posterior is given by Bayes' theorem,

$$P(\boldsymbol{s}|\boldsymbol{d}) = \frac{P(\boldsymbol{d}|\boldsymbol{s})\,P(\boldsymbol{s})}{P(\boldsymbol{d})}, \tag{1}$$

as a combination of the likelihood $P(\boldsymbol{d}|\boldsymbol{s})$, the signal prior $P(\boldsymbol{s})$, and the evidence $P(\boldsymbol{d})$, which serves as a normalization. The likelihood characterizes how likely it is to measure data set $\boldsymbol{d}$ from a given signal field $\boldsymbol{s}$. It covers all processes that are relevant for the measurement of $\boldsymbol{d}$. The prior describes the knowledge about $\boldsymbol{s}$ without considering the data, and should, in general, be less restrictive than the likelihood.

IFT is a Bayesian framework for the inference of signal fields exploiting mathematical methods for theoretical physics. A signal field, $\boldsymbol{s} = s(x)$, is a function of a continuous position x in some position space Ω. In order to avoid a dependence of the reconstruction on the partition of Ω, the according calculus regarding fields is geared to preserve the continuum limit (Enßlin 2013, 2014; Selig et al. 2013). In general, we are interested in the *a posteriori* mean estimate $\boldsymbol{m}$ of the signal field given the data, and its (uncertainty) covariance $\boldsymbol{D}$, defined as

$$\boldsymbol{m} = \langle \boldsymbol{s} \rangle_{(\boldsymbol{s}|\boldsymbol{d})} = \int \mathcal{D}\boldsymbol{s}\, \boldsymbol{s}\, P(\boldsymbol{s}|\boldsymbol{d}), \tag{2}$$

$$\boldsymbol{D} = \left\langle (\boldsymbol{m}-\boldsymbol{s})(\boldsymbol{m}-\boldsymbol{s})^{\dagger} \right\rangle_{(\boldsymbol{s}|\boldsymbol{d})}, \tag{3}$$

where $\dagger$ denotes adjunction and $\langle \cdot \rangle_{(\boldsymbol{s}|\boldsymbol{d})}$ the expectation value with respect to the posterior probability distribution $P(\boldsymbol{s}|\boldsymbol{d})$.[3]

In the following, the posterior of the physical photon flux distribution of two morphologically different source components given a data set of photon counts is build up piece by piece according to Eq. (1).

2.2. Poissonian likelihood

The images provided by astronomical high energy telescopes typically consist of integer photon counts that are binned spatially into pixels. Let d_i be the number of detected photons, also called events, in pixel i, where $i \in \{1, \ldots, N_{\mathrm{pix}}\} \subset \mathbb{N}$.

The kind of signal field we would like to infer from such data is the causative photon flux distribution. The photon flux, $\boldsymbol{\rho} = \rho(x)$, is defined for each position x on the observational space Ω. In astrophysics, this space Ω is typically the $\mathcal{S}^2$ sphere representing an all-sky view, or a region within $\mathbb{R}^2$ representing an approximately plane patch of the sky. The flux $\boldsymbol{\rho}$ might express different morphological features, which can be classified into a diffuse and point-like component. The exact definitions of the diffuse and point-like flux should be specified *a priori*, without knowledge of the data, and are addressed in Sec. 2.3.1 and 2.3.3, respectively. At this point it shall suffices to say that the diffuse flux varies smoothly on large spatial scales, while the flux originating from point sources is fairly local. These two flux components are superimposed,

$$\boldsymbol{\rho} = \boldsymbol{\rho}_{\mathrm{diffuse}} + \boldsymbol{\rho}_{\mathrm{point-like}} = \rho_0 \left(\mathrm{e}^{\boldsymbol{s}} + \mathrm{e}^{\boldsymbol{u}} \right), \tag{4}$$

where we introduced the dimensionless diffuse and point-like signal fields, $\boldsymbol{s}$ and $\boldsymbol{u}$, and the constant ρ_0 which absorbs the physical dimensions of the photon flux; i.e., events per area per energy and time interval. The exponential function in Eq. (4) is applied componentwise. In this way, we naturally account for the strict positivity of the photon flux at the price of a non-linear change of variables, from the flux to its natural logarithm.

A measurement apparatus observing the photon flux $\boldsymbol{\rho}$ is expected to detect a certain number of photons $\boldsymbol{\lambda}$. This process can be modeled by a linear response operator $\boldsymbol{R}_0$ as follows,

$$\boldsymbol{\lambda} = \boldsymbol{R}_0 \boldsymbol{\rho} = \boldsymbol{R} \left(\mathrm{e}^{\boldsymbol{s}} + \mathrm{e}^{\boldsymbol{u}} \right), \tag{5}$$

where $\boldsymbol{R} = \boldsymbol{R}_0 \rho_0$. This reads for pixel i,

$$\lambda_i = \int_{\Omega} \mathrm{d}x\, R_i(x) \left(\mathrm{e}^{s(x)} + \mathrm{e}^{u(x)} \right). \tag{6}$$

The response operator $\boldsymbol{R}_0$ comprises all aspects of the measurement process; i.e., all instrument response functions. This includes the survey coverage, which describes the instrument's overall exposure to the observational area, and the instrument's PSF, which describes how a point source is imaged by the instrument.

The superposition of different components and the transition from continuous coordinates to some discrete pixelization, cf. Eq. (6), cause a severe loss of information

---

[3] This expectation value is computed by a path integral, $\int \mathcal{D}\boldsymbol{s}$, over the complete phase space of the signal field $\boldsymbol{s}$; i.e., all possible field configurations.

about the original signal fields. In addition to that, measurement noise distorts the signal's imprint in the data. The individual photon counts per pixel can be assumed to follow a Poisson distribution $\mathcal{P}$ each. Therefore, the likelihood of the data $\boldsymbol{d}$ given an expected number of events $\boldsymbol{\lambda}$ is modeled as a product of statistically independent Poisson processes,

$$P(\boldsymbol{d}|\boldsymbol{\lambda}) = \prod_i \mathcal{P}(d_i, \lambda_i) = \prod_i \frac{1}{d_i!} \lambda_i^{d_i} e^{-\lambda_i}. \tag{7}$$

The Poisson distribution has a signal-to-noise ratio of $\sqrt{\boldsymbol{\lambda}}$ which scales with the expected number of photon counts. Therefore, Poissonian shot noise is most severe in regions with low photon fluxes. This makes the detection of faint sources in high energy astronomy a particularly challenging task, as X- and γ-ray photons are sparse.

The likelihood of photon count data given a two component photon flux is hence described by the Eqs. (5) and (7). Rewriting this likelihood $P(\boldsymbol{d}|\boldsymbol{s}, \boldsymbol{u})$ in form of its negative logarithm yields the information Hamiltonian $H(\boldsymbol{d}|\boldsymbol{s}, \boldsymbol{u})$,[4]

$$H(\boldsymbol{d}|\boldsymbol{s}, \boldsymbol{u}) = -\log P(\boldsymbol{d}|\boldsymbol{s}, \boldsymbol{u}) \tag{8}$$

$$= H_0 + \mathbf{1}^\dagger \boldsymbol{\lambda} - \boldsymbol{d}^\dagger \log(\boldsymbol{\lambda}) \tag{9}$$

$$= H_0 + \mathbf{1}^\dagger \boldsymbol{R}\left(e^{\boldsymbol{s}} + e^{\boldsymbol{u}}\right) - \boldsymbol{d}^\dagger \log\left(\boldsymbol{R}\left(e^{\boldsymbol{s}} + e^{\boldsymbol{u}}\right)\right), \tag{10}$$

where the ground state energy H_0 comprises all terms constant in $\boldsymbol{s}$ and $\boldsymbol{u}$, and $\mathbf{1}$ is a constant data vector being 1 everywhere.

2.3. Prior assumptions

The diffuse and point-like signal fields, $\boldsymbol{s}$ and $\boldsymbol{u}$, contribute equally to the likelihood defined by Eq. (10), and thus leaving it completely degenerate. On the mere basis of the likelihood, the full data set could be explained by the diffuse signal alone, or only by point-sources, or any other conceivable combination. In order to downweight intuitively implausible solutions, we introduce priors. The priors discussed in the following address the morphology of the different photon flux contributions, and define diffuse and point-like in the first place. These priors aid the reconstruction by providing some remedy for the degeneracy of the likelihood. The likelihood describes noise and convolution properties, and the prior describe the individual morphological properties. Therefore, the denoising and deconvolution of the data towards the total photon flux $\boldsymbol{\rho}$ is primarily likelihood driven, but for a decomposition of the total photon flux into $\boldsymbol{\rho}^{(s)}$ and $\boldsymbol{\rho}^{(u)}$, the signal priors are imperative.

2.3.1. Diffuse component

The diffuse photon flux, $\boldsymbol{\rho}^{(s)} = \rho_0 e^{\boldsymbol{s}}$, is strictly positive and might vary in intensity over several orders of magnitude. Its morphology shows cloudy patches with smooth fluctuations across spatial scales; i.e., one expects similar values of the diffuse flux in neighboring locations. In other words, the diffuse component exhibits spatial correlations. A log-normal model for $\boldsymbol{\rho}^{(s)}$ satisfies those requirements according to the

maximum entropy principle (Oppermann et al. 2012; Kinney 2013). If the diffuse photon flux follows a multivariate log-normal distribution, the diffuse signal field $\boldsymbol{s}$ obeys a multivariate Gaussian distribution $\mathcal{G}$,

$$P(\boldsymbol{s}|\boldsymbol{S}) = \mathcal{G}(\boldsymbol{s}, \boldsymbol{S}) = \frac{1}{\sqrt{\det[2\pi\boldsymbol{S}]}} \exp\left(-\frac{1}{2}\boldsymbol{s}^\dagger \boldsymbol{S}^{-1}\boldsymbol{s}\right), \tag{11}$$

with a given covariance $\boldsymbol{S} = \left\langle \boldsymbol{s}\boldsymbol{s}^\dagger \right\rangle_{(\boldsymbol{s}|\boldsymbol{S})}$. This covariance describes the strength of the spatial correlations, and thus the smoothness of the fluctuations.

A convenient parametrization of the covariance $\boldsymbol{S}$ can be found, if the signal field $\boldsymbol{s}$ is *a priori* not known to distinguish any position or orientation axis; i.e., its correlations only depend on relative distances. This is equivalent to assume $\boldsymbol{s}$ to be statistically homogeneous and isotropic. Under this assumption, $\boldsymbol{S}$ is diagonal in the harmonic basis[5] of the position space Ω such that

$$\boldsymbol{S} = \sum_k e^{\tau_k} \boldsymbol{S}_k, \tag{12}$$

where τ_k are spectral parameters and $\boldsymbol{S}_k$ are projections onto a set of disjoint harmonic subspaces of Ω. These subspaces are commonly denoted as spectral bands or harmonic modes. The set of spectral parameters, $\boldsymbol{\tau} = \{\tau_k\}_k$, is then the logarithmic power spectrum of the diffuse signal field $\boldsymbol{s}$ with respect to the chosen harmonic basis denoted by k.

However, the diffuse signal covariance is in general unknown *a priori*. This requires the introduction of another prior for the covariance, or for the set of parameters $\boldsymbol{\tau}$ describing it adequately. This approach of hyperpriors on prior parameters creates a hierarchical parameter model.

2.3.2. Unknown power spectrum

The lack of knowledge of the power spectrum, requires its reconstruction from the same data the signal is inferred from (Wandelt et al. 2004; Jasche et al. 2010; Enßlin & Frommert 2011; Jasche & Wandelt 2013). Therefore, two *a priori* constraints for the spectral parameters $\boldsymbol{\tau}$, which describe the logarithmic power spectrum, are incorporated in the model.

The power spectrum is unknown and might span over several orders of magnitude. This implies a logarithmically uniform prior for each element of the power spectrum, and a uniform prior for each spectral parameter τ_k, respectively. We initially assume independent inverse-Gamma distributions $\mathcal{I}$ for the individual elements,

$$P(e^{\boldsymbol{\tau}}|\boldsymbol{\alpha}, \boldsymbol{q}) = \prod_k \mathcal{I}(e^{\tau_k}, \alpha_k, q_k) \tag{13}$$

$$= \prod_k \frac{q_k^{\alpha_k - 1}}{\Gamma(\alpha_k - 1)} e^{-\left(\alpha_k \tau_k + q_k e^{-\tau_k}\right)}, \tag{14}$$

and hence

$$P_{\mathrm{un}}(\boldsymbol{\tau}|\boldsymbol{\alpha}, \boldsymbol{q}) = \prod_k \mathcal{I}(e^{\tau_k}, \alpha_k, q_k) \left|\frac{de^{\tau_k}}{d\tau_k}\right| \tag{15}$$

$$\propto \exp\left(-(\boldsymbol{\alpha} - \mathbf{1})^\dagger \boldsymbol{\tau} - \boldsymbol{q}^\dagger e^{-\boldsymbol{\tau}}\right), \tag{16}$$

---

[4] Throughout this work we define $H(\cdot) = -\log P(\cdot)$, and absorb constant terms into a normalization constant H_0 in favor of clarity.

[5] The basis in which the Laplace operator is diagonal is denoted harmonic basis. If Ω is a n-dimensional Euclidean space $\mathbb{R}^n$ or Torus $\mathcal{T}^n$, the harmonic basis is the Fourier basis; if Ω is the $\mathcal{S}^2$ sphere, the harmonic basis is the spherical harmonics basis.

where $\boldsymbol{\alpha} = \{\alpha_k\}_k$ and $\boldsymbol{q} = \{q_k\}_k$ are the shape and scale parameters, and Γ denotes the Gamma function. In the limit of $\alpha_k \to 1$ and $q_k \to 0 \; \forall k$, the inverse-Gamma distributions become asymptotically flat on a logarithmic scale, and thus P_{un} constant.[6] Small non-zero scale parameters, $0 < q_k$, provide lower limits for the power spectrum that, in practice, lead to more stable inference algorithms.

So far, the variability of the individual elements of the power spectrum is accounted for, but the question about their correlations has not been addressed. Empirically, power spectra of a diffuse signal field do not exhibit wild fluctuation or change drastically over neighboring modes. They rather show some sort of spectral smoothness. Moreover, for diffuse signal fields that were shaped by local and causal processes, we might expect a finite correlation support in position space. This translates into a smooth power spectrum. In order to incorporate spectral smoothness, we employ a prior introduced by Enßlin & Frommert (2011); Oppermann et al. (2012). This prior is based on the second logarithmic derivative of the spectral parameters $\boldsymbol{\tau}$, and favors power spectra that obey a power law. It reads

$$P_{\mathrm{sm}}(\boldsymbol{\tau}|\boldsymbol{\sigma}) \propto \exp\left(-\frac{1}{2}\boldsymbol{\tau}^{\dagger}\boldsymbol{T}\boldsymbol{\tau}\right), \qquad (17)$$

with

$$\boldsymbol{\tau}^{\dagger}\boldsymbol{T}\boldsymbol{\tau} = \int \mathrm{d}(\log k)\frac{1}{\sigma_k^2}\left(\frac{\partial^2 \tau_k}{\partial(\log k)^2}\right)^2, \qquad (18)$$

where $\boldsymbol{\sigma} = \{\sigma_k\}_k$ are Gaussian standard deviations specifying the tolerance against deviation from a power-law behavior of the power spectrum. A choice of $\sigma_k = 1 \; \forall k$ would typically allow for a change in the power law's slope of 1 per e-fold in k. In the limit of $\sigma_k \to \infty \; \forall k$, no smoothness is enforced upon the power spectrum.

The resulting prior for the spectral parameters is given by the product of the priors discussed above,

$$P(\boldsymbol{\tau}|\boldsymbol{\alpha},\boldsymbol{q},\boldsymbol{\sigma}) = P_{\mathrm{un}}(\boldsymbol{\tau}|\boldsymbol{\alpha},\boldsymbol{q}) \, P_{\mathrm{sm}}(\boldsymbol{\tau}|\boldsymbol{\sigma}). \qquad (19)$$

The parameters $\boldsymbol{\alpha}, \boldsymbol{q}$ and $\boldsymbol{\sigma}$ are considered to be given as part of the hierarchical Bayesian model, and provide a flexible handle to model our knowledge on the scaling and smoothness of the power spectrum.

2.3.3. Point-like component

The point-like photon flux, $\boldsymbol{\rho}^{(u)} = \rho_0 e^{\boldsymbol{u}}$, is supposed to originate from very distant astrophysical sources. These sources appear morphologically point-like to an observer because their actual extent is negligible given the extreme distances. This renders point sources to be spatially local phenomena. The photon flux contributions of neighboring point sources can (to zeroth order approximation) be assumed to be statistically independent of each other. Even if the two sources are very close on the observational plane, their physical distance might be huge. Even in practice, the spatial cross-correlation of point sources is negligible. Therefore, statistically independent priors for the photon flux contribution of each point-source are assumed in the following.

Because of the spatial locality of a point source, the corresponding photon flux signal is supposed to be confined to a single spot, too. If the point-like signal field, defined over a continuous position space Ω, is discretized properly[7], this spot is sufficiently identified by an image pixel in the reconstruction. A discretization, $\rho(x \in \Omega) \to (\rho_x)_x$, is an inevitable step since the algorithm is to be implemented in a computer environment anyway. Nevertheless, we have to ensure that the *a priori* assumptions do not depend on the chosen discretization but satisfy the continuous limit.

Therefore, the prior for the point-like signal component factorizes spatially,

$$P(\boldsymbol{\rho}^{(u)}) = \prod_x P(\rho_x^{(u)}), \qquad (20)$$

but the functional form of the priors are yet to be determined. This model allows the point-like signal field to host one point source in every pixel. Most of these point sources are expected to be invisibly faint contributing negligibly to the total photon flux. However, the point sources which are just identifiable from the data are pinpointed in the reconstruction. In this approach, there is no necessity for a complicated determination of the number and position of sources.

For the construction of a prior, that the photon flux is a strictly positive quantity also needs to be considered. Thus, a simple exponential prior,

$$P(\rho_x^{(u)}) \propto \exp\left(-\rho_x^{(u)}/\rho_0\right), \qquad (21)$$

has been suggested (e.g., Guglielmetti et al. 2009). It has the advantage of being (easily) analytically treatable, but its physical implications are questionable. This distribution strongly suppresses high photon fluxes in favor of lower ones. The maximum entropy prior, which is also often applied, is even worse because it corresponds to a brightness distribution,[8]

$$P(\rho_x^{(u)}) \propto \left(\rho_x^{(u)}/\rho_0\right)^{\left(-\rho_x^{(u)}/\rho_0\right)}. \qquad (22)$$

The following (rather crude) consideration might motivate a more astrophysical prior. Say the universe hosts a homogeneous distribution of point sources. The number of point sources would therefore scale with the observable volume; i.e., with distance cubed. Their apparent brightness, which is reduced because of the spreading of the light rays; i.e., a proportionality to the distance squared. Consequently, a power-law behavior between the number of point sources and their brightness with a slope $\beta = \frac{3}{2}$ is to be expected (Fomalont 1968; Malyshev & Hogg 2011). However, such a plain power law diverges at 0, and is not necessarily normalizable. Furthermore, Galactic and extragalactic sources cannot be found in arbitrary distances owing to the finite size of the Galaxy and the cosmic (past) light cone. Imposing an exponential cut-off above 0 onto the power law yields an inverse-Gamma distribution, which has been shown to be an appropriate prior for point-like photon fluxes (Guglielmetti et al. 2009; Carvalho et al. 2009, 2012).

---

[6] If $P(\tau_k = \log z) = \mathrm{const.}$, then a substitution yields $P(z) = P(\log z) \, |\mathrm{d}(\log z)/\mathrm{d}z| \propto z^{-1} \sim \mathcal{I}(z, \alpha \to 1, q \to 0)$.

[7] The numerical discretization of information fields is described in great detail in Selig et al. (2013).

[8] The so-called maximum entropy regularization $\sum_x (\rho_x^{(u)}/\rho_0) \log(\rho_x^{(u)}/\rho_0)$ of the log-likelihood can be regarded as log-prior, cf. Eqs. (20) and (22).

The prior for the point-like signal field is therefore derived from a product of independent inverse-Gamma distributions,[9]

$$P(\boldsymbol{\rho}^{(u)}|\boldsymbol{\beta},\boldsymbol{\eta}) = \prod_x \mathcal{I}(\rho_x^{(u)}, \beta_x, \rho_0\eta_x) \qquad (23)$$

$$= \prod_x \frac{(\rho_0\eta_x)^{\beta_x-1}}{\Gamma(\beta_x-1)} \left(\rho_x^{(u)}\right)^{-\beta_x} \exp\left(-\frac{\rho_0\eta_x}{\rho_x^{(u)}}\right), \qquad (24)$$

yielding

$$P(\boldsymbol{u}|\boldsymbol{\beta},\boldsymbol{\eta}) = \prod_x \mathcal{I}(\rho_0\mathrm{e}^{u_x}, \beta_k, \rho_0\eta_k) \left|\frac{\mathrm{d}\rho_0\mathrm{e}^{u_x}}{\mathrm{d}u_x}\right| \qquad (25)$$

$$\propto \exp\left(-(\boldsymbol{\beta}-\mathbf{1})^\dagger\boldsymbol{u} - \boldsymbol{\eta}^\dagger\mathrm{e}^{-\boldsymbol{u}}\right), \qquad (26)$$

where $\boldsymbol{\beta} = \{\beta_x\}_x$ and $\boldsymbol{\eta} = \{\eta_x\}_x$ are the shape and scale parameters. The latter is responsible for the cut-off of vanishing fluxes, and should be chosen adequately small in analogy to the spectral scale parameters $\boldsymbol{q}$. The determination of the shape parameters is more difficile. The geometrical argument above suggests a universal shape parameter, $\beta_x = \frac{3}{2} \, \forall x$. A second argument for this value results from demanding *a priori* independence of the discretization. If we choose a coarser resolution that would add up the flux from two point sources at merged pixels, then our prior should still be applicable. The universal value of $\frac{3}{2}$ indeed fulfills this requirement as shown in Appendix A. There it is also shown that η has to be chosen resolution dependent, though.

2.4. Parameter model

Figure 2 gives an overview of the parameter hierarchy of the suggested Bayesian model. The data $\boldsymbol{d}$ is given, and the diffuse signal field $\boldsymbol{s}$ and the point-like signal field $\boldsymbol{u}$ shall be reconstructed from that data. The logarithmic power spectrum $\boldsymbol{\tau}$ is a set of nuisance parameters that also need to be reconstructed from the data in order to accurately model the diffuse flux contributions. The model parameters form the top layer of this hierarchy and are given to the reconstruction algorithm. This set of model parameters can be boiled down to five scalars, namely α, q, σ, β, and η, if one defines $\boldsymbol{\alpha} = \alpha\mathbf{1}$, etc. The incorporation of the scalars in the inference is possible in theory, but this would increase the computational complexity dramatically.

We discussed reasonable values for these scalars to be chosen *a priori*. If additional information sources, such as theoretical power spectra or object catalogs, are available the model parameters can be adjusted accordingly. In Sec. 4, different parameter choices for the analysis of simulated data are investigated.

3. Denoising, deconvolution, and decomposition

The likelihood model, describing the measurement process, and the prior assumptions for the signal fields and the power

---

[9] A possible extension of this prior model that includes spatial correlations would be an inverse-Wishart distribution for $\mathrm{diag}[\boldsymbol{\rho}^{(u)}]$.
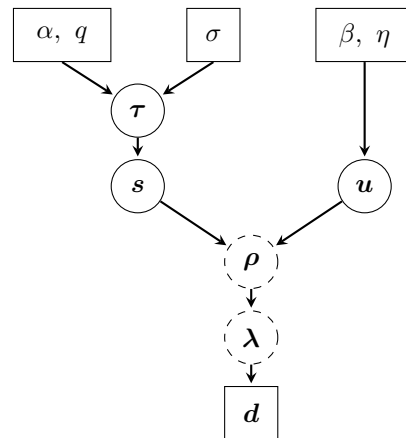

Fig. 2. Graphical model of the model parameters α, q, σ, β, and η, the logarithmic spectral parameters $\boldsymbol{\tau}$, the diffuse signal field $\boldsymbol{s}$, the point-like signal field $\boldsymbol{u}$, the total photon flux $\boldsymbol{\rho}$, the expected number of photons $\boldsymbol{\lambda}$, and the observed photon count data $\boldsymbol{d}$.

spectrum of the diffuse component yield a well-defined inference problem. The corresponding posterior is given by

$$P(\boldsymbol{s},\boldsymbol{\tau},\boldsymbol{u}|\boldsymbol{d}) = \frac{P(\boldsymbol{d}|\boldsymbol{s},\boldsymbol{u})\,P(\boldsymbol{s}|\boldsymbol{\tau})\,P(\boldsymbol{\tau}|\alpha,q,\sigma)\,P(\boldsymbol{u}|\beta,\eta)}{P(\boldsymbol{d})}, \qquad (27)$$

which is a complex form of Bayes' theorem (1).

Ideally, we would now calculate the *a posteriori* expectation values and uncertainties according to Eqs. (2) and (3) for the diffuse and point-like signal fields, $\boldsymbol{s}$ and $\boldsymbol{u}$, as well as for the logarithmic spectral parameters $\boldsymbol{\tau}$. However, an analytical evaluation of these expectation values is not possible because of the complexity of the posterior.

The posterior is non-linear in the signal fields and, except for artificially constructed data, non-convex. It, however, is more flexible and therefore allows for a more comprehensive description of the parameters to be inferred (Kirkpatrick et al. 1983; Geman & Geman 1984).

Numerical approaches involving Markov chain Monte Carlo methods (Metropolis & Ulam 1949; Metropolis et al. 1953) are possible, but hardly feasible because of the huge parameter phase space. Nevertheless, similar problems have been addressed by elaborate sampling techniques (Wandelt et al. 2004; Jasche et al. 2010; Jasche & Kitaura 2010; Jasche & Wandelt 2013).

Here, two approximative algorithms with lower computational costs are derived. The first one uses the maximum *a posteriori* (MAP) approximation, the second one minimizes the Gibbs free energy of an approximate posterior ansatz in the spirit of variational Bayesian methods. The fidelity and accuracy of these two algorithms are compared in a numerical application in Sec. 4.

3.1. Posterior maximum

The posterior maximum and mean coincide, if the posterior distribution is symmetric and single peaked. In practice, this often holds – at least in good approximation –, so that the maximum *a posteriori* approach can provide suitable estimators. This can either be achieved using a δ-distribution

at the posterior's mode,

$$\langle \boldsymbol{s} \rangle_{(\boldsymbol{s}|\boldsymbol{d})} \overset{\text{MAP-}\delta}{\approx} \int \mathcal{D}\boldsymbol{s} \; \boldsymbol{s} \; \delta(\boldsymbol{s} - \boldsymbol{s}_{\text{mode}}), \tag{28}$$

or using a Gaussian approximation around this point,

$$\langle \boldsymbol{s} \rangle_{(\boldsymbol{s}|\boldsymbol{d})} \overset{\text{MAP-}\mathcal{G}}{\approx} \int \mathcal{D}\boldsymbol{s} \; \boldsymbol{s} \; \mathcal{G}(\boldsymbol{s} - \boldsymbol{s}_{\text{mode}}, \boldsymbol{D}_{\text{mode}}), \tag{29}$$

Both approximations require us to find the mode, which is done by extremizing the posterior.

Instead of the complex posterior distribution, it is convenient to consider the information Hamiltonian, defined by its negative logarithm,

$$H(\boldsymbol{s}, \boldsymbol{\tau}, \boldsymbol{u}|\boldsymbol{d}) = -\log P(\boldsymbol{s}, \boldsymbol{\tau}, \boldsymbol{u}|\boldsymbol{d}) \tag{30}$$

$$= H_0 + \mathbf{1}^\dagger \boldsymbol{R} \left(\mathrm{e}^{\boldsymbol{s}} + \mathrm{e}^{\boldsymbol{u}}\right) - \boldsymbol{d}^\dagger \log\left(\boldsymbol{R}\left(\mathrm{e}^{\boldsymbol{s}} + \mathrm{e}^{\boldsymbol{u}}\right)\right)$$

$$+ \frac{1}{2} \log\left(\det[\boldsymbol{S}]\right) + \frac{1}{2} \boldsymbol{s}^\dagger \boldsymbol{S}^{-1} \boldsymbol{s} \tag{31}$$

$$+ (\boldsymbol{\alpha} - \mathbf{1})^\dagger \boldsymbol{\tau} + \boldsymbol{q}^\dagger \mathrm{e}^{-\boldsymbol{\tau}} + \frac{1}{2} \boldsymbol{\tau}^\dagger \boldsymbol{T} \boldsymbol{\tau}$$

$$+ (\boldsymbol{\beta} - \mathbf{1})^\dagger \boldsymbol{u} + \boldsymbol{\eta}^\dagger \mathrm{e}^{-\boldsymbol{u}},$$

where all terms constant in $\boldsymbol{s}$, $\boldsymbol{\tau}$, and $\boldsymbol{u}$ have been absorbed into a ground state energy H_0, cf. Eqs. (7), (11), (19), and (26), respectively.

The MAP solution, which maximizes the posterior, minimizes the Hamiltonian. This minimum can thus be found by taking the first (functional) derivatives of the Hamiltonian with respect to $\boldsymbol{s}$, $\boldsymbol{\tau}$, and $\boldsymbol{u}$ and equating them with zero. Unfortunately, this yields a set of implicit, self-consistent equations rather than an explicit solution. However, these equations can be solved by an iterative minimization of the Hamiltonian using a steepest descent method for example, see Sec. 3.4 for details.

In order to better understand the structure of the MAP solution, we consider the minimum $(\boldsymbol{s}, \boldsymbol{\tau}, \boldsymbol{u}) = (\boldsymbol{m}^{(s)}, \boldsymbol{\tau}^\star, \boldsymbol{m}^{(u)})$. The resulting filter formulas for the diffuse and point-like signal field read

$$\left.\frac{\partial H}{\partial \boldsymbol{s}}\right|_{\text{min}} = \boldsymbol{0} = (\mathbf{1} - \boldsymbol{d}/\boldsymbol{l})^\dagger \boldsymbol{R} * \mathrm{e}^{\boldsymbol{m}^{(s)}} + \boldsymbol{S}^{\star -1} \boldsymbol{m}^{(s)}, \tag{32}$$

$$\left.\frac{\partial H}{\partial \boldsymbol{u}}\right|_{\text{min}} = \boldsymbol{0} = (\mathbf{1} - \boldsymbol{d}/\boldsymbol{l})^\dagger \boldsymbol{R} * \mathrm{e}^{\boldsymbol{m}^{(u)}} + \boldsymbol{\beta} - \mathbf{1} - \boldsymbol{\eta} * \mathrm{e}^{-\boldsymbol{m}^{(u)}}, \tag{33}$$

with

$$\boldsymbol{l} = \boldsymbol{R}\left(\mathrm{e}^{\boldsymbol{m}^{(s)}} + \mathrm{e}^{\boldsymbol{m}^{(u)}}\right), \tag{34}$$

$$\boldsymbol{S}^\star = \sum_k \mathrm{e}^{\tau_k^\star} \boldsymbol{S}_k. \tag{35}$$

Here, $*$ and $/$ denote componentwise multiplication and division, respectively. The first term in Eq. (32) and (33), which comes from the likelihood, vanishes in case $\boldsymbol{l} = \boldsymbol{d}$. We note that $\boldsymbol{l} = \boldsymbol{\lambda}|_{\text{min}}$ describes the most likely number of photon counts, not the expected number of photon counts $\boldsymbol{\lambda} = \langle \boldsymbol{d} \rangle_{(\boldsymbol{d}|\boldsymbol{s},\boldsymbol{u})}$, cf. Eqs. (5) and (7). Disregarding the regularization by the priors, the solution would overfit; i.e., noise features are partly assigned to the signal fields in order to achieve an unnecessarily close agreement with the

data. However, the *a priori* regularization suppresses this tendency to some extend.

The second derivative of the Hamiltonian describes the curvature around the minimum, and therefore approximates the (inverse) uncertainty covariance,

$$\left.\frac{\partial^2 H}{\partial \boldsymbol{s} \partial \boldsymbol{s}^\dagger}\right|_{\text{min}} \approx \boldsymbol{D}^{(s)\,-1}, \qquad \left.\frac{\partial^2 H}{\partial \boldsymbol{u} \partial \boldsymbol{u}^\dagger}\right|_{\text{min}} \approx \boldsymbol{D}^{(u)\,-1}. \tag{36}$$

The closed form of $\boldsymbol{D}^{(s)}$ and $\boldsymbol{D}^{(u)}$ is given explicitly in Appendix B.

The filter formula for the power spectrum, which is derived from a first derivative of the Hamiltonian with respect to $\boldsymbol{\tau}$, yields

$$\mathrm{e}^{\boldsymbol{\tau}^\star} = \frac{\boldsymbol{q} + \frac{1}{2}\left(\mathrm{tr}\left[\boldsymbol{m}^{(s)} \boldsymbol{m}^{(s)\,\dagger} \boldsymbol{S}_k^{-1}\right]\right)_k}{\boldsymbol{\gamma} + \boldsymbol{T}\boldsymbol{\tau}^\star}, \tag{37}$$

where $\boldsymbol{\gamma} = (\boldsymbol{\alpha} - \mathbf{1}) + \frac{1}{2}\left(\mathrm{tr}\left[\boldsymbol{S}_k \boldsymbol{S}_k^{-1}\right]\right)_k$. This formula is in accordance with the results by Enßlin & Frommert (2011); Oppermann et al. (2012). It has been shown by the former authors that such a filter exhibits a perception threshold; i.e., on scales where the signal-response-to-noise ratio drops below a certain bound the reconstructed signal power becomes vanishingly low. This threshold can be cured by a better capture of the *a posteriori* uncertainty structure.

3.2. Posterior approximation

In order to overcome the analytical infeasibility as well as the perception threshold, we seek an approximation to the true posterior. Instead of approximating the expectation values of the posterior, approximate posteriors are investigated in this section. In case the approximation is good, the expectation values of the approximate posterior should then be close to the real ones.

The posterior given by Eq. (27) is inaccessible because of the entanglement of the diffuse signal field $\boldsymbol{s}$, its logarithmic power spectrum $\boldsymbol{\tau}$, and the point-like signal field $\boldsymbol{u}$. The involvement of $\boldsymbol{\tau}$ can been simplified by a mean field approximation,

$$P(\boldsymbol{s}, \boldsymbol{\tau}, \boldsymbol{u}|\boldsymbol{d}) \approx Q = Q_s(\boldsymbol{s}, \boldsymbol{u}|\boldsymbol{\mu}, \boldsymbol{d})\, Q_\tau(\boldsymbol{\tau}|\boldsymbol{\mu}, \boldsymbol{d}), \tag{38}$$

where $\boldsymbol{\mu}$ denotes an abstract mean field mediating some information between the signal field tuple $(\boldsymbol{s}, \boldsymbol{u})$ and $\boldsymbol{\tau}$ that are separated by the product ansatz in Eq. (38). This mean field is fully determined by the problem, as it represents effective (rather than additional) degrees of freedom. It is only needed implicitly for the derivation, an explicit formula can be found in Appendix C.3, though.

Since the *a posteriori* mean estimates for the signal fields and their uncertainty covariances are of primary interest, a Gaussian approximation for Q_s that accounts for correlation between $\boldsymbol{s}$ and $\boldsymbol{u}$ would be sufficient. Hence, our previous approximation is extended by setting

$$Q_s(\boldsymbol{s}, \boldsymbol{u}|\boldsymbol{\mu}, \boldsymbol{d}) = \mathcal{G}(\boldsymbol{\varphi}, \boldsymbol{D}), \tag{39}$$

with

$$\boldsymbol{\varphi} = \begin{pmatrix} \boldsymbol{s} - \boldsymbol{m}^{(s)} \\ \boldsymbol{u} - \boldsymbol{m}^{(u)} \end{pmatrix}, \qquad \boldsymbol{D} = \begin{pmatrix} \boldsymbol{D}^{(s)} & \boldsymbol{D}^{(su)} \\ \boldsymbol{D}^{(su)\,\dagger} & \boldsymbol{D}^{(u)} \end{pmatrix}. \tag{40}$$

This Gaussian approximation is also a convenient choice in terms of computational complexity because of its simple analytic structure.

The goodness of the approximation $P \approx Q$ can be quantified by an information theoretical measure, see Appendix C.1. The Gibbs free energy of the inference problem,

$$G = \langle H(\boldsymbol{s}, \boldsymbol{\tau}, \boldsymbol{u}|\boldsymbol{d}) \rangle_Q - \langle -\log Q(\boldsymbol{s}, \boldsymbol{\tau}, \boldsymbol{u}|\boldsymbol{d}) \rangle_Q, \qquad (41)$$

which is equivalent to the Kullback-Leibler divergence $D_{\mathrm{KL}}(Q, P)$, is chosen as such a measure (Enßlin & Weig 2010).

In favor of comprehensibility, we suppose the solution for the logarithmic power spectrum $\boldsymbol{\tau}^\star$ is known for the moment. The Gibbs free energy is then calculated by plugging in the Hamiltonian, and evaluating the expectation values[10],

$$G = G_0 + \langle H(\boldsymbol{s}, \boldsymbol{u}|\boldsymbol{d}) \rangle_{Q_s} - \frac{1}{2}\log\left(\det\left[\boldsymbol{D}\right]\right) \qquad (42)$$

$$= G_1 + \boldsymbol{1}^\dagger \boldsymbol{l} - \boldsymbol{d}^\dagger \left\{ \log(\boldsymbol{l}) - \sum_{\nu=2}^{\infty} \frac{(-1)^\nu}{\nu} \left\langle (\boldsymbol{\lambda}/\boldsymbol{l} - 1)^\nu \right\rangle_{Q_s} \right\}$$

$$+ \frac{1}{2}\boldsymbol{m}^{(s)\dagger}\boldsymbol{S}^{\star-1}\boldsymbol{m}^{(s)} + \frac{1}{2}\mathrm{tr}\left[\boldsymbol{D}^{(s)}\boldsymbol{S}^{\star-1}\right] \qquad (43)$$

$$+ (\boldsymbol{\beta} - 1)^\dagger \boldsymbol{m}^{(u)} + \boldsymbol{\eta}^\dagger \mathrm{e}^{-\boldsymbol{m}^{(u)}+\frac{1}{2}\widehat{\boldsymbol{D}}^{(u)}}$$

$$- \frac{1}{2}\log\left(\det\left[\boldsymbol{D}\right]\right),$$

with

$$\boldsymbol{\lambda} = \boldsymbol{R}\left(\mathrm{e}^{\boldsymbol{s}} + \mathrm{e}^{\boldsymbol{u}}\right), \qquad (44)$$

$$\boldsymbol{l} = \langle \boldsymbol{\lambda} \rangle_{Q_s} = \boldsymbol{R}\left(\mathrm{e}^{\boldsymbol{m}^{(s)}+\frac{1}{2}\widehat{\boldsymbol{D}}^{(s)}} + \mathrm{e}^{\boldsymbol{m}^{(u)}+\frac{1}{2}\widehat{\boldsymbol{D}}^{(u)}}\right), \qquad (45)$$

$$\boldsymbol{S}^\star = \sum_k \mathrm{e}^{\tau_k^\star}\boldsymbol{S}_k, \text{ and} \qquad (46)$$

$$\widehat{\boldsymbol{D}} = \mathrm{diag}\left[\boldsymbol{D}\right]. \qquad (47)$$

Here, G_0 and G_1 carry all terms independent of $\boldsymbol{s}$ and $\boldsymbol{u}$. In comparison to the Hamiltonian given in Eq. (31), there are a number of correction terms that now also consider the uncertainty covariances of the signal estimates properly. For example, the expectation values of the photon fluxes differ comparing $\boldsymbol{l}$ in Eq. (34) and (45) where it now describes the expectation value of $\boldsymbol{\lambda}$ over the approximate posterior. In case $\boldsymbol{l} = \boldsymbol{\lambda}$ the explicit sum in Eq. (43) vanishes. Since this sum includes powers of $\langle \boldsymbol{\lambda}^{\nu>2} \rangle_{Q_s}$ its evaluation would require all entries of $\boldsymbol{D}$ to be known explicitly. In order to keep the algorithm computationally feasible, this sum shall hereafter be neglected. This is equivalent to truncating the corresponding expansion at second order; i.e., $\nu = 2$. It can be shown that, in consequence of this approximation, the cross-correlation $\boldsymbol{D}^{(su)}$ equals zero, and $\boldsymbol{D}$ becomes block diagonal.

[10] The second likelihood term in Eq. (43), $\boldsymbol{d}^\dagger \log(\boldsymbol{\lambda})$, is thereby expanded according to

$$\log(x) = \log\langle x \rangle - \sum_{\nu=2}^{\infty} \frac{(-1)^\nu}{\nu} \left\langle \left(\frac{x}{\langle x \rangle} - 1\right)^\nu \right\rangle$$

$$\approx \log\langle x \rangle + \mathcal{O}\left(\langle x^2 \rangle\right),$$

under the assumption $x \approx \langle x \rangle$.

Without these second order terms, the Gibbs free energy reads

$$G = G_1 + \boldsymbol{1}^\dagger \boldsymbol{l} - \boldsymbol{d}^\dagger \log(\boldsymbol{l})$$

$$+ \frac{1}{2}\boldsymbol{m}^{(s)\dagger}\boldsymbol{S}^{\star-1}\boldsymbol{m}^{(s)} + \frac{1}{2}\mathrm{tr}\left[\boldsymbol{D}^{(s)}\boldsymbol{S}^{\star-1}\right] \qquad (48)$$

$$+ (\boldsymbol{\beta} - 1)^\dagger \boldsymbol{m}^{(u)} + \boldsymbol{\eta}^\dagger \mathrm{e}^{-\boldsymbol{m}^{(u)}+\frac{1}{2}\widehat{\boldsymbol{D}}^{(u)}}$$

$$- \frac{1}{2}\log\left(\det\left[\boldsymbol{D}^{(s)}\right]\right) - \frac{1}{2}\log\left(\det\left[\boldsymbol{D}^{(u)}\right]\right).$$

Minimizing the Gibbs free energy with respect to $\boldsymbol{m}^{(s)}$, $\boldsymbol{m}^{(u)}$, $\boldsymbol{D}^{(s)}$, and $\boldsymbol{D}^{(u)}$ would optimize the fitness of the posterior approximation $P \approx Q$. Filter formulas for the Gibbs solution can be derived by taking the derivative of G with respect to the approximate mean estimates,

$$\frac{\partial G}{\partial \boldsymbol{m}^{(s)}} = \boldsymbol{0} = (\boldsymbol{1} - \boldsymbol{d}/\boldsymbol{l})^\dagger \boldsymbol{R} * \mathrm{e}^{\boldsymbol{m}^{(s)}+\frac{1}{2}\widehat{\boldsymbol{D}}^{(s)}} + \boldsymbol{S}^{\star-1}\boldsymbol{m}^{(s)}, \qquad (49)$$

$$\frac{\partial G}{\partial \boldsymbol{m}^{(u)}} = \boldsymbol{0} = (\boldsymbol{1} - \boldsymbol{d}/\boldsymbol{l})^\dagger \boldsymbol{R} * \mathrm{e}^{\boldsymbol{m}^{(u)}+\frac{1}{2}\widehat{\boldsymbol{D}}^{(u)}} \qquad (50)$$

$$+ \boldsymbol{\beta} - \boldsymbol{1} - \boldsymbol{\eta} * \mathrm{e}^{-\boldsymbol{m}^{(u)}+\frac{1}{2}\widehat{\boldsymbol{D}}^{(u)}},$$

This filter formulas again account for the uncertainty of the mean estimates in comparison to the MAP filter formulas in Eq. (32) and (33). The uncertainty covariances can be constructed by either taking the second derivatives,

$$\frac{\partial^2 G}{\partial \boldsymbol{m}^{(s)}\partial \boldsymbol{m}^{(s)\dagger}} \approx \boldsymbol{D}^{(s)-1}, \qquad \frac{\partial^2 G}{\partial \boldsymbol{m}^{(u)}\partial \boldsymbol{m}^{(u)\dagger}} \approx \boldsymbol{D}^{(u)-1}, \quad (51)$$

or setting the first derivatives of G with respect to the uncertainty covariances equal to zero matrices,

$$\frac{\partial G}{\partial D_{xy}^{(s)}} = 0, \qquad\qquad \frac{\partial G}{\partial D_{xy}^{(u)}} = 0. \qquad (52)$$

The closed form of $\boldsymbol{D}^{(s)}$ and $\boldsymbol{D}^{(u)}$ is given explicitly in Appendix B.

So far, the logarithmic power spectrum $\boldsymbol{\tau}^\star$, and with it $\boldsymbol{S}^\star$, have been supposed to be known. The mean field approximation in Eq. (38) does not specify the approximate posterior $Q_\tau(\boldsymbol{\tau}|\boldsymbol{\mu}, \boldsymbol{d})$, but it can be retrieved by variational Bayesian methods (Jordan et al. 1999; Wingate & Weber 2013), according to the procedure detailed in Appendix C.2. The subsequent Appendix C.3 discusses the derivation of an solution for $\boldsymbol{\tau}$ by extremizing Q_τ. This result, which was also derived in Oppermann et al. (2012), applies to the inference problem discussed here, yielding

$$\mathrm{e}^{\boldsymbol{\tau}^\star} = \frac{\boldsymbol{q} + \frac{1}{2}\left(\mathrm{tr}\left[\left(\boldsymbol{m}^{(s)}\boldsymbol{m}^{(s)\dagger} + \boldsymbol{D}^{(s)}\right)\boldsymbol{S}_k^{-1}\right]\right)_k}{\boldsymbol{\gamma} + \boldsymbol{T}\boldsymbol{\tau}^\star}. \qquad (53)$$

Again, this solution includes a correction term in comparison to the MAP solution in Eq. (37). Since $\boldsymbol{D}^{(s)}$ is positive definite, it contributes positive to the (logarithmic) power spectrum, and therefore reduces the possible perception threshold further.

We note that this is a minimal Gibbs free energy solution that maximizes Q_τ. A proper calculation of $\langle \boldsymbol{\tau} \rangle_{Q_\tau}$ might include further correction terms, but their derivation

is not possible in closed form. Moreover, the above used diffuse signal covariance $\boldsymbol{S}^{\star-1}$ should be replaced by $\langle \boldsymbol{S}^{-1} \rangle_{Q_\tau}$ adding further correction terms to the filter formulas.

In order to keep the computational complexity on a feasible level, all these higher order corrections are not considered here. The detailed characterization of their implications and implementation difficulties is left for future investigation.

3.3. Physical flux solution

To perform calculations on the logarithmic fluxes is convenient for numerical reasons, but it is the physical fluxes that are actually of interest to us. Given the chosen approximation, we can compute the posterior expectation values of the diffuse and point-like photon flux, $\boldsymbol{\rho}^{(s)}$ and $\boldsymbol{\rho}^{(u)}$, straight forwardly,

$$\left\langle \boldsymbol{\rho}^{(\cdot)} \right\rangle_P \overset{\text{MAP-}\delta}{\approx} \left\langle \boldsymbol{\rho}^{(\cdot)} \right\rangle_\delta = \rho_0 e^{\boldsymbol{m}^{(\cdot)}_{\text{mode}}}, \tag{54}$$

$$\overset{\text{MAP-}\mathcal{G}}{\approx} \left\langle \boldsymbol{\rho}^{(\cdot)} \right\rangle_\mathcal{G} = \rho_0 e^{\boldsymbol{m}^{(\cdot)}_{\text{mode}} + \frac{1}{2}\widehat{\boldsymbol{D}}^{(\cdot)}_{\text{mode}}}, \tag{55}$$

$$\overset{\text{Gibbs}}{\approx} \left\langle \boldsymbol{\rho}^{(\cdot)} \right\rangle_Q = \rho_0 e^{\boldsymbol{m}^{(\cdot)}_{\text{mean}} + \frac{1}{2}\widehat{\boldsymbol{D}}^{(\cdot)}_{\text{mean}}}, \tag{56}$$

in accordance with Eq. (28), (29), or (38), respectively. Those solutions differ from each other in terms of the involvement of the posterior's mode or mean, and in terms of the inclusion of the uncertainty information, see subscripts.

In general, the mode approximation holds for symmetric, single peaked distributions, but can perform poorly in other cases (e.g., Enßlin & Frommert 2011). The exact form of the posterior considered here is highly complex because of the many degrees of freedom. In a dimensionally reduced frame, however, the posterior appears single peaked and exhibits a negative skewness.[11] Although this is not necessarily generalizable, it suggest a superiority of the posterior mean compared to the MAP because of the asymmetry of the distribution. Nevertheless, the MAP approach is computationally cheaper compared to the Gibbs approach that requires permanent knowledge of the uncertainty covariance.

The uncertainty of the reconstructed photon flux can be approximated as for an ordinary log-normal distribution,

$$\left\langle \boldsymbol{\rho}^{(\cdot)2} \right\rangle_P - \left\langle \boldsymbol{\rho}^{(\cdot)} \right\rangle_P^2 \overset{\text{MAP}}{\approx} \left\langle \boldsymbol{\rho}^{(\cdot)} \right\rangle_\mathcal{G}^2 \left(e^{\widehat{\boldsymbol{D}}^{(\cdot)}_{\text{mode}}} - 1 \right), \tag{57}$$

$$\overset{\text{Gibbs}}{\approx} \left\langle \boldsymbol{\rho}^{(\cdot)} \right\rangle_Q^2 \left(e^{\widehat{\boldsymbol{D}}^{(\cdot)}_{\text{mean}}} - 1 \right), \tag{58}$$

where the square root of the latter term would describe the relative uncertainty.

3.4. Imaging algorithm

The problem of denoising, deconvolving, and decomposing photon observations is a non-trivial task. Therefore, this section discusses the implementation of the D³PO algorithm given the two sets of filter formulas derived in Sec. 3.1 and 3.2, respectively.

---

[11] For example, the posterior $P(s|d)$ for a one-dimensional diffuse signal is proportional to $\exp(-\frac{1}{2}s^2 + ds - \exp(s))$, whereby all other parameters are fixed to unity. Analogously, $P(u|d) \propto \exp(du - 2\cosh(u))$.

The information Hamiltonian, or equivalently the Gibbs free energy, are scalar quantities defined over a huge phase space of possible field and parameter configurations including, among others, the elements of $\boldsymbol{m}^{(s)}$ and $\boldsymbol{m}^{(u)}$. If we only consider those, and no resolution refinement from data to signal space, two numbers need to be inferred from one data value each. Including $\boldsymbol{\tau}$ and the uncertainty covariances $\boldsymbol{D}^{(s)}$ and $\boldsymbol{D}^{(u)}$ in the inference, the problem of underdetermined degrees of freedom gets worse. This is reflected in the possibility of a decent number of local minima in the non-convex manifold landscape of the codomain of the Hamiltonian, or Gibbs free energy, respectively (Kirkpatrick et al. 1983; Geman & Geman 1984; Giovannelli & Coulais 2005). The complexity of the inference problem goes back to the, in general, non-linear entanglement between the individual parameters.

The D³PO algorithm is based on an iterative optimization scheme, where certain subsets of the problem are optimized alternately instead of the full problem at once. Each subset optimization is designed individually, see below. The global optimization cycle is in some degree sensitive to the starting values because of the non-convexity of the considered potential; i.e., the information Hamiltonian or Gibbs free energy, respectively. We can find such appropriate starting values by solving the inference problem in a reduced frame in advance, see below. So far, a step-by-step guide of the algorithm looks like the following.

1. Initialize the algorithm with primitive starting values; e.g., $m_x^{(s)} = m_x^{(u)} = 0$, $D_{xy}^{(s)} = D_{xy}^{(u)} = \delta_{xy}$, and $\tau_k^\star = \log(k^{-2})$. – Those values are arbitrary. Although the optimization is rather insensitive to them, inappropriate values can cripple the algorithm for numerical reasons because of the high non-linearity of the inference problem.

2. Optimize $\boldsymbol{m}^{(s)}$, the diffuse signal field, coarsely. – The preliminary optimization shall yield a rough estimate of the diffuse only contribution. This can be achieved by reconstructing a coarse screened diffuse signal field that only varies on large scales; i.e., limiting the bandwidth of the diffuse signal in its harmonic basis. Alternatively, obvious point sources in the data could be masked out by introducing an artificial mask into the response, if feasible.

3. Optimize $\boldsymbol{m}^{(u)}$, the point-like signal field, locally. – This initial optimization shall approximate the brightest, most obvious, point sources that are visible in the data image by eye. Their current disagreement with the data dominates the considered potential, and introduces some numerical stiffness. The gradient of the potential can be computed according to Eq. (33) or (50), and its minima will be at the expected position of the brightest point source which has not been reconstructed, yet. It is therefore very efficient to increase $\boldsymbol{m}^{(u)}$ at this location directly until the sign of the gradient flips, and repeat this procedure until the obvious point sources are fit.

4. Optimize $\boldsymbol{m}^{(u)}$, the point-like signal field. – This task can be done by a steepest descent minimization of the potential combined with a line search following the Wolfe conditions (Nocedal & Wright 2006). The potentials can be computed according to Eq. (31) or (43) neglecting terms independent of $\boldsymbol{m}^{(u)}$, and the gradient according to Eq. (33) or (50). A more sophisticated minimization scheme, such as a non-linear conjugate gradi-
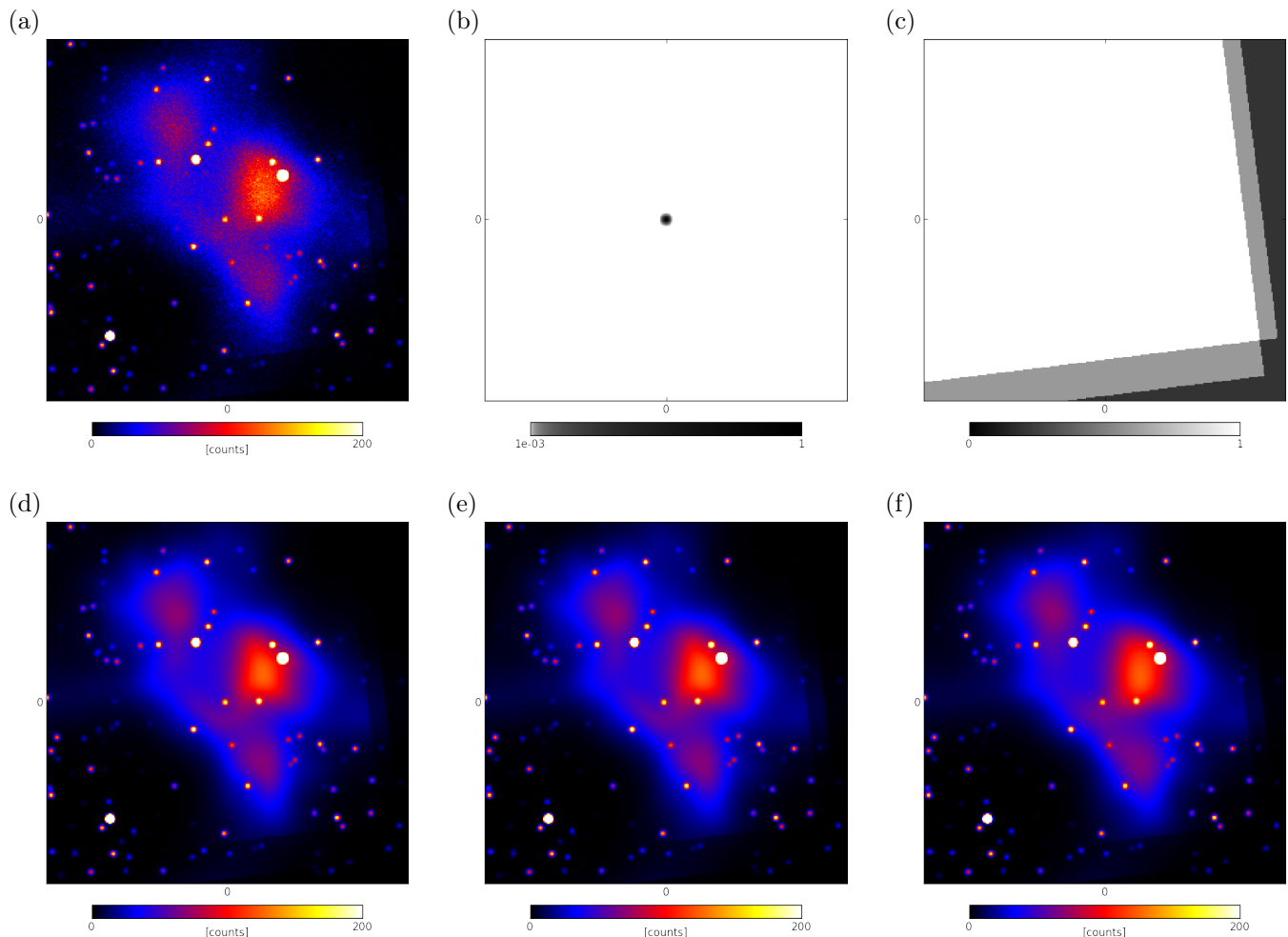
(a)

(b)

(c)

(d)

(e)

(f)

Fig. 3. Illustration of the data and noiseless, but reconvolved, signal responses of the reconstructions. Panel (a) shows the data from a mock observation of a $32 \times 32 \, \mathrm{arcmin}^2$ patch of the sky with a resolution of $0.1 \, \mathrm{arcmin}$ corresponding to a total of $102\,400$ pixels. The data had been convolved with a Gaussian-like PSF (FWHM $\approx 0.2 \, \mathrm{arcmin} = 2$ pixels, finite support of $1.1 \, \mathrm{arcmin}$ = 11 pixels) and masked because of an uneven exposure. Panel (b) shows the centered convolution kernel. Panel (c) shows the exposure mask. The bottom panels show the reconvolved signal response $\boldsymbol{R} \langle \boldsymbol{\rho} \rangle$ of a reconstruction using a different approach each, namely (d) MAP-δ, (e) MAP-$\mathcal{G}$, and (f) Gibbs. All reconstructions shown here and in the following figures used the same model parameters: $\alpha = 1$, $q = 10^{-12}$, $\sigma = 10$, $\beta = \frac{3}{2}$, and $\eta = 10^{-4}$.

ent (Shewchuk 1994), is conceivable but would require the application of the full Hessian, cf. step 5. In the first run, it might be sufficient to restrict the optimization to the locations identified in step 3.

5. Update $\widehat{\boldsymbol{D}}^{(u)}$, the point-like uncertainty variance, in case of a Gibbs approach. – It is not feasible to compute the full uncertainty covariance $\boldsymbol{D}^{(u)}$ explicitly in order to extract its diagonal. A more elegant way is to apply a probing technique relying on the application of $\boldsymbol{D}^{(u)}$ to random fields $\boldsymbol{\xi}$ that project out the diagonal (Hutchinson 1989; Selig et al. 2012). The uncertainty covariance is given as the inverse Hessian by Eq. (36) or (51), and should be symmetric and positive definite. For that reason, it can be applied to a field using a conjugate gradient (Shewchuk 1994); i.e., solving $(\boldsymbol{D}^{(u)})^{-1}\boldsymbol{y} = \boldsymbol{\xi}$ for $\boldsymbol{y}$. However, if the current phase space position is far away from the minimum, the Hessian is not necessarily positive definite. One way to overcome this temporal instability, would be to introduce a Levenberg damp-

ing in the Hessian (inspired by Transtrum et al. 2010; Transtrum & Sethna 2012).

6. Optimize $\boldsymbol{m}^{(s)}$, the diffuse signal field. – An analog scheme as in step 4 using steepest descent and Wolfe conditions is effective. The potentials can be computed according to Eq. (31) or (43) neglecting terms independent of $\boldsymbol{m}^{(s)}$, and the gradient according to Eq. (32) or (49), respectively. It has proven useful to first ensure a convergence on large scales; i.e., small harmonic modes k. This can be done repeating steps 6, 7, and 8 for all $k < k_{\max}$ with growing $k_{\max}$ using the corresponding projections $\boldsymbol{S}_k$.

7. Update $\widehat{\boldsymbol{D}}^{(s)}$, the diffuse uncertainty variance, in case of a Gibbs approach in analogy to step 5.

8. Optimize $\boldsymbol{\tau}^\star$, the logarithmic power spectrum. – This is done by solving Eq. (37) or (53). The trace term can be computed analog to the diagonal; e.g., by probing. Given this, the equation can be solved efficiently by a Newton-Raphson method.

9. Repeat the steps 4 to 8 until convergence. – This scheme will take several cycles until the algorithm reaches the
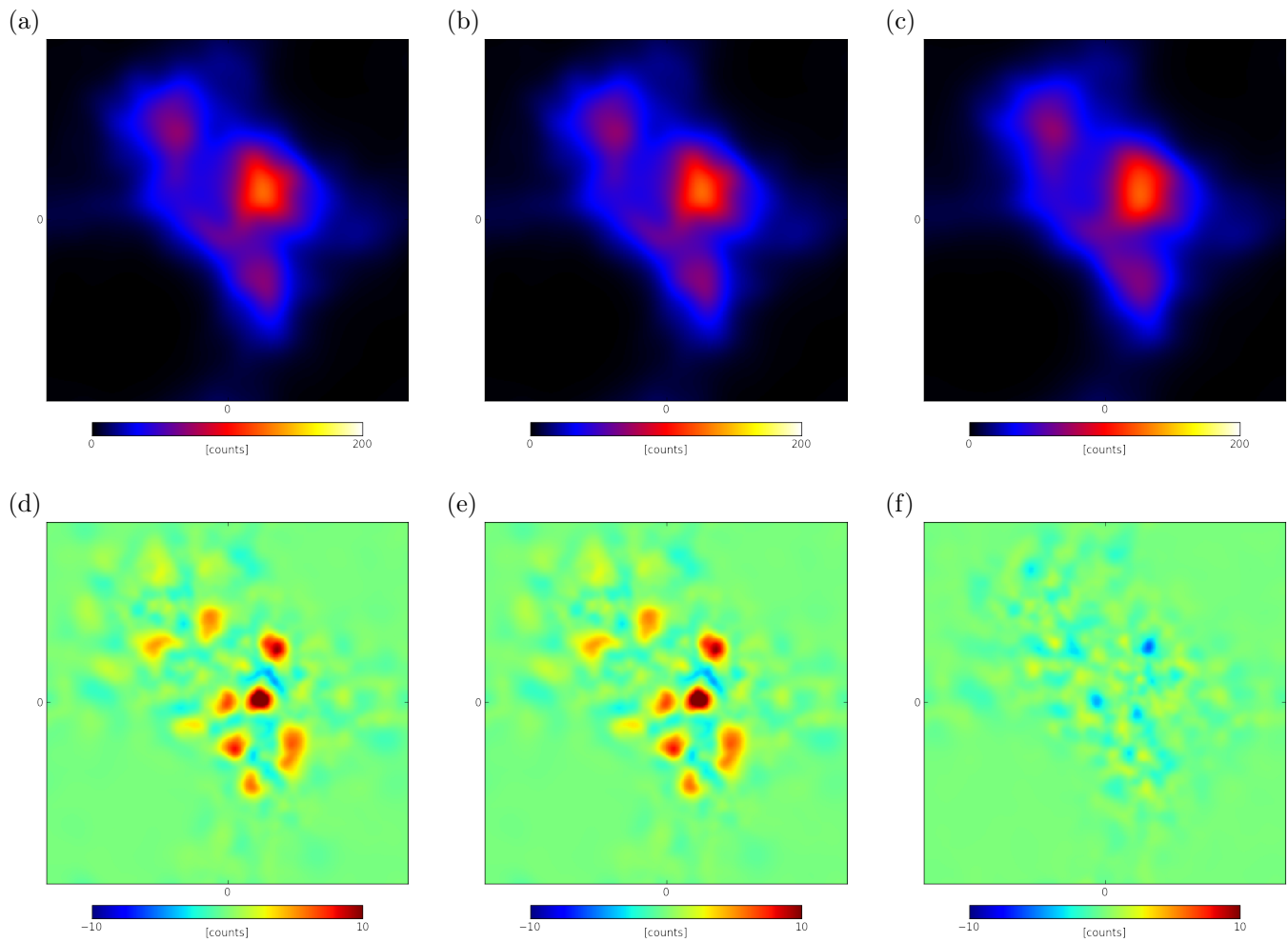
Fig. 4. Illustration of the diffuse reconstruction. The top panels show the denoised and deconvolved diffuse contribution $\langle\boldsymbol{\rho}^{(s)}\rangle/\rho_0$ reconstructed using a different approach each, namely (d) MAP-δ, (e) MAP-$\mathcal{G}$, and (f) Gibbs. The bottom panels (d) to (f) show the difference between the originally simulated signal and the respective reconstruction.

desired convergence level. Therefore, it is not required to achieve a convergence to the final accuracy level in all subsets in all cycles. It is advisable to start with weak convergence criteria in the first loop and increase them gradually.

A few remarks are in order.

The phase space of possible signal field configurations is tremendously huge. It is therefore impossible to judge if the algorithm has converged to the global or some local minima, but this does not matter if both yield reasonable results that do not differ substantially.

In general, the converged solution is also subject to the choice of starting values. Solving a non-convex, non-linear inference problem without proper initialization can easily lead to nonsensical results, such as fitting (all) diffuse features by point sources. Therefore, the D³PO algorithm essentially creates its own starting values executing the initial steps 1 to 3. The primitive starting values are thereby processed to rough estimates that cover coarsely resolved diffuse and prominent point-like features. These estimates serve then as actual starting values for the optimization cycle.

Because of the iterative optimization scheme starting with the diffuse component in step 2, the algorithm might be prone to explaining some point-like features by diffuse sources. Starting with the point-like component instead would give rise to the opposite bias. To avoid such biases, it is advisable to restart the algorithm partially. To be more precise, we propose to discard the current reconstruction of $\boldsymbol{m}^{(u)}$ after finishing step 8 for the first time, then start the second iteration again with step 3, and to discard the current $\boldsymbol{m}^{(s)}$ before step 6.

The above scheme exploits a few numerical techniques, such as probing or Levenberg damping, that are described in great detail in the given references. The code of our implementation of the D³PO algorithm will be made public in the future under `http://www.mpa-garching.mpg.de/ift/d3po/`.

4. Numerical application

Exceeding the simple 1D scenario illustrated in Fig. 1, the D³PO algorithm is now applied to a realistic, but simulated, data set. The data set represents a high energy observation with a field of view of $32 \times 32\,\mathrm{arcmin}^2$ and a resolution of 0.1 arcmin; i.e., the photon count image comprises 102 400 pixels. The instrument response includes the convolution with a Gaussian-like PSF with a FWHM of roughly
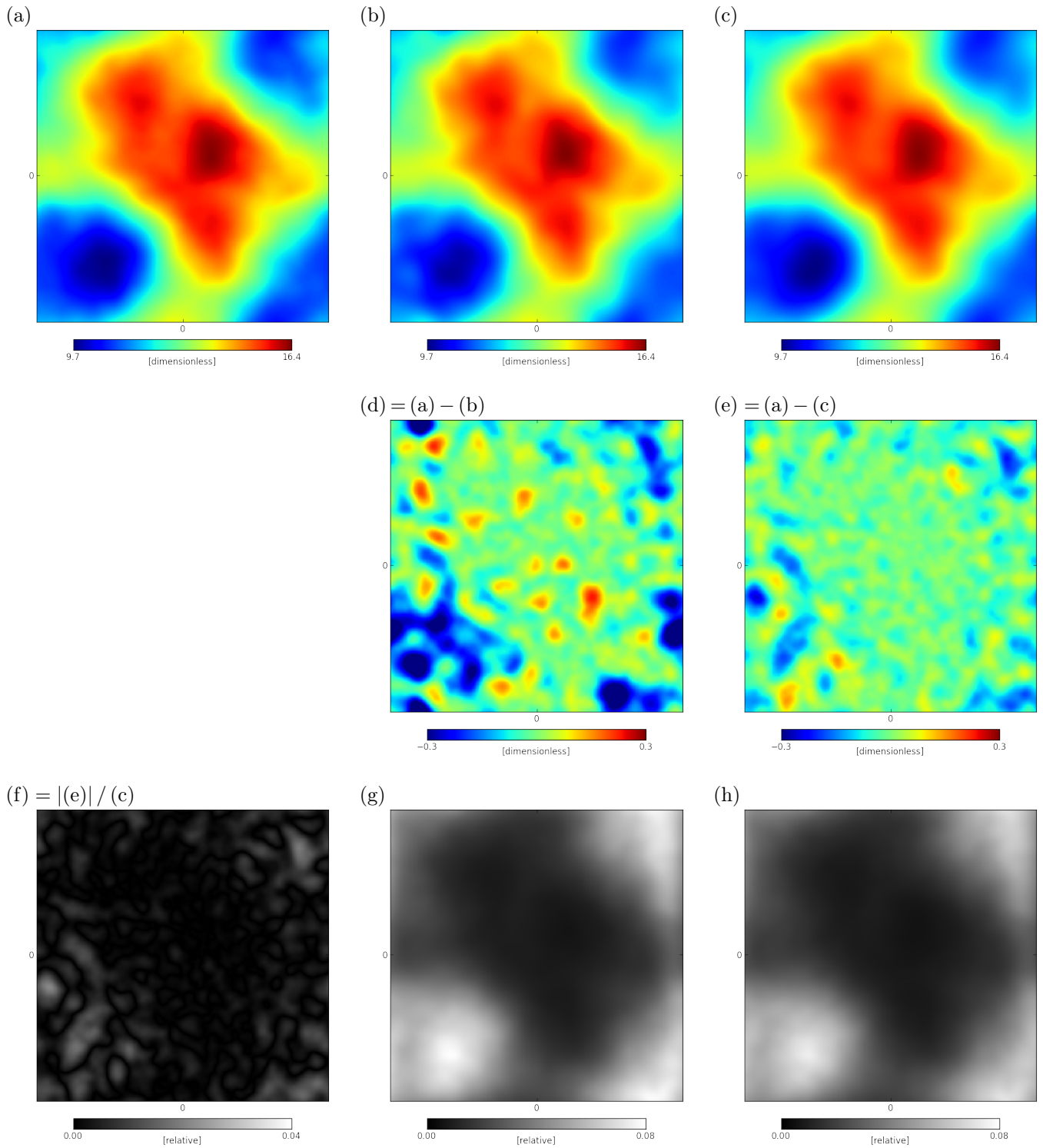
(a)

(b)

(c)

(d) = (a) − (b)

(e) = (a) − (c)

(f) = |(e)| / (c)

(g)

(h)

Fig. 5. Illustration of the reconstruction of the diffuse signal field $s = \log \rho^{(s)}$ and its uncertainty. The top panels show diffuse signal fields. Panel (a) shows the original simulation s, panel (b) the reconstruction $m_{\mathrm{mode}}^{(s)}$ using a MAP approach, and panel (c) the reconstruction $m_{\mathrm{mean}}^{(s)}$ using a Gibbs approach. The panels (d) and (e) show the differences between original and reconstruction. Panel (f) shows the relative difference. The panels (g) and (h) show the relative uncertainty of the above reconstructions.

0.2 arcmin, and an uneven survey mask attributable to the inhomogeneous exposure of the virtual instrument. The data image and those characteristics are shown in Fig. 3.

In addition, the top panels of Fig. 3 show the reproduced signal responses of the reconstructed (total) photon flux.

The reconstructions used the same model parameters, $\alpha = 1$, $q = 10^{-12}$, $\sigma = 10$, $\beta = \frac{3}{2}$, and $\eta = 10^{-4}$ in a MAP-δ, MAP-$\mathcal{G}$ and a Gibbs approach, respectively. They all show a very good agreement with the actual data, and differences are barely visible by eye. We note that only the quality
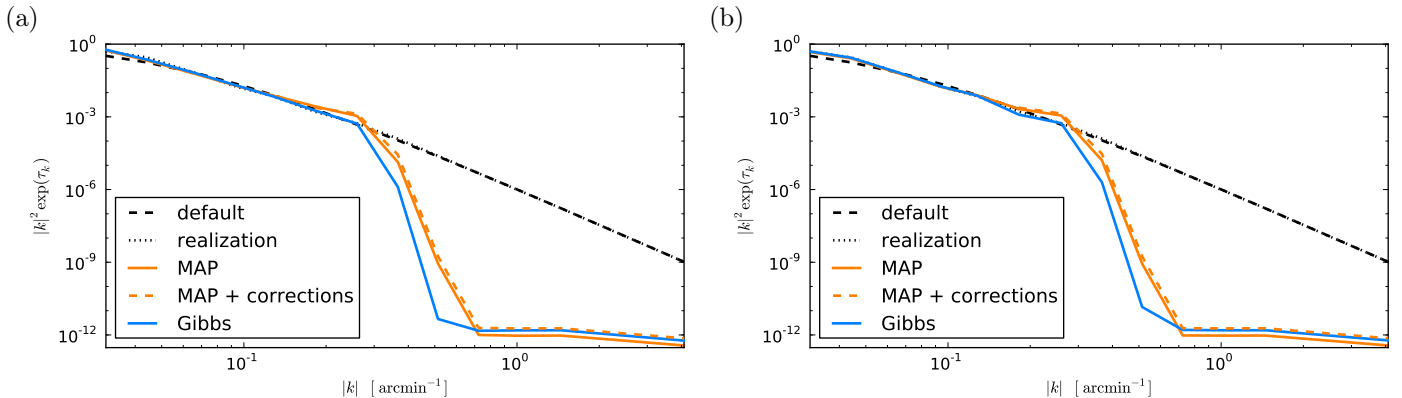
(a)

(b)

Fig. 6. Illustration of the reconstruction of the logarithmic power spectrum $\boldsymbol{\tau}$. Both panels show the default power spectrum (black dashed line), and the simulated realization (black dotted line), as well as the reconstructed power spectra using a MAP (orange solid line), plus second order corrections (orange dashed line), and a Gibbs approach (blue solid line). Panel (a) shows the reconstruction for a chosen σ parameter of 10, panel (b) for a σ of 1000.

of denoising is visible, since the signal response shows the convolved and superimposed signal fields.

The diffuse contribution to the deconvolved photon flux is shown Fig. 4 for all three estimators, cf. Eqs. (54) to (56). There, all point-like contributions as well as noise and instrumental effects have been removed presenting a denoised, deconvolved and decomposed reconstruction result for the diffuse photon flux. Figure 4 also shows the absolute difference to the original flux. Although the differences in the MAP estimators are insignificant, the Gibbs solution seems to be slightly better.

In order to have a quantitative statement about the goodness of the reconstruction, we define a relative residual error $\epsilon^{(s)}$ for the diffuse contribution as follows,

$$\epsilon^{(s)} = \left| \boldsymbol{\rho}^{(s)} - \left\langle \boldsymbol{\rho}^{(s)} \right\rangle \right|_2 \left| \boldsymbol{\rho}^{(s)} \right|_2^{-1}, \tag{59}$$

where $|\cdot|_2$ is the Euclidean L^2-norm. For the point-like contribution, however, we have to consider an error in brightness and position. For this purpose we define,

$$\epsilon^{(u)} = \int_1^N \mathrm{d}n \, \left| \boldsymbol{R}_{\mathrm{PSF}}^n \boldsymbol{\rho}^{(u)} - \boldsymbol{R}_{\mathrm{PSF}}^n \left\langle \boldsymbol{\rho}^{(u)} \right\rangle \right|_2 \left| \boldsymbol{R}_{\mathrm{PSF}}^n \boldsymbol{\rho}^{(u)} \right|_2^{-1}, \tag{60}$$

where $\boldsymbol{R}_{\mathrm{PSF}}$ is a (normalized) convolution operator, such that $\boldsymbol{R}_{\mathrm{PSF}}^N$ becomes the identity for large N. These errors are listed in Table 1. When comparing the MAP-δ and MAP-$\mathcal{G}$ approach, the incorporation of uncertainty corrections seems to improve the results slightly. The full regularization treatment within the Gibbs approach outperforms MAP solutions in terms of the chosen error measure $\epsilon^{(\cdot)}$. For a discussion of how such measures can change the view on certain Bayesian estimators, we refer to the work by Burger & Lucka (2014).

Table 1. Overview of the relative residual errors in the photon flux reconstructions for the respective approaches, all using the same model parameters, cf. text.

MAP-δ	MAP-$\mathcal{G}$	Gibbs
$\epsilon^{(s)} = 4.442\%$	$\epsilon^{(s)} = 4.441\%$	$\epsilon^{(s)} = 2.078\%$
$\epsilon^{(u)} = 1.540\%$	$\epsilon^{(u)} = 1.540\%$	$\epsilon^{(u)} = 1.089\%$

Figure 5 illustrates the reconstruction of the diffuse signal field, now in terms of logarithmic flux. The original and the reconstructions agree well, and the strongest deviations are found in the areas with low amplitudes. With regard to the exponential ansatz in Eq. (4), it is not surprising that the inference on the signal fields is more sensitive to higher values than to lower ones. For example, a small change in the diffuse signal field, $\boldsymbol{s} \rightarrow (1 \pm \epsilon)\boldsymbol{s}$, translates into a factor in the photon flux, $\boldsymbol{\rho}^{(s)} \rightarrow \boldsymbol{\rho}^{(s)}\mathrm{e}^{\pm \epsilon \boldsymbol{s}}$, that scales exponentially with the amplitude of the diffuse signal field. The Gibbs solution shows less deviation from the original signal than the MAP solution. Since the latter lacks the regularization by the uncertainty covariance it exhibits a stronger tendency to overfitting compared to the former. This includes overestimates in noisy regions with low flux intensities, as well as underestimates at locations where point-like contributions dominate the total flux.

The reconstruction of the power spectrum, as shown in Fig. 6, gives further indications of the reconstruction quality of the diffuse component. The simulation used a default power spectrum of

$$\exp(\tau_k) = 42 \, (k+1)^{-7}. \tag{61}$$

This power spectrum was on purpose chosen to deviate from a strict power law supposed by the smoothness prior.

From Fig. 6 it is apparent that the reconstructed power spectra track the original well up to a harmonic mode k of roughly $0.4 \, \mathrm{arcmin}^{-1}$. Beyond that point, the reconstructed power spectra fall steeply until they hit a lower boundary set by the model parameter $q = 10^{-12}$. This drop-off point at $0.4 \, \mathrm{arcmin}^{-1}$ corresponds to a physical wavelength of roughly $2.5 \, \mathrm{arcmin}$, and thus (half-phase) fluctuations on a spatial distances below $1.25 \, \mathrm{arcmin}$. The Gaussian-like PSF of the virtual observatory has a finite support of $1.1 \, \mathrm{arcmin}$. The lack of reconstructed power indicates that the algorithm assigns features on spatial scales smaller than the PSF support preferably to the point-like component. This behavior is reasonable because solely the point-like signal can cause PSF-like shaped imprints in the data image. However, there is no strict threshold in the distinction between the components on the mere basis of their spatial extend. We rather observe a continuous transition from assigning flux to the diffuse component to assigning it to the point-like component while reaching smaller spatial scales because
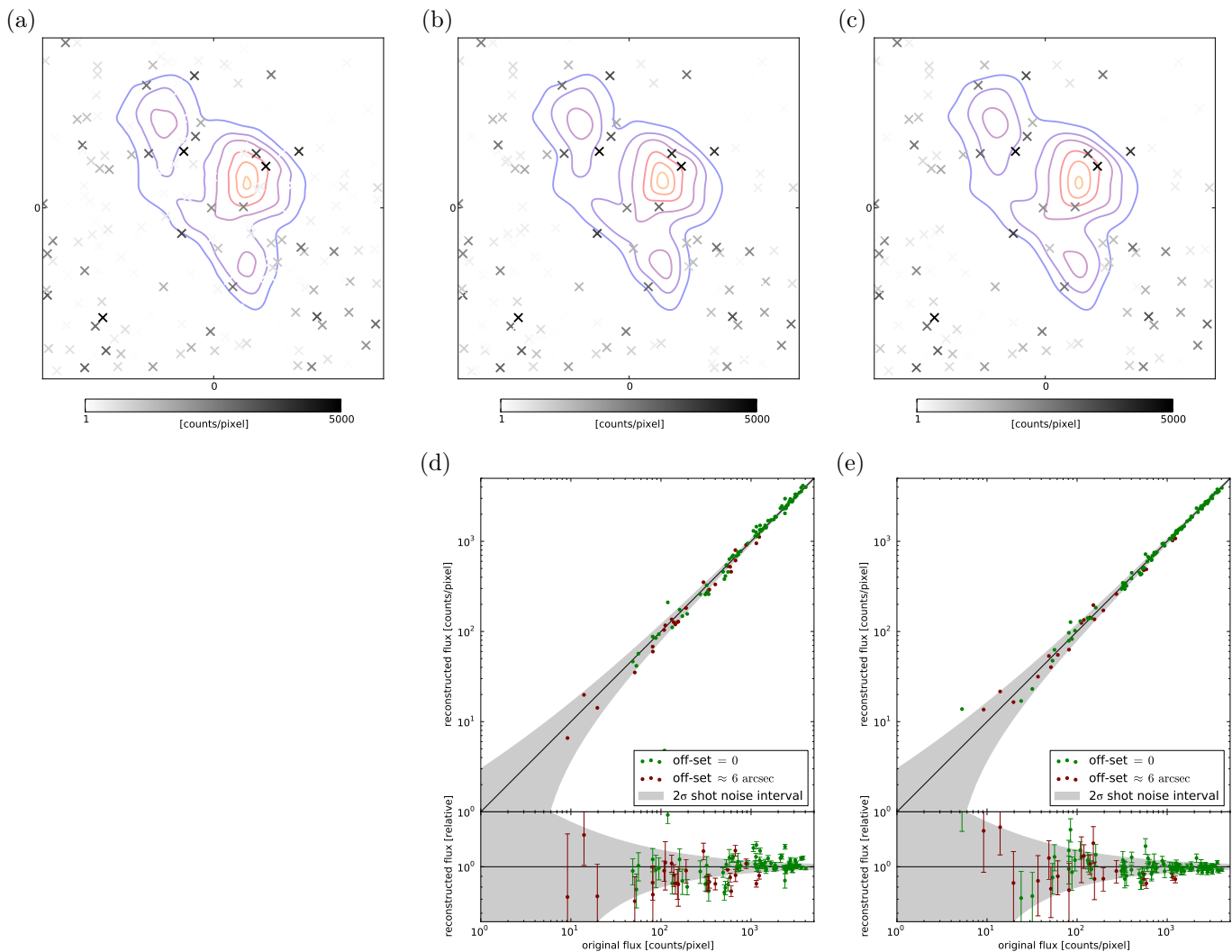
Fig. 7. Illustration of the reconstruction of the point-like signal field $\boldsymbol{u} = \log \boldsymbol{\rho}^{(u)}$ and its uncertainty. The top panels show the location (markers) and intensity (gray scale) of the point-like photon fluxes, underlaid is the respective diffuse contribution (contours) to guide the eye, cf. Fig 4. Panel (a) shows the original simulation, panel (b) the reconstruction using a MAP approach, and panel (c) the reconstruction using a Gibbs approach. The bottom panels (d) and (e) show the match between original and reconstruction in absolute and relative fluxes, the 2σ shot noise interval (gray contour), as well as some reconstruction uncertainty estimate (error bars).

strict boundaries are blurred out under the consideration of noise effects.

The differences between the reconstruction using a MAP and a Gibbs approach are subtle. The difference in the reconstruction formulas given by Eqs. (37) and (53) is an additive trace term involving $\boldsymbol{D}^{(s)}$, which is positive definite. Therefore, a reconstructed power spectrum regularized by uncertainty corrections is never below the one with out given the same $\boldsymbol{m}^{(s)}$. However, the reconstruction of the signal field follows different filter formulas, respectively. Since the Gibbs approach considers the uncertainty covariance $\boldsymbol{D}^{(s)}$ properly in each cycle, it can present a more conservative solution. The drop-off point is apparently at higher k for the MAP approach, leading to higher power on scales between roughly 0.3 and 0.7 arcmin⁻¹. In turn, the MAP solution tends to overfit by absorbing some noise power into $\boldsymbol{m}^{(s)}$ as discussed in Sec. 3. Thus, the higher MAP power spectrum in Fig. 6 seems to be caused by a higher level of noise remnants in the signal estimate.

The influence of the choice of the model parameter σ is also shown in Fig. 6. Neither a smoothness prior with $\sigma = 10$, nor a weak one with $\sigma = 1000$ influences the reconstruction of the power spectrum substantially in this case.[12] The latter choice, however, exhibits some more fluctuations in order to better track the concrete realization.

The results for the reconstruction of the point-like component are illustrated in Fig. 7. Overall, the reconstructed point-like signal field and the corresponding photon flux are in good agreement with the original ones. The point-sources have been located with an accuracy of ±0.1 arcmin, which is less than the FWHM of the PSF. The localization tends to be more precise for higher flux values because of the higher signal-to-noise ratio. The reconstructed intensities match the simulated ones well, although the MAP solution shows a spread that exceeds the expected shot noise uncertainty interval. This is again an indication of the overfitting known

---

[12] For a discussion of further log-normal reconstruction scenarios please refer to the work by Oppermann et al. (2012).

for MAP solutions. Moreover, neither reconstruction shows a bias towards higher or lower fluxes.

The uncertainty estimates for the point-like photon flux $\rho^{(u)}$ obtained from $\boldsymbol{D}^{(u)}$ according to Eqs. (57) and (58) are, in general, consistent with the deviations from the original and the shot noise uncertainty, cf. Fig. 7. They show a reasonable scaling being higher for lower fluxes and vice versa. However, some uncertainties seem to be underestimated. There are different reasons for this.

On the one hand, the Hessian approximation for $\boldsymbol{D}^{(u)}$ in Eq. (36) or (51) is in individual cases in so far poor as that the curvature of the considered potential does not describe the uncertainty of the point-like component adequately. The data admittedly constrains the flux intensity of a point source sufficiently, especially if it is a bright one. However, the rather narrow dip in the manifold landscape of the considered potential can be asymmetric, and thus not always well described by the quadratic approximation of Eq. (36) or (51), respectively.

On the other hand, the approximation leading to vanishing cross-correlation $\boldsymbol{D}^{(su)}$, takes away the possibility of communicating uncertainties between diffuse and point-like components. However, omitting the used simplification or incorporating higher order corrections would render the algorithm too computationally expensive. The fact that the Gibbs solution, which takes $\boldsymbol{D}^{(u)}$ into account, shows improvements backs up this argument.

The reconstructions shown in Fig. 5 and 7 used the model parameters $\sigma = 10$, $\beta = \frac{3}{2}$, and $\eta = 10^{-4}$. In order to reflect the influence of the choice of σ, β, and η, Table 2 summarizes the results from several reconstructions carried out with varying model parameters. Accordingly, the best parameters seem to be $\sigma = 10$, $\beta = \frac{5}{4}$, and $\eta = 10^{-4}$, although we caution that the total error is difficult to determine as the residual errors, $\epsilon^{(s)}$ and $\epsilon^{(u)}$, are defined differently. Although the errors vary significantly, 2–15% for $\epsilon^{(s)}$, we like to stress that the model parameters were changed drastically, partly even by orders of magnitude. The impact of the prior clearly exists, but is moderate. We note that the case of $\sigma \to \infty$ corresponds to neglecting the smoothness prior completely. The $\beta = 1$ case that corresponds to a logarithmically flat prior on $\boldsymbol{u}$ showed a tendency to fit more noise features by point-like contributions.

In summary, the D³PO algorithm is capable of denoising, deconvolving and decomposing photon observations by reconstructing the diffuse and point-like signal field, and the logarithmic power spectrum of the former. The reconstruction using MAP and Gibbs approaches perform flawlessly, except for a little underestimation of the uncertainty of the point-like component. The MAP approach shows signs of overfitting, but those are not overwhelming. Considering the simplicity of the MAP approach that goes along with a numerically faster performance, this shortcoming seems acceptable.

Because of the iterative scheme of the algorithm, a combination of the MAP approach for the signal fields and a Gibbs approach for the power spectrum is possible.

5. Conclusions & summary

The D³PO algorithm for the denoising, deconvolving and decomposing photon observations has been derived. It allows for the simultaneous but individual reconstruction of the diffuse and point-like photon fluxes, as well as the harmonic power spectrum of the diffuse component, from a single data image that is exposed to Poissonian shot noise and effects of the instrument response functions. Moreover, the D³PO algorithm can provide *a posteriori* uncertainty information on the reconstructed signal fields. With these capabilities, D³PO surpasses previous approaches that address only subsets of these complications.

The theoretical foundation is a hierarchical Bayesian parameter model embedded in the framework of IFT. The model comprises *a priori* assumptions for the signal fields that account for the different statistics and correlations of the morphologically different components. The diffuse photon flux is assumed to obey multivariate log-normal statistics, where the covariance is described by a power spectrum. The power spectrum is *a priori* unknown and reconstructed from the data along with the signal. Therefore, hyperpriors on the (logarithmic) power spectra have been introduced, including a spectral smoothness prior (Enßlin & Frommert 2011; Oppermann et al. 2012). The point-like photon flux, in contrast, is assumed to factorize spatially in independent inverse-Gamma distributions implying a (regularized) power-law behavior of the amplitudes of the flux.

An adequate description of the noise properties in terms of a likelihood, here a Poisson distribution, and the incorporation of all instrumental effects into the response operator renders the denoising and deconvolution task possible. The strength of the proposed approach is the performance of the additional decomposition task, which especially exploits the *a priori* description of diffuse and point-like. The model comes down to five scalar parameters, for which all *a priori* defaults can be motivated, and of which none is driving the inference predominantly.

We discussed maximum *a posteriori* (MAP) and Gibbs free energy approaches to solve the inference problem. The derived solutions provide optimal estimators that, in the considered examples, yielded equivalently excellent results. The Gibbs solution slightly outperforms MAP solutions (in terms of the considered L²-residuals) thanks to the full regularization treatment, however, for the price of a computationally more expensive optimization. Which approach is to be preferred in general might depend on the concrete problem at hand and the trade-off between reconstruction precision against computational effort.

The performance of the D³PO algorithm has been demonstrated in realistic simulations carried out in 1D and 2D. The implementation relies on the NIFTY package (Selig et al. 2013), which allows for the application regardless of the underlying position space.

In the 2D application example, a high energy observation of a $32 \times 32 \, \text{arcmin}^2$ patch of a simulated sky with a $0.1 \, \text{arcmin}$ resolution has been analyzed. The D³PO algorithm successfully denoised, deconvolved and decomposed the data image. The analysis yielded a detailed reconstruction of the diffuse photon flux and its logarithmic power spectrum, the precise localization of the point sources and accurate determination of their flux intensities, as well as *a posteriori* estimates of the reconstructed fields.

The D³PO algorithm should be applicable to a wide range of inference problems appearing in astronomical imaging and related fields. Concrete applications in high energy astrophysics, for example, the analysis of data from the Chandra X-ray observatory or the Fermi γ-ray space

telescope, are currently considered by the authors. In this regard, the public release of the D³PO code is planned.

Acknowledgments

References

Bertin, E. & Arnouts, S., SExtractor: Software for source extraction., A&AS **117** (Jun. 1996) 393–404, ADS

Bobin, J., Starck, J.-L., Fadili, J. M., Moudden, Y., & Donoho, D., Morphological Component Analysis: An Adaptive Thresholding Strategy, Image Processing, IEEE Transactions on **16** no. 11, (2007) 2675–2681

Bouchet, L., Amestoy, P., Buttari, A., Rouet, F.-H., & Chauvin, M., Simultaneous analysis of large INTEGRAL/SPI¹ datasets: Optimizing the computation of the solution and its variance using sparse matrix algorithms, Astronomy and Computing **1** (Feb. 2013) 59–69, `arXiv:1305.5683` `[astro-ph.IM]`

Burger, M. & Lucka, F., Maximum-A-Posteriori Estimates in Linear Inverse Problems with Log-concave Priors are Proper Bayes Estimators, ArXiv e-prints (Feb. 2014), `arXiv:1402.5297` `[math.ST]`

Carvalho, P., Rocha, G., & Hobson, M. P., A fast Bayesian approach to discrete object detection in astronomical data sets - PowellSnakes I, MNRAS **393** (Mar. 2009) 681–702, `arXiv:0802.3916`

Carvalho, P., Rocha, G., Hobson, M. P., & Lasenby, A., PowellSnakes II: a fast Bayesian approach to discrete object detection in multi-frequency astronomical data sets, MNRAS **427** (Dec. 2012) 1384–1400, `arXiv:1112.4886` `[astro-ph.CO]`

Caticha, A., Lectures on Probability, Entropy, and Statistical Physics, ArXiv e-prints physics.data-an/0808.0012 (Jul. 2008), `arXiv:0808.0012` `[physics.data-an]`

Caticha, A., Entropic Inference, in American Institute of Physics Conference Series, Vol. 1305, American Institute of Physics Conference Series, ed. A. Mohammad-Djafari, J.-F. Bercher, & P. Bessiére. 2011, 20–29, `arXiv:1011.0723` `[physics.data-an]`

Chapman, E., Abdalla, F. B., Bobin, J., et al., The scale of the problem: recovering images of reionization with Generalized Morphological Component Analysis, MNRAS **429** (Feb. 2013) 165–176, `arXiv:1209.4769` `[astro-ph.CO]`

Cornwell, T. J., Multiscale CLEAN Deconvolution of Radio Synthesis Images, IEEE Journal of Selected Topics in Signal Processing **2** (Nov. 2008) 793–801

Dupé, F.-X., Fadili, J., & Starck, J.-L., Deconvolution under Poisson noise using exact data fidelity and synthesis or analysis sparsity priors, ArXiv e-prints (Mar. 2011), `arXiv:1103.2213` `[stat.AP]`

Dupe, F.-X., Fadili, J. M., & Starck, J.-L., A Proximal Iteration for Deconvolving Poisson Noisy Images Using Sparse Representations, IEEE Transactions on Image Processing **18** (Feb. 2009) 310–321, `arXiv:0803.2623` `[stat.AP]`

Enßlin, T., Information field theory, in American Institute of Physics Conference Series, Vol. 1553, American Institute of Physics Conference Series, ed. U. von Toussaint. 2013, 184–191, `arXiv:1301.2556` `[astro-ph.IM]`

Enßlin, T., Astrophysical data analysis with information field theory, in American Institute of Physics Conference Series, American Institute of Physics Conference Series. 2014, `arXiv:1405.7701` `[astro-ph.IM]`

Enßlin, T. A. & Frommert, M., Reconstruction of signals with unknown spectra in information field theory with parameter uncertainty, Phys. Rev. D **83** no. 10, (May 2011) 105014–+, `arXiv:1002.2928` `[astro-ph.IM]`

Enßlin, T. A., Frommert, M., & Kitaura, F. S., Information field theory for cosmological perturbation reconstruction and nonlinear signal analysis, Phys. Rev. D **80** no. 10, (Nov. 2009) 105005–+, `arXiv:0806.3474`

Enßlin, T. A. & Weig, C., Inference with minimal Gibbs free energy in information field theory, Phys. Rev. E **82** no. 5, (Nov. 2010) 051112–+, `arXiv:1004.2868` `[astro-ph.IM]`

Enßlin, T. A. & Weig, C., Reply to "Comment on 'Inference with minimal Gibbs free energy in information field theory' ", Phys. Rev. E **85** no. 3, (Mar. 2012) 033102

Figueiredo, M. A. T. & Bioucas-Dias, J. M., Restoration of Poissonian Images Using Alternating Direction Optimization, IEEE Transactions on Image Processing **19** (Dec. 2010) 3133–3145, `arXiv:1001.2244` `[math.OC]`

Fomalont, E. B., A two-component log N-log S relationship, Bull. Astron. Inst. Netherlands **20** (Aug. 1968) 69, ADS

Geman, S. & Geman, D., Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images., IEEE Trans. Pattern Anal. Mach. Intell. **6** no. 6, (1984) 721–741, `http://dblp.uni-trier.de/db/journals/pami/pami6.html#GemanG84`

Giovannelli, J.-F. & Coulais, A., Positive deconvolution for superimposed extended source and point sources, A&A **439** (Aug. 2005) 401–412, `astro-ph/0507691`

Giron, Francisco Javier, C. C. d., A note on the convolution of inverted-gamma distributions with applications to the Behrens-Fisher distribution., RACSAM **95** no. 1, (2001) 39–44, `http://eudml.org/doc/40894`

González-Nuevo, J., Argüeso, F.and Lopez-Caniego, M., Toffolatti, L., et al., The Mexican hat wavelet family: application to point-source detection in cosmic microwave background maps, Notices of the Royal Astronomical Society (2006) 1603–1610

Guglielmetti, F., Fischer, R., & Dose, V., Background-source separation in astronomical images with Bayesian probability theory - I. The method, MNRAS **396** (Jun. 2009) 165–190, `arXiv:0903.2342` `[astro-ph.IM]`

Haar, A., Zur Theorie der orthogonalen Funktionensysteme (Erste Mitteilung), Mathematische Annalen **69** (1910) 331–371

Haar, A., Zur Theorie der orthogonalen Funktionensysteme (Zweite Mitteilung), Mathematische Annalen **71** (1911) 38–53

Hensley, B. S., Pavlidou, V., & Siegal-Gaskins, J. M., Novel techniques for decomposing diffuse backgrounds, MNRAS **433** (Jul. 2013) 591–602, `arXiv:1210.7239` `[astro-ph.CO]`

Högbom, J. A., Aperture Synthesis with a Non-Regular Distribution of Interferometer Baselines, A&AS **15** (Jun. 1974) 417, ADS

Hutchinson, M. F., A stochastic estimator of the trace of the influence matrix for laplacian smoothing splines, Communications in Statistics - Simulation and Computation **18** (1989) 1059–1076

Iatsenko, D., Stefanovska, A., & McClintock, P. V. E., Comment on "Inference with minimal Gibbs free energy in information field theory", Phys. Rev. E **85** no. 3, (Mar. 2012) 033101

Jasche, J. & Kitaura, F. S., Fast Hamiltonian sampling for large-scale structure inference, MNRAS **407** (Sep. 2010) 29–42, `arXiv:0911.2496` `[astro-ph.CO]`

Jasche, J., Kitaura, F. S., Wandelt, B. D., & Enßlin, T. A., Bayesian power-spectrum inference for large-scale structure data, MNRAS **406** (Jul. 2010) 60–85, `arXiv:0911.2493` `[astro-ph.CO]`

Jasche, J. & Wandelt, B. D., Methods for Bayesian Power Spectrum Inference with Galaxy Surveys, ApJ **779** (Dec. 2013) 15, `arXiv:1306.1821` `[astro-ph.CO]`

Jaynes, E. T., Information Theory and Statistical Mechanics, I and II, Physical Reviews **106 and 108** (1957) 620–630 and 171–190

Jordan, M. I., Ghahramani, Z., Jaakkola, T. S., & Saul, L. K., An Introduction to Variational Methods for Graphical Models, Machine Learning **37** no. 2, (Nov. 1999) 183–233

Junklewitz, H., Bell, M. R., Selig, M., & Enßlin, T. A., RESOLVE: A new algorithm for aperture synthesis imaging of extended emission in radio astronomy, ArXiv e-prints (Nov. 2013), `arXiv:1311.5282` `[astro-ph.IM]`

Kinney, J. B., Estimation of probability densities using scale-free field theories, ArXiv e-prints (Dec. 2013), `arXiv:1312.6661` `[physics.data-an]`

Kirkpatrick, S., Gelatt, C. D., & Vecchi, M. P., Optimization by simulated annealing, Science **220** (1983) 671–680

Kullback, S. & Leibler, R. A., On Information and Sufficiency, The Annals of Mathematical Statistics **22** no. 1, (Mar. 1951) 79–86

Malyshev, D. & Hogg, D. W., Statistics of Gamma-Ray Point Sources below the Fermi Detection Limit, ApJ **738** (Sep. 2011) 181, `arXiv:1104.0010` `[astro-ph.CO]`

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., & Teller, E., Equation of State Calculations by Fast Computing Machines, J. Chem. Phys. **21** (Jun. 1953) 1087–1092

Metropolis, N. & Ulam, S., The Monte Carlo method, J. Am. Stat. Assoc. **44** (1949) 335

Nocedal, J. & Wright, S. J. 2006, Numerical optimization, `http://site.ebrary.com/id/10228772`

Oppermann, N., Selig, M., Bell, M. R., & Enßlin, T. A., Reconstruction of Gaussian and log-normal fields with spectral smoothness, `arXiv:1210.6866 [astro-ph.IM]`

Planck Collaboration, Ade, P. A. R., Aghanim, N., et al., Planck early results. VII. The Early Release Compact Source Catalogue, A&A **536** (Dec. 2011) A7, `arXiv:1101.2041 [astro-ph.CO]`

Rau, U. & Cornwell, T. J., A multi-scale multi-frequency deconvolution algorithm for synthesis imaging in radio interferometry, A&A **532** (Aug. 2011) A71, `arXiv:1106.2745 [astro-ph.IM]`

Schmitt, J., Starck, J. L., Casandjian, J. M., Fadili, J., & Grenier, I., Poisson denoising on the sphere: application to the Fermi gamma ray space telescope, A&A **517** (Jul. 2010) A26, `arXiv:1003.5613 [astro-ph.IM]`

Schmitt, J., Starck, J. L., Casandjian, J. M., Fadili, J., & Grenier, I., Multichannel Poisson denoising and deconvolution on the sphere: application to the Fermi Gamma-ray Space Telescope, A&A **546** (Oct. 2012) A114, `arXiv:1206.2787 [astro-ph.IM]`

Selig, M., Bell, M. R., Junklewitz, H., et al., NIFTY - Numerical Information Field Theory - a versatile Python library for signal inference, A&A **554** (Apr. 2013) A26, `arXiv:1301.4499 [astro-ph.IM]`

Selig, M., Oppermann, N., & Enßlin, T. A., Improving stochastic estimates with inference methods: calculating matrix diagonals, Phys. Rev. E **85** no. 2, (Feb. 2012) 021134, `arXiv:1108.0600 [astro-ph.IM]`

Shewchuk, J. R., An Introduction to the Conjugate Gradient Method Without the Agonizing Pain, Technical report, Carnegie Mellon University, Pittsburgh, PA (1994)

Strong, A. W., Maximum Entropy imaging with INTEGRAL/SPI data, A&A **411** (Nov. 2003) L127–L129

Transtrum, M. K., Machta, B. B., & Sethna, J. P., Why are Nonlinear Fits to Data so Challenging?, Physical Review Letters **104** no. 6, (Feb. 2010) 060201, `arXiv:0909.3884 [cond-mat.stat-mech]`

Transtrum, M. K. & Sethna, J. P., Improvements to the Levenberg-Marquardt algorithm for nonlinear least-squares minimization, ArXiv e-prints (Jan. 2012), `arXiv:1201.5885 [physics.data-an]`

Valdes, F., Resolution classifier, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 331, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series. 1982, 465–472, ADS

Wandelt, B. D., Larson, D. L., & Lakshminarayanan, A., Global, exact cosmic microwave background data analysis using Gibbs sampling, Phys. Rev. D **70** no. 8, (Oct. 2004) 083511, `arXiv:astro-ph/0310080`

Willett, R. & Nowak, R., Multiscale Poisson Intensity and Density Estimation, Information Theory, IEEE Transactions on **53** no. 9, (2007) 3171–3187

Wingate, D. & Weber, T., Automated Variational Inference in Probabilistic Programming, ArXiv e-prints (Jan. 2013), `arXiv:1301.1299 [stat.ML]`

Appendix A: Point source stacking

In Sec. 2.3.3, a prior for the point-like signal field has been derived under the assumption that the photon flux of point sources is independent between different pixels and identically inverse-Gamma distributed,

$$\rho_x^{(u)} \curvearrowleft \mathcal{I}\left(\rho_x^{(u)}, \beta = \frac{3}{2}, \rho_0 \eta\right) \quad \forall x, \tag{A.1}$$

with the shape and scale parameters, β and η. It can be shown that, for $\beta = \frac{3}{2}$, the sum of N such variables still obeys an inverse-Gamma distribution,

$$\rho_N^{(u)} = \sum_x^N \rho_x^{(u)} \tag{A.2}$$

$$\rho_N^{(u)} \curvearrowleft \mathcal{I}\left(\rho_N^{(u)}, \beta = \frac{3}{2}, N^2 \rho_0 \eta\right). \tag{A.3}$$

For a proof see (Giron 2001).

In the case of $\beta = \frac{3}{2}$, the power-law behavior of the prior becomes independent of the discretization of the continuous position space. This means that the slope of the distribution of $\rho_x^{(u)}$ remains unchanged notwithstanding that we refine or coarsen the resolution of the reconstruction. However, the scale parameter η needs to be adapted for each resolution; i.e., $\eta \to N^2 \eta$ if N pixels are merged.

Appendix B: Covariance & curvature.

The covariance $\boldsymbol{D}$ of a Gaussian $\mathcal{G}(\boldsymbol{s} - \boldsymbol{m}, \boldsymbol{D})$ describes the uncertainty associated with the mean $\boldsymbol{m}$ of the distribution. It can be computed by second moments or cumulants according to Eq. (3), or in this Gaussian case as the inverse Hessian of the corresponding information Hamiltonian,

$$\left.\frac{\partial^2 H}{\partial \boldsymbol{s} \partial \boldsymbol{s}^\dagger}\right|_{\boldsymbol{s}=\boldsymbol{m}} = \left.\frac{\partial^2}{\partial \boldsymbol{s} \partial \boldsymbol{s}^\dagger}\left(\frac{1}{2}(\boldsymbol{s}-\boldsymbol{m})^\dagger \boldsymbol{D}^{-1}(\boldsymbol{s}-\boldsymbol{m})\right)\right|_{\boldsymbol{s}=\boldsymbol{m}}$$
$$= \boldsymbol{D}^{-1}. \tag{B.1}$$

In Sec. 3, uncertainty covariances for the diffuse signal field $\boldsymbol{s}$ and the point-like signal field $\boldsymbol{u}$ have been derived that are here given in closed form.

The MAP uncertainty covariances introduced in Sec. 3.1 are approximated by inverse Hessians. According to Eq. (36), they read

$$D_{xy}^{(s)}{}^{-1} \approx \left\{\sum_i \left(1 - \frac{d_i}{l_i}\right) R_{ix} e^{m_x^{(s)}}\right\} \delta_{xy} \tag{B.2}$$
$$+ \sum_i \frac{d_i}{l_i^2}\left(R_{ix} e^{m_x^{(s)}}\right)\left(R_{iy} e^{m_y^{(s)}}\right) + S_{xy}^{\star}{}^{-1},$$

and

$$D_{xy}^{(u)}{}^{-1} \approx \left\{\sum_i \left(1 - \frac{d_i}{l_i}\right) R_{ix} e^{m_x^{(u)}} + \eta\, e^{-m_x^{(u)}}\right\} \delta_{xy}$$
$$+ \sum_i \frac{d_i}{l_i^2}\left(R_{ix} e^{m_x^{(u)}}\right)\left(R_{iy} e^{m_y^{(u)}}\right), \tag{B.3}$$

with

$$l_i = \int \mathrm{d}x\, R_{ix}\left(e^{m_x^{(s)}} + e^{m_x^{(u)}}\right). \tag{B.4}$$

The corresponding covariances derived in the Gibbs approach according to Eq. (51), yield

$$D_{xy}^{(s)}{}^{-1} \approx \left\{\sum_i \left(1 - \frac{d_i}{l_i}\right) R_{ix} e^{m_x^{(s)} + \frac{1}{2} D_{xx}^{(s)}}\right\} \delta_{xy} \tag{B.5}$$
$$+ \sum_i \frac{d_i}{l_i^2}\left(R_{ix} e^{m_x^{(s)} + \frac{1}{2} D_{xx}^{(s)}}\right)\left(R_{iy} e^{m_y^{(s)} + \frac{1}{2} D_{yy}^{(s)}}\right)$$
$$+ S_{xy}^{\star}{}^{-1},$$

and

$$D_{xy}^{(u)}{}^{-1} \approx \left\{\sum_i \left(1 - \frac{d_i}{l_i}\right) R_{ix} e^{m_x^{(u)} + \frac{1}{2} D_{xx}^{(u)}}\right.$$
$$\left. + \eta\, e^{-m_x^{(u)} + \frac{1}{2} D_{xx}^{(u)}}\right\} \delta_{xy} \tag{B.6}$$
$$+ \sum_i \frac{d_i}{l_i^2}\left(R_{ix} e^{m_x^{(u)} + \frac{1}{2} D_{xx}^{(u)}}\right)\left(R_{iy} e^{m_y^{(u)} + \frac{1}{2} D_{yy}^{(u)}}\right),$$

with

$$l_i = \int dx\, R_{ix} \left(e^{m_x^{(s)} + \frac{1}{2} D_{xx}^{(s)}} + e^{m_x^{(u)} + \frac{1}{2} D_{xx}^{(u)}} \right). \tag{B.7}$$

They are identical up to the $+\frac{1}{2} D_{xx}$ terms in the exponents. On the one hand, this reinforces the approximations done in Sec. 3.2. On the other hand, this shows that higher order correction terms might alter the uncertainty covariances further, cf. Eq. (43). The concrete impact of these correction terms is difficult to judge, since they introduce terms involving D_{xy} that couple all elements of $\boldsymbol{D}$ in an implicit manner.

We note that the inverse Hessian describes the curvature of the potential, its interpretation as uncertainty is, strictly speaking, only valid for quadratic potentials. However, in most cases it is a sufficient approximation.

The Gibbs approach provides an alternative by equating the first derivative of the Gibbs free energy with respect to the covariance with zero. Following Eq. (52), the covariances read

$$D_{xy}^{(s)\,-1} = \left\{ \sum_i \left(1 - \frac{d_i}{l_i} \right) R_{ix} e^{m_x^{(s)} + \frac{1}{2} D_{xx}^{(s)}} \right\} \delta_{xy} \tag{B.8}$$

$$+ S_{xy}^{\star\,-1}, \tag{B.9}$$

and

$$D_{xy}^{(u)\,-1} = \left\{ \sum_i \left(1 - \frac{d_i}{l_i} \right) R_{ix} e^{m_x^{(u)} + \frac{1}{2} D_{xx}^{(u)}} \right. \tag{B.10}$$

$$\left. + \eta\, e^{-m_x^{(u)} + \frac{1}{2} D_{xx}^{(u)}} \right\} \delta_{xy}. \tag{B.11}$$

Compared to the above solutions, there is one term missing indicating that they already lack first order corrections. For this reasons, the solutions obtained from the inverse Hessians are used in the D³PO algorithm.

Appendix C: Posterior approximation

Appendix C.1: Information theoretical measure

If the full posterior $P(\boldsymbol{z}|\boldsymbol{d})$ of an inference problem is so complex that an analytic handling is infeasible, an approximate posterior Q might be used instead. The fitness of such an approximation can be quantified by an asymmetric measure for which different terminologies appear in the literature.

First, the Kullback-Leibler divergence,

$$D_{\mathrm{KL}}(Q, P) = \int \mathcal{D}\boldsymbol{z}\, Q(\boldsymbol{z}|\boldsymbol{d}) \log \frac{Q(\boldsymbol{z}|\boldsymbol{d})}{P(\boldsymbol{z}|\boldsymbol{d})} \tag{C.1}$$

$$= \left\langle \log \frac{Q(\boldsymbol{z}|\boldsymbol{d})}{P(\boldsymbol{z}|\boldsymbol{d})} \right\rangle_Q, \tag{C.2}$$

defines mathematically an information theoretical distance, or divergence, which is minimal if a maximal cross information between P and Q exists (Kullback & Leibler 1951).

Second, the information entropy,

$$S_{\mathrm{E}}(Q, P) = - \int \mathcal{D}\boldsymbol{z}\, P(\boldsymbol{z}|\boldsymbol{d}) \log \frac{P(\boldsymbol{z}|\boldsymbol{d})}{Q(\boldsymbol{z}|\boldsymbol{d})} \tag{C.3}$$

$$= \left\langle -\log \frac{P(\boldsymbol{z}|\boldsymbol{d})}{Q(\boldsymbol{z}|\boldsymbol{d})} \right\rangle_P \tag{C.4}$$

$$= -D_{\mathrm{KL}}(P, Q),$$

is derived under the maximum entropy principle (Jaynes 1957) from fundamental axioms demanding locality, coordinate invariance and system independence Caticha (see e.g., 2008, 2011).

Third, the (approximate) Gibbs free energy (Enßlin & Weig 2010),

$$G = \left\langle H(\boldsymbol{z}|\boldsymbol{d}) \right\rangle_Q - S_{\mathrm{B}}(Q) \tag{C.5}$$

$$= \left\langle -\log P(\boldsymbol{z}|\boldsymbol{d}) \right\rangle_Q - \left\langle -\log Q(\boldsymbol{z}|\boldsymbol{d}) \right\rangle_Q \tag{C.6}$$

$$= D_{\mathrm{KL}}(Q, P),$$

describes the difference between the internal energy $\left\langle H(\boldsymbol{z}|\boldsymbol{d}) \right\rangle_Q$ and the Boltzmann-Shannon entropy $S_{\mathrm{B}}(Q) = S_{\mathrm{E}}(1, Q)$. The derivation of the Gibbs free energy is based on the principles of thermodynamics[13].

The Kullback-Leibler divergence, information entropy, and the Gibbs free energy are equivalent measures that allow one to assess the approximation $Q \approx P$. Alternatively, a parametrized proposal for Q can be pinned down by extremizing the measure of choice with respect to the parameters.

Appendix C.2: Calculus of variations

The information theoretical measure can be interpreted as an action to which the principle of least action applies. This concept is the basis for variational Bayesian methods (Jordan et al. 1999; Wingate & Weber 2013), which enable among others the derivation of approximate posterior distributions.

We suppose that $\boldsymbol{z}$ is a set of multiple signal fields, $\boldsymbol{z} = \{\boldsymbol{z}^{(i)}\}_{i\in\mathbb{N}}$, $\boldsymbol{d}$ a given data set, and $P(\boldsymbol{z}|\boldsymbol{d})$ the posterior of interest. In practice, such a problem is often addressed by a mean field approximation that factorizes the variational posterior Q,

$$P(\boldsymbol{z}|\boldsymbol{d}) \approx Q = \prod_i Q_i(\boldsymbol{z}^{(i)}|\boldsymbol{\mu}, \boldsymbol{d}). \tag{C.7}$$

Here, the mean field $\boldsymbol{\mu}$, which mimics the effect of all $\boldsymbol{z}^{(i\neq j)}$ onto $\boldsymbol{z}^{(j)}$, has been introduced. The approximation in Eq. (C.7) shifts any possible entanglement between the $\boldsymbol{z}^{(i)}$ within P into the dependence of $\boldsymbol{z}^{(i)}$ on $\boldsymbol{\mu}$ within Q_i. Hence, the mean field $\boldsymbol{\mu}$ is well determined by the inference problem at hand, as demonstrated in the subsequent Sect. C.3. We note that $\boldsymbol{\mu}$ represents effective rather than additional degrees of freedom.

Following the principle of least action, any variation of the Gibbs free energy must vanish. We consider a variation $\delta_j = \delta/\delta Q_j(\boldsymbol{z}^{(j)}|\boldsymbol{\mu}, \boldsymbol{d})$ with respect to one approximate posterior $Q_j(\boldsymbol{z}^{(j)}|\boldsymbol{\mu}, \boldsymbol{d})$. It holds,

$$\frac{\delta Q_i(\tilde{\boldsymbol{z}}^{(i)}|\boldsymbol{\mu}, \boldsymbol{d})}{\delta Q_j(\boldsymbol{z}^{(j)}|\boldsymbol{\mu}, \boldsymbol{d})} = \delta_{ij}\, \delta(\boldsymbol{z}^{(i)} - \tilde{\boldsymbol{z}}^{(j)}). \tag{C.8}$$

Computing the variation of the Gibbs free energy yields

[13] In Eq. (C.5), a unit temperature is implied, see discussion by Enßlin & Weig (2010); Iatsenko et al. (2012); Enßlin & Weig (2012)

$$\delta_j G = 0 = \frac{\delta}{\delta Q_j(\boldsymbol{z}^{(j)}|\boldsymbol{\mu},\boldsymbol{d})} \left\{ \left\langle H(\boldsymbol{z}|\boldsymbol{d})\right\rangle_Q - \left\langle -\log Q\right\rangle_Q \right\} \tag{C.9}$$

$$= \frac{\delta}{\delta Q_j(\boldsymbol{z}^{(j)}|\boldsymbol{\mu},\boldsymbol{d})} \left\{ \left\langle H(\boldsymbol{z}|\boldsymbol{d})\right\rangle_Q + \sum_i \left\langle \log Q_i(\tilde{\boldsymbol{z}}^{(i)}|\boldsymbol{\mu},\boldsymbol{d})\right\rangle_{Q_i} \right\} \tag{C.10}$$

$$= \frac{\delta}{\delta Q_j(\boldsymbol{z}^{(j)}|\boldsymbol{\mu},\boldsymbol{d})} \int \mathcal{D}\tilde{\boldsymbol{z}}^{(j)} \, Q_j(\tilde{\boldsymbol{z}}^{(j)}|\boldsymbol{\mu},\boldsymbol{d}) \left\{ \left\langle H(\boldsymbol{z}|\boldsymbol{d})\right\rangle_{\prod Q_{i\neq j}} + \log Q_j(\tilde{\boldsymbol{z}}^{(j)}|\boldsymbol{\mu},\boldsymbol{d}) \right\} + \underbrace{\frac{\delta}{\delta Q_j(\boldsymbol{z}^{(j)}|\boldsymbol{\mu},\boldsymbol{d})} \sum_{i\neq j} \cdots}_{=0}$$

$$= \int \mathcal{D}\tilde{\boldsymbol{z}}^{(j)} \, \delta(\boldsymbol{z}^{(i)} - \tilde{\boldsymbol{z}}^{(j)}) \left\{ \left\langle H(\boldsymbol{z}|\boldsymbol{d})\right\rangle_{\prod Q_{i\neq j}} + \log Q_j(\tilde{\boldsymbol{z}}^{(j)}|\boldsymbol{\mu},\boldsymbol{d}) + 1 \right\}$$

$$= \left\langle H(\boldsymbol{z}|\boldsymbol{d})\Big|_{\boldsymbol{z}^{(j)}} \right\rangle_{\prod Q_{i\neq j}} + \log Q_j(\boldsymbol{z}^{(j)}|\boldsymbol{\mu},\boldsymbol{d}) + \mathrm{const.} \tag{C.11}$$

This defines a solution for the approximate posterior Q_j, where the constant term in Eq. (C.11) ensures the correct normalization[14] of Q_j,

$$Q_j(\boldsymbol{z}^{(j)}|\boldsymbol{\mu},\boldsymbol{d}) \propto \exp\left(-\left\langle H(\boldsymbol{z}|\boldsymbol{d})\Big|_{\boldsymbol{z}^{(j)}} \right\rangle_{\prod Q_{i\neq j}} \right). \tag{C.12}$$

Although the parts $\boldsymbol{z}^{(i\neq j)}$ are integrated out, Eq. (C.12) is no marginalization since the integration is performed on the level of the (negative) logarithm of a probability distribution. The success of the mean field approach might be that this integration is often more well-behaved in comparison to the corresponding marginalization. However, the resulting equations for the Q_i depend on each other, and thus need to be solved self-consistently.

A maximum *a posteriori* solution for $\boldsymbol{z}^{(j)}$ can then be found by minimizing an effective Hamiltonian,

$$\underset{\boldsymbol{z}^{(j)}}{\mathrm{argmax}}\, P(\boldsymbol{z}|\boldsymbol{d}) = \underset{\boldsymbol{z}^{(j)}}{\mathrm{argmin}}\, H(\boldsymbol{z}|\boldsymbol{d}) \tag{C.13}$$

$$\approx \underset{\boldsymbol{z}^{(j)}}{\mathrm{argmin}}\, \left\langle H(\boldsymbol{z}|\boldsymbol{d})\Big|_{\boldsymbol{z}^{(j)}} \right\rangle_{\prod Q_{i\neq j}}. \tag{C.14}$$

Since the posterior is approximated by a product, the Hamiltonian is approximated by a sum, and each summand depends on solely one variable in the partition of the latent variable $\boldsymbol{z}$.

Appendix C.3: Example

In this section, the variational method is demonstrated with an exemplary posterior of the following form,

$$P(\boldsymbol{s},\boldsymbol{\tau}|\boldsymbol{d}) = \frac{P(\boldsymbol{d}|\boldsymbol{s})}{P(\boldsymbol{d})}\, P(\boldsymbol{s}|\boldsymbol{\tau})\, P(\boldsymbol{\tau}) \tag{C.15}$$

$$= \frac{P(\boldsymbol{d}|\boldsymbol{s})}{P(\boldsymbol{d})}\, \mathcal{G}(\boldsymbol{s},\boldsymbol{S})\, P_{\mathrm{un}}(\boldsymbol{\tau}|\boldsymbol{\alpha},\boldsymbol{q})\, P_{\mathrm{sm}}(\boldsymbol{\tau}|\boldsymbol{\sigma}), \tag{C.16}$$

where $P(\boldsymbol{d}|\boldsymbol{s})$ stands for an arbitrary likelihood describing how likely the data $\boldsymbol{d}$ can be measured from a signal $\boldsymbol{s}$, and $\boldsymbol{S} = \sum_k \mathrm{e}^{\tau_k} \boldsymbol{S}_k$ for a parametrization of the signal covariance. This posterior is equivalent to the one derived in Sec. 2 in order to find a solution for the logarithmic power
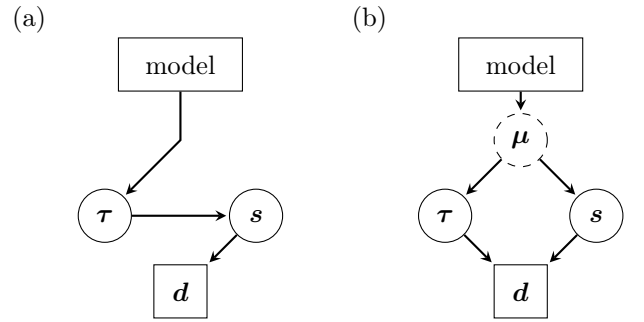

Fig. C.1. Graphical model for the variational method applied to the example posterior in Eq. (C.15). Panel (a) shows the graphical model without, and panel (b) with the mean field $\boldsymbol{\mu}$.

spectrum $\boldsymbol{\tau}$. Here, any explicit dependence on the point-like signal field $\boldsymbol{u}$ is veiled in favor of clarity.

The corresponding Hamiltonian reads

$$H(\boldsymbol{s},\boldsymbol{\tau}|\boldsymbol{d}) = -\log P(\boldsymbol{s},\boldsymbol{\tau}|\boldsymbol{d}) \tag{C.17}$$

$$= H_0 + \frac{1}{2}\sum_k \left(\varrho_k \tau_k + \mathrm{tr}\left[\boldsymbol{s}\boldsymbol{s}^\dagger \boldsymbol{S}_k^{-1}\right] \mathrm{e}^{-\tau_k} \right) \tag{C.18}$$

$$+ (\boldsymbol{\alpha}-\mathbf{1})^\dagger \boldsymbol{\tau} + \boldsymbol{q}^\dagger \mathrm{e}^{-\boldsymbol{\tau}} + \frac{1}{2}\boldsymbol{\tau}^\dagger \boldsymbol{T}\boldsymbol{\tau},$$

where $\varrho_k = \mathrm{tr}\left[\boldsymbol{S}_k \boldsymbol{S}_k^{-1}\right]$ and all terms constant in $\boldsymbol{\tau}$, including the likelihood $P(\boldsymbol{d}|\boldsymbol{s})$, have been absorbed into H_0.

For an arbitrary likelihood it might not be possible to marginalize the posterior over $\boldsymbol{s}$ analytically. However, an integration of the Hamiltonian over $\boldsymbol{s}$ might be feasible since the only relevant term is quadratic in $\boldsymbol{s}$. As, on the one hand, the prior $P(\boldsymbol{s}|\boldsymbol{\tau})$ is Gaussian and, on the other hand, a posterior mean $\boldsymbol{m}$ and covariance $\boldsymbol{D}$ for the signal field $\boldsymbol{s}$ suffice, cf. Eq. (2) and (3), we assume a Gaussian approximation for Q_s; i.e., $Q_s = \mathcal{G}(\boldsymbol{s}-\boldsymbol{m},\boldsymbol{D})$.

We now introduce a mean field approximation, denoted by $\boldsymbol{\mu}$, by changing the causal structure as depicted in Fig. C.1. With the consequential approximation of the posterior,

$$P(\boldsymbol{s},\boldsymbol{\tau}|\boldsymbol{d}) \approx \mathcal{G}(\boldsymbol{s}-\boldsymbol{m},\boldsymbol{D})\, Q_\tau(\boldsymbol{\tau}|\boldsymbol{\mu},\boldsymbol{d}), \tag{C.19}$$

---

[14] The normalization could be included by usage of Lagrange multipliers; i.e., by adding a term $\sum_i \lambda_i \left(1 - \int \mathcal{D}\boldsymbol{z}^{(i)}\, Q_i(\boldsymbol{z}^{(i)}|\boldsymbol{\mu},\boldsymbol{d})\right)$ to the Gibbs free energy in Eq. (C.9).

we can calculate the effective Hamiltonian for $\boldsymbol{\tau}$ as

$$\left\langle H(\boldsymbol{s}, \boldsymbol{\tau}|\boldsymbol{d})\Big|_{\boldsymbol{\tau}}\right\rangle_{Q_s} = H_0 + \boldsymbol{\gamma}^\dagger \boldsymbol{\tau} + \frac{1}{2}\boldsymbol{\tau}^\dagger \boldsymbol{T}\boldsymbol{\tau} + \boldsymbol{q}^\dagger \mathrm{e}^{-\boldsymbol{\tau}} \quad (C.20)$$

$$+ \frac{1}{2}\sum_k \mathrm{tr}\left[\langle \boldsymbol{s}\boldsymbol{s}^\dagger\rangle_{Q_s} \boldsymbol{S}_k^{-1}\right]\mathrm{e}^{-\tau_k}$$

$$= H_0 + \boldsymbol{\gamma}^\dagger \boldsymbol{\tau} + \frac{1}{2}\boldsymbol{\tau}^\dagger \boldsymbol{T}\boldsymbol{\tau} + \boldsymbol{q}^\dagger \mathrm{e}^{-\boldsymbol{\tau}} \quad (C.21)$$

$$+ \frac{1}{2}\sum_k \mathrm{tr}\left[\left(\boldsymbol{m}\boldsymbol{m}^\dagger + \boldsymbol{D}\right) \boldsymbol{S}_k^{-1}\right]\mathrm{e}^{-\tau_k},$$

where $\boldsymbol{\gamma} = (\boldsymbol{\alpha} - \boldsymbol{1}) + \frac{1}{2}\boldsymbol{\varrho}$.

The nature of the mean field $\boldsymbol{\mu}$ can be derived from the coupling term in Eq. (C.18) that ensures an information flow between $\boldsymbol{s}$ and $\boldsymbol{\tau}$,

$$\boldsymbol{\mu} = \begin{pmatrix} \left\langle \mathrm{tr}\left[\boldsymbol{s}\boldsymbol{s}^\dagger \boldsymbol{S}_k^{-1}\right]\right\rangle_{Q_s} \\ \left\langle \sum_k \mathrm{e}^{-\tau_k} \boldsymbol{S}_k^{-1}\right\rangle_{Q_\tau} \end{pmatrix} = \begin{pmatrix} \mathrm{tr}\left[\left(\boldsymbol{m}\boldsymbol{m}^\dagger + \boldsymbol{D}\right)\boldsymbol{S}_k^{-1}\right] \\ \left\langle \boldsymbol{S}^{-1}\right\rangle_{Q_\tau} \end{pmatrix}$$

$$(C.22)$$

Hence, the mean field effect on τ_k is given by the above trace, and the mean field effect on $\boldsymbol{s}$ is described by $\left\langle \boldsymbol{S}^{-1}\right\rangle_{Q_\tau}$.

Extremizing Eq. (C.21) yields

$$\mathrm{e}^{\boldsymbol{\tau}} = \frac{\boldsymbol{q} + \frac{1}{2}\left(\mathrm{tr}\left[\left(\boldsymbol{m}\boldsymbol{m}^\dagger + \boldsymbol{D}\right)\boldsymbol{S}_k^{-1}\right]\right)_k}{\boldsymbol{\gamma} + \boldsymbol{T}\boldsymbol{\tau}}. \quad (C.23)$$

This formula is in agreement with the critical filter formula (Enßlin & Frommert 2011; Oppermann et al. 2012). In case a Gaussian likelihood and no smoothness prior is assumed, it is the exact maximum of the true posterior with respect to the (logarithmic) power spectrum.

Table 2. Overview of the relative residual error in the photon flux reconstructions for a MAP-δ approach with varying model parameters σ, β, and η. The parameters α and q were fixed. The best and worst residuals are printed in bold face.

$\alpha=1$	$q=10^{-12}$	$\sigma=1$	$\sigma=10$	$\sigma=100$	$\sigma=1000$	$\sigma\to\infty$
$\beta=1$	$\eta=10^{-6}$	$\epsilon^{(s)}=0.06710$	$\epsilon^{(s)}=0.05406$	$\epsilon^{(s)}=0.05323$	$\epsilon^{(s)}=0.05383$	$\epsilon^{(s)}=0.05359$
		$\epsilon^{(u)}=0.02000$	$\epsilon^{(u)}=0.01941$	$\epsilon^{(u)}=0.01602$	$\epsilon^{(u)}=0.01946$	$\epsilon^{(u)}=0.01898$
$\beta=\frac{5}{4}$	$\eta=10^{-6}$	$\epsilon^{(s)}=0.02874$	$\epsilon^{(s)}=0.01929$	$\epsilon^{(s)}=0.01974$	$\epsilon^{(s)}=0.02096$	$\epsilon^{(s)}=0.01991$
		$\epsilon^{(u)}=0.01207$	$\epsilon^{(u)}=0.01102$	$\epsilon^{(u)}=\mathbf{0.01090}$	$\epsilon^{(u)}=0.01123$	$\epsilon^{(u)}=0.01104$
$\beta=\frac{3}{2}$	$\eta=10^{-6}$	$\epsilon^{(s)}=0.05890$	$\epsilon^{(s)}=0.02237$	$\epsilon^{(s)}=0.02318$	$\epsilon^{(s)}=0.02238$	$\epsilon^{(s)}=0.02344$
		$\epsilon^{(u)}=0.02741$	$\epsilon^{(u)}=0.01343$	$\epsilon^{(u)}=0.01346$	$\epsilon^{(u)}=0.01342$	$\epsilon^{(u)}=0.01351$
$\beta=\frac{7}{4}$	$\eta=10^{-6}$	$\epsilon^{(s)}=0.10864$	$\epsilon^{(s)}=0.04304$	$\epsilon^{(s)}=0.03234$	$\epsilon^{(s)}=0.03248$	$\epsilon^{(s)}=0.03263$
		$\epsilon^{(u)}=0.04840$	$\epsilon^{(u)}=0.02767$	$\epsilon^{(u)}=0.02142$	$\epsilon^{(u)}=0.02143$	$\epsilon^{(u)}=0.02167$
$\beta=2$	$\eta=10^{-6}$	$\epsilon^{(s)}=0.11870$	$\epsilon^{(s)}=0.04614$	$\epsilon^{(s)}=0.04527$	$\epsilon^{(s)}=0.04522$	$\epsilon^{(s)}=0.04500$
		$\epsilon^{(u)}=0.05360$	$\epsilon^{(u)}=0.02926$	$\epsilon^{(u)}=0.02924$	$\epsilon^{(u)}=0.02926$	$\epsilon^{(u)}=0.02915$
$\beta=1$	$\eta=10^{-4}$	$\epsilon^{(s)}=0.06660$	$\epsilon^{(s)}=0.05474$	$\epsilon^{(s)}=0.05377$	$\epsilon^{(s)}=0.05474$	$\epsilon^{(s)}=0.05423$
		$\epsilon^{(u)}=0.02157$	$\epsilon^{(u)}=0.01903$	$\epsilon^{(u)}=0.01657$	$\epsilon^{(u)}=0.01986$	$\epsilon^{(u)}=0.02055$
$\beta=\frac{5}{4}$	$\eta=10^{-4}$	$\epsilon^{(s)}=0.02874$	$\epsilon^{(s)}=\mathbf{0.01929}$	$\epsilon^{(s)}=0.01974$	$\epsilon^{(s)}=0.02096$	$\epsilon^{(s)}=0.01991$
		$\epsilon^{(u)}=0.01207$	$\epsilon^{(u)}=0.01100$	$\epsilon^{(u)}=0.01103$	$\epsilon^{(u)}=0.01123$	$\epsilon^{(u)}=0.01102$
$\beta=\frac{3}{2}$	$\eta=10^{-4}$	$\epsilon^{(s)}=0.05890$	$\epsilon^{(s)}=0.02237$	$\epsilon^{(s)}=0.02318$	$\epsilon^{(s)}=0.02238$	$\epsilon^{(s)}=0.02344$
		$\epsilon^{(u)}=0.02743$	$\epsilon^{(u)}=0.01343$	$\epsilon^{(u)}=0.01346$	$\epsilon^{(u)}=0.01340$	$\epsilon^{(u)}=0.01352$
$\beta=\frac{7}{4}$	$\eta=10^{-4}$	$\epsilon^{(s)}=0.10864$	$\epsilon^{(s)}=0.04304$	$\epsilon^{(s)}=0.03234$	$\epsilon^{(s)}=0.03248$	$\epsilon^{(s)}=0.03263$
		$\epsilon^{(u)}=0.04840$	$\epsilon^{(u)}=0.02766$	$\epsilon^{(u)}=0.02145$	$\epsilon^{(u)}=0.02142$	$\epsilon^{(u)}=0.02166$
$\beta=2$	$\eta=10^{-4}$	$\epsilon^{(s)}=0.11870$	$\epsilon^{(s)}=0.04614$	$\epsilon^{(s)}=0.04527$	$\epsilon^{(s)}=0.04522$	$\epsilon^{(s)}=0.04500$
		$\epsilon^{(u)}=0.05358$	$\epsilon^{(u)}=0.02926$	$\epsilon^{(u)}=0.02926$	$\epsilon^{(u)}=0.02927$	$\epsilon^{(u)}=0.02916$
$\beta=1$	$\eta=10^{-2}$	$\epsilon^{(s)}=0.07271$	$\epsilon^{(s)}=0.06209$	$\epsilon^{(s)}=0.06192$	$\epsilon^{(s)}=0.06291$	$\epsilon^{(s)}=0.06265$
		$\epsilon^{(u)}=0.02252$	$\epsilon^{(u)}=0.02047$	$\epsilon^{(u)}=0.02109$	$\epsilon^{(u)}=0.01764$	$\epsilon^{(u)}=0.02068$
$\beta=\frac{5}{4}$	$\eta=10^{-2}$	$\epsilon^{(s)}=0.02335$	$\epsilon^{(s)}=0.01934$	$\epsilon^{(s)}=0.02042$	$\epsilon^{(s)}=0.01999$	$\epsilon^{(s)}=0.01930$
		$\epsilon^{(u)}=0.01139$	$\epsilon^{(u)}=0.01112$	$\epsilon^{(u)}=0.01097$	$\epsilon^{(u)}=0.01124$	$\epsilon^{(u)}=0.01102$
$\beta=\frac{3}{2}$	$\eta=10^{-2}$	$\epsilon^{(s)}=0.05999$	$\epsilon^{(s)}=0.02227$	$\epsilon^{(s)}=0.02347$	$\epsilon^{(s)}=0.02266$	$\epsilon^{(s)}=0.02274$
		$\epsilon^{(u)}=0.02745$	$\epsilon^{(u)}=0.01341$	$\epsilon^{(u)}=0.01356$	$\epsilon^{(u)}=0.01332$	$\epsilon^{(u)}=0.01351$
$\beta=\frac{7}{4}$	$\eta=10^{-2}$	$\epsilon^{(s)}=0.10715$	$\epsilon^{(s)}=0.04304$	$\epsilon^{(s)}=0.03254$	$\epsilon^{(s)}=0.03264$	$\epsilon^{(s)}=0.03258$
		$\epsilon^{(u)}=0.04833$	$\epsilon^{(u)}=0.02766$	$\epsilon^{(u)}=0.02140$	$\epsilon^{(u)}=0.02144$	$\epsilon^{(u)}=0.02163$
$\beta=2$	$\eta=10^{-2}$	$\epsilon^{(s)}=0.12496$	$\epsilon^{(s)}=0.04614$	$\epsilon^{(s)}=0.04497$	$\epsilon^{(s)}=0.04528$	$\epsilon^{(s)}=0.04500$
		$\epsilon^{(u)}=\mathbf{0.05361}$	$\epsilon^{(u)}=0.02927$	$\epsilon^{(u)}=0.02915$	$\epsilon^{(u)}=0.02914$	$\epsilon^{(u)}=0.02915$
$\beta=1$	$\eta=1$	$\epsilon^{(s)}=0.15328$	$\epsilon^{(s)}=0.14544$	$\epsilon^{(s)}=0.14138$	$\epsilon^{(s)}=0.14181$	$\epsilon^{(s)}=0.14185$
		$\epsilon^{(u)}=0.03250$	$\epsilon^{(u)}=0.03291$	$\epsilon^{(u)}=0.02905$	$\epsilon^{(u)}=0.03087$	$\epsilon^{(u)}=0.02876$
$\beta=\frac{5}{4}$	$\eta=1$	$\epsilon^{(s)}=\mathbf{0.15473}$	$\epsilon^{(s)}=0.14406$	$\epsilon^{(s)}=0.14357$	$\epsilon^{(s)}=0.14465$	$\epsilon^{(s)}=0.13964$
		$\epsilon^{(u)}=0.03217$	$\epsilon^{(u)}=0.03166$	$\epsilon^{(u)}=0.03089$	$\epsilon^{(u)}=0.03101$	$\epsilon^{(u)}=0.03160$
$\beta=\frac{3}{2}$	$\eta=1$	$\epsilon^{(s)}=0.15360$	$\epsilon^{(s)}=0.14216$	$\epsilon^{(s)}=0.14248$	$\epsilon^{(s)}=0.14208$	$\epsilon^{(s)}=0.14233$
		$\epsilon^{(u)}=0.03262$	$\epsilon^{(u)}=0.03063$	$\epsilon^{(u)}=0.02534$	$\epsilon^{(u)}=0.02872$	$\epsilon^{(u)}=0.03095$
$\beta=\frac{7}{4}$	$\eta=1$	$\epsilon^{(s)}=0.15206$	$\epsilon^{(s)}=0.14156$	$\epsilon^{(s)}=0.13772$	$\epsilon^{(s)}=0.14160$	$\epsilon^{(s)}=0.14390$
		$\epsilon^{(u)}=0.03262$	$\epsilon^{(u)}=0.03065$	$\epsilon^{(u)}=0.03174$	$\epsilon^{(u)}=0.03141$	$\epsilon^{(u)}=0.03178$
$\beta=2$	$\eta=1$	$\epsilon^{(s)}=0.06421$	$\epsilon^{(s)}=0.05479$	$\epsilon^{(s)}=0.05365$	$\epsilon^{(s)}=0.05499$	$\epsilon^{(s)}=0.05429$
		$\epsilon^{(u)}=0.02043$	$\epsilon^{(u)}=0.01966$	$\epsilon^{(u)}=0.01676$	$\epsilon^{(u)}=0.02070$	$\epsilon^{(u)}=0.01996$