Information Theory Based High Energy Photon Imaging

Marco Selig

SIGILL·UNIVERSITAT·LUDOVIC·MAXIM·

München 2014

Information Theory Based High Energy Photon Imaging

Marco Selig

Dissertation
an der Fakultät für Physik
der Ludwig-Maximilians-Universität
München

vorgelegt von
Marco Selig
aus Karl-Marx-Stadt jetzt Chemnitz

München, 31. Oktober 2014

Erstgutachter: PD Dr. Torsten A. Enßlin
Zweitgutachter: Prof. Dr. Jochen Weller
Tag der mündlichen Prüfung: 20. Januar 2015

For my source of inspiration and valour, my beloved

Marie

≫ Captain, the most elementary and valuable statement in science,
the beginning of wisdom, is, › I do not know ‹ . ≪

Lieutenant Commander Data, star date 42193.6

Contents

List of Figures

List of Tables

Zusammenfassung

Die Schwerpunkte dieser Dissertation liegen in der Entwicklung und Anwendung bildgebender Verfahren im Bereich der Gammaastronomie und der medizinischen Röntgentomographie. Die Analyse der vom Fermi Gamma-ray Space Telescope gesammelten Daten kann uns einiges über den Ursprung und die Zusammensetzung der kosmischen Gammastrahlung lehren. Die hierfür notwendigen numerischen Methoden nutzen Wahrscheinlichkeitslogik und sind vielseitig anwendbar.

Die Flussdichte der extraterrestrischen Gammastrahlung ist, abgesehen von Punktquellen und Linienemission, eine kontinuierliche Funktion des Ortes und der Energie. Ausgehend von dem fundamentalen Problem, ein solches Signalfeld aus verrauschten Daten herauszufiltern, konstruieren wir einen Algorithmus zum "Entrauschen, Entfalten und Entwirren von Photonenbeobachtungen" (engl. Abk. D^{3}PO). Eine gegebene Beobachtung zerlegt der D^{3}PO-Algorithmus simultan in Schätzer für den diffusen und punktförmigen Photonenfluss, und liefert zudem Informationen über deren Unsicherheit. Dabei nutzt D^{3}PO Vorwissen innerhalb der zugrundeliegenden Bayes'schen Parameterhierarchie aus. Der D^{3}PO-Algorithmus basiert auf "Numerischer Informationsfeldtheorie" (engl. Abk. NIFTy), welche es ermöglicht Berechnungen unabhängig von räumlichen Gittern und deren Auflösung durchzuführen.

Wir nutzen den D^{3}PO-Algorithmus zur Analyse der 5,5-Jahres-Daten des Fermi Large Area Telescope (LAT). Mit unserem schablonenfreien, nichtparametrischen Rekonstruktionsverfahren enthüllen wir den diffusen Gammastrahlungsfluss auf Skalen der *Giant Fermi Bubbles* bis hinunter zu dem kleinskaligen Gammaglühen um *Centaurus A*. Am Himmel finden sich morphologisch und spektral unterschiedliche Quellen. Anhand ihrer Morphologie können wir diffuse Emission und punktförmige Quellen unterscheiden. Wir stellen einen vorläufigen Punktquellenkatalog vor, der 2.522 Kandidaten zählt, von denen wir 1.269 mit bekannten Quellen aus dem zweiten Fermi LAT Katalog assoziieren können. Eine phänomenologische Analyse der diffusen Emission offenbart zwei morphologisch und spektral unterscheidbare Komponenten. Zum Einen entdecken wir eine weiche Komponente, die Strukturen des interstellaren Mediums folgt und vermutlich durch hadronische Wechselwirkungen induziert wurde. Zum Anderen gibt es eine harte Komponente, die heiße Ausflüsse, wie die Fermi Bubbles, nachzeichnet. Wir zeigen weiterhin die erste Ganzhimmelskarte des spektralen Index des anisotropen, diffusen Gammastrahlungsflusses.

Des Weiteren übertragen wir die für die Gammaastronomie entwickelten bildgebenden Verfahren auf medizinische Röntgencomputertomographie (CT). Im Speziellen erforschen wir, ob die mit Vorwissen ausgestattete Wahrscheinlichkeitslogik bessere Bilder für diagnostische Zwecke zu liefern vermag. Solch eine Verbesserung könnte, auf lange Sicht, zu einer Verminderung der Strahlendosis bei CT-Untersuchungen führen. Wir demonstrieren die Konkurrenzfähigkeit unseres Verfahrens anhand der Rekonstruktion von anonymisierten CT-Daten.

Summary

This thesis is focused on the development of imaging techniques for high energy photon observations and their applications in γ-ray astronomy and medical X-ray computed tomography. By analyzing data from the Fermi Gamma-ray Space Telescope, we advance the knowledge on the origin and composition of the celestial γ-ray sky. The numerical method required for this purpose is derived using probabilistic reasoning and implemented in an abstract way so that it can be used in a wide range of applications.

The astrophysical γ-ray flux density is, except for point sources and line emission, a continuous function of position and energy. Starting from the fundamental problem of inferring such a continuous signal field from noisy measurement data, we construct an inference algorithm capable of "Denoising, Deconvolving, and Decomposing Photon Observations" (D^{3}PO) in the language of information field theory. Given a single photon count image, the D^{3}PO algorithm simultaneously infers individual estimates for the diffuse and point-like photon flux, and additionally provides valuable uncertainty information on the reconstruction. As a probabilistic algorithm, D^{3}PO exploits prior information throughout a hierarchical Bayesian parameter model. The D^{3}PO algorithm is based on a new software package called "Numerical Information Field Theory" (NIFTy) that ensures the algorithm's applicability irrespective of the underlying spatial grid and its resolution. To demonstrate the fidelity of the algorithm, we analyze simulated high energy photon count images in one- and two-dimensional settings.

We apply the D^{3}PO inference algorithm to the 5.5 year all-sky data from the Fermi Large Area Telescope (LAT). With our template-free, non-parametric reconstruction, we uncover the morphology of the diffuse γ-ray flux up to a few hundred GeV revealing features that scale from the *Giant Fermi Bubbles* to the γ-ray glow around *Centaurus A*. The γ-ray sky is composed of several morphologically and spectrally distinct sources. The difference in morphology allows us to distinguish diffuse γ-ray emission and contributions from point-sources. We present a preliminary point source catalog that lists 2,522 candidates of which 1,269 can be associated with known sources from the second Fermi LAT source catalog. A phenomenological analysis of the diffuse emission reveals two spectrally and morphologically distinct components. Firstly, we find a soft component that traces the cold and dense interstellar medium and is presumably induced by hadronic interactions. Secondly, there is a hard component that follows the hot and dilute gas, in particular in outflows such as the Fermi bubbles. We also present the first all-sky spectral index map for the anisotropic diffuse γ-ray flux.

Furthermore, we transfer the high energy photon imaging techniques developed for γ-ray astronomy to medical X-ray computed tomography (CT). In particular, we explore whether probabilistic methods aided by prior knowledge can yield improved images for diagnosing. Such an advance could, in the long term, reduce the radiation dose patients are exposed to when undergoing CT screenings. We demonstrate the competitiveness of our probabilistic techniques by reconstructing slices of anonymized CT data.

Chapter 1

Introduction

In this thesis, high energy photon observations, in particular the astronomical γ-ray sky as seen by the Large Area Telescope on board the Fermi satellite, are studied. The aim is to learn about the origin, composition, and properties of cosmic γ-rays and related particle populations. For this purpose, we investigate the fundamental challenge of gaining information on continuous and point-like physical quantities of interest from noisy observational data. Using probabilistic methods, we derive a versatile technique to address high energy photon imaging in an astrophysical setting. As this technique is easily transferable to other areas of application, we further demonstrate its suitability to medical imaging.

This chapter introduces the basic physics of high energy photons by reviewing relevant processes related to the production of cosmic γ-rays. Afterwards, we explain the challenges of high energy photon imaging and give an introduction to information theory. We conclude this introductory chapter by outlining the remainder of this cumulative thesis.

1.1 High Energy Photon Imaging

In the standard model of particle physics, the photon is the gauge boson of the electromagnetic force. This quantum of light has no mass, spin 1, and carries an energy $E_\nu = h\nu$, where h is the Planck constant and ν the frequency of the associated electromagnetic wave. While optical photons have energies around $2\,\mathrm{eV}$, the high energy regime starts with X-rays of energies above $100\,\mathrm{eV}$ with a seamless transition to γ-rays above $500{,}000\,\mathrm{eV}$.

In terms of thermodynamic temperature, the cosmic γ-ray energies exceed $10^{11}\,\mathrm{K}$ excluding thermal sources as their origin. Nevertheless, strong solar flares can release X-rays and even soft γ-rays. About 90% of the observed γ-ray emission is of diffuse origin (Unsöld & Baschek 2001). The main contribution in the soft regime below $10^{8}\,\mathrm{eV}$ is bremsstrahlung emitted by cosmic ray (CR) electrons. Hard γ-rays are the result of hadronic interactions of CR nuclei with the interstellar medium (ISM) or leptonic interactions, such as inverse Compton scattering of CR electrons with low energy photons (Unsöld & Baschek 2001; Longair 2011).

In this section, a discussion of the physical processes that release high energy photons is followed by an overview of the experimental challenges of astrophysical and medical photon imaging.

1.1.1 Physical Processes

Pion Production and Decay

The majority of diffuse γ-ray emission from within the Milky Way comes from hadronic interactions of CR nuclei with the interstellar gas content. The CR population at a fixed energy, say $\sim 10^9\,\mathrm{eV}$, consists of roughly 90% protons, 9% heavier nuclei, and 1% electrons.

CRs with energies up to $10^{21}\,\mathrm{eV}$ have been observed. In fact, their spectrum has been measured over an energy range of 10^8–$10^{20}\,\mathrm{eV}$ and exhibits a multiply broken power-law. We can approximate the CR spectrum by individual power-laws with different spectral indices α,

$$N(E)\,\mathrm{d}E \propto E^{-\alpha}\,\mathrm{d}E, \tag{1.1}$$

where $N(E)\mathrm{d}E$ is the number flux of CRs with energies between E and $E + \mathrm{d}E$. Up to the so-called "knee" at $\sim 10^{14}\,\mathrm{eV}$ the empirical spectral index is 2.7, above it the spectrum steepens ($\alpha \approx 3$), and finally flattens again beyond the so-called "ankle" at $\sim 10^{18}\,\mathrm{eV}$ (Unsöld & Baschek 2001).

As the proton-proton scattering becomes inelastic for center of mass energies above $10^8\,\mathrm{eV}$, the collision of relativistic CR protons with thermal gas targets can produce new particles. Typical channels result in the production of neutral or charged π-mesons; e.g.,

$$p + p \to p + p + \pi^0. \tag{1.2}$$

In a monoenergetic approximation, the production rate q_{π^0} of neutral pions with energy E_{π^0} in such a proton-proton collision,

$$q_{\pi^0} = \int \mathrm{d}E_p\; N_p(E_p)\, \delta(E_{\pi^0} - fE_p)\, \sigma_{pp} n_{\mathrm{gas}} c, \tag{1.3}$$

depends on number spectrum $N_p(E_p) \propto E_p^{-\alpha}$ of the CR protons, the total scattering cross-section σ_{pp}, and the mean number density of the gas targets n_{gas}. Further, c is the speed of light and f the fraction of the CR protons' kinetic energy transferred to the pion, which is empirically found to be 0.17 (B. & Lightman 1985; Unsöld & Baschek 2001). Since the proton-proton cross-section becomes constant above 1 GeV, the rate of pion production follows the CR proton spectrum; i.e., $q_{\pi^0} \propto E_{\pi^0}^{-\alpha}$.

In the primary decay channel of the neutral pion, $\pi^0 \to 2\gamma^*$, two γ-ray photons are released, which, in the pion rest frame, are isotropically emitted with energies $E_{\nu^*} = \frac{1}{2} m_{\pi^0} c^2$, where m_{π^0} is the mass of the neutral pion.

Since this chain of events is induced by CRs, we can derive the photon spectrum from the CR spectrum. In the observer's frame, the photons produced during the neutral pion's decay are beamed in the direction of motion of the pion. It can be shown that the resulting spectrum of the γ-ray emission is flat, but has a finite support with a logarithmic mean of $\frac{1}{2} m_{\pi^0}$. Taking this and the pion production into account, the spectrum of γ-rays produced by CR protons exhibits a bump at $\frac{1}{2} m_{\pi^0}$ and falls off with a power-law,

$$q_{\nu^*} \propto E_{\nu^*}^{-\alpha}. \tag{1.4}$$

The subdominant Dalitz decay, $\pi^0 \to \gamma + \mathrm{e}^- + \mathrm{e}^+$, as well as the decay cascade of a charged pion, generates secondary electrons (and neutrinos) that, still having relativistic velocities, can induce further γ-ray emission.

Inverse Compton Scattering

Inverse Compton (IC) scattering,

$$\mathrm{e}^- + \gamma \rightarrow \mathrm{e}^- + \gamma^*, \tag{1.5}$$

describes the collision of a relativistic electron with a low energy photon, whereby the former transfers a substantial fraction of its energy to the latter. Typical photon target fields are the cosmic microwave background or star light.

If a photon with Energy E_ν collides with an electron with energy E_e, its energy increases on average to

$$E_{\nu^*} = E_\nu\, \gamma^2 \left(1 + \frac{1}{3}\beta^2\right), \tag{1.6}$$

where $\gamma = 1/\sqrt{1-\beta^2} = E_\mathrm{e}/(m_\mathrm{e}c^2)$ is the Lorentz factor associated with the CR electron and m_e its mass. In the relativistic limit, $\beta \rightarrow 1$, we find $E_{\nu^*} \propto E_\nu E_\mathrm{e}^2$. In this way, a relativistic electron can boost a photon into the high energy regime.

In order to obtain the production spectrum of the upscattered photons, we need to integrate over the energies of the CR electrons E_e and the low energy photons E_ν,

$$q_{\nu^*} = \int \mathrm{d}E_\nu \int \mathrm{d}E_\mathrm{e}\; N_\mathrm{e}(E_\mathrm{e})\, \delta(E_{\nu^*} - E_\nu\gamma^2)\, \sigma(E_\mathrm{e}, E_\nu, E_{\nu^*})\, n_\nu(E_\nu)\, c, \tag{1.7}$$

where $N_\mathrm{e}(E_p) \propto E_\mathrm{e}^{-\alpha}$ is the number spectrum of the CR electrons, $\sigma(E_\mathrm{e}, E_\nu, E_{\nu^*})$ the total scattering cross-section, and $n_\nu(E_\nu)$ the number density of photons with energy E_ν. For simplicity, we assume the background photon field to have a uniform energy E_0; i.e., $n_\nu(E_\nu) = n_0\, \delta(E_\nu - E_0)$. In case $E_0 \ll m_\mathrm{e}c^2$, the IC scattering can be described by Thompson scattering in the electron rest frame. Since the Thompson scattering cross-section is independent of the photon's energy, we can compute the production spectrum analytically in this approximation,

$$q_{\nu^*} \propto E_0^{\frac{\alpha-1}{2}}\, E_{\nu^*}^{-\frac{\alpha+1}{2}}. \tag{1.8}$$

This approximation does not hold for CR electrons with ultrarelativistic momenta. Instead, we would need to apply the Klein-Nishina cross-section, which introduces a high energy cut-off in the photon spectrum. Nevertheless, our approximation suffices for estimating the spectral index of the photon spectrum induced by CR electrons. Over an energy range of 10–1000 GeV, experiments have shown that the CR electron (and positron) spectrum follows a power-law with a spectral index $\alpha \approx 3$ (see e.g., Strong et al. 2011). As a result, the upscattered photon spectrum would have a spectral index of $\frac{\alpha+1}{2} = 2$. This yields a distinctly harder spectrum compared to the hadronic case discussed earlier, and thus offers an approach for distinguishing γ-ray emission induced by hadronic or leptonic processes.

Bremsstrahlung and Electron-Positron Pairs

The electromagnetic radiation emitted by charged particles that are decelerated (or accelerated) due to the presence of electromagnetic fields is termed bremsstrahlung. This includes special cases such as synchrotron and cyclotron radiation.

The bremsstrahlung emitted by relativistic CR electrons in the Coulomb field of nuclei within the ISM is the dominant contribution to the soft γ-ray emission. This is a form of free-free emission, where relativistic CR electrons lose a part of their kinetic energy. It can be shown that the resulting photon spectrum, $q_{\nu^*} \propto E_{\nu^*}^{-\alpha}$, has the same spectral index $\alpha \approx 3$ as the CR electron spectrum. Therefore, bremsstrahlung becomes negligible in the hard γ-ray regime, where hadronic processes and IC scattering dominate.

If the energy of the emitted photon is much larger than the electron mass, $E_{\nu^*} \gg m_\mathrm{e}c^2$, a electron-positron pair can be created,

$$\gamma + \gamma^* \to e^- + e^+, \tag{1.9}$$

which eventually annihilates into two or three photons.[1] However, before this might happen, the electron and the positron can yet again emit bremsstrahlung. This leads to an electromagnetic cascade that ends when the photon's energy is too low for pair production or when the particles propagate into more dilute regions of the ISM.

Line emission

The above mentioned annihilation of an electron-positron pair with antiparallel spins releases two photons, which, in the center of mass rest frame, have energies equivalent to the electron mass, $E_\nu = m_\mathrm{e}c^2$. In the observer's frame, we can therefore observe beamed line emission at this energy.

In theory, the annihilation (or decay) of dark matter particles could also result in line-like γ-ray emission corresponding to the mass of the dark matter particle. In practice, however, such a "smoking gun" signal for dark matter annihilation has not (yet) been confirmed.

Another source of γ-ray lines is the eponymous γ-decay of nuclei. A nucleus in an excited state, often reached through a preceding electron capture, α-, or (inverse) β-decay, can relax to its ground state by emission of a γ-ray photon. Hence, astronomical γ-ray spectroscopy enables the study of nucleosynthesis in a Galactic context.

Medical X-rays

For medical purposes, X-rays can be generated inside a so-called X-ray tube. In this vacuum tube, electrons that were thermally emitted from a cathode are accelerated by high voltage towards the anode, which they hit producing X-ray emission. This emission consists of a bremsstrahlung continuum and line emission characteristic for the anode material. A filter absorbs (most of) the continuous spectrum leaving the X-ray lines for the actual medical screening.

X-rays interact with matter, primarily through photoelectric absorption (inducing bound-bound transitions in the electron shell of an atom) or Compton scattering. The associated (differential) loss in intensity when propagating through a medium is given by the Bouguer-Lambert-Beer law (Kak & Slaney 1988; Hsieh 2009),

$$I = I_0\,\mathrm{e}^{-a(E_\nu)\,r}, \tag{1.10}$$

where I is the intensity of the X-ray beam with an initial intensity I_0 after propagating a distance r through the medium. The X-ray absorption coefficient a is medium specific, but also depends on the energy E_ν of the X-rays. Hence, it is possible to distinguish different

---

[1] Para-positronium, with antiparallel spins, decays predominately into two photons, ortho-positronium into three.

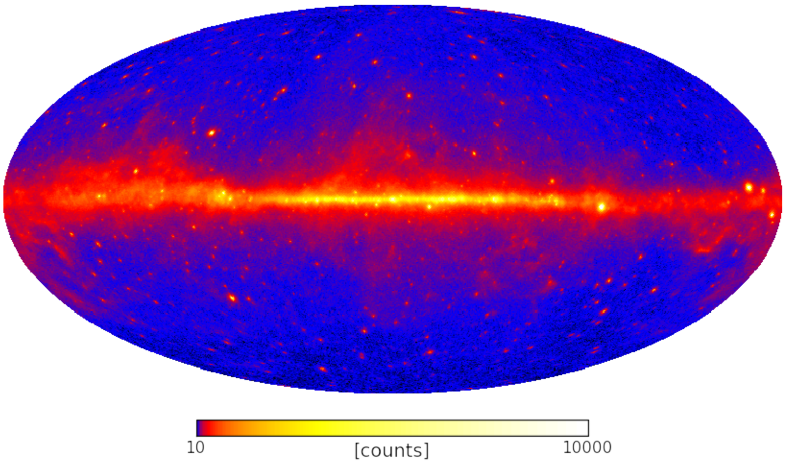


Figure 1.1: Illustration of the γ-ray sky seen by the Fermi LAT on logarithmic false color scale in Mollweide projection. Shown is the photon count data of 5.5 years mission elapsed time in the energy range from 0.6 to 307.2 GeV.

media by their transmissivity. While, on the one hand, this medium dependence is the basic working principle of medical X-ray screenings, the energy deposited inside a subject undergoing a computed tomography (CT) screening poses a certain health risk, mainly due to the possible corruption of *Deoxyribonucleic acid* (DNA) molecules.

1.1.2 Challenges in γ-ray Astronomy

As the Earth's atmosphere shields us from high energy particles, γ-ray observatories are required to be mounted on balloons or spacecrafts. The latest in a long series of successful γ-ray observatories is the Fermi Gamma-ray Space Telescope hosting the Large Area Telescope (LAT), its main instrument, and the Gamma-ray Burst Monitor.

The Fermi LAT monitors the celestial sky with unprecedented sensitivity covering an energy range from 0.02 GeV up to above 300 GeV. Figure 1.1 shows the all-sky photon data gathered during 5.5 years of operation.

The analysis of such data sets is a non-trivial task posing a number of challenges:

- **The proper treatment of noise in experimental data.** In high energy astronomy, individual events that each summarize the arrival time, incident angle, and energy of the photon reaching the detector are recorded. The measurement of these quantities, even if accurately calibrated, is subject to noise. We cannot say with certainty where the photon came from or what its energy was. The same holds for the event itself, which could just be noise – be it a CR, instead of a photon, hitting the detector or electronic noise inside the measurement circuit.

- **The rectification of instrumental imprints and artifacts.** In practice, the execution of an observation strategy often leads to spatial irregularities, as the instrument might, for example, unevenly scan the survey area or cover it incompletely. In addition, there are instrument characteristics, such as point spread functions, the original signal might be convolved with. All these instrument response functions need to be taken into account when analyzing the data.

- **The superposition of different sources.** An observation commonly captures objects that are morphologically diverse, exhibit different spectral features, and might even vary with time. In the data, such objects appear superimposed, though. In order to study these different sources, we are required to distinguish them by means of their emission properties. For example, point-like and diffuse sources have very distinct spatial characteristics. While the former are fairly local features, the latter can cover large areas exhibiting smooth fluctuation.

In order to master these challenges, we have to coherently denoise, deconvolve, and decompose high energy photon observations. The signal, the quantity we are interested in, is the γ-ray photon flux density. Applying a morphological decomposition, we obtain a diffuse and a point-like signal reconstruction.

From the point-like photon flux, we can infer point source candidates in order to build a source catalog. Most of the Galactic point sources, of which the brightest are the γ-ray active pulsars *Vela*, *Geminga*, and *Crab*, are located in or close to the Milky Way's disk. Although these three are seen at radio frequency, there are also radio-quiet gamma-pulsars. Further Galactic point sources are supernova remnants, neutron stars with very strong magnetic fields, so-called magnetars, or binary systems involving massive, compact objects like neutron stars or black holes. Extragalactic point sources seem to be homogeneously distributed over the entire sky. Some of them are so bright that they appear "quasi stellar", so-called quasars, although they are active nuclei of distant galaxies hosting a supermassive black hole. A prominent example is the active galactic nucleus *Centaurus A*, which is so bright that it became part of the eponymous constellation.

However, there are not only bright sources in the sky. Indeed some are so faint that advanced image processing needs to be applied to detect them. Once identified, the sources' flux energy spectrum and its behavior over time can be used to classify it. A point source emitting γ-rays is likely to also be visible in observations at different energies. It is therefore useful to collect all information on the source (candidate) in a catalog, and to update existing catalogs when either more data or improved imaging techniques become available. In turn, objects that have been seen at other frequencies can be searched for, and, even if they are not verified, we can provide upper limits for their γ-ray fluxes.

Most of the diffuse γ-ray emission is of Galactic origin and captured in the diffuse signal. The spectral properties of the diffuse flux can, as discussed in Sect. 1.1.1, give some indication of the astrophysical emission processes. For example, the strongest diffuse flux is found along the Galactic disk, where the ISM is dense and hadronic interactions of CRs and gas targets can induce γ-rays. The reconstruction of the diffuse γ-ray emission can therefore provide insights on the distribution and composition of CRs. The origin of the *Giant Fermi Bubbles*, two recently discovered, symmetric outflows perpendicular to the disk, is not yet fully understood. The spectral analysis of these diffuse structures might contribute to the resolution of this open question.

On the other hand, the accurate reconstruction of diffuse Galactic structures is essential for extragalactic astrophysics, where such Galactic foregrounds are a nuisance. If we are able to reconstruct the γ-ray emission from hadronic interactions that traces the ISM, this reconstruction could, for example, be used as a foreground template in studies of the cosmic

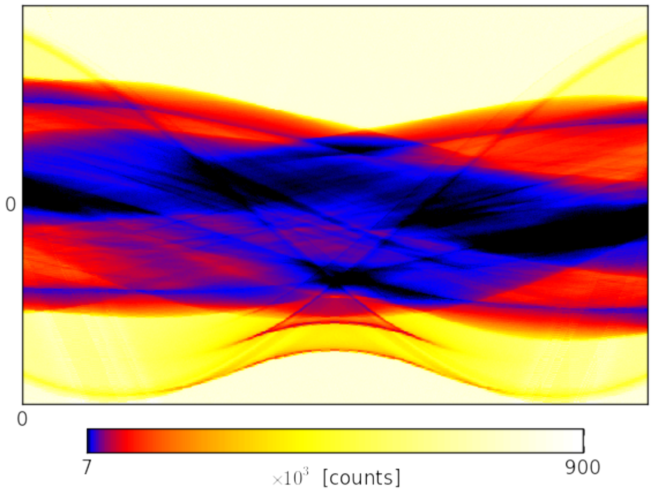


Figure 1.2: Illustration of X-ray CT data on logarithmic false color scale. Shown is a 1π scan of a single CT slice. The detector pixels are plotted along the ordinate, the 1152 projections along the axis of abscissas in angular steps of $\pi/1152$.

microwave background. Moreover, comparing such a template to other constituents of the ISM might yield insights on their dynamics.

1.1.3 Challenges in X-ray Computed Tomography

X-ray CT produces a three-dimensional image of a subject by performing a cylindrical scan. Thereby, each circular slice is screened by radial projections along certain angular increments. A typical CT data set, also called "sinogram", of a single slice is shown in Fig. 1.2.

Apart from the measurement geometry, a CT is not too different from an astronomical observation. We therefore face the same principle challenges as discussed in the previous subsection. The fact that the screened subject is exposed to an artificial X-ray source, however, raises another challenge:

- **The minimally necessary radiation dose.** The absorption of X-rays can damage living tissue and the radiation dose should therefore be kept at a minimum. However, lowering the X-ray intensity shrinks the signal-to-noise ratio that, in turn, reduces the image quality. Since the image needs still to be suitable for medical diagnosing, we have to find a trade-off between necessary image quality and unnecessary exposure.

Hence, we are challenged to advance the medical imaging techniques in order to ensure (or even improve) the image quality while diminishing the required exposure. Applying a lower current to the X-ray tube reduces the emitted intensity and results in a lower dose for the subject. The reduced number of X-ray photons is, however, reflected in a lower signal-to-noise ratio. In order to address this issue, detailed knowledge of the processes that introduce noise to the measurement is required. Alternatively, we could record fewer tomographic projections per slice. In this way, the redundancy of the CT scan is reduced, as we effectively waive parts of the measurement data. Since less data means weaker constraints, we are then confronted with a tougher inverse problem. Either way, reducing the X-ray exposure puts increased demands on the applied imaging algorithm.

1.2 Information Theory

1.2.1 Bayesics

Information can be defined as the cause of change of rational belief. In other words, we learn by updating our state of knowledge according to relevant information. However, information needs to be neither consistent nor reliable. For this reason, we require a consistent measure to judge the plausibility of information.

It can be shown that probabilities provide such a measure. Let for example the probability $P(A)$ describe the degree of rational belief in the argument A in terms of a positive real number[2], conventionally in the interval $[0,1]$. Furthermore, probabilities are conditional, so that $P(B|A)$ describes the probability of statement B given statement A. The before mentioned consistency of this measure of plausibility is warranted by, first, the sum rule for disjoint statements $\{C_i\}_i$ (such as C and its negation $\bar{C}$),

$$P(B|A) = \sum_i P(C_i, B|A) = P(C, B|A) + P(\bar{C}, B|A), \tag{1.11}$$

and second, the product rule for conditional statements,

$$P(C, B|A) = P(C|B, A) \cdot P(B|A) = P(B|C, A) \cdot P(C|A). \tag{1.12}$$

Here, commas denote conjunctions in terms of a logical “AND”.

In physics, the fundamental learning principle comes down to the confrontation of experiment and theory. Since experimental data are a fundamental source of information, their proper analysis is mandatory.

In general, we attempt to learn about a physical quantity of interest, the signal s, from measurement data d. Before the experiment, we are in an *a priori* state of knowledge that we can express in terms of a prior probability distribution $P(s)$. After gathering the data, we would like to infer information on the signal from our updated, *a posteriori* knowledge state described by the posterior probability distribution $P(s|d)$. Rewriting the product rule yields Bayes’ Theorem,

$$P(s|d) = \frac{P(s,d)}{P(d)} = \frac{P(d|s)\,P(s)}{P(d)}, \tag{1.13}$$

detailing the updating from prior to posterior given some data. This introduces the likelihood $P(d|s)$ that describes how likely it is to measure the data d when observing the signal s, and the evidence $P(d)$ that describes the probability of recording the data d considering any possible signal; i.e., marginalizing with respect to the signal,

$$P(d) = \int \mathrm{d}s\; P(s,d) = \int \mathrm{d}s\; P(d|s)\,P(s). \tag{1.14}$$

---

[2] The usage of complex valued probabilities provides a statistical approach to quantum theory (A. 1993).

The measured data do not single out a unique signal, though. In order to infer the signal, we thus have to resort to the posterior. The *a posteriori* expectation value of the signal s, which is defined as

$$\left\langle s \right\rangle_{(s|d)} = \int \mathrm{d}s\; s^1 P(s|d), \tag{1.15}$$

yields the most plausible signal (in terms of the posterior weighted mean signl) out of all possible ones under the constraints of the data at hand. One reason for this signal ambiguity is the involvement of noise that all experiments are prone to. Further, we can assess higher order moments or cumulants of the posterior with respect to the signal. The second cumulant, for example, describes the signal variance,

$$\left\langle s^2 \right\rangle^{\mathrm{c}}_{(s|d)} = \left\langle s^2 \right\rangle_{(s|d)} - \left\langle s \right\rangle^2_{(s|d)} = \int \mathrm{d}s\; s^2 P(s|d) - \left(\int \mathrm{d}s\; s^1 P(s|d) \right)^2 . \tag{1.16}$$

The variance describes the uncertainty margin around the inferred signal estimate, and hence provides crucial information. Without it, we cannot reason on how significant the signal estimate we computed is.

1.2.2 Information Fields

We only considered scalar signals, so far. If we intend to infer physical quantities like the photon flux or matter density, we have to extend our formalism to scalar fields. Let the signal $\boldsymbol{s} = s(x)$ be a field defined over some continuous space $\mho$ with $x \in \mho$. Notice that $s : \mho \to \mathbb{C}$ hence describes a continuous function, which we can evaluate at any location. In the context of parameter inference, such a signal field $\boldsymbol{s}$ is also described as "non-parametric", because it is free of an *a priori* fixed model structure defined by a limited set of parameters. The resulting subtleties in the calculus of information fields are detailed in Ch. 2. At this point, we neglect the mathematical details and focus on conceptionally new aspects.

When computing the *a posteriori* expectation value for the signal $\boldsymbol{s}$,

$$\left\langle \boldsymbol{s} \right\rangle_{(\boldsymbol{s}|\boldsymbol{d})} = \int \mathcal{D}\boldsymbol{s}\; \boldsymbol{s}\; P(\boldsymbol{s}|\boldsymbol{d}), \tag{1.17}$$

the integral $\int \mathrm{d}s$ is replaced by the phase space integral $\int \mathcal{D}\boldsymbol{s}$. This phase space is huge, since the signal is a continuous function with an infinite number of degrees of freedom and can, in principle, attain any possible value at any position.

However, the signal field representing a physical quantity constrained by finite data is often found to exhibit an effective, finite number of degrees of freedom. This restriction of the phase space volume to physically plausible signal configurations can be explained by spatial or spectral continuities and the field's internal correlation structure. One way of introducing such a regularization is the formulation of an *a priori* correlation; e.g., we could choose a prior with a non-zero second cumulant,

$$\left\langle \boldsymbol{s}\boldsymbol{s}^\dagger \right\rangle^{\mathrm{c}}_{(\boldsymbol{s})} = \boldsymbol{S}. \tag{1.18}$$

Here, $\boldsymbol{S} = S(x, y)$ is the prior covariance of the signal field $\boldsymbol{s}$. It is by definition a hermitian and positive definite linear operator. If our *a priori* state of knowledge, for example, assumes the signal to be statistically homogeneous and isotropic, we suppose the covariance to only depend on absolute distances; i.e., $\boldsymbol{S} = S(x,y) = S(|x-y|)$. It can be shown that in this case, the covariance is diagonal in the harmonic basis of $\mho$. If we further impose a certain diagonal in order to fix the covariance, we constrain the statistical fluctuations the signal is expected to exhibit. Hence, a certain continuity of the signal field is enforced.

However, such prior regularizations are merely supposed to remedy the inverse problem by limiting the degrees of freedom. Imposing too strict priors would render the inference ignorant to constraints by the data.

Arriving at the *a posteriori* knowledge state after the experiment, we can determine the second cumulant of the posterior,

$$\left\langle \boldsymbol{s}\boldsymbol{s}^{\dagger} \right\rangle^{\mathrm{c}}_{(\boldsymbol{s}|\boldsymbol{d})} = \boldsymbol{D}. \tag{1.19}$$

The operator $\boldsymbol{D} = D(x,y)$, which is also referred to as information propagator (Enßlin et al. 2009), represents a covariance. Its diagonal $D(x,x)$ therefore describes the variance of the reconstructed signal, cf. Eq. (1.16). In this way, we can assess uncertainty information on inferred information fields with infinitely many degrees of freedom.

1.3 Outline

The remainder of this cumulative thesis is structured as follows.

Chapter 2 is devoted to numerical information field theory (NIFTy). There, we introduce the basic concepts of inferring continuous signal fields from noisy measurement data with an emphasis on the formulation of inverse problems in the language of information field theory (IFT). Furthermore, we discuss and demonstrate the implementation of inference algorithms in a computer environment. The focus, hereby, lies on the flexibility of the NIFTy library to accommodate algorithms on various spatial grids and at any resolution.

In Ch. 3, we derive an inference algorithm that addresses the challenging task of denoising, deconvolving, and decomposing photon observations ($\mathrm{D}^3\mathrm{PO}$). This derivation introduces an elaborate measurement model, which can, not exclusively, cope with astronomical high energy observations. We also detail all prior assumptions fold into the hierarchical Bayesian parameter model. The fidelity of the $\mathrm{D}^3\mathrm{PO}$ algorithm is verified in a series of applications to one- and two-dimensional data sets. This is possible, because the $\mathrm{D}^3\mathrm{PO}$ code is based on the NIFTy library.

We apply the $\mathrm{D}^3\mathrm{PO}$ inference algorithm to the high energy photon data gathered by the Fermi LAT in Ch. 4. The data set we investigate comprises γ-ray events with energies between 0.6–307.2 GeV. Separating diffuse and point-like emission, we build a preliminary catalog of source candidates that lists 2,522 sources. Comparing those with the second Fermi LAT source catalog, whereby we find 1,269 associations, leaves 1,253 potentially new candidates. Our analysis also permits us to put upper limits on the total γ-ray flux that is to be expected from galaxy clusters. Moreover, we analyze the energy spectrum of the diffuse γ-ray emission and show an all-sky spectral index map for its anisotropic component. We also find that the diffuse component is phenomenologically well described by two spectrally and morphologically distinct components. We argue that the soft component is dominated by emission induced by hadronic processes since it traces the dense, cold ISM, as we find by comparing it to maps showing Galactic thermal dust emission. The hard component, in turn, seems to predominately come from leptonic processes, such as IC scattering. It seems to trace the dilute, hot gas as well as bubble- and mushroom-like outflows. We also find the angular power spectrum of the diffuse γ-ray flux to exhibit a power-law behavior on large scales.

In Ch. 5, we transfer the inference techniques, which were derived in an abstract, mathematical way, to the field of medical imaging. We describe how the D^{3}PO algorithm can easily be used to analyze medical CT data. In this context, the signal field of interest is the physical matter density, not the astrophysical photon flux density. The reconstruction of different slices from anonymized CT data demonstrates the functionality of our inference techniques. This emphasizes the conceptional and practical transferability of astronomical imaging techniques to further technologically relevant ares of application.

Finally, we conclude and give a brief outlook on future work in Ch.6.

Chapter 2

NIFTY – Numerical Information Field Theory

A versatile PYTHON library for signal inference

Note : This chapter, as well as App. A, has been accepted for publication in Astronomy & Astrophysics (Selig et al. 2013).

2.1 Introduction

In many signal inference problems, one tries to reconstruct a continuous signal field from a finite set of experimental data. The finiteness of data sets is due to their incompleteness, resolution, and the sheer duration of the experiment. A further complication is the inevitability of experimental noise, which can arise from various origins. Numerous methodological approaches to such inference problems are known in modern information theory founded by Cox (1946); Shannon (1948); Wiener (1949).

Signal inference methods are commonly formulated in an abstract, mathematical way to be applicable in various scenarios; i.e., the method itself is independent, or at least partially independent, of resolution, geometry, physical size, or even dimensionality of the inference problem. It then is up to the user to apply the appropriate method correctly to the problem at hand.

In practice, signal inference problems are solved numerically, rather than analytically. Numerical algorithms should try to preserve as much of the universality of the underlying inference method as possible, given the limitations of a computer environment, so that the code is reusable. For example, an inference algorithm developed in astrophysics that reconstructs the photon flux on the sky from high energy photon counts might also serve the purpose of reconstructing two- or three-dimensional medical images obtained from tomographical X-rays. The desire for multi-purpose, problem-independent inference algorithms is one motivation for the NIFTY package presented here. Another is to facilitate the implementation of problem specific algorithms by providing many of the essential operations in a convenient way.

NIFTY stands for “Numerical Information Field Theory”. It is a software package written in PYTHON[1][2], however, it also incorporates CYTHON[3] (Behnel et al. 2009; Seljebotn 2009), C++, and C libraries for efficient computing.

The purpose of the NIFTY library is to provide a toolkit that enables users to implement their algorithms as abstractly as they are formulated mathematically. NIFTY's field of application is kept broad and not bound to one specific methodology. The implementation of maximum entropy (Jaynes 1957, 1989), likelihood-free, maximum likelihood, or full Bayesian inference methods (Bayes 1763; Laplace 1795/1951; Cox 1946) are feasible, as well as the implementation of posterior sampling procedures based on Markov chain Monte Carlo procedures (Metropolis & Ulam 1949; Metropolis et al. 1953).

Although NIFTY is versatile, the original intention was the implementation of inference algorithms that are formulated methodically in the language of information field theory[4] (IFT). The idea of IFT is to apply information theory to the problem of signal field inference, where "field" is the physicist's term for a continuous function over a continuous space. The recovery of a field that has an infinite number of degrees of freedom from finite data can be achieved by exploiting the spatial continuity of fields and their internal correlation structures. The framework of IFT is detailed in the work by Enßlin et al. (2009) where the focus lies on a field theoretical approach to inference problems based on Feynman diagrams. An alternative approach using entropic matching based on the formalism of the Gibbs free energy can be found in the work by Enßlin & Weig (2010). IFT based methods have been developed to reconstruct signal fields without *a priori* knowledge of signal and noise correlation structures (Enßlin & Frommert 2011; Oppermann et al. 2011). Furthermore, IFT has been applied to a number of problems in astrophysics, namely to recover the large scale structure in the cosmic matter distribution using galaxy counts (Kitaura et al. 2009; Jasche et al. 2010b; Jasche & Kitaura 2010; Jasche et al. 2010a; Weig & Enßlin 2010), and to reconstruct the Faraday rotation of the Milky Way (Oppermann et al. 2012a). A more abstract application has been shown to improve stochastic estimates such as the calculation of matrix diagonals by sample averages (Selig et al. 2012).

One natural requirement of signal inference algorithms is their independence of the choice of a particular grid and a specific resolution, so that the code is easily transferable to problems that are similar in terms of the necessary inference methodology but might differ in terms of geometry or dimensionality. In response to this requirement, NIFTY comprises several commonly used pixelization schemes and their corresponding harmonic bases in an object-oriented framework. Furthermore, NIFTY preserves the continuous limit by taking care of the correct normalization of operations like scalar products, matrix-vector multiplications, and grid transformations; i.e., all operations involving position integrals over continuous domains.

The remainder of this chapter is structured as follows. In Sec. 2.2 an introduction to signal inference is given, with the focus on the representation of continuous information fields in the discrete computer environment. Sec. 2.3 provides an overview of the class hierarchy and features of the NIFTY package. The implementation of a Wiener filter algorithm demonstrates the basic functionality of NIFTY in Sec. 2.4. We conclude in Sec. 2.5.

---

[1] PYTHON homepage `http://www.python.org/`

[2] NIFTY is written in PYTHON 2 which is supported by all platforms and compatible to existing third party packages. A PYTHON 3 compliant version is left for a future upgrade.

[3] CYTHON homepage `http://cython.org/`

[4] IFT homepage `http://www.mpa-garching.mpg.de/ift/`

2.2 Concepts of Signal Inference

2.2.1 Fundamental Problem

Many signal inference problems can be reduced to a single model equation,

$$\boldsymbol{d} = f(\boldsymbol{s}, \dots), \tag{2.1}$$

where the data set $\boldsymbol{d}$ is the outcome of some function f being applied to a set of unknowns.[5] Some of the unknowns are of interest and form the signal $\boldsymbol{s}$, whereas the remaining are considered as nuisance parameters. The goal of any inference algorithm is to obtain an approximation for the signal that is "best" supported by the data. Which criteria define this "best" is answered differently by different inference methodologies.

There is in general no chance of a direct inversion of Eq. (2.1). Any realistic measurement involves random processes summarized as noise and, even for deterministic or noiseless measurement processes, the number of degrees of freedom of a signal typically outnumbers those of a finite data set measured from it, because the signal of interest might be a continuous field; e.g., some physical flux or density distribution.

In order to clarify the concept of measuring a continuous signal field, let us consider a linear measurement by some response $\boldsymbol{R}$ with additive and signal independent noise $\boldsymbol{n}$,

$$\boldsymbol{d} = \boldsymbol{R}\boldsymbol{s} + \boldsymbol{n}, \tag{2.2}$$

which reads for the individual data points,

$$d_i = \int_{\mho} \mathrm{d}x \; R_i(x) s(x) + n_i. \tag{2.3}$$

Here we introduced the discrete index $i \in \{1, \dots, N\} \subset \mathbb{N}$ and the continuous position $x \in \mho$ of some abstract position space $\mho$. For example, in the context of image reconstruction, i could label the N image pixels and x would describe real space positions.

The model given by Eq. (2.2) already poses a full inference problem since it involves an additive random process and a non-invertible signal response. As a consequence, there are many possible field configurations in the signal phase space that could explain a given data set. The approach used to single out the "best" estimate of the signal field from the data at hand is up to the choice of inference methodology. However, the implementation of any derived inference algorithm needs a proper discretization scheme for the fields defined on $\mho$. Since one might want to extend the domain of application of a successful algorithm, it is worthwhile to keep the implementation flexible with respect to the characteristics of $\mho$.

2.2.2 Discretized Continuum

The representation of fields that are mathematically defined on a continuous space in a finite computer environment is a common necessity. The goal hereby is to preserve the continuum limit in the calculus in order to ensure a resolution independent discretization.

---

[5] An alternative notation commonly found in the literature is $\boldsymbol{y} = f[\boldsymbol{x}]$. We do not use this notation in order to avoid confusion with coordinate variables, which in physics are commonly denoted by x and y.

Any partition of the continuous position space $\mho$ (with volume V) into a set of Q disjoint, proper subsets $\mho_q$ (with volumes V_q) defines a pixelization,

$$\mho = \dot{\bigcup_q} \mho_q \qquad \text{with} \quad q \in \{1, \dots, Q\} \subset \mathbb{N}, \tag{2.4}$$

$$V = \int_{\mho} \mathrm{d}x = \sum_{q=1}^{Q} \int_{\mho_q} \mathrm{d}x = \sum_{q=1}^{Q} V_q. \tag{2.5}$$

Here the number Q characterizes the resolution of the pixelization, and the continuum limit is described by $Q \to \infty$ and $V_q \to 0$ for all $q \in \{1, \dots, Q\}$ simultaneously. Moreover, Eq. (2.5) defines a discretization of continuous integrals, $\int_{\mho} \mathrm{d}x \mapsto \sum_q V_q$.

Any valid discretization scheme for a field $\boldsymbol{s}$ can be described by a mapping,

$$s(x \in \mho_q) \mapsto s_q = \int_{\mho_q} \mathrm{d}x \; w_q(x) s(x), \tag{2.6}$$

if the weighting function $w_q(x)$ is chosen appropriately. In order for the discretized version of the field to converge to the actual field in the continuum limit, the weighting functions need to be normalized in each subset; i.e., $\forall q : \int_{\mho_q} \mathrm{d}x \; w_q(x) = 1$. Choosing such a weighting function that is constant with respect to x yields

$$s_q = \frac{\int_{\mho_q} \mathrm{d}x \; s(x)}{\int_{\mho_q} \mathrm{d}x} = \langle s(x) \rangle_{\mho_q}, \tag{2.7}$$

which corresponds to a discretization of the field by spatial averaging. Another common and equally valid choice is $w_q(x) = \delta(x - x_q)$, which distinguishes some position $x_q \in \mho_q$, and evaluates the continuous field at this position,

$$s_q = \int_{\mho_q} \mathrm{d}x \; \delta(x - x_q) s(x) = s(x_q). \tag{2.8}$$

In practice, one often makes use of the spatially averaged pixel position, $x_q = \langle x \rangle_{\mho_q}$; cf. Eq. (2.7). If the resolution is high enough to resolve all features of the signal field $\boldsymbol{s}$, both of these discretization schemes approximate each other, $\langle s(x) \rangle_{\mho_q} \approx s(\langle x \rangle_{\mho_q})$, since they approximate the continuum limit by construction.[6]

All operations involving position integrals can be normalized in accordance with Eqs. (2.5) and (2.7). For example, the scalar product between two fields $\boldsymbol{s}$ and $\boldsymbol{u}$ is defined as

$$\boldsymbol{s}^{\dagger} \boldsymbol{u} = \int_{\mho} \mathrm{d}x \; s^{*}(x) u(x) \approx \sum_{q=1}^{Q} V_q \; s_q^{*} u_q, \tag{2.9}$$

where $\dagger$ denotes adjunction and $*$ complex conjugation. Since the approximation in Eq. (2.9) becomes an equality in the continuum limit, the scalar product is independent of the pixelization scheme and resolution, if the latter is sufficiently high.

---

[6] The approximation of $\langle s(x) \rangle_{\mho_q} \approx s(x_q \in \mho_q)$ marks a resolution threshold beyond which further refinement of the discretization reveals no new features; i.e., no new information content of the field $\boldsymbol{s}$.

NIFTy subclass	corresponding grid	conjugate space class
`point_space`	unstructured list of points	(none)
`rg_space`	n-dimensional regular Euclidean grid over $\mathcal{T}^n$	`rg_space`
`lm_space`	spherical harmonics	`gl_space` or `hp_space`
`gl_space`	Gauss-Legendre grid on the $\mathcal{S}^2$ sphere	`lm_space`
`hp_space`	HEALPix grid on the $\mathcal{S}^2$ sphere	`lm_space`
`nested_space`	(arbitrary product of grids)	(partial conjugation)

Table 2.1: Overview of derivatives of the NIFTy space class, the corresponding grids, and conjugate space classes.

The above line of argumentation analogously applies to the discretization of operators. For a linear operator $\boldsymbol{A}$ acting on some field $\boldsymbol{s}$ as $\boldsymbol{A}\boldsymbol{s} = \int_{\mho} \mathrm{d}y\; A(x,y)s(y)$, a matrix representation discretized in analogy to Eq. (2.7) is given by

$$A(x \in \mho_p, y \in \mho_q) \mapsto A_{pq} = \frac{\iint_{\mho_p \mho_q} \mathrm{d}x\,\mathrm{d}y\; A(x,y)}{\iint_{\mho_p \mho_q} \mathrm{d}x\,\mathrm{d}y} = \big\langle \big\langle A(x,y) \big\rangle_{\mho_p} \big\rangle_{\mho_q}. \qquad (2.10)$$

Consequential subtleties regarding operators are addressed in App. A.1.

The proper discretization of spaces, fields, and operators, as well as the normalization of position integrals, is essential for the conservation of the continuum limit. Their consistent implementation in NIFTy allows a pixelization independent coding of algorithms.

2.3 Class and Feature Overview

The NIFTy library features three main classes: spaces that represent certain grids, fields that are defined on spaces, and operators that apply to fields. In the following, we will introduce the concept of these classes and comment on further NIFTy features such as operator probing.

2.3.1 Spaces

The `space` class is an abstract class from which all other specific space subclasses are derived. Each subclass represents a grid type and replaces some of the inherited methods with its own methods that are unique to the respective grid. This framework ensures an abstract handling of spaces independent of the underlying geometrical grid and the grid's resolution.

An instance of a space subclass represents a geometrical space approximated by a specific grid in the computer environment. Therefore, each subclass needs to capture all structural and dimensional specifics of the grid and all computationally relevant quantities such as the data type of associated field values. These parameters are stored as properties of an instance of the class at its initialization, and they do not need to be accessed explicitly by the user thereafter. This prevents the writing of grid or resolution dependent code.

Spatial symmetries of a system can be exploited by corresponding coordinate transformations. Often, transformations from one basis to its harmonic counterpart can greatly reduce the computational complexity of algorithms. The harmonic basis is defined by the eigenbasis of the Laplace operator; e.g., for a flat position space it is the Fourier basis.[7] This conjuga-

---

[7] The covariance of a Gaussian random field that is statistically homogeneous in position space becomes diagonal in the harmonic basis.

tion of bases is implemented in NIFTy by distinguishing conjugate space classes, which can be obtained by the instance method `get_codomain` (and checked for by `check_codomain`). Moreover, transformations between conjugate spaces are performed automatically if required.

Thus far, NIFTy has six classes that are derived from the abstract space class. These subclasses are described here, and an overview can be found in Tab. 2.1.

- The `point_space` class merely embodies a geometrically unstructured list of points. This simplest possible kind of grid has only one parameter, the total number of points. This space is thought to be used as a default data space and neither has a conjugate space nor matches any continuum limit.
- The `rg_space` class comprises all regular Euclidean grids of arbitrary dimension and periodic boundary conditions. Such a grid is described by the number of grid points per dimension, the edge lengths of one n-dimensional pixel and a few flags specifying the origin of ordinates, internal symmetry, and basis type; i.e., whether the grid represents a position or Fourier basis. The conjugate space of a `rg_space` is another `rg_space` that is obtained by a fast Fourier transformation of the position basis yielding a Fourier basis or vice versa by an inverse fast Fourier transformation.
- The spherical harmonics basis is represented by the `lm_space` class which is defined by the maximum of the angular and azimuthal quantum numbers, ℓ and m, where $m_{\mathrm{max}} \leq \ell_{\mathrm{max}}$ and equality is the default. It serves as the harmonic basis for the instance of both the `gl_space` and the `hp_space` class.
- The `gl_space` class describes a Gauss-Legendre grid on an $\mathcal{S}^2$ sphere, where the pixels are centered at the roots of Gauss-Legendre polynomials. A grid representation is defined by the number of latitudinal and longitudinal bins, n_{lat} and n_{lon}.
- The hierarchical equal area isolatitude pixelization of an $\mathcal{S}^2$ sphere (abbreviated as HEALPix[8]) is represented by the `hp_space` class. The grid is characterized by twelve basis pixels and the n_{side} parameter that specifies how often each of them is quartered.
- The `nested_space` class is designed to comprise all possible product spaces constructed out of those described above. Therefore, it is defined by an ordered list of space instances that are meant to be multiplied by an outer product. Conjugation of this space is conducted separately for each subspace.

  For example, a 2D regular grid can be cast to a nesting of two 1D regular grids that would then allow for separate Fourier transformations along one of the two axes.

2.3.2 Fields

The second fundamental NIFTy class is the `field` class whose purpose is to represent discretized fields. Each field instance has not only a property referencing an array of field values, but also `domain` and `target` properties. The domain needs to be stated during initialization to clarify in which space the field is defined. Optionally, one can specify a target space as codomain for transformations; by default the conjugate space of the domain is used as the target space.

---

[8] HEALPix homepage `http://sourceforge.net/projects/healpix/`

method name	description
`cast_domain`	alters the field's domain without altering the field values or the codomain.
`conjugate`	complex conjugates the field values.
`dot`	applies the scalar product between two fields, returns a scalar.
`tensor_dot`	applies a tensor product between two fields, returns a field defined in the product space.
`pseudo_dot`	applies a scalar product between two fields on a certain subspace of a product space, returns a scalar or a field, depending on the subspace.
`dim`	returns the dimensionality of the field.
`norm`	returns the L^2-norm of the field.
`plot`	draws a figure illustrating the field.
`set_target`	alters the field's codomain without altering the domain or the field values.
`set_val`	alters the field values without altering the domain or codomain.
`smooth`	smooths the field values in position space by convolution with a Gaussian kernel.
`transform`	applies a transformation from the field's domain to some codomain.
`weight`	multiplies the field with the grid's volume factors (to a given power).
(and more)	

Table 2.2: Selection of instance methods of the NIFTy field class.

In this way, a field is not only implemented as a simple array, but as a class instance carrying an array of values and information about the geometry of its domain. Calling field methods then invokes the appropriate methods of the respective space without any additional input from the user. For example, the scalar product, computed by `field.dot`, applies the correct weighting with volume factors as addressed in Sec. 2.2.2 and performs basis transformations if the two fields to be scalar-multiplied are defined on different but conjugate domains.[9] The same is true for all other methods applicable to fields; see Tab. 2.2 for a selection of those instance methods.

Furthermore, NIFTy overloads standard operations for fields in order to support a transparent implementation of algorithms. Thus, it is possible to combine field instances by $+, -, *, /, \dots$ and to apply trigonometric, exponential, and logarithmic functions componentwise to fields in their current domain.

2.3.3 Operators

Up to this point, we abstracted fields and their domains leaving us with a toolkit capable of performing normalizations, field-field operations, and harmonic transformations. Now, we introduce the generic `operator` class from which other, concrete operators can be derived.

In order to have a blueprint for operators capable of handling fields, any application of operators is split into a general and a concrete part. The general part comprises the correct involvement of normalizations and transformations, necessary for any operator type, while the

---

[9] Since the scalar product by discrete summation approximates the integration in its continuum limit, it does not matter in which basis it is computed.

NIFTy subclass	description
`operator`	
↪ `diagonal_operator`	representing diagonal matrices in a specified space.
↪ `power_operator`	representing covariance matrices that are defined by a power spectrum of a statistically homogeneous and isotropic random field.
↪ `projection_operator`	representing projections onto subsets of the basis of a specified space.
↪ `vecvec_operator`	representing matrices of the form $\boldsymbol{A} = \boldsymbol{a}\boldsymbol{a}^\dagger$, where $\boldsymbol{a}$ is a field.
↪ `response_operator`	representing an exemplary response including a convolution, masking and projection.
↪ `propagator_operator`	representing the Wiener information propagator.
↪ `explicit_operator`	representing linear operators with by explicit matrices.
(and more)	

Table 2.3: Overview of derivatives of the NIFTy operator class.

concrete part is unique for each operator subclass. In analogy to the field class, any operator instance has a set of properties that specify its domain and target as well as some additional flags.

For example, the application of an operator $\boldsymbol{A}$ to a field $\boldsymbol{s}$ is coded as `A(s)`, or equivalently `A.times(s)`. The instance method `times` then invokes `_briefing`, `_multiply` and `_debriefing` consecutively. The briefing and debriefing are generic methods in which in- and output are checked; e.g., the input field might be transformed automatically during the briefing to match the operators domain. The `_multiply` method, being the concrete part, is the only contribution coded by the user. This can be done both explicitly by multiplication with a complete matrix or implicitly by a computer routine.

There are a number of basic operators that often appear in inference algorithms and are therefore pre-implemented in NIFTy. An overview of pre-implemented derivatives of the `operator` class can be found in Tab. 2.3.

2.3.4 Operator Probing

While properties of a linear operator, such as its diagonal, are directly accessible in case of an explicitly given matrix, there is no direct approach for implicitly stated operators. Even a brute force approach to calculate the diagonal elements one by one may be prohibited in such cases by the high dimensionality of the problem.

That is why the NIFTy library features a generic `probing` class. The basic idea of probing (Hutchinson 1989) is to approximate properties of implicit operators that are only accessible at a high computational expense by using sample averages. Individual samples are generated by a random process constructed to project the quantity of interest. For example, an approximation of the trace or diagonal of a linear operator $\boldsymbol{A}$ (neglecting the discretization subtleties) can be obtained by

$$\mathrm{tr}[\boldsymbol{A}] \approx \left\langle \boldsymbol{\xi}^\dagger \boldsymbol{A} \boldsymbol{\xi} \right\rangle_{\{\boldsymbol{\xi}\}} = \sum_{pq} A_{pq} \left\langle \xi_p \xi_q \right\rangle_{\{\boldsymbol{\xi}\}} \rightarrow \sum_{p} A_{pp}, \tag{2.11}$$

$$\Big(\mathrm{diag}[\boldsymbol{A}]\Big)_p \approx \Big(\langle \boldsymbol{\xi} * \boldsymbol{A}\boldsymbol{\xi} \rangle_{\{\boldsymbol{\xi}\}} \Big)_p = \sum_{q} A_{pq} \left\langle \xi_p \xi_q \right\rangle_{\{\boldsymbol{\xi}\}} \rightarrow A_{pp}, \tag{2.12}$$

where $\langle\,\cdot\,\rangle_{\{\boldsymbol{\xi}\}}$ is the sample average of a sample of random fields $\boldsymbol{\xi}$ with the property $\langle \xi_p \xi_q \rangle_{\{\boldsymbol{\xi}\}} \rightarrow \delta_{pq}$ for $|\{\boldsymbol{\xi}\}| \rightarrow \infty$ and $*$ denotes componentwise multiplication, cf. (Selig et al. 2012, and references therein). One of many possible choices for the random values of $\boldsymbol{\xi}$ are equally probable values of ±1 as originally suggested by Hutchinson (1989). Since the residual error of the approximation decreases with the number of used samples, one obtains the exact result in the limit of infinitely many samples. In practice, however, one has to find a trade-off between acceptable numerical accuracy and affordable computational cost.

The NIFTy probing class allows for the implementation of arbitrary probing schemes. Because each sample can be computed independently, all probing operations take advantage of parallel processing for reasons of efficiency, by default. There are two derivatives of the probing class implemented in NIFTy, the `trace_probing` and `diagonal_probing` subclasses, which enable the probing of traces and diagonals of operators, respectively.

An extension to improve the probing of continuous operators by exploiting their internal correlation structure as suggested in the work by Selig et al. (2012) is planned for a future version of NIFTy.

2.3.5 Parallelization

The parallelization of computational tasks is supported. NIFTy itself uses a shared memory parallelization provided by the Python standard library `multiprocessing`[10] for probing. If parallelization within NIFTy is not desired or needed, it can be turned off by the global setting flag `about.multiprocessing`.

Nested parallelization is not supported by Python; i.e., the user has to decide between the usage of parallel processing either within NIFTy or within dependent libraries such as HEALPix.

2.4 Demonstration

An established and widely used inference algorithm is the Wiener filter (Wiener 1949) whose implementation in NIFTy shall serve as a demonstration example.

The underlying inference problem is the reconstruction of a signal, $\boldsymbol{s}$, from a data set, $\boldsymbol{d}$, that is the outcome of a measurement process (2.2), where the signal response, $\boldsymbol{R}\,\boldsymbol{s}$, is linear in the signal and the noise, $\boldsymbol{n}$, is additive. The statistical properties of signal and noise are both assumed to be Gaussian,

$$\boldsymbol{s} \curvearrowleft \mathcal{G}(\boldsymbol{s}, \boldsymbol{S}) \propto \exp\left(-\tfrac{1}{2}\boldsymbol{s}^\dagger \boldsymbol{S}^{-1} \boldsymbol{s}\right), \tag{2.13}$$

$$\boldsymbol{n} \curvearrowleft \mathcal{G}(\boldsymbol{n}, \boldsymbol{N}). \tag{2.14}$$

---

[10] Python documentation `http://docs.python.org/2/library/multiprocessing.html`

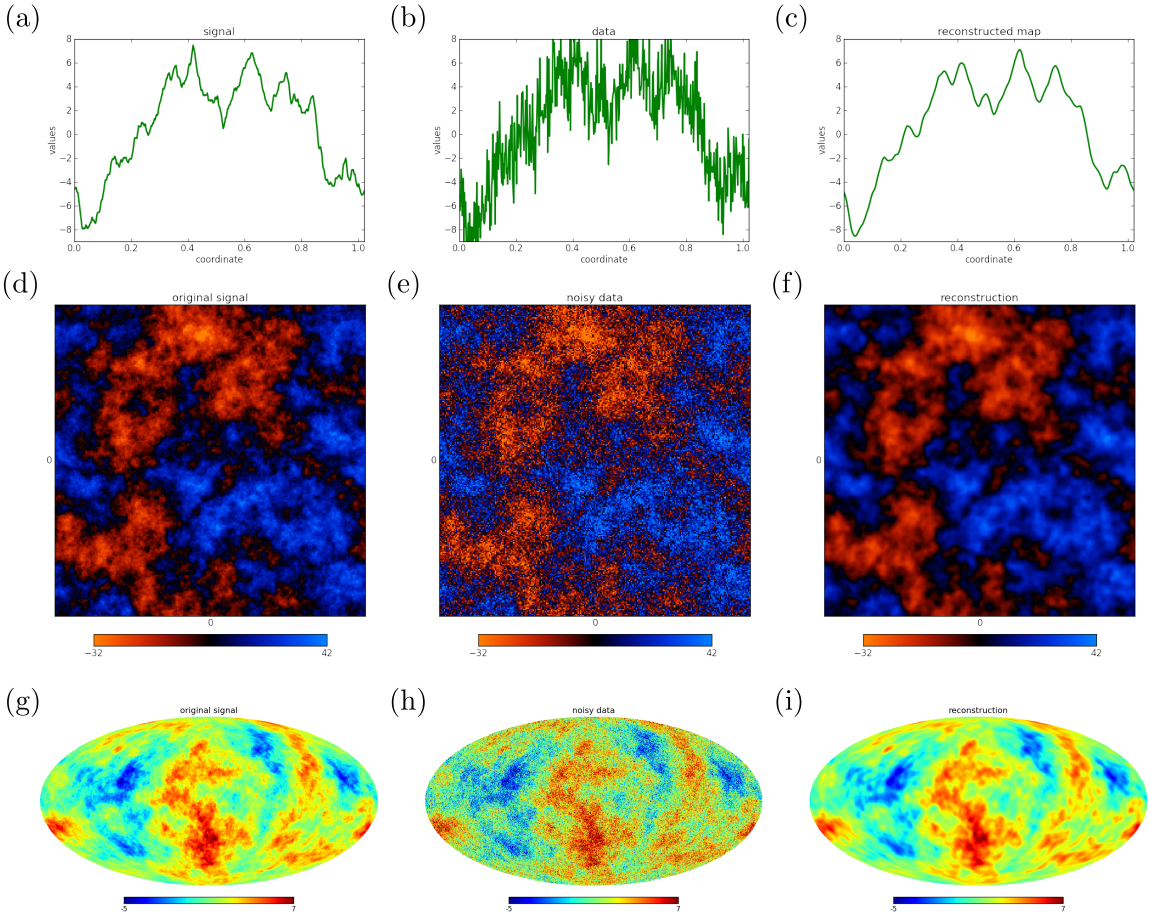


Figure 2.1: Illustration of the Wiener filter code example showing (left to right) a Gaussian random signal (a,d,g), the data including noise (b,e,h), and the reconstructed map (c,f,i). The additive Gaussian white noise has a variance σ_n^2 that sets a signal-to-noise ratio $\langle \sigma_s \rangle_{\Omega} / \sigma_n$ of roughly 2. The same code has been applied to three different spaces (top to bottom), namely a 1D regular grid with 512 pixels (a,b,c), a 2D regular grid with 256 × 256 pixels (d,e,f), and a HEALPix grid with $n_{\mathrm{side}} = 128$ corresponding to 196,608 pixels on the $\mathcal{S}^2$ sphere (g,h,i). (All figures have been created by NIFTy using the `field.plot` method.)

Here, the signal and noise covariances, $\boldsymbol{S}$ and $\boldsymbol{N}$, are known *a priori*. The *a posteriori* solution for this inference problem can be found in the expectation value for the signal $\boldsymbol{m} = \langle \boldsymbol{s} \rangle_{(\boldsymbol{s}|\boldsymbol{d})}$ weighted by the posterior $P(\boldsymbol{s}|\boldsymbol{d})$. This map can be calculated with the Wiener filter equation,

$$\boldsymbol{m} = \underbrace{\left(\boldsymbol{S}^{-1} + \boldsymbol{R}^{\dagger}\boldsymbol{N}^{-1}\boldsymbol{R}\right)^{-1}}_{\boldsymbol{D}} \underbrace{\left(\boldsymbol{R}^{\dagger}\boldsymbol{N}^{-1}\boldsymbol{d}\right)}_{\boldsymbol{j}}, \tag{2.15}$$

which is linear in the data. In the IFT framework, this scenario corresponds to a free theory as discussed in the work by Enßlin et al. (2009), where a derivation of Eq. (2.15) can be found. In analogy to quantum field theory, the posterior covariance, $\boldsymbol{D}$, is referred to as the information propagator and the data dependent term, $\boldsymbol{j}$, as the information source.

The NIFTy based implementation is given in App. A.3, where a unit response and noise

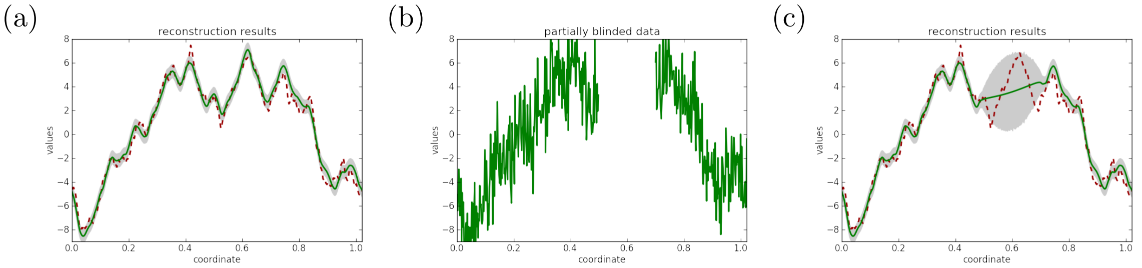


Figure 2.2: Illustration of the 1D reconstruction results. Panel (a) summarizes the results from Fig. 2.1 by showing the original signal (red dashed line), the reconstructed map (green solid line), and the 1σ-confidence interval (gray contour) obtained from the square root of the diagonal of the posterior covariance $\boldsymbol{D}$ that has been computed using probing; cf. Eq. (2.12). Panel (b) shows the 1D data set from Fig. 2.1 with a blinded region in the interval $[0.5, 0.7]$. Panel (c) shows again the original signal (red, dashed line), the map reconstructed from the partially blinded data (green solid line), and the corresponding 1σ-interval (gray contour) which is significantly enlarged in the blinded region indicating the uncertainty of the interpolation therein.

covariance are used.[11] This implementation is not only easily readable, but it also solves for $\boldsymbol{m}$ regardless of the chosen signal space; i.e., regardless of the underlying grid and its resolution. The functionality of the code for different signal spaces is illustrated in Fig. 2.1. The performance of this implementation is exemplified in Fig. 2.3 for different signal spaces and sizes of data sets. A qualitative power law behavior is apparent, but the quantitative performance depends strongly on the used machine.

The confidence in the quality of the reconstruction can be expressed in terms of a 1σ-confidence interval that is related to the diagonal of $\boldsymbol{D}$ as follows,

$$\sigma^{(\boldsymbol{m})} = \sqrt{\mathrm{diag}[\boldsymbol{D}]}. \tag{2.16}$$

The operator $\boldsymbol{D}$ defined in Eq. (2.15) may involve inversions in different bases and thus is accessible explicitly only with major computational effort. However, its diagonal can be approximated efficiently by applying operator probing (2.12). Figure 2.2 illustrates the 1D reconstruction results in order to visualize the estimates obtained with probing and to emphasize the importance of *a posteriori* uncertainties.

The Wiener filter code example given in App. A.3 can easily be modified to handle more complex inference problems. In Fig. 2.4, this is demonstrated for the image reconstruction problem of the classic “Moon Surface” image[12]. During the data generation (2.2), the signal is convolved with a Gaussian kernel, multiplied with some structured mask, and finally, contaminated by inhomogeneous Gaussian noise. Despite these complications, the Wiener filter is able to recover most of the original signal field.

NIFTy can also be applied to non-linear inference problems, as has been demonstrated in the reconstruction of log-normal fields with *a priori* unknown covariance and spectral smoothness (Oppermann et al. 2012b; Greiner & Enßlin 2013). Since its release, a wide range of scientific projects have made use of NIFTy, including, among others, studies of primordial non-Gaussianities (Dorn et al. 2013, 2014), inference of Galactic and extragalactic Faraday

---

[11] The Wiener filter demonstration is also part of the NIFTy package; see `nifty/demos/demo_excaliwir.py` for an extended version.

[12] Source taken from the USC-SIPI image database at `http://sipi.usc.edu/database/`

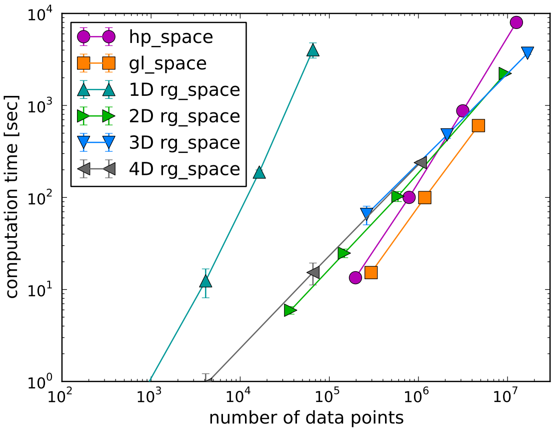


Figure 2.3: Illustration of the performance of the Wiener filter code given in App. A.3 showing computation time against the size of the data set (ranging from 512 to 256 × 256 × 256 data points) for different signal spaces (see legend). The markers show the average runtime of multiple runs, and the error bars indicate their variation. (Related markers are solely connected to guide the eye.)

rotation (Oppermann et al. 2014), and the development of novel self-calibration schemes (Enßlin et al. 2013), as well as applications in the area of Galactic tomography (Greiner et al. 2014) and weak gravitational lensing (Böhm et al. 2014), which are currently in preparation. Moreover, NIFTy is the foundation of the RESOLVE code for aperture synthesis imaging in radio astronomy (Junklewitz et al. 2013, 2014) and the D^{3}PO code for denoising, deconvolving and decomposing photon observations discussed in Ch. 3 (Selig & Enßlin 2013).

2.5 Conclusions & Summary

The NIFTy library enables the programming of grid and resolution independent algorithms. In particular for signal inference algorithms, where a continuous signal field is to be recovered, this freedom is desirable. This is achieved with an object-oriented infrastructure that comprises, among others, abstract classes for spaces, fields, and operators. NIFTy supports a consistent discretization scheme that preserves the continuum limit. Proper normalizations are applied automatically, which makes considerations by the user concerning this matter (almost) superfluous. NIFTy offers a straightforward transition from formulas to implemented algorithms thereby speeding up the development cycle. Inference algorithms that have been coded using NIFTy are reusable for similar inference problems even though the underlying geometrical space may differ.

The application areas of NIFTy are widespread and include inference algorithms derived within both information field theory and other frameworks. The successful application of a Wiener filter to non-trivial inference problems illustrates the flexibility of NIFTy. The very same code runs successfully whether the signal domain is an n-dimensional regular or a spherical grid. Moreover, NIFTy has already been applied to the reconstruction of Gaussian and log-normal fields (Oppermann et al. 2012b) and many other projects.

The NIFTy source code and online documentation is publicly available on the project homepage `http://www.mpa-garching.mpg.de/ift/nifty/`.

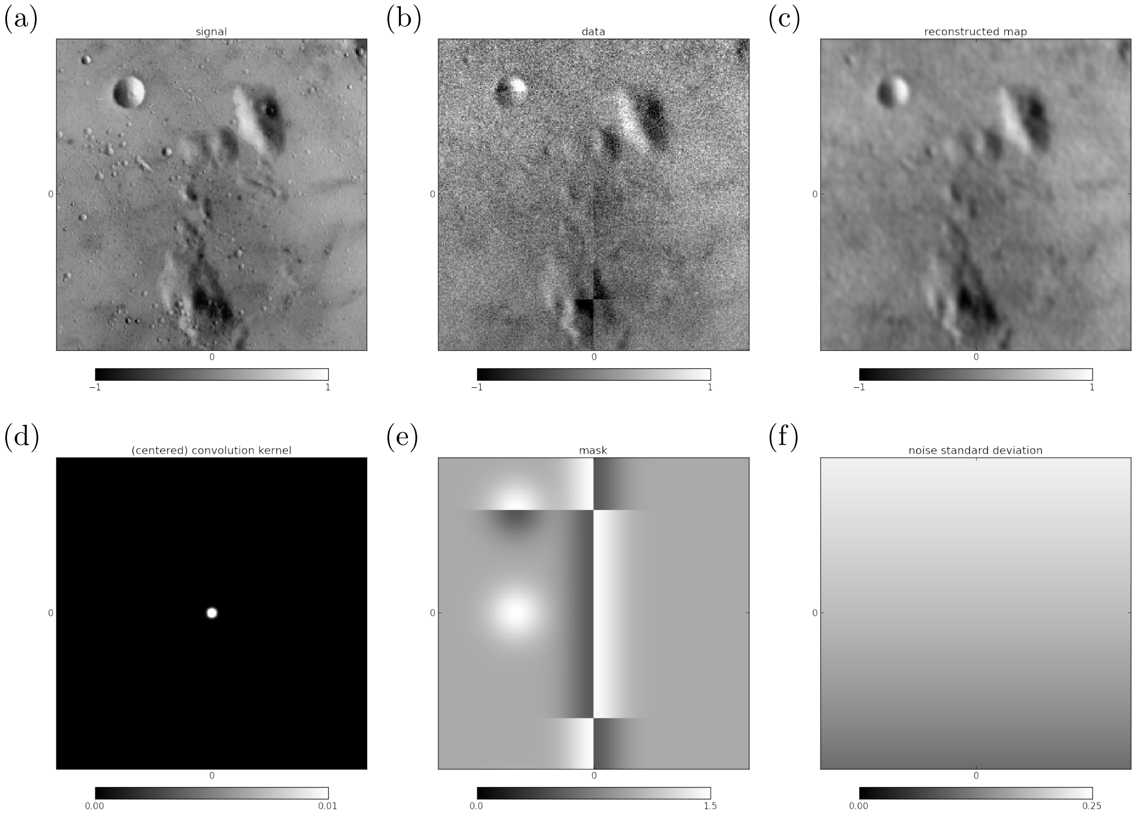


Figure 2.4: Application of a Wiener filter to the classic "Moon Surface" image on a 2D regular grid with 256 × 256 pixels showing (top, left to right) the original "Moon Surface" signal (a), the data including noise (b), and the reconstructed map (c). The response operator involves a convolution with a Gaussian kernel (d) and a masking (e). The additive noise is Gaussian white noise with an inhomogeneous standard deviation (f) that approximates an overall signal-to-noise ratio $\langle \sigma_s \rangle_{\mho} \, / \, \langle \sigma_n \rangle_{\mho}$ of roughly 1. (All figures have been created by NIFTy using the `field.plot` method.)

Chapter 3

Denoising, Deconvolving, and Decomposing Photon Observations

Derivation of the D^3PO Algorithm

Note : This chapter, as well as App. B, has been accepted for publication in Astronomy & Astrophysics (Selig & Enßlin 2013).

3.1 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 suffering from shot noise and distortions due to 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 cumbers 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. Especially point-like sources are 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, 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 an unique solution. There are a number of heuristic and probabilistic approaches that address the problem of denoising, deconvolution, and decomposition in parts or simpler settings.

SExtractor (Bertin & Arnouts 1996) is one of the heuristic kind and the most prominent tool 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. 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.

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 D^{3}PO algorithm presented here targets the simultaneous denoising, deconvolution, and decomposition of photon observations into two signals, the diffuse and point-like photon flux.

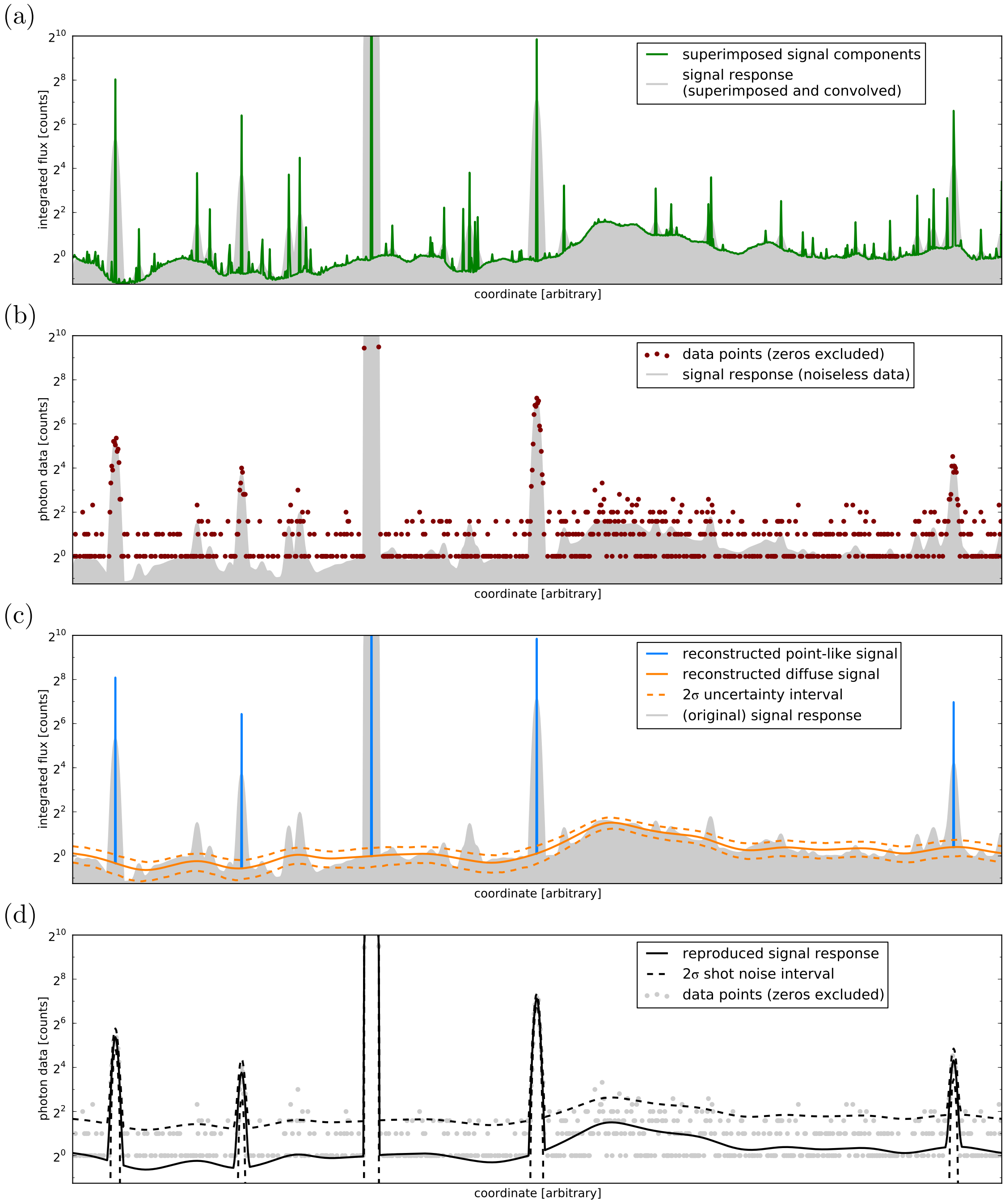


Figure 3.1: Illustration of a 1D reconstruction scenario with 1024 pixels. Panel (a) shows the superimposed diffuse and point-like signal components (green solid line) and its observational response (gray contour). Panel (b) shows again the signal response representing noiseless data (gray contour) and 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 5 are not invisibly 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).

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] Notice 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 D^{3}PO algorithm.

Figure 3.1 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 that an algorithm can 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 chapter is structured as follows. Sec. 3.2 discusses the inference on photon observations; i.e., the underlying model and prior assumptions. The D^{3}PO algorithm solving this inference problem by denoising, deconvolution, and decomposition is derived in Sec. 3.3. In Sec. 3.4 the algorithm is demonstrated in a numerical application on simulated high energy photon data. We conclude in Sec. 3.5.

---

[1] If the resolution of the reconstruction would be 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/`

3.2 Inference on Photon Observations

3.2.1 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 due to 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,

$$P(\boldsymbol{s}|\boldsymbol{d}) = \frac{P(\boldsymbol{d}|\boldsymbol{s})\,P(\boldsymbol{s})}{P(\boldsymbol{d})}, \tag{3.1}$$

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 $\mho$. In order to avoid a dependence of the reconstruction on the partition of $\mho$, the according calculus regarding fields is geared to preserve the continuum limit, cf. (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

$$\boldsymbol{m} = \langle \boldsymbol{s} \rangle_{(\boldsymbol{s}|\boldsymbol{d})} = \int \mathcal{D}\boldsymbol{s}\ \boldsymbol{s}\ P(\boldsymbol{s}|\boldsymbol{d}), \tag{3.2}$$

$$\boldsymbol{D} = \left\langle (\boldsymbol{m}-\boldsymbol{s})(\boldsymbol{m}-\boldsymbol{s})^{\dagger} \right\rangle_{(\boldsymbol{s}|\boldsymbol{d})}, \tag{3.3}$$

where † 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]

---

[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.

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. (3.1).

3.2.2 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 $i \in \{1, \dots, N_{\mathrm{pix}}\} \subset \mathbb{N}$.

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 $\mho$. In astrophysics, this space $\mho$ 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. 3.2.3 and 3.2.3, 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,

$$\boldsymbol{\rho} = \boldsymbol{\rho}_{\mathrm{diffuse}} + \boldsymbol{\rho}_{\mathrm{point-like}} = \rho_0 \left(\mathrm{e}^{\boldsymbol{s}} + \mathrm{e}^{\boldsymbol{u}} \right), \tag{3.4}$$

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. (3.4) 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,

$$\boldsymbol{\lambda} = \boldsymbol{R}_0 \boldsymbol{\rho} = \boldsymbol{R} \left(\mathrm{e}^{\boldsymbol{s}} + \mathrm{e}^{\boldsymbol{u}} \right), \tag{3.5}$$

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

$$\lambda_i = \int_{\mho} \mathrm{d}x \; R_i(x) \left(\mathrm{e}^{s(x)} + \mathrm{e}^{u(x)} \right). \tag{3.6}$$

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. (3.6), cause a severe loss of information 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,

$$P(\boldsymbol{d}|\boldsymbol{\lambda}) = \prod_i \mathcal{P}(d_i, \lambda_i) = \prod_i \frac{1}{d_i!} \, \lambda_i^{d_i} \mathrm{e}^{-\lambda_i}. \tag{3.7}$$

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. (3.5) and (3.7). 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]

$$\begin{aligned} H(\boldsymbol{d}|\boldsymbol{s},\boldsymbol{u}) &= -\log P(\boldsymbol{d}|\boldsymbol{s},\boldsymbol{u}) && (3.8)\\ &= H_0 + \boldsymbol{1}^\dagger \boldsymbol{\lambda} - \boldsymbol{d}^\dagger \log(\boldsymbol{\lambda}) && (3.9)\\ &= H_0 + \boldsymbol{1}^\dagger \boldsymbol{R}\left(\mathrm{e}^{\boldsymbol{s}} + \mathrm{e}^{\boldsymbol{u}}\right) - \boldsymbol{d}^\dagger \log\left(\boldsymbol{R}\left(\mathrm{e}^{\boldsymbol{s}} + \mathrm{e}^{\boldsymbol{u}}\right)\right), && (3.10) \end{aligned}$$

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

3.2.3 Prior Assumptions

The diffuse and point-like signal fields, $\boldsymbol{s}$ and $\boldsymbol{u}$, contribute equally to the likelihood defined by Eq. (3.10), 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 by point-sources only, 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.

Diffuse Component

The diffuse photon flux, $\boldsymbol{\rho}^{(s)} = \rho_0 \mathrm{e}^{\boldsymbol{s}}$, 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. 2012b; Kinney 2013). If the diffuse photon flux follows a multivariate log-normal distribution, the diffuse signal field $\boldsymbol{s}$ obeys a multivariate Gaussian distribution $\mathcal{G}$,

$$P(\boldsymbol{s}|\boldsymbol{S}) = \mathcal{G}(\boldsymbol{s},\boldsymbol{S}) = \frac{1}{\sqrt{\det[2\pi \boldsymbol{S}]}} \exp\left(-\frac{1}{2}\boldsymbol{s}^\dagger \boldsymbol{S}^{-1}\boldsymbol{s}\right), \tag{3.11}$$

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.

---

[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.

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 $\mho$ such that

$$\boldsymbol{S} = \sum_k \mathrm{e}^{\tau_k} \boldsymbol{S}_k, \tag{3.12}$$

where τ_k are spectral parameters and $\boldsymbol{S}_k$ are projections onto a set of disjoint harmonic subspaces of $\mho$. 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.

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. 2010b; 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. Let us initially assume independent inverse-Gamma distributions $\mathcal{I}$ for the individual elements,

$$P(\mathrm{e}^{\boldsymbol{\tau}} | \boldsymbol{\alpha}, \boldsymbol{q}) = \prod_k \mathcal{I}(\mathrm{e}^{\tau_k}, \alpha_k, q_k) = \prod_k \frac{q_k^{\alpha_k - 1}}{\Gamma(\alpha_k - 1)} \, \mathrm{e}^{-\left(\alpha_k \tau_k + q_k \mathrm{e}^{-\tau_k}\right)}, \tag{3.13}$$

and hence

$$P_{\mathrm{un}}(\boldsymbol{\tau} | \boldsymbol{\alpha}, \boldsymbol{q}) = \prod_k \mathcal{I}(\mathrm{e}^{\tau_k}, \alpha_k, q_k) \left| \frac{\mathrm{d}\mathrm{e}^{\tau_k}}{\mathrm{d}\tau_k} \right| \propto \exp\Big(- (\boldsymbol{\alpha} - \boldsymbol{1})^{\dagger} \boldsymbol{\tau} - \boldsymbol{q}^{\dagger} \mathrm{e}^{-\boldsymbol{\tau}} \Big), \tag{3.14}$$

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

---

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

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

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. (2012b). 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

$$P_{\mathrm{sm}}(\boldsymbol{\tau}|\boldsymbol{\sigma}) \propto \exp\left(-\frac{1}{2}\boldsymbol{\tau}^{\dagger}\boldsymbol{T}\boldsymbol{\tau}\right), \tag{3.15}$$

with

$$\boldsymbol{\tau}^{\dagger}\boldsymbol{T}\boldsymbol{\tau} = \int \mathrm{d}(\log k)\frac{1}{\sigma_k^2}\left(\frac{\partial^2\tau_k}{\partial(\log k)^2}\right)^2, \tag{3.16}$$

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 of 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,

$$P(\boldsymbol{\tau}|\boldsymbol{\alpha},\boldsymbol{q},\boldsymbol{\sigma}) = P_{\mathrm{un}}(\boldsymbol{\tau}|\boldsymbol{\alpha},\boldsymbol{q})\ P_{\mathrm{sm}}(\boldsymbol{\tau}|\boldsymbol{\sigma}). \tag{3.17}$$

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.

Point-like Component

The point-like photon flux, $\boldsymbol{\rho}^{(u)} = \rho_0 \mathrm{e}^{\boldsymbol{u}}$, is supposed to originate from very distant astrophysical sources. These sources appear morphologically point-like to an observer because their actual extent is negligible due to 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.

Due to 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 $\mho$, is discretized properly[7], this spot is sufficiently identified by an image pixel in the reconstruction. A discretization, $\rho(x \in \mho) \to (\rho_x)_x$, 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.

---

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

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

$$P(\boldsymbol{\rho}^{(u)}) = \prod_x P(\rho_x^{(u)}), \tag{3.18}$$

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, it further needs to be considered that the photon flux is a strictly positive quantity. Thus, a simple exponential prior,

$$P(\rho_x^{(u)}) \propto \exp\left(-\rho_x^{(u)}/\rho_0\right), \tag{3.19}$$

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]

$$P(\rho_x^{(u)}) \propto \left(\rho_x^{(u)}/\rho_0\right)^{\left(-\rho_x^{(u)}/\rho_0\right)}. \tag{3.20}$$

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 due to 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 can not be found in arbitrary distances due 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).

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

$$P(\boldsymbol{\rho}^{(u)}|\boldsymbol{\beta},\boldsymbol{\eta}) = \prod_x \mathcal{I}(\rho_x^{(u)}, \beta_x, \rho_0\eta_x) = \prod_x \frac{(\rho_0\eta_x)^{\beta_x - 1}}{\Gamma(\beta_x - 1)} \left(\rho_x^{(u)}\right)^{-\beta_x} \exp\left(-\frac{\rho_0\eta_x}{\rho_x^{(u)}}\right), \tag{3.21}$$

yielding

$$P(\boldsymbol{u}|\boldsymbol{\beta},\boldsymbol{\eta}) = \prod_x \mathcal{I}(\rho_0 \mathrm{e}^{u_x}, \beta_k, \rho_0\eta_k) \left|\frac{\mathrm{d}\rho_0\mathrm{e}^{u_x}}{\mathrm{d}u_x}\right| \propto \exp\Big(-(\boldsymbol{\beta}-\boldsymbol{1})^\dagger\boldsymbol{u} - \boldsymbol{\eta}^\dagger\mathrm{e}^{-\boldsymbol{u}}\Big), \tag{3.22}$$

---

[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. (3.18) and (3.20).

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

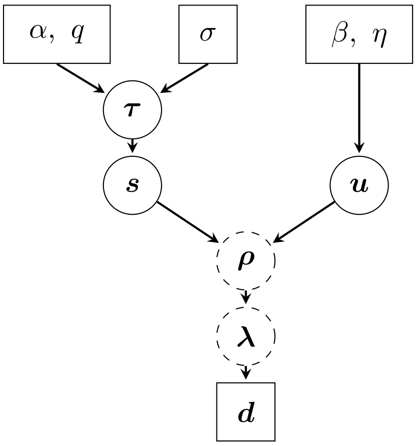


Figure 3.2: Graphical model of the model parameters α, q, σ, β, and η, the logarithmic spectral parameters $\boldsymbol{\tau}$, the diffuse signal field $\boldsymbol{s}$, the point-like signal field $\boldsymbol{u}$, the total photon flux $\boldsymbol{\rho}$, the expected number of photons $\boldsymbol{\lambda}$, and the observed photon count data $\boldsymbol{d}$.

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 App. B.1. There it is also shown that η has to be chosen resolution dependent, though.

3.2.4 Parameter Model

Figure 3.2 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. 3.4, different parameter choices for the analysis of simulated data are investigated.

3.3 Denoising, Deconvolution, and Decomposition

The likelihood model, describing the measurement process, and the prior assumptions for the signal fields and the power spectrum of the diffuse component yield a well-defined inference problem. The corresponding posterior is given by

$$P(\boldsymbol{s},\boldsymbol{\tau},\boldsymbol{u}|\boldsymbol{d}) = \frac{P(\boldsymbol{d}|\boldsymbol{s},\boldsymbol{u})\ P(\boldsymbol{s}|\boldsymbol{\tau})\ P(\boldsymbol{\tau}|\alpha,q,\sigma)\ P(\boldsymbol{u}|\beta,\eta)}{P(\boldsymbol{d})}, \tag{3.23}$$

which is a complex form of Bayes' theorem (3.1).

Ideally, we would now calculate the *a posteriori* expectation values and uncertainties according to Eqs. (3.2) and (3.3) 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 due to 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 due to the huge parameter phase space. Nevertheless, similar problems have been addressed by elaborate sampling techniques (Wandelt et al. 2004; Jasche et al. 2010b; 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. 3.4.

3.3.1 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,

$$\langle \boldsymbol{s} \rangle_{(\boldsymbol{s}|\boldsymbol{d})} \overset{\text{MAP-}\delta}{\approx} \int \mathcal{D}\boldsymbol{s}\ \boldsymbol{s}\ \delta(\boldsymbol{s} - \boldsymbol{s}_{\text{mode}}), \tag{3.24}$$

or using a Gaussian approximation around this point,

$$\langle \boldsymbol{s} \rangle_{(\boldsymbol{s}|\boldsymbol{d})} \overset{\text{MAP-}\mathcal{G}}{\approx} \int \mathcal{D}\boldsymbol{s}\ \boldsymbol{s}\ \mathcal{G}(\boldsymbol{s} - \boldsymbol{s}_{\text{mode}}, \boldsymbol{D}_{\text{mode}}), \tag{3.25}$$

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,

$$\begin{aligned} H(\boldsymbol{s},\boldsymbol{\tau},\boldsymbol{u}|\boldsymbol{d}) &= -\log P(\boldsymbol{s},\boldsymbol{\tau},\boldsymbol{u}|\boldsymbol{d}) && (3.26)\\ &= H_0 + \boldsymbol{1}^\dagger \boldsymbol{R}\left(\mathrm{e}^{\boldsymbol{s}} + \mathrm{e}^{\boldsymbol{u}}\right) - \boldsymbol{d}^\dagger \log\left(\boldsymbol{R}\left(\mathrm{e}^{\boldsymbol{s}} + \mathrm{e}^{\boldsymbol{u}}\right)\right) \\ &\quad + \frac{1}{2}\log\left(\det\left[\boldsymbol{S}\right]\right) + \frac{1}{2}\boldsymbol{s}^\dagger \boldsymbol{S}^{-1} \boldsymbol{s} && (3.27)\\ &\quad + (\boldsymbol{\alpha} - \boldsymbol{1})^\dagger \boldsymbol{\tau} + \boldsymbol{q}^\dagger \mathrm{e}^{-\boldsymbol{\tau}} + \frac{1}{2}\boldsymbol{\tau}^\dagger \boldsymbol{T} \boldsymbol{\tau} \\ &\quad + (\boldsymbol{\beta} - \boldsymbol{1})^\dagger \boldsymbol{u} + \boldsymbol{\eta}^\dagger \mathrm{e}^{-\boldsymbol{u}}, \end{aligned}$$

where all terms constant in $\boldsymbol{s}$, $\boldsymbol{\tau}$, and $\boldsymbol{u}$ have been absorbed into a ground state energy H_0, cf. Eqs. (3.7), (3.11), (3.17), and (3.22), 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. 3.3.4 for details.

In order to better understand the structure of the MAP solution, let us 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

$$\left.\frac{\partial H}{\partial \boldsymbol{s}}\right|_{\mathrm{min}} = \boldsymbol{0} = (\boldsymbol{1} - \boldsymbol{d}/\boldsymbol{l})^{\dagger}\, \boldsymbol{R} * \mathrm{e}^{\boldsymbol{m}^{(s)}} + {\boldsymbol{S}^{\star}}^{-1} \boldsymbol{m}^{(s)}, \tag{3.28}$$

$$\left.\frac{\partial H}{\partial \boldsymbol{u}}\right|_{\mathrm{min}} = \boldsymbol{0} = (\boldsymbol{1} - \boldsymbol{d}/\boldsymbol{l})^{\dagger}\, \boldsymbol{R} * \mathrm{e}^{\boldsymbol{m}^{(u)}} + \boldsymbol{\beta} - \boldsymbol{1} - \boldsymbol{\eta} * \mathrm{e}^{-\boldsymbol{m}^{(u)}}, \tag{3.29}$$

with

$$\boldsymbol{l} = \boldsymbol{R}\left(\mathrm{e}^{\boldsymbol{m}^{(s)}} + \mathrm{e}^{\boldsymbol{m}^{(u)}}\right), \tag{3.30}$$

$$\boldsymbol{S}^{\star} = \sum_k \mathrm{e}^{\tau^{\star}_k} \boldsymbol{S}_k. \tag{3.31}$$

Here, $*$ and $/$ denote componentwise multiplication and division, respectively. The first term in Eq. (3.28) and (3.29), which comes from the likelihood, vanishes in case $\boldsymbol{l} = \boldsymbol{d}$. Notice that $\boldsymbol{l} = \boldsymbol{\lambda}|_{\mathrm{min}}$ 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. (3.5) and (3.7). 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 closer 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,

$$\left.\frac{\partial^2 H}{\partial \boldsymbol{s} \partial \boldsymbol{s}^{\dagger}}\right|_{\mathrm{min}} \approx {\boldsymbol{D}^{(s)}}^{-1}, \qquad \left.\frac{\partial^2 H}{\partial \boldsymbol{u} \partial \boldsymbol{u}^{\dagger}}\right|_{\mathrm{min}} \approx {\boldsymbol{D}^{(u)}}^{-1}. \tag{3.32}$$

The closed form of $\boldsymbol{D}^{(s)}$ and $\boldsymbol{D}^{(u)}$ is given explicitly in App. B.2.

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

$$\mathrm{e}^{\boldsymbol{\tau}^{\star}} = \frac{\boldsymbol{q} + \frac{1}{2}\left(\mathrm{tr}\left[\boldsymbol{m}^{(s)} {\boldsymbol{m}^{(s)}}^{\dagger} \boldsymbol{S}_k^{-1}\right]\right)_k}{\boldsymbol{\gamma} + \boldsymbol{T}\boldsymbol{\tau}^{\star}}, \tag{3.33}$$

where $\boldsymbol{\gamma} = (\boldsymbol{\alpha} - \boldsymbol{1}) + \frac{1}{2}\left(\mathrm{tr}\left[\boldsymbol{S}_k {\boldsymbol{S}_k}^{-1}\right]\right)_k$. This formula is in accordance with the results by Enßlin & Frommert (2011); Oppermann et al. (2012b). 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.

3.3.2 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. (3.23) is inaccessible due to 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,

$$P(\boldsymbol{s}, \boldsymbol{\tau}, \boldsymbol{u}|\boldsymbol{d}) \approx Q = Q_s(\boldsymbol{s}, \boldsymbol{u}|\boldsymbol{\mu}, \boldsymbol{d})\; Q_\tau(\boldsymbol{\tau}|\boldsymbol{\mu}, \boldsymbol{d}), \tag{3.34}$$

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. (3.34). 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 App. B.3.3, 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

$$Q_s(\boldsymbol{s}, \boldsymbol{u}|\boldsymbol{\mu}, \boldsymbol{d}) = \mathcal{G}(\boldsymbol{\varphi}, \boldsymbol{D}), \tag{3.35}$$

with

$$\boldsymbol{\varphi} = \begin{pmatrix} \boldsymbol{s} - \boldsymbol{m}^{(s)} \\ \boldsymbol{u} - \boldsymbol{m}^{(u)} \end{pmatrix}, \qquad \boldsymbol{D} = \begin{pmatrix} \boldsymbol{D}^{(s)} & \boldsymbol{D}^{(su)} \\ \boldsymbol{D}^{(su)\dagger} & \boldsymbol{D}^{(u)} \end{pmatrix}. \tag{3.36}$$

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

The goodness of the approximation $P \approx Q$ can be quantified by an information theoretical measure, see App. B.3.1. The Gibbs free energy of the inference problem,

$$G = \big\langle H(\boldsymbol{s}, \boldsymbol{\tau}, \boldsymbol{u}|\boldsymbol{d}) \big\rangle_Q - \big\langle -\log Q(\boldsymbol{s}, \boldsymbol{\tau}, \boldsymbol{u}|\boldsymbol{d}) \big\rangle_Q, \tag{3.37}$$

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, let us 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],

$$\begin{aligned} G &= G_0 + \left\langle H(\boldsymbol{s}, \boldsymbol{u}|\boldsymbol{d}) \right\rangle_{Q_s} - \frac{1}{2} \log\left(\det\left[\boldsymbol{D}\right]\right) \qquad (3.38)\\ &= G_1 + \boldsymbol{1}^\dagger \boldsymbol{l} - \boldsymbol{d}^\dagger \left\{ \log(\boldsymbol{l}) - \sum_{\nu=2}^{\infty} \frac{(-1)^\nu}{\nu} \left\langle \left(\boldsymbol{\lambda}/\boldsymbol{l} - 1\right)^\nu \right\rangle_{Q_s} \right\} \\ &\quad + \frac{1}{2} \boldsymbol{m}^{(s)\dagger} \boldsymbol{S}^{\star -1} \boldsymbol{m}^{(s)} + \frac{1}{2} \mathrm{tr}\left[\boldsymbol{D}^{(s)} \boldsymbol{S}^{\star -1}\right] \qquad (3.39)\\ &\quad + (\boldsymbol{\beta} - \boldsymbol{1})^\dagger \boldsymbol{m}^{(u)} + \boldsymbol{\eta}^\dagger \mathrm{e}^{-\boldsymbol{m}^{(u)} + \frac{1}{2}\widehat{\boldsymbol{D}}^{(u)}} \\ &\quad - \frac{1}{2} \log\left(\det\left[\boldsymbol{D}\right]\right), \end{aligned}$$

with

$$\begin{aligned} \boldsymbol{\lambda} &= \boldsymbol{R}\left(\mathrm{e}^{\boldsymbol{s}} + \mathrm{e}^{\boldsymbol{u}}\right), \qquad (3.40)\\ \boldsymbol{l} &= \left\langle \boldsymbol{\lambda} \right\rangle_{Q_s} = \boldsymbol{R}\left(\mathrm{e}^{\boldsymbol{m}^{(s)} + \frac{1}{2}\widehat{\boldsymbol{D}}^{(s)}} + \mathrm{e}^{\boldsymbol{m}^{(u)} + \frac{1}{2}\widehat{\boldsymbol{D}}^{(u)}}\right), \qquad (3.41)\\ \boldsymbol{S}^\star &= \sum_k \mathrm{e}^{\tau_k^\star} \boldsymbol{S}_k, \qquad \text{and} \qquad (3.42)\\ \widehat{\boldsymbol{D}} &= \mathrm{diag}\left[\boldsymbol{D}\right]. \qquad (3.43) \end{aligned}$$

Here, G_0 and G_1 carry all terms independent of $\boldsymbol{s}$ and $\boldsymbol{u}$. In comparison to the Hamiltonian given in Eq. (3.27), 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. (3.30) and (3.41) 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. (3.39) vanishes. Since this sum includes powers of $\left\langle \boldsymbol{\lambda}^{\nu>2} \right\rangle_{Q_s}$ its evaluation would require all entries of $\boldsymbol{D}$ to be known explicitly. In order to keep the algorithm computationally feasible, this sum shall furthermore 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.

Without these second order terms, the Gibbs free energy reads

$$\begin{aligned} G = G_1 &+ \boldsymbol{1}^\dagger \boldsymbol{l} - \boldsymbol{d}^\dagger \log(\boldsymbol{l}) \\ &+ \frac{1}{2} \boldsymbol{m}^{(s)\dagger} \boldsymbol{S}^{\star -1} \boldsymbol{m}^{(s)} + \frac{1}{2} \mathrm{tr}\left[\boldsymbol{D}^{(s)} \boldsymbol{S}^{\star -1}\right] \\ &+ (\boldsymbol{\beta} - \boldsymbol{1})^\dagger \boldsymbol{m}^{(u)} + \boldsymbol{\eta}^\dagger \mathrm{e}^{-\boldsymbol{m}^{(u)} + \frac{1}{2}\widehat{\boldsymbol{D}}^{(u)}} \\ &- \frac{1}{2} \log\left(\det\left[\boldsymbol{D}^{(s)}\right]\right) - \frac{1}{2} \log\left(\det\left[\boldsymbol{D}^{(u)}\right]\right). \end{aligned} \qquad (3.44)$$

---

[10] The second likelihood term in Eq. (3.39), $\boldsymbol{d}^\dagger \log(\boldsymbol{\lambda})$, 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 $x \approx \langle x \rangle$.

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,

$$\frac{\partial G}{\partial \boldsymbol{m}^{(s)}} = \boldsymbol{0} = (\boldsymbol{1} - \boldsymbol{d}/\boldsymbol{l})^{\dagger}\, \boldsymbol{R} * \mathrm{e}^{\boldsymbol{m}^{(s)} + \frac{1}{2}\widehat{\boldsymbol{D}}^{(s)}} + \boldsymbol{S}^{\star-1}\boldsymbol{m}^{(s)}, \tag{3.45}$$

$$\frac{\partial G}{\partial \boldsymbol{m}^{(u)}} = \boldsymbol{0} = (\boldsymbol{1} - \boldsymbol{d}/\boldsymbol{l})^{\dagger}\, \boldsymbol{R} * \mathrm{e}^{\boldsymbol{m}^{(u)} + \frac{1}{2}\widehat{\boldsymbol{D}}^{(u)}} + \boldsymbol{\beta} - \boldsymbol{1} - \boldsymbol{\eta} * \mathrm{e}^{-\boldsymbol{m}^{(u)} + \frac{1}{2}\widehat{\boldsymbol{D}}^{(u)}}, \tag{3.46}$$

This filter formulas again account for the uncertainty of the mean estimates in comparison to the MAP filter formulas in Eq. (3.28) and (3.29). The uncertainty covariances can be constructed by either taking the second derivatives,

$$\frac{\partial^2 G}{\partial \boldsymbol{m}^{(s)} \partial \boldsymbol{m}^{(s)\dagger}} \approx \boldsymbol{D}^{(s)^{-1}}, \qquad \frac{\partial^2 G}{\partial \boldsymbol{m}^{(u)} \partial \boldsymbol{m}^{(u)\dagger}} \approx \boldsymbol{D}^{(u)^{-1}}, \tag{3.47}$$

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

$$\frac{\partial G}{\partial D^{(s)}_{xy}} = 0, \qquad \frac{\partial G}{\partial D^{(u)}_{xy}} = 0. \tag{3.48}$$

The closed form of $\boldsymbol{D}^{(s)}$ and $\boldsymbol{D}^{(u)}$ is given explicitly in App. B.2.

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. (3.34) 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 App. B.3.2. The subsequent App. B.3.3 discusses the derivation of an solution for $\boldsymbol{\tau}$ by extremizing Q_{τ}. This result, which was also derived in Oppermann et al. (2012b), applies to the inference problem discussed here, yielding

$$\mathrm{e}^{\boldsymbol{\tau}^{\star}} = \frac{\boldsymbol{q} + \frac{1}{2}\left(\mathrm{tr}\left[\left(\boldsymbol{m}^{(s)}\boldsymbol{m}^{(s)\dagger} + \boldsymbol{D}^{(s)}\right)\boldsymbol{S}_k^{-1}\right]\right)_k}{\boldsymbol{\gamma} + \boldsymbol{T}\boldsymbol{\tau}^{\star}}. \tag{3.49}$$

Again, this solution includes a correction term in comparison to the MAP solution in Eq. (3.33). Since $\boldsymbol{D}^{(s)}$ is positive definite, it contributes positive to the (logarithmic) power spectrum, and therefore reduces the possible perception threshold further.

Notice 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 $\left\langle \boldsymbol{S}^{-1} \right\rangle_{Q_{\tau}}$ 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.

3.3.3 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,

$$\left\langle \boldsymbol{\rho}^{(\cdot)} \right\rangle_P \overset{\text{MAP-}\delta}{\approx} \left\langle \boldsymbol{\rho}^{(\cdot)} \right\rangle_\delta = \rho_0 \mathrm{e}^{\boldsymbol{m}^{(\cdot)}_{\text{mode}}}, \tag{3.50}$$

$$\overset{\text{MAP-}\mathcal{G}}{\approx} \left\langle \boldsymbol{\rho}^{(\cdot)} \right\rangle_\mathcal{G} = \rho_0 \mathrm{e}^{\boldsymbol{m}^{(\cdot)}_{\text{mode}} + \frac{1}{2}\widehat{\boldsymbol{D}}^{(\cdot)}_{\text{mode}}}, \tag{3.51}$$

$$\overset{\text{Gibbs}}{\approx} \left\langle \boldsymbol{\rho}^{(\cdot)} \right\rangle_Q = \rho_0 \mathrm{e}^{\boldsymbol{m}^{(\cdot)}_{\text{mean}} + \frac{1}{2}\widehat{\boldsymbol{D}}^{(\cdot)}_{\text{mean}}}, \tag{3.52}$$

in accordance with Eq. (3.24), (3.25), or (3.34), 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 due to 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,

$$\left\langle \boldsymbol{\rho}^{(\cdot)2} \right\rangle_P - \left\langle \boldsymbol{\rho}^{(\cdot)} \right\rangle_P^2 \overset{\text{MAP}}{\approx} \left\langle \boldsymbol{\rho}^{(\cdot)} \right\rangle_\mathcal{G}^2 \left(\mathrm{e}^{\widehat{\boldsymbol{D}}^{(\cdot)}_{\text{mode}}} - 1 \right), \tag{3.53}$$

$$\overset{\text{Gibbs}}{\approx} \left\langle \boldsymbol{\rho}^{(\cdot)} \right\rangle_Q^2 \left(\mathrm{e}^{\widehat{\boldsymbol{D}}^{(\cdot)}_{\text{mean}}} - 1 \right), \tag{3.54}$$

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

3.3.4 Imaging Algorithm

The problem of denoising, deconvolving, and decomposing photon observations is a non-trivial task. Therefore, this section discusses the implementation of the D^{3}PO algorithm given the two sets of filter formulas derived in Sec. 3.3.1 and 3.3.2, respectively.

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.

---

[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 D^{3}PO 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 due to 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., $m^{(s)}_x = m^{(u)}_x = 0$, $D^{(s)}_{xy} = D^{(u)}_{xy} = \delta_{xy}$, and $\tau^\star_k = \log(k^{-2})$. – 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. (3.29) or (3.46), 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. (3.27) or (3.39) neglecting terms independent of $\boldsymbol{m}^{(u)}$, and the gradient according to Eq. (3.29) or (3.46). A more sophisticated minimization scheme, such as a non-linear conjugate gradient (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. (3.32) or (3.47), 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 damping 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. (3.27) or (3.39) neglecting terms independent of $\boldsymbol{m}^{(s)}$, and the gradient according to Eq. (3.28) or (3.45), 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. (3.33) or (3.49). 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 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 D^{3}PO 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 raise 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 D^{3}PO algorithm will be made public in the future under `http://www.mpa-garching.mpg.de/ift/d3po/`.

3.4 Numerical Application

Exceeding the simple 1D scenario illustrated in Fig. 3.1, the D^{3}PO algorithm is now applied to a realistic, but simulated, data set. The data set represents a high energy observation

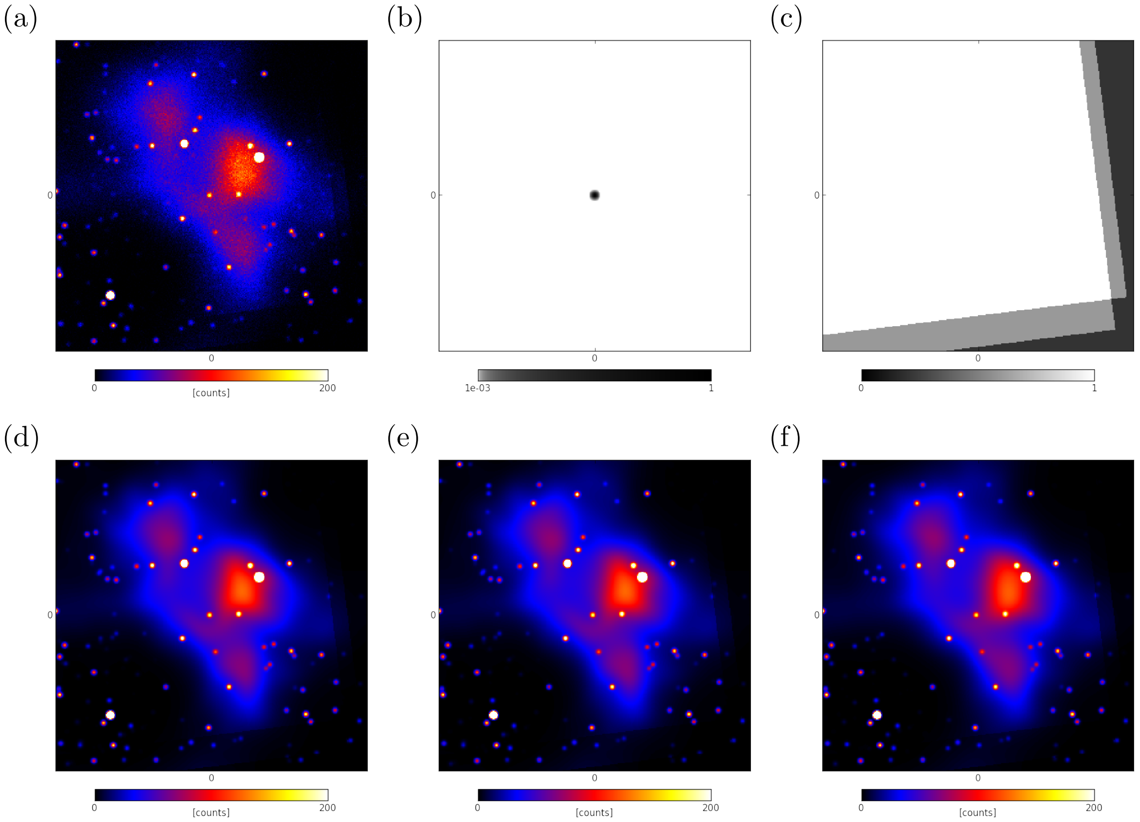


Figure 3.3: Illustration of the data and noiseless, but reconvolved, signal responses of the reconstructions. Panel (a) shows the data from a mock observation of a $32 \times 32\,\mathrm{arcmin}^2$ patch of the sky with a resolution of 0.1 arcmin corresponding to a total of 102 400 pixels. The data had been convolved with a Gaussian-like PSF (FWHM ≈ 0.2 arcmin $= 2$ pixels, finite support of 1.1 arcmin $= 11$ pixels) and masked due to 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 $\eta = 10^{-4}$.

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 0.2 arcmin, and an uneven survey mask due to the inhomogeneous exposure of the virtual instrument. The data image and those characteristics are shown in Fig. 3.3. In addition, the top panels of Fig. 3.3 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. Notice that only the quality 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. 3.4 for all three estimators, cf. Eqs. (3.50) to (3.52). There, all point-like contributions as well as noise and instrumental effects have been removed presenting a denoised, deconvolved and decomposed

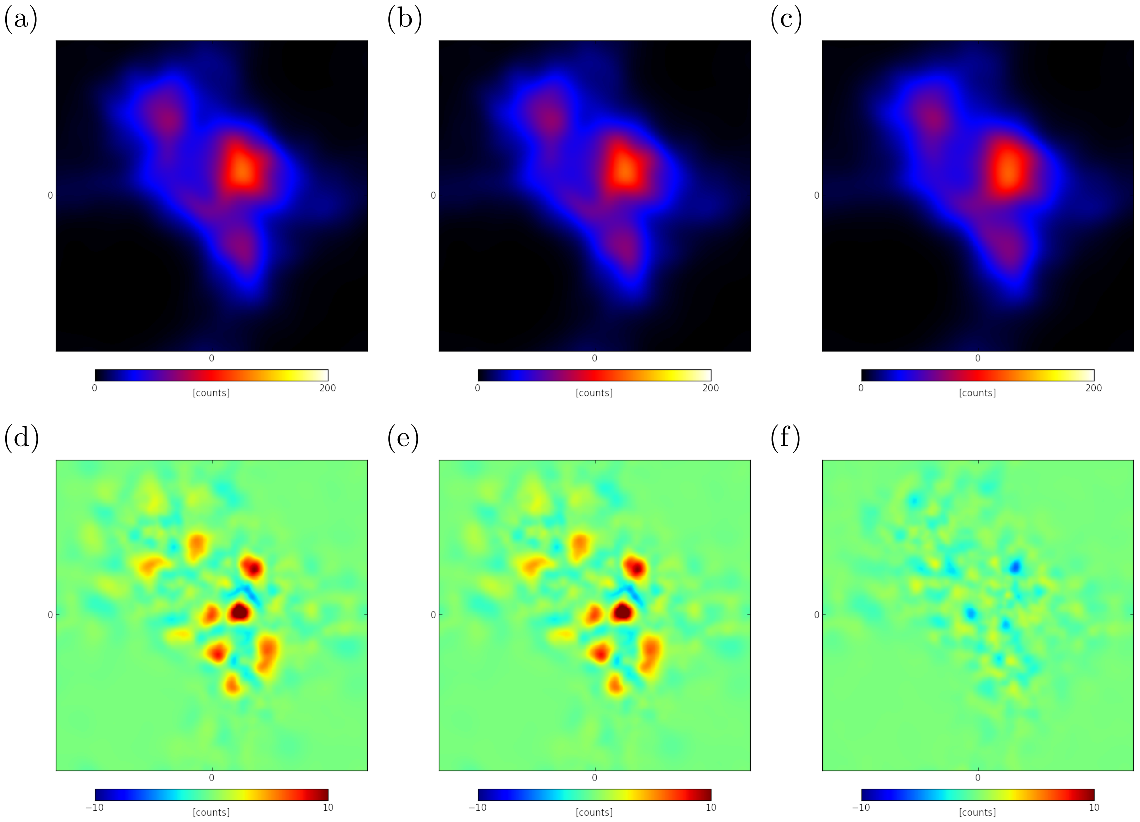


Figure 3.4: Illustration of the diffuse reconstruction. The top panels show the denoised and deconvolved diffuse contribution $\langle \boldsymbol{\rho}^{(s)} \rangle / \rho_0$ reconstructed using a different approach each, namely (d) MAP-δ, (e) MAP-$\mathcal{G}$, and (f) Gibbs. The bottom panels (d) to (f) show the difference between the originally simulated signal and the respective reconstruction.

reconstruction result for the diffuse photon flux. Figure 3.4 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,

$$\epsilon^{(s)} = \left| \boldsymbol{\rho}^{(s)} - \left\langle \boldsymbol{\rho}^{(s)} \right\rangle \right|_2 \left| \boldsymbol{\rho}^{(s)} \right|_2^{-1} , \tag{3.55}$$

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,

$$\epsilon^{(u)} = \int_1^N \mathrm{d}n \, \left| \boldsymbol{R}_{\mathrm{PSF}}^n \boldsymbol{\rho}^{(u)} - \boldsymbol{R}_{\mathrm{PSF}}^n \left\langle \boldsymbol{\rho}^{(u)} \right\rangle \right|_2 \left| \boldsymbol{R}_{\mathrm{PSF}}^n \boldsymbol{\rho}^{(u)} \right|_2^{-1} , \tag{3.56}$$

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 Tab. 3.1. 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

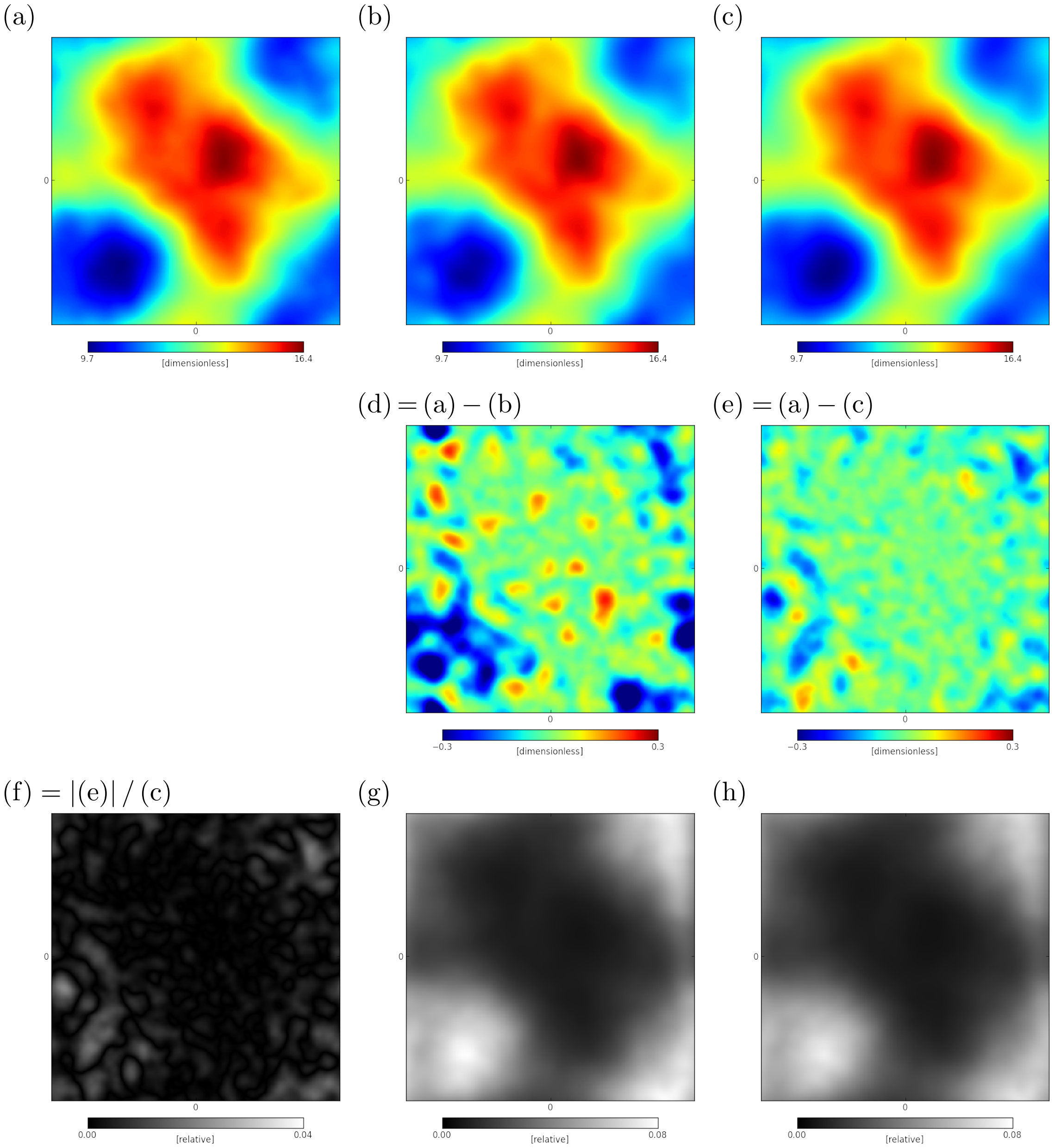


Figure 3.5: Illustration of the reconstruction of the diffuse signal field $\boldsymbol{s} = \log \boldsymbol{\rho}^{(s)}$ and its uncertainty. The top panels show diffuse signal fields. Panel (a) shows the original simulation $\boldsymbol{s}$, panel (b) the reconstruction $\boldsymbol{m}^{(s)}_{\mathrm{mode}}$ using a MAP approach, and panel (c) the reconstruction $\boldsymbol{m}^{(s)}_{\mathrm{mean}}$ 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.

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).

Figure 3.5 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

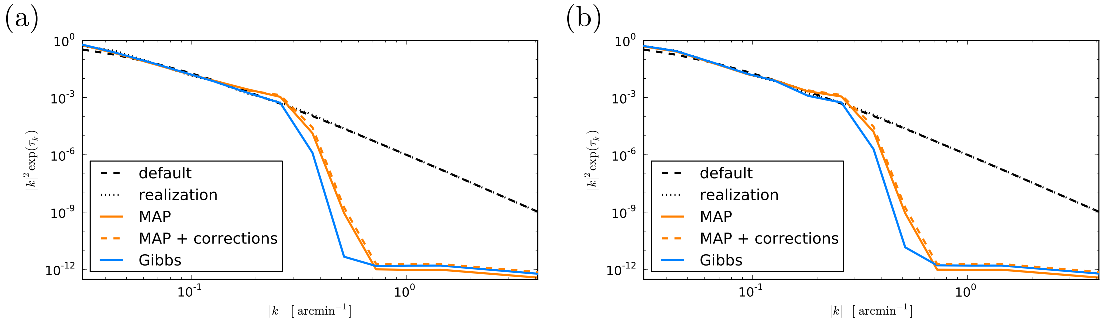


Figure 3.6: Illustration of the reconstruction of the logarithmic power spectrum $\boldsymbol{\tau}$. Both panels show the default power spectrum (black dashed line), and the simulated realization (black dotted line), as well as the reconstructed power spectra using a MAP (orange solid line), plus second 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.

found in the areas with low amplitudes. With regard to the exponential ansatz in Eq. (3.4), 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 deviations 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. 3.6, gives further indications of the reconstruction quality of the diffuse component. The simulation used a default power spectrum of

$$\exp(\tau_k) = 42\,(k+1)^{-7}. \tag{3.57}$$

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

From Fig. 3.6 it is apparent that the reconstructed power spectra track the original well up to a harmonic mode k of roughly 0.4 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 arcmin^{-1} corresponds to a physical wavelength of roughly 2.5 arcmin, and thus (half-phase) fluctuations on a spatial distances below 1.25 arcmin. The Gaussian-like PSF of the virtual observatory has a finite support of 1.1 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

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

Table 3.1: Overview of the relative residual errors of the photon flux reconstructions for the respective approaches, all using the same model parameters, cf. text.

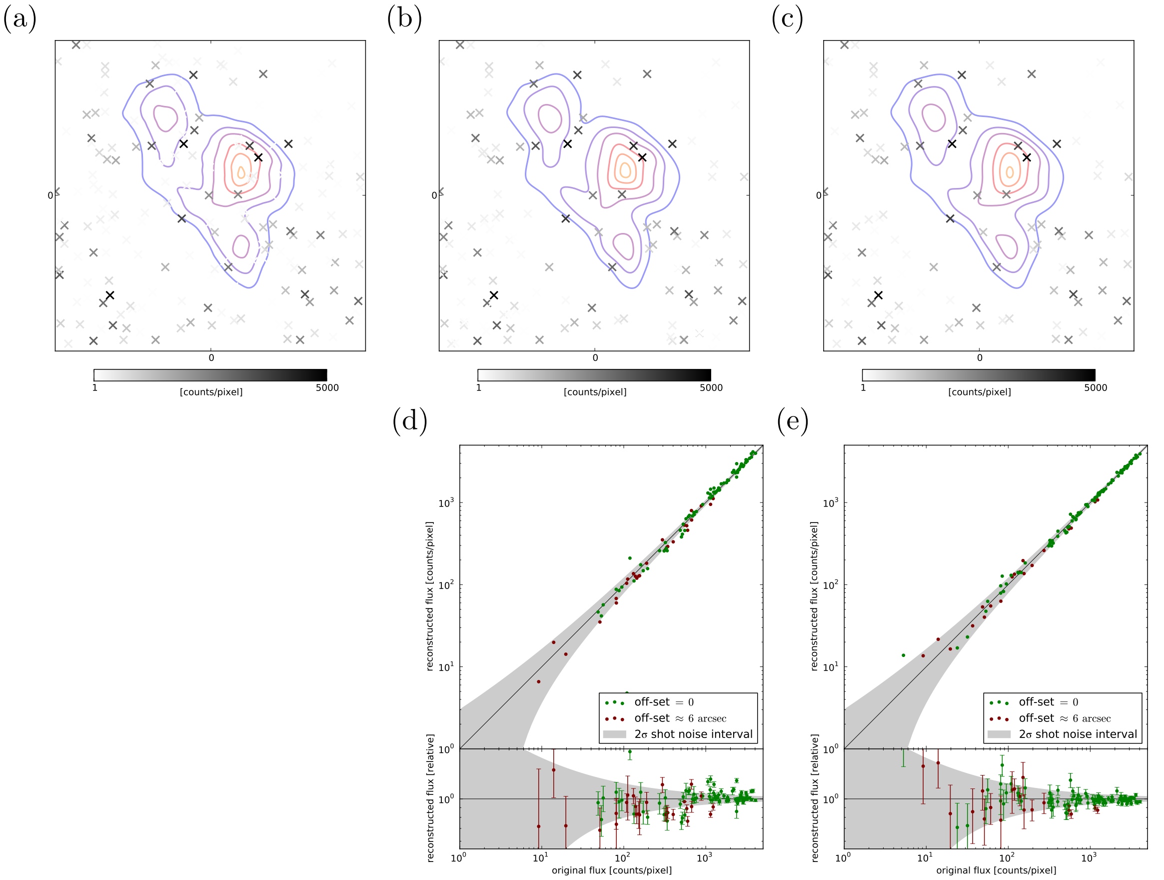


Figure 3.7: Illustration of the reconstruction of the point-like signal field $\boldsymbol{u} = \log \boldsymbol{\rho}^{(u)}$ and its uncertainty. The top panels show the location (markers) and intensity (gray scale) of the point-like photon fluxes, underlaid is the respective diffuse contribution (contours) to guide the eye, cf. Fig. 3.4. 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).

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 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. (3.33) and (3.49) 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 $0.7\,\mathrm{arcmin}^{-1}$. In turn, the MAP solution tends to overfit by absorbing some noise power into

$\boldsymbol{m}^{(s)}$ as discussed in Sec. 3.3. Thus, the higher MAP power spectrum in Fig. 3.6 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. 3.6. 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. 3.7. 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 $\pm 0.1\,\mathrm{arcmin}$, 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 for MAP solutions. Moreover, neither reconstruction shows a bias towards higher or lower fluxes.

The uncertainty estimates for the point-like photon flux $\boldsymbol{\rho}^{(u)}$ obtained from $\boldsymbol{D}^{(u)}$ according to Eqs. (3.53) and (3.54) are, in general, consistent with the deviations from the original and the shot noise uncertainty, cf. Fig. 3.7. 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. (3.32) or (3.47) 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. (3.32) or (3.47), 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. 3.5 and 3.7 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, Tab. B.1 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 difficile 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. Notice 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 D^{3}PO algorithm is capable of denoising, deconvolving and decomposing photon observations by reconstructing the diffuse and point-like signal field, and the loga-

---

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

rithmic 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.

Due to 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.

3.5 Conclusions & Summary

The D^{3}PO 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 D^{3}PO algorithm can provide *a posteriori* uncertainty information on the reconstructed signal fields. With these capabilities, D^{3}PO 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. 2012b). 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 L^2-residuals) due 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 D^{3}PO 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 $32 \times 32\,\mathrm{arcmin}^2$ patch of a simulated sky with a $0.1\,\mathrm{arcmin}$ resolution has been analyzed. The D^{3}PO 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 D^{3}PO 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 D^{3}PO code is planned.

Chapter 4

The Denoised, Deconvolved, and Decomposed Fermi γ-ray Sky

An Application of the D^{3}PO Algorithm

Note: This chapter, as well as App. C, has been submitted for publication in Astronomy & Astrophysics (Selig et al. 2014).

4.1 Introduction

Since August 2008 the Fermi Gamma-ray Space Telescope has observed the γ-ray sky with its main instrument, the Large Area Telescope (LAT) (Atwood et al. 2009), which is sensitive to photons with energies ranging from around 20 MeV to above 300 GeV.

There is a diversity of astrophysical contributions to the total γ-ray flux. Most of the photons in the GeV-range are induced by cosmic rays (CRs), charged particles moving at (ultra-)relativistic speeds, through hadronic interactions of CR nuclei with the interstellar medium (ISM) or inverse Compton scattering (IC) of electrons with background light (Ackermann et al. 2012d; Dermer et al. 2013). In addition, there is emission from an isotropic diffuse background, which is commonly denoted as "extragalactic" background (Dermer 2007, and references therein), and from sources that appear point-like.

The diffuse and point-like γ-ray fluxes appear superimposed to an observer. An observation through an instrument, like the Fermi LAT for example, additionally convolves the total flux with the instrument response functions (IRFs). The gathered data are, lastly, subject to noise; i.e., Poissonian shot noise in the case of integer photon counts. In order to retrieve the physical photon flux from observations, we would need to reverse those processes. Unfortunately, neither a direct inversion of the convolution nor the separation of noise and signal components is feasible exactly, so that we have to resort to alternative approaches.

One possibility is "forward" modeling, whereby parametrized models of different emission components are fit to the data; e.g., by a maximum-likelihood procedure as suggested by Ackermann et al. (2008) for the analysis of Fermi LAT data. By inspection of residuals between the data and the best fitting model(s), new features might be discovered. A famous example are the *Giant Fermi Bubbles* revealed by Su et al. (2010) using templates. Such templates are commonly constructed in accordance with surveys at lower energies or by modeling of the relevant CR physics (Ackermann et al. 2008; Su et al. 2010; Su & Finkbeiner 2012; Ackermann et al. 2012d, 2014b, and references therein).

In this work, we investigate the "backward" reconstruction of flux contributions using Bayesian inference methods (Bayes 1763; Cox 1946; Shannon 1948; Wiener 1949). The idea is to obtain signal estimates (and uncertainties) from a probabilistic model that denoises, deconvolves, and decomposes the input data. This probabilistic model includes prior constraints to remedy the complexity of the inverse problem. Assuming a sparsity based regularization, for example, Schmitt et al. (2010, 2012) proposed an analysis strategy using waveforms, which they applied to simulated Fermi data. For the analysis of X-ray images, which pose the same challenges as γ-ray images, a Bayesian background-source separation technique was proposed by Guglielmetti et al. (2009).

We deploy the D^{3}PO inference algorithm (Selig & Enßlin 2013) derived within the framework of information field theory (IFT) (Enßlin et al. 2009; Enßlin 2013, 2014). It simultaneously provides non-parametric estimates for the diffuse and the point-like photon flux given a photon count map. This challenging inverse problem is thereby regularized by prior assumptions that provide a statistical description of the morphologically different components; i.e., the priors define our naive understanding of "diffuse" and "point-like". D^{3}PO considers Poissonian shot noise, without Gaussian approximations, and takes the provided IRFs of the Fermi LAT fully into account. Furthermore, we can retrieve uncertainty information on the estimates.

All this allows us to present a continuous reconstruction of the diffuse γ-ray sky up to around 300 GeV, as well as an estimate of the point-like contribution, from which we derive a preliminary source catalog. By analyzing the spectral behavior of the diffuse component, it is possible to investigate the underlying processes, especially with regard to the CRs responsible for the emission.

The remainder of this chapter is structured as follows. Sec. 4.2 summarizes the analysis procedure, a more detailed description is gven in App. C.1. We present and discuss our findings in Sec. 4.3, and conclude in Sec. 4.4.

band	E^{min} [GeV]	E^{mid} [GeV]	E^{max} [GeV]
1	0.60	0.85	1.20
2	1.20	1.70	2.40
3	2.40	3.40	4.80
4	4.80	6.79	9.60
5	9.60	13.58	19.20
6	19.20	27.15	38.40
7	38.40	54.31	76.80
8	76.80	108.61	153.60
9	153.60	217.22	307.20

Table 4.1: Overview of the energy binning. Listed are minimum, logarithmic mean, and maximum energy for each band.

4.2 Analysis Summary

We analyze the photon count data collected by the Fermi LAT within its 5.5 years of operation. All data selection criteria, the working principle of the applied D^{3}PO inference algorithm (Selig & Enßlin 2013), and a description of the analysis procedure are detailed in App. C.1.

In summary, we make use of the reprocessed Pass 7 available within the Fermi Science Tools in order to retrieve the data, as well as the corresponding instrument response functions and exposure of the Fermi LAT (Atwood et al. 2009; Abdo et al. 2009; Ackermann et al. 2012a). We consider nine logarithmically spaced energy bands ranging from 0.6 to 307.2 GeV, cf. Tab. 4.1. For each band, we spatially bin all events classified as `CLEAN` in count maps, whereby we distinguish the front or back conversion of the photon within the LAT. Throughout this work, we discretize the sky using the HEALPix scheme with $n_\text{side} = 128$, which corresponds to an angular resolution of approximately 0.46°.

This data set is the input for the D^{3}PO algorithm. In order to denoise, deconvolve and decompose the data, we suppose the data $\boldsymbol{d}$ to be the result of a Poisson process with an expectation value given by the convolved sum of the diffuse and point-like flux, $\boldsymbol{\phi}^{(s)}$ and $\boldsymbol{\phi}^{(u)}$; i.e.,

$$\boldsymbol{d} \curvearrowleft \mathcal{P}(\boldsymbol{d}|\boldsymbol{R}(\boldsymbol{\phi}^{(s)} + \boldsymbol{\phi}^{(u)})), \tag{4.1}$$

where the operator $\boldsymbol{R}$ describes the full instrument response of the LAT. Under this model assumption, and with the aid of prior regularizations, D^{3}PO computes estimates for the photon fluxes. The inference is performed iteratively until convergence and for each energy band separately. Further details regarding the inference can be found in App. C.1.

The results of this analysis – including, among others, the reconstructed fluxes, uncertainties, and the preliminary D^{3}PO Fermi (pDF) catalog of source candidates – are publicly available at `http://www.mpa-garching.mpg.de/ift/fermi/`.

4.3 Results & Discussion

4.3.1 The γ-ray Sky

The Fermi LAT has detected millions of γ-ray photons within the first 5.5 years of its mission. We can stack them in a binned all-sky count map disregarding the energy of the photons.

Figure 4.1 shows such all-sky maps for the stacked data and the result of the reconstruction. D^{3}PO denoises, deconvolves, and decomposes photon count maps yielding a reconstruction of the diffuse and point-like photon flux. A reconvolution of the reconstructed γ-ray flux with the instrument response functions demonstrates the quality of the performed denoising; i.e., the removal of Poissonian shot noise.

The fractional residual exhibits no significant structures and the mean of its absolute value is below 6%.[1] As we compare the data with a denoised reproduction, the major difference is due to shot noise. We observe a weak ringing around the Galactic plane, which is a

---

[1] If we approximate the Poissonian $\mathcal{P}(\boldsymbol{d}|\boldsymbol{\lambda})$ by a Gaussian $\mathcal{G}$ with mean $\boldsymbol{\lambda}$ and variance diag$[\boldsymbol{\lambda}]$, we can expect the mean of the absolute fractional residual to be

$$\frac{1}{N}\sum_n^N \frac{|d_n - \lambda_n|}{\lambda_n} \approx \frac{1}{N}\sum_n^N \frac{|\mathcal{G}(0,1)|}{\sqrt{\lambda_n}} \approx \frac{1}{N}\sum_n^N \sqrt{\frac{2}{\pi d_n}}.$$

For the considered data set, this computes to around 10%.

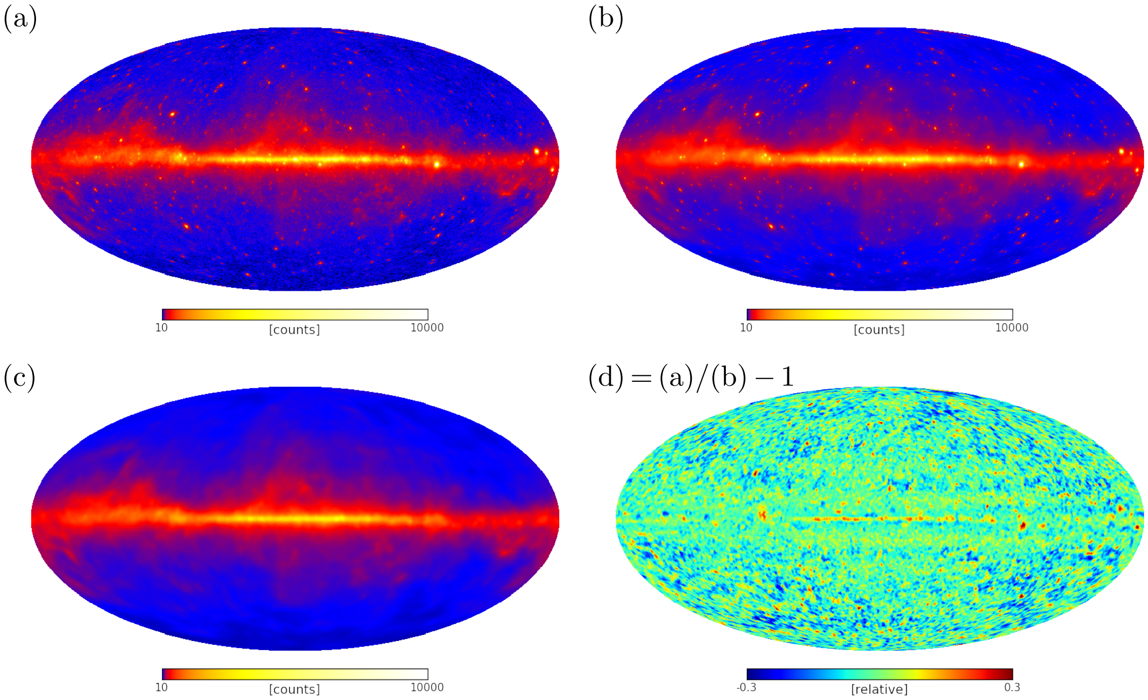


Figure 4.1: Illustration of the γ-ray sky seen by the Fermi LAT in Mollweide projection. Panel (a) shows the photon count data of 5.5 years mission elapsed time in the energy range from 0.6 to 307.2 GeV. Panel (b) shows the reconstructed (total) photon flux reconvolved with the LAT's IRFs. Panel (c) shows only the reconvolved diffuse contribution. Panel (d) shows the fractional residual map smoothed with a 0.5° Gaussian kernel.

numerical artifact due to imperfections of spherical harmonic transformations applied during the inference. Previous comparisons with best fitting templates created by the GALPROP code (Moskalenko & Strong 2000; Strong et al. 2000; Ackermann et al. 2008, 2012d, and references therein) often show significant residuals indicating features lacked by the respective models; e.g., cf. Fig. 6 in Ackermann et al. (2012d). Since our inference machinery, on the contrary, is free of *a priori* assumption regarding the existence of any Galactic or extragalactic features, significant residuals are not to be expected.

Excluding the point-like contribution to the reconvolved count map, the diffuse γ-ray sky becomes fully revealed, see Fig. 4.1c. The diffuse count map clearly displays Galactic features and substructures within the ISM. While the reconvolved photon count image appears somewhat smoothed, its deconvolved counterpart displays the Milky Way in more detail. The diffuse γ-ray fluxes in the individual energy bands are shown in Fig. 4.2. The coarseness of the images increases with energy because the number of detected photons, and thus the signal-to-noise ratio, drops drastically. The uncertainties of the reconstructions are illustrated in Fig. 4.3. Nevertheless, the Galactic disk and bulge are clearly visible at all energies.

Pseudocolor Images

In order to obtain a better view on the spectral characteristics of the γ-ray sky, we combine the maps at different energies by a pseudocolor scheme. This scheme is designed to mimic

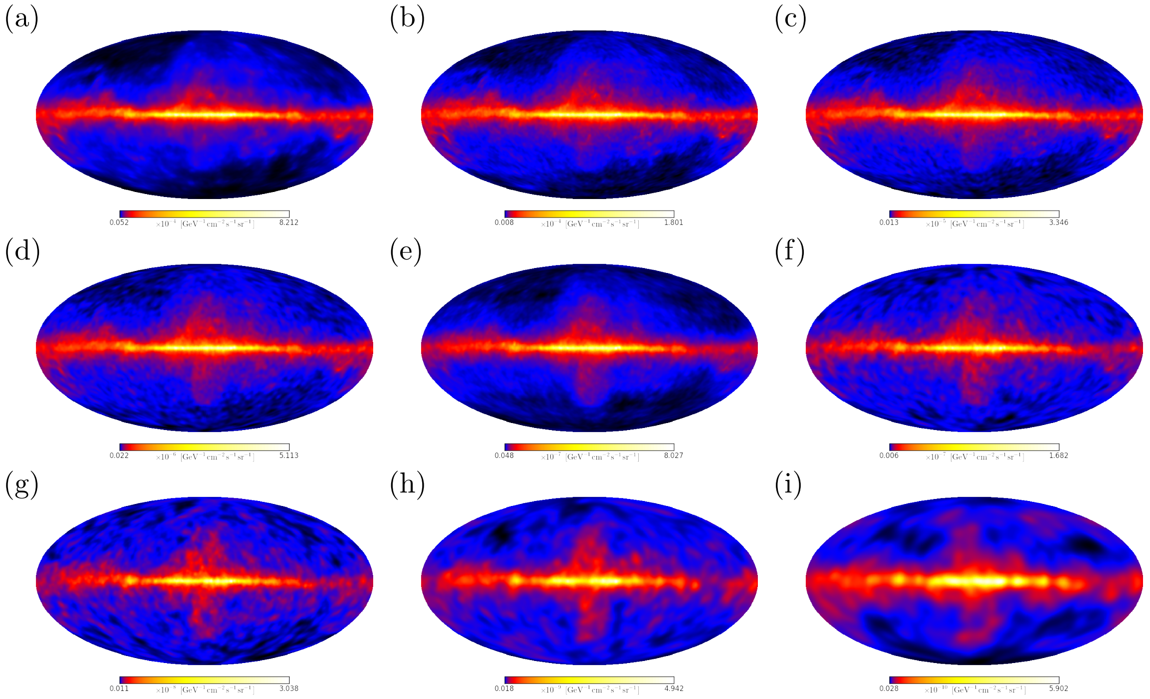


Figure 4.2: Illustration of the diffuse γ-ray flux on a logarithmic scale at different energies in Mollweide projection. The panels show the reconstructed diffuse photon flux for energy bands 1–9, cf. Tab. 4.1. The photon flux is given in units of $\mathrm{GeV}^{-1}\mathrm{cm}^{-2}\mathrm{s}^{-1}\mathrm{sr}^{-1}$. Notice that the color scale varies.

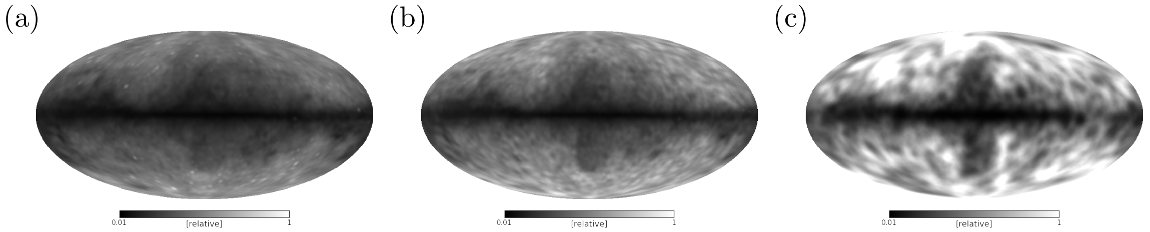


Figure 4.3: Illustration of the relative reconstruction uncertainty on a logarithmic scale at different energies in Mollweide projection. The panels show the relative uncertainty on the reconstructed diffuse photon flux at $\sim 2\,\mathrm{GeV}$, $\sim 8\,\mathrm{GeV}$, $\sim 100\,\mathrm{GeV}$ corresponding to panels (b), (d), and (h) in Fig. 4.2. The uncertainties for all energy bands are contained in the online material.

the human perception of optical light in the γ-ray range. Intensity indicates the (logarithmic) brightness of the flux, and red colors correspond to low energy γ-rays around 1 GeV and blue colors to γ-rays up to 300 GeV. The resulting pseudocolor maps of the γ-ray sky are presented in Fig. 4.4. Thanks to a suitably tuned color response, spectrally different regions can easily be identified by the human eye. At a first glance, we can recognize the bright bulge of the Milky Way, the Fermi bubbles as two greenish blue, roundish areas, and red to yellowish cloud-like structures at low and intermediate latitudes, in particular around the Galactic anticenter.

The upper panels (a–d) illustrate the functionality of the D^{3}PO inference algorithm showing the raw data and the denoised, deconvolved, and decomposed reconstruction, respectively. The denoising applies most strongly to the high energy bands, appearing green- to blueish, where the signal-to-noise ratios are worst. The deconvolution effect is most evident for point-like contributions in lower energy bands, appearing reddish, because of the increasing width of the point spread function (PSF) for these bands. Finally, the decomposition reveals the purely diffuse γ-ray sky.

This view reveals many interesting features beyond the Galactic disk and bulge, which we will discuss in the following.

Bubbles, Features, and Radio

The most striking features recovered by our reconstruction are the *Giant Fermi Bubbles* first found by Su et al. (2010). The bubbles extend up to $|b| \lesssim 50°$ in latitude and $|l| \lesssim 20°$ in longitude. They appear to emerge from the Galactic center, however, their astrophysical origin is still under discussion (Su et al. 2010; Cheng et al. 2011; Dogiel et al. 2011; Su & Finkbeiner 2012; Yang et al. 2014; Ackermann et al. 2014b, and references therein). In agreement with previous studies, we find the bubbles to have relatively sharp edges and an overall homogeneous surface brightness, appearing greenish blue in Fig. 4.4. Yang et al. (2014) report an energy dependent morphology of the southern bubble, which is, in particular, more extended to the Galactic south and west at high energies. Our results confirm this extension, as can be seen in the reconstruction for the highest energy band in Fig. 4.2.

Figure 4.4 also shows the *North Arc*, *Donut* and *Cocoon* (Su et al. 2010; Su & Finkbeiner 2012). However, we do not find evidence for a jet-like structure as reported by Su & Finkbeiner (2012).

Moreover, there is a correlation with structures seen at radio frequencies. For example, a comparison with the synchrotron map from Haslam et al. (1982) taken at 408 MHz reveals γ-ray counter parts of the *Radio Loop I* (Large et al. 1962) and smaller objects like the *Large Magellanic Cloud* at $(l, b) \approx (-80°, -30°)$, as well as the γ-ray glow around *Centaurus A* at $(l, b) \approx (-50°, 20°)$. However, the resolution of the all-sky reconstruction is too coarse to detail the morphology of such small sources. A reconstruction of a more focused field of view would be necessary to that end.

4.3.2 Energy Spectra & Spectral Indices

In order to get a more quantitative view on the different contributions to the γ-ray flux, we now investigate photon flux energy spectra.

Figure 4.5 shows the measured[2] and reconstructed energy spectra in four different regions of interest (ROI) specified in Tab. 4.2. The ROIs are typical areas in- or excluding the Galactic center or bulge. The errors are dominated by systematics; i.e., by the uncertainty in the absolute energy scale $\Delta\mathcal{E}/\mathcal{E} = (+2\%, -5\%)$ (Ackermann et al. 2012c) and in the normalization of the effective area, which is ±10% up to 100 GeV and increases linearly with $\log(E)$ to ±15% at 1 TeV (Bregeon et al. 2013)[3]. The statistical uncertainties determined

---

[2] The photon data $\boldsymbol{d}$ can be converted into flux units by division by the respective exposure ε, solid angle Ω, and width of the energy band, according to

$$\phi_j = \phi(E_j^{\mathrm{mid}}) \equiv \frac{1}{(E_j^{\mathrm{max}} - E_j^{\mathrm{min}})} \sum_{i \in \mathrm{ROI}} \frac{1}{2\Omega^{\mathrm{ROI}}} \left(\frac{d_{ij}^{\texttt{FRONT}}}{\varepsilon_{ij}^{\texttt{FRONT}}} + \frac{d_{ij}^{\texttt{BACK}}}{\varepsilon_{ij}^{\texttt{BACK}}} \right),$$

where the indices i and j label pixels and energy bands, respectively. Notice that front and back converted data are averaged accordingly.

[3] See also `http://fermi.gsfc.nasa.gov/ssc/data/analysis/LAT_caveats.html`.

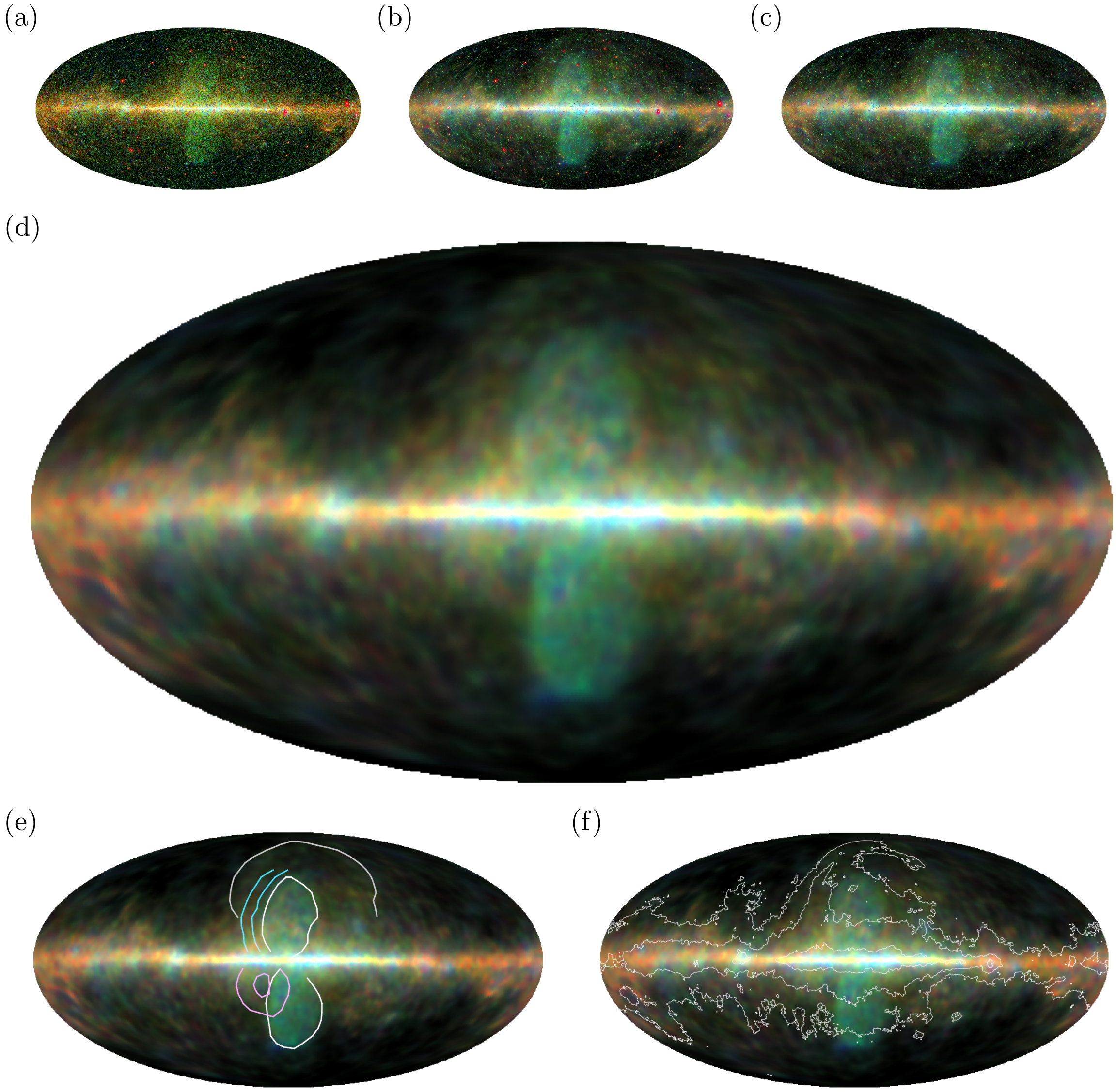


Figure 4.4: Illustration of the γ-ray sky in pseudocolor in Mollweide projection. Panel (a) shows the 5.5 year data. The panels (b) and (c) show the reconstructed (total) photon flux that in (b) is reconvolved with the IRFs. Panel (d) shows the reconstructed diffuse photon flux. The panels (e) and (f) reproduce the latter but are overlaid with the feature contours found by Su et al. (2010) (white: *Giant Fermi Bubbles*, light magenta: *Donut*, light blue: *North Arc*, light gray: *Radio Loop I*) and contours of the 408 MHz radio map from Haslam et al. (1982), respectively.

from the inference are shown in Fig. 4.3, and tend to track the signal-to-noise ratio.

We split the total energy spectrum into a diffuse and a point-like contribution, whereby we additionally distinguish between isotropic[4] and anisotropic diffuse components. Overall, the spectra from the reconstructed fluxes agree well with the data, except for the highest energy bin, where the point-like component seems to be strongly overestimated. There are two reasons for this. On the one hand, the signal-to-noise ratio is lowest, and on the other

---

[4] In this context, “isotropic” means spatially constant.

ROI	inclusion cuts	Ω [sr]	$\gamma^{(s)}$	χ^2/DOF	$\gamma^{(s)}_{\text{iso}}$	χ^2_{iso}/DOF
1	all-sky	4π	2.52 ± 0.05	0.10	2.59 ± 0.05	0.15
2	$\lvert b\rvert > 10° \vee \lvert l\rvert < 10°$	10.5	2.50 ± 0.05	0.12	2.59 ± 0.05	0.15
3	$\lvert b\rvert > 10°$	10.4	2.50 ± 0.05	0.14	2.59 ± 0.05	0.15
4	$\lvert b\rvert < 10° \wedge \lvert l\rvert < 80°$	4.5	2.51 ± 0.05	0.13	2.56 ± 0.05	0.15

Table 4.2: Overview of the regions of interest. Listed are inclusion cuts in Galactic longitude l and latitude b, covered solid angle Ω, fitted spectral indices $\gamma^{(s)}$, and χ^2 divided by the degrees of freedom (DOF); i.e., 9 bands - 2 unknowns = 7. The latter two are given for the anisotropic and the isotropic diffuse photon flux, respectively.

hand, the PSF is sharpest. Therefore, the distinction between point-like sources, noise peaks, and weak diffuse emission breaks down.

The diffuse γ-ray flux amounts to $\sim$ 90% of the total flux with the majority being anisotropic contributions of Galactic origin. Both diffuse contributions, isotropic and anisotropic, are consistent with featureless power-laws, $\phi \propto E^{-\gamma}$. The results of the power-law fits are given in Tab. 4.2.

For the anisotropic component we find a spectral index $\gamma^{(s)} = 2.50 \pm 0.05$, which is in agreement with the index of 2.44 ± 0.01 reported by Ackermann et al. (2012b). In low energy bands, the spectrum is dominated by the production and decay of π^0-mesons induced by CR protons, while IC emission becomes increasingly important at the highest energies (Ackermann et al. 2012b,d). The declining tail of the pion bump, peaking at $\frac{1}{2}m_{\pi^0} \approx 0.07\,\text{GeV}\,\text{c}^{-2}$, is visible.

The isotropic background is often referred to as "extragalactic" because it comprises unresolved extragalactic sources and might include possible signatures from large-scale structure formation or dark matter decay (Dermer 2007; Abdo et al. 2010b, and references therein). The isotropic diffuse background also follows a featureless power-law with a spectral index $\gamma^{(s)}_{\text{iso}} = 2.59 \pm 0.05$. Notice that the excess of isotropic emission around $\sim 14\,\text{GeV}$ is rather insignificant with regard to the uncertainties.[5] Abdo et al. (2010b) derive a spectral index of 2.41 ± 0.05 from 1 year Fermi LAT data in the energy range of 0.03–100 GeV. This indicates a slight spectral softening of the isotropic background towards higher energies. In the same energy range, observations with the Energetic Gamma Ray Experiment Telescope (EGRET) yield a spectral index of 2.10 ± 0.03 that is considerably smaller (Sreekumar et al. 1998). This discrepancy, which might be an instrumental issue, is not yet clarified.

The recent analysis by Ackermann et al. (2014a) investigating an energy range of 0.1–820 GeV reports the isotropic γ-ray background to be consistent with a power-law with exponential cut-off at 280 GeV having a spectral index of 2.32 ± 0.02.

If we consider smaller ROIs, we find fluctuations in the spectral index of the diffuse γ-ray flux. These fluctuations could give some indication about the CR spectrum, the composition of the local ISM, etc.

Since the D^{3}PO algorithm provides a continuous estimate of the diffuse photon flux, we can perform a spectral analysis in individual pixels. Although the energy spectra vary with location, we, for simplicity, assume a general power-law behavior everywhere, but with varying spectral index. Figure 4.6 shows the obtained spectral index map for the anisotropic γ-ray

---

[5] Such a tentative excess could be a line-like signal from dark matter decay/annihilation (Conrad 2012), however, this is highly speculative considering the available data.

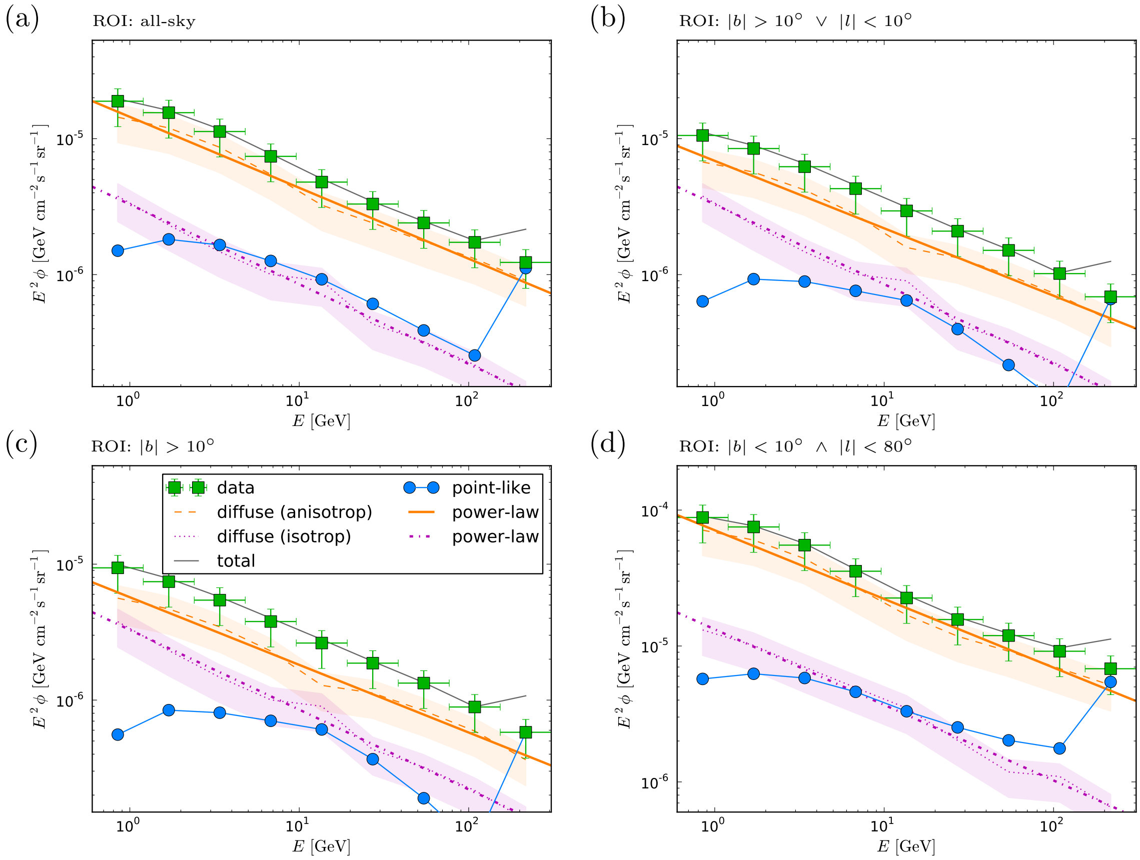


Figure 4.5: Illustration of energy spectra considering different ROIs. The legend in panel (c) applies to all other panels. Shown are the data (green squares) converted to flux units, and spectra from the reconstructed total (gray), anisotropic diffuse (dashed orange), isotropic diffuse (dotted magenta), and point-like photon flux (blue circles). Furthermore, power-law fits for the anisotropic (thick solid orange) and isotropic emission (thick dash-dotted magenta) are shown. The errors include statistical and systematic uncertainties and are only shown for data and diffuse contributions for reasons of clarity.

sky, centered on the average index of 2.5. The spatial smoothness of the spectral index map reflects that there are no discontinuities between neighboring pixels in the reconstructions.

From this spectral index map it is apparent that the Galactic disk is spectrally softer compared to the all-sky average. The same holds for the extensive structures that trace interstellar gas. These regions are dominated by hadronic interactions releasing γ-ray photons; e.g. π^0 production and decay (cf. e.g., Ackermann et al. 2012d)

In the region overlapping with the *Giant Fermi Bubbles* we find overall similar spectra that are, however, harder in contrast to the all-sky average. This is in agreement with the results of Ackermann et al. (2014b), although they found a log-parabola to fit best. The strong hardening towards the high latitude edge of the southern bubble comes from its increased spatial extent compared to lower energies, cf. Yang et al. (2014). Further local spots inside the bubble region appear insignificant within the statistical and systematic uncertainties. Although the morphology and spectra of the bubbles can be explained with hadronic and leptonic CR processes, IC scenarios give the preference of also reproducing the microwave haze observed with WMAP and Planck (Planck Collaboration 2013b; Yang et al.

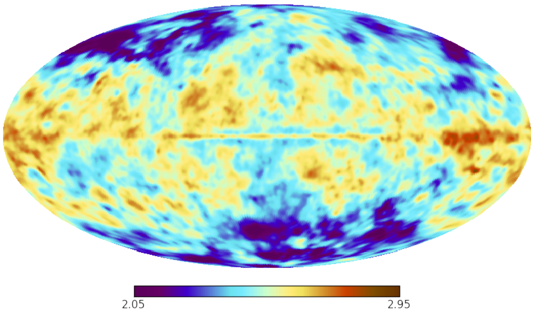


Figure 4.6: Illustration of the spectral indices of featureless power-law-fits, $\phi_{\mathrm{aniso}} \propto E^{-\gamma}$, at all positions in the sky using energy bands 1–8.

2013; Ackermann et al. 2014b). Furthermore, the low target densities at higher Galactic latitudes render the hadronic scenario as not very compelling.

4.3.3 Diffuse Emission Components

The pseudocolor scheme introduced in Sec. 4.3.1 already allows us to visually inspect the continuous reconstruction of the diffuse γ-ray sky. By eye, we can make out the Galactic bulge, the Fermi bubbles, and also cloudy structures around the Galactic anticenter.

In order to confirm this impression the visualization gives, we retrieve energy spectra from three characteristic regions: “bulge”-like ($|l| < 40°, |b| < 1.5°$), “cloud”-like ($-150° < l < -120°, |b| < 3°$), and “bubble”-like (for which we select the southern bubble up to latitudes $b < -27.5°$). The contours of those regions are shown in Fig. 4.7a.

Figure 4.8a shows the energy spectra retrieved from the three regions. The “cloud”-like spectrum is rather soft ($\gamma^{(s)} \approx 2.6$) and features the tail of the pion bump. It is not surprising that the “cloud”-like spectrum is dominated by emission from π^0 decay, because the cloudy structures trace the gas content of the ISM that provides the target protons for π^0 production. The “bubble”-like spectrum is significantly harder ($\gamma^{(s)} \approx 2.4$) indicating the dominance of hard processes like IC emission. The “bulge”-like region exhibits a spectrum that, besides having a higher absolute scale, can be described as a linear combination of the former two spectra, cf. Fig. 4.8a.

As the “bulge”-like spectrum is found to be a linear combination of “cloud”- and “bubble”-like, we can try to decompose the whole diffuse sky into those two components. For this purpose, we fit the spectrum in individual pixels by the “cloud”-, and “bubble”-like component.[6] The fit coefficients then indicate the strength of the “cloud”- or “bubble”-like contribution at all locations. Multiplying the fit coefficients with the respective spectra, we obtain a pseudocolor visualization of the “cloud”- or “bubble”-like emission components as shown in Fig. 4.7.

In spite of the simplicity of this two component model, we find a good agreement between the total diffuse emission and the sum of the two components. The relative residuals are around 5–19%, except for the highest energy band that was excluded from the fitting procedure, where the error is approximately 35%. Our findings demonstrate that the γ-ray sky

---

[6] In case the fit suggests a negative coefficient for one component, the fitting procedure is repeated ignoring this component. In this way, we ensure the positivity of the components.

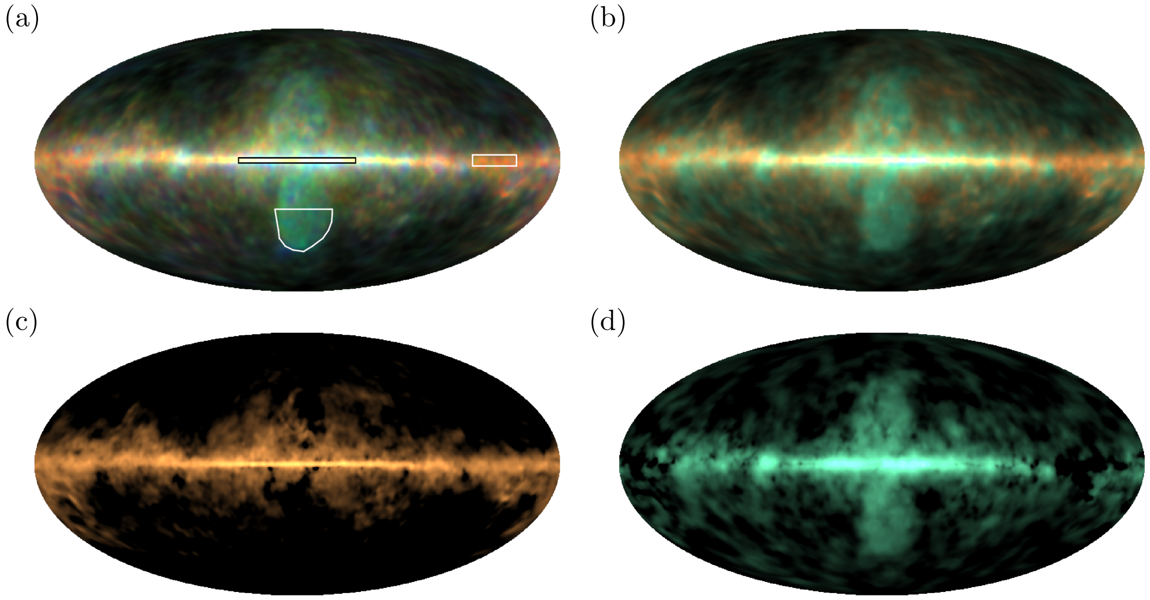


Figure 4.7: Illustration of the γ-ray sky in pseudocolor in Mollweide projection. Panel (a) reproduces Fig. 4.4d and highlights the contours defining the “bulge”-, “cloud”-, and “bubble”-like region. Panel (b) shows the linear fit of the latter two to the total diffuse flux. The Panels (c) and (d) show the “cloud”- and “bubble”-like diffuse components.

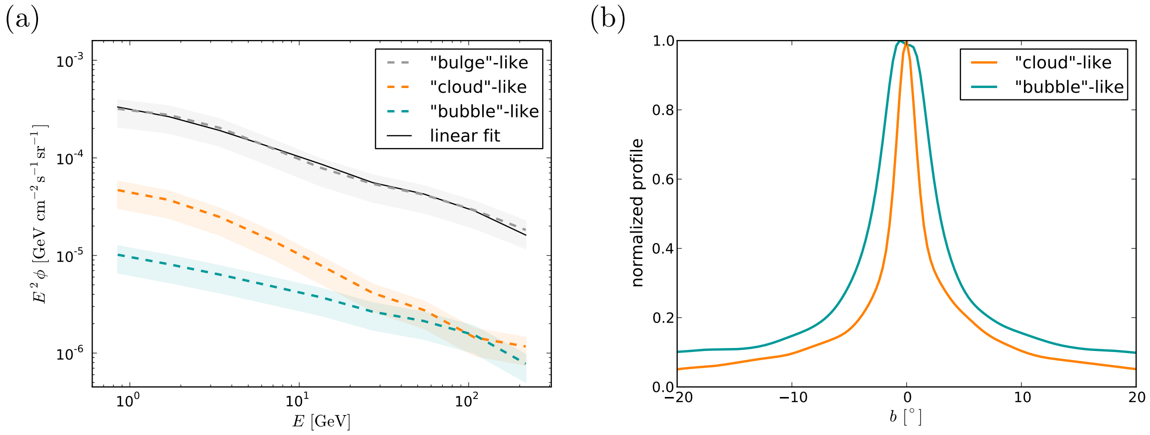


Figure 4.8: Illustration of “cloud”-, and “bubble”-like component. Panel (a) shows energy spectra from ROIs defined in Fig. 4.7a, cf. text. In addition to the spectra retrieved from the different regions, a linear combination of the “cloud”-, and “bubble”-like is fit to the “bulge”-like component, cf. legend. Panel (b) shows the normalized latitude profiles of the “cloud”-, and “bubble”-like component.

in the energy range from 0.6 to 307.2 GeV can with high precision be described by “cloud”- and “bubble”-like emission components only.

From the shape of the energy spectrum of the “cloud”-like component, we deduced that it is dominated by hadronic processes. We can also compare its morphology with other ISM tracers. For this, we compute the brightness of the “cloud”-like component and show it in Fig. 4.9a. The resulting map agrees with the thermal dust emission seen by Planck (Planck Collaboration 2013a) at 353 GHz shown in Fig. 4.9b. We like to stress how similar the morphology of thermal dust microwave/IR emission and the “cloud”-like γ-ray component are. In particular in the cloudy region in the Galactic east, the structures agree well, although we took the spectrum from a region in the west. As the thermal dust emission traces the ISM, and hence the target gas for the CR protons, we are confident that the “cloud”-like component is indeed dominated by hadronic emission processes.

The *Chamaeleon* complex, around $(l, b) \sim (-60°, -20°)$, hosting a number of star forming clouds is visible in the “cloud”-like component, cf. Fig. 4.11a. Recent work by the Planck Collaboration (2014) used γ-, radio- and dust data to map the local gas content of the clouds.

The “cloud”-like component is, however, morphologically not exactly identical to the dust emission. For example, in the latitude profile shown in Fig. 4.10, the dust-to-gamma ratio decreases with increasing latitude. This profile was computed including only pixels outside the bubble region ($|l| > 30°$), where both, the “cloud”- and “bubble”-like component, contribute, and the estimated Planck dust emission is positive. The dust seems to be preferentially in the Galactic disk compared to the thermal gas traced by γ-rays.

The morphology of the “bubble”-like component is very different, cf. Fig. 4.7. Since the spectrum of the Fermi bubbles is spatially relatively constant, the shape of the northern bubble is well recovered based on the spectrum of the southern one. We also find an excess in the “bubble”-like emission in the bulge region, as well as in the star forming region *Cygnus-*

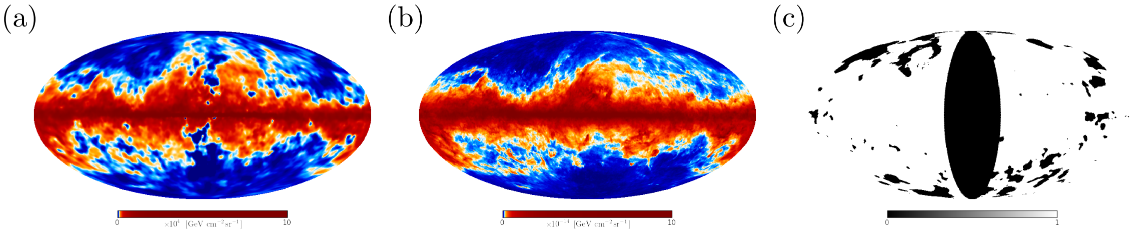


Figure 4.9: Illustration of the brightness in Mollweide view. Panel (a) shows the integrated brightness of the “cloud”-like component. Panel (b) shows the monochromatic brightness of thermal dust emission at 353 GHz (Planck Collaboration 2013a). Panel (c) shows in white the area included in computing the latitude profile in Fig. 4.10.

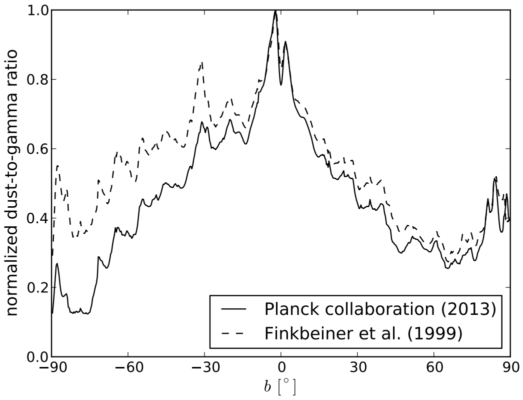


Figure 4.10: Illustration of the latitude profile of the dust-to-“cloud”-like-gamma ratio. Thermal dust emission from Planck Collaboration (2013a); Finkbeiner et al. (1999) has been smoothed with a 0.7° Gaussian kernel to match the coarseness of the γ-ray map.

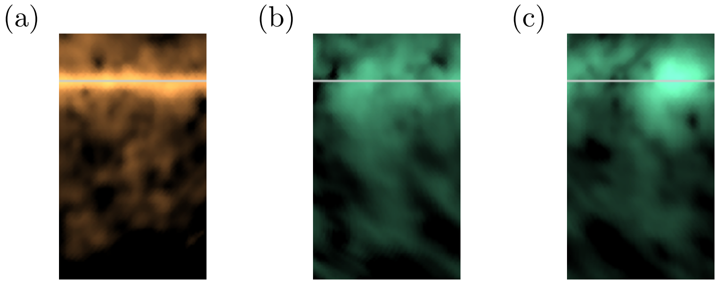


Figure 4.11: Panel (a) shows a magnification of the *Chamaeleon* complex from Fig. 4.7c. Panel (b) and (c) show magnifications of tentative outflows from Fig. 4.7d. The light gray line marks the Galactic plane ($l = 0°$).

X around $(l, b) \sim (80°, 0°)$. Excess of emission following the "bubble"-like spectrum is also visible at intermediate latitudes exhibiting mushroom-like shapes typical for hot outflows; e.g., south of the *Cygnus-X* region along $(l, b) \sim (90°, < -15°)$, or along $(l, b) \sim (130°, < -15°)$, cf. Fig. 4.11b and c. Those are likely candidates for outflows from active star forming regions of the Milky Way. Furthermore, the latitude profile shown in Fig. 4.8b indicates that the "bubble"-like disk is roughly twice as thick as the "cloud"-like component.

Since the "bubble"-like γ-ray emission is morphologically so distinct, and sets itself apart from the "cloud"-like component, we suppose that the two components are dominated by different emission processes. The "bubble"-like spectrum is distinctly harder and less structured, therefore a leptonic emission process, in particular IC scattering, seems more convincing in causing the "bubble"-like diffuse component.

The CR populations producing these two γ-ray emission components need not to be very different. It might be that we are just seeing two different phases of the ISM:

- a cold and condensed one that carries most of the Galactic dust and has a sufficient nuclei target density to be predominately revealed through hadronic interactions with CR protons. Hence, the resulting γ-ray emission mostly traces the highly structured gas distribution.

- a hot, dilute, and voluminous one that tends to flow out of the Galactic disc. The γ-ray emission from within is dominated by IC upscattering of the Galactic photon field by CR electrons. As the photon field is relatively homogeneous, the morphology of the "bubble"-like component is probably shaped by the spatial distribution of the CRs.

This simple two component model of the diffuse γ-ray emission supports scenarios in which the Fermi bubbles are just outflows of the hot ISM (Yang et al. 2013; Cheng et al. 2011; Dogiel et al. 2011; Chernyshov 2011; Carretti et al. 2013).

4.3.4 Angular Power Spectra

Under the assumption of statistical isotropy and homogeneity the second moments of a diffuse signal field are defined by its angular power spectrum. Studying power spectra gives some indication of the strength of typical fluctuations on respective angular scales described by the angular quantum number ℓ. According to our chosen HEALPix discretization, we examine spectra up to a maximum scale set by $\ell_{\mathrm{max}} = 2\, n_{\mathrm{side}} = 256$.

Figure 4.12 shows the angular power spectra of the diffuse photon flux $\boldsymbol{\phi}^{(s)}$ – the quantity we are interested in – and its logarithm, the diffuse signal field $\boldsymbol{s}$. The power spectrum, which

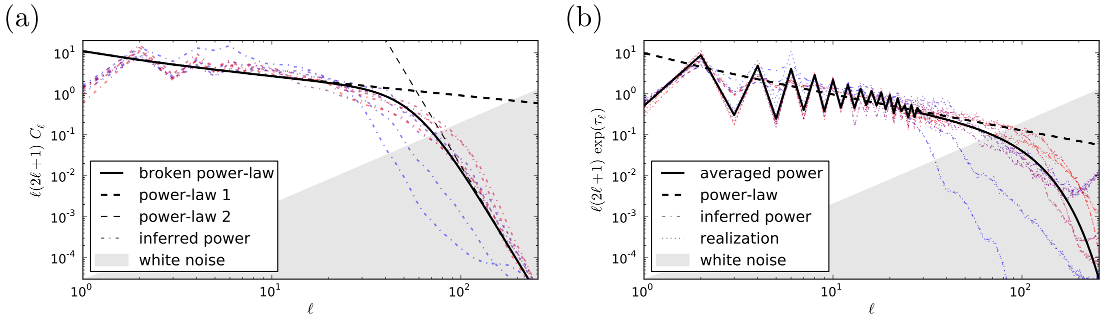


Figure 4.12: Illustration of angular power spectra over angular quantum number ℓ. Panel (a) shows the spectra of the diffuse photon flux $\boldsymbol{\phi}^{(s)}$ in each energy band and power-law fits thereof, cf. legend. Panel (b) shows the direct realization spectra of the diffuse signal, as well as the inferred, fitted and averaged spectra, cf. legend. The colors (from red over violet to blue) indicate energy (from band 1 to 9) in both panels, and a white noise spectrum with arbitrary normalization is included for comparison.

is *a priori* unknown, needs to be reconstructed from the data alongside the diffuse signal field (Wandelt et al. 2004; Jasche et al. 2010b; Enßlin & Frommert 2011; Oppermann et al. 2012b; Jasche & Wandelt 2013). This is done for each energy band separately. Further details on the inference procedure can be found in App. C.1.

The inferred power spectra of $\boldsymbol{s}$ show similar power-law behavior at all energies with some remarkable deviations. On large scales, $0 < \ell \lesssim 28$, the spectra exhibit a strong distinction between even and odd ℓ-modes. The reason for this is the dominant contribution of the Galactic disk centered around $b = 0°$ to the diffuse photon flux, which excites/suppresses even/odd ℓ-modes in the reconstruction. On smaller scales, on the other hand, the power spectra start to fall off because small-scale features cannot be resolved due to the finite exposure of the Fermi LAT. This effect has a clear energy dependence. Since events with higher energies are more rare, the decline of spectra from high energy bands begins at lower ℓ. Notice that the threshold set by the PSF is on very small scales; e.g., the 68% angular containment radius above 10 GeV is smaller than $0.2°$ corresponding to $\ell \gtrsim 900$.

Let us now consider the actual angular power spectrum C_ℓ of the diffuse photon flux $\boldsymbol{\phi}^{(s)} \propto \exp(\boldsymbol{s})$. This can be computed by transforming[7] the inferred (final) power spectra of $\boldsymbol{s}$ according to Greiner & Enßlin (2013). This transformation shifts power between different ℓ-modes, in particular towards smaller scales (larger ℓ). Again, we find a power-law behavior of the angular power spectrum, as can be seen in Fig. 4.12a. We fit a broken power-law with indices $\gamma_{\ell 1}^{(s)} = 2.45 \pm 0.01$ and $\gamma_{\ell 2}^{(s)} = 9.73 \pm 0.16$. The break point, though, is subject to the selected energy bands, and should shift to higher ℓ with increasing observation time.

4.3.5 Point Sources

Another result of our analysis of the 5.5 year Fermi data is a reconstruction of the point-like contribution to the photon flux, $\boldsymbol{\phi}^{(u)}$, which consists of Galactic and extragalactic point sources.

Figure 4.13 shows an all-sky map of all point source candidates in the pseudocolor scheme introduced in Sec. 4.3.1. Markers (and their opacity) in the map indicate the position (and

---

[7] Here, we disregard the respective monopole (mode with $\ell = 0$) for convenience. As a consequence, the absolute scale of the power spectrum of the diffuse γ-ray flux $\boldsymbol{\phi}^{(s)}$ becomes ambiguous.

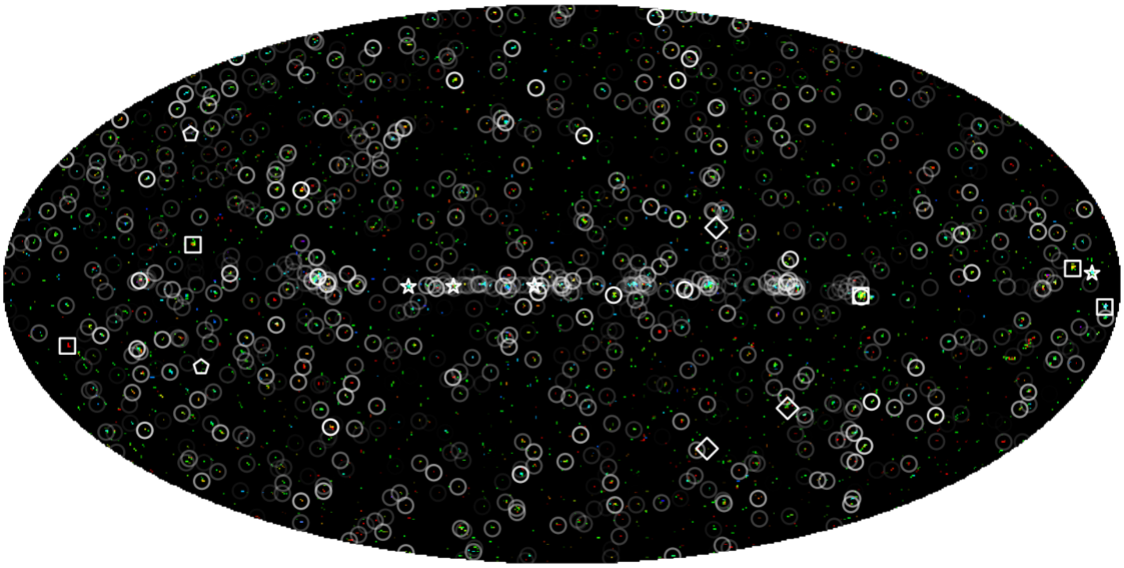


Figure 4.13: Illustration of the point sources in the γ-ray sky in pseudocolor and in Mollweide projection. Markers show point sources from the second Fermi LAT source catalog (Nolan et al. 2012) for comparison, whereby the gray scale indicates their average detection significance as listed in the catalog. Special markers show a selection of pulsars (squares), local SNRs (stars), and well-known galaxies (pentagons), as well as famous extragalactic objects (diamonds), cf. text.

candidate ID	$l[^\circ]$	$b[^\circ]$	$F^{(u)}$ $[\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}]$	contributing energy bands	dist. [°]	primary association
pDF091157	−164.88	4.27	2.8×10^{-6}	✓✓✓✓✓✓✓	0.01	2FGL J0633.9+1746
pDF009247	179.79	64.92	1.4×10^{-7}	✓✓✓✓✓✓✓✓✓	0.11	2FGL J1104.4+3812
pDF103031	−96.40	−2.76	6.0×10^{-6}	✓✓✓✓✓✓✓✓✓	0.05	2FGL J0835.3−4510
pDF108550	−175.42	−5.79	7.4×10^{-7}	✓✓✓✓✓✓✓✓✓	0.03	2FGL J0534.5+2201
pDF036698	63.60	38.89	3.5×10^{-8}	✓✓✓✓✓✓✓✓✓	0.04	2FGL J1653.9+3945
pDF029553	21.85	44.08	4.1×10^{-8}	✓✓✓✓✓✓✓✓✓	0.14	2FGL J1555.7+1111
pDF176023	17.65	−52.32	8.3×10^{-8}	✓✓✓✓✓✓✓✓✓	0.09	2FGL J2158.8−3013
pDF161378	−111.41	−39.89	3.5×10^{-8}	✓✓✓✓✓✓✓✓✓	0.04	2FGL J0449.4−4350
pDF006863	29.44	68.27	4.7×10^{-8}	✓✓✓✓✓✓✓✓✓	0.07	2FGL J1427.0+2347
pDF158586	86.09	−38.18	4.0×10^{-7}	✓✓✓✓✓✓✓✓	0.02	2FGL J2253.9+1609
⋮						

Table 4.3: Extract from the pDF catalog. Listed are candidate's ID (derived from the HEALPix index), position in Galactic longitude l and latitude b, total flux between 1–100 GeV, contributing energy bands, and distance to the associated source in the second Fermi LAT source catalog (Nolan et al. 2012). The full catalog including uncertainties is available online at `http://www.mpa-garching.mpg.de/ift/fermi/` as a FITS table.

detection significance) of point sources from the second Fermi LAT source catalog (Nolan et al. 2012).[8] There is a diversity of sources, which is why we highlight a selection by special markers. The brightest γ-ray sources are pulsars (PSRs) like *Vela* (PSR J0835-4510), *Geminga* (PSR J0633+1746), and *Crab* (PSR J0534+2200), but there are also pulsars that

---

[8] See also `http://fermi.gsfc.nasa.gov/ssc/data/access/lat/2yr_catalog/`.

have first been detected in γ-ray; e.g., LAT PSR J0007+7303 and LAT PSR J0357+3205. The Galactic disk, and especially the bulge, is clustered with point sources including, among others, supernova remnants (SNRs) like W51C, W44, W30, and IC443. There are also many homogeneously distributed extragalactic sources, for example, the starburst galaxy *Cigar* (M82) or our neighboring galaxy *Andromeda* (M31). Furthermore, the core of *Centaurus A* (NGC 5128) and the *Small* and *Large Magellanic Cloud* are visible in γ-rays.

Deriving a catalog of source candidates from the point-like flux is difficult because a point source might, for example, appear in neighboring pixels at different energies due to different noise realizations and the aforementioned energy dependence of the signal-to-noise ratio and the PSF. In order to nominate a candidate, we check if the point-like contribution exceeds 5σ above the diffuse emission in at least two of the energy bands 1–8, which is a simple but conservative criterion taking the diffuse reconstruction uncertainty σ into account, cf. Fig. 4.3. Notice that we exclude the highest energy band from our search, since the point-like flux in this band seems to be contaminated as discussed in Sec. 4.3.2. We refer to the compiled point source catalog as preliminary D^{3}PO Fermi (pDF) catalog.

Qualitatively, the point-like flux found by D^{3}PO agrees with the second Fermi LAT source catalog as shown in Fig. 4.13. A few sources appear slightly off-center or smeared out over two or more pixels. The reason for this is that such sources are positioned in-between grid points of the chosen HEALPix grid. Notice that image pixels in Fig. 4.13 do not represent HEALPix pixels.

We find 2,522 source candidates, cf. Tab. 4.3. For comparison, the 1 and 2 year Fermi LAT source catalog comprises 1,451 and 1,873 sources, respectively (Abdo et al. 2010a; Nolan et al. 2012). 1,269 of our sources can be associated with known LAT sources as the angular distance between pDF candidate and catalog source is less than the angular resolution of our reconstruction. The reason why we do not confirm all objects in the second Fermi LAT source catalog is the conservative criterion we apply. This still leaves 1,253 new source candidates to be confirmed by future work.

We caution that a more detailed study is necessary to confirm or reject those candidates. The analysis of the individual sources, best done on more constrained ROIs and with higher angular resolution, is left for future work.

In the following, we compare the pDF candidates for which we find an unique association within the second Fermi LAT source catalog (Nolan et al. 2012) by means of their spectral index and total flux. Notice that both studies are based on different data, exposure, calibration, and analysis algorithms.[9]

The spectral index of a sources' energy spectra should not (or at least not strongly) be influenced by those differences. For each source, we attempt to fit three different spectral shapes: a plain power-law,

$$\phi^{(u)}(E) = K \left(\frac{E}{E_0} \right)^{-\gamma^{(u)}}, \tag{4.2}$$

a log-parabola,

$$\phi^{(u)}(E) = K \left(\frac{E}{E_0} \right)^{-\gamma^{(u)} - \beta \log(E/E_0)}, \tag{4.3}$$

---

[9] The main differences are the selection of `SOURCE` (`CLEAN`) events, the 2 (5.5) years of observation, and the usage of the `P7_V6` (`P7REP_V15`) IRFs in the second Fermi LAT source catalog (Nolan et al. 2012) (the candidate catalog presented here).

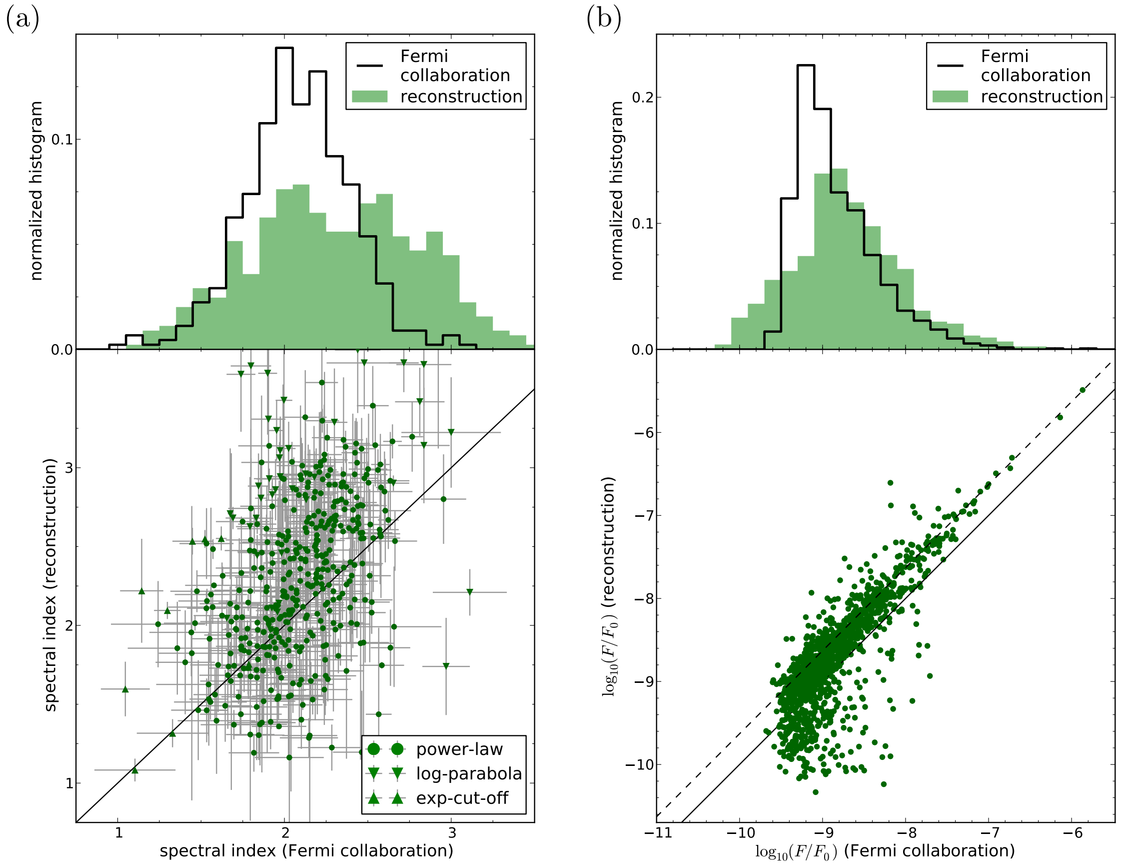


Figure 4.14: Comparison of the the second Fermi LAT source catalog (Nolan et al. 2012) and candidates from the reconstruction. Panel (a) shows the histogram and scatter plot of the spectral indices $\gamma^{(u)}$ of candidates versus catalog sources. In the latter, a 1 : 1 line (black solid) is plotted for comparison, and the markers indicate the fit spectral shape, cf. legend. Panel (b) shows the histogram and scatter plot of the logarithmic total fluxes $\log_{10}(F/F_0)$, where $F_0 = 1\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$. The scatter plot hosts an 1 : 1 line (black solid) and 1 : 2.3 line (black dashed) for comparison.

and a power-law with exponential cut-off,

$$\phi^{(u)}(E) = K \left(\frac{E}{E_0}\right)^{-\gamma^{(u)}} \exp\left(-\frac{E-E_0}{E_\mathrm{cut}}\right). \tag{4.4}$$

Here $E_0 = 1\,\mathrm{GeV}$ serves as a reference energy, and the spectral index $\gamma^{(u)}$ is a fit parameter as are K, β, and E_cut. The preliminary source catalog contains the best fit parameters for all shapes if applicable.

Figure 4.14a shows the comparison of the pDF spectral indices and the ones listed in the second Fermi LAT source catalog (corrected to E_0 if need be). The scatter of spectral indices is large, but comparable to the uncertainties. We find a rough agreement, although our distribution is broadened towards higher indices. Most of the outliers yielding a high (low) spectral index are modeled by a log-parabola (exponential cut-off) that has an additional degree of freedom compared to the plain power-law. This implies that the versatility of point source spectra might not be covered by the considered spectral shapes.

Figure 4.14b shows a comparison of the total fluxes $F^{(u)}$, defined as

$$F^{(u)} = \int \mathrm{d}\Omega \sum_{j=1}^{8} \int_{\widetilde{E}_j^{\min}}^{\widetilde{E}_j^{\max}} \mathrm{d}E \left(\frac{E}{E_j^{\mathrm{mid}}} \right)^{-\gamma^{(u)}} \phi^{(u)}(E_j^{\mathrm{mid}}) \tag{4.5}$$

with j labeling the energy bands and

$$\begin{aligned} \widetilde{E}_j^{\max} &= \min\{1\,\mathrm{GeV},\, E_j^{\min}\} \\ \widetilde{E}_j^{\min} &= \max\{E_j^{\max},\, 100\,\mathrm{GeV}\}, \end{aligned}$$

in a histogram and a scatter plot. The differences in the event selection, exposure, and calibration seem to amount to an off-set in the total flux of an factor ~ 2.3. Taking this factor into account, the fluxes show a good agreement. At the faint end, our fluxes tend to come below the fluxes reported in the second Fermi LAT source catalog (Nolan et al. 2012). Since our analysis benefits from a higher exposure and improved calibration, the fluxes from the Fermi collaboration might rather be considered as upper limits in this comparison.

Galaxy Cluster

Some galaxy clusters exhibit diffuse, extended radio emission, so-called radio halos, which proves the existence of relativistic electrons therein. If relativistic protons are present as well, γ-ray emission is to be expected due to hadronic interactions (Ackermann et al. 2010, and references therein).

Feretti et al. (2012) provide a collection of cluster hosting radio halos. We investigate the presence of γ-ray emission in the direction of those clusters, which would in our reconstruction appear point-like due to the pixelization of our reconstruction. Tab. 4.4 lists our upper limits of the total flux F^{up} and the level of the diffuse emission $F^{(s)}$ at the cluster locations. The upper limit flux is computed according to $F^{\mathrm{up}} = F^{(u)} + 2\sigma_{F^{(s)}}$, where $\sigma_{F^{(s)}}$ is the uncertainty of the total diffuse flux $F^{(s)}$.[10] This is the largest possible flux hidden under the diffuse γ-ray emission.

We find upper limit fluxes between $10^{-10}\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$ and $10^{-8}\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$ for the energy range 1–100 GeV. A few clusters contain or are in projection to known γ-ray point-sources (active galaxies), which affects the upper limits for those clusters. These are: A1758a, RXJ0107.7+5408, *Perseus* (NGC 1275 and IC310), A2029, A2390.

Ackermann et al. (2010) provide upper limits for nearby clusters above 0.1 GeV. Some clusters are in both samples, namely A2256, A2319, *Coma* (A1656), *Ophiuchus*, *Perseus* (A0426), A1914, A2029, A2142, A2163, A2744, *Bullet* (1E 0657-56), MACSJ0717.5+3745. We find comparable or slightly lower upper limit fluxes. For example, for the *Coma* (A1656) cluster Ackermann et al. (2010) report about $10^{-8}\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$ above 0.1 GeV. Assuming a spectral index of 2.5 this translates to $3 \times 10^{-10}\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$ above 1 GeV, which is slightly larger than the $1.2 \times 10^{-10}\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$ we obtain.

However, we caution that our upper limits are not strict 95% confidence intervals, as they are not a direct outcome of our inference, but estimated as described above. A closer investigation is left for future work.

---

[10] Unfortunately, D^{3}PO does not converge on an all-sky point-like uncertainty map.

4.4 Conclusions & Summary

We analyze the Fermi LAT 5.5 year photon data in the energy range from 0.6 to 307.2 GeV. Applying the D^{3}PO inference algorithm, the data is effectively denoised, deconvolved, and decomposed with Bayesian inference methods. In contrast to previous approaches by the Fermi collaboration and others, our non-parametric reconstruction does not rely on emission templates.

We obtain estimates for the diffuse and point-like contributions to the γ-ray flux. Furthermore, D^{3}PO also provides uncertainty information and the *a priori* unknown angular power spectrum of the diffuse contribution.

The inferred diffuse photon flux reveals the diversity of the γ-ray sky. We clearly reproduce the structure of the Galactic disk, bulge and local interstellar gas, all of which are primarily illuminated by photons induced by hadronic interactions of CRs with the ISM. We confirm the existence of the *Giant Fermi Bubbles*, as well as their homogeneous morphology, sharp edges, and hard spectra. Moreover, we are also able to resolve small diffuse structures; e.g., the γ-ray glow around *Centaurus A*.

The continuous reconstruction of the diffuse flux allows us to present the first spectral index map of the diffuse γ-ray sky, as well as a pseudocolor composite visualizing the spectrally different regions. Furthermore, the large-scale angular power spectrum of the diffuse emission seems to obey a pow-law with index 2.45 ± 0.01 across all energy bands.

Inspired by the pseudocolor visualization, we decompose the diffuse γ-ray sky into a “cloud”-like and “bubble”-like emission component. The former, tracing the dense, cold ISM, is dominated by hadronic emission processes, while the latter, being morphologically and spectrally distinct, seems to be dominated by leptonic processes in hot, dilute parts of the ISM and outflows thereof. In particular, our findings indicate a preference for IC emission from the Fermi bubbles and support scenarios (Yang et al. 2013; Cheng et al. 2011; Dogiel et al. 2011; Chernyshov 2011; Carretti et al. 2013) in which the Fermi bubbles are explained by hot outflows powered by strong activities in the Galactic center region. We report further, smaller outflows of a similar population of relativistic particles at other locations.

The reconstruction of the point-like photon flux qualitatively confirms most of the sources from the second Fermi LAT source catalog. Quantitatively, we derive a preliminary catalog of source candidates that comprises 2,522 point sources. A more detailed analysis of this catalog is left for future research.

Finally, we provide upper limit fluxes for galaxy clusters with known radio halos. Our limits, although also preliminary, are comparable or tighter than the upper limits from the Fermi collaboration’s study on galaxy clusters

cluster name	$F^{\mathrm{up}}[\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}]$	$F^{(s)}[\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}]$
A209	1.4×10^{-10}	1.6×10^{-10}
A399	1.8×10^{-10}	4.4×10^{-10}
A401	1.8×10^{-10}	4.4×10^{-10}
A520	2.1×10^{-10}	3.1×10^{-10}
A521	1.6×10^{-10}	3.0×10^{-10}
A523	2.0×10^{-10}	4.8×10^{-10}
A545	2.0×10^{-10}	5.1×10^{-10}
A665	1.2×10^{-10}	2.6×10^{-10}
A697	1.3×10^{-10}	2.4×10^{-10}
A746	1.2×10^{-10}	1.9×10^{-10}
A754	1.6×10^{-10}	2.7×10^{-10}
A773	1.2×10^{-10}	2.0×10^{-10}
A781	1.4×10^{-10}	1.9×10^{-10}
A851	1.3×10^{-10}	1.8×10^{-10}
A1213	1.8×10^{-10}	2.0×10^{-10}
A1300	1.9×10^{-10}	3.6×10^{-10}
A1351	1.0×10^{-10}	1.9×10^{-10}
A1656	1.2×10^{-10}	1.6×10^{-10}
A1689	1.6×10^{-10}	2.6×10^{-10}
A1758a	2.3×10^{-10}	2.0×10^{-10}
A1914	1.4×10^{-10}	1.9×10^{-10}
A1995	1.1×10^{-10}	1.7×10^{-10}
A2034	1.2×10^{-10}	1.9×10^{-10}
A2163	2.3×10^{-10}	7.5×10^{-10}
A2218	1.1×10^{-10}	2.3×10^{-10}
A2219	1.3×10^{-10}	2.1×10^{-10}
A2254	1.6×10^{-10}	3.6×10^{-10}
A2255	1.1×10^{-10}	2.3×10^{-10}
A2256	1.1×10^{-10}	2.6×10^{-10}
A2294	1.3×10^{-10}	3.6×10^{-10}
A2319	1.6×10^{-10}	4.5×10^{-10}
A2744	1.8×10^{-10}	1.7×10^{-10}
A3562	1.7×10^{-10}	3.9×10^{-10}
RXJ0107.7+5408	2.3×10^{-10}	8.9×10^{-10}
CL0016+16	1.6×10^{-10}	2.7×10^{-10}
CL0217+70	2.3×10^{-10}	1.4×10^{-9}
1E0657-56	1.5×10^{-10}	3.6×10^{-10}
MACSJ0717.5+3745	1.5×10^{-10}	3.1×10^{-10}
RXCJ1314.4-2515	1.9×10^{-10}	4.7×10^{-10}
RXCJ1514.9-1523	3.1×10^{-10}	5.1×10^{-10}
RXCJ2003.5-2323	3.1×10^{-10}	5.1×10^{-10}
CIZAJ2242.8+5301	2.2×10^{-10}	9.8×10^{-10}
Perseus	4.4×10^{-8}	7.5×10^{-10}
Ophiuchus	3.1×10^{-10}	1.7×10^{-9}
A1835	1.7×10^{-10}	2.7×10^{-10}
A2029	9.9×10^{-10}	3.5×10^{-10}
A2142	1.4×10^{-10}	2.7×10^{-10}
A2390	3.9×10^{-10}	4.3×10^{-10}
A2626	1.6×10^{-10}	3.3×10^{-10}
MRC0116+111	1.5×10^{-10}	2.6×10^{-10}
RBS797	1.0×10^{-10}	2.0×10^{-10}
RXJ1347.5-1145	1.8×10^{-10}	3.6×10^{-10}
RXCJ1504.1-0248	2.9×10^{-10}	5.1×10^{-10}

Table 4.4: Overview of total flux upper limits for clusters hosting a radio (mini-)halo from the collection by Feretti et al. (2012).

Chapter 5

X-ray Computed Tomography

5.1 Motivation

X-ray computed tomography (CT) has become a standard tool for medical diagnosing as it provides a non-invasive option for examining the interior of a subject.

During a CT scan, the emitting X-ray tube and the opposing detector are rotated around the subject. Thereby, a series of projections along the radial axis in small angular increments is performed. The intensity loss of the emitted X-rays due to interactions with matter is measured along the lines of sight between the X-ray source and each detector element. In this way, a thin tomographic slice of the subject is recorded, then the subject is moved along the cylinder axis in order to scan the next circular slice, and so on, until the whole volume to be examined is covered (Kak & Slaney 1988; Herman 2009).

The absorption of X-rays propagating through matter happens primarily through Compton scattering and (partial) ionization of atoms by inducing bound-bound transitions in its electron shell. Such interactions can damage living tissue, in particular the *Deoxyribonucleic acid* (DNA) molecules whose corruption can lead to cancer. For the subject's protection, it is therefore imperative to keep the X-ray exposure at a necessary minimum such that the image quality suffices for medical diagnoses. In practice, radiation protection is ensured by many regulations, mechanisms, and alternative options; e.g., the administration of contrast agents.

Another way of diminishing the required dose is the improvement of the imaging techniques used to reconstruct (two- or) three-dimensional images from tomographic data. The filtered back projection (FBP) is a standard, but in its basic form outdated, technique based on a Radon transformation (Kak & Slaney 1988; Hsieh 2009). It offers a practical and especially fast, heuristic approach, but its performance depends on the high quality and redundancy of the CT data.

We propose using probabilistic methods to derive an inference algorithm that properly treats noise in the data and can exploit prior knowledge. Accounting for noise statistics and remedying the ambiguity of inverse problems by prior information can significantly reduce the appearance of imaging artifacts, and thus improves the quality of the reconstructed medical images.

Furthermore, keeping the level of image quality unchanged, we might be able to process noisier data that require a smaller dose for the subject. This is the main motivation of this study.

5.2 Inference

A medical CT scan is a highly complex procedure. Here, we derive an inference model in the language of information field theory (IFT) that captures all essential aspects of the underlying inverse problem, but we also introduce a number of simplifications in order to keep the complexity of the model minimal.

5.2.1 Simplified Data Model

The raw CT data is denoted as "sinogram" and provides a logarithmic intensity ratio (Kak & Slaney 1988),

$$\text{sinogram} \equiv -a \log\left(\frac{I}{I_0}\right), \tag{5.1}$$

where a is a coefficient specific to the CT scanner, I the received, and I_0 the emitted monochromatic X-ray intensity. Since the intensities are proportional to the respective number of X-ray photons, we are free to exchange these quantities.

Following the same line of arguments as in Sect. 3.2.2, we assume the photon data $\boldsymbol{d}$ to suffer from Poissonian shot noise,

$$\boldsymbol{d} \curvearrowleft P(\boldsymbol{d}|\boldsymbol{\lambda}) = \prod_i \mathcal{P}(d_i, \lambda_i) = \prod_i \frac{1}{d_i!}\, \lambda_i^{d_i} \mathrm{e}^{-\lambda_i}, \tag{5.2}$$

where i indexes the data entries; i.e., it labels detector pixels and projections consistently. In accordance with Eq. (5.1), the expected number of photons $\lambda_i \propto I_i$ is given by

$$\lambda_i = \lambda_0 \exp\left(-a \int \mathrm{d}r_i\; \rho(x)\right). \tag{5.3}$$

Here, we introduced the integral along the line of sight (LOS) r_i over the quantity of interest $\boldsymbol{\rho} = \rho(x)$, which is defined at real space positions $x \in \mho$. The X-ray absorption inside each line increment $\mathrm{d}r_i$ depends on the present medium and its density; i.e., on the charge number and number density of the present atoms (or molecules). Therefore, $\boldsymbol{\rho}$ describes the density mixture of different media, which are air, soft tissue, bones, etc. At this point, we adopt the simplification of considering only one average medium, and understand $\boldsymbol{\rho}$ as its physical matter density. Extending this model to involve multiple components is conceivable, but increases the complexity beyond the scope of this study.

Moreover, we can rewrite the integration over all LOS as a linear operation represented by the response $\boldsymbol{R}$,

$$\int \mathrm{d}r_i\; \rho(x) = \int_{\mho} \mathrm{d}x\; R_i(x)\, \rho(x) = (\boldsymbol{R}\,\boldsymbol{\rho})_i\,, \tag{5.4}$$

and therefore

$$\boldsymbol{\lambda} = \lambda_0 \exp\left(-a\boldsymbol{R}\,\boldsymbol{\rho}\right). \tag{5.5}$$

As the geometry of the scanning strategy is fully absorbed into the response operator, any configuration can be handled with this abstract approach.

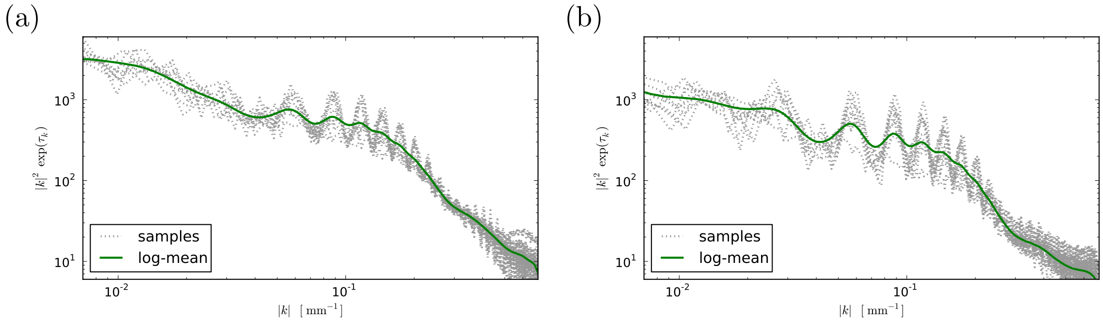


Figure 5.1: Illustration of the prior Fourier power spectra for the lung slice (a) and the hip slice (b). Both panels show the direct realization spectra of the samples (gray dotted line) and their logarithmic mean (green solid line).

Notice that we, so far, have only considered monochromatic X-rays. In reality, however, neither the emitted nor the detected photons have a single universal energy (see e.g., Alvarez & Macovski 1976; Johnson et al. 2007). We refrain from extending the data model to incorporate a dual- or multi-spectral response at this point. To explore the potential of such approaches is left for future research.

5.2.2 Prior Knowledge

The matter density $\boldsymbol{\rho}$ is supposed to be strictly positive (with zero representing a vacuum). In analogy to the discussion in Sect. 3.2.1, we enforce this positivity constraint on $\boldsymbol{\rho}$ by introducing the signal field $\boldsymbol{s} = \log \boldsymbol{\rho}$ as its logarithm. As a next step, we have to formulate our *a priori* state of knowledge on this signal.

When physicians examine medical images, it is their prior knowledge built on their education and experience that allows them to identify bones and distinguish organs from connective tissue. Acquiring such skills from a medical image data base poses a typical machine learning problem. Similarly, we can construct a prior from such a data base.

For each data set, we are given a total of 10 images showing reconstructions of similar CT slices. Assuming the signal field to be a statistically homogeneous and isotropic Gaussian random field, we are able to sufficiently constrain the signal covariance $\boldsymbol{S}$ by its Fourier power spectrum using this limited data base; i.e.,

$$\boldsymbol{s} \curvearrowleft P(\boldsymbol{s}|\boldsymbol{S}) = \mathcal{G}(\boldsymbol{s}, \boldsymbol{S}) = \frac{1}{\sqrt{\det[2\pi \boldsymbol{S}]}} \exp\left(-\frac{1}{2}\boldsymbol{s}^\dagger \boldsymbol{S}^{-1} \boldsymbol{s}\right). \tag{5.6}$$

This prior assumption is functionally adequate, though it is a rather crude approximation to reality. As discussed above, we know that there are distinct media that absorb X-rays differently. Metals and bones are strong absorbers, while soft tissue absorbs less X-rays, and air almost none. Introducing a multi-component signal model, which would allow us to construct more realistic priors, is not conceptionally different, though. Moreover, we could drop the assumption of statistical isotropy and homogeneity and build a prior covariance in position space. For this purpose, we would, however, need to impose other approximations, such as a sparsity assumption. Or we would need to construct the covariance explicitly, which is computationally unattractive and demands a spatial alignment of the images from the data base.

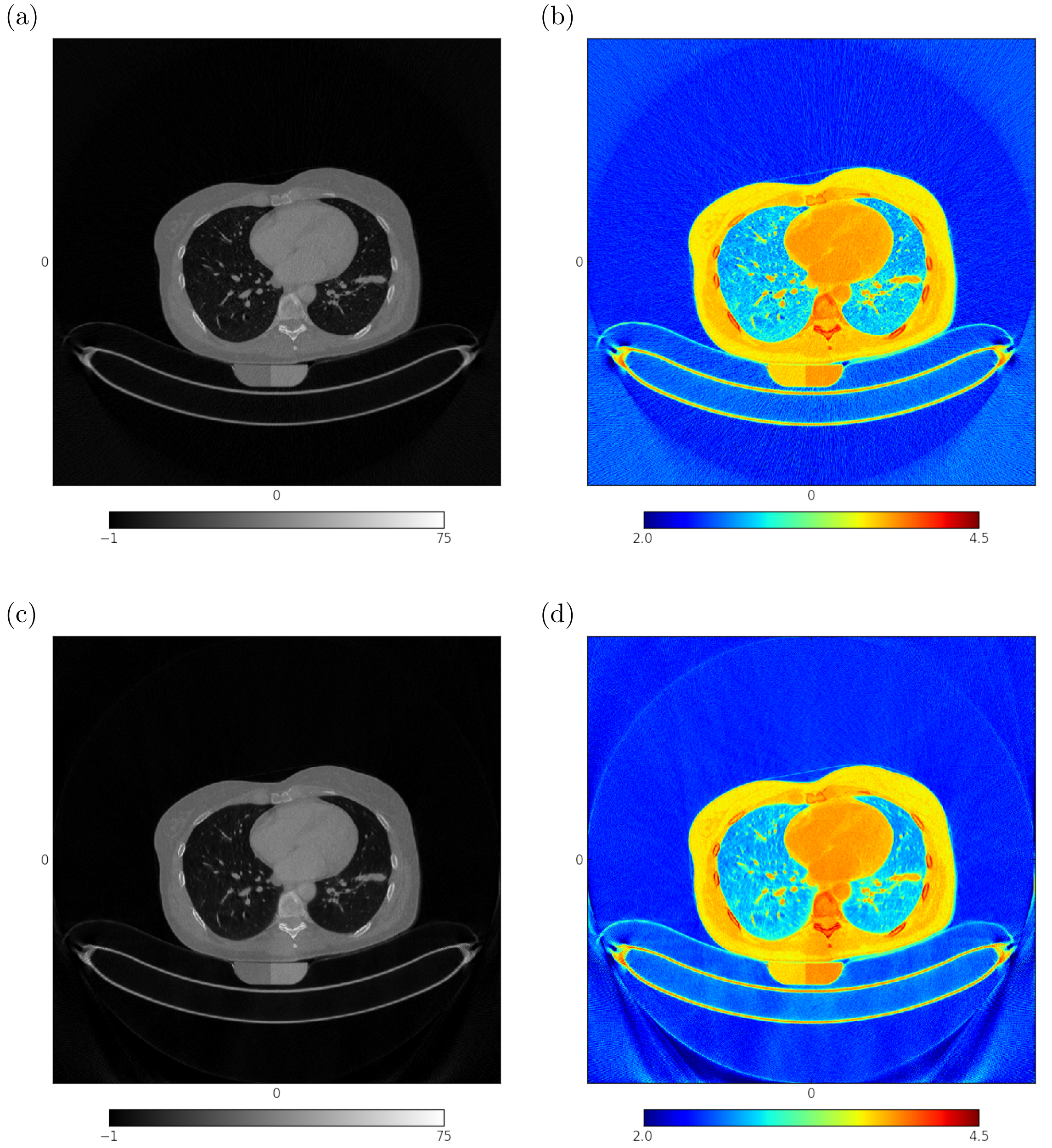
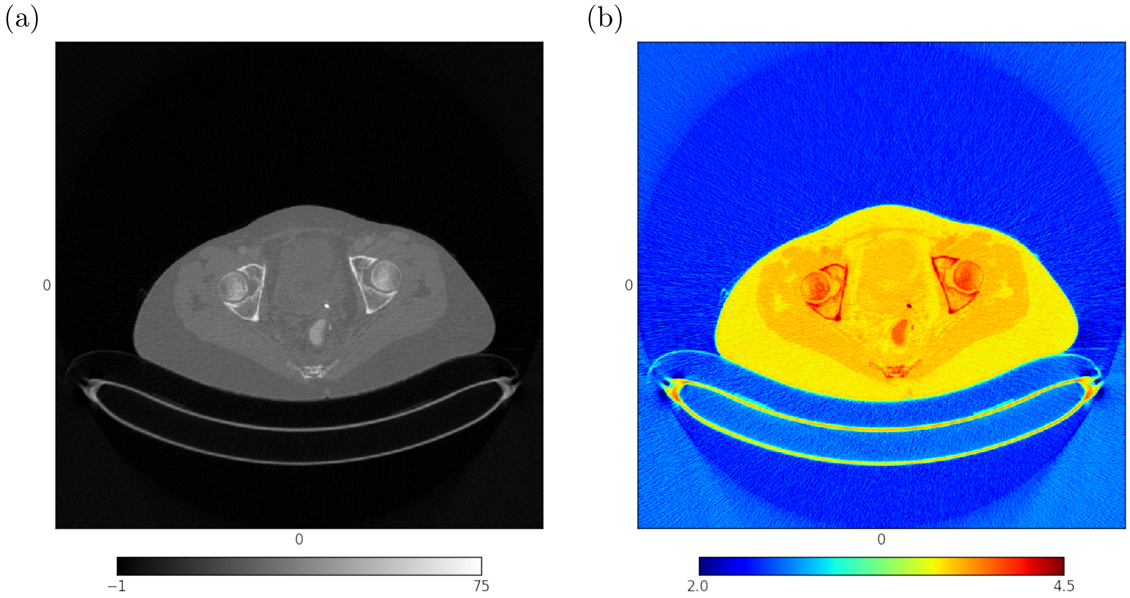


Figure 5.2: Illustration of the reconstruction of the lung slice. The top panels (a) and (b) show FBPs, while the bottom panels (c) and (d) show the reconstruction using IFT. The left panels (a) and (c) show the actual medical image in gray scale, the right panels (b) and (d) show the corresponding reconstructed signal field on a logarithmic false color scale.

5.2.3 Algorithm

The simplest approach to solve the inverse problem of reconstructing the (logarithmic) matter density from CT data is to maximize the posterior distribution. This is equivalent to



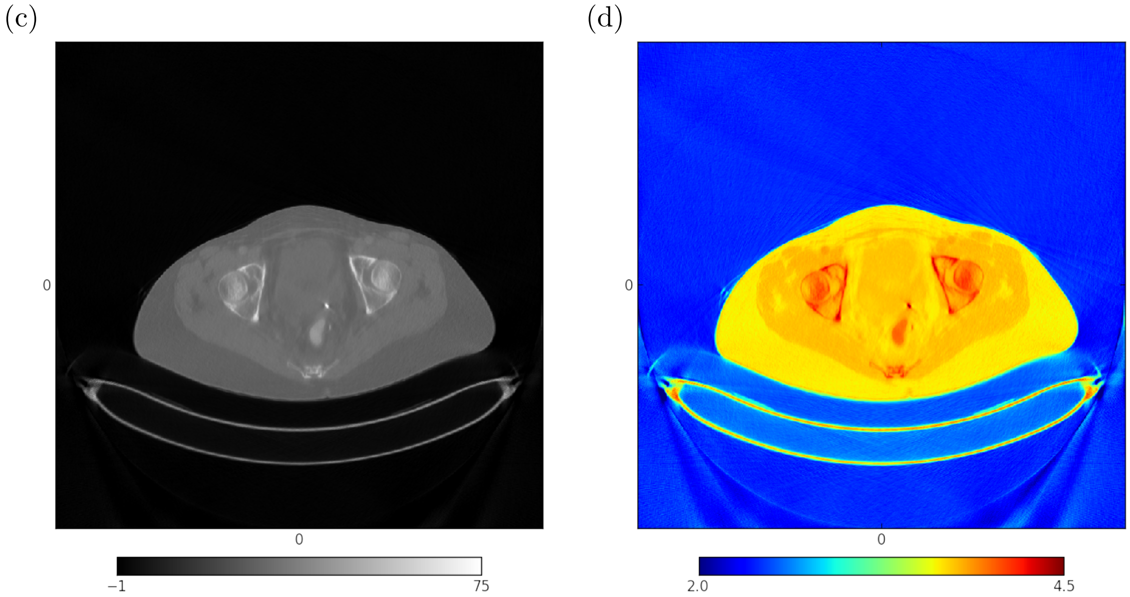


Figure 5.3: Illustration of the reconstruction of the hip slice. Labeling as in Fig. 5.2

minimizing the corresponding information Hamiltonian,

$$H(\boldsymbol{s}|\boldsymbol{d}) = -\log P(\boldsymbol{s}|\boldsymbol{d}) \tag{5.7}$$

$$= H_0 + \lambda_0\, \mathbf{1}^\dagger \exp\left(-a\boldsymbol{R}\mathrm{e}^{\boldsymbol{s}}\right) + a\, \boldsymbol{d}^\dagger \boldsymbol{R}\mathrm{e}^{\boldsymbol{s}} + \frac{1}{2}\boldsymbol{s}^\dagger \boldsymbol{S}^{-1}\boldsymbol{s}, \tag{5.8}$$

where all terms independent of $\boldsymbol{s}$ are absorbed into the term H_0. This Hamiltonian is similar to the one derived for the D^{3}PO problem in Eq. (3.27), where, in addition, a point-like signal field and further parameters are present. A “light” version of the D^{3}PO inference algorithm, which was developed for high energy astronomy, can thus be used to infer medical images with few modifications. All that effectively needs to be modified is the definition of the expected photon counts $\boldsymbol{\lambda}$, which, according to Eq. (5.5), reads

$$\boldsymbol{\lambda} = \lambda_0 \exp\left(-a\boldsymbol{R}\mathrm{e}^{\boldsymbol{s}}\right), \tag{5.9}$$

in comparison to Eq. (3.5). Other differences, such as the completely changed measurement geometry, the replacement of the response operator, or the omission of a point-like component, are all covered by input specifications for the algorithm. This is possible because the D^{3}PO algorithm is based on the NIFTY library and can therefore with ease switch from the celestial sphere to a 2D tomographic plane.

5.3 Applications

In this study, we apply the "light" D^{3}PO algorithm to the X-ray CT scans of two different slices through the torso: one at the level of the lungs, the other at the level of the hip joints.

We compare our reconstructions to FBPs done with the same tomographic projector (Fehringer et al. 2014) using a parallel-beam projection and an angular separation of $\pi/1152$ over a half sphere. All reconstructions have a resolution of 736×736 pixels of which each is $0.675 \times 0.675\,\mathrm{mm}^2$ in size. According to the manufacturer the coefficient a equals 2294.5, however, the emitted number of photons λ_0 is confidential and therefore needs to be calibrated.

For both, the "lung" and the "hip" slice, 10 images are available for determining the prior covariance. We compute the Fourier power spectrum of each sample image and the logarithmic mean of the samples. The power spectra we obtain are shown in Fig. 5.1. The individual sample spectra exhibit similar patterns on the same scales, and those are also represented in their logarithmic mean, although with reduced amplitudes. This implies that the corresponding features in the images are of similar size, although their position information is lost when we condense the images into power spectra. Features that are present in both samples relate to common features, such as the CT table.

Figure 5.2 shows the reconstructions of the lung slice in gray scale and on logarithmic false color scale. While the former gives an impression of the medical image used for diagnosing, the latter shows the reconstructed signal $\boldsymbol{s}$. The FBP and our reconstruction are competitive, as both show the subject's interior with great detail so that different tissues are clearly visible. The FBP exhibits slight radial artifacts, while the IFT reconstruction, on the contrary, shows mostly large-scale artifacts, especially at the lower quarter of the image, below the CT table. Since those are barely present in the FBP, we conclude that they come from the forward projection, which is not needed in the FBP, or from a discrepancy between forward and back projection. D^{3}PO is an iterative algorithm that sequentially applies the projections in form of the response $\boldsymbol{R}$ and its adjoint $\boldsymbol{R}^\dagger$. If forward and back projection are not exactly adjoint, a tiny discrepancy can escalate and cause artifacts or corrupt the whole reconstruction.

The results for the hip slice are shown in Fig. 5.3. The above mentioned applies also here, as we find a general agreement between the two reconstructions. The IFT reconstruction, though, appears blurred towards a spot located near the image center. This spot, which is similar to a point source, apparently disagrees with our model of a single continuous signal field.

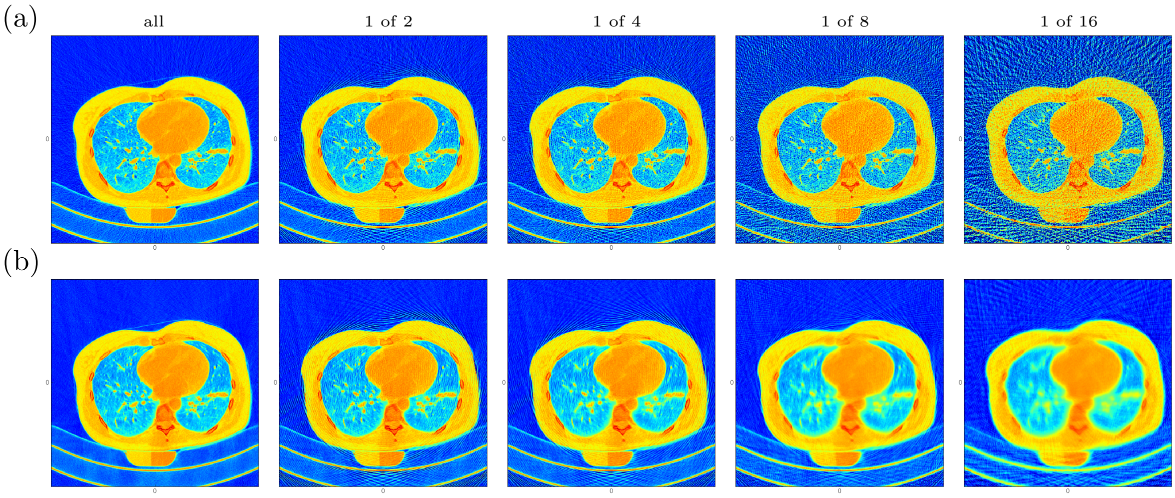


Figure 5.4: Illustration of the reconstruction of the central 512 × 512 pixels of the lung slice on a logarithmic false color scale. The top panels are FBPs, the bottom panels are reconstructions using IFT. From left to right the number of projections considered decreases by a factor of 2 each.

Less Data

One way of reducing the radiation dose for the subject is to record less projections. So far, we used all 1152 projections in the reconstruction of the lung slice. Figure 5.4 shows reconstructions that used only every 2nd, 4th, 8th, and 16th projection of the same data set. Fewer projections mean less redundancy in the data, and thus pose a tougher inverse problem.

The FBPs suffer from less available data by exhibiting stronger radial artifacts. This results in grainier images in which small details perish or are picked to pieces. The IFT reconstructions display only a few radial artifacts, but appear increasingly blurred as the number of used projections diminishes. This behavior is to be expected, since the importance of the prior grows if the constraints of the data are weakened.

In comparison, the FBP and IFT reconstructions using the full and halved data set are competitive. When only 1 out of 4 projections is available, the preference seems to lie with the IFT reconstruction. Beyond that a comparison of the reconstructions is difficult. An assessment of the actual diagnostic value of those images should, however, be left to the experienced eyes of physicians.

5.4 Conclusion & Summary

We have derived a probabilistic imaging technique based on the D^{3}PO inference algorithm that is capable of reconstructing CT slices from real data. The underlying model includes several simplifications, especially with respect to the prior assumptions. A single signal field that is *a priori* assumed to be a statistically homogeneous and isotropic Gaussian random field suffices, but shows its limitations; e.g., when encountering point-like features caused by small pieces of metal. We discussed potential extensions introducing multiple signal components for different classes of media; i.e., air, soft tissue, bones, metals, etc.

The derived imaging technique was successfully applied to anonymized CT data. The resulting reconstructions are competitive with standard FBPs. Furthermore, we test the fidelity of the reconstructions when fewer projections are available, since this means exposing the subject to less radiation. Although judging the diagnostic value of the reconstructions is difficult, we rank the images obtained using IFT as competitive, if not slightly better in comparison to FBPs. We conclude that this study serves as a proof of concept for information theory based medical imaging and provides a promising prospect for future research.

Chapter 6

Conclusions & Future Perspectives

We advanced the imaging techniques for high energy photon observations and successfully applied them to studies of the γ-ray sky seen by the Fermi Large Area Telescope (LAT). We presented a novel approach that simultaneously provides estimates for the continuous, diffuse emission and the point-like γ-ray flux using information field theory (IFT).

After developing NIFTy, a versatile library for signal inference, we derived the D^{3}PO algorithm tasked with denoising, deconvolving, and decomposing photon observations. It addresses the removal or suppression of Poissonian shot noise in photon count images, the rectification of instrumental imprints or artifacts, and the separation of diffuse and point-like components. The D^{3}PO inference algorithm is based on a Bayesian parameter model that describes the multivariate statistics of high dimensional signal fields and their non-linear entanglement in the language of IFT. To solve this complex inverse problem, we exploited *a priori* knowledge on the morphological differences of the diffuse and point-like signal by assuming appropriate statistics and correlation structures. However, D^{3}PO does not only reconstruct the diffuse and point-like photon flux from photon count images, it also estimates the harmonic power spectrum of the diffuse component and provides *a posteriori* uncertainty information on the reconstruction.

Using the D^{3}PO inference algorithm, we analyzed 5.5 years of observational data from the Fermi LAT and presented a continuous reconstruction of the diffuse γ-ray sky, in particular in the GeV energy range, for the first time. This non-parametric estimate excels previous approaches that rely on fitting and subtracting emission templates. The all-sky maps of the diffuse γ-ray emission we obtained and publicly released open up various possibilities for scientific studies. Exploring the spectral properties of the diffuse emission, we presented the first all-sky spectral index map of anisotropic diffuse γ-ray emission and uncovered two distinct components. We concluded that the "cloud"-like component traces the cold, dense interstellar medium, since it is morphologically and spectrally consistent with hadronic interactions of cosmic ray (CR) nuclei with interstellar gas. On the contrary, the "bubble"-like component capturing the morphology of the *Giant Fermi Bubbles* exhibits a harder spectrum, which can be explained by inverse Compton scattering of CR electrons with low energy photons. This leads us to the conclusion that a leptonic scenario explaining the γ-ray emission observed from the bubbles is more compelling.

Our analysis of the Fermi γ-ray sky also yielded a reconstruction of the point-like flux from which we built a point source catalog listing 2,522 candidates. 1,269 of these were confirmed comparing with the second Fermi LAT source catalog, which leaves the potential discovery of 1,253 new sources. Furthermore, we provided upper limit fluxes for a selection of galaxy clusters that are known to have a radio halo.

Finally, we proved the portability of inference algorithms developed using the NIFTy library by applying the D^{3}PO algorithm in the area of medical X-ray computed tomography (CT). We showed that such advanced imaging techniques developed for astronomical imaging are absolutely competitive with standard medical imaging techniques. The successful reconstruction of typical CT slices we carried out provides promising prospects for analyzing CT data sets with less redundancy that would require less X-ray exposure for a subject.

Outlook

Thanks to the great success of the NIFTy library, the D^{3}PO inference algorithm, and the applications presented in this thesis, there are multiple directions of future research. In the following, a brief outlook on the most promising pathways is given.

- **The interplay of spatial and spectral correlations.** Extending the information theoretical methodology and numerical implementation to a coherent treatment of spatial and spectral correlation structures would be the next logical step. This would allow us to exploit the correlation of spatial features across the energy domain, and vice versa, when reconstructing position and energy dependent signals from noisy data cubes. The consequent extension of the D^{3}PO algorithm would render the inference of multiple, multivariate components from multi-channel high energy photon data possible.

- **A refined analysis of the Fermi γ-ray sky.** In case of a simple reanalysis, we should involve further low energy bands down to the LAT limit of 0.02 GeV and refine the resolution of the spherical grid. Such a reanalysis might also profit from, in the meantime, improved instrument calibrations, especially with respect to the LAT's point spread function. We should also reassess our preliminary point source catalog, ideally by studying the individual sources in localized regions of interest.

- **Expansion of the model for medical imaging.** In the first place, this includes the incorporation of a multi-component model for the different absorber media and the construction of priors from a sufficiently large medical data base. Next, we should adopt a realistic tomographic measurement scenario into our model that accounts for energy dispersion, not strictly straight projections, etc. In the end, medical and Galactic tomography might both profit from these improvements.

Appendix A

NIFTy supplement

A.1 Remark On Matrices

The discretization of an operator that is defined on a continuum is a necessity for its computational implementation and is analogous to the discretization of fields; cf. Sec. 2.2.2. However, the involvement of volume weights can cause some confusion concerning the interpretation of the corresponding matrix elements. For example, the discretization of the continuous identity operator, which equals a δ-distribution $\delta(x-y)$, yields a weighted Kronecker-Delta δ_{pq},

$$\mathrm{id} \equiv \delta(x-y) \mapsto \big\langle \big\langle \delta(x-y) \big\rangle_{\mho_p} \big\rangle_{\mho_q} = \frac{\delta_{pq}}{V_q}, \tag{A.1}$$

where $x \in \mho_p$ and $y \in \mho_q$. Say a field $\boldsymbol{\xi}$ is drawn from a zero-mean Gaussian with a covariance that equals the identity, $\mathcal{G}(\boldsymbol{\xi}, \mathrm{id})$. The intuitive assumption that the field values of $\boldsymbol{\xi}$ have a variance of 1 is not true. The variance is given by

$$\langle \xi_p \xi_q \rangle_{\{\boldsymbol{\xi}\}} = \frac{\delta_{pq}}{V_q}, \tag{A.2}$$

and scales with the inverse of the volume V_q. Moreover, the identity operator is the result of the multiplication of any operator with its inverse, $\mathrm{id} = \boldsymbol{A}^{-1}\boldsymbol{A}$. It is trivial to show that, if $A(x,y) \mapsto A_{pq}$ and $\sum_q A^{-1}_{pq} A_{qr} = \delta_{pr}$, the inverse of $\boldsymbol{A}$ maps as follows,

$$\boldsymbol{A}^{-1} \mapsto \big\langle \big\langle A^{-1}(x-y) \big\rangle_{\mho_p} \big\rangle_{\mho_q} = (A^{-1})_{pq} = \frac{A^{-1}_{pq}}{V_p V_q}, \tag{A.3}$$

where A^{-1}_{pq} in comparison to $(A^{-1})_{pq}$ is inversely weighted with the volumes V_p and V_q.

Since all those weightings are implemented in NIFTy, users need to concern themselves with these subtleties only if they intend to extend the functionality of NIFTy.

A.2 Libraries

NIFTy depends on a number of other libraries which are listed here for completeness and in order to give credit to the authors.

- NumPy, SciPy[1] (Oliphant 2006), MatPlotLib (Hunter 2007), and several other Python standard libraries
- GFFT[2] for generalized fast Fourier transformations on regular and irregular grids; of which the latter are currently considered for implementation in a future version of NIFTy
- HEALPy[3] and HEALPix (Górski et al. 2005) for spherical harmonic transformations on the HEALPix grid which are based on the LibPSHT (Reinecke 2011) library or its recent successor LibSHARP[4] (Reinecke & Seljebotn 2013), respectively
- Another Python wrapper[5] for the performant LibSHARP library supporting further spherical pixelizations and the corresponding transformations

These libraries have been selected because they have either been established as standards or they are performant and fairly general.

The addition of alternative numerical libraries is most easily done by the introduction of new derivatives of the `space` class. Replacements of libraries that are already used in NIFTy are possible, but require detailed code knowledge.

---

[1] NumPy and SciPy homepage `http://numpy.scipy.org/`

[2] GFFT homepage `https://github.com/mrbell/gfft`

[3] HEALPy homepage `https://github.com/healpy/healpy`

[4] LibSHARP homepage `http://sourceforge.net/projects/libsharp/`

[5] libsharp-wrapper homepage `https://github.com/mselig/libsharp-wrapper`

A.3 Wiener Filter Code Example

A.3.1 Release Version

```
from nifty import *                                                   # version 0.3.0
from scipy.sparse.linalg import LinearOperator as lo
from scipy.sparse.linalg import cg


class propagator(operator):                                           # define propagator class

    _matvec = (lambda self, x: self.inverse_times(x).val.flatten())

    def _multiply(self, x):
        # some numerical invertion technique; here, conjugate gradient
        A = lo(shape=tuple(self.dim()), matvec=self._matvec, dtype=self.domain.datatype)
        b = x.val.flatten()
        x_, info = cg(A, b, M=None)
        return x_

    def _inverse_multiply(self, x):
        S, N, R = self.para
        return S.inverse_times(x) + R.adjoint_times(N.inverse_times(R.times(x)))


# some signal space; e.g., a one-dimensional regular grid
s_space = rg_space(512, zerocenter=False, dist=0.002)                 # define signal space
# or      rg_space([256, 256])
# or      hp_space(128)

k_space = s_space.get_codomain()                                      # get conjugate space
kindex, rho = k_space.get_power_index(irreducible=True)

# some power spectrum
power = [42 / (kk + 1) ** 3 for kk in kindex]

S = power_operator(k_space, spec=power)                               # define signal covariance
s = S.get_random_field(domain=s_space)                                # generate signal

R = response_operator(s_space, sigma=0.0, mask=1.0, assign=None)      # define response
d_space = R.target                                                    # get data space

# some noise variance; e.g., 1
N = diagonal_operator(d_space, diag=1, bare=True)                     # define noise covariance
n = N.get_random_field(domain=d_space)                                # generate noise

d = R(s) + n                                                          # compute data

j = R.adjoint_times(N.inverse_times(d))                               # define source
D = propagator(s_space, sym=True, imp=True, para=[S,N,R])             # define propagator

m = D(j)                                                              # reconstruct map

s.plot(title="signal")                                                # plot signal
d.cast_domain(s_space)
d.plot(title="data", vmin=s.val.min(), vmax=s.val.max())              # plot data
m.plot(title="reconstructed map", vmin=s.val.min(), vmax=s.val.max()) # plot map
```

A.3.2 Current Version

```
from __future__ import division
from nifty import *                                                # version 0.8.0


# some signal space; e.g., a two-dimensional regular grid
x_space = rg_space([128, 128])                                     # define signal space

k_space = x_space.get_codomain()                                   # get conjugate space

# some power spectrum
power = (lambda k: 42 / (k + 1) ** 3)

S = power_operator(k_space, spec=power)                            # define signal covariance
s = S.get_random_field(domain=x_space)                             # generate signal

R = response_operator(x_space, sigma=0.0, mask=1.0, assign=None) # define response
d_space = R.target                                                 # get data space

# some noise variance; e.g., signal-to-noise ratio of 1
N = diagonal_operator(d_space, diag=s.var(), bare=True)            # define noise covariance
n = N.get_random_field(domain=d_space)                             # generate noise

d = R(s) + n                                                       # compute data

j = R.adjoint_times(N.inverse_times(d))                            # define information source
D = propagator_operator(S=S, N=N, R=R)                             # define information propagator

m = D(j, W=S, tol=1E-3, note=True)                                 # reconstruct map

s.plot(title="signal")                                             # plot signal
d_ = field(x_space, val=d.val, target=k_space)
d_.plot(title="data", vmin=s.min(), vmax=s.max())                  # plot data
m.plot(title="reconstructed map", vmin=s.min(), vmax=s.max())      # plot map
```

Appendix B

D^{3}PO supplement

B.1 Point Source Stacking

In Sec. 3.2.3, 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,

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

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,

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

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

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.

B.2 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. (3.3), or in this Gaussian case as the inverse Hessian of the corresponding information Hamiltonian,

$$\left.\frac{\partial^2 H}{\partial \boldsymbol{s} \partial \boldsymbol{s}^\dagger}\right|_{\boldsymbol{s}=\boldsymbol{m}} = \left.\frac{\partial^2}{\partial \boldsymbol{s} \partial \boldsymbol{s}^\dagger} \left(\frac{1}{2}(\boldsymbol{s} - \boldsymbol{m})^\dagger \boldsymbol{D}^{-1} (\boldsymbol{s} - \boldsymbol{m})\right)\right|_{\boldsymbol{s}=\boldsymbol{m}} = \boldsymbol{D}^{-1}. \tag{B.4}$$

In Sec. 3.3, 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. 3.3.1 are approximated by inverse Hessians. According to Eq. (3.32), they read

$$D^{(s)\,-1}_{xy} \approx \left\{ \sum_i \left(1 - \frac{d_i}{l_i}\right) R_{ix} \mathrm{e}^{m^{(s)}_x} \right\} \delta_{xy} + S^{\star\;-1}_{xy} + \sum_i \frac{d_i}{l_i^2} \left(R_{ix} \mathrm{e}^{m^{(s)}_x}\right) \left(R_{iy} \mathrm{e}^{m^{(s)}_y}\right),$$

and

$$D^{(u)\,-1}_{xy} \approx \left\{ \sum_i \left(1 - \frac{d_i}{l_i}\right) R_{ix} \mathrm{e}^{m^{(u)}_x} + \eta\, \mathrm{e}^{-m^{(u)}_x} \right\} \delta_{xy} + \sum_i \frac{d_i}{l_i^2} \left(R_{ix} \mathrm{e}^{m^{(u)}_x}\right) \left(R_{iy} \mathrm{e}^{m^{(u)}_y}\right), \tag{B.5}$$

with

$$l_i = \int \mathrm{d}x \; R_{ix} \left(\mathrm{e}^{m^{(s)}_x} + \mathrm{e}^{m^{(u)}_x} \right). \tag{B.6}$$

The corresponding covariances derived in the Gibbs approach according to Eq. (3.47), yield

$$\begin{aligned} D^{(s)\,-1}_{xy} \approx & \left\{ \sum_i \left(1 - \frac{d_i}{l_i}\right) R_{ix} \mathrm{e}^{m^{(s)}_x + \frac{1}{2} D^{(s)}_{xx}} \right\} \delta_{xy} + S^{\star\;-1}_{xy} \\ & + \sum_i \frac{d_i}{l_i^2} \left(R_{ix} \mathrm{e}^{m^{(s)}_x + \frac{1}{2} D^{(s)}_{xx}}\right) \left(R_{iy} \mathrm{e}^{m^{(s)}_y + \frac{1}{2} D^{(s)}_{yy}}\right), \end{aligned} \tag{B.7}$$

and

$$\begin{aligned} D^{(u)\,-1}_{xy} \approx & \left\{ \sum_i \left(1 - \frac{d_i}{l_i}\right) R_{ix} \mathrm{e}^{m^{(u)}_x + \frac{1}{2} D^{(u)}_{xx}} + \eta\, \mathrm{e}^{-m^{(u)}_x + \frac{1}{2} D^{(u)}_{xx}} \right\} \delta_{xy} \\ & + \sum_i \frac{d_i}{l_i^2} \left(R_{ix} \mathrm{e}^{m^{(u)}_x + \frac{1}{2} D^{(u)}_{xx}}\right) \left(R_{iy} \mathrm{e}^{m^{(u)}_y + \frac{1}{2} D^{(u)}_{yy}}\right), \end{aligned} \tag{B.8}$$

with

$$l_i = \int \mathrm{d}x \; R_{ix} \left(\mathrm{e}^{m^{(s)}_x + \frac{1}{2} D^{(s)}_{xx}} + \mathrm{e}^{m^{(u)}_x + \frac{1}{2} D^{(u)}_{xx}} \right). \tag{B.9}$$

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. 3.3.2. On the other hand, this shows that higher order correction terms might alter the uncertainty covariances further, cf. Eq. (3.39). 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.

Notice 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. (3.48), the covariances read

$$D^{(s)\,-1}_{xy} = \left\{ \sum_i \left(1 - \frac{d_i}{l_i} \right) R_{ix} \mathrm{e}^{m^{(s)}_x + \frac{1}{2} D^{(s)}_{xx}} \right\} \delta_{xy} + S^{\star\;-1}_{xy}, \tag{B.10}$$

and

$$D^{(u)\,-1}_{xy} = \left\{ \sum_i \left(1 - \frac{d_i}{l_i} \right) R_{ix} \mathrm{e}^{m^{(u)}_x + \frac{1}{2} D^{(u)}_{xx}} + \eta \, \mathrm{e}^{-m^{(u)}_x + \frac{1}{2} D^{(u)}_{xx}} \right\} \delta_{xy}. \tag{B.11}$$

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 D^{3}PO algorithm.

B.3 Posterior Approximation

B.3.1 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,

$$D_{\mathrm{KL}}(Q,P) = \int \mathcal{D}\boldsymbol{z}\; Q(\boldsymbol{z}|\boldsymbol{d}) \log \frac{Q(\boldsymbol{z}|\boldsymbol{d})}{P(\boldsymbol{z}|\boldsymbol{d})} = \left\langle \log \frac{Q(\boldsymbol{z}|\boldsymbol{d})}{P(\boldsymbol{z}|\boldsymbol{d})} \right\rangle_Q, \tag{B.12}$$

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,

$$S_{\mathrm{E}}(Q,P) = -\int \mathcal{D}\boldsymbol{z}\; P(\boldsymbol{z}|\boldsymbol{d}) \log \frac{P(\boldsymbol{z}|\boldsymbol{d})}{Q(\boldsymbol{z}|\boldsymbol{d})} = \left\langle -\log \frac{P(\boldsymbol{z}|\boldsymbol{d})}{Q(\boldsymbol{z}|\boldsymbol{d})} \right\rangle_P = -D_{\mathrm{KL}}(P,Q),$$

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

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

$$\begin{aligned} G &= \big\langle H(\boldsymbol{z}|\boldsymbol{d}) \big\rangle_Q - S_{\mathrm{B}}(Q) \\ &= \big\langle -\log P(\boldsymbol{z}|\boldsymbol{d}) \big\rangle_Q - \big\langle -\log Q(\boldsymbol{z}|\boldsymbol{d}) \big\rangle_Q = D_{\mathrm{KL}}(Q,P), \end{aligned} \tag{B.13}$$

describes the difference between the internal energy $\langle H(\boldsymbol{z}|\boldsymbol{d}) \rangle_Q$ 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[1].

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.

---

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

B.3.2 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.

Let $\boldsymbol{z}$ be 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,

$$P(\boldsymbol{z}|\boldsymbol{d}) \approx Q = \prod_i Q_i(\boldsymbol{z}^{(i)}|\boldsymbol{\mu},\boldsymbol{d}). \tag{B.14}$$

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. (B.14) 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. B.3.3. Notice, 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. Let us 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,

$$\frac{\delta Q_i(\tilde{\boldsymbol{z}}^{(i)}|\boldsymbol{\mu},\boldsymbol{d})}{\delta Q_j(\boldsymbol{z}^{(j)}|\boldsymbol{\mu},\boldsymbol{d})} = \delta_{ij}\ \delta(\boldsymbol{z}^{(i)} - \tilde{\boldsymbol{z}}^{(j)}). \tag{B.15}$$

Computing the variation of the Gibbs free energy yields

$$\begin{aligned}
\delta_j G = 0 &= \frac{\delta}{\delta Q_j(\boldsymbol{z}^{(j)}|\boldsymbol{\mu},\boldsymbol{d})} \left\{ \big\langle H(\boldsymbol{z}|\boldsymbol{d}) \big\rangle_Q - \big\langle -\log Q \big\rangle_Q \right\} && \text{(B.16)} \\
&= \frac{\delta}{\delta Q_j(\boldsymbol{z}^{(j)}|\boldsymbol{\mu},\boldsymbol{d})} \left\{ \big\langle H(\boldsymbol{z}|\boldsymbol{d}) \big\rangle_Q + \sum_i \big\langle \log Q_i(\tilde{\boldsymbol{z}}^{(i)}|\boldsymbol{\mu},\boldsymbol{d}) \big\rangle_{Q_i} \right\} && \text{(B.17)} \\
&= \frac{\delta}{\delta Q_j(\boldsymbol{z}^{(j)}|\boldsymbol{\mu},\boldsymbol{d})} \int \mathcal{D}\tilde{\boldsymbol{z}}^{(j)}\ Q_j(\tilde{\boldsymbol{z}}^{(j)}|\boldsymbol{\mu},\boldsymbol{d}) \left\{ \Big\langle H(\boldsymbol{z}|\boldsymbol{d}) \Big\rangle_{\prod Q_{i\neq j}} + \log Q_j(\tilde{\boldsymbol{z}}^{(j)}|\boldsymbol{\mu},\boldsymbol{d}) \right\} \\
&\quad \underbrace{+ \frac{\delta}{\delta Q_j(\boldsymbol{z}^{(j)}|\boldsymbol{\mu},\boldsymbol{d})} \sum_{i\neq j} \big\langle \log Q_i(\tilde{\boldsymbol{z}}^{(i)}|\boldsymbol{\mu},\boldsymbol{d}) \big\rangle_{Q_i}}_{=0} \\
&= \int \mathcal{D}\tilde{\boldsymbol{z}}^{(j)}\ \delta(\boldsymbol{z}^{(i)} - \tilde{\boldsymbol{z}}^{(j)}) \left\{ \Big\langle H(\boldsymbol{z}|\boldsymbol{d}) \Big\rangle_{\prod Q_{i\neq j}} + \log Q_j(\tilde{\boldsymbol{z}}^{(j)}|\boldsymbol{\mu},\boldsymbol{d}) + 1 \right\} \\
&= \Big\langle H(\boldsymbol{z}|\boldsymbol{d}) \Big|_{\boldsymbol{z}^{(j)}} \Big\rangle_{\prod Q_{i\neq j}} + \log Q_j(\boldsymbol{z}^{(j)}|\boldsymbol{\mu},\boldsymbol{d}) + \text{const.} && \text{(B.18)}
\end{aligned}$$

This defines a solution for the approximate posterior Q_j, where the constant term in Eq. (B.18) ensures the correct normalization[2] of Q_j,

$$Q_j(\boldsymbol{z}^{(j)}|\boldsymbol{\mu},\boldsymbol{d}) \propto \exp\left(- \Big\langle H(\boldsymbol{z}|\boldsymbol{d}) \Big|_{\boldsymbol{z}^{(j)}} \Big\rangle_{\prod Q_{i\neq j}} \right). \tag{B.19}$$

---

[2] The normalization could be included by usage of Lagrange multipliers; i.e., by adding a term $\sum_i \lambda_i \big(1 - \int \mathcal{D}\boldsymbol{z}^{(i)}\ Q_i(\boldsymbol{z}^{(i)}|\boldsymbol{\mu},\boldsymbol{d})\big)$ to the Gibbs free energy in Eq. (B.16).

Although the parts $\boldsymbol{z}^{(i \neq j)}$ are integrated out, Eq. (B.19) 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,

$$\underset{\boldsymbol{z}^{(j)}}{\operatorname{argmax}}\; P(\boldsymbol{z}|\boldsymbol{d}) = \underset{\boldsymbol{z}^{(j)}}{\operatorname{argmin}}\; H(\boldsymbol{z}|\boldsymbol{d}) \approx \underset{\boldsymbol{z}^{(j)}}{\operatorname{argmin}}\; \left\langle H(\boldsymbol{z}|\boldsymbol{d})\Big|_{\boldsymbol{z}^{(j)}} \right\rangle_{\prod Q_{i \neq j}} . \tag{B.20}$$

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 $\boldsymbol{z}$.

B.3.3 Example

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

$$P(\boldsymbol{s}, \boldsymbol{\tau}|\boldsymbol{d}) = \frac{P(\boldsymbol{d}|\boldsymbol{s})}{P(\boldsymbol{d})}\; P(\boldsymbol{s}|\boldsymbol{\tau})\; P(\boldsymbol{\tau}) \tag{B.21}$$

$$= \frac{P(\boldsymbol{d}|\boldsymbol{s})}{P(\boldsymbol{d})}\; \mathcal{G}(\boldsymbol{s}, \boldsymbol{S})\; P_{\mathrm{un}}(\boldsymbol{\tau}|\boldsymbol{\alpha}, \boldsymbol{q})\; P_{\mathrm{sm}}(\boldsymbol{\tau}|\boldsymbol{\sigma}), \tag{B.22}$$

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. 3.2 in order to find a solution for the logarithmic power 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

$$H(\boldsymbol{s}, \boldsymbol{\tau}|\boldsymbol{d}) = -\log P(\boldsymbol{s}, \boldsymbol{\tau}|\boldsymbol{d}) \tag{B.23}$$

$$= H_0 + \frac{1}{2} \sum_k \left(\varrho_k \tau_k + \operatorname{tr}\left[\boldsymbol{s}\boldsymbol{s}^\dagger \boldsymbol{S}_k^{-1} \right] \mathrm{e}^{-\tau_k} \right) \tag{B.24}$$

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

where $\varrho_k = \operatorname{tr}\left[\boldsymbol{S}_k {\boldsymbol{S}_k}^{-1} \right]$ 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. (3.2) and (3.3), let us 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. B.1. With the consequential approximation of the posterior,

$$P(\boldsymbol{s}, \boldsymbol{\tau}|\boldsymbol{d}) \approx \mathcal{G}(\boldsymbol{s} - \boldsymbol{m}, \boldsymbol{D})\; Q_\tau(\boldsymbol{\tau}|\boldsymbol{\mu}, \boldsymbol{d}), \tag{B.25}$$

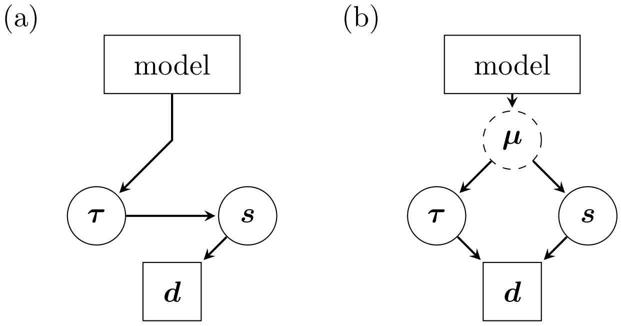


Figure B.1: Graphical model for the variational method applied to the example posterior in Eq. (B.21). Panel (a) shows the graphical model without, and panel (b) with the mean field $\boldsymbol{\mu}$.

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

$$\begin{aligned}\left\langle H(\boldsymbol{s},\boldsymbol{\tau}|\boldsymbol{d})\Big|_{\boldsymbol{\tau}}\right\rangle_{Q_s} &= H_0 + \boldsymbol{\gamma}^\dagger\boldsymbol{\tau} + \frac{1}{2}\boldsymbol{\tau}^\dagger\boldsymbol{T}\boldsymbol{\tau} + \boldsymbol{q}^\dagger \mathrm{e}^{-\boldsymbol{\tau}} + \frac{1}{2}\sum_k \mathrm{tr}\left[\left\langle \boldsymbol{s}\boldsymbol{s}^\dagger\right\rangle_{Q_s}\boldsymbol{S}_k^{-1}\right]\mathrm{e}^{-\tau_k} \\ &= H_0 + \boldsymbol{\gamma}^\dagger\boldsymbol{\tau} + \frac{1}{2}\boldsymbol{\tau}^\dagger\boldsymbol{T}\boldsymbol{\tau} + \boldsymbol{q}^\dagger \mathrm{e}^{-\boldsymbol{\tau}} + \frac{1}{2}\sum_k \mathrm{tr}\Big[\Big(\boldsymbol{m}\boldsymbol{m}^\dagger + \boldsymbol{D}\Big)\boldsymbol{S}_k^{-1}\Big]\mathrm{e}^{-\tau_k},\end{aligned} \tag{B.26}$$

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. (B.24) that ensures an information flow between $\boldsymbol{s}$ and $\boldsymbol{\tau}$,

$$\boldsymbol{\mu} = \begin{pmatrix} \left\langle \mathrm{tr}\left[\boldsymbol{s}\boldsymbol{s}^\dagger\boldsymbol{S}_k^{-1}\right]\right\rangle_{Q_s} \\ \left\langle \sum_k \mathrm{e}^{-\tau_k}\boldsymbol{S}_k^{-1}\right\rangle_{Q_\tau} \end{pmatrix} = \begin{pmatrix} \mathrm{tr}\left[\left(\boldsymbol{m}\boldsymbol{m}^\dagger + \boldsymbol{D}\right)\boldsymbol{S}_k^{-1}\right] \\ \left\langle \boldsymbol{S}^{-1}\right\rangle_{Q_\tau} \end{pmatrix} \tag{B.27}$$

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. (B.26) yields

$$\mathrm{e}^{\boldsymbol{\tau}} = \frac{\boldsymbol{q} + \frac{1}{2}\left(\mathrm{tr}\left[\left(\boldsymbol{m}\boldsymbol{m}^\dagger + \boldsymbol{D}\right)\boldsymbol{S}_k^{-1}\right]\right)_k}{\boldsymbol{\gamma} + \boldsymbol{T}\boldsymbol{\tau}}. \tag{B.28}$$

This formula is in agreement with the critical filter formula (Enßlin & Frommert 2011; Oppermann et al. 2012b). 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.

B.4 Model Parameter Sampling

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

Table B.1: Overview of the relative residual error of 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.

Appendix C

Fermi LAT Data analysis

C.1 Data Analysis

C.1.1 Data Selection

In this work, we analyze the 5.5 years of observational data taken by the Fermi LAT; i.e., data from mission weeks 9 to 296 (mission elapsed time 239,557,417 s to 413,199,625 s). The data are subject to multiple restrictions and cuts detailed in the following.

For our analysis we exclusively consider events classified as `P7REP_CLEAN_V15` in the reprocessed Pass 7 data set. The `CLEAN` events, which are "cleaned" of CR interactions with the instrument, are recommended for studies of the diffuse γ-ray emission.[1] Events with inclination angles above 52° and zenith angles above 100° are excluded in order to suppress contaminations from CRs and the Earth's limb (Atwood et al. 2009).[2] In addition, we apply a (non-standard) cut with respect to the angular distance to the Sun that we require to exceed 20°. This way, almost all Solar γ-rays are rejected at the price of reducing the total number of events by less than 3%. A similar procedure regarding the moon is conceivable but ignored, because its contribution is negligible.

The individual events are labeled `FRONT` or `BACK` according to whether the photon has been converted in the front or back section of the LAT instrument. We retain this labeling, but combine those data vectors by a direct sum; i.e.,

$$\boldsymbol{d} = \boldsymbol{d}^{\texttt{FRONT}} \oplus \boldsymbol{d}^{\texttt{BACK}} = \left(\boldsymbol{d}^{\texttt{FRONT}}, \boldsymbol{d}^{\texttt{BACK}}\right)^{\mathsf{T}} . \tag{C.1}$$

The selected events are binned in nine (logarithmically equally spaced) energy bins ranging from 0.6 to 307.2 GeV, cf. Tab. 4.1. We also apply a spatial binning of the events into all-sky count maps using a HEALPix discretization with $n_{\text{side}} = 128$, which corresponds to 196,608 pixels with a size of roughly $64\,\mu\text{sr} \approx (0.46°)^2$ each.

For a proper deconvolution, our analysis has to take the LAT's PSF and exposure into account. The instrumental response functions of the Fermi LAT (Atwood et al. 2009; Abdo et al. 2009; Ackermann et al. 2012a), which are essential therefor, have been improved in the reprocessed Pass 7 release, and are available within the Fermi Science Tools. According to our event selection, we make use of the `P7REP_CLEAN_V15::FRONT` and `BACK` IRFs, respectively.

---

[1] For further details regarding the Fermi LAT data products see `http://fermi.gsfc.nasa.gov/ssc/data/`.

[2] The exact filter expression reads `"DATA_QUAL>0 && LAT_CONFIG==1 && ABS(ROCK_ANGLE)<52 && ZENITH<100"`.

These can be assumed to be accurately calibrated, although further improvements, especially at low energies, are under discussion (Portillo & Finkbeiner 2014). Given the IRFs, the exposure $(\varepsilon_{ij}^{\texttt{FRONT}}, \varepsilon_{ij}^{\texttt{BACK}})$ for each HEALPix pixel i and each energy band j can be retrieved from the data archive. In order to compute ε, the Sun exposure is subtracted from the "standard" exposure due to the chosen rejection of potential Solar events (Johannesson et al. 2013). The PSF of the Fermi LAT is a function of position and energy E. Its shape varies slightly with spatial translation and sharpens strongly with increasing energy. The forward application of the PSF is a linear operation that can be implemented in form of a convolution matrix evaluating the PSF at each pixel center and for each energy band. This matrix is fairly sparse because of the vanishing tails of the PSF, and is computed beforehand. The exposure and the PSF define the instrument response operator $\boldsymbol{R}$,

$$R_{ij}(x) \propto \frac{1}{(E_j^{\max} - E_j^{\min})} \begin{pmatrix} \varepsilon_{ij}^{\texttt{FRONT}} \times \mathrm{PSF}_i^{\texttt{FRONT}}(E_j^{\mathrm{mid}}, x) \\ \varepsilon_{ij}^{\texttt{BACK}} \times \mathrm{PSF}_i^{\texttt{BACK}}(E_j^{\mathrm{mid}}, x) \end{pmatrix}, \tag{C.2}$$

up to a proportionality constant that can absorb numerical factors and physical units. Notice that this definition does not include a spectral convolution; i.e., there is no cross-talk between different energy bands.

The primary target of our analysis is the physical photon flux $\boldsymbol{\phi} = \phi(x)$, which is a function of position $x \in \Omega$. Here, the position space Ω is the observational sphere, and the position x might be given in spherical coordinates (φ, θ), or in Galactic longitude and latitude (l, b).

The response operator $\boldsymbol{R}$ describes the mapping of a photon flux $\boldsymbol{\phi}$ to $\boldsymbol{\lambda} = \boldsymbol{R\phi}$ by a convolution with the IRFs; i.e.,

$$\lambda_{ij} = \int_\Omega \mathrm{d}x \; R_{ij}(x)\phi(x), \tag{C.3}$$

where $\boldsymbol{\lambda}$ describes the noiseless (non-integer) number of photons one expects to observe through the IRFs given some photon flux in the sky. This expected number of counts $\boldsymbol{\lambda}$ relates to the observed (integer) photon counts $\boldsymbol{d}$ by a noise process, which is part of the statistical model detailed in the next section.

C.1.2 Inference Algorithm

The foundation of the analysis presented in this work is the D^{3}PO inference algorithm derived by Selig & Enßlin (2013) that targets the denoising, deconvolution and decomposition of photon observations. Without going into technical details, we briefly review the underlying assumptions and characteristics of the D^{3}PO algorithm in the following.

The observed photon count data $\boldsymbol{d}$ carries information about the astrophysical photon flux $\boldsymbol{\phi}$, as well as noise and instrumental imprints. In order to optimally reconstruct $\boldsymbol{\phi}$ given $\boldsymbol{d}$, we incorporate our knowledge about the actual measurement in a data model that consists of deterministic relations and probabilistic processes. Hence, D^{3}PO is a probabilistic algorithm conducting Bayesian inference.

We can assume the photon counts to suffer from Poissonian shot noise; i.e., the data entries d_{ij} are the outcomes of statistically independent Poisson processes given an expected number of counts λ_{ij} each. Especially in low photon flux regions and at high energies, where the signal-to-noise ratios are worst, using a Poissonian likelihood allows for an accurate noise treatment whereas Gaussian noise approximations often fail.

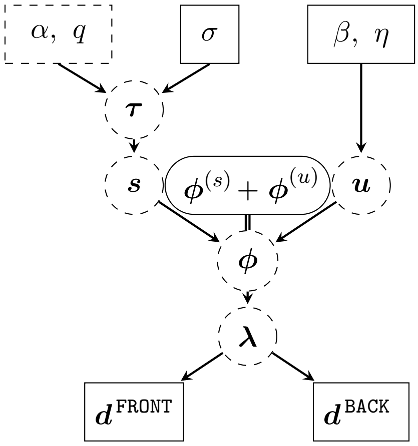


Figure C.1: Graphical model of the model parameters α, q, σ, β, and η, the logarithmic angular power spectrum $\boldsymbol{\tau}$, the diffuse and point-like signal fields, $\boldsymbol{s}$ and $\boldsymbol{u}$, the photon fluxes, $\boldsymbol{\phi}$, $\boldsymbol{\phi}^{(s)}$, and $\boldsymbol{\phi}^{(u)}$, and the expected and observed number of photons, $\boldsymbol{\lambda}$ and $\boldsymbol{d}$.

D^{3}PO's deconvolution task covers the correction of all effects that trace back to the instrumental response $\boldsymbol{R}$. As discussed in Sect. C.1.1, this response establishes a relation between the astrophysical photon flux $\boldsymbol{\phi}$ and the expected counts $\boldsymbol{\lambda}$ by taking the IRFs of the Fermi LAT fully into account. Since we suppose the IRFs to be thoroughly calibrated (to the best of our knowledge), this relation is deterministic.

The total photon flux $\boldsymbol{\phi}$ consists of many different contributions that can be divided into two morphological classes, diffuse and point-like contributions. Diffuse emission, which is produced by the interaction of CRs with the ISM, unresolved point sources and extragalactic background, is characterized by spatially smooth fluctuations. On the contrary, point-like emission is fairly local originating primarily from resolved point sources. The D^{3}PO algorithm reconstructs the total photon flux as the sum of a diffuse and point-like flux contribution; i.e.,

$$\boldsymbol{\phi} = \boldsymbol{\phi}^{(s)} + \boldsymbol{\phi}^{(u)} = \phi_0 \left(\mathrm{e}^{\boldsymbol{s}} + \mathrm{e}^{\boldsymbol{u}} \right), \tag{C.4}$$

where ϕ_0 is a constant absorbing numerical factors and flux units, and the exponentiation is applied pixelwise to the diffuse and point-like signal fields, $\boldsymbol{s}$ and $\boldsymbol{u}$. Those signal fields describe the dimensionless logarithmic flux ensuring the positivity of the physical photon flux in a natural way. Although the algorithm deals with the $\boldsymbol{s}$ and $\boldsymbol{u}$ fields for numerical reasons, we only regard the fluxes $\boldsymbol{\phi}^{(s)}$ and $\boldsymbol{\phi}^{(u)}$ in the following as they are physical.

We can incorporate our naive understanding of “diffuse” and “point-like” by introducing prior assumptions. Embedding *a priori* knowledge on the solution, of course, biases the inference. However, priors also remedy the degeneracy of the inference problem as they suppress counterintuitive solutions.

The diffuse photon flux $\boldsymbol{\phi}^{(s)}$, being spatially smooth, is expected to exhibit spatial correlation. Without enforcing concrete spatial features, such as a Galactic profile, we assume $\boldsymbol{\phi}^{(s)}$ to obey multivariate log-normal statistics. Assuming, furthermore, statistical homogeneity and isotropy, the underlying covariance is determined by an *a priori* unknown angular power spectrum. In order to retain a flexible handle on this power spectrum, we further introduce hyperpriors (cf. e.g., Enßlin & Frommert 2011; Oppermann et al. 2012b; Selig & Enßlin 2013). We assume a (asymptotically) uniform prior for the logarithmic angular power and a spectral smoothness prior as suggested by Oppermann et al. (2012b). This introduces scalar model parameters α, q, and σ.

The point-like photon flux $\boldsymbol{\phi}^{(u)}$ exhibits strong features that appear to be fairly rare and local. We assume $\boldsymbol{\phi}^{(u)}$ to follow statistically independent inverse-Gamma distributions described by the model parameters β and η.

For a detailed derivation and discussion of the probabilistic model D^{3}PO is based on, we refer to Selig & Enßlin (2013). An illustrative graphical model of the introduced hierarchy of Bayesian parameters is shown in Fig. C.1.

C.1.3 Analysis Procedure

In theory, we could apply the D^{3}PO algorithm to the whole data set at once. However, it is computationally more efficient to exploit the spectral separability of the response model by applying the algorithm to each energy band individually, cf. Eq. C.2. In order to exploit spectral correlations, we propose to align the priors of the diffuse component after an initial inference run by defining a common angular power spectrum. This yields a three-step analysis procedure detailed in the following.

Initial Inference: The D^{3}PO algorithm is applied to each energy band separately, which can be done in parallel.

We fix the model parameters with fairly soft constraints, by setting $\alpha = 1$, $q = 10^{-12}$, $\sigma = 10$, $\beta = \frac{3}{2}$, and $\eta = 10^{-2}$. The limit of $(\alpha, q) \rightarrow (1, 0)$ leads to uniform prior for the logarithmic angular power spectrum $\boldsymbol{\tau}$ of the diffuse photon flux (Enßlin & Frommert 2011), but choosing a non-zero q is numerically more stable. The spectral smoothness parameter σ is the *a priori* standard deviation of the second derivative of $\boldsymbol{\tau} = \tau_\ell$ with respect to the logarithm of the angular quantum number ℓ; i.e., σ describes the tolerance of deviations from a power-law shape (Oppermann et al. 2012b). The parameter tuple (β, η) determines the slope and scale of the inverse-Gamma prior of the point-like photon flux. While a slope of $\frac{3}{2}$ is generally applicable, the scale, for which we find 10^{-2} fitting, has to be adapted according to the chosen resolution (Selig & Enßlin 2013).

D^{3}PO solves an inference problem that is non-linear and, in general, non-convex. To circumvent a dependence on its initialization, D^{3}PO can generate suitable starting values by solving a coarse grained inference problem first. For this purpose only, we provide a binary exposure masking the most prominent point sources.

Prior Alignment: The prior of the diffuse component describes our *a priori* expectation of how spatially smooth the emission is. If we find diffuse structures of a certain size at one energy band, we can expect to find similar structures at neighboring bands, especially since most diffuse emission processes exhibit power-law-like energy spectra. Thus, we expect significant spectral correlations, in particular for prominent features such as the Galactic bulge, for example. Since the incorporation of a spectral convolution in the response is computationally infeasible, we impose an aligning of the diffuse priors to exploit spectral correlations, nonetheless.

The diffuse prior is defined by the logarithmic angular power spectrum $\boldsymbol{\tau}$. As discussed in Sec. 4.3.4, we find a rough power-law behavior of the power spectra with deviations due to the Galactic disk and finite exposure. Fig. 4.12 shows the results of the initial inference, in particular including a power spectrum fit averaged across the energy bands. This average spectrum defines the aligned prior. Notice that the apparent excess of small-scale power for high energy bands, comparing inferred and aligned power spectra, remedies potential perception thresholds occurring during the inference (Enßlin & Frommert 2011).

Afterwards, we also align the diffuse maps within the initially masked regions according to the aligned prior in order to avoid artifacts due to initialization.[3]

The alignment of the point-like priors, through β and η, has proven ineffective in tests and is therefore omitted.

Final Inference: The D^{3}PO algorithm is applied to each energy band separately, again. For this run, however, we keep the (aligned) angular power spectrum fixed and provide the aligned diffuse maps as starting values. Hence, the initially used binary mask is not required any more. Notice that a fixed angular power spectrum renders the model parameters α, q, and σ obsolete.

---

[3] We minimize the prior term, $\boldsymbol{s}^\dagger \boldsymbol{S}^{-1} \boldsymbol{s}$, but only for $s(x|x = (\varphi, \theta) \in \text{mask} \wedge x \sim (\ell, m) \neq (2\mathbb{N}, 0))$ ensuring the preservation of the Galactic profile and the reconstructed field values outside the mask.

Acknowledgments

During the three years of my PhD studies, I came in contact with many people who in one way or another contributed to the success of this thesis. In the following, I would like to acknowledge the involvement of the collaborators, colleagues, and friends deserving my special thanks:

- Niels and Henrik – my "big PhD brothers" – with whom I faced countless puzzling problems at the blackboard and could always discuss science, fantasy, and worldly matters;
- Torsten – my "PhD father" – who I can without hesitation nominate as "Best Thesis Supervisor" for guiding, encouraging, advising, supporting, and pushing me;
- Mike and Valentina – "The PostDocs" – who taught me how to live as a scientist and that there is more to life than living for science;
- Philipp, Mahsa, Maksim, Vanessa, Sebastian, and Theo – my "little PhD siblings" – whom I more than anything wish well for their thesis to come;
- Lars, Carlos, Ashmeet, Gašper, Christian, Daniel, JP, and David – our undergraduate and graduate students – who never stopped to amaze me;
- Martin, Jörg, Wolfgang, and Uwe – "The Planck guys" – who shared their expertise in computer science with me;
- Tiago, who granted me a little insight into biophysics;
- Peter, Andreas, and Radin, who supported me throughout the medical imaging project;
- Michael, Zhongli, Ming, Dimitri, and Qingbo, with whom it was a pleasure to share office number 161;
- and all the silent helpers whom I would like to express my gratitude.

Finally, I would like to thank my whole family and closest friends for their lasting support and encouragement by word and deed.

Thank you.

Bibliography

A., M. P. 1993, Quantum probability for probabilists, , Springer

Abdo, A. A., Ackermann, M., Ajello, M., et al., Fermi Large Area Telescope First Source Catalog, ApJS **188** (Jun. 2010) 405–436, `arXiv:1002.2280 [astro-ph.HE]`

Abdo, A. A., Ackermann, M., Ajello, M., et al., The on-orbit calibration of the Fermi Large Area Telescope, Astroparticle Physics **32** (Oct. 2009) 193–219, `arXiv:0904.2226 [astro-ph.IM]`

Abdo, A. A., Ackermann, M., Ajello, M., et al., Spectrum of the Isotropic Diffuse Gamma-Ray Emission Derived from First-Year Fermi Large Area Telescope Data, Physical Review Letters **104** no. 10, (Mar. 2010) 101101, `arXiv:1002.3603 [astro-ph.HE]`

Ackermann, M., Ajello, M., Albert, A., et al., The Fermi Large Area Telescope on Orbit: Event Classification, Instrument Response Functions, and Calibration, ApJS **203** (Nov. 2012) 4, `arXiv:1206.1896 [astro-ph.IM]`

Ackermann, M., Ajello, M., Albert, A., Atwood, W. B., & Fermi LAT Collaboration, The spectrum of isotropic diffuse gamma-ray emission between 100 MeV and 820 GeV, ArXiv e-prints (Oct. 2014), `arXiv:1410.3696 [astro-ph.HE]`

Ackermann, M., Ajello, M., Albert, A., et al., Fermi LAT search for dark matter in gamma-ray lines and the inclusive photon spectrum, Phys. Rev. D **86** no. 2, (Jul. 2012) 022002, `arXiv:1205.2739 [astro-ph.HE]`

Ackermann, M., Ajello, M., Allafort, A., et al., In-flight measurement of the absolute energy scale of the Fermi Large Area Telescope, Astroparticle Physics **35** (Jan. 2012) 346–353, `arXiv:1108.0201 [astro-ph.IM]`

Ackermann, M., Ajello, M., Allafort, A., et al., GeV Gamma-ray Flux Upper Limits from Clusters of Galaxies, ApJ **717** (Jul. 2010) L71–L78, `arXiv:1006.0748 [astro-ph.HE]`

Ackermann, M., Ajello, M., Atwood, W. B., et al., Fermi-LAT Observations of the Diffuse Gamma-Ray Emission: Implications for Cosmic Rays and the Interstellar Medium, ArXiv e-prints (Feb. 2012), `arXiv:1202.4039 [astro-ph.HE]`

Ackermann, M., Albert, A., Atwood, W. B., & Fermi LAT Collaboration, The Spectrum and Morphology of the Fermi Bubbles, ArXiv e-prints (Jul. 2014), `arXiv:1407.7905 [astro-ph.HE]`

Ackermann, M., Jóhannesson, G., Digel, S., et al., A method to analyze the diffuse gamma-ray emission with the Fermi Large Area Telescope, in American Institute of Physics Conference Series, Vol. 1085, American Institute of Physics Conference Series, ed. F. A. Aharonian, W. Hofmann, & F. Rieger. 2008, 763–766

Alvarez, R. E. & Macovski, A., Energy-selective reconstructions in X-ray computerized tomography, Phys Med Biol. (Sep. 1976) 733–44

Atwood, W. B., Abdo, A. A., Ackermann, M., et al., The Large Area Telescope on the Fermi Gamma-Ray Space Telescope Mission, ApJ **697** (Jun. 2009) 1071–1102, `arXiv:0902.1089 [astro-ph.IM]`

B., R. & Lightman, A. P. 1985, Radiative Processes in Astrophysics, , Wiley-VCH

Bayes, T., An Essay towards Solving a Problem in the Doctrine of Chances, Philosophical Transactions of the Royal Society **35** (1763) 370–418, `http://journals.royalsociety.org/content/w24177406t384j06/`

Behnel, S., Bradshaw, R. W., & Seljebotn, D. S., Cython tutorial, in Proceedings of the 8th Python in Science Conference, Pasadena, CA USA. 2009, 4 – 14

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

Böhm et al., in prep. (2014)

Bouchet, L., Amestoy, P., Buttari, A., Rouet, F.-H., & Chauvin, M., Simultaneous analysis of large INTEGRAL/SPI[1] 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]`

Bregeon, J., Charles, E., & M. Wood for the Fermi-LAT collaboration, Fermi-LAT data reprocessed with updated calibration constants, ArXiv e-prints (Apr. 2013), `arXiv:1304.5456 [astro-ph.HE]`

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]`

Carretti, E., Crocker, R. M., Staveley-Smith, L., et al., Giant magnetized outflows from the centre of the Milky Way, Nature **493** (Jan. 2013) 66–69, `arXiv:1301.0512 [astro-ph.GA]`

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]`

Cheng, K.-S., Chernyshov, D. O., Dogiel, V. A., Ko, C.-M., & Ip, W.-H., Origin of the Fermi Bubble, ApJ **731** (Apr. 2011) L17, `arXiv:1103.1002 [astro-ph.HE]`

Chernyshov, D., Particle acceleration and the origin of gamma-ray emission from Fermi Bubbles, International Cosmic Ray Conference **7** (2011) 11–18, `arXiv:1109.2619 [astro-ph.GA]`

Conrad, J., Searches for particle dark matter with gamma-rays, in American Institute of Physics Conference Series, Vol. 1505, American Institute of Physics Conference Series, ed. F. A. Aharonian, W. Hofmann, & F. M. Rieger. 2012, 166–176, `arXiv:1210.4392 [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

Cox, R. T., Probability, Frequency and Reasonable Expectation, American Journal of Physics **14** (Jan. 1946) 1

Dermer, C. D., The Extragalactic γ Ray Background, in American Institute of Physics Conference Series, Vol. 921, The First GLAST Symposium, ed. S. Ritz, P. Michelson, & C. A. Meegan. 2007, 122–126, `arXiv:0704.2888`

Dermer, C. D., Strong, A. W., Orlando, E., Tibaldo, L., & for the Fermi Collaboration, Determining the Spectrum of Cosmic Rays in Interstellar Space from the Diffuse Galactic Gamma-Ray Emissivity, ArXiv e-prints (Jul. 2013), `arXiv:1307.0497 [astro-ph.HE]`

Dogiel, V. A., Chernyshov, D. O., Cheng, K., et al., Fermi Bubbles as a Result of Star Capture in the Galactic Center, ArXiv e-prints (Sep. 2011), `arXiv:1109.6087 [astro-ph.HE]`

Dorn, S., Oppermann, N., Khatri, R., Selig, M., & Enßlin, T. A., A fast and precise way to calculate the posterior for the local non-Gaussianity parameter f_{NL} from Cosmic Microwave Background observations, Phys. Rev. D **88** (Nov. 2013) 103516, `arXiv:1307.3884 [astro-ph.CO]`

Dorn, S., Ramirez, E., Kunze, K. E., Hofmann, S., & Enßlin, T. A., Generic inference of inflation models by non-Gaussianity and primordial power spectrum reconstruction, J. Cosmology Astropart. Phys. **6** (Jun. 2014) 48, `arXiv:1403.5067`

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., Junklewitz, H., Winderling, L., & Selig, M., On the art and theory of self-calibration, `arXiv:1312.1349 [astro-ph.IM]`

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

Fehringer, A., Lasser, T., Zanette, I., Noël, P. B., & F., P., A versatile tomographic forward- and back-projection approach on multi-GPUs, in Proc. SPIE, Vol. 9034, Medical Imaging 2014: Image Processing. 2014

Feretti, L., Giovannini, G., Govoni, F., & Murgia, M., Clusters of galaxies: observational properties of the diffuse radio emission, A&A Rev. **20** (May 2012) 54, `arXiv:1205.1919 [astro-ph.CO]`

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]`

Finkbeiner, D. P., Davis, M., & Schlegel, D. J., Extrapolation of Galactic Dust Emission at 100 Microns to Cosmic Microwave Background Radiation Frequencies Using FIRAS, ApJ **524** (Oct. 1999) 867–886, `astro-ph/9905128`

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

Górski, K. M., Hivon, E., Banday, A. J., et al., HEALPix: A Framework for High-Resolution Discretization and Fast Analysis of Data Distributed on the Sphere, ApJ **622** (Apr. 2005) 759–771, `arXiv:astro-ph/0409513`

Greiner, M. & Enßlin, T. A., Conversion between the power spectra of a field and its logarithm, ArXiv e-prints (Dec. 2013), `arXiv:1312.1354 [astro-ph.CO]`

Greiner et al., in prep. (2014)

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

Haslam, C. G. T., Salter, C. J., Stoffel, H., & Wilson, W. E., A 408 MHz all-sky continuum survey. II - The atlas of contour maps, A&AS **47** (Jan. 1982) 1, ADS

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]`

Herman, G. T. 2009, Fundamentals of Computerized Tomography: Image Reconstruction from Projections, , Springer

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

Hsieh, J. 2009, Computed Tomography Principles, Design, Artifacts, and Recent Advances, , Wiley-VCH

Hunter, J. D., Matplotlib: A 2D graphics environment, Computing In Science & Engineering **9** no. 3, (2007) 90–95

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., Li, C., & Enßlin, T. A., Bayesian non-linear large-scale structure inference of the Sloan Digital Sky Survey Data Release 7, MNRAS **409** (Nov. 2010) 355–370, `arXiv:0911.2498 [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

Jaynes, E. T. 1989, Dordrecht

Johannesson, G., Orlando, E., & for the Fermi-LAT collaboration, Accounting for the Sun and the Moon in Fermi-LAT Analysis, ArXiv e-prints (Jun. 2013), `arXiv:1307.0197 [astro-ph.IM]`

Johnson, T. R., Krauss, B., Sedlmair, M., et al., Material differentiation by dual energy CT: initial experience, Eur Radiol. (Jun. 2007)

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. A., & Enßlin, T., A new approach to multi-frequency synthesis in radio interferometry, ArXiv e-prints (Jan. 2014), `arXiv:1401.4711 [astro-ph.IM]`

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]`

Kak, A. C. & Slaney, M. 1988, Principles of Computerized Tomographic Imaging, , IEEE Press

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

Kitaura, F. S., Jasche, J., Li, C., et al., Cosmic cartography of the large-scale structure with Sloan Digital Sky Survey data release 6, MNRAS **400** (Nov. 2009) 183–203, `arXiv:0906.3978 [astro-ph.CO]`

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

Laplace, P. S. 1795/1951, A philosophical essay on probabilities, Dover, New York

Large, M. I., Quigley, M. J. S., & Haslam, C. G. T., A new feature of the radio sky, MNRAS **124** (1962) 405

Longair, M. S. 2011, High Energy Astrophysics, , Cambridge University Press

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

Moskalenko, I. V. & Strong, A. W., Anisotropic Inverse Compton Scattering in the Galaxy, ApJ **528** (Jan. 2000) 357–367, `astro-ph/9811284`

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

Nolan, P. L., Abdo, A. A., Ackermann, M., et al., Fermi Large Area Telescope Second Source Catalog, ApJS **199** (Apr. 2012) 31, `arXiv:1108.1435 [astro-ph.HE]`

Oliphant, T. 2006, A Guide to NumPy, , Trelgol Publishing

Oppermann, N., Junklewitz, H., Greiner, M., et al., Estimating extragalactic Faraday rotation, ArXiv e-prints (Apr. 2014), `arXiv:1404.3701 [astro-ph.IM]`

Oppermann, N., Junklewitz, H., Robbers, G., et al., An improved map of the Galactic Faraday sky, A&A **542** (Jun. 2012) A93, `arXiv:1111.6186 [astro-ph.GA]`

Oppermann, N., Robbers, G., & Enßlin, T. A., Reconstructing signals from noisy data with unknown signal and noise covariance, Phys. Rev. E **84** no. 4, (Oct. 2011) 041118, `arXiv:1107.2384 [astro-ph.IM]`

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, Planck 2013 results. XI. All-sky model of thermal dust emission, ArXiv e-prints (Dec. 2013), `arXiv:1312.1300`

Planck Collaboration, Planck intermediate results. IX. Detection of the Galactic haze with Planck, A&A **554** (Jun. 2013) A139, `arXiv:1208.5483 [astro-ph.GA]`

Planck Collaboration, Planck intermediate results. XXVIII. Interstellar gas and dust in the Chamaeleon clouds as seen by Fermi LAT and Planck, ArXiv e-prints (Sep. 2014), `arXiv:1409.3268 [astro-ph.HE]`

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]`

Portillo, S. K. N. & Finkbeiner, D. P., Sharper Fermi LAT Images: instrument response functions for an improved event selection, ArXiv e-prints (Jun. 2014), `arXiv:1406.0507 [astro-ph.IM]`

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]`

Reinecke, M., Libpsht - algorithms for efficient spherical harmonic transforms, A&A **526** (Feb. 2011) A108, `arXiv:1010.2084 [astro-ph.IM]`

Reinecke, M. & Seljebotn, D. S., Libsharp - spherical harmonic transforms revisited, ArXiv e-prints (Mar. 2013), `arXiv:1303.4945 [physics.comp-ph]`

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. & Enßlin, T. A., Denoising, Deconvolving, and Decomposing Photon Observations, `arXiv:1311.1888 [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]`

Selig, M., Vacca, V., Oppermann, N., & Enßlin, T. A., The Denoised, Deconvolved, and Decomposed Fermi gamma-ray Sky, `arXiv:1410.4562 [astro-ph.HE]`

Seljebotn, D. S., Fast numerical computations with Cython, in Proceedings of the 8th Python in Science Conference, Pasadena, CA USA. 2009, 15 – 22

Shannon, C. E., A mathematical theory of communication, Bell System Technical Journal **27** (1948) 379–423

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

Sreekumar, P., Bertsch, D. L., Dingus, B. L., et al., EGRET Observations of the Extragalactic Gamma-Ray Emission, ApJ **494** (Feb. 1998) 523–534, `astro-ph/9709257`

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

Strong, A. W., Moskalenko, I. V., & Reimer, O., Diffuse Continuum Gamma Rays from the Galaxy, ApJ **537** (Jul. 2000) 763–784, `astro-ph/9811296`

Strong, A. W., Orlando, E., & Jaffe, T. R., The interstellar cosmic-ray electron spectrum from synchrotron radiation and direct measurements, A&A **534** (Oct. 2011) A54, `arXiv:1108.4822 [astro-ph.HE]`

Su, M. & Finkbeiner, D. P., Evidence for Gamma-Ray Jets in the Milky Way, ApJ **753** (Jul. 2012) 61, `arXiv:1205.5852 [astro-ph.HE]`

Su, M., Slatyer, T. R., & Finkbeiner, D. P., Giant Gamma-ray Bubbles from Fermi-LAT: Active Galactic Nucleus Activity or Bipolar Galactic Wind?, ApJ **724** (Dec. 2010) 1044–1082, `arXiv:1005.5480 [astro-ph.HE]`

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]`

Unsöld, A. & Baschek, B. 2001, The New Cosmos – An Introduction to Astronomy and Astrophysics, ed. W. Brewer, , Springer

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`

Weig, C. & Enßlin, T. A., Bayesian analysis of spatially distorted cosmic signals from Poissonian data, MNRAS **409** (Dec. 2010) 1393–1411, `arXiv:1003.1311 [astro-ph.CO]`

Wiener, N. 1949, Extrapolation, Interpolation and Smoothing of Stationary Time Series, with Engineering Applications, Technology Press and Wiley, New York, note: Originally issued in Feb 1942 as a classified Nat. Defense Res. Council Rep.

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]`

Yang, H.-Y. K., Ruszkowski, M., & Zweibel, E., The Fermi bubbles: gamma-ray, microwave and polarization signatures of leptonic AGN jets, MNRAS **436** (Dec. 2013) 2734–2746, `arXiv:1307.3551 [astro-ph.GA]`

Yang, R.-z., Aharonian, F., & Crocker, R., The Fermi Bubbles Revisited, ArXiv e-prints (Feb. 2014), `arXiv:1402.0403 [astro-ph.HE]`

Selbstständigkeitserklärung

Ich versichere hiermit, dass diese kumulative Dissertation von mir selbständig und ohne unerlaubte Hilfe angefertigt wurde.

Marco Selig 31. Oktober 2014