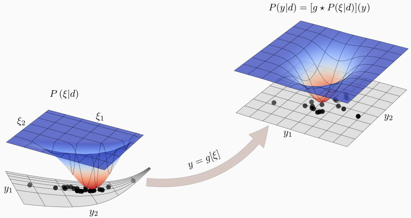


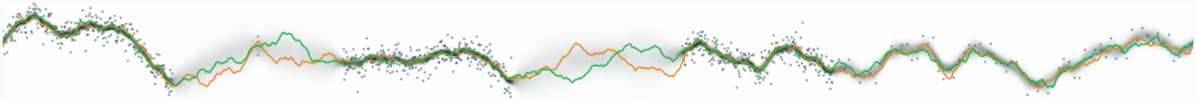
Geometric Variational Inference

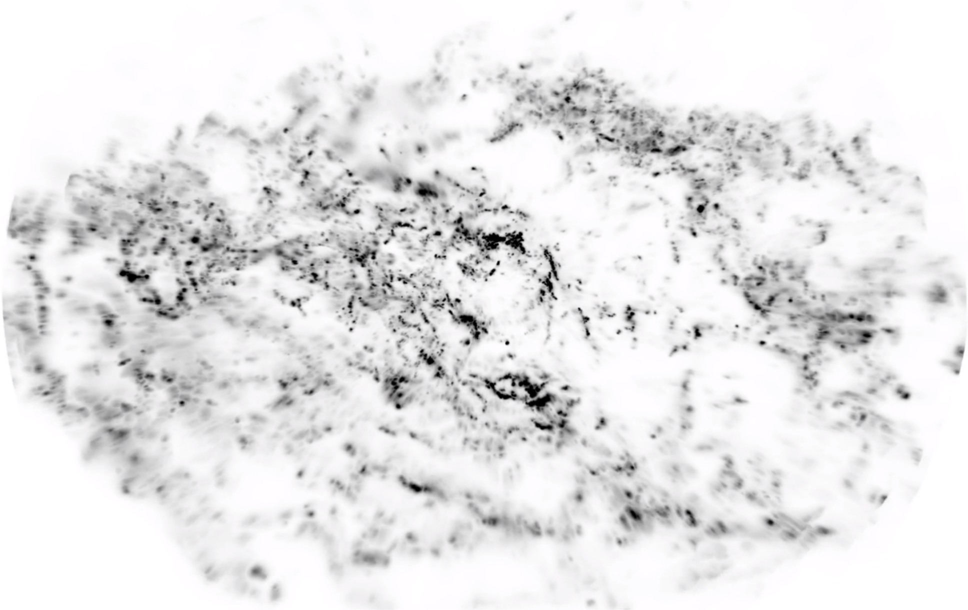
Approximate Bayesian inference in high dimensions

Philipp Frank[1], Reimar Leike[1,3], Torsten Ensslin[1,2]

Seminar: Signal reconstruction with python, MPA, Garching, September 26, 2023

(1) Max-Planck Institute for Astrophysics MPA, Garching, Germany
(2) Ludwig-Maximilians University LMU, Munich, Germany
(3) OpenAI, San Francisco, California, U.S.





Imaging problems

Product Rule of Probabilities aka Bayes' theorem

$$\mathcal{P}(s|d, \mathcal{M}) = \frac{\mathcal{P}(d|s, \mathcal{M})\,\mathcal{P}(s|\mathcal{M})}{\mathcal{P}(d|\mathcal{M})}$$

Definitions: $s :=$ parameters, $d :=$ data, $\mathcal{M}$: model assumptions.

Imaging problems

Product Rule of Probabilities aka Bayes' theorem

$$\mathcal{P}(s|d,\mathcal{M}) = \frac{\mathcal{P}(d|s,\mathcal{M})\,\mathcal{P}(s|\mathcal{M})}{\mathcal{P}(d|\mathcal{M})}$$

Definitions: $s :=$ parameters, $d :=$ data, $\mathcal{M}$: model assumptions.

Generative prior model: $s(\xi) = F_M(\xi) \quad \text{with} \quad P(\xi) = \mathcal{N}(\xi|0,\mathbb{1})$

Product Rule of Probabilities aka Bayes' theorem

$$\mathcal{P}(\xi|d,\mathcal{M}) = \frac{\mathcal{P}(d|s(\xi),\mathcal{M})\,\mathcal{N}(\xi|0,\mathbb{1})}{\mathcal{P}(d|\mathcal{M})}$$

Definitions: ξ := parameters, d := data, $\mathcal{M}$: model assumptions.

Generative prior model: $s(\xi) = F_M(\xi)$ with $P(\xi) = \mathcal{N}(\xi|0,\mathbb{1})$

Probability measures

Posterior expectation

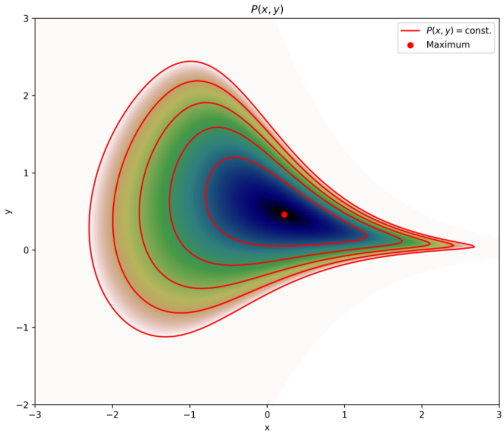
$$\langle f(\xi)\rangle_{\mathcal{P}(\xi|d)} = \int f \; \mathrm{d}\mathcal{P} = \int f(\xi) \; \mathcal{P}(\xi|d) \; \mathrm{d}\xi$$

Function: $f(\xi)$; Posterior: $\mathcal{P}(\xi|d)$; parameters: ξ; data: d.

Posterior expectation

$$\langle f(\xi)\rangle_{\mathcal{P}(\xi|d)} = \int f \, \mathrm{d}\mathcal{P} = \int f(\xi) \, \mathcal{P}(\xi|d) \, \mathrm{d}\xi$$

Function: $f(\xi)$; Posterior: $\mathcal{P}(\xi|d)$; parameters: ξ; data: d.

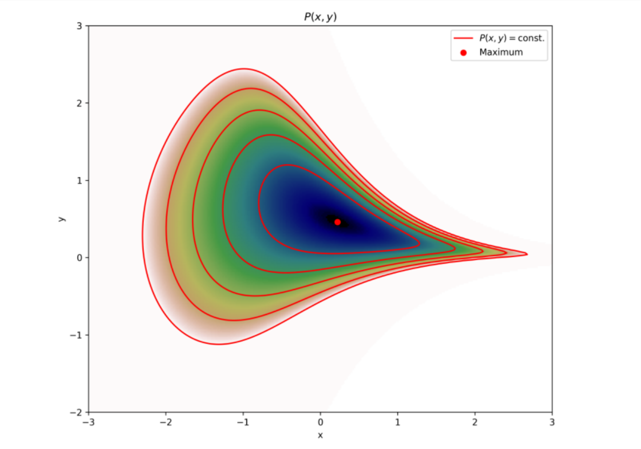


Probability measures

Posterior expectation

$$\langle f(\xi)\rangle_{\mathcal{P}(\xi|d)} = \int f \; \mathrm{d}\mathcal{P} = \int f(\xi)\; \mathcal{P}(\xi|d)\; \mathrm{d}\xi \approx f\left(\xi_{\mathrm{MAX}}\right)$$

Function: $f(\xi)$; Posterior: $\mathcal{P}(\xi|d)$; parameters: ξ; data: d.

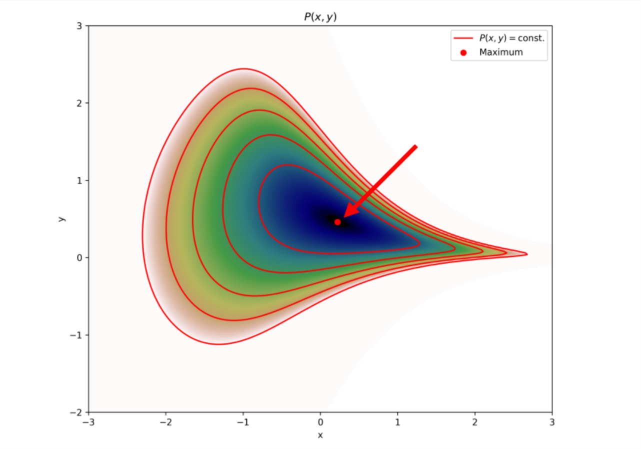


Probability measures

Posterior expectation

$$\langle f(\xi) \rangle_{\mathcal{P}(\xi|d)} = \int f \, \mathrm{d}\mathcal{P} = \int f(\xi) \, \mathcal{P}(\xi|d) \, \mathrm{d}\xi \approx f\left(\xi_{\mathrm{MAX}}\right)$$

Function: $f(\xi)$; Posterior: $\mathcal{P}(\xi|d)$; parameters: ξ; data: d.

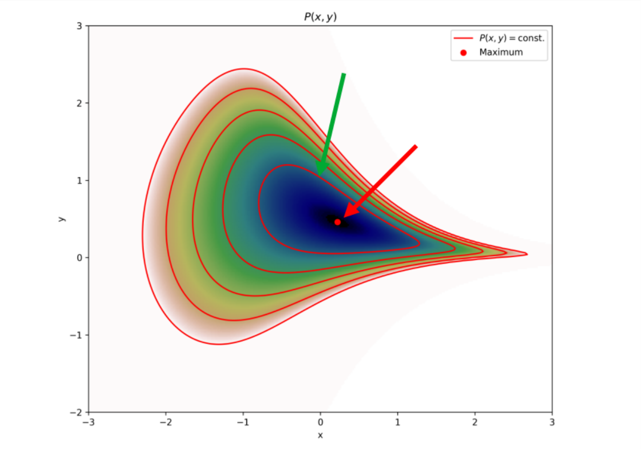


Probability measures

Posterior expectation

$$\langle f(\xi)\rangle_{\mathcal{P}(\xi|d)} = \int f \, \mathrm{d}\mathcal{P} = \int f(\xi)\, \mathcal{P}(\xi|d)\, \mathrm{d}\xi \approx f\left(\xi_{\mathrm{MAX}}\right)$$

Function: $f(\xi)$; Posterior: $\mathcal{P}(\xi|d)$; parameters: ξ; data: d.



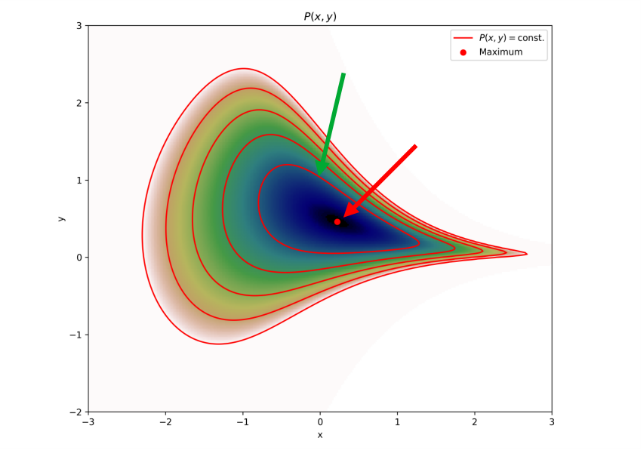
Variational Inference

Probability measures

Posterior expectation

$$\langle f(\xi) \rangle_{\mathcal{P}(\xi|d)} = \int f \; \mathrm{d}\mathcal{P} = \int f(\xi) \; \mathcal{P}(\xi|d) \; \mathrm{d}\xi \approx f\left(\xi_{\mathrm{MAX}}\right)$$

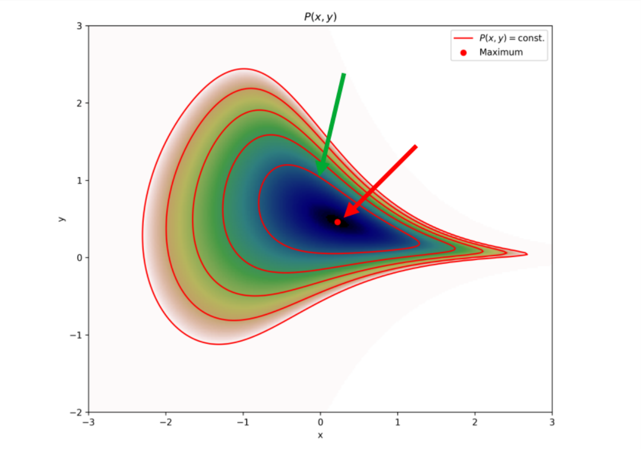
Function: $f(\xi)$; Posterior: $\mathcal{P}(\xi|d)$; parameters: ξ; data: d.



Posterior expectation

$$\langle f(\xi)\rangle_{\mathcal{P}(\xi|d)} = \int f \, \mathrm{d}\mathcal{P} = \int f(\xi)\, \mathcal{P}(\xi|d)\, \mathrm{d}\xi \approx \int f(\xi)\, \mathcal{Q}(\xi|d)\, \mathrm{d}\xi$$

Function: $f(\xi)$; Posterior: $\mathcal{P}(\xi|d)$; parameters: ξ; data: d.

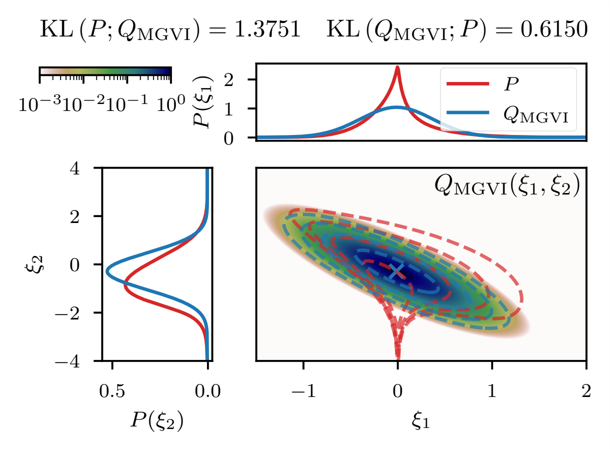


Probability measures

Posterior expectation

$$\langle f(\xi) \rangle_{\mathcal{P}(\xi|d)} = \int f \ \mathrm{d}\mathcal{P} = \int f(\xi)\ \mathcal{P}(\xi|d)\ \mathrm{d}\xi \approx \int f(\xi)\ \mathcal{Q}(\xi|d)\ \mathrm{d}\xi$$

Function: $f(\xi)$; Posterior: $\mathcal{P}(\xi|d)$; parameters: ξ; data: d.



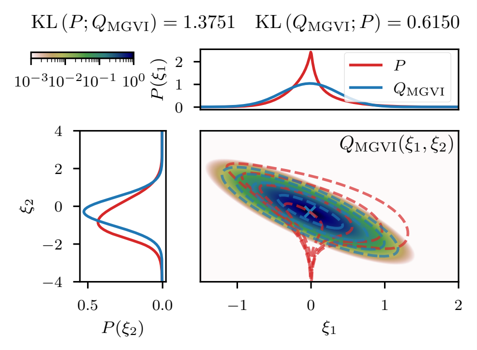
Variational Inference (VI)

Kullback-Leibler divergence

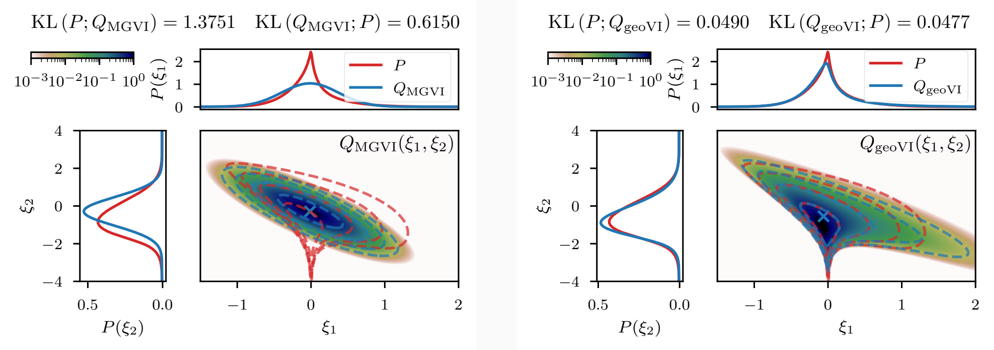
$$\mathrm{KL}\left[\mathcal{Q}_\sigma || \mathcal{P}\right] = -\int \log\left(\frac{\mathcal{P}(\xi|d)}{\mathcal{Q}_\sigma(\xi)}\right)\ \mathcal{Q}_\sigma(\xi)\ \mathrm{d}\xi$$

Posterior: $\mathcal{P}(\xi|d)$; Approximation: $\mathcal{Q}_\sigma(\xi)$; Variational parameters: σ.

Variational Inference (VI)



Variational Inference (VI)



Variational Inference (VI)

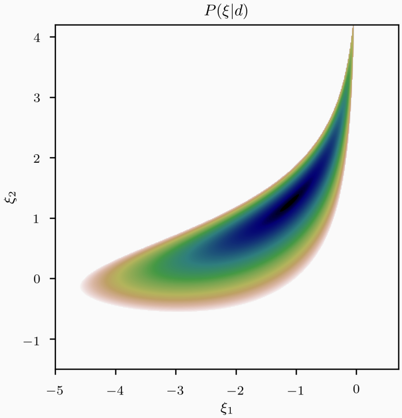
Kullback-Leibler divergence

$$\mathrm{KL}\left[\mathcal{Q}_\sigma || \mathcal{P}\right] = -\int \log\left(\frac{\mathcal{P}(\xi|d)}{\mathcal{Q}_\sigma(\xi)}\right) \mathcal{Q}_\sigma(\xi)\, \mathrm{d}\xi$$

Posterior: $\mathcal{P}(\xi|d)$; Approximation: $\mathcal{Q}_\sigma(\xi)$; Variational parameters: σ.

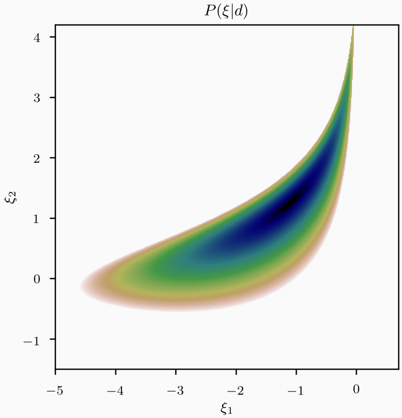
Approximate distribution $\mathcal{Q}$: $\mathcal{Q}(y) = \mathcal{N}(y; 0, \mathbb{1})$
Coordinate system $y = g_\sigma(\xi)$ such that the *posterior* is close to Normal.

Geometric Variational Inference (geoVI) [FLE21]



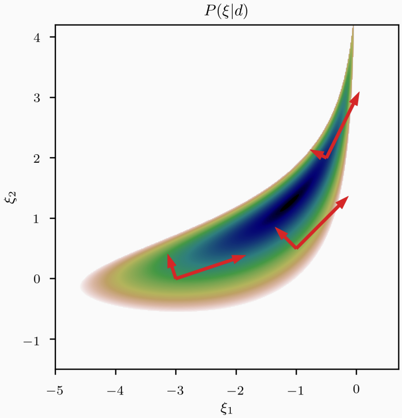
Information Hamiltonian $\mathcal{H}(\xi|d)$: $\quad -\log\left(\mathcal{P}(\xi|d)\right)$

Geometric Variational Inference (geoVI) [FLE21]



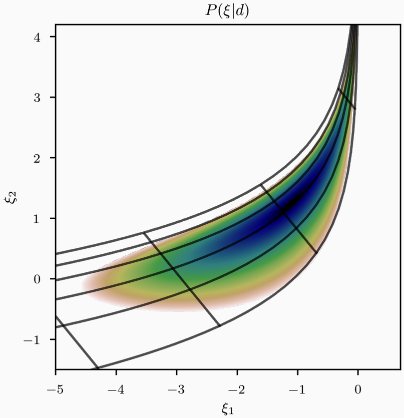
Information Hamiltonian $\mathcal{H}(\xi\|d)$:	$-\log\left(\mathcal{P}(\xi\|d)\right)$
Posterior metric $\mathcal{M}(\xi)$:	$\mathcal{M}_{\mathrm{lh}}(\xi) + \mathbb{1}$
Fisher information metric $\mathcal{M}_{\mathrm{lh}}(\xi)$:	$\left\langle \frac{\partial^2 \mathcal{H}(d\|\xi)}{\partial\xi\partial\xi'} \right\rangle_{\mathcal{P}(d\|\xi)}$

Geometric Variational Inference (geoVI) [FLE21]



Information Hamiltonian $\mathcal{H}(\xi\|d)$:	$-\log\left(\mathcal{P}(\xi\|d)\right)$
Posterior metric $\mathcal{M}(\xi)$:	$\mathcal{M}_{\mathrm{lh}}(\xi) + \mathbb{1}$
Fisher information metric $\mathcal{M}_{\mathrm{lh}}(\xi)$:	$\left\langle \frac{\partial^2 \mathcal{H}(d\|\xi)}{\partial\xi\partial\xi'} \right\rangle_{\mathcal{P}(d\|\xi)}$

Geometric Variational Inference (geoVI) [FLE21]

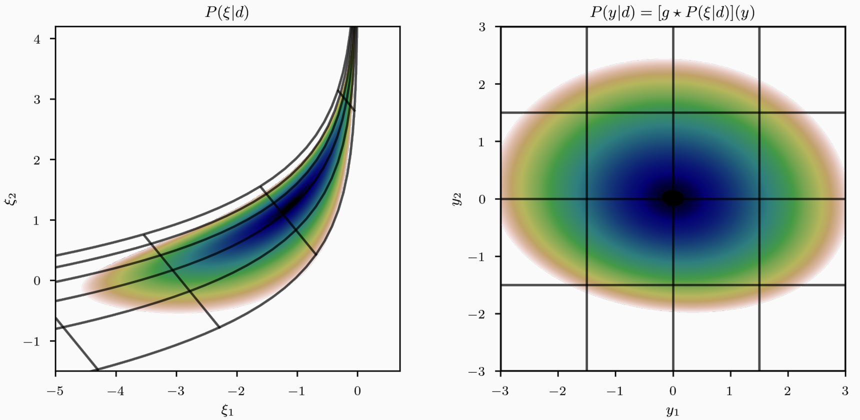


Information Hamiltonian $\mathcal{H}(\xi|d)$: $-\log\left(\mathcal{P}(\xi|d)\right)$

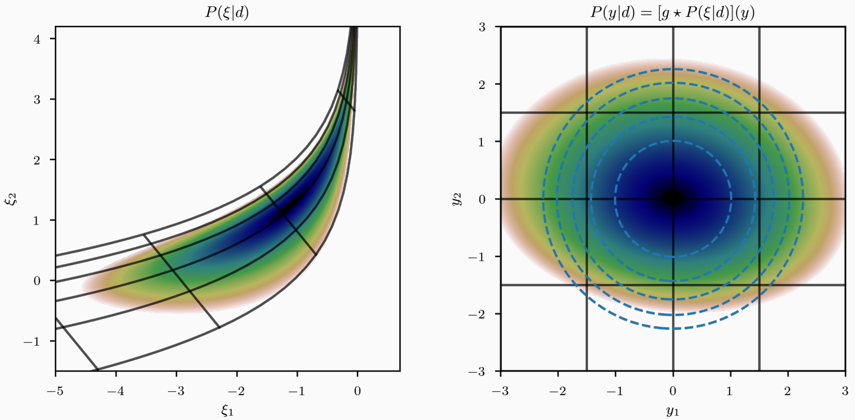
Posterior metric $\mathcal{M}(\xi)$: $\mathcal{M}_{\mathrm{lh}}(\xi) + \mathbb{1}$

Fisher information metric $\mathcal{M}_{\mathrm{lh}}(\xi)$: $\left\langle \frac{\partial^2 \mathcal{H}(d|\xi)}{\partial\xi\partial\xi'} \right\rangle_{\mathcal{P}(d|\xi)}$

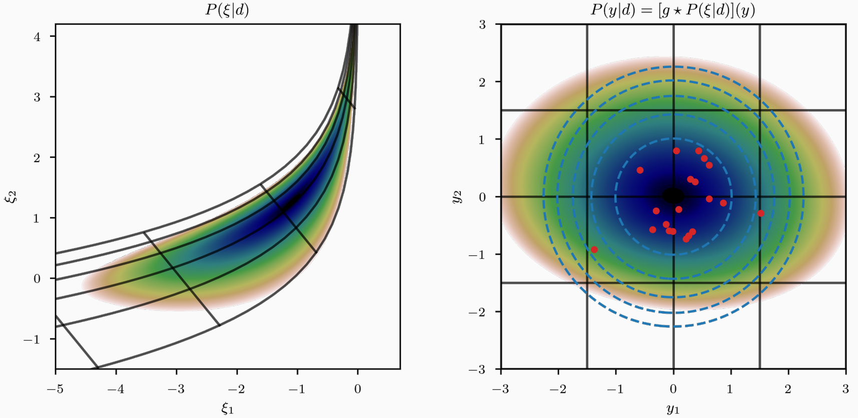
Geometric Variational Inference (geoVI) [FLE21]



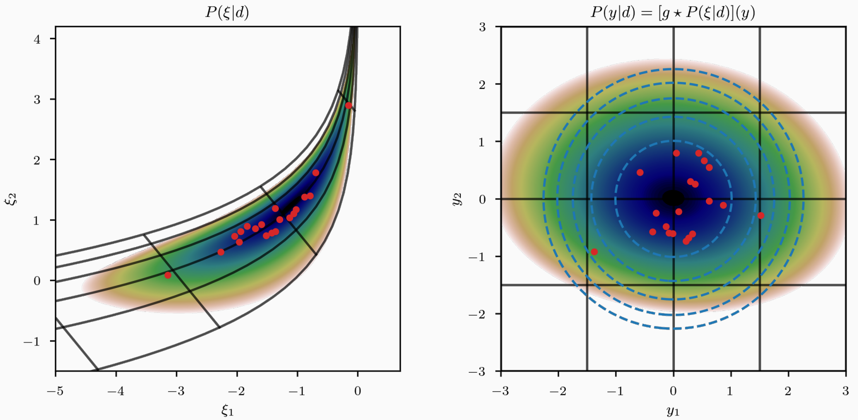
Geometric Variational Inference (geoVI) [FLE21]



Geometric Variational Inference (geoVI) [FLE21]



Geometric Variational Inference (geoVI) [FLE21]



Geometric Variational Inference (geoVI) [FLE21]

Local Euclidean isometry around $\bar{\xi}$

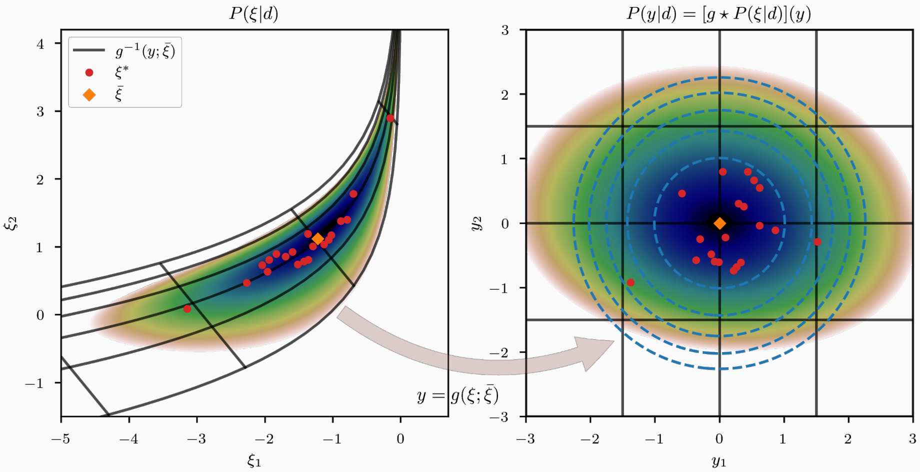
$$y = g(\xi; \bar{\xi}) = \mathcal{M}(\bar{\xi})^{-1/2} \left[\xi - \bar{\xi} + \left(\frac{\partial x}{\partial \xi} \right)^T \Bigg|_{\xi = \bar{\xi}} \left(x(\xi) - x(\bar{\xi}) \right) \right]$$

Likelihood transformation: $x(\xi) = x(s(\xi))$, expansion point: $\bar{\xi}$.

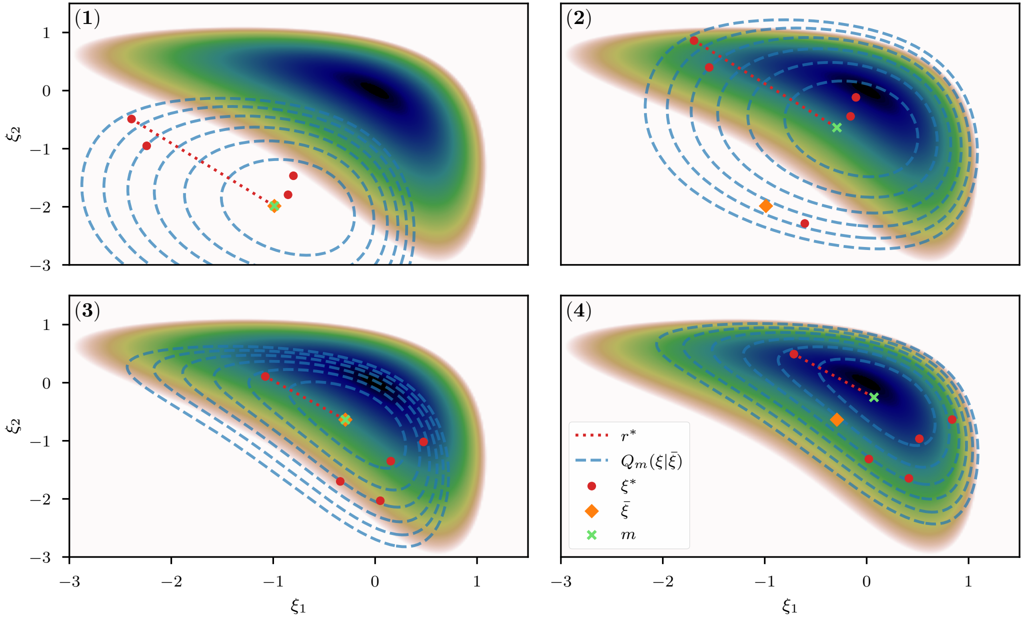
Variational approximation with transformed distribution $\mathcal{Q}$

$$\mathcal{Q}_{\bar{\xi}}(\xi) = \mathcal{N}(y|0, \mathbb{1})|_{y = g(\xi; \bar{\xi})} \left\| \frac{\partial g(\xi; \bar{\xi})}{\partial \xi} \right\|$$

Geometric Variational Inference (geoVI) [FLE21]

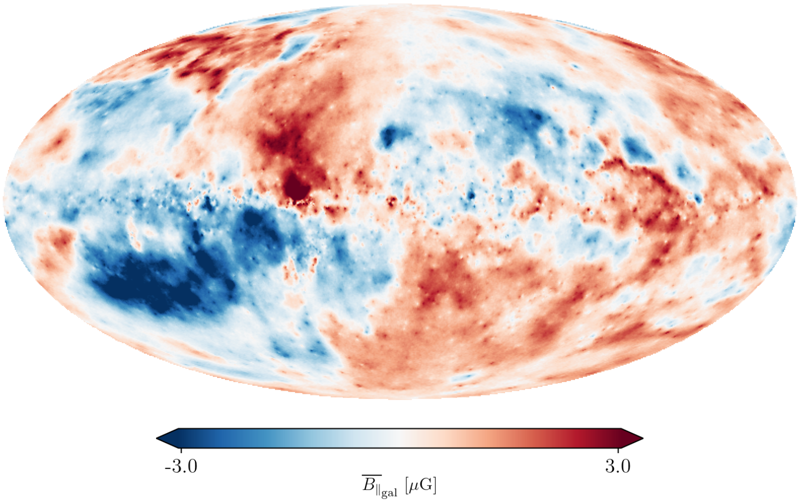


Geometric Variational Inference (geoVI) [FLE21]

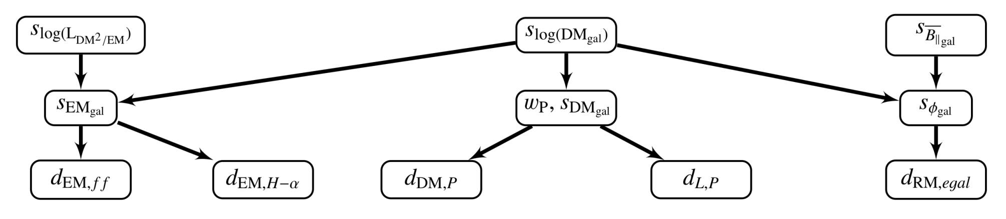


Applications

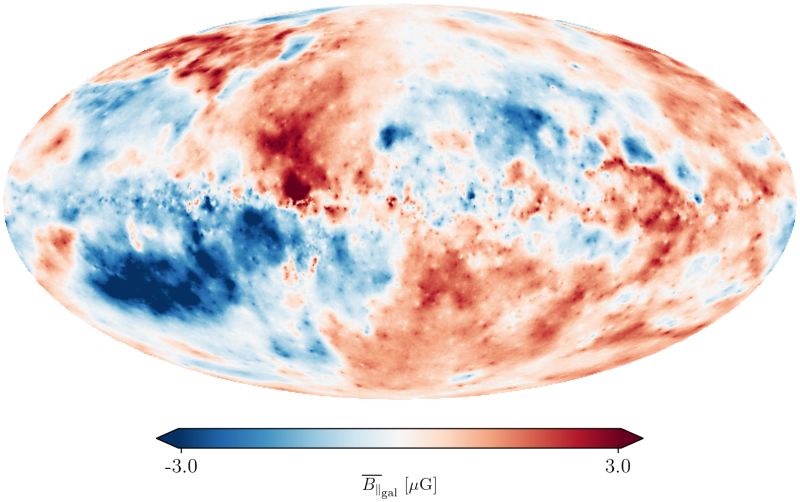
Faraday sky & LOS magnetic field [HHF+23]



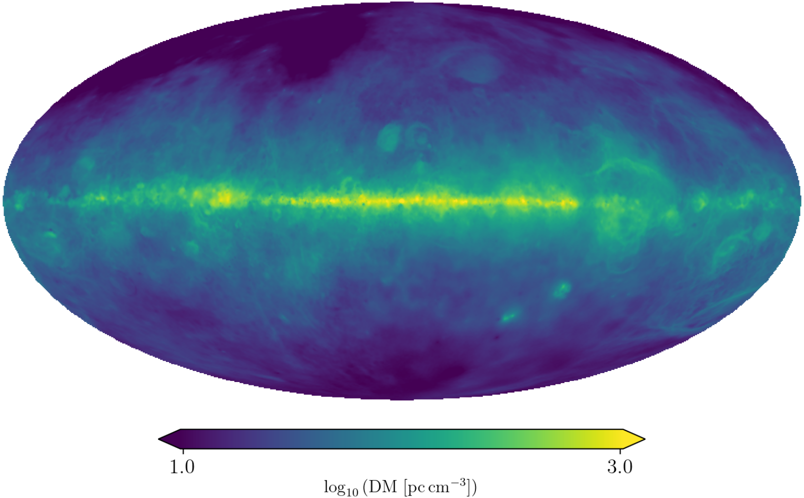
Faraday sky & LOS magnetic field [HHF+23]



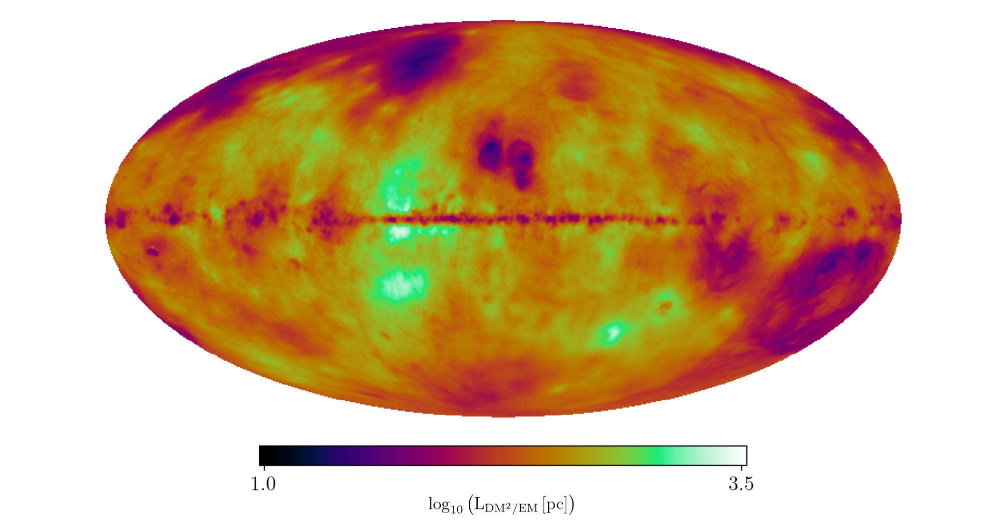
Faraday sky & LOS magnetic field [HHF+23]



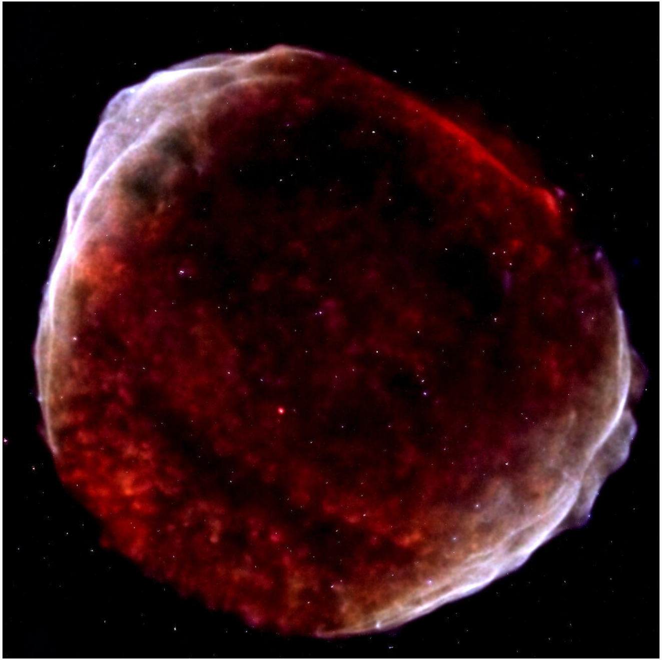
Faraday sky & LOS magnetic field [HHF+23]



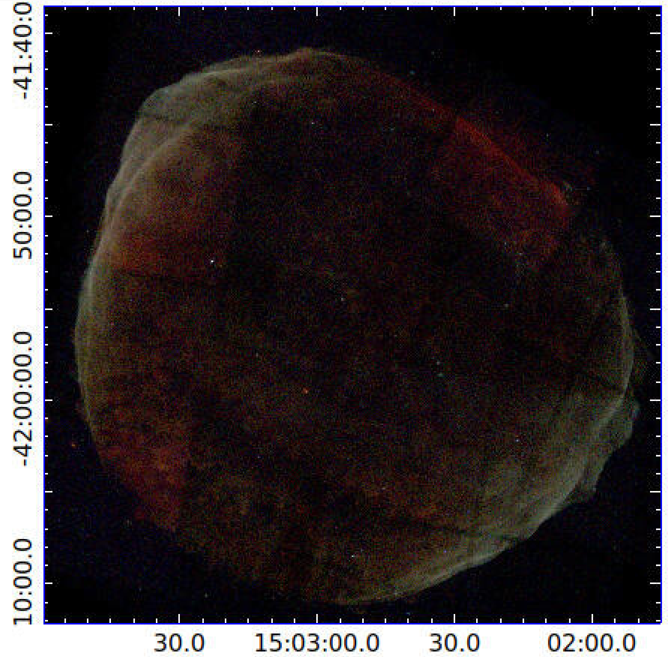
Faraday sky & LOS magnetic field [HHF+23]



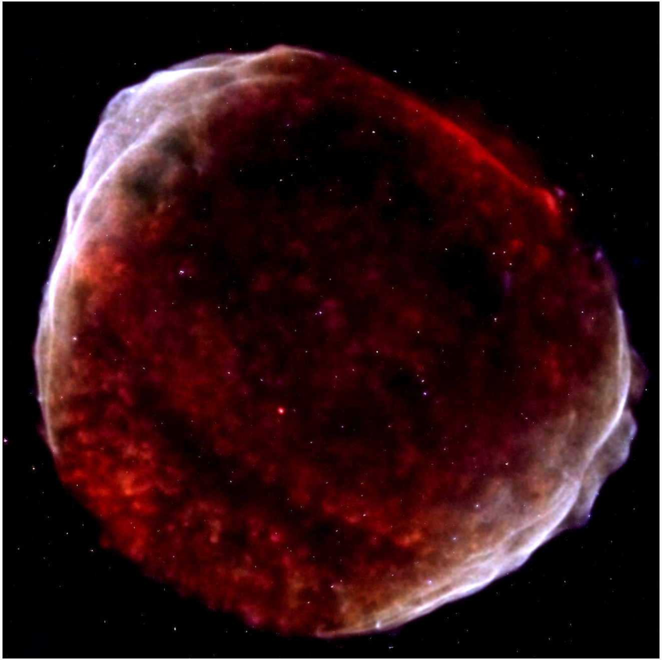
Chandra X-ray imaging [WEG+23]



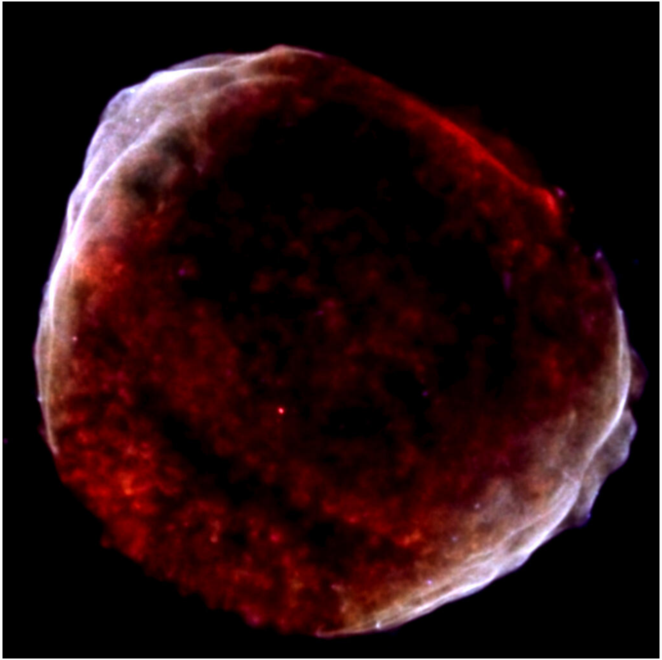
Chandra X-ray imaging [WEG+23]



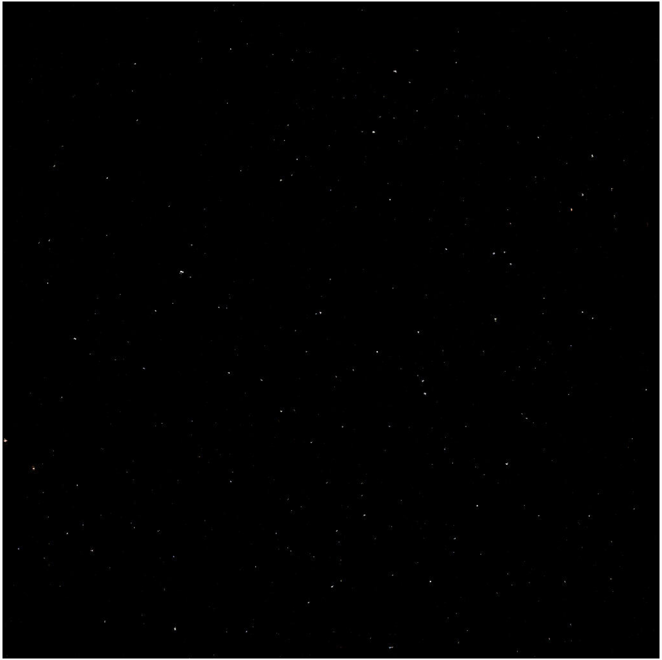
Chandra X-ray imaging [WEG+23]



Chandra X-ray imaging [WEG+23]

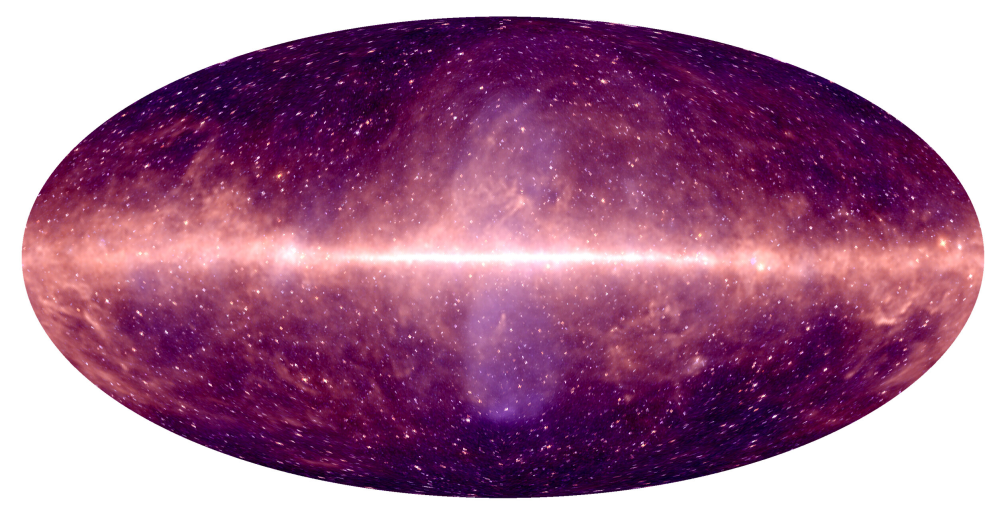


Chandra X-ray imaging [WEG+23]

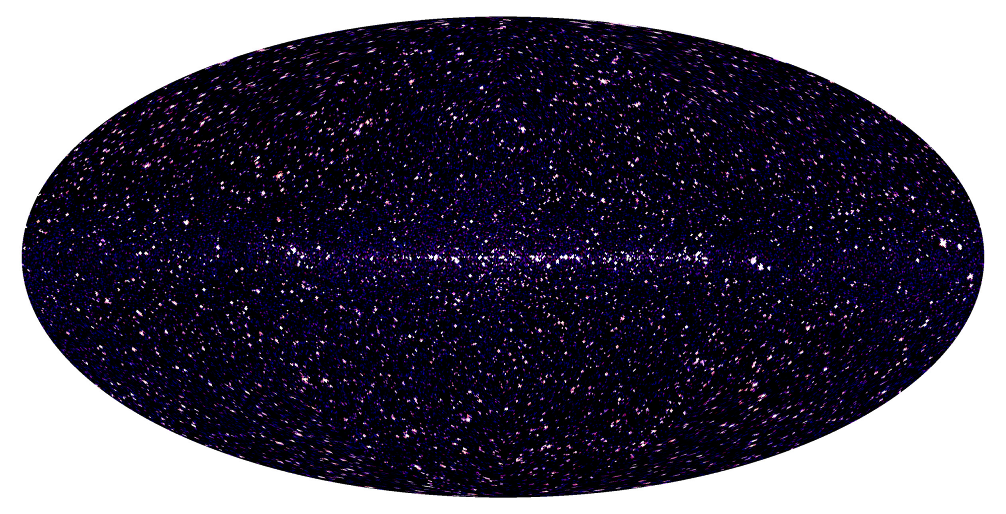


Fermi γ-ray sky

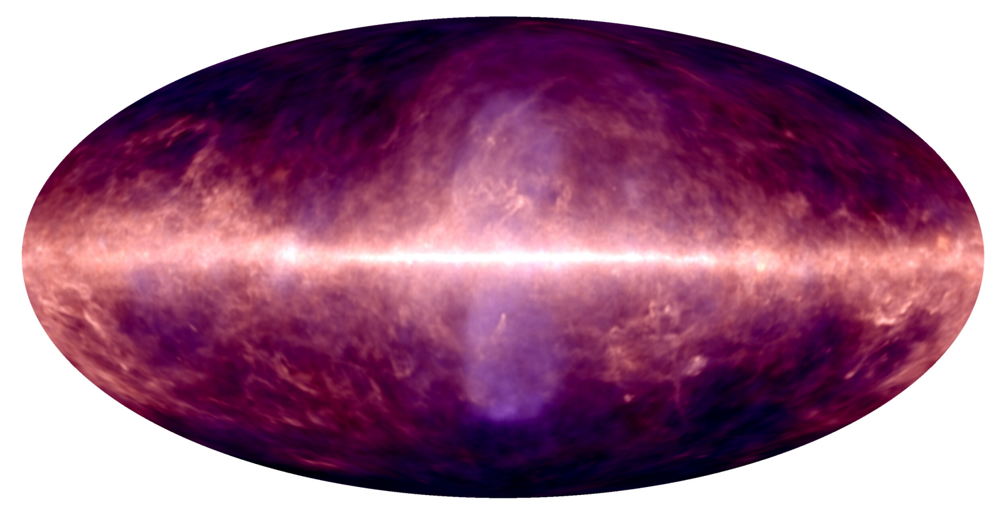
Fermi γ-ray sky



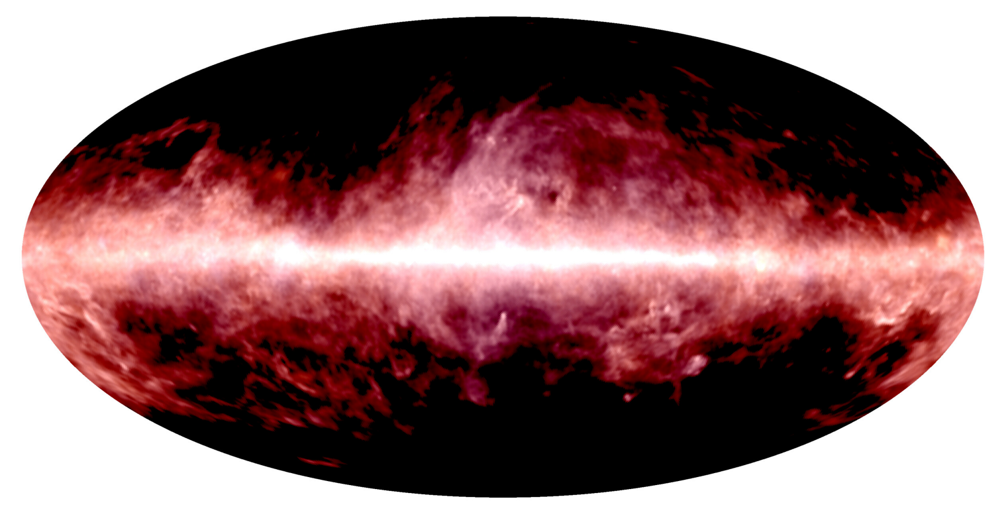
Fermi γ-ray sky



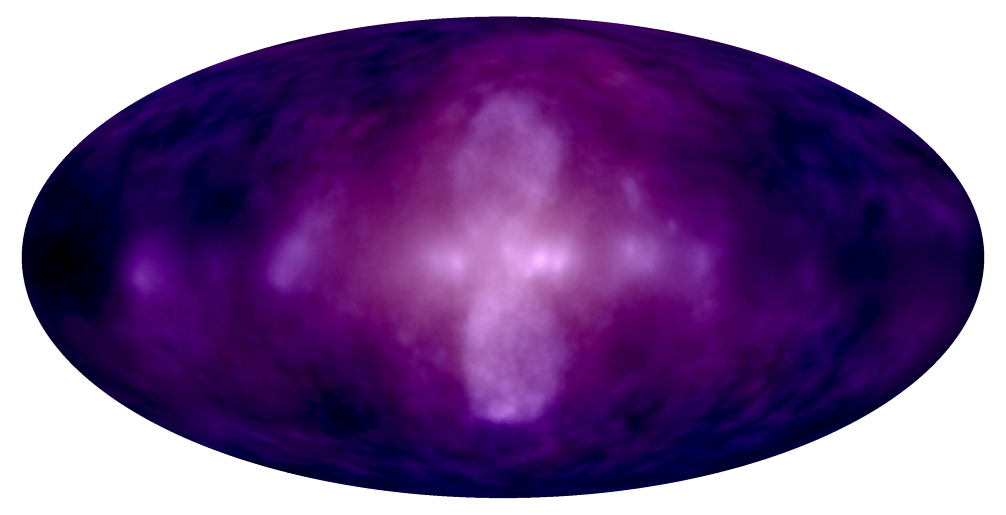
Fermi γ-ray sky



Fermi γ-ray sky



Fermi γ-ray sky



References

- Gordian Edenhofer, Catherine Zucker, Philipp Frank, Andrew K. Saydjari, Joshua S. Speagle, Douglas Finkbeiner, and Torsten Enßlin.
  A Parsec-Scale Galactic 3D Dust Map out to 1.25 kpc from the Sun.
  arXiv e-prints, page arXiv:2308.01295, August 2023.
- Philipp Frank, Reimar Leike, and Torsten A. Enßlin.
  Geometric variational inference.
  Entropy, 23(7), 2021.
- Sebastian Hutschenreuter, Marijke Haverkorn, Philipp Frank, Nergis C. Raycheva, and Torsten A. Enßlin.
  Disentangling the Faraday rotation sky.
  arXiv e-prints, page arXiv:2304.12350, April 2023.
- Margret Westerkamp, Vincent Eberle, Matteo Guardiani, Philipp Frank, Lukas Platz, Philipp Arras, Jakob Knollmüller, Julia Stadler, and Torsten Enßlin.
  First spatio-spectral Bayesian imaging of SN1006 in X-ray.
  arXiv e-prints, page arXiv:2308.09176, August 2023.

Appendix

Conclusion - Code

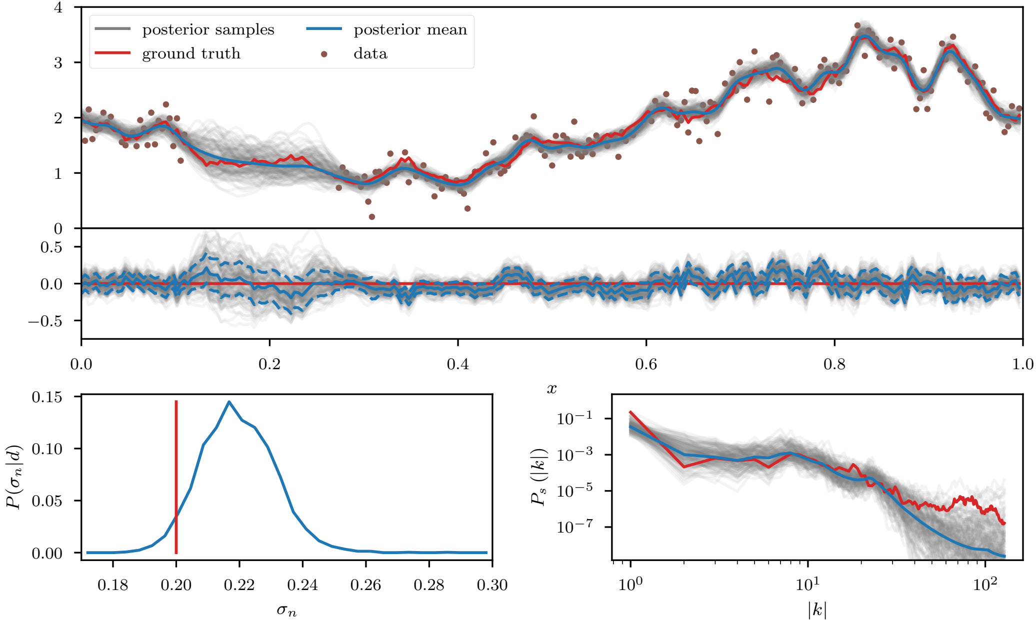
Numerical information field theory

https://gitlab.mpcdf.mpg.de/ift/nifty

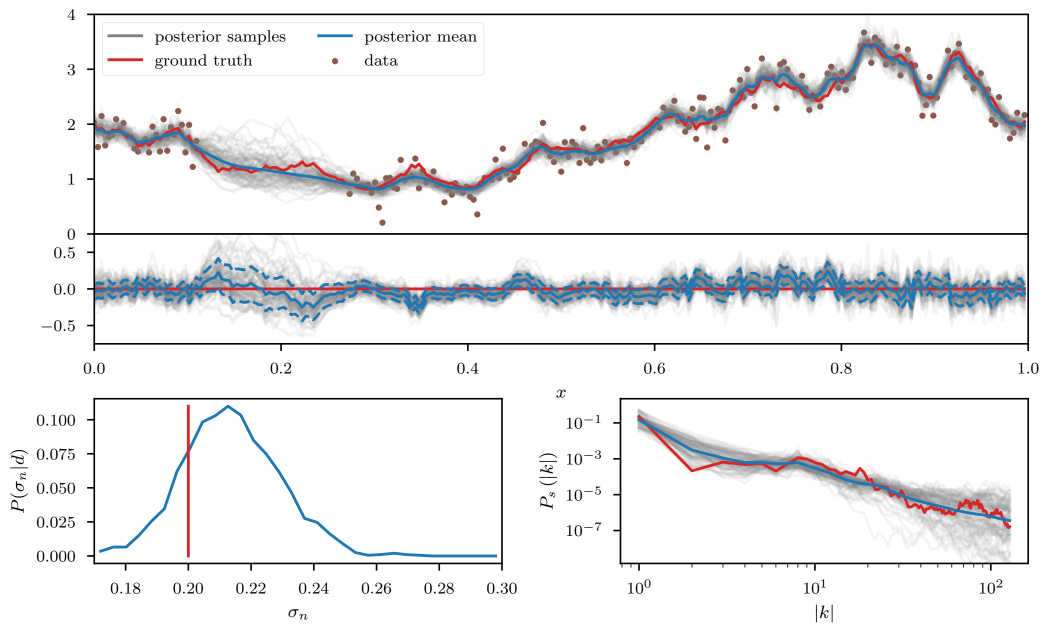
NIFTy tutorial - this afternoon!

https://gitlab.mpcdf.mpg.de/ift/tutorial_nifty_resolve

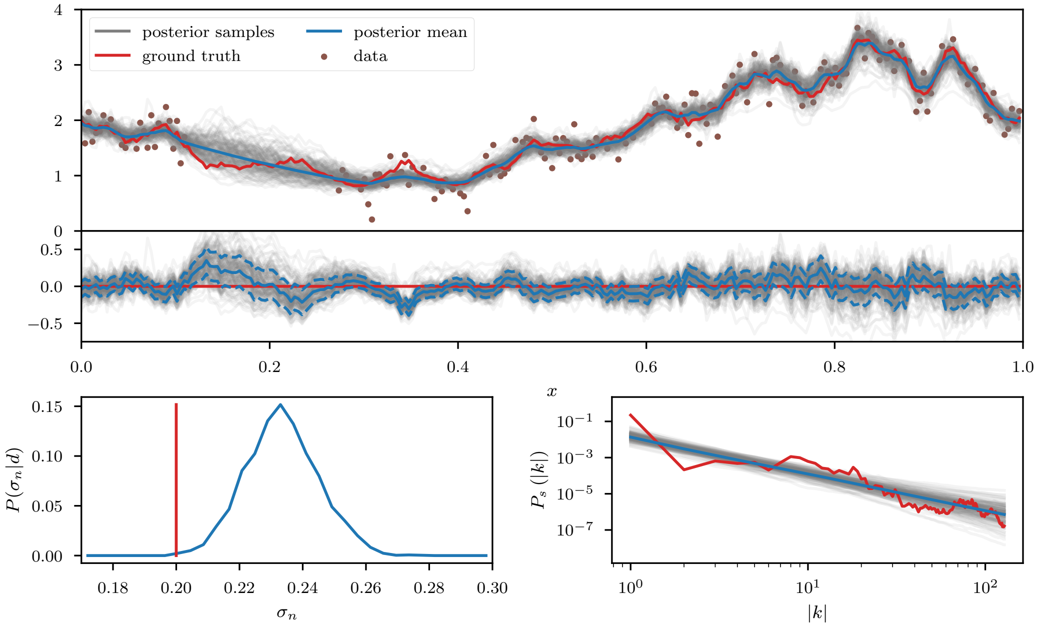
Applications



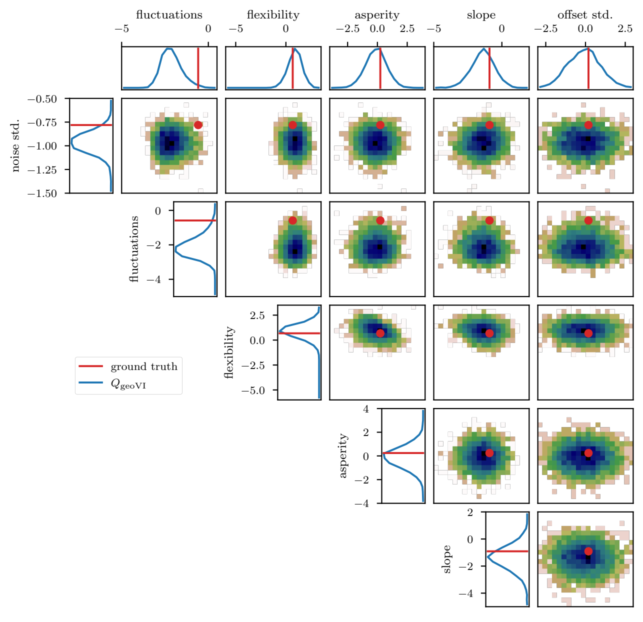
Applications



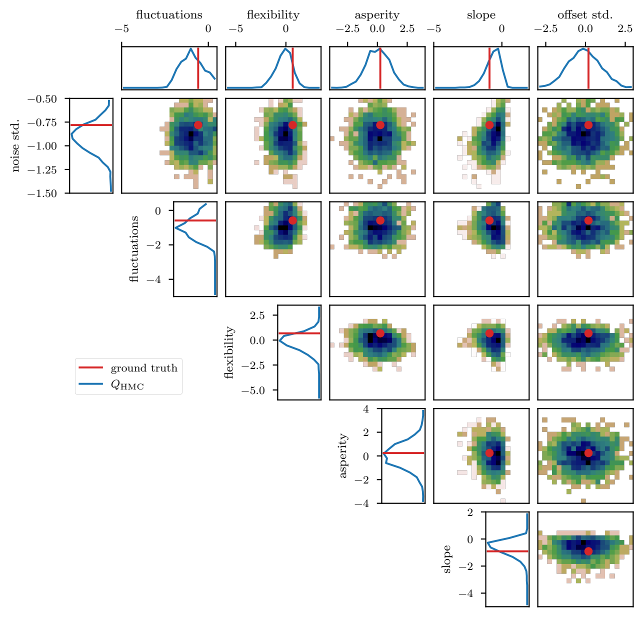
Applications



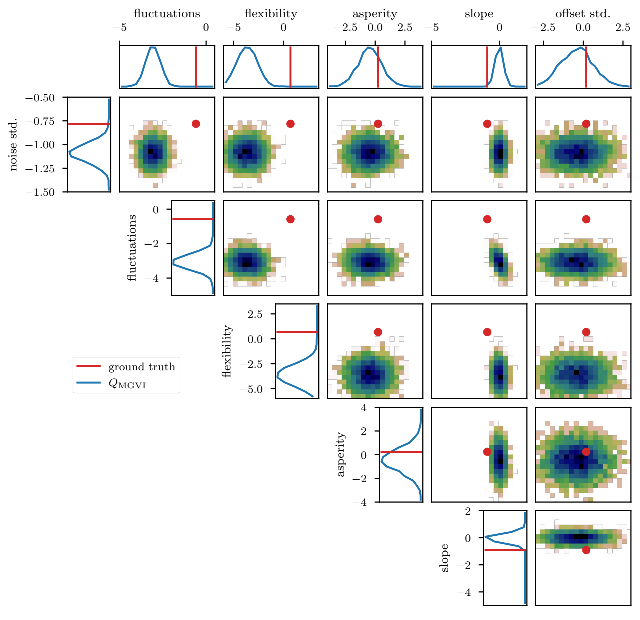
Applications



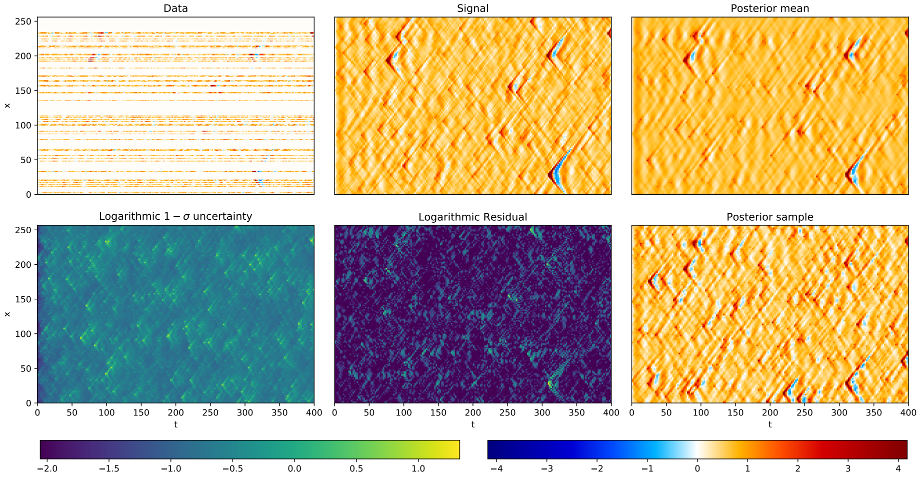
Applications



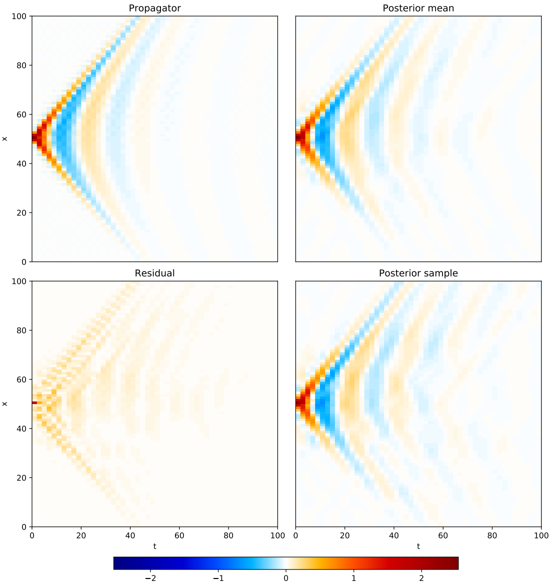
Applications



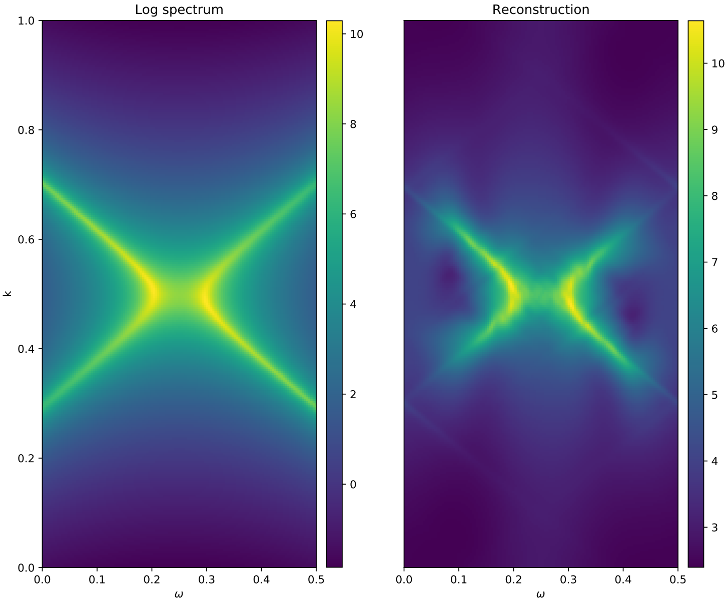
Application



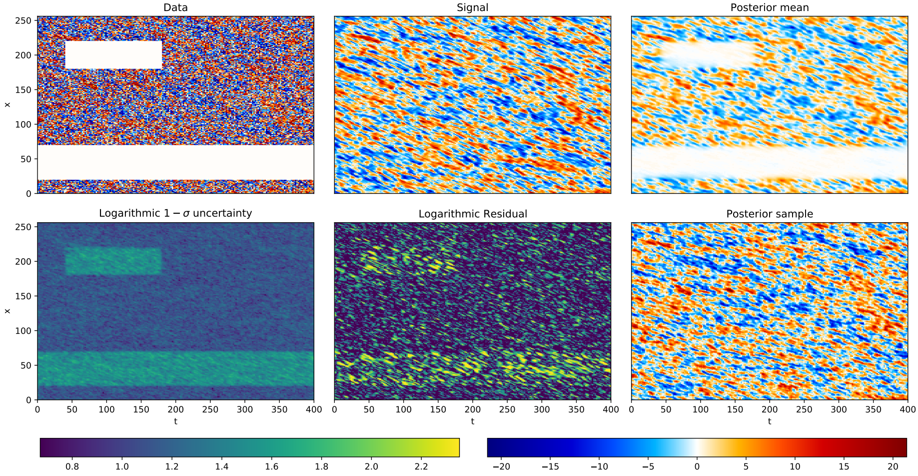
Application



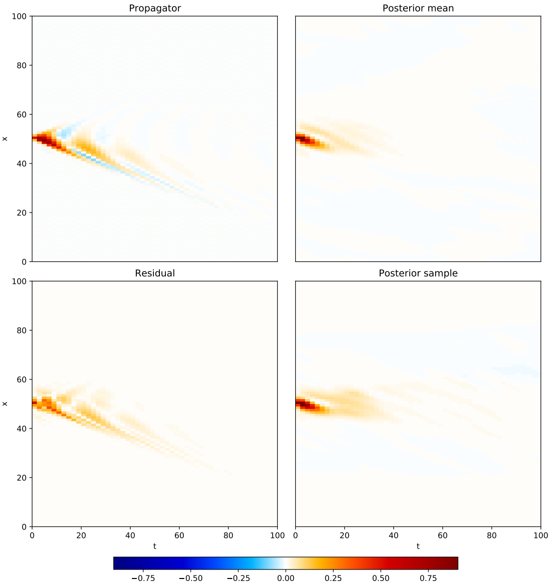
Application



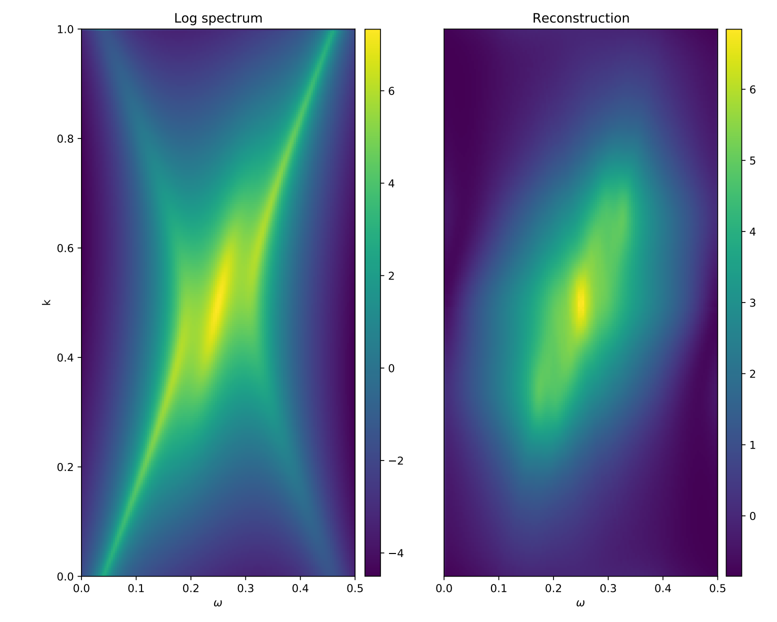
Application



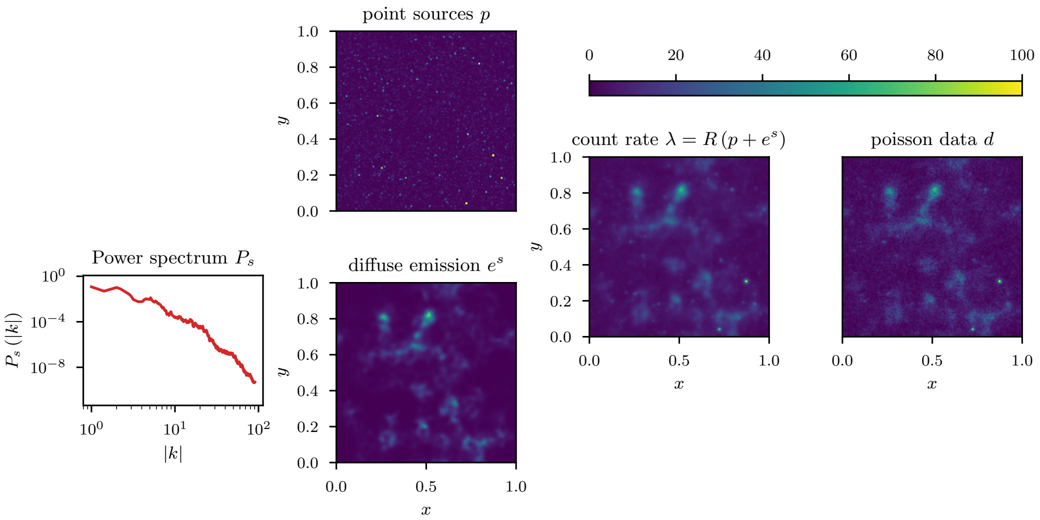
Application



Application

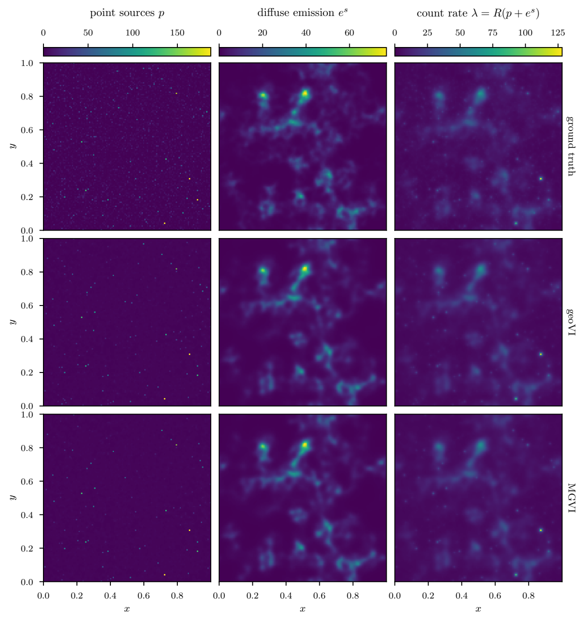


Appendix



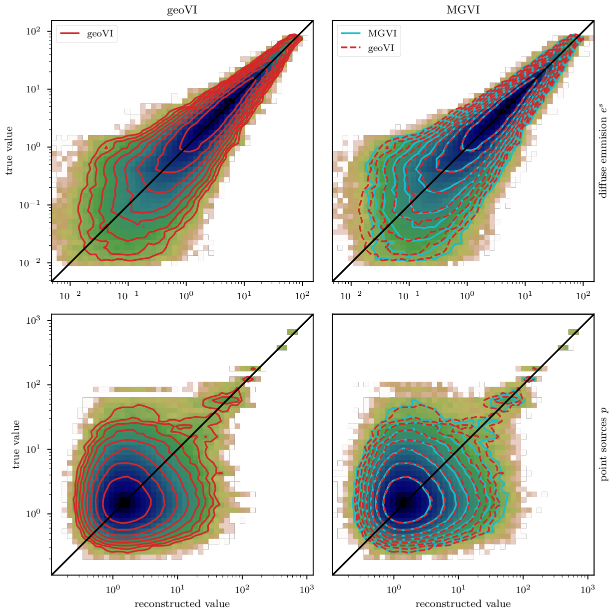
point sources p
1.0
0.8
0.6
0.4
0.2
0.0
y
0
20
40
60
80
100
count rate $\lambda = R\,(p + e^s)$
poisson data d
Power spectrum P_s
10^0
10^{-4}
10^{-8}
$P_s\,(|k|)$
10^0
10^1
10^2
$|k|$
diffuse emission e^s
0.0
0.5
1.0
x

Appendix



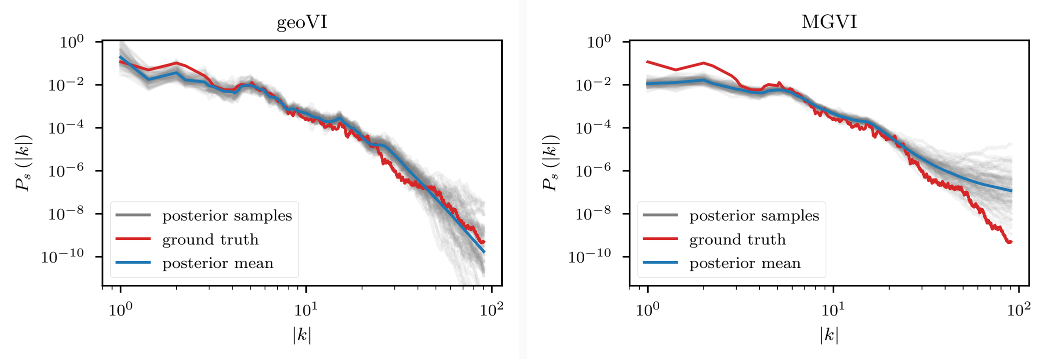
point sources p
diffuse emission e^s
count rate $\lambda = R(p + e^s)$
0 50 100 150
0 20 40 60
0 25 50 75 100 125
ground truth
geoVI
MGVI
y
x
0.0 0.2 0.4 0.6 0.8 1.0

Appendix

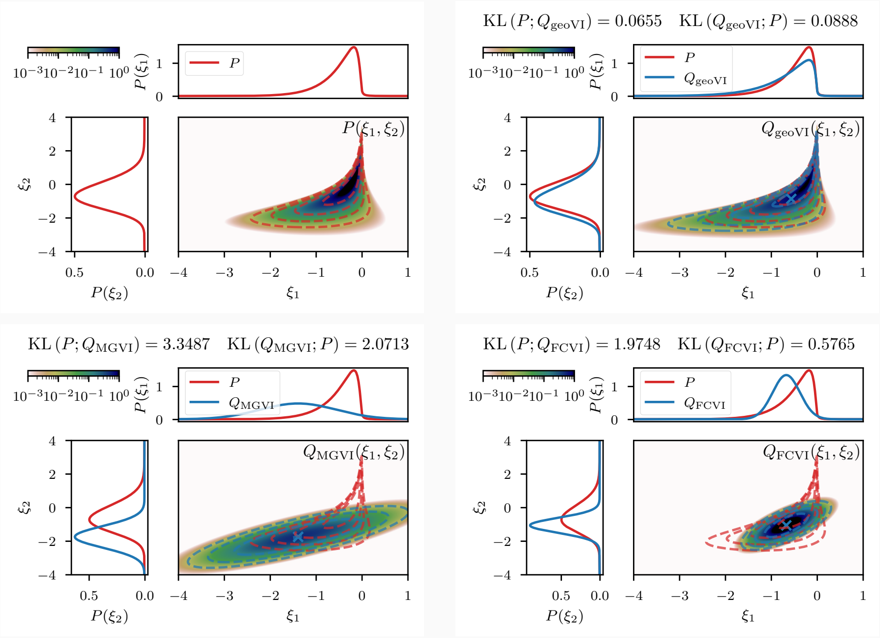


geoVI
MGVI
geoVI
MGVI
geoVI
true value
true value
diffuse emmision e^s
point sources p
reconstructed value
reconstructed value
10^{-2}
10^{-1}
10^0
10^1
10^2
10^3

Appendix

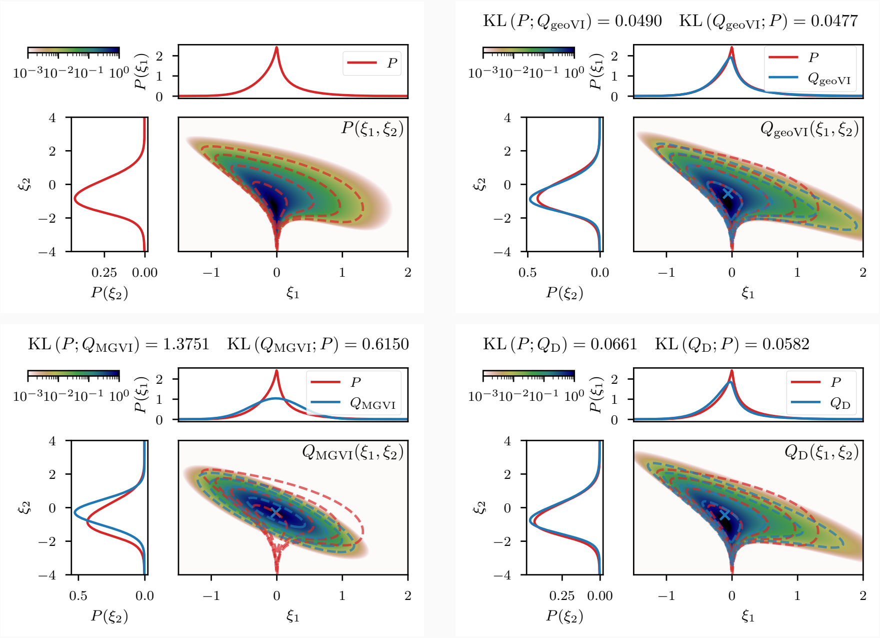

geoVI
MGVI
$P_s(|k|)$
$|k|$
10^0
10^{-2}
10^{-4}
10^{-6}
10^{-8}
10^{-10}
10^0
10^1
10^2
posterior samples
ground truth
posterior mean

Appendix

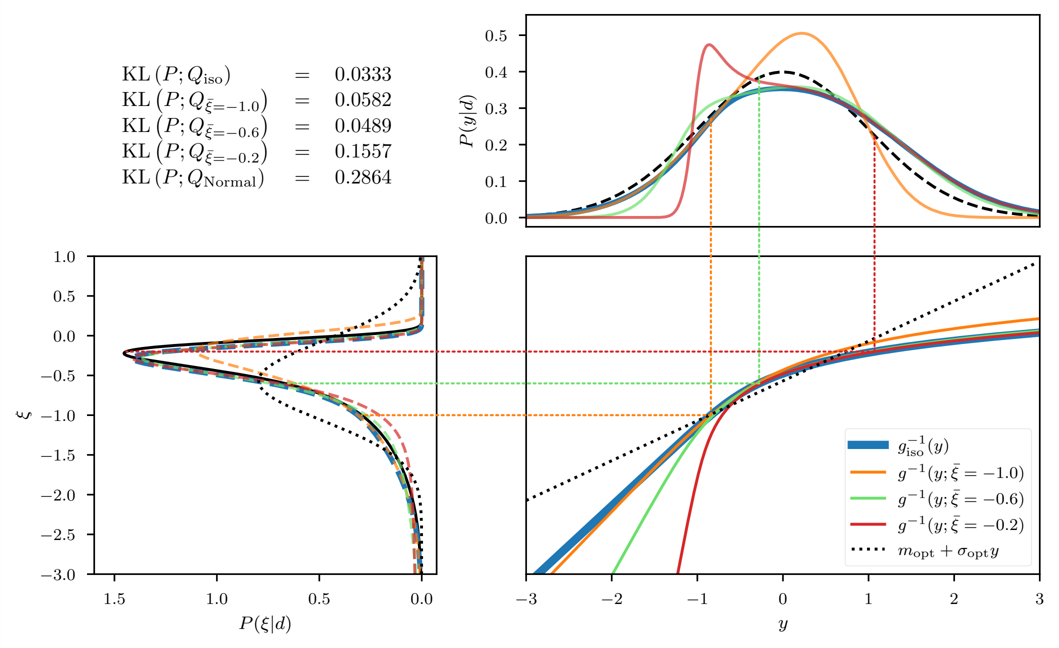


KL (P; Q_geoVI) = 0.0655   KL (Q_geoVI; P) = 0.0888
KL (P; Q_MGVI) = 3.3487   KL (Q_MGVI; P) = 2.0713
KL (P; Q_FCVI) = 1.9748   KL (Q_FCVI; P) = 0.5765
P(ξ1, ξ2)
Q_geoVI(ξ1, ξ2)
Q_MGVI(ξ1, ξ2)
Q_FCVI(ξ1, ξ2)
P
Q_geoVI
Q_MGVI
Q_FCVI
P(ξ1)
P(ξ2)
ξ1
ξ2

Appendix

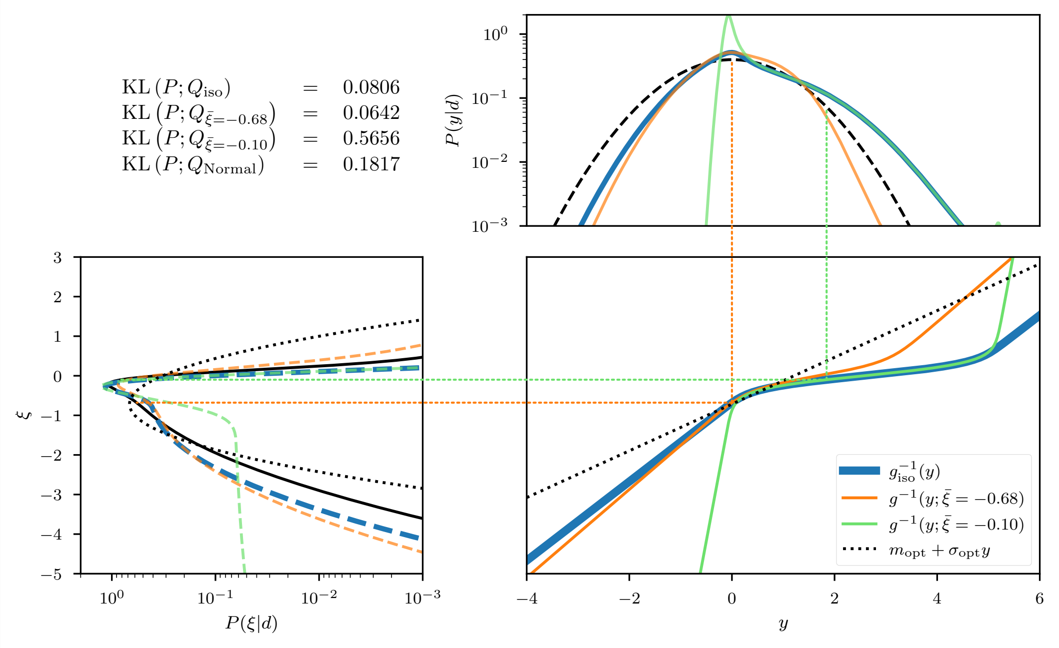

KL (P; Q_geoVI) = 0.0490 KL (Q_geoVI; P) = 0.0477
KL (P; Q_MGVI) = 1.3751 KL (Q_MGVI; P) = 0.6150
KL (P; Q_D) = 0.0661 KL (Q_D; P) = 0.0582
P(ξ1, ξ2)
Q_geoVI(ξ1, ξ2)
Q_MGVI(ξ1, ξ2)
Q_D(ξ1, ξ2)
P
Q_geoVI
Q_MGVI
Q_D
P(ξ1)
P(ξ2)
ξ1
ξ2
10^-3 10^-2 10^-1 10^0

Appendix



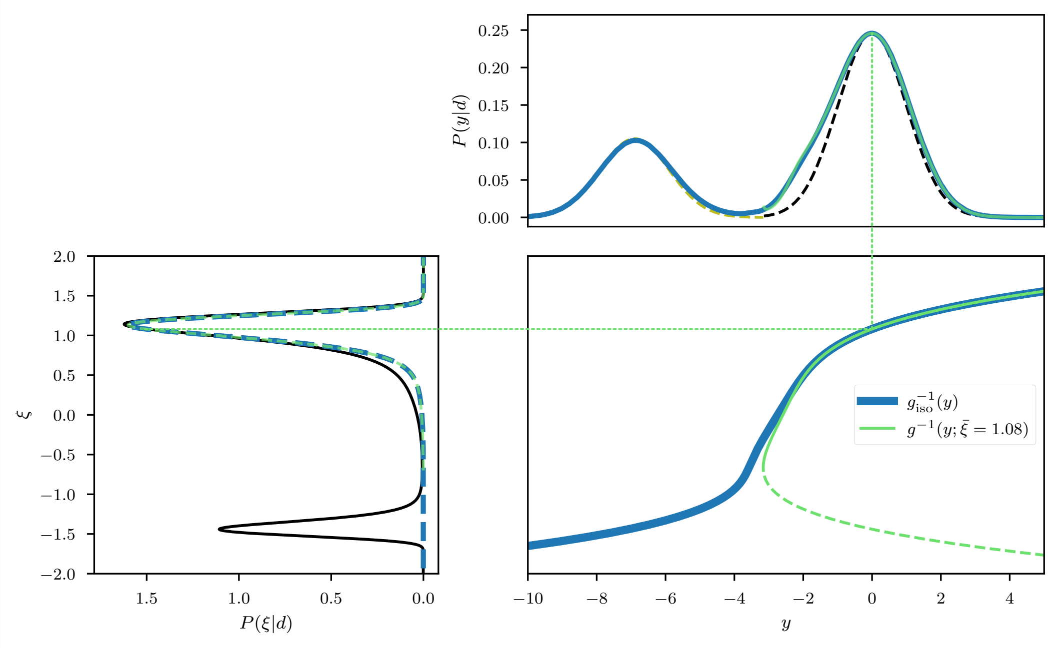
$\mathrm{KL}\left(P; Q_{\mathrm{iso}}\right) = 0.0333$
$\mathrm{KL}\left(P; Q_{\bar{\xi}=-1.0}\right) = 0.0582$
$\mathrm{KL}\left(P; Q_{\bar{\xi}=-0.6}\right) = 0.0489$
$\mathrm{KL}\left(P; Q_{\bar{\xi}=-0.2}\right) = 0.1557$
$\mathrm{KL}\left(P; Q_{\mathrm{Normal}}\right) = 0.2864$
$P(y|d)$
y
ξ
$P(\xi|d)$
$g_{\mathrm{iso}}^{-1}(y)$
$g^{-1}(y; \bar{\xi} = -1.0)$
$g^{-1}(y; \bar{\xi} = -0.6)$
$g^{-1}(y; \bar{\xi} = -0.2)$
$m_{\mathrm{opt}} + \sigma_{\mathrm{opt}} y$

Appendix



$\mathrm{KL}(P; Q_{\mathrm{iso}}) = 0.0806$
$\mathrm{KL}(P; Q_{\bar{\xi}=-0.68}) = 0.0642$
$\mathrm{KL}(P; Q_{\bar{\xi}=-0.10}) = 0.5656$
$\mathrm{KL}(P; Q_{\mathrm{Normal}}) = 0.1817$
$P(y|d)$
10^0
10^{-1}
10^{-2}
10^{-3}
ξ
3
2
1
0
−1
−2
−3
−4
−5
$P(\xi|d)$
y
−4
−2
0
2
4
6
$g_{\mathrm{iso}}^{-1}(y)$
$g^{-1}(y; \bar{\xi} = -0.68)$
$g^{-1}(y; \bar{\xi} = -0.10)$
$m_{\mathrm{opt}} + \sigma_{\mathrm{opt}} y$

Appendix



0.25
0.20
0.15
0.10
0.05
0.00
P(y|d)
2.0
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
−2.0
ξ
1.5
1.0
0.5
0.0
P(ξ|d)
−10
−8
−6
−4
−2
0
2
4
y
$g_{\mathrm{iso}}^{-1}(y)$
$g^{-1}(y; \bar{\xi} = 1.08)$