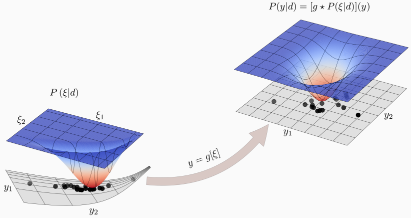
Geometric Variational Inference



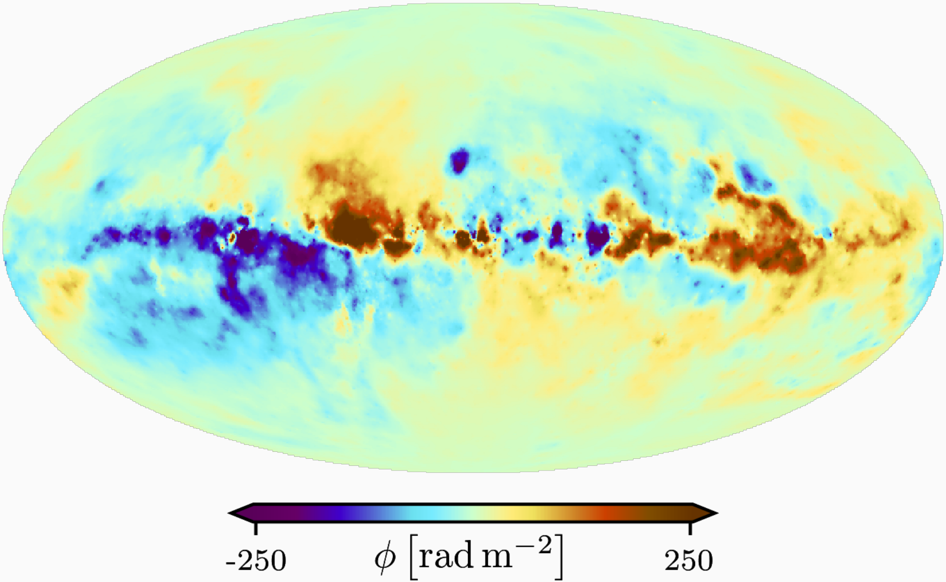
Using Riemannian geometry to enhance variational methods

Philipp Frank

September 30, 2021

Max-Planck Institute for Astrophysics, Garching, Germany

Imaging problems



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

Imaging problems

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

Imaging problems

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

Imaging problems:

1. Very high dimensional (in both, the parameter space and the data space)

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

Imaging problems:

1. Very high dimensional (in both, the parameter space and the data space)
   1.1 No explicit representation of matrices (e.g. covariance, curvature)
   1.2 Inference algorithm must scale (almost) linear with problem size

Imaging problems

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

Imaging problems:

1. Very high dimensional (in both, the parameter space and the data space)
   1.1 No explicit representation of matrices (e.g. covariance, curvature)
   1.2 Inference algorithm must scale (almost) linear with problem size
2. Non-linear (non-Gaussian) posterior distributions
   2.1 Point estimates (e.g. MAP) not sufficient

Variational Inference

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)

Kullback-Leibler divergence

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

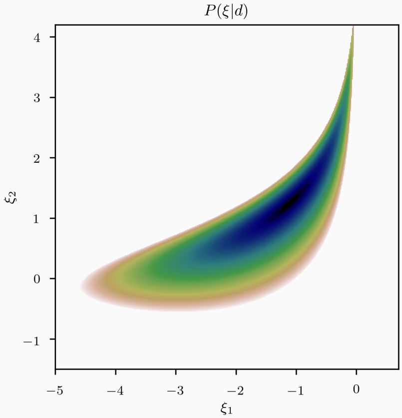
Posterior: $\mathcal{P}(\xi|d)$; approximation: $\mathcal{Q}_\sigma(\xi)$; variational parameters: σ.

Use a transformed standard normal distribution for $\mathcal{Q}$: $\mathcal{Q}(y) = \mathcal{N}(y; 0, \mathbb{1})$

Choose a coordinate system $y = g(\xi)$ such that the *posterior* distribution is close to a standard Normal distribution.

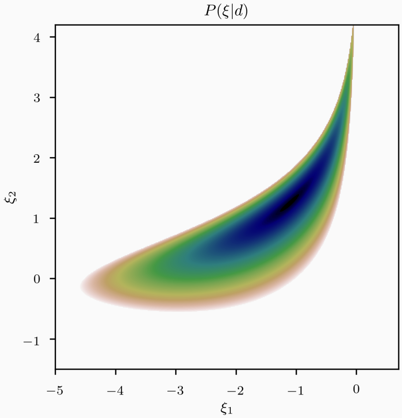
Geometric Variational Inference

Geometric Variational Inference (geoVI)



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

Geometric Variational Inference (geoVI)

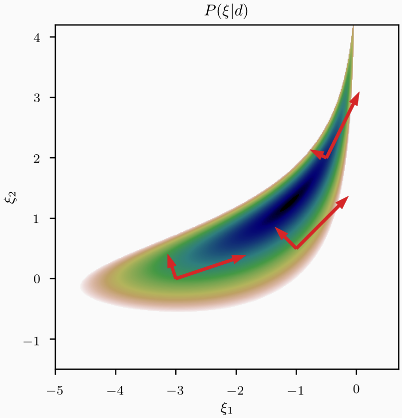


Information Hamiltonian $\mathcal{H}(\xi|d)$: $-\log\left(\mathcal{P}(\xi|d)\right)$

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

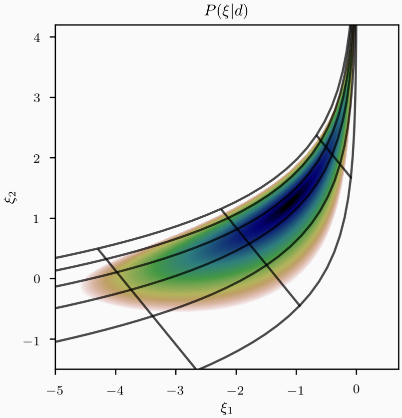
Fisher information metric $\mathcal{M}_{\mathrm{lh}}(\xi)$: $\left\langle \frac{\partial^2 \mathcal{H}(d|\xi)}{\partial\xi\partial\xi'} \right\rangle_{\mathcal{P}(d|\xi)}$

Geometric Variational Inference (geoVI)



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)

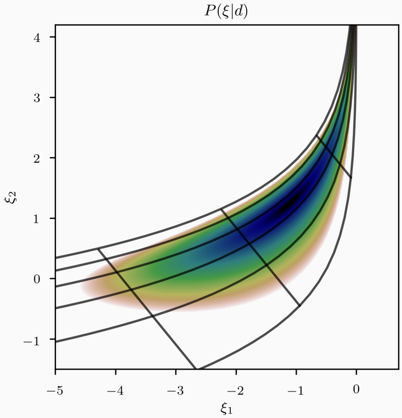


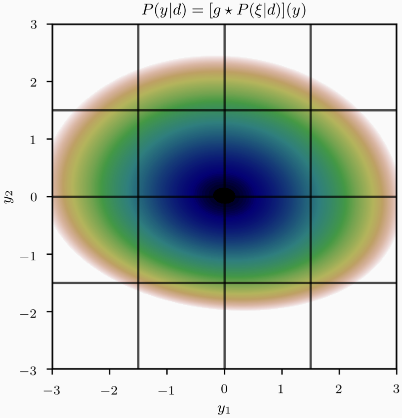
Information Hamiltonian $\mathcal{H}(\xi|d)$: $-\log\left(\mathcal{P}(\xi|d)\right)$

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

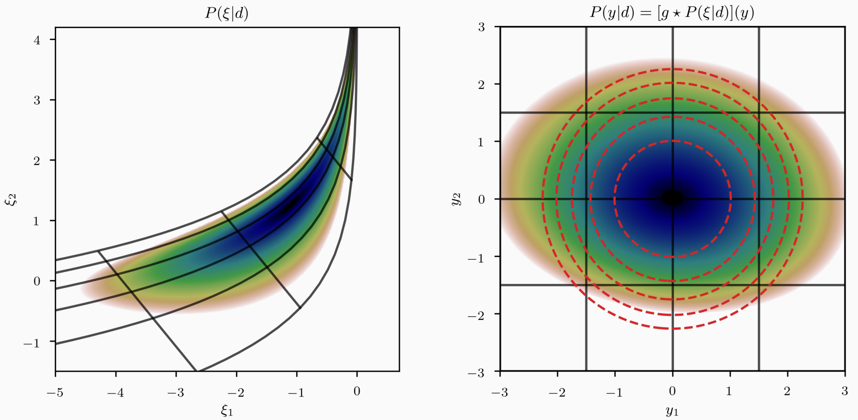
Fisher information metric $\mathcal{M}_{\mathrm{lh}}(\xi)$: $\left\langle \frac{\partial^2 \mathcal{H}(d|\xi)}{\partial\xi\partial\xi'} \right\rangle_{\mathcal{P}(d|\xi)}$

Geometric Variational Inference (geoVI)

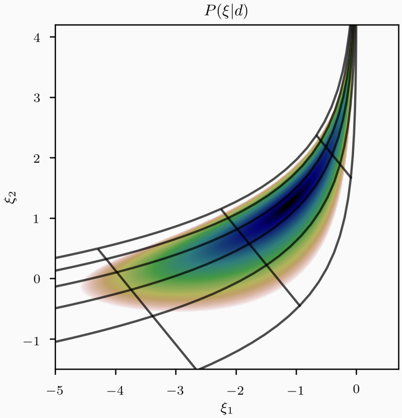


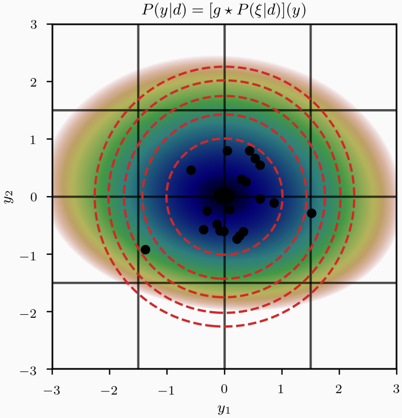


Geometric Variational Inference (geoVI)

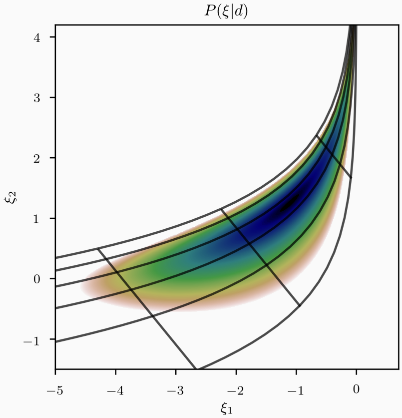


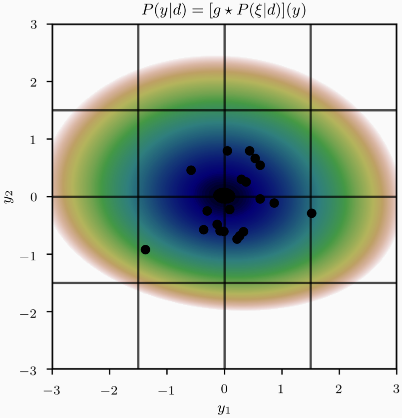
Geometric Variational Inference (geoVI)



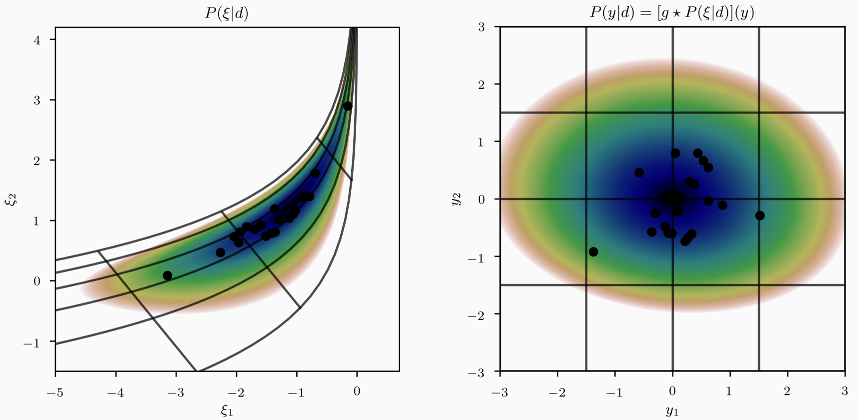


Geometric Variational Inference (geoVI)





Geometric Variational Inference (geoVI)



Geometric Variational Inference (geoVI)

- Write Fisher metric as pullback: $\mathcal{M}(\xi) = \mathcal{M}_{\mathrm{lh}}(\xi) + \mathbb{1} = \left(\frac{\partial x}{\partial \xi}\right)^T \frac{\partial x}{\partial \xi} + \mathbb{1}$.
- Find coordinate transformation g such that $\mathcal{M}(\xi) \approx \left(\frac{\partial g}{\partial \xi}\right)^T \frac{\partial g}{\partial \xi}$

Geometric Variational Inference (geoVI)

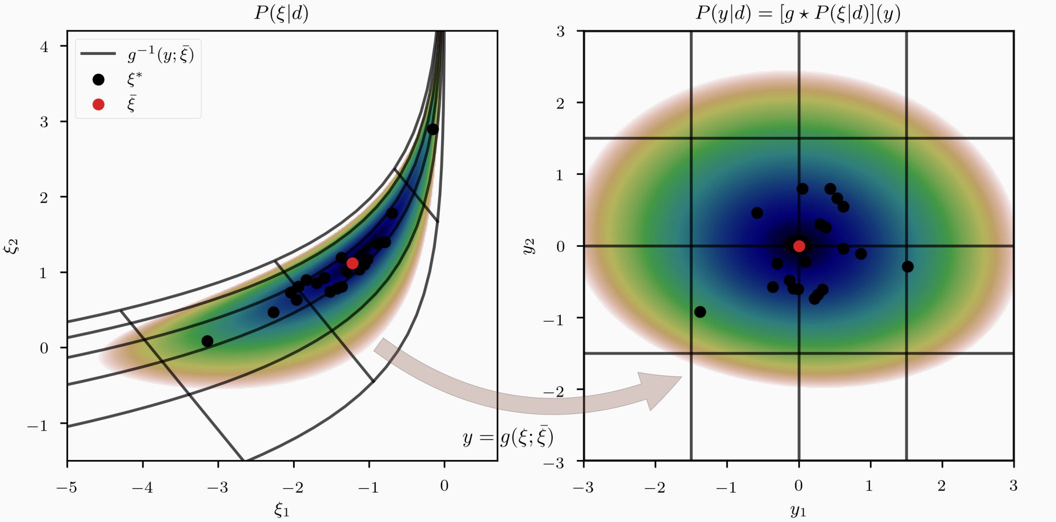
- Write Fisher metric as pullback: $\mathcal{M}(\xi) = \mathcal{M}_{\mathrm{lh}}(\xi) + \mathbb{1} = \left(\frac{\partial x}{\partial \xi}\right)^T \frac{\partial x}{\partial \xi} + \mathbb{1}$.
- Find coordinate transformation g such that $\mathcal{M}(\xi) \approx \left(\frac{\partial g}{\partial \xi}\right)^T \frac{\partial g}{\partial \xi}$

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

$$y = g(\xi; \bar{\xi}) = \xi - \bar{\xi} + \left(\frac{\partial x}{\partial \xi}\right)^T \Bigg|_{\xi=\bar{\xi}} \left(x(\xi) - x(\bar{\xi})\right)$$

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

Geometric Variational Inference (geoVI)



Geometric Variational Inference (geoVI)

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

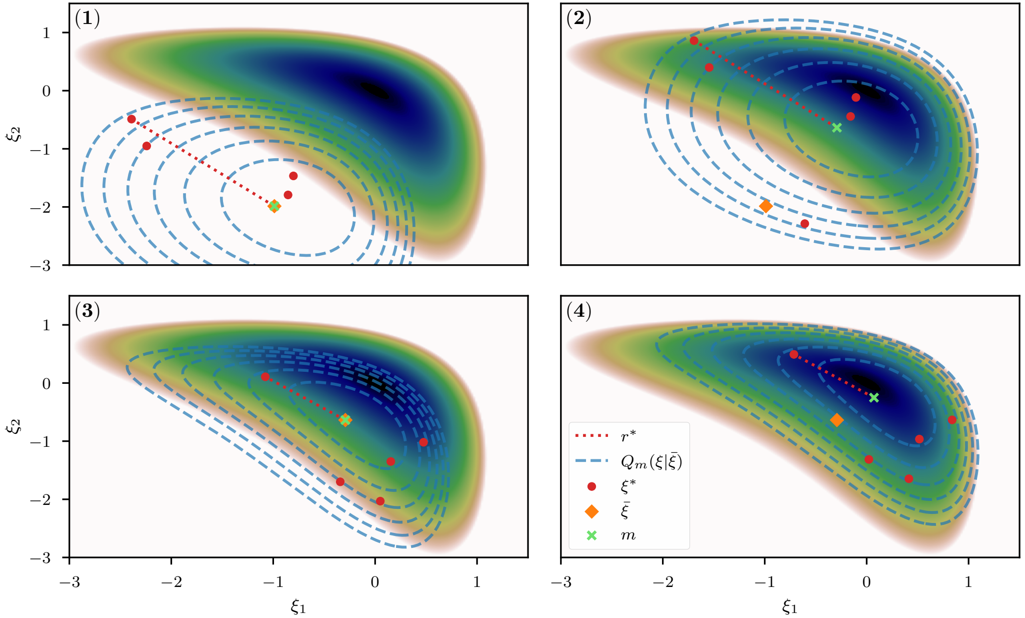
$$y = g(\xi; \bar{\xi}) = \xi - \bar{\xi} + \left(\frac{\partial x}{\partial \xi}\right)^T \Bigg|_{\xi=\bar{\xi}} \left(x(\xi) - x(\bar{\xi})\right)$$

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

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

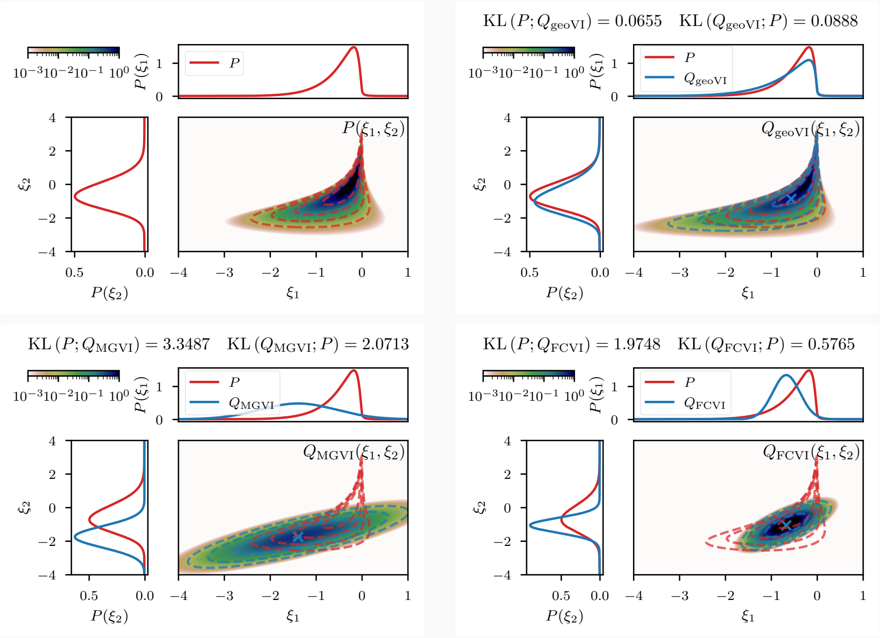
$$\mathcal{Q}_{\bar{\xi}}(\xi) = \mathcal{N}(y|0, \mathbb{1})|_{y=g(\xi;\bar{\xi})} \left|\left|\frac{\partial g(\xi; \bar{\xi})}{\partial \xi}\right|\right|$$

Geometric Variational Inference (geoVI)

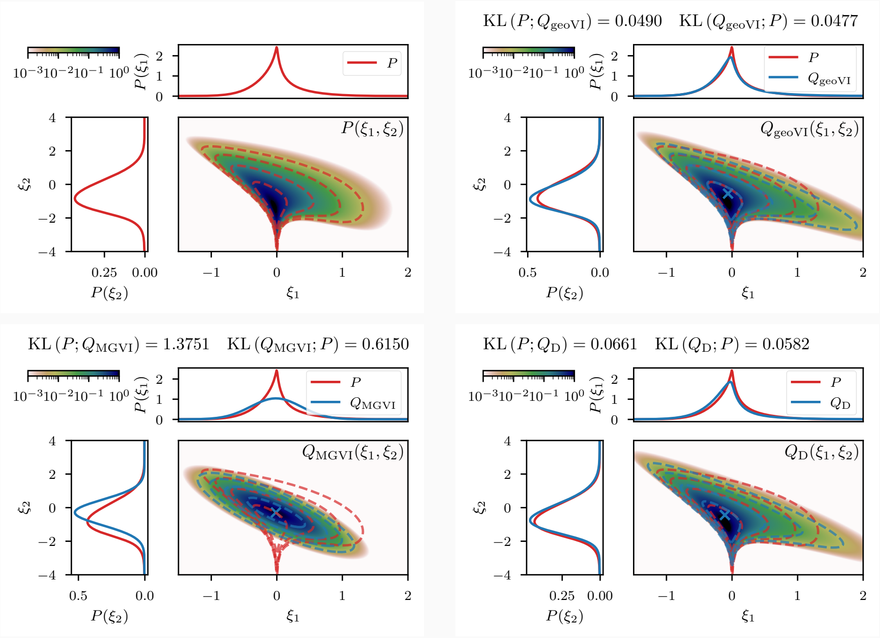


Examples

Examples

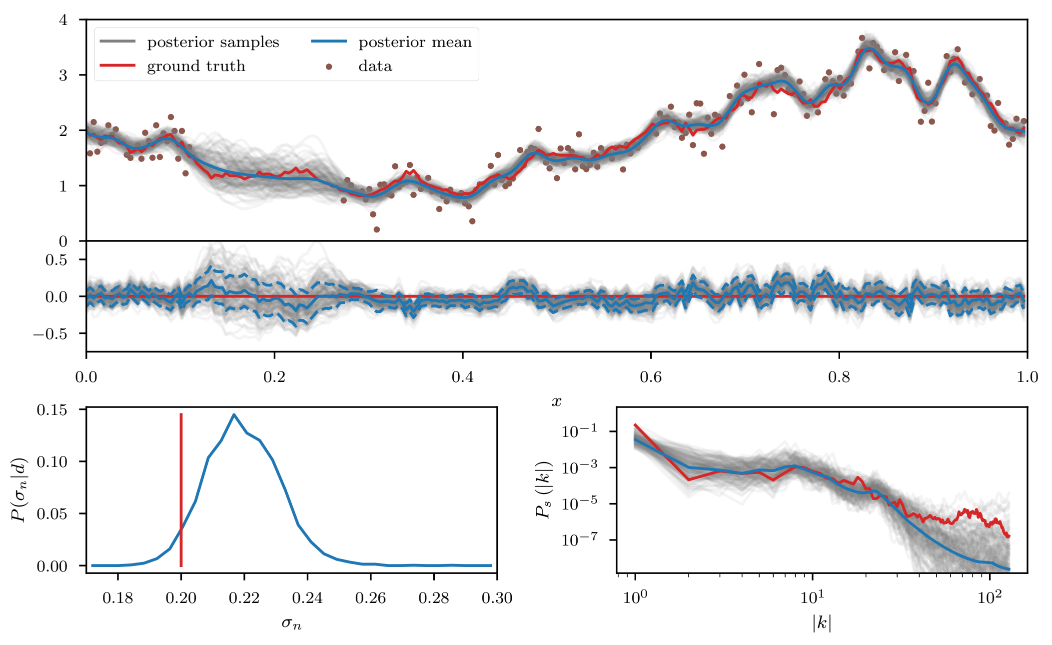

KL $(P; Q_{\mathrm{geoVI}}) = 0.0655$ KL $(Q_{\mathrm{geoVI}}; P) = 0.0888$
KL $(P; Q_{\mathrm{MGVI}}) = 3.3487$ KL $(Q_{\mathrm{MGVI}}; P) = 2.0713$
KL $(P; Q_{\mathrm{FCVI}}) = 1.9748$ KL $(Q_{\mathrm{FCVI}}; P) = 0.5765$
$P(\xi_1, \xi_2)$
$Q_{\mathrm{geoVI}}(\xi_1, \xi_2)$
$Q_{\mathrm{MGVI}}(\xi_1, \xi_2)$
$Q_{\mathrm{FCVI}}(\xi_1, \xi_2)$
P
Q_{geoVI}
Q_{MGVI}
Q_{FCVI}
$P(\xi_1)$
$P(\xi_2)$
ξ_1
ξ_2
10^{-3} 10^{-2} 10^{-1} 10^{0}

Examples



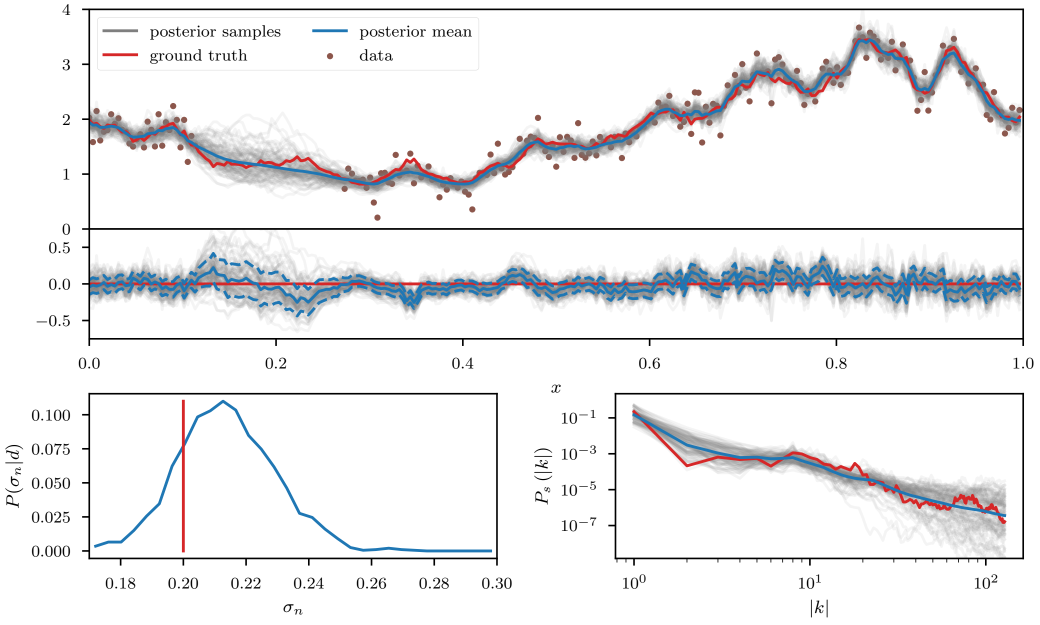
KL $(P; Q_{\text{geoVI}}) = 0.0490$ KL $(Q_{\text{geoVI}}; P) = 0.0477$
KL $(P; Q_{\text{MGVI}}) = 1.3751$ KL $(Q_{\text{MGVI}}; P) = 0.6150$
KL $(P; Q_{\text{D}}) = 0.0661$ KL $(Q_{\text{D}}; P) = 0.0582$
$P(\xi_1, \xi_2)$
$Q_{\text{geoVI}}(\xi_1, \xi_2)$
$Q_{\text{MGVI}}(\xi_1, \xi_2)$
$Q_{\text{D}}(\xi_1, \xi_2)$
$P(\xi_1)$
$P(\xi_2)$
ξ_1
ξ_2

Examples



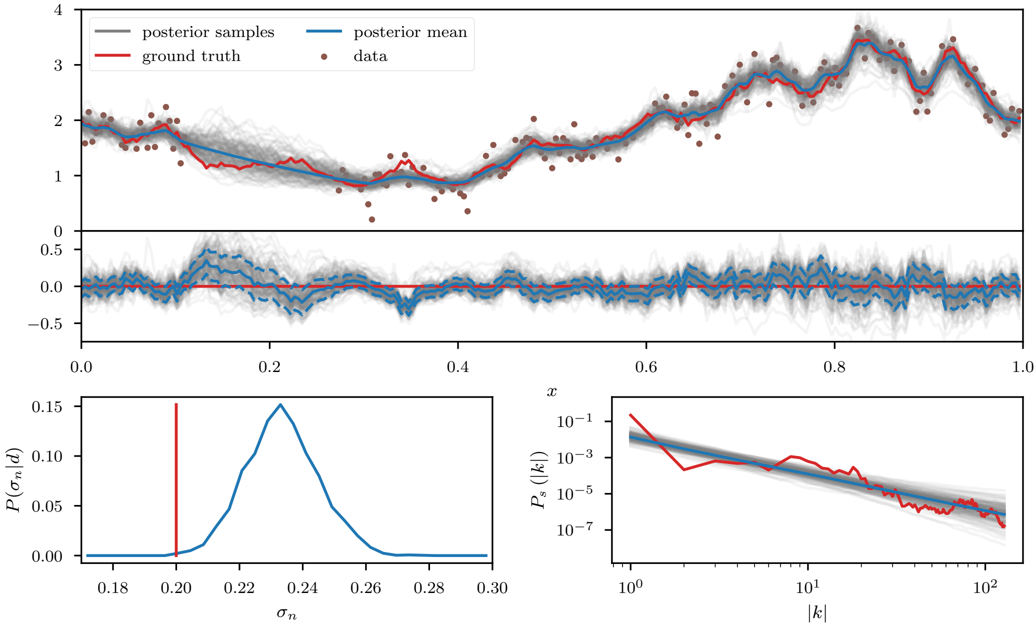
posterior samples
ground truth
posterior mean
data
4
3
2
1
0
0.5
0.0
−0.5
0.0
0.2
0.4
0.6
0.8
1.0
x
0.15
0.10
0.05
0.00
$P(\sigma_n|d)$
0.18
0.20
0.22
0.24
0.26
0.28
0.30
σ_n
10^{-1}
10^{-3}
10^{-5}
10^{-7}
$P_s(|k|)$
10^0
10^1
10^2
$|k|$

Examples



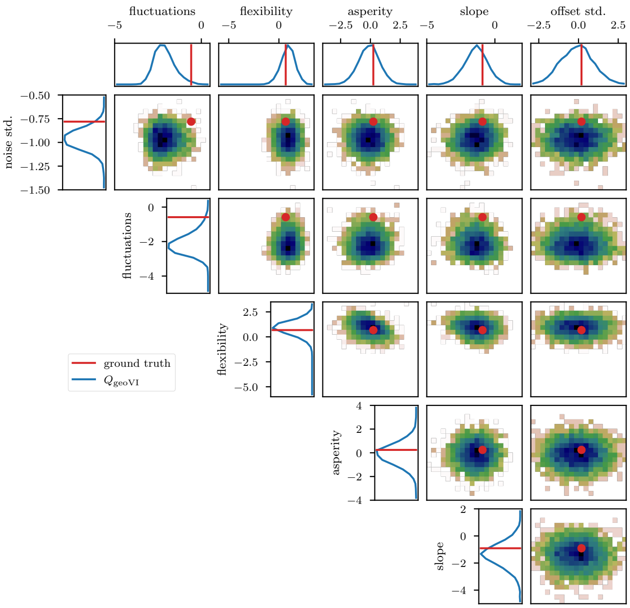
posterior samples
posterior mean
ground truth
data
x
$P(\sigma_n|d)$
σ_n
$P_s(|k|)$
$|k|$

Examples



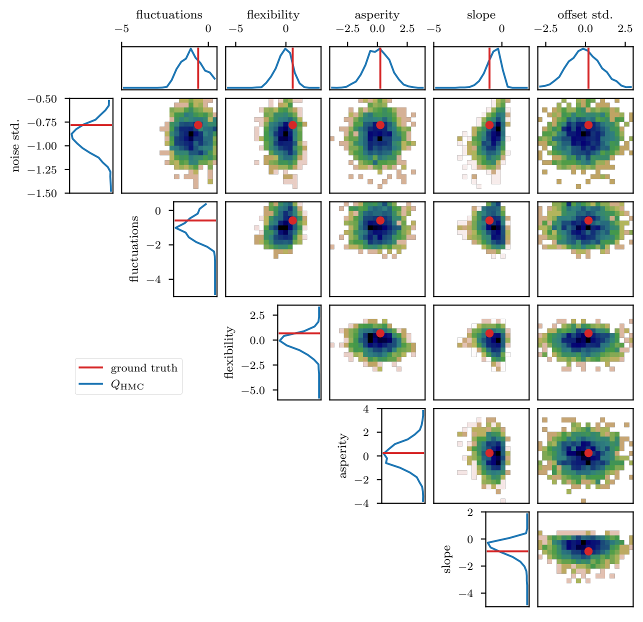
posterior samples
posterior mean
ground truth
data
4
3
2
1
0
0.5
0.0
−0.5
0.0
0.2
0.4
0.6
0.8
1.0
x
0.15
0.10
0.05
0.00
$P(\sigma_n|d)$
0.18
0.20
0.22
0.24
0.26
0.28
0.30
σ_n
10^{-1}
10^{-3}
10^{-5}
10^{-7}
$P_s(|k|)$
10^0
10^1
10^2
$|k|$

Examples



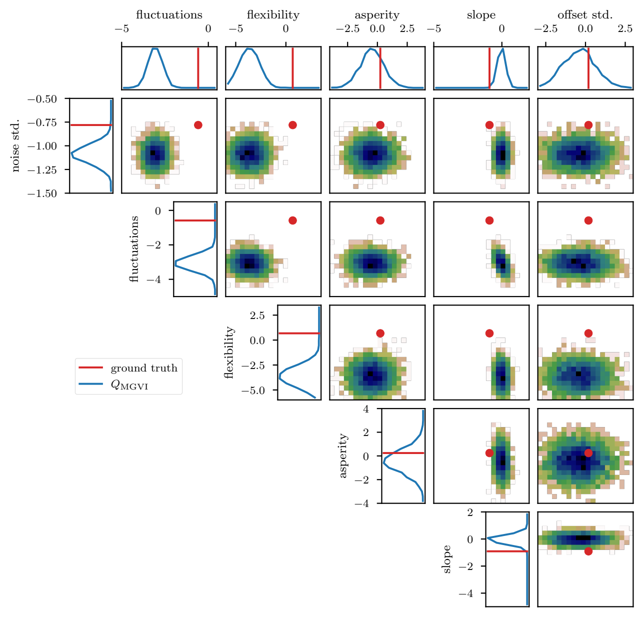
fluctuations
flexibility
asperity
slope
offset std.
noise std.
ground truth
Q_{geoVI}

Examples

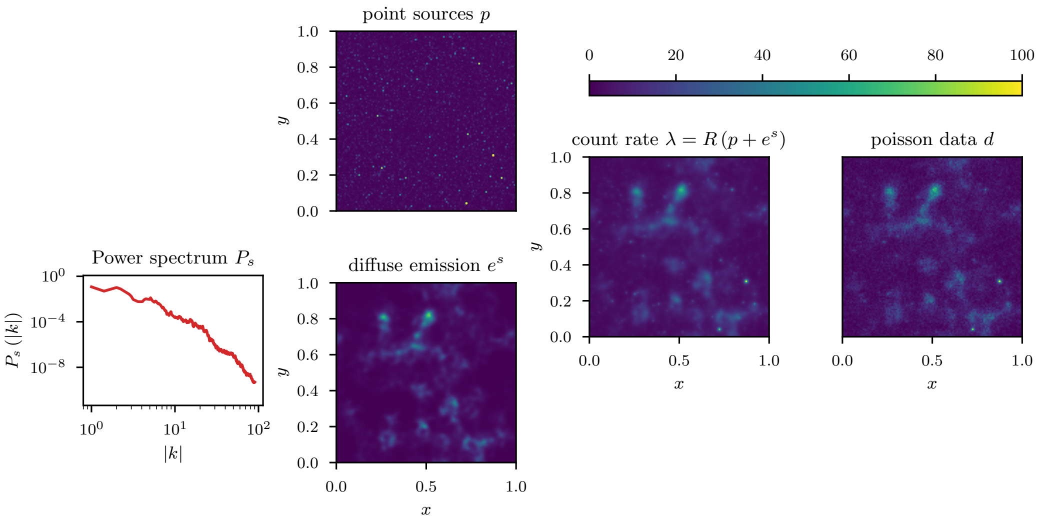


fluctuations
flexibility
asperity
slope
offset std.
noise std.
ground truth
Q_{HMC}

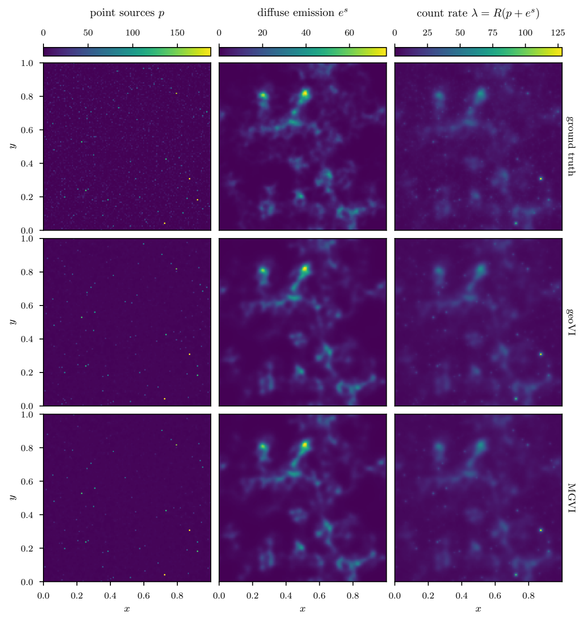
Examples



fluctuations
flexibility
asperity
slope
offset std.
−5
0
−5
0
−2.5
0.0
2.5
−5
0
−2.5
0.0
2.5
noise std.
−0.50
−0.75
−1.00
−1.25
−1.50
fluctuations
0
−2
−4
flexibility
2.5
0.0
−2.5
−5.0
asperity
4
2
0
−2
−4
slope
2
0
−2
−4
ground truth
Q_{MGVI}

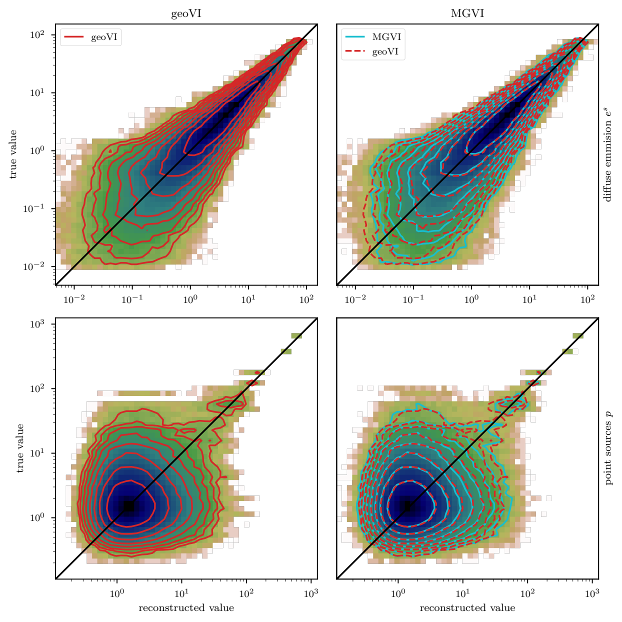

point sources p
diffuse emission e^s
Power spectrum P_s
count rate $\lambda = R\,(p + e^s)$
poisson data d
$P_s\,(|k|)$
$|k|$
x
y
0
20
40
60
80
100
1.0
0.8
0.6
0.4
0.2
0.0
0.5
10^0
10^{-4}
10^{-8}
10^1
10^2

Examples



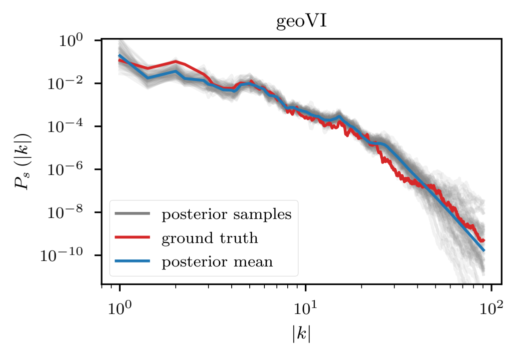
point sources p
diffuse emission e^s
count rate $\lambda = R(p + e^s)$
0 50 100 150
0 20 40 60
0 25 50 75 100 125
ground truth
geoVI
MGVI
y
x
0.0 0.2 0.4 0.6 0.8 1.0

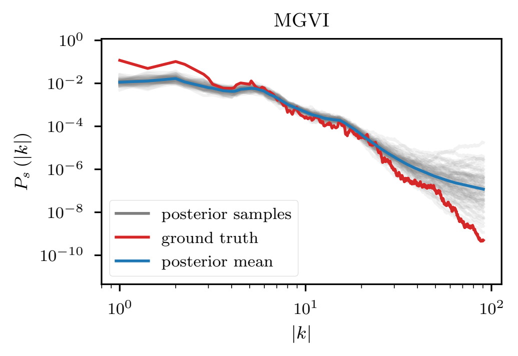
Examples



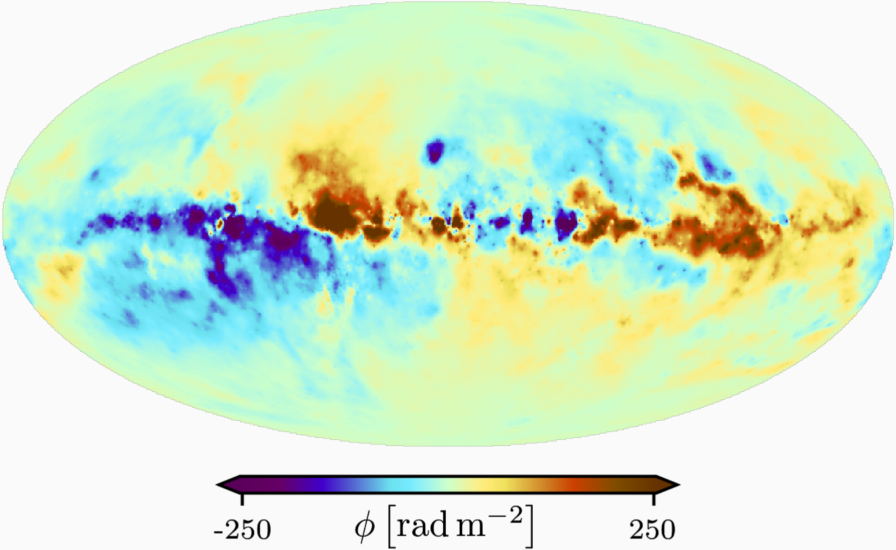
geoVI
MGVI
geoVI
MGVI
geoVI
true value
true value
diffuse emmision e^s
point sources p
reconstructed value
reconstructed value
10^{-2}
10^{-1}
10^{0}
10^{1}
10^{2}
10^{3}

Examples

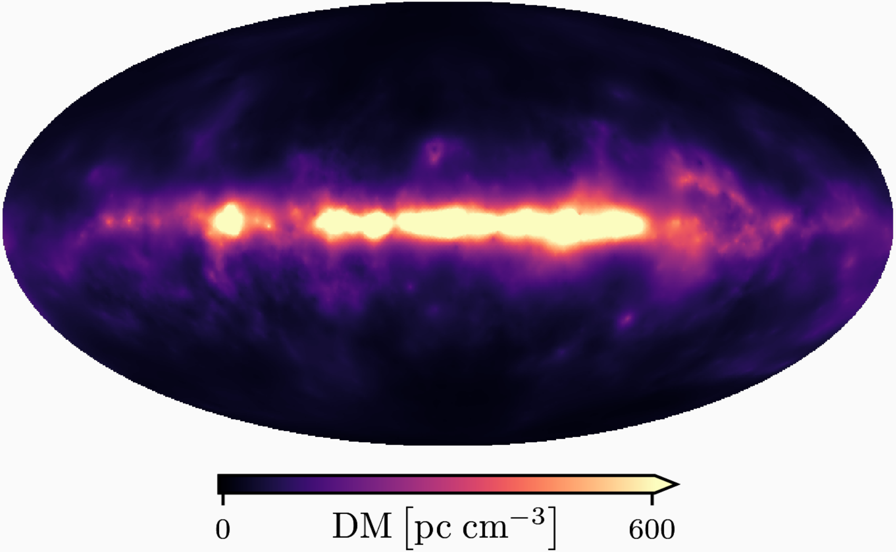

geoVI
10^0
10^{-2}
10^{-4}
10^{-6}
10^{-8}
10^{-10}
$P_s(|k|)$
posterior samples
ground truth
posterior mean
10^0
10^1
10^2
$|k|$


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

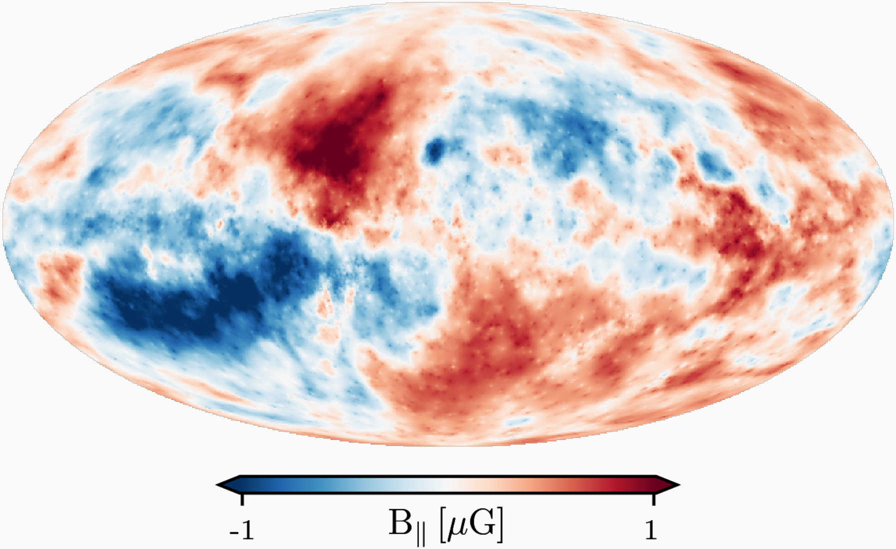
Application (preliminary)



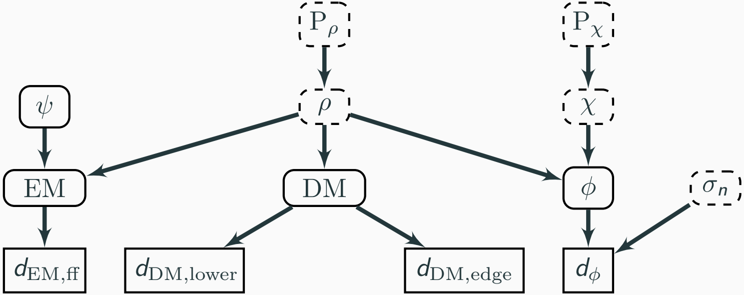
Application (preliminary)



Application (preliminary)



Application (preliminary)

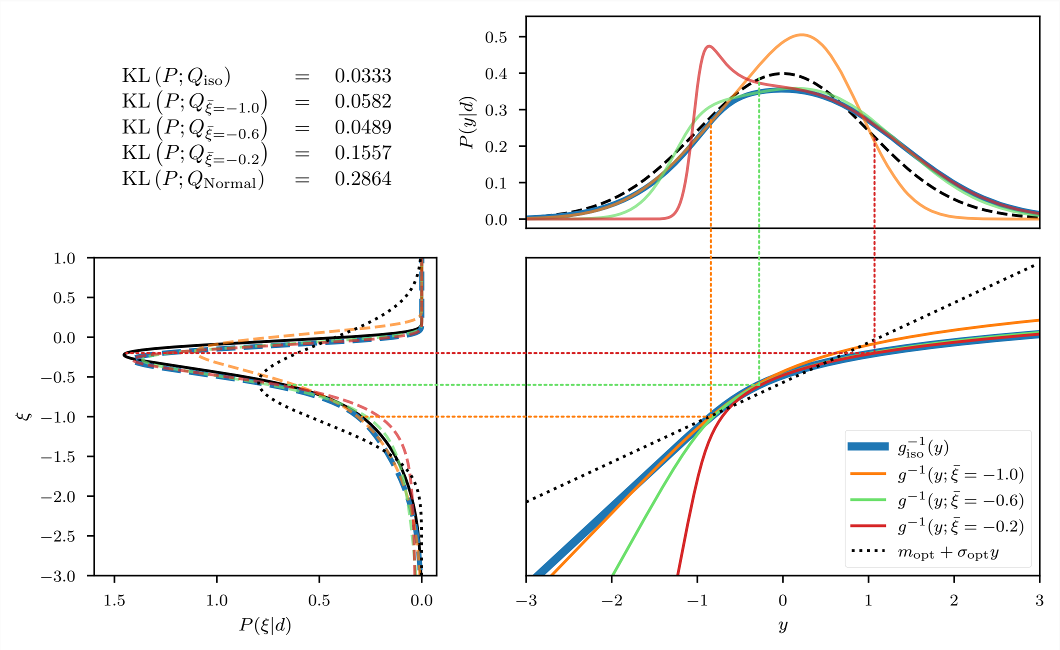


References

- **Geometric variational inference**; Frank, Leike, Enßlin; https://arxiv.org/abs/2105.10470
- **Numerical information field theory (Nifty7)**; https://gitlab.mpcdf.mpg.de/ift/nifty

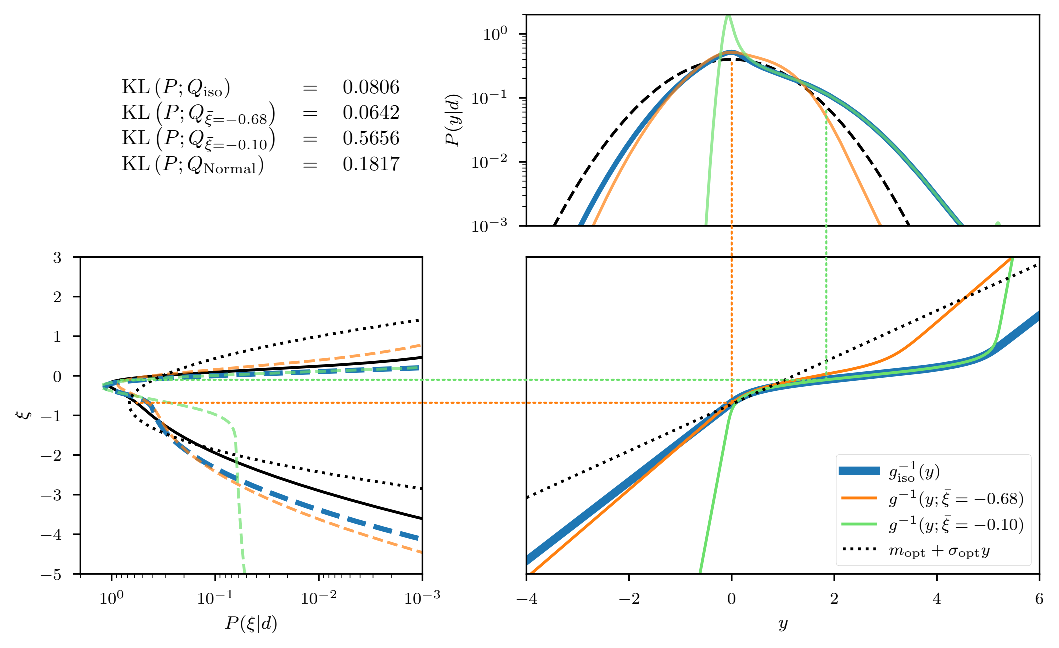
Appendix

Appendix



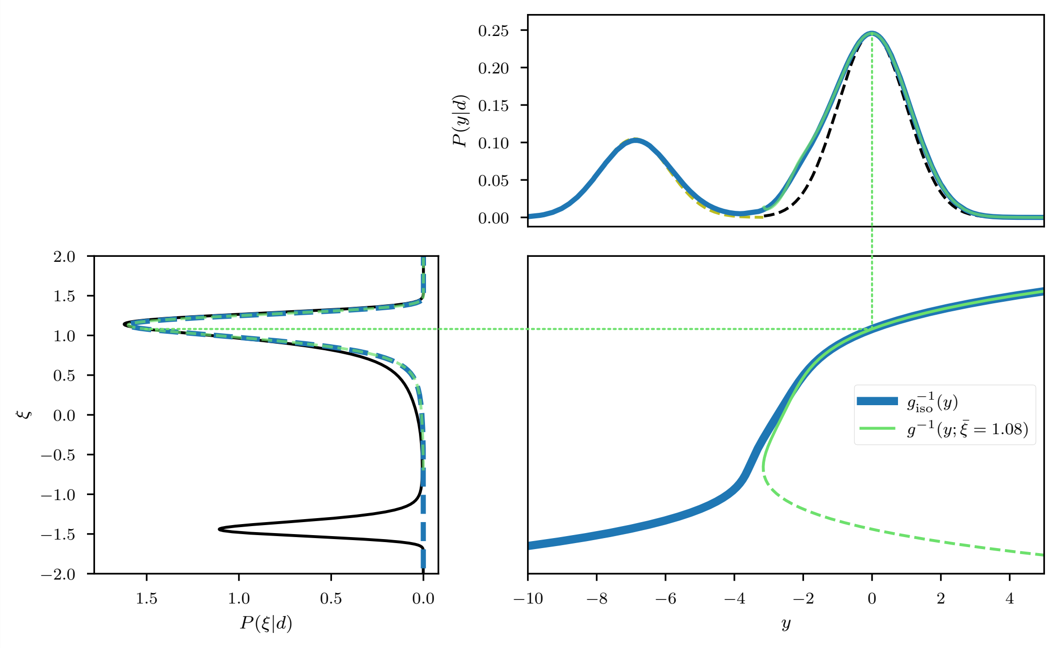
$\mathrm{KL}\left(P; Q_{\mathrm{iso}}\right) = 0.0333$
$\mathrm{KL}\left(P; Q_{\bar{\xi}=-1.0}\right) = 0.0582$
$\mathrm{KL}\left(P; Q_{\bar{\xi}=-0.6}\right) = 0.0489$
$\mathrm{KL}\left(P; Q_{\bar{\xi}=-0.2}\right) = 0.1557$
$\mathrm{KL}\left(P; Q_{\mathrm{Normal}}\right) = 0.2864$
$P(y|d)$
ξ
$P(\xi|d)$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

Appendix



KL (P; Q_iso) = 0.0806
KL (P; Q_ξ̄=−0.68) = 0.0642
KL (P; Q_ξ̄=−0.10) = 0.5656
KL (P; Q_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(ξ|d)
y
−4
−2
0
2
4
6
g_iso^−1(y)
g^−1(y; ξ̄ = −0.68)
g^−1(y; ξ̄ = −0.10)
m_opt + σ_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
$P(\xi|d)$
−10
−8
−6
−4
−2
0
2
4
y
$g_{\text{iso}}^{-1}(y)$
$g^{-1}(y; \bar{\xi} = 1.08)$