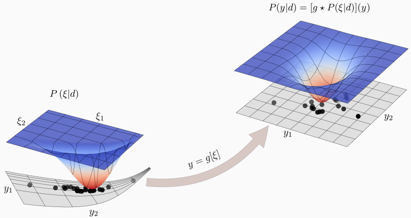
Geometric Variational Inference

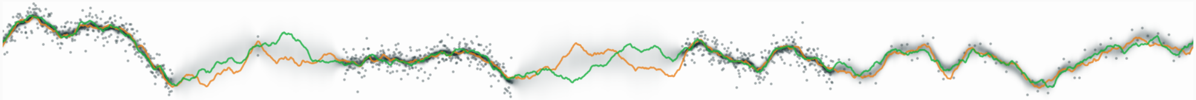


Riemannian geometry for approximate Bayesian imaging

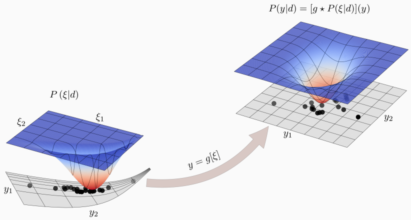
Philipp Frank[1], Reimar Leike[3], Torsten Ensslin[1,2]

41st MaxEnt2022 Conference, IHP, Paris, July 19, 2022

(1) Max-Planck Institute for Astrophysics MPA, Garching, Germany
(2) Ludwig-Maximilians University LMU, Munich, Germany
(3) Amazon Development Center, Cambridge, United Kingdom

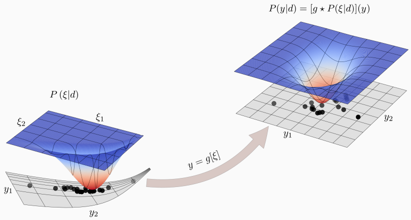


Disclaimer



Geometry ...

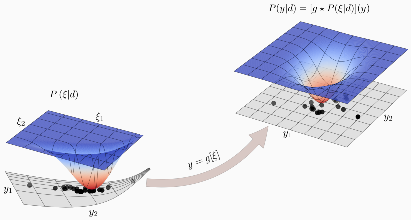
Disclaimer



Geometry ...

- ... is great!

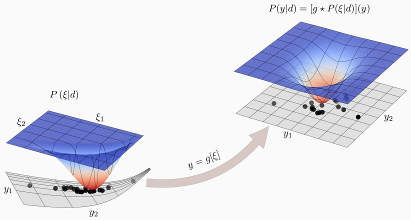
Disclaimer



Geometry ...

- ... is great!
- ... is not fundamental to reasoning

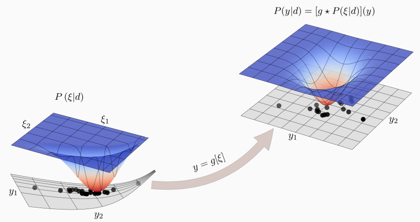
Disclaimer



Geometry ...

- ... is great!
- ... is not fundamental to reasoning
- ... in very high-dimensions is unintuitive

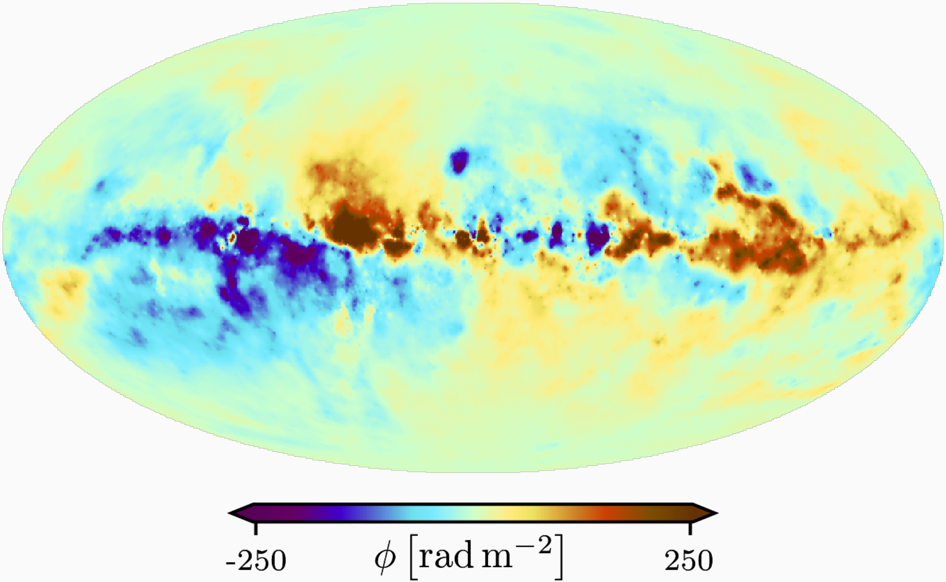
Disclaimer



Geometry ...

- ... is great!
- ... is not fundamental to reasoning
- ... in very high-dimensions is unintuitive
- ... used for algorithms is subject to computational constraints

Imaging problems [HE20]



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

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

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

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

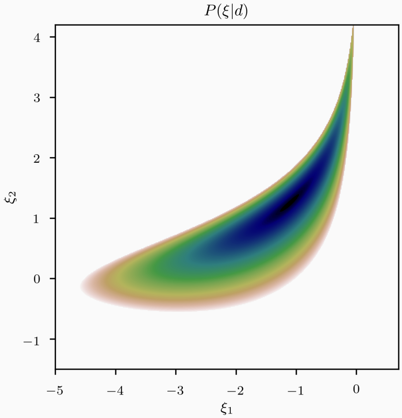
Posterior: $\mathcal{P}(\xi|d)$; Approximation: $\mathcal{Q}_\sigma(\xi)$; Variational parameters: σ.

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.

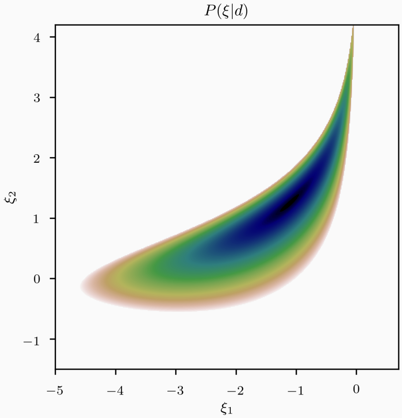
Geometric Variational Inference

Geometric Variational Inference (geoVI) [FLE21]



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

Geometric Variational Inference (geoVI) [FLE21]

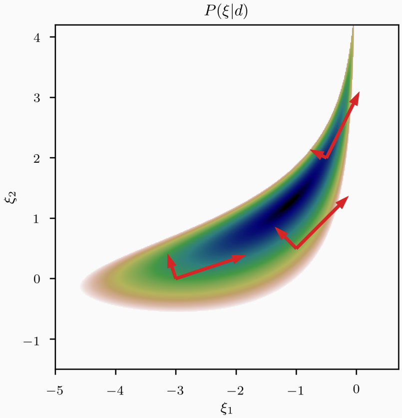


Information Hamiltonian $\mathcal{H}(\xi|d)$: $-\log\left(\mathcal{P}(\xi|d)\right)$

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]

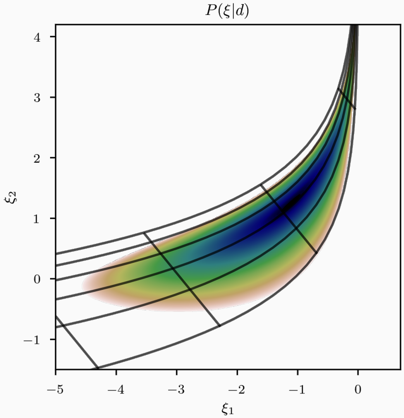


Information Hamiltonian $\mathcal{H}(\xi|d)$: $\quad -\log\left(\mathcal{P}(\xi|d)\right)$

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

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

Geometric Variational Inference (geoVI) [FLE21]

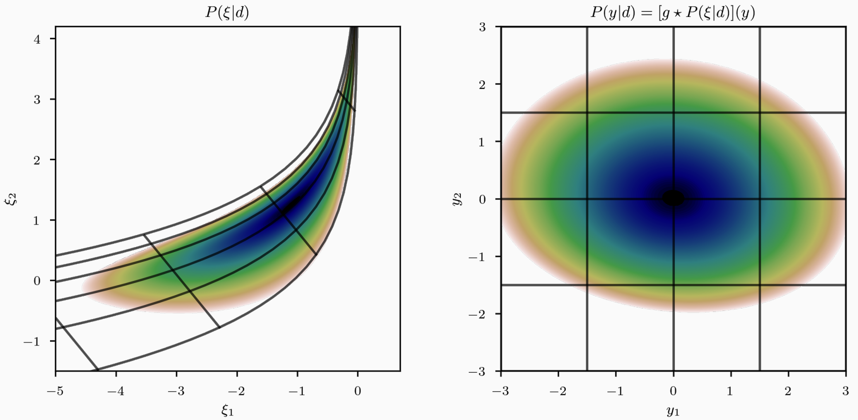


Information Hamiltonian $\mathcal{H}(\xi|d)$: $-\log\left(\mathcal{P}(\xi|d)\right)$

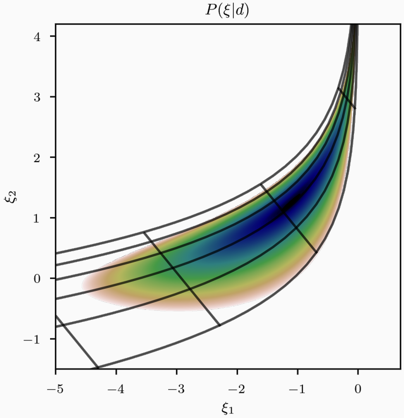
Posterior metric $\mathcal{M}(\xi)$: $\mathcal{M}_{\mathrm{lh}}(\xi) + \mathbb{1}$

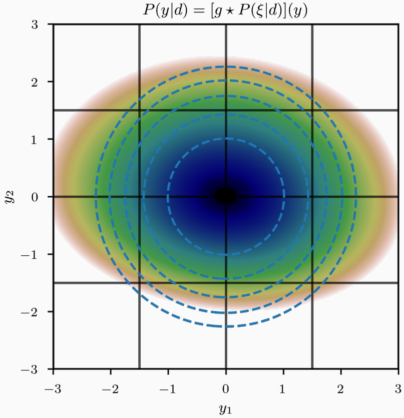
Fisher information metric $\mathcal{M}_{\mathrm{lh}}(\xi)$: $\left\langle \frac{\partial^2 \mathcal{H}(d|\xi)}{\partial\xi\partial\xi'} \right\rangle_{\mathcal{P}(d|\xi)}$

Geometric Variational Inference (geoVI) [FLE21]

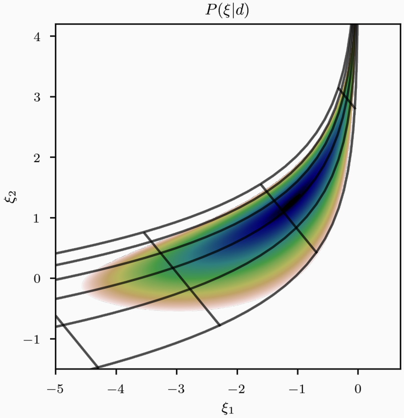


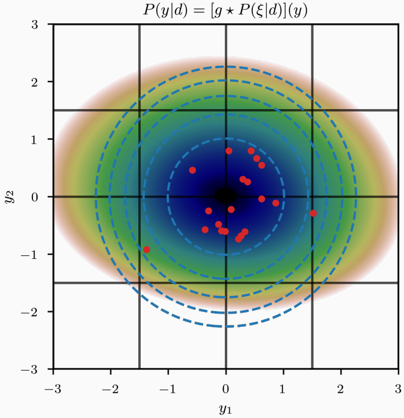
Geometric Variational Inference (geoVI) [FLE21]



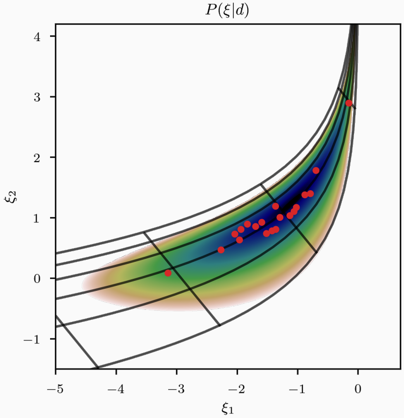


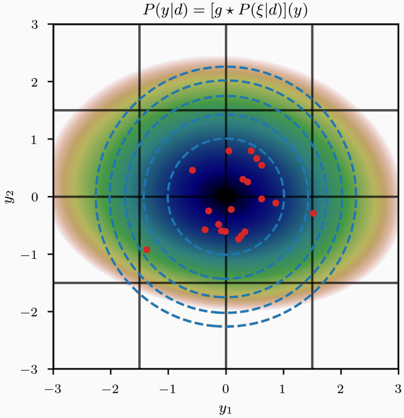
Geometric Variational Inference (geoVI) [FLE21]





Geometric Variational Inference (geoVI) [FLE21]





Geometric Variational Inference (geoVI) [FLE21]

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

Geometric Variational Inference (geoVI) [FLE21]

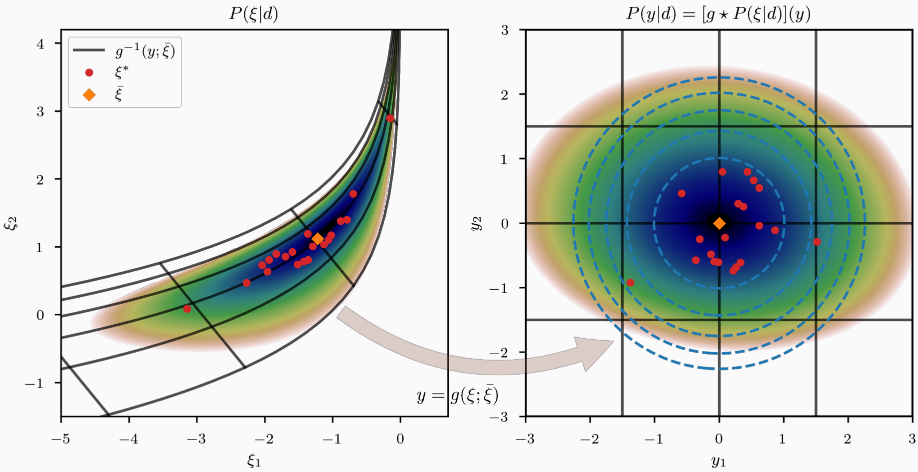
- Approximate coordinate transformation g such that $\mathcal{M}(\xi) \approx \left(\frac{\partial g}{\partial \xi}\right)^T \frac{\partial g}{\partial \xi}$
- Fisher metric via pullback: $\mathcal{M}(\xi) = \mathcal{M}_{\mathrm{lh}}(\xi) + \mathbb{1} = \left(\frac{\partial x}{\partial \xi}\right)^T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

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: $\bar{\xi}$.

Geometric Variational Inference (geoVI) [FLE21]



Geometric Variational Inference (geoVI) [FLE21]

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

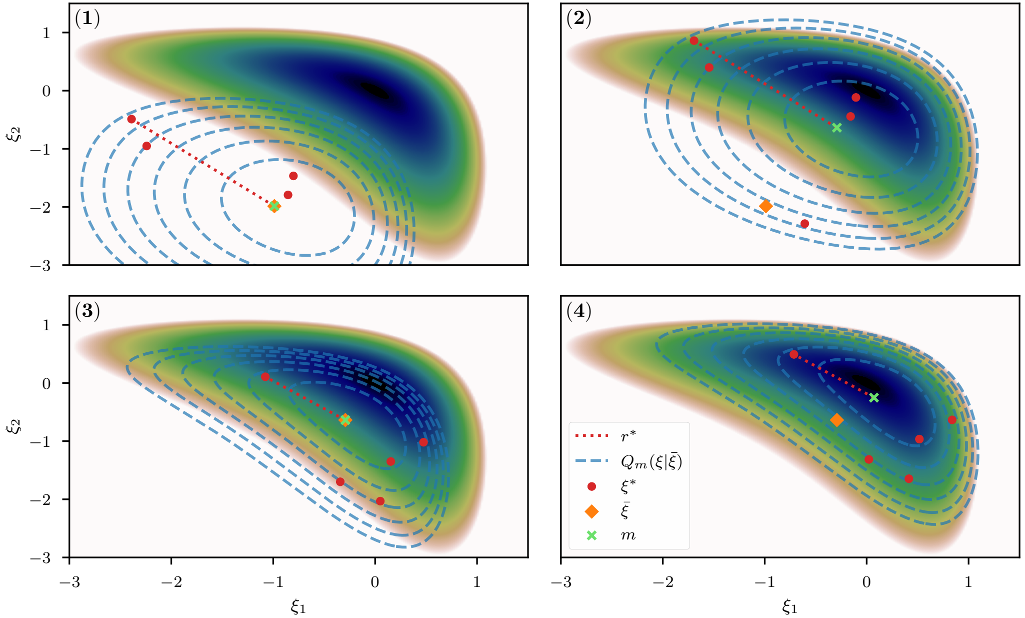
$$y = g(\xi; \bar{\xi}) = \mathcal{M}(\bar{\xi})^{-1/2} \left[\xi - \bar{\xi} + \left(\frac{\partial x}{\partial \xi} \right)^T \Bigg|_{\xi = \bar{\xi}} \left(x(\xi) - x(\bar{\xi}) \right) \right]$$

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

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

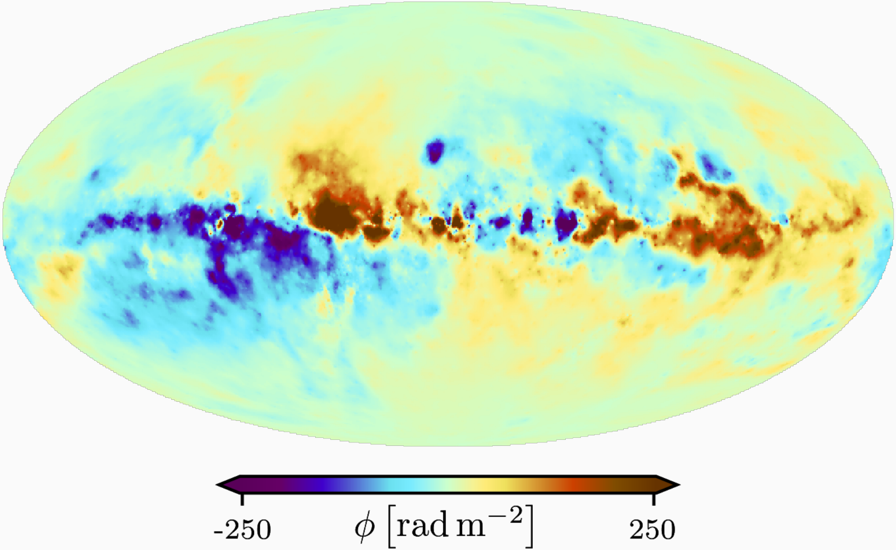
$$\mathcal{Q}_{\bar{\xi}}(\xi) = \mathcal{N}(y|0, \mathbb{1})|_{y = g(\xi; \bar{\xi})} \left\| \frac{\partial g(\xi; \bar{\xi})}{\partial \xi} \right\|$$

Geometric Variational Inference (geoVI) [FLE21]

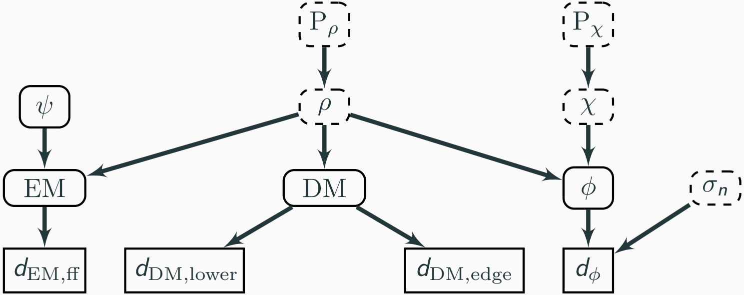


Application (preliminary)

Application (preliminary)



Application (preliminary)



Conclusion - Code

Numerical information field theory

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

References

Philipp Arras, Philipp Frank, Philipp Haim, Jakob Knollmüller, Reimar Leike, Martin Reinecke, and Torsten Enßlin.
Variable structures in m87* from space, time and frequency resolved interferometry.
Nature Astronomy, 6(2):259–269, 2022.

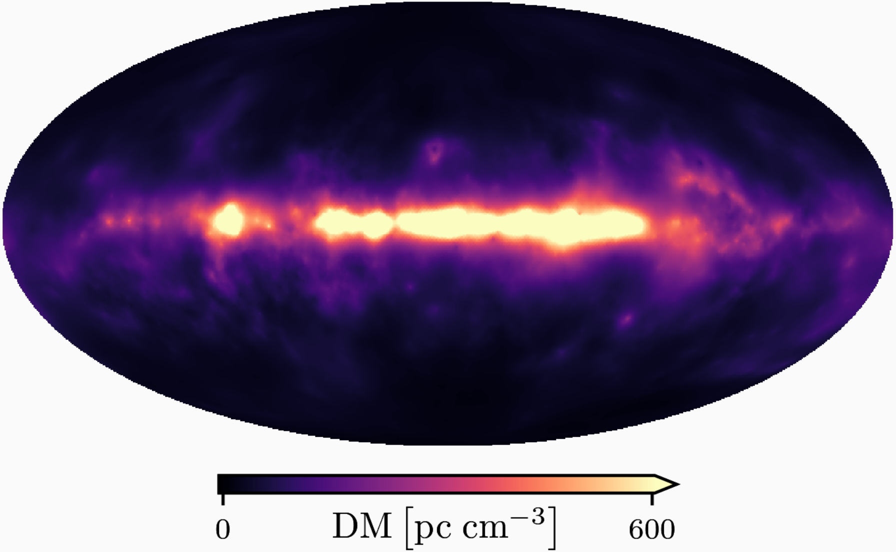
Philipp Frank, Reimar Leike, and Torsten A. Enßlin.
Geometric variational inference.
Entropy, 23(7), 2021.

Sebastian Hutschenreuter and Torsten A. Enßlin.
The galactic faraday depth sky revisited.
A&A, 633:A150, 2020.

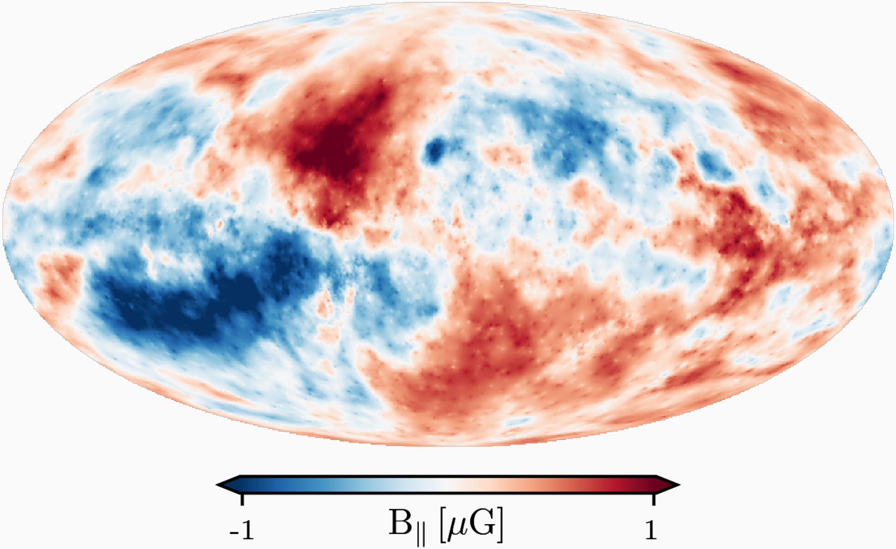
Reimar Leike, Gordian Edenhofer, Jakob Knollmüller, Christian Alig, Philipp Frank, and Torsten A. Enßlin.
The galactic 3d large-scale dust distribution via gaussian process regression on spherical coordinates.
arXiv, 2204.11715, 2022.

Appendix

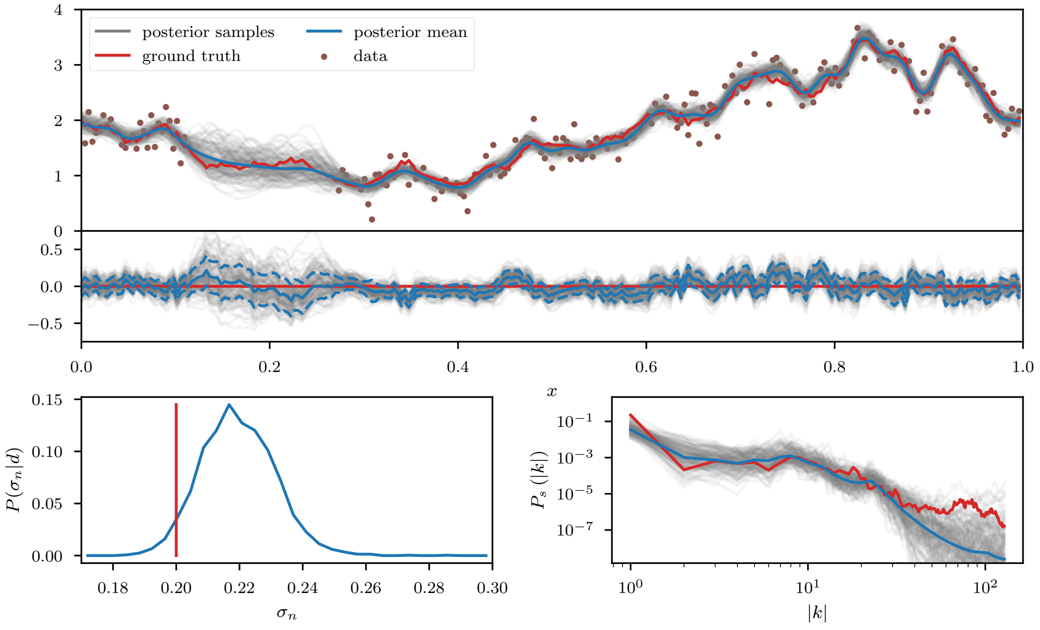
Appendix



Appendix

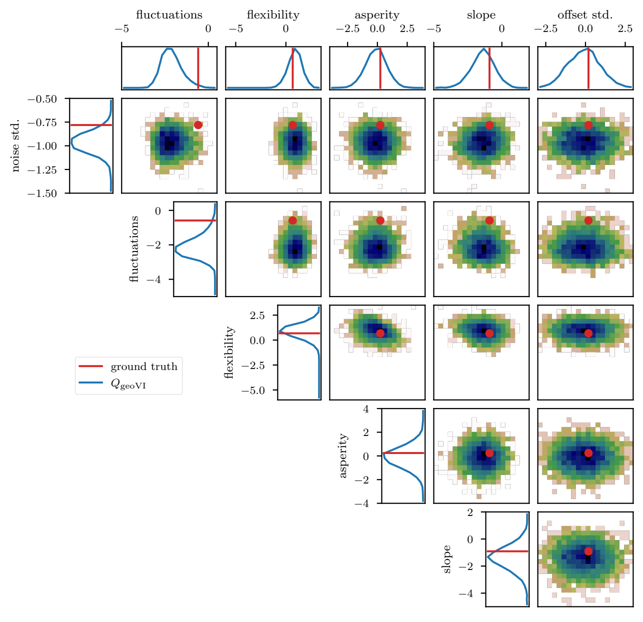


Appendix



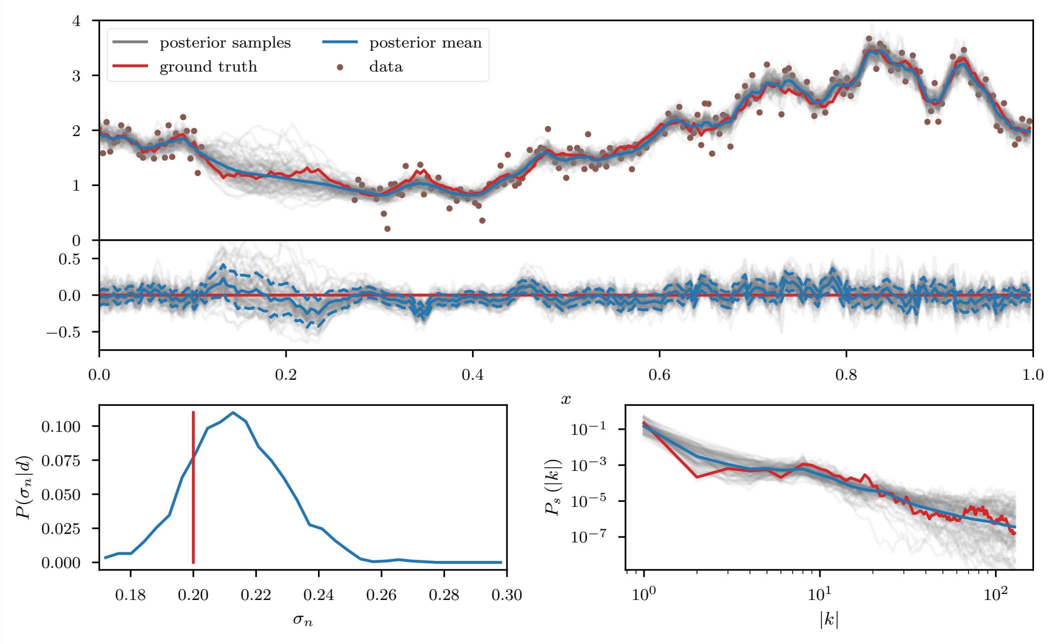
posterior samples
ground truth
posterior mean
data
4
3
2
1
0
0.5
0.0
−0.5
0.0
0.2
0.4
0.6
0.8
1.0
x
0.15
0.10
0.05
0.00
$P(\sigma_n|d)$
0.18
0.20
0.22
0.24
0.26
0.28
0.30
σ_n
10^{-1}
10^{-3}
10^{-5}
10^{-7}
$P_s(|k|)$
10^0
10^1
10^2
$|k|$

Appendix



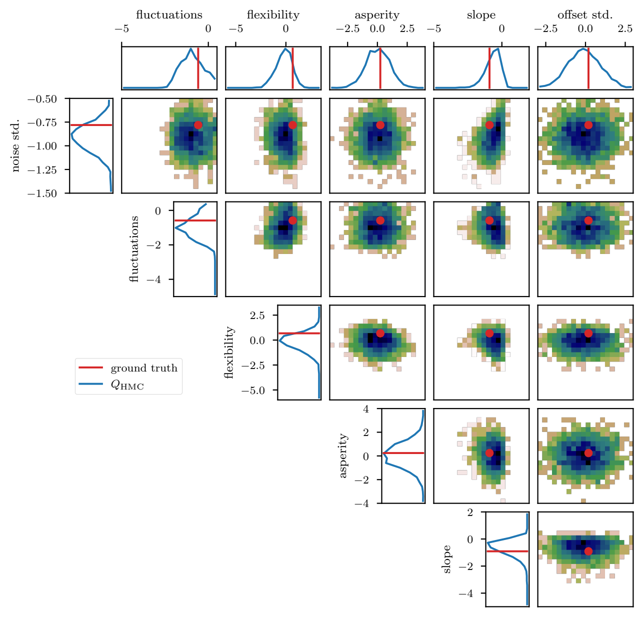
fluctuations
flexibility
asperity
slope
offset std.
noise std.
fluctuations
flexibility
asperity
slope
−5
0
−2.5
0.0
2.5
−0.50
−0.75
−1.00
−1.25
−1.50
−2
−4
−5.0
4
2
ground truth
Q_{geoVI}

Appendix



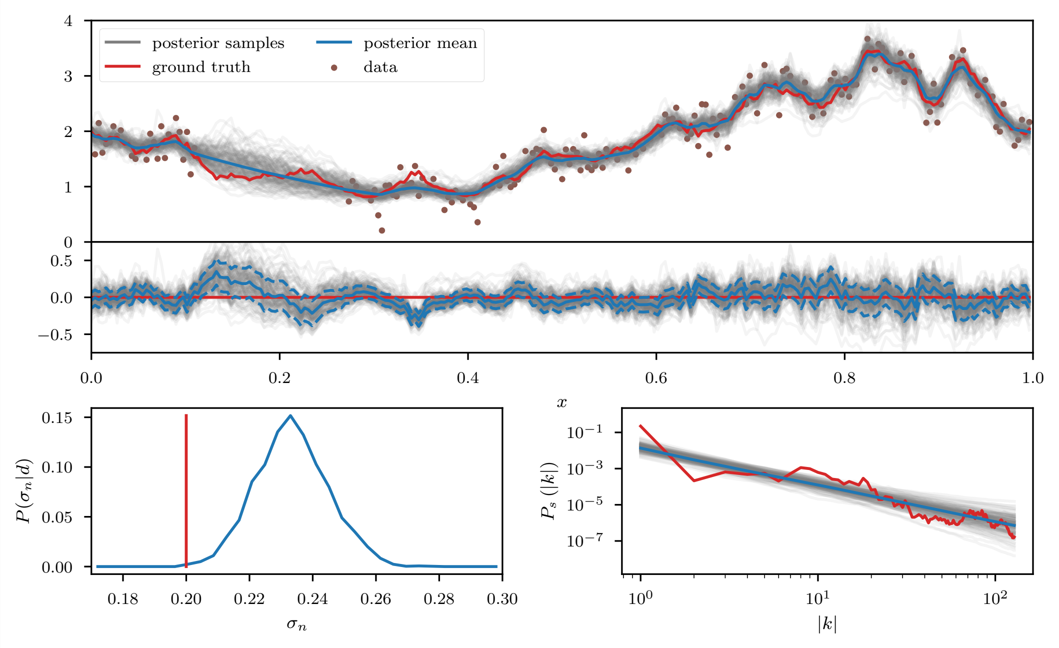
posterior samples
posterior mean
ground truth
data
x
$P(\sigma_n|d)$
σ_n
$P_s(|k|)$
$|k|$

Appendix



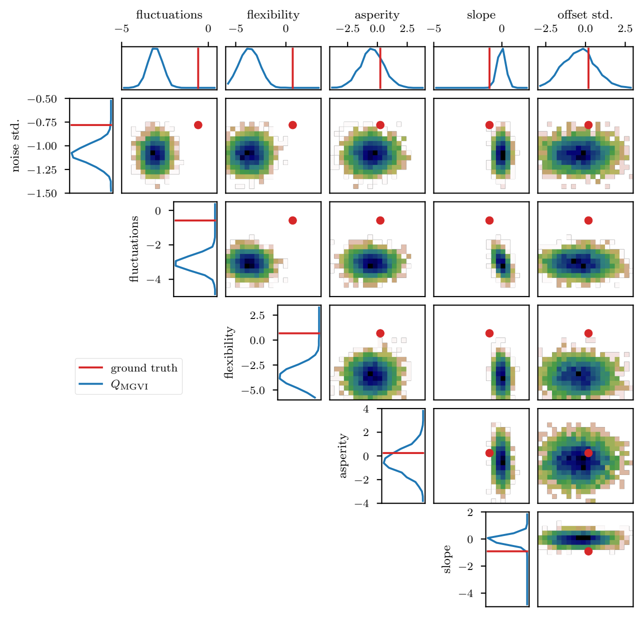
fluctuations
flexibility
asperity
slope
offset std.
noise std.
ground truth
Q_{HMC}

Appendix

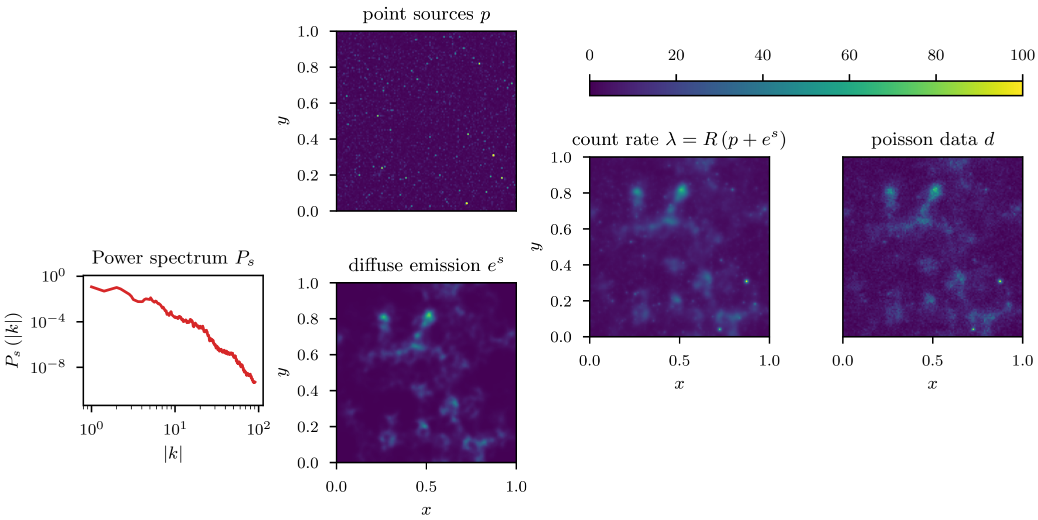


posterior samples
ground truth
posterior mean
data
4
3
2
1
0
0.5
0.0
−0.5
0.0
0.2
0.4
0.6
0.8
1.0
x
0.15
0.10
0.05
0.00
$P(\sigma_n|d)$
0.18
0.20
0.22
0.24
0.26
0.28
0.30
σ_n
10^{-1}
10^{-3}
10^{-5}
10^{-7}
$P_s(|k|)$
10^0
10^1
10^2
$|k|$

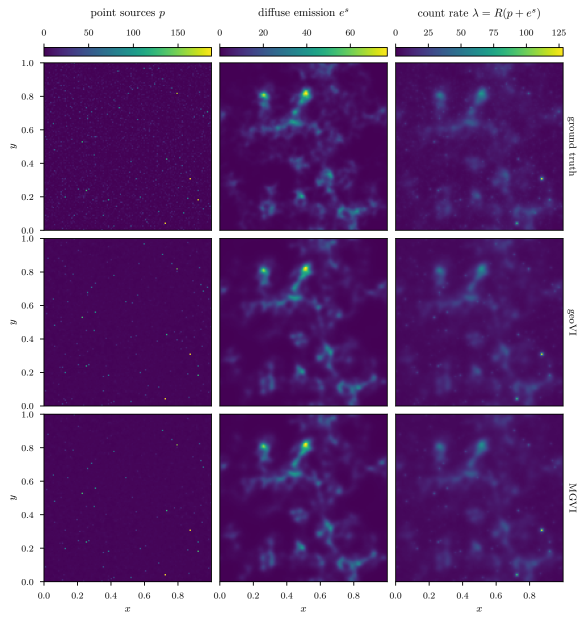
Appendix


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

Appendix

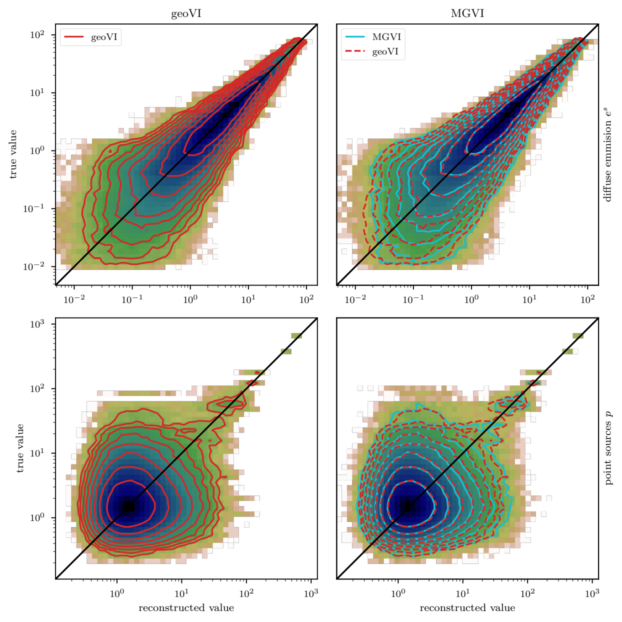

point sources p
count rate $\lambda = R\,(p + e^s)$
poisson data d
Power spectrum P_s
diffuse emission e^s
$P_s\,(|k|)$
$|k|$
x
y
0
20
40
60
80
100
10^0
10^{-4}
10^{-8}
10^1
10^2
0.0
0.2
0.4
0.6
0.8
1.0
0.5

Appendix



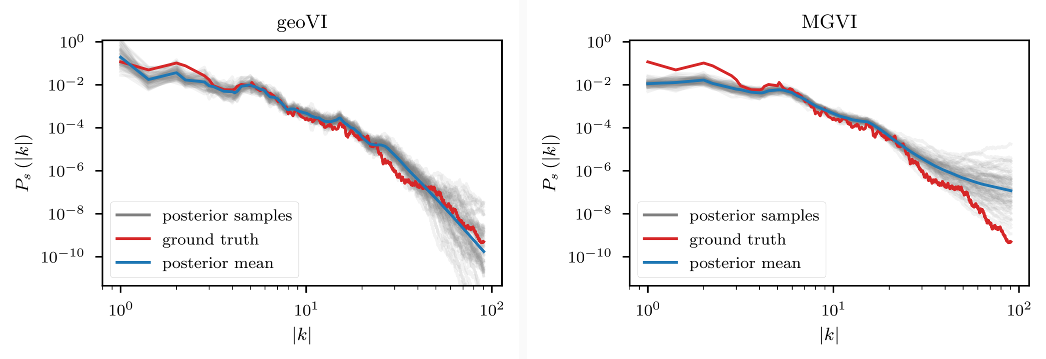
point sources p
diffuse emission e^s
count rate $\lambda = R(p + e^s)$
0 50 100 150
0 20 40 60
0 25 50 75 100 125
ground truth
geoVI
MGVI
y
x
0.0 0.2 0.4 0.6 0.8 1.0

Appendix



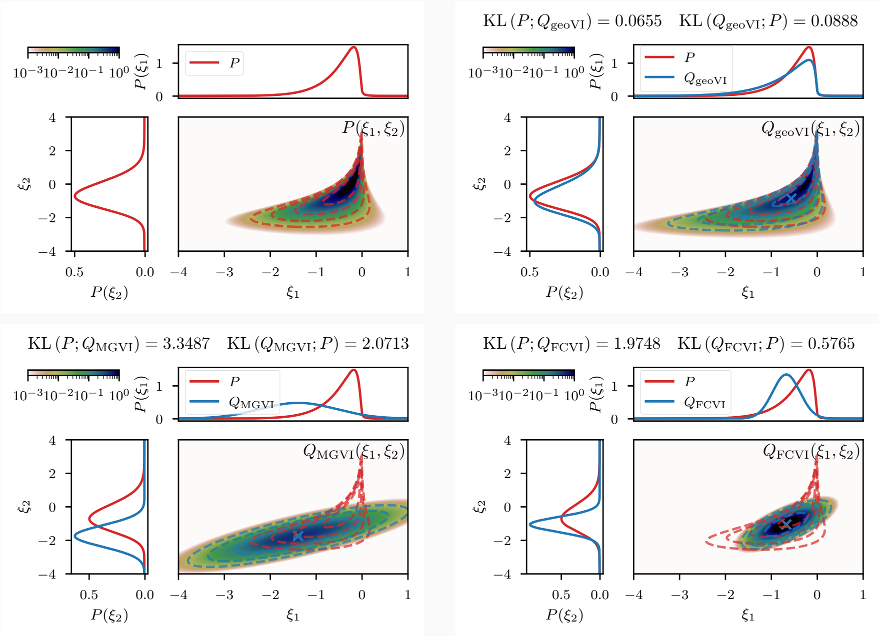
geoVI
MGVI
geoVI
MGVI
geoVI
true value
true value
diffuse emmision e^s
point sources p
reconstructed value
reconstructed value

Appendix



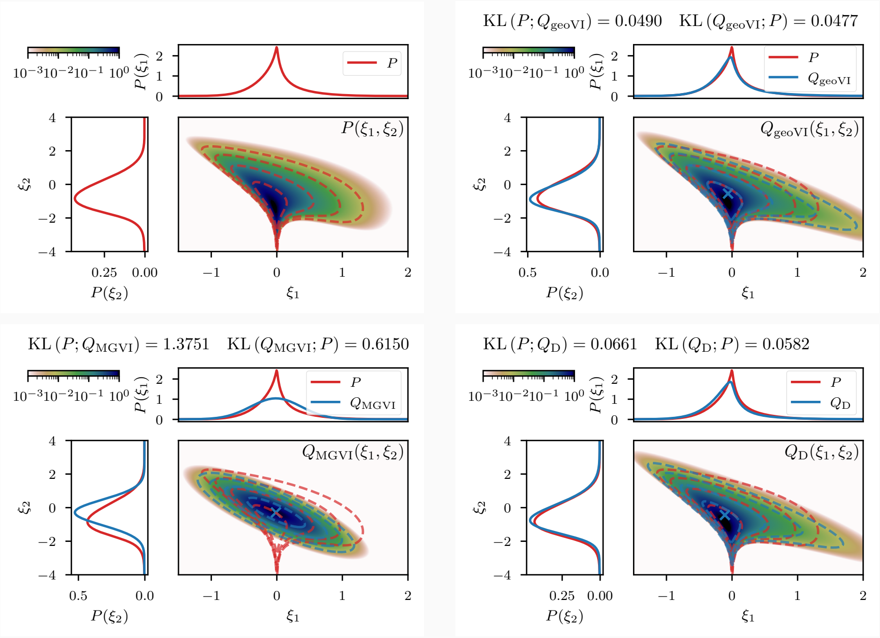
geoVI
MGVI
10^{0}
10^{-2}
10^{-4}
10^{-6}
10^{-8}
10^{-10}
$P_s(|k|)$
10^{0}
10^{1}
10^{2}
$|k|$
posterior samples
ground truth
posterior mean

Appendix



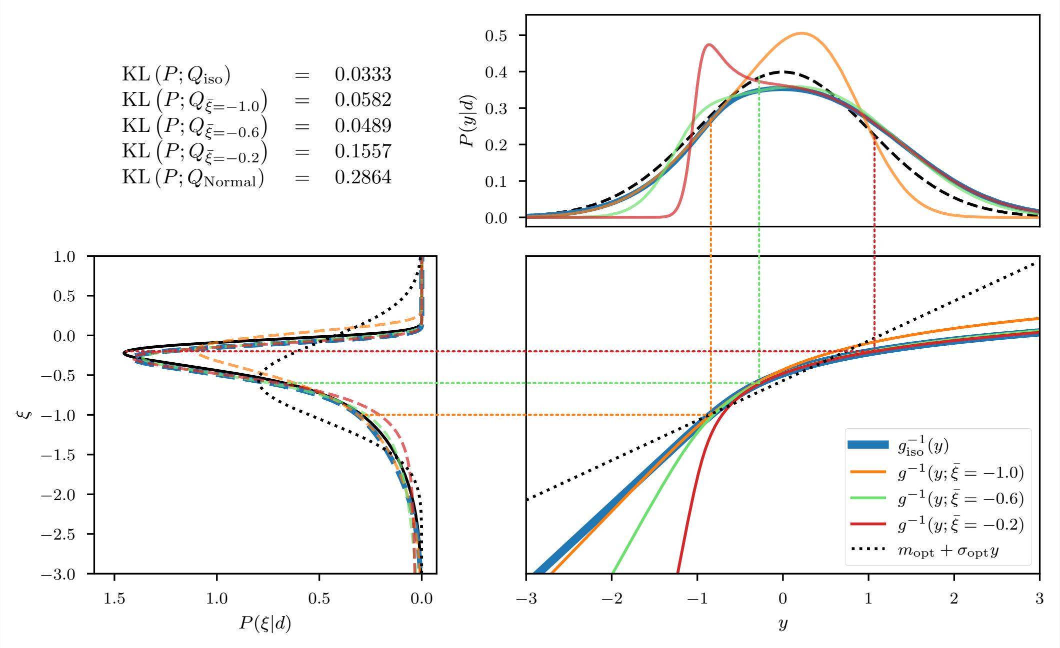
KL $(P; Q_{\mathrm{geoVI}}) = 0.0655$ KL $(Q_{\mathrm{geoVI}}; P) = 0.0888$
KL $(P; Q_{\mathrm{MGVI}}) = 3.3487$ KL $(Q_{\mathrm{MGVI}}; P) = 2.0713$
KL $(P; Q_{\mathrm{FCVI}}) = 1.9748$ KL $(Q_{\mathrm{FCVI}}; P) = 0.5765$
P
Q_{geoVI}
Q_{MGVI}
Q_{FCVI}
$P(\xi_1, \xi_2)$
$Q_{\mathrm{geoVI}}(\xi_1, \xi_2)$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

Appendix



$\mathrm{KL}\,(P; Q_{\mathrm{geoVI}}) = 0.0490$ $\mathrm{KL}\,(Q_{\mathrm{geoVI}}; P) = 0.0477$
$\mathrm{KL}\,(P; Q_{\mathrm{MGVI}}) = 1.3751$ $\mathrm{KL}\,(Q_{\mathrm{MGVI}}; P) = 0.6150$
$\mathrm{KL}\,(P; Q_{\mathrm{D}}) = 0.0661$ $\mathrm{KL}\,(Q_{\mathrm{D}}; P) = 0.0582$
$P(\xi_1, \xi_2)$
$Q_{\mathrm{geoVI}}(\xi_1, \xi_2)$
$Q_{\mathrm{MGVI}}(\xi_1, \xi_2)$
$Q_{\mathrm{D}}(\xi_1, \xi_2)$
P
Q_{geoVI}
Q_{MGVI}
Q_{D}
10^{-3} 10^{-2} 10^{-1} 10^{0}
$P(\xi_1)$
$P(\xi_2)$
ξ_1
ξ_2

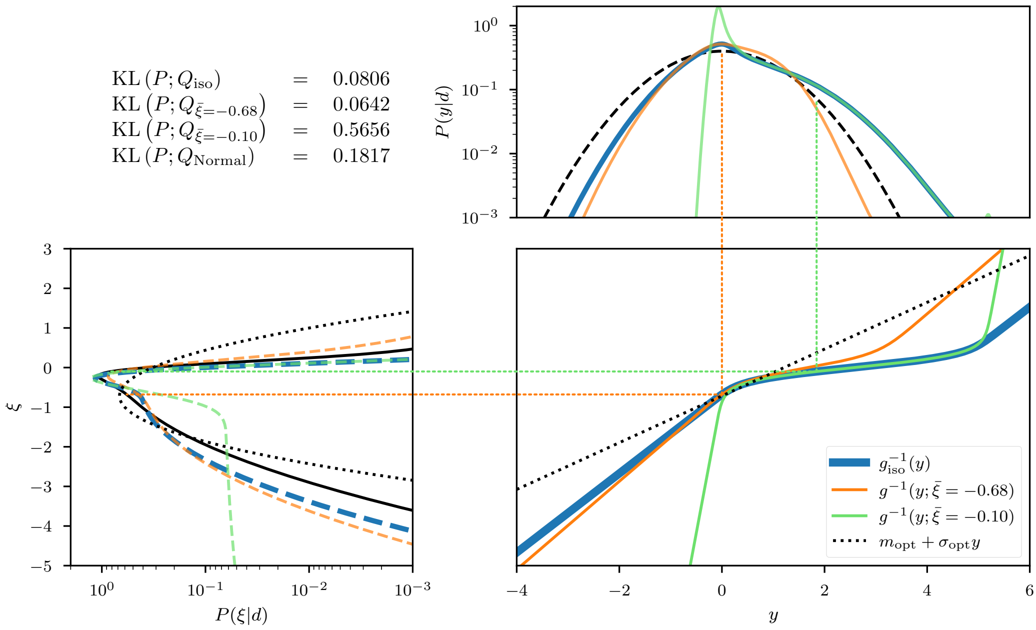
Appendix

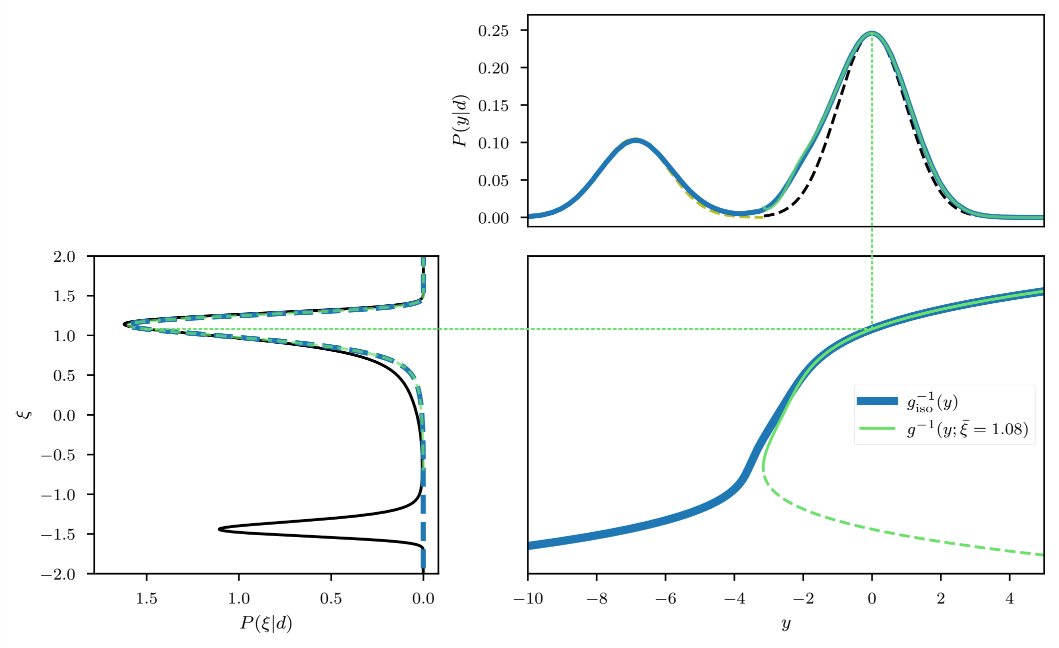


KL (P; Q_iso) = 0.0333
KL (P; Q_ξ̄=−1.0) = 0.0582
KL (P; Q_ξ̄=−0.6) = 0.0489
KL (P; Q_ξ̄=−0.2) = 0.1557
KL (P; Q_Normal) = 0.2864
P(y|d)
ξ
P(ξ|d)
y
g_iso^−1(y)
g^−1(y; ξ̄ = −1.0)
g^−1(y; ξ̄ = −0.6)
g^−1(y; ξ̄ = −0.2)
m_opt + σ_opt y

Appendix

$$
\begin{aligned}
\mathrm{KL}\left(P; Q_{\mathrm{iso}}\right) &= 0.0806 \\
\mathrm{KL}\left(P; Q_{\bar{\xi}=-0.68}\right) &= 0.0642 \\
\mathrm{KL}\left(P; Q_{\bar{\xi}=-0.10}\right) &= 0.5656 \\
\mathrm{KL}\left(P; Q_{\mathrm{Normal}}\right) &= 0.1817
\end{aligned}
$$





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
$P(\xi|d)$
$g_{\mathrm{iso}}^{-1}(y)$
$g^{-1}(y; \bar{\xi} = 1.08)$
−10
−8
−6
−4
−2
0
2
4
y