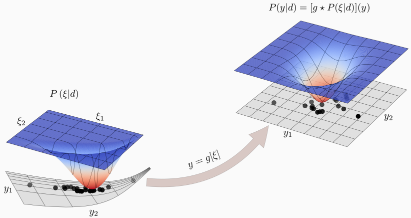


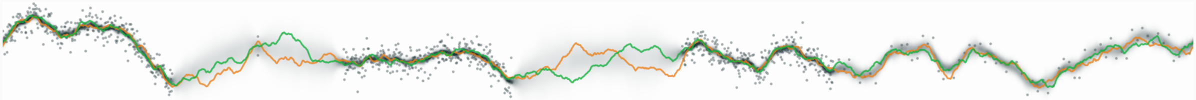
Variational Inference

Approximate Bayesian inference in high dimensions

Philipp Frank[1], Torsten Ensslin[1,2]

IMAGINE workshop: Towards a comprehensive model of the galactic magnetic field, NORDITA, Stockholm, April 5, 2023

(1) Max-Planck Institute for Astrophysics MPA, Garching, Germany
(2) Ludwig-Maximilians University LMU, Munich, Germany



Probability measures

Posterior expectation

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

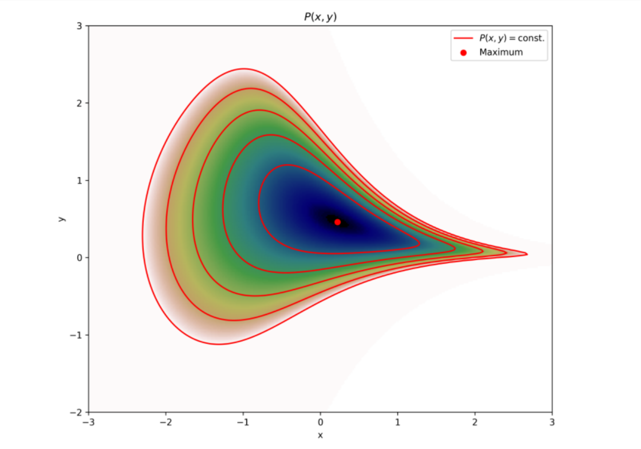
Function: $f(\xi)$; Posterior: $\mathcal{P}(\xi|d)$; parameters: ξ; data: d.

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.

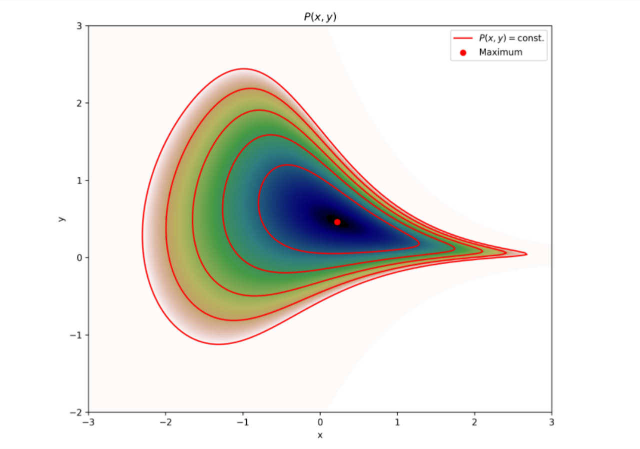


Probability measures

Posterior expectation

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

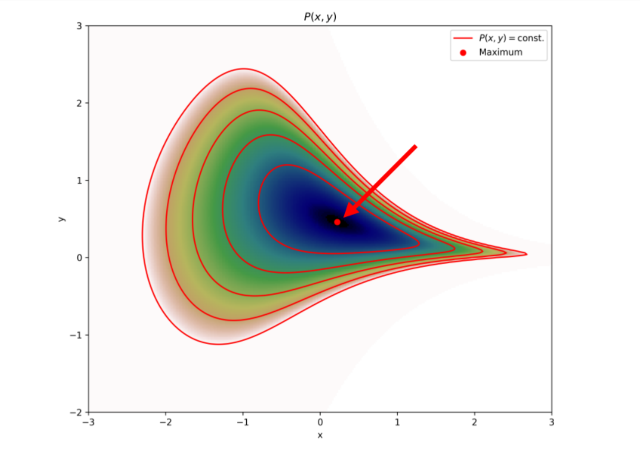
Function: $f(\xi)$; Posterior: $\mathcal{P}(\xi|d)$; parameters: ξ; data: d.



Posterior expectation

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

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

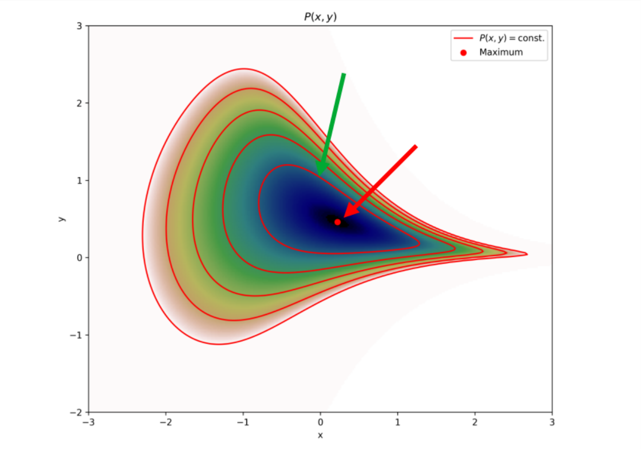


Probability measures

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.

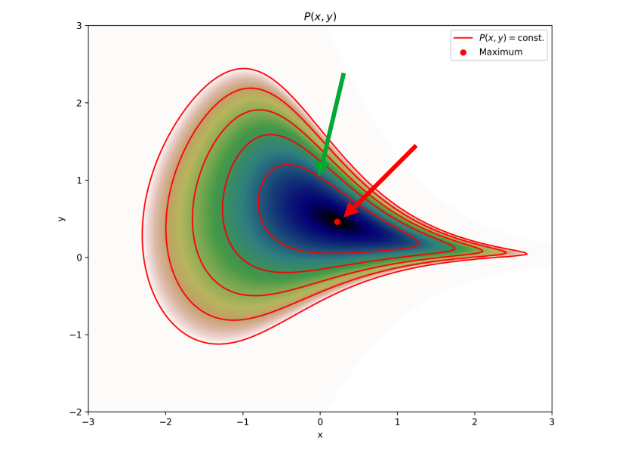


Variational Inference

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.

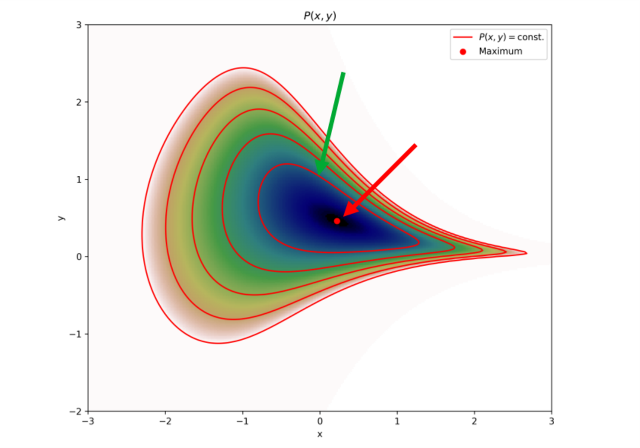


Probability measures

Posterior expectation

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

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

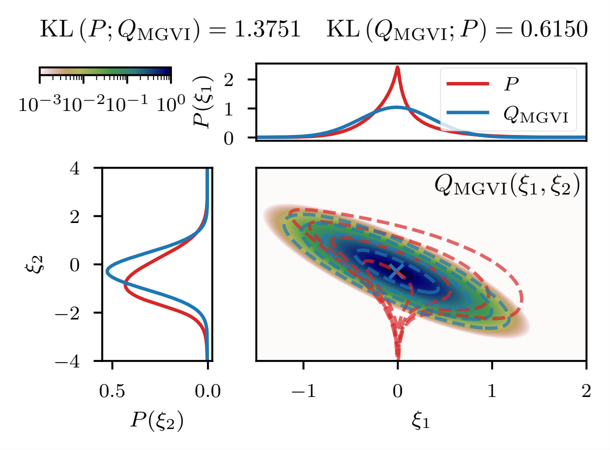


Probability measures

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.



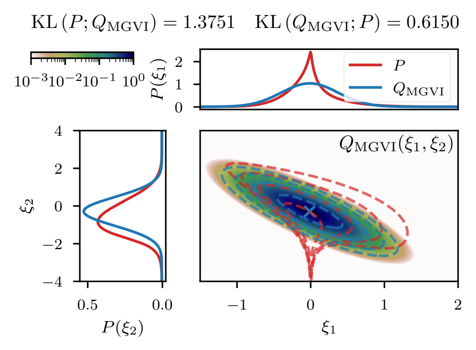
Variational Inference (VI)

Kullback-Leibler divergence

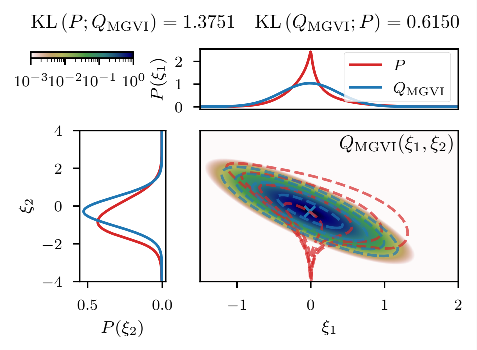
$$\mathrm{KL}\left[\mathcal{Q}_\sigma || \mathcal{P}\right] = -\int \log\left(\frac{\mathcal{P}(\xi|d)}{\mathcal{Q}_\sigma(\xi)}\right) \mathcal{Q}_\sigma(\xi)\ \mathrm{d}\xi$$

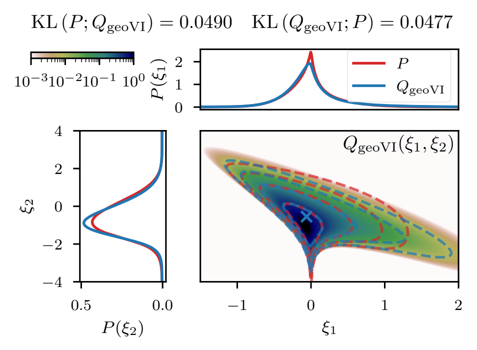
Posterior: $\mathcal{P}(\xi|d)$; Approximation: $\mathcal{Q}_\sigma(\xi)$; Variational parameters: σ.

Variational Inference (VI)



Variational Inference (VI)





Variational Inference (VI)

Kullback-Leibler divergence

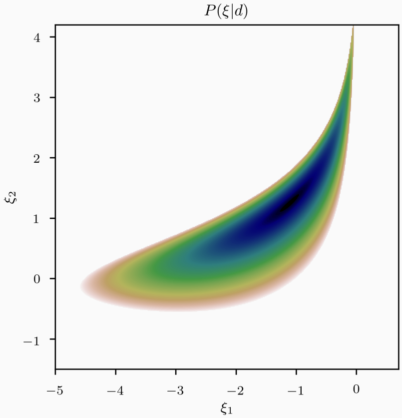
$$\mathrm{KL}\left[\mathcal{Q}_\sigma || \mathcal{P}\right] = -\int \log\left(\frac{\mathcal{P}(\xi|d)}{\mathcal{Q}_\sigma(\xi)}\right) \mathcal{Q}_\sigma(\xi)\, \mathrm{d}\xi$$

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

Approximate distribution $\mathcal{Q}$: $\mathcal{Q}(y) = \mathcal{N}(y; 0, \mathbb{1})$

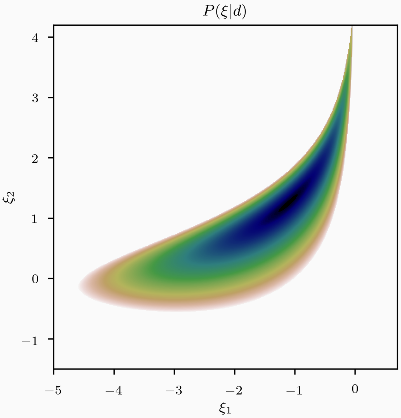
Coordinate system $y = g_\sigma(\xi)$ such that the *posterior* is close to Normal.

Geometric Variational Inference (geoVI) [FLE21]



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

Geometric Variational Inference (geoVI) [FLE21]

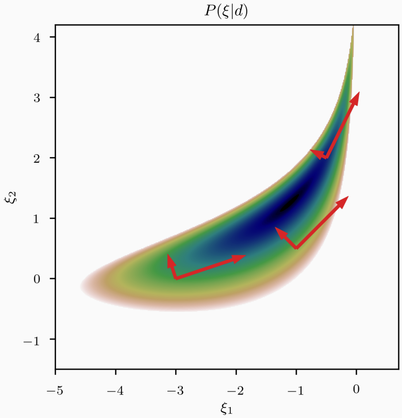


Information Hamiltonian $\mathcal{H}(\xi|d)$: $-\log\left(\mathcal{P}(\xi|d)\right)$

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]

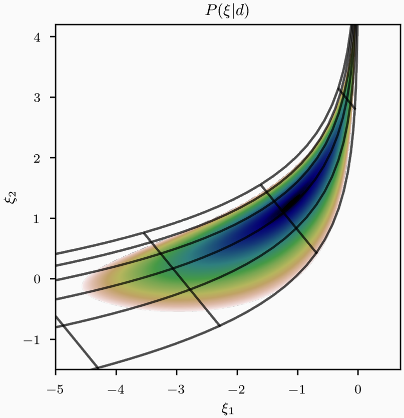


Information Hamiltonian $\mathcal{H}(\xi|d)$: $-\log\left(\mathcal{P}(\xi|d)\right)$

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

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

Geometric Variational Inference (geoVI) [FLE21]

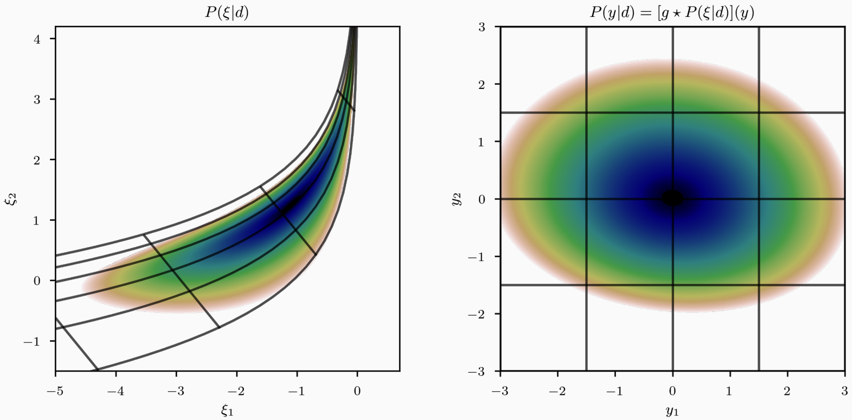


Information Hamiltonian $\mathcal{H}(\xi|d)$: $-\log(\mathcal{P}(\xi|d))$

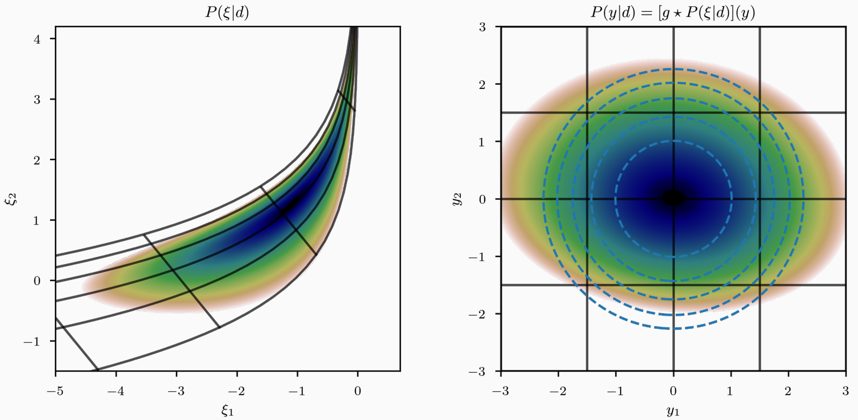
Posterior metric $\mathcal{M}(\xi)$: $\mathcal{M}_{\mathrm{lh}}(\xi) + \mathbb{1}$

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

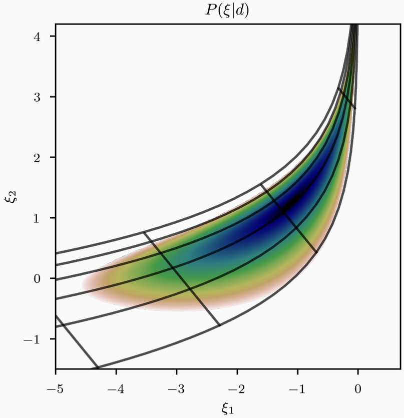
Geometric Variational Inference (geoVI) [FLE21]

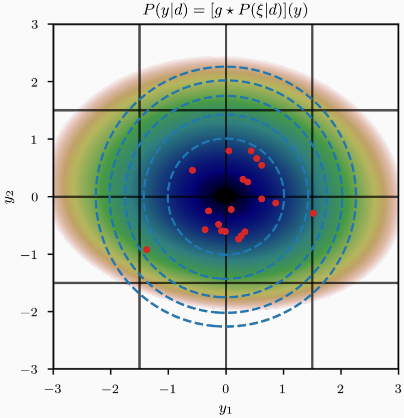


Geometric Variational Inference (geoVI) [FLE21]

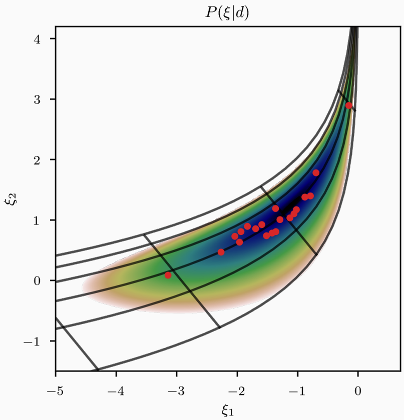


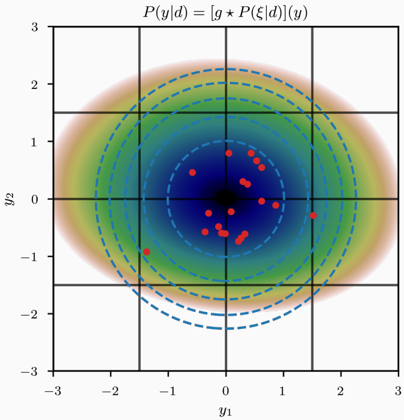
Geometric Variational Inference (geoVI) [FLE21]





Geometric Variational Inference (geoVI) [FLE21]





Geometric Variational Inference (geoVI) [FLE21]

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

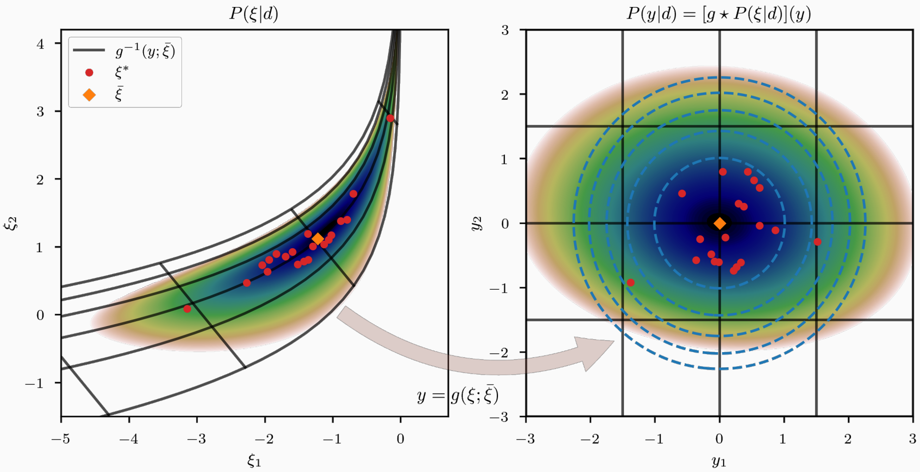
$$y = g(\xi; \bar{\xi}) = \mathcal{M}(\bar{\xi})^{-1/2} \left[\xi - \bar{\xi} + \left(\frac{\partial x}{\partial \xi} \right)^{T} \Bigg|_{\xi = \bar{\xi}} \left(x(\xi) - x(\bar{\xi}) \right) \right]$$

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

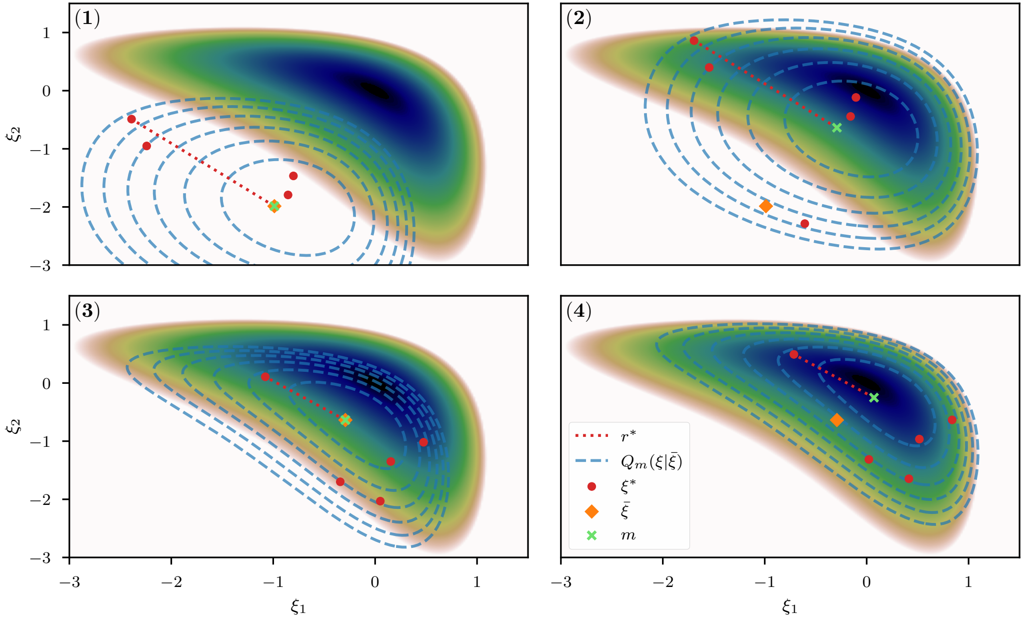
Variational approximation with transformed distribution $\mathcal{Q}$

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

Geometric Variational Inference (geoVI) [FLE21]

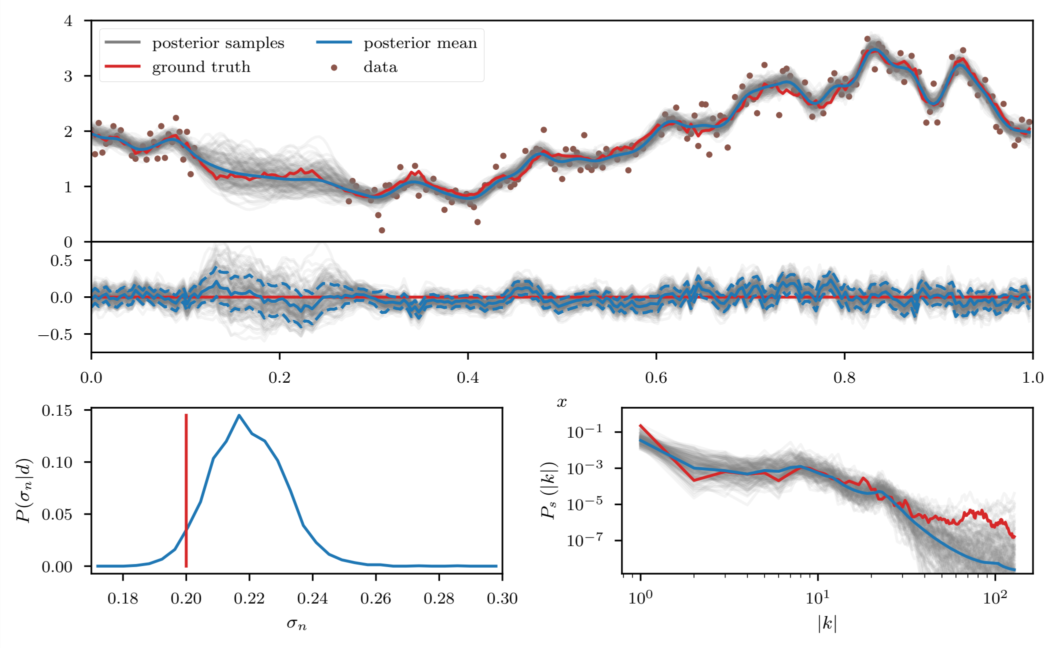


Geometric Variational Inference (geoVI) [FLE21]

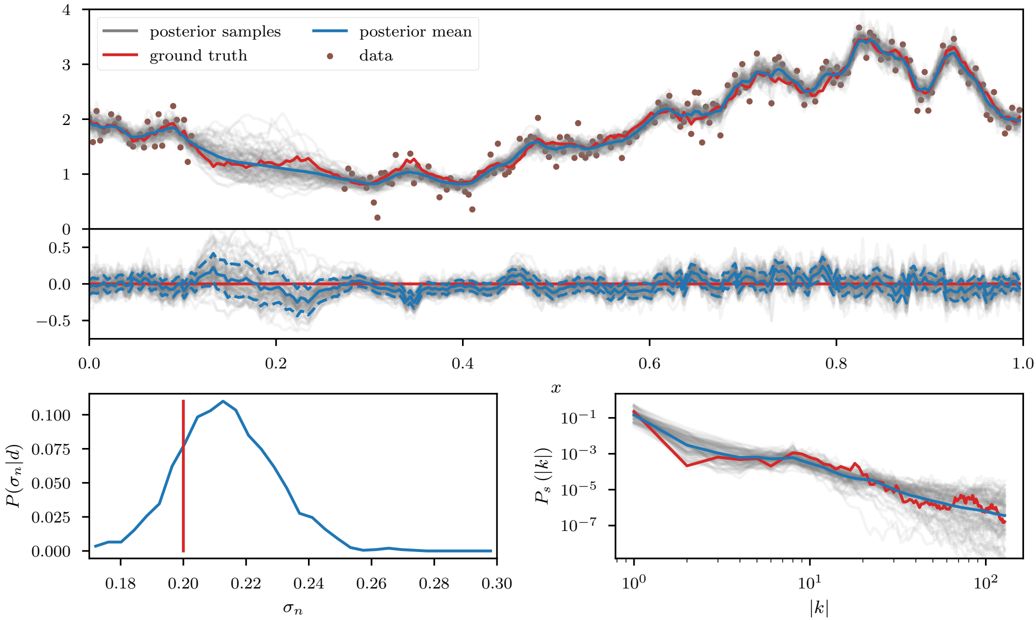


Applications

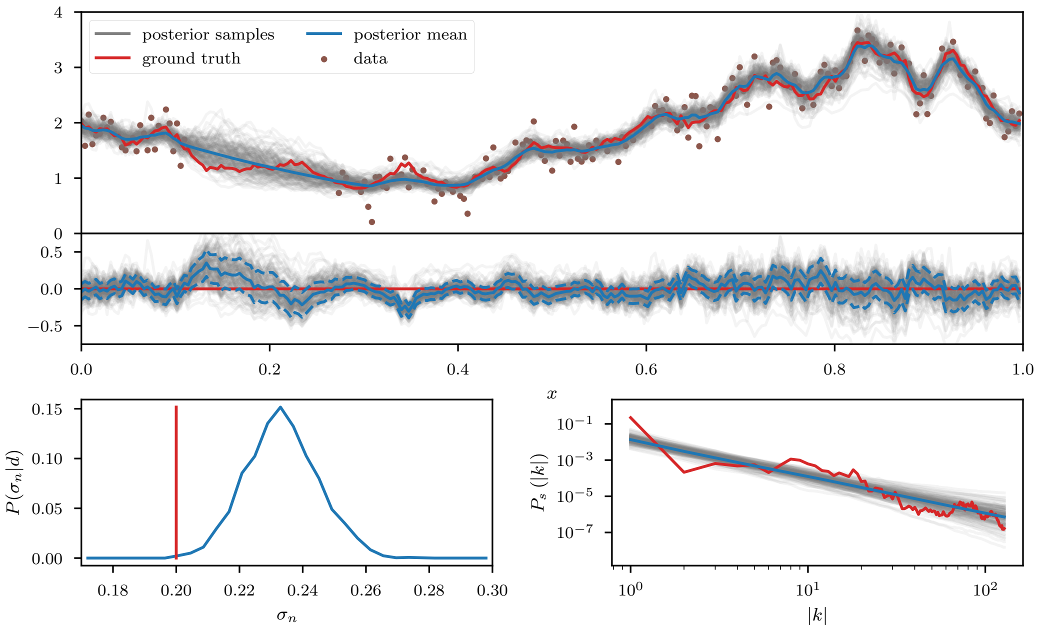
Applications



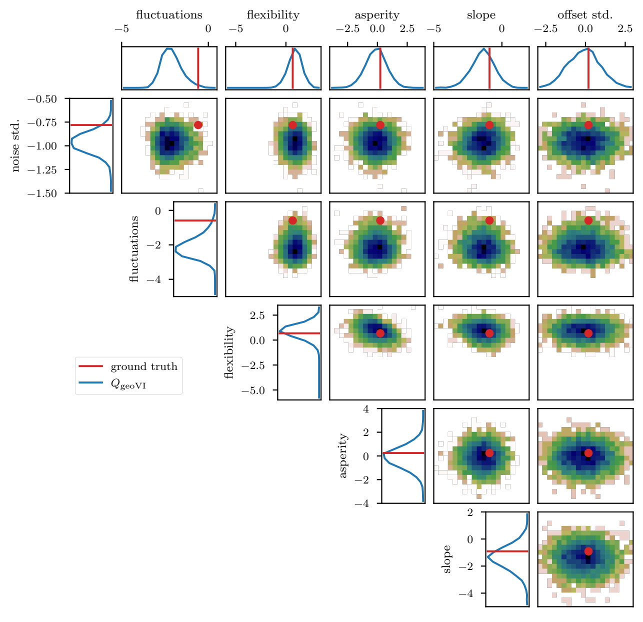
Applications



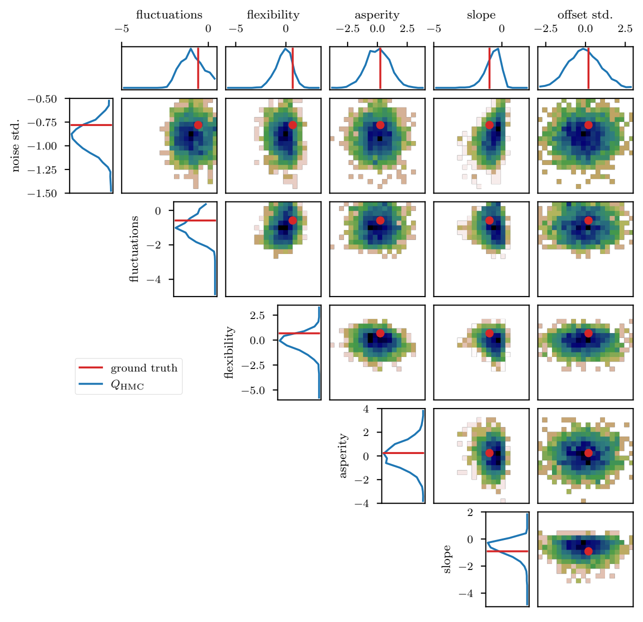
Applications



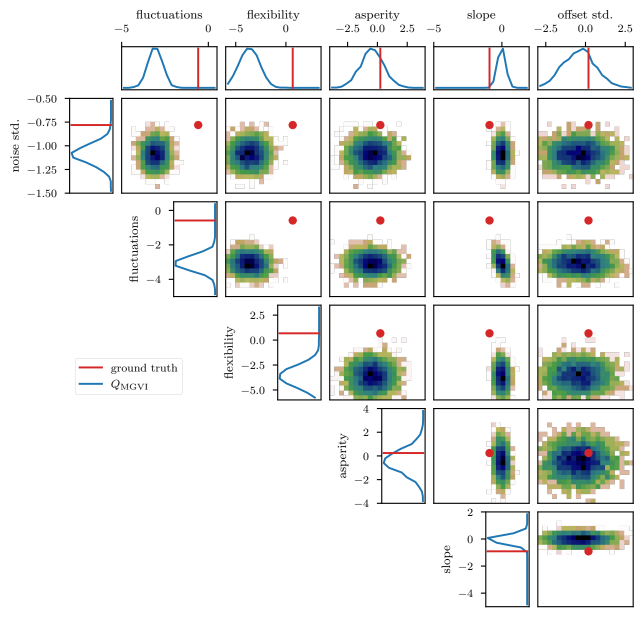
Applications



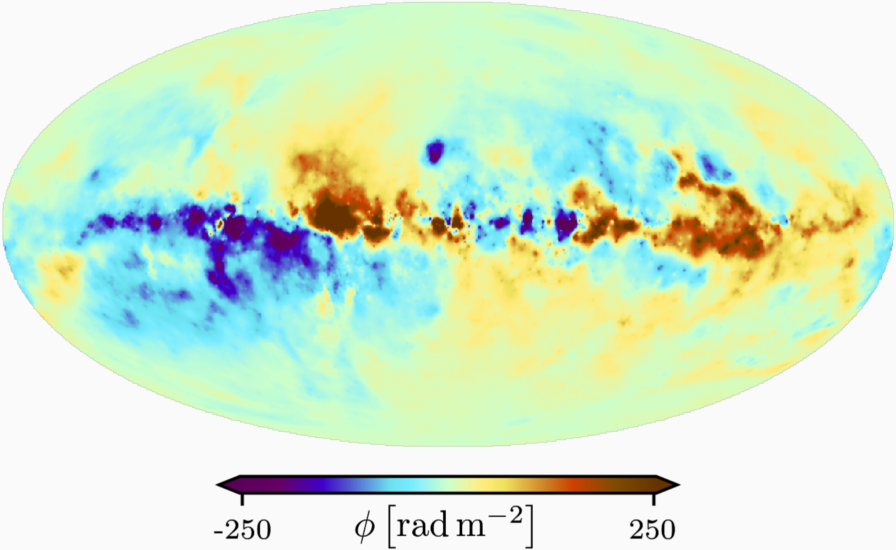
Applications



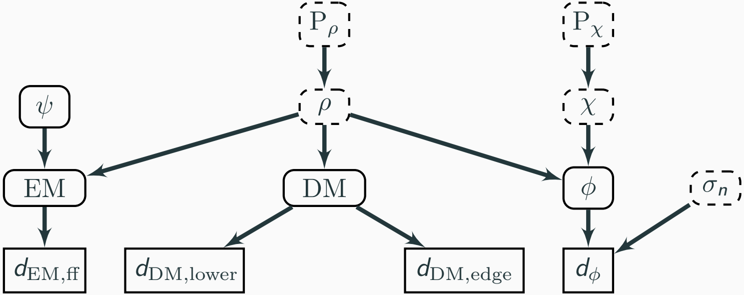
Applications



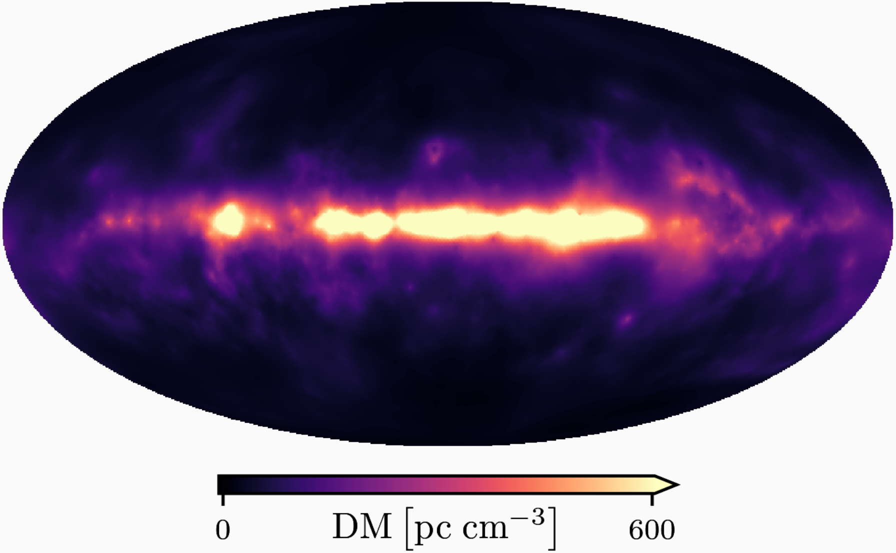
Applications [HE20]



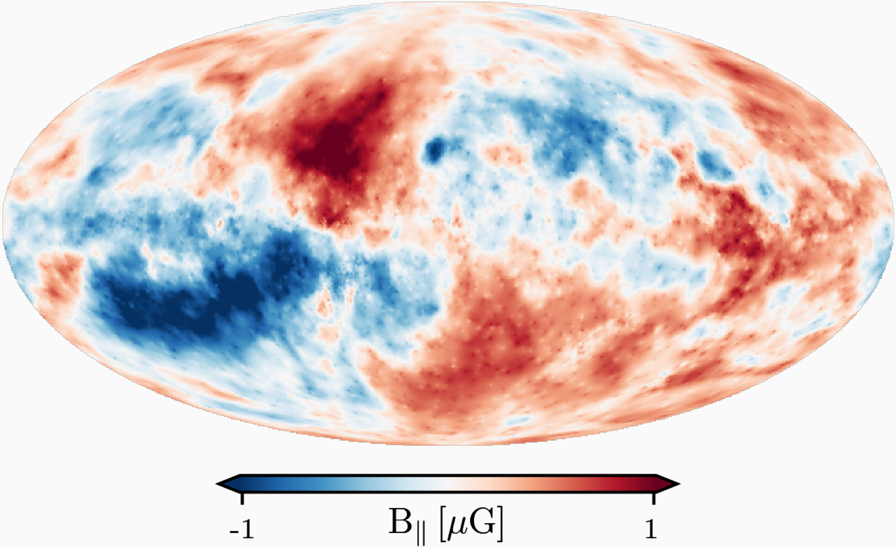
Applications [HE20]



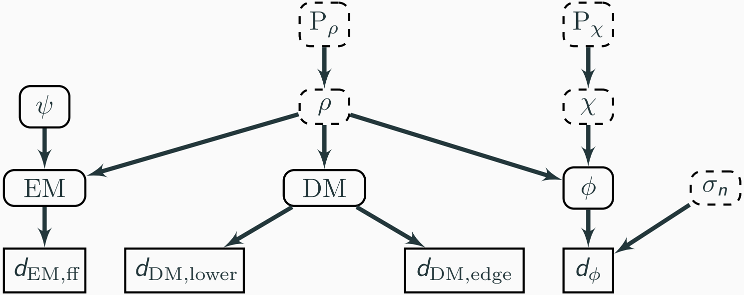
Applications [HE20]



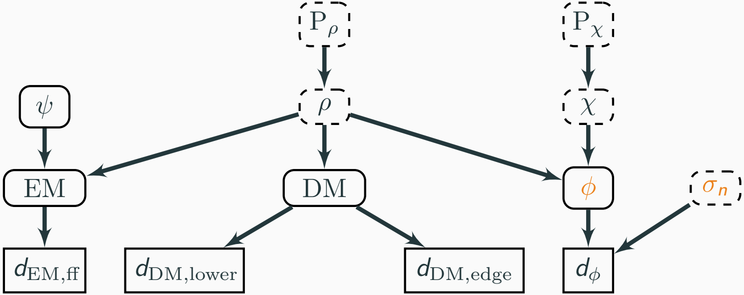
Applications [HE20]



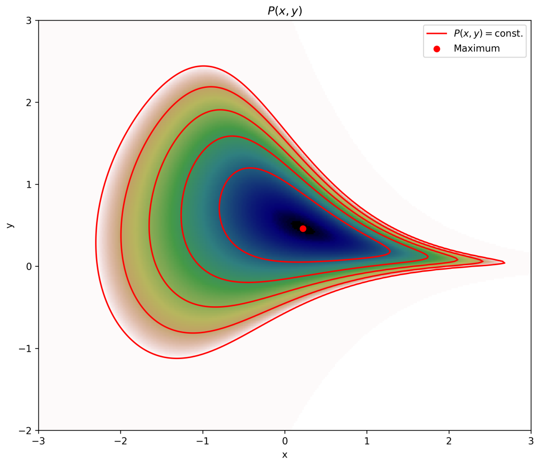
Applications [HE20]



Applications [HE20]



Probability measures



References

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.

Appendix

Conclusion - Code

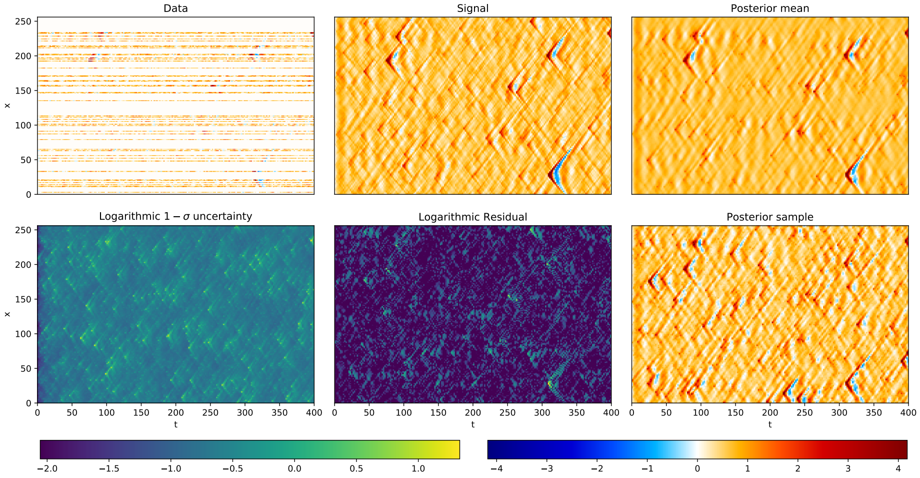
Numerical information field theory

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

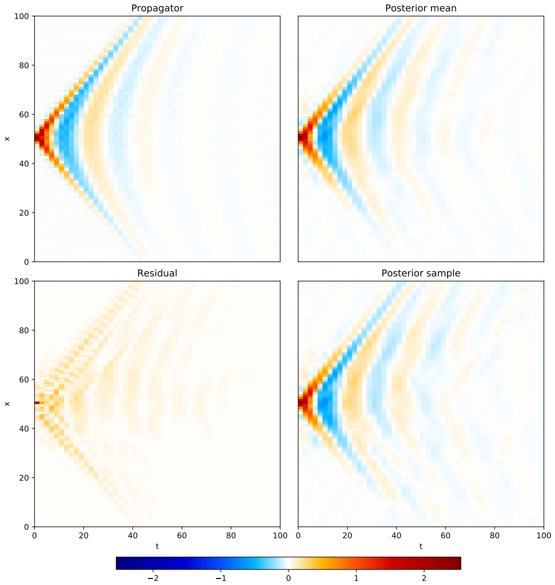
NIFTy tutorial - this afternoon!

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

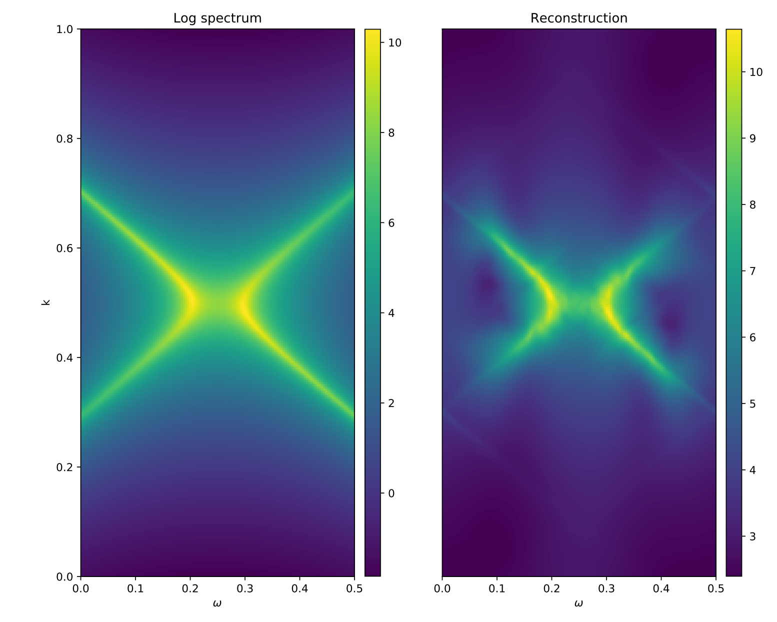
Application



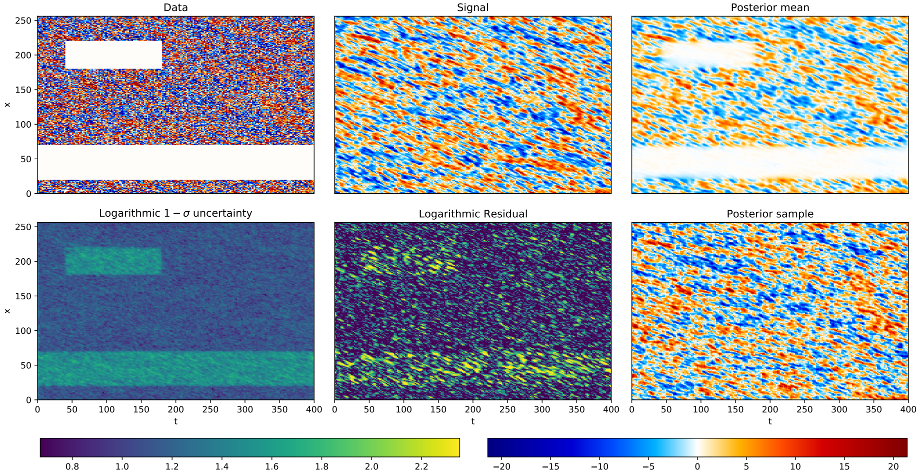
Application



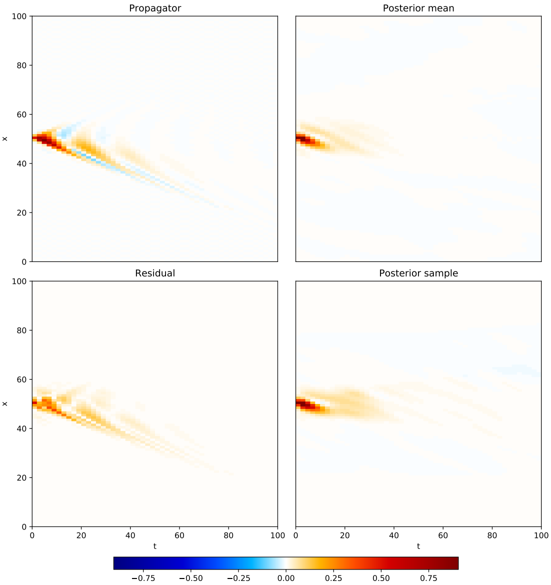
Application



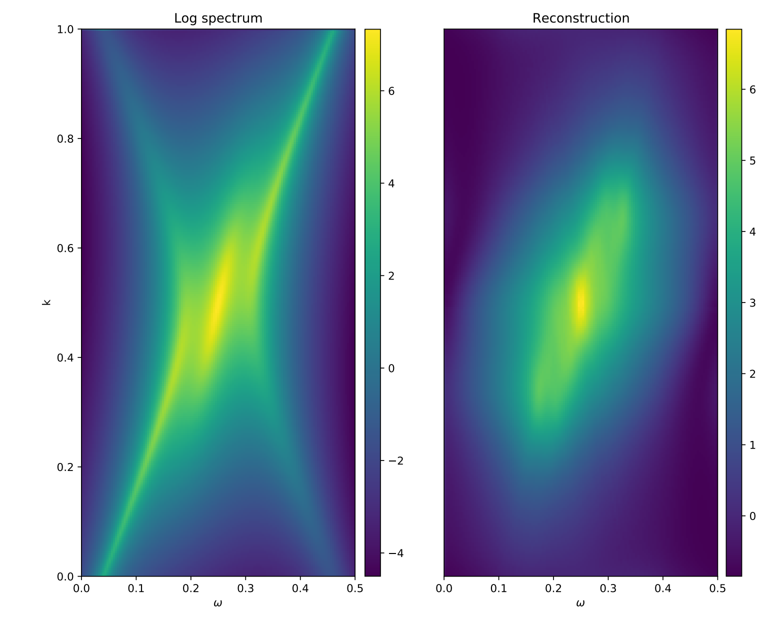
Application



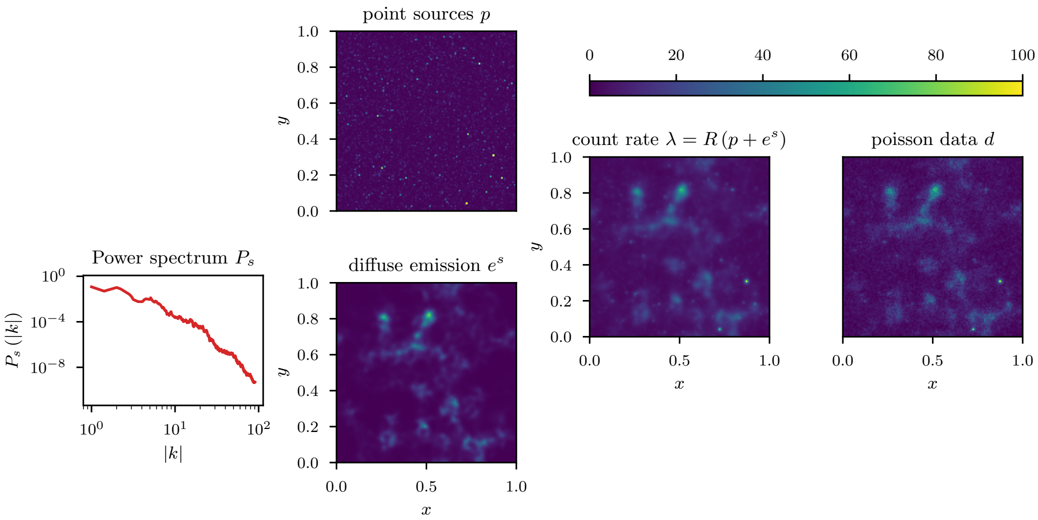
Application



Application

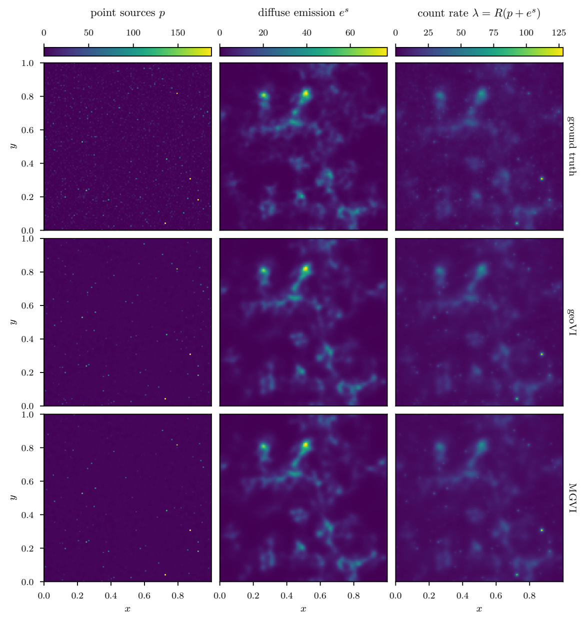


Appendix



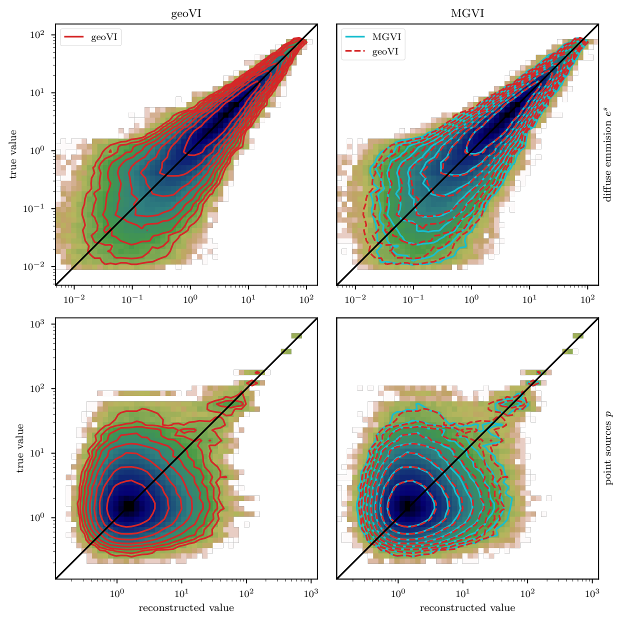
Power spectrum P_s
$P_s(|k|)$
10^0
10^{-4}
10^{-8}
10^0
10^1
10^2
$|k|$
point sources p
diffuse emission e^s
count rate $\lambda = R(p + e^s)$
poisson data d
0
20
40
60
80
100
1.0
0.8
0.6
0.4
0.2
0.0
0.5
x
y

Appendix



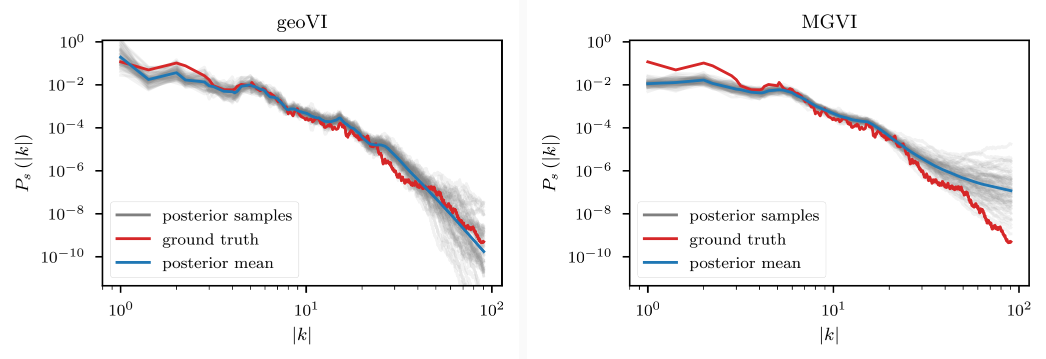
point sources p
diffuse emission e^s
count rate $\lambda = R(p + e^s)$
0 50 100 150
0 20 40 60
0 25 50 75 100 125
ground truth
geoVI
MGVI
y
x
0.0 0.2 0.4 0.6 0.8 1.0

Appendix

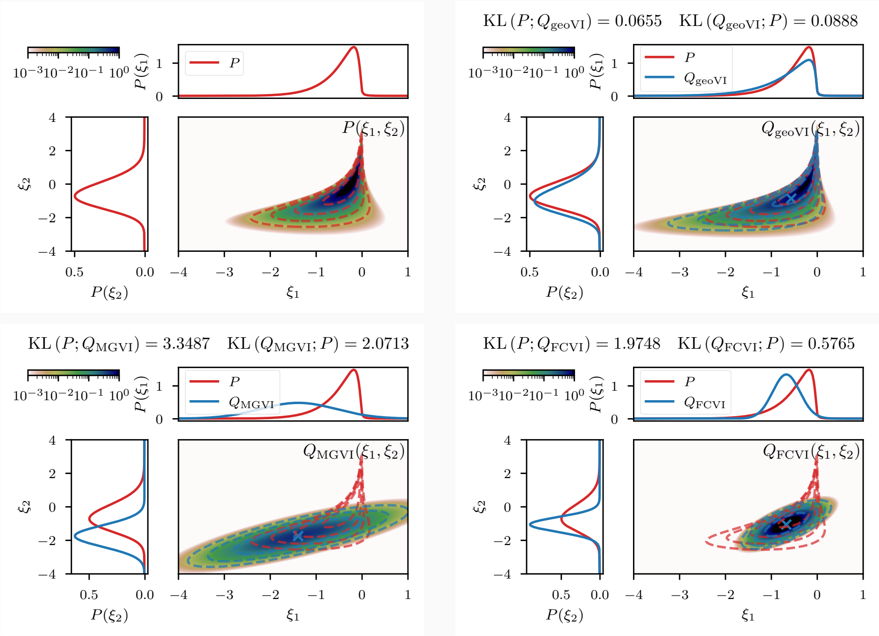


geoVI
MGVI
geoVI
MGVI
geoVI
true value
diffuse emmision e^s
point sources p
reconstructed value
reconstructed value

Appendix

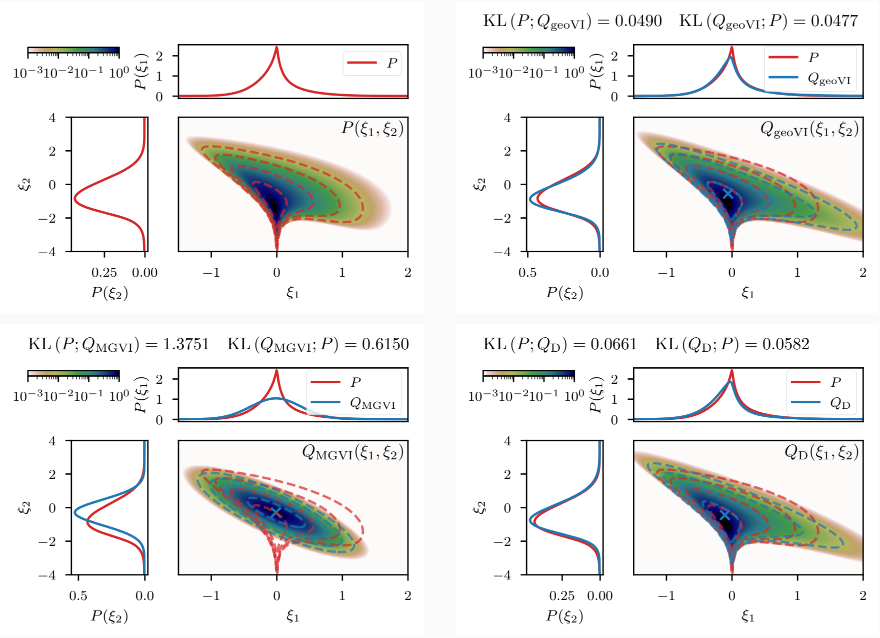

geoVI
MGVI
10^0
10^{-2}
10^{-4}
10^{-6}
10^{-8}
10^{-10}
$P_s(|k|)$
$|k|$
10^0
10^1
10^2
posterior samples
ground truth
posterior mean

Appendix



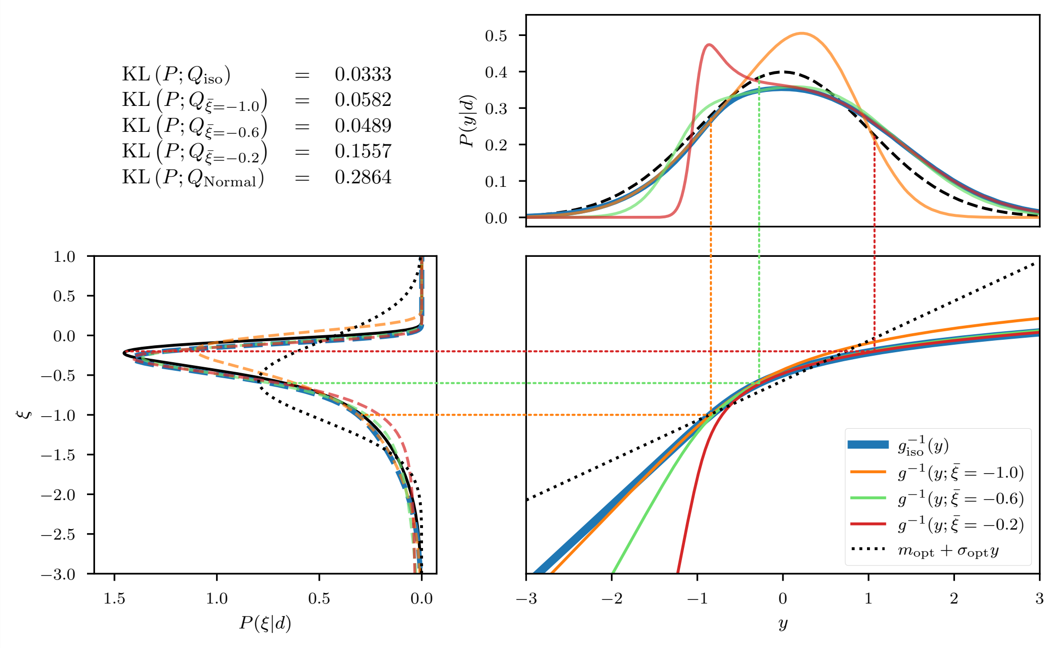
KL $(P; Q_{\mathrm{geoVI}}) = 0.0655$ KL $(Q_{\mathrm{geoVI}}; P) = 0.0888$
KL $(P; Q_{\mathrm{MGVI}}) = 3.3487$ KL $(Q_{\mathrm{MGVI}}; P) = 2.0713$
KL $(P; Q_{\mathrm{FCVI}}) = 1.9748$ KL $(Q_{\mathrm{FCVI}}; P) = 0.5765$
10^{-3} 10^{-2} 10^{-1} 10^{0}
P
Q_{geoVI}
Q_{MGVI}
Q_{FCVI}
$P(\xi_1)$
$P(\xi_2)$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

Appendix



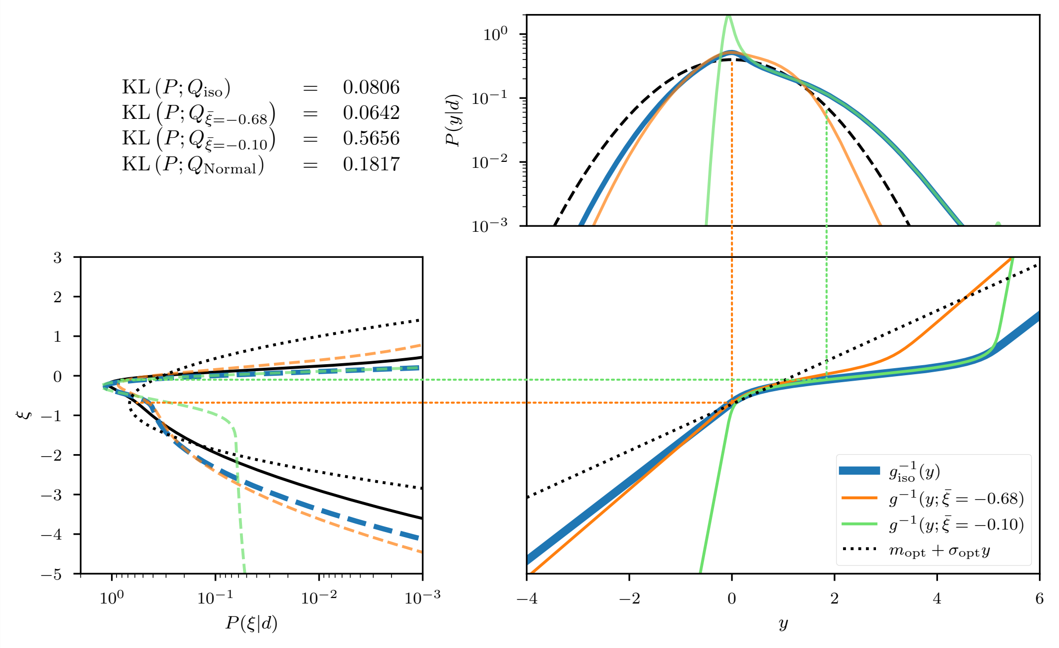
KL $(P; Q_{\mathrm{geoVI}}) = 0.0490$ KL $(Q_{\mathrm{geoVI}}; P) = 0.0477$
KL $(P; Q_{\mathrm{MGVI}}) = 1.3751$ KL $(Q_{\mathrm{MGVI}}; P) = 0.6150$
KL $(P; Q_{\mathrm{D}}) = 0.0661$ KL $(Q_{\mathrm{D}}; P) = 0.0582$
$P(\xi_1, \xi_2)$
$Q_{\mathrm{geoVI}}(\xi_1, \xi_2)$
$Q_{\mathrm{MGVI}}(\xi_1, \xi_2)$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

Appendix



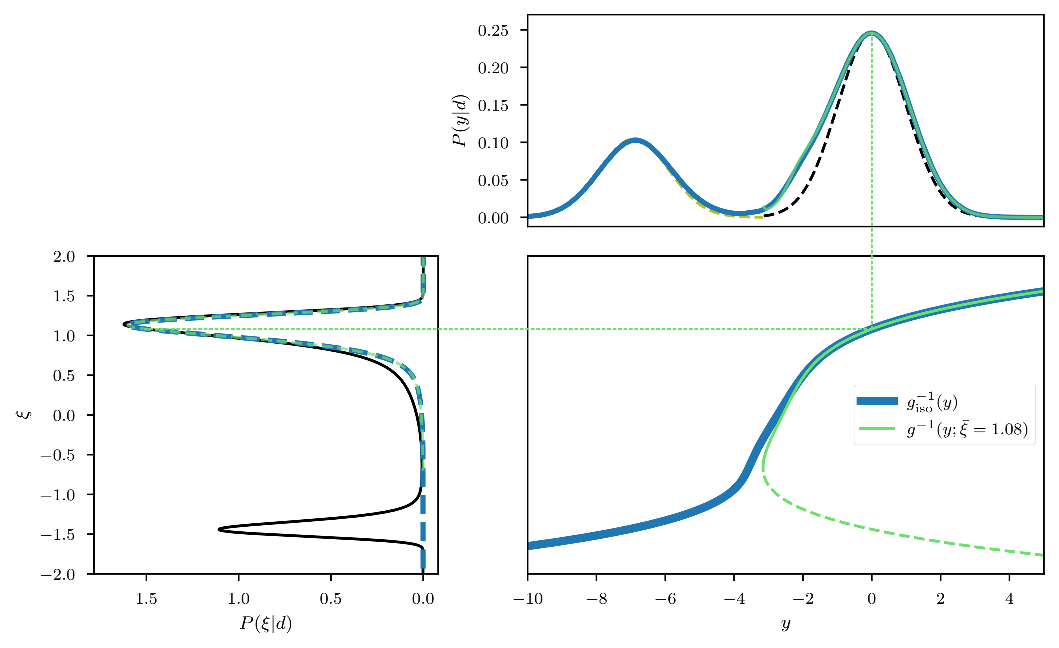
KL (P; Q_iso) = 0.0333
KL (P; Q_ξ̄=−1.0) = 0.0582
KL (P; Q_ξ̄=−0.6) = 0.0489
KL (P; Q_ξ̄=−0.2) = 0.1557
KL (P; Q_Normal) = 0.2864
P(y|d)
0.5
0.4
0.3
0.2
0.1
0.0
1.0
0.5
0.0
−0.5
−1.0
−1.5
−2.0
−2.5
−3.0
ξ
1.5
1.0
0.5
0.0
P(ξ|d)
−3
−2
−1
0
1
2
3
y
g_iso^−1(y)
g^−1(y; ξ̄ = −1.0)
g^−1(y; ξ̄ = −0.6)
g^−1(y; ξ̄ = −0.2)
m_opt + σ_opt y

Appendix



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