Mixture of noises and sampling non log-concave distributions

Pierre Palud[1,2], F. Le Petit[1], P. Chainais[2], E. Bron[1], P.-A. Thouvenin[2], ORIONB consortium
[1] LERMA, Obs. de Paris, PSL Research Univ., CNRS, Meudon
[2] Univ. Lille, CNRS, Centrale Lille, UMR 9189 - CRIStAL, France

1/ Aims and contributions

Real life inverse problems: **multiple noises**, **non-linear** models, need for **uncertainty** quantification.

Bayesian inference with
- ➙ Mixture of noises
- ➙ Non-log-concave posterior (multimodal)
- ➙ Non-gradient-Lipschitz log-posterior

★ Likelihood approximation for a mixture of noises,
★ New kernel to efficiently sample from posterior,
★ Application to astrophysical data.

2/ Observation model

➙ **Forward model** $f : \boldsymbol{\Theta} \in \mathbb{R}^{ND} \mapsto \boldsymbol{Y} \in \mathbb{R}^{NL}$, strictly positive, twice differentiable, covers multiple decades.

➙ **Observation model**:

$$y_{n,\ell} = \max \left\{ \omega,\ \epsilon^{(m)}_{n,\ell} f_\ell(\boldsymbol{\theta}_n) + \epsilon^{(a)}_{n,\ell} \right\}$$

$$\text{with } \begin{cases} \omega > 0 \\ \epsilon^{(a)}_{n,\ell} \sim \mathcal{N}\left(0, \sigma_a^2\right) \text{ i.i.d.} \\ \epsilon^{(m)}_{n,\ell} \sim \log\mathcal{N}\left(0, \sigma_m^2\right) \text{ i.i.d.} \end{cases}$$

$\Longrightarrow$ untractable likelihood $\pi\left(\boldsymbol{Y} \mid \boldsymbol{\Theta}\right)$.

➙ Literature often neglects one noise [1]. But when f covers many decades: dominant noise depends on $\boldsymbol{\theta}_n$.
$\Longrightarrow$ need to address full mixture model.

How to **approximate the likelihood** with **controlled error**? → See 4/

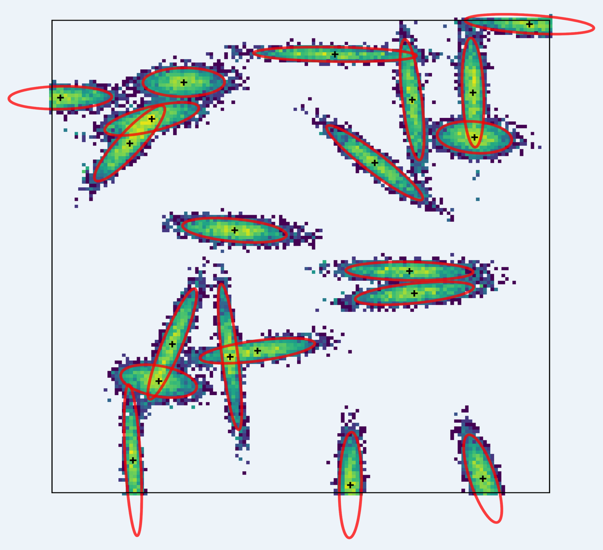
3/ Bayesian approach & sampling

➙ Combine likelihood with informative **prior** $\pi(\boldsymbol{\Theta})$: yields **Posterior** distribution:

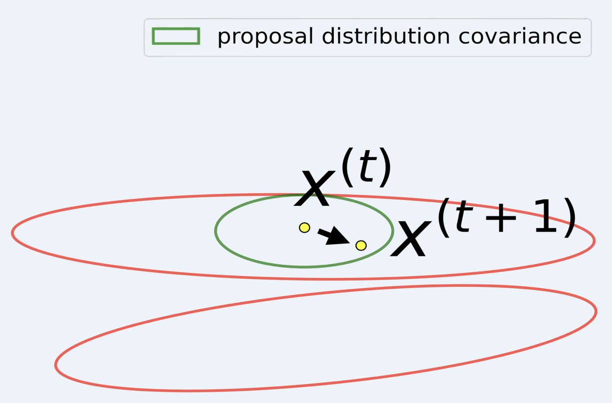
$$\pi(\boldsymbol{\Theta} \mid \boldsymbol{Y}) \propto \pi(\boldsymbol{Y} \mid \boldsymbol{\Theta})\ \pi(\boldsymbol{\Theta})$$

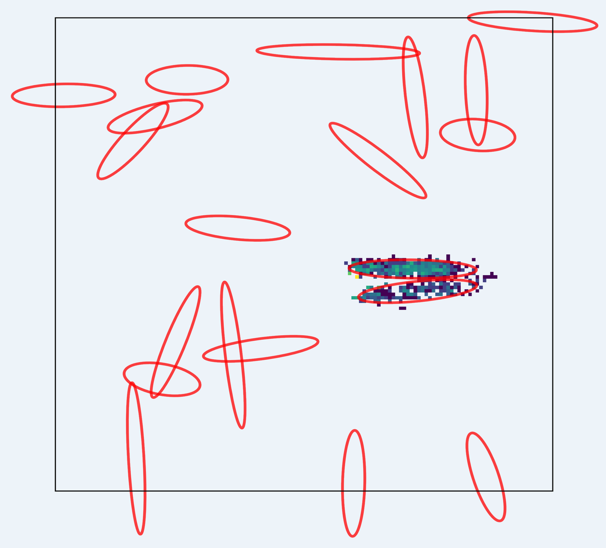
➙ Sampling: MCMC
$\Longrightarrow$ uncertainty quantification

➙ Model: Non-linear
$\Longrightarrow$ **multimodal** distribution

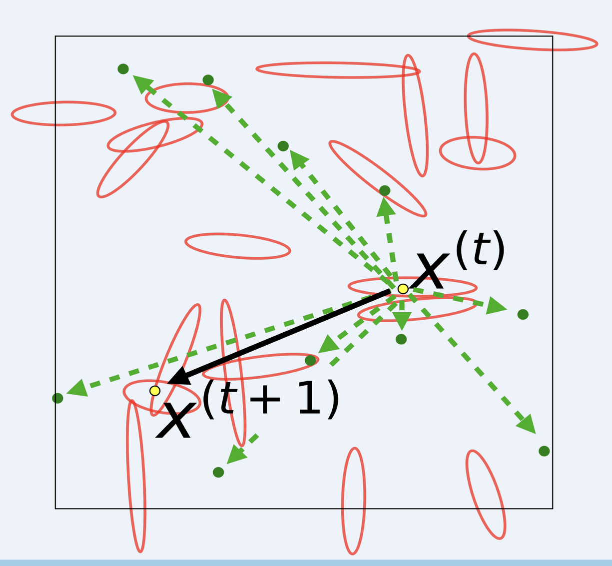


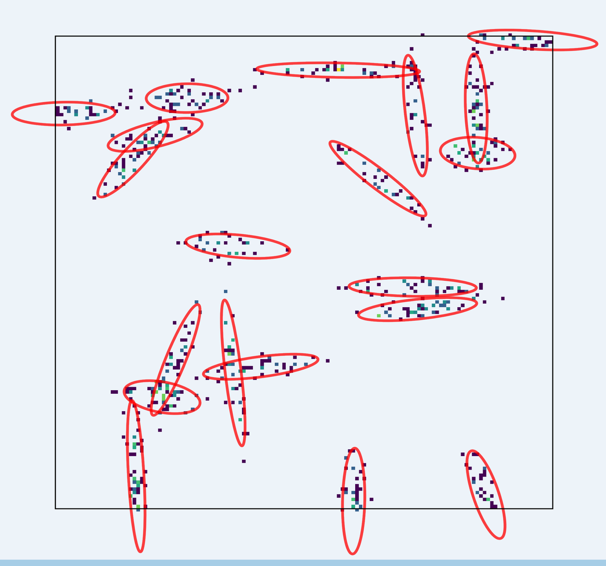
✓ Preconditioned-MALA [2]: local exploration





✓ MTM [3]: jumps between minima.





4/ Mixture of noises: Likelihood approximation

1. Build 2 likelihood approximations with moment matching:

Gaussian approx $\pi^{(a)}$	lognormal approx $\pi^{(m)}$
$y_{n,\ell} \simeq f_\ell(\boldsymbol{\theta}_n) + e^{(a)}_{n,\ell}$	$y_{n,\ell} \simeq e^{(m)}_{n,\ell} f_\ell(\boldsymbol{\theta}_n)$
$e^{(a)}_{n,\ell} \sim \mathcal{N}(m_{a,n,\ell}, s^2_{a,n,\ell})$	$e^{(m)}_{n,\ell} \sim \log\mathcal{N}(m_{m,n,\ell}, s^2_{m,n,\ell})$

2. Combine approximations with weight function $\lambda_\ell^{\boldsymbol{a}_\ell}$:

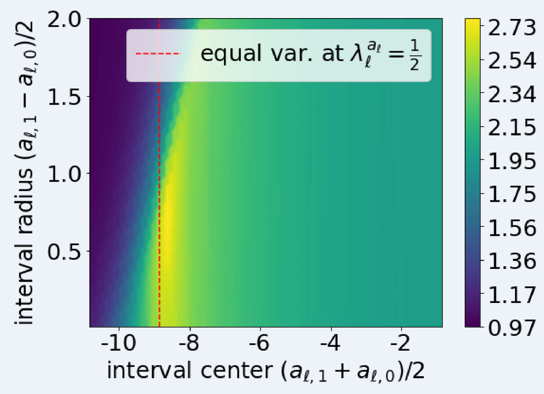
$$\tilde{\pi}_{\boldsymbol{a}_\ell}(y_{n,\ell}|\boldsymbol{\theta}_n) \propto \underbrace{\pi^{(a)}(y_{n,\ell}|\boldsymbol{\theta}_n)^{1-\lambda_\ell^{\boldsymbol{a}_\ell}(\boldsymbol{\theta}_n)}}_{\text{Gaussian approx}} \underbrace{\pi^{(m)}(y_{n,\ell}|\boldsymbol{\theta}_n)^{\lambda_\ell^{\boldsymbol{a}_\ell}(\boldsymbol{\theta}_n)}}_{\text{lognormal approx}}$$

with $\lambda_\ell^{\boldsymbol{a}_\ell} \in [0,1]$, twice differentiable, parametrized with $\boldsymbol{a}_\ell \in \mathbb{R}^2$
$\Longrightarrow$ negative log of this approx: easy to work with.

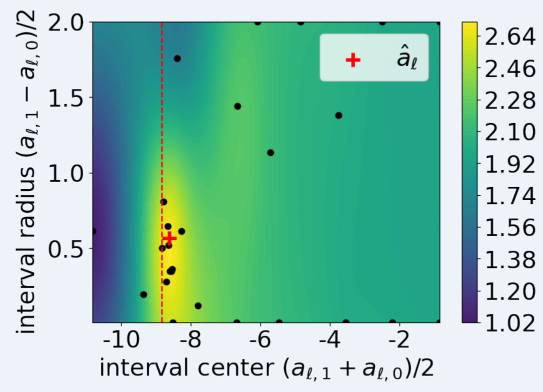
Tuning $\boldsymbol{a}_\ell$:
Approx error = Kolmogorov-Smirnov-based metric $\varphi(\boldsymbol{a}_\ell)$.
φ minimized with Bayesian Optimization (BO).

➙ From approx error: $\tilde{\pi}_{\boldsymbol{a}_\ell}$ **better than** $\pi^{(a)}$ **or** $\pi^{(m)}$ **alone**.

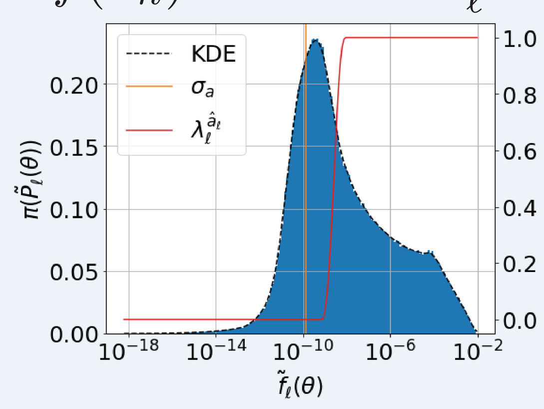
grid eval of $-\log_{10}\varphi$



BO, 25 points



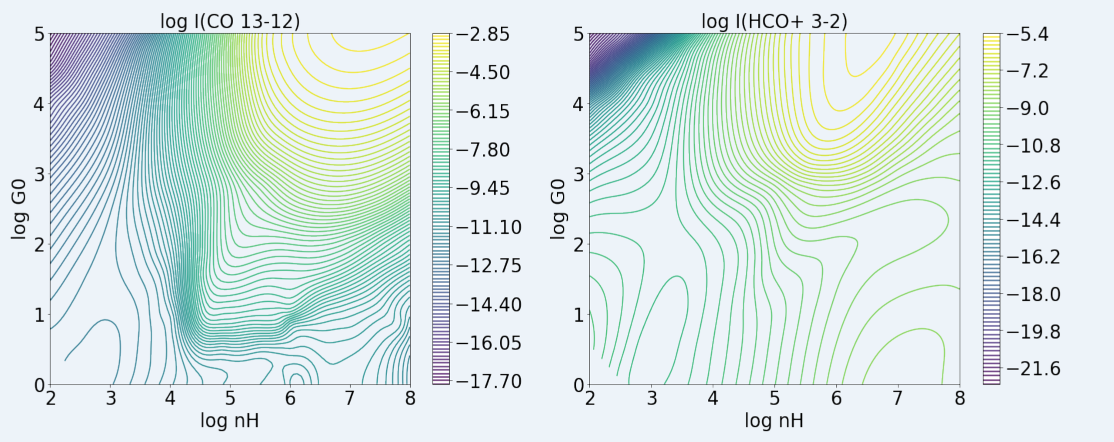
$f(\boldsymbol{\theta}_n)$ hist. and $\lambda_\ell^{\boldsymbol{a}_\ell}$



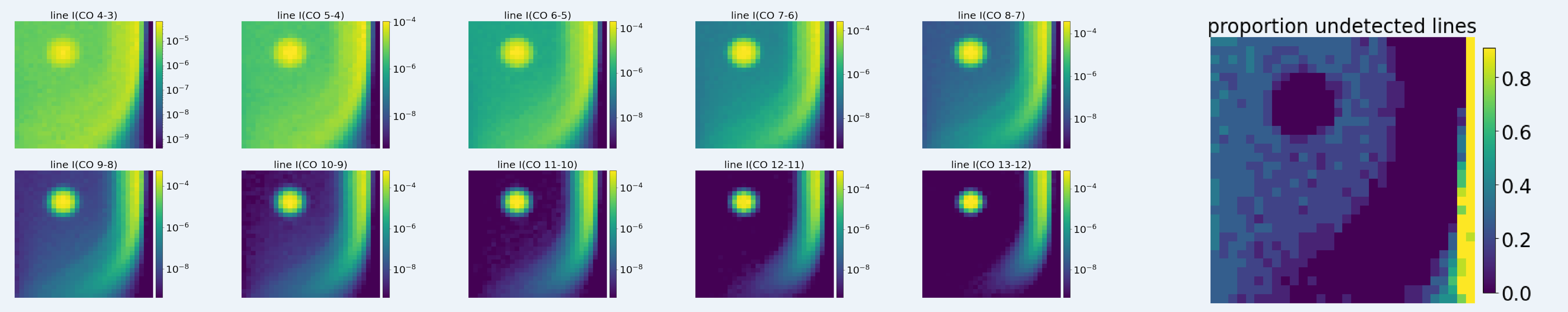
5/ Astrophysics synthetic dataset

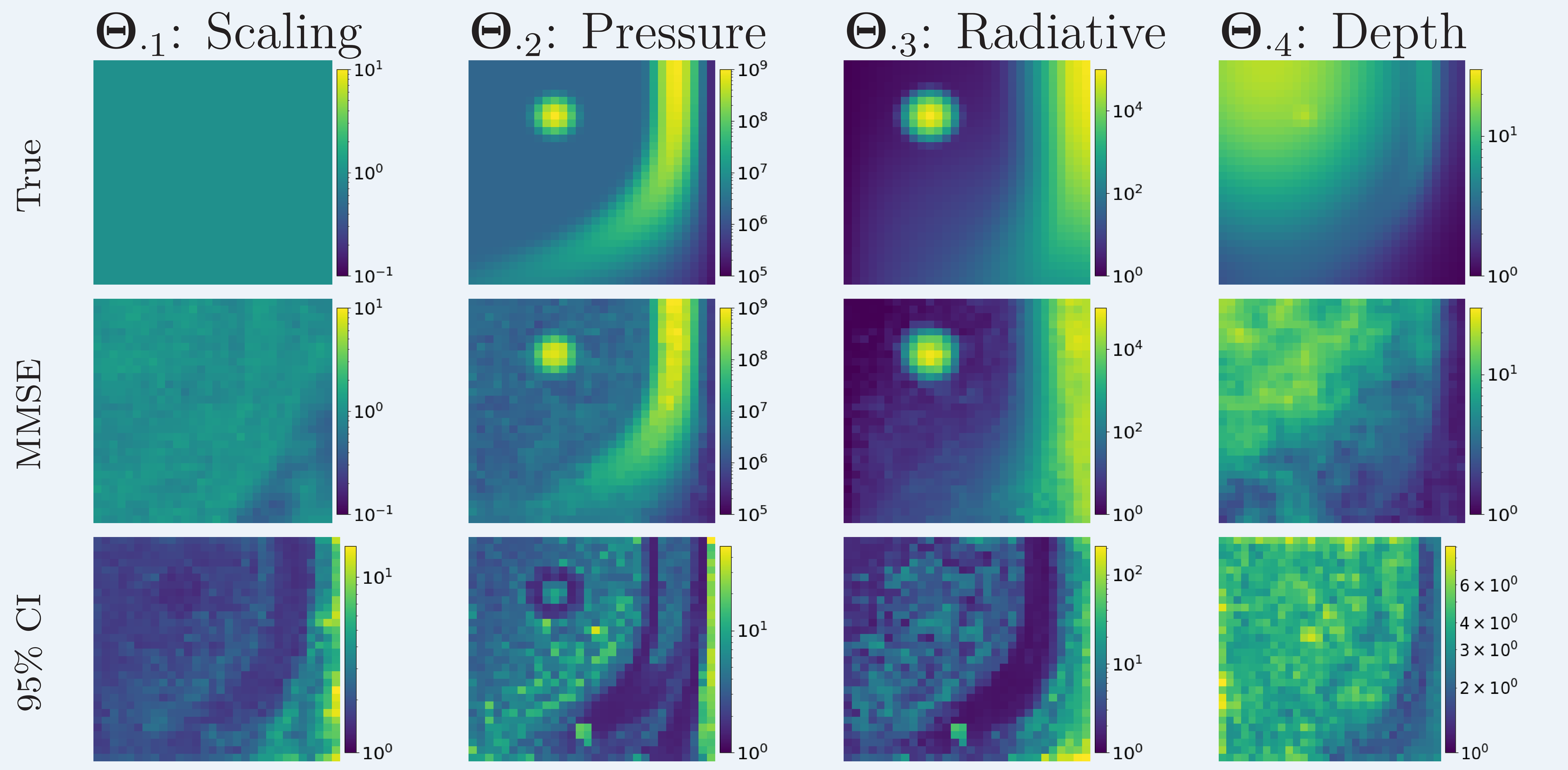
➙ $\boldsymbol{\Theta} \in \mathbb{R}^{900 \times 4}$: high dimensional
➙ $\pi(\boldsymbol{\Theta})$: spatial (Laplacian L_2-norm)

Forward model: Meudon PDR code [4]
✗ non-linear
✗ non-gradient-Lipschitz



Synthetic observations $\boldsymbol{Y} \in \mathbb{R}^{900 \times 10}$: integrated intensities of excited lines of CO





6/ Conclusion

✓ **Mixture of noises** $\Longrightarrow$ likelihood approx.
✓ **Non-log-concave** posterior $\Longrightarrow$ MTM kernel.
✓ **Non-gradient-Lipschitz** log-posterior $\Longrightarrow$ P-MALA kernel.
✓ Application on astrophysical inverse problem.
➙ Application to Orion-B data, James Webb Spatial Telescope, etc.

References

[1] Nicholson and Kaipio, *An additive Approximation to Multiplicative Noise*, 2018
[2] Xifara *et al.*, *Langevin diffusions and the Metropolis-adjusted Langevin algorithm*, 2014
[3] Martino, *A review of multiple try MCMC algorithms for signal processing*, 2018
[4] Le Petit *et al.*, *A Model for Atomic and Molecular Interstellar Gas: The Meudon PDR Code*, 2006