Mixture of noises and sampling non-log-concave posterior distributions
– EUSIPCO 2022 –

Pierre Palud

Pierre Chainais, Franck Le Petit

Emeric Bron, Pierre-Antoine Thouvenin

Ecole Centrale de Lille, CRIStAL, LERMA
Observatory of Paris, Meudon,
funded by CNRS via 80|Prime

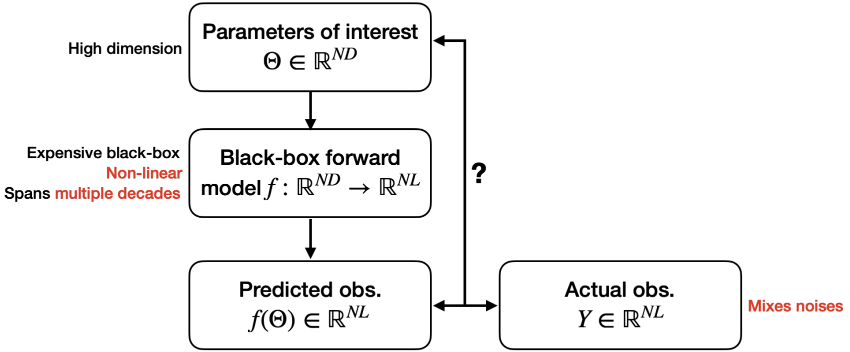
Real life inverse problems

Considered general inverse problem

	Log-likelihood	**Proposition**
Mixture of noises	Untractable	Approx.
Forward model:		
➙ expensive black-box	-	Model reduction
➙ Non-linear	Non-concave	MTM kernel
➙ Spans multiple decades	Non-grad-Lipschitz	P-MALA kernel
Uncertainty quantification	-	MCMC

Limitation: Sampler restricted to smooth log-posterior because of P-MALA (see slide 9)

Considered general inverse problem



Observation model

$$y_{n,\ell} = \max\left\{\omega,\ \epsilon_{n,\ell}^{(m)}\, f_{n,\ell}(\boldsymbol{\Theta}) + \epsilon_{n,\ell}^{(a)}\right\}$$

$\boldsymbol{\Theta}$	parameters to infer
$f_{n,\ell}$	black-box, spans multiple decades (element (n,ℓ))
$\epsilon_{n,\ell}^{(a)} \sim \mathcal{N}(0, \sigma_a^2)$	e.g., instruments noise
$\epsilon_{n,\ell}^{(m)} \sim \log\mathcal{N}(0, \sigma_m^2)$	e.g., calibration error
$\omega > 0$	instrument detectability limit

How to deal with
- black-box forward map f ?
- mixture of additive and multiplicative noises?

Addressing likelihood difficulties

How to deal with black-box forward map f?

➙ Model reduction

How to deal with mixture of additive and multiplicative noises?

➙ likelihood approximation with controlled error

Deriving a new Likelihood (uncensored)

$$\boldsymbol{Y} = \boldsymbol{\epsilon^{(m)}} f(\boldsymbol{\Theta}) + \boldsymbol{\epsilon^{(a)}}$$

	Additive approx	**Multiplicative approx**
Approximated model	$\boldsymbol{Y} \simeq f(\boldsymbol{\Theta}) + \boldsymbol{e^{(a)}}$	$\boldsymbol{Y} \simeq \boldsymbol{e^{(m)}} f(\boldsymbol{\Theta})$
Moment matching	$\boldsymbol{e^{(a)}} \sim \mathcal{N}(\boldsymbol{m}_a, \boldsymbol{s}_a^2)$	$\boldsymbol{e^{(m)}} \sim \log\mathcal{N}(\boldsymbol{m}_m, \boldsymbol{s}_m^2)$
likelihood approx (1 elt)	$\pi^{(a)}(y_{n,\ell} \vert \boldsymbol{\Theta})$	$\pi^{(m)}(y_{n,\ell} \vert \boldsymbol{\Theta})$

Deriving a new Likelihood (uncensored)

$$\tilde{\pi}(y_{n,\ell}|\mathbf{\Theta}) \quad \propto \quad \pi^{(a)}(y_{n,\ell}|\mathbf{\Theta})^{1-\lambda_{n,\ell}} \quad \pi^{(m)}(y_{n,\ell}|\mathbf{\Theta})^{\lambda_{n,\ell}}$$

with $\lambda_{n,\ell} = \lambda\left(f_{n,\ell}(\mathbf{\Theta})\right)$: **controls the mixing of the two approx**
sigmoid parametrized by $\boldsymbol{a}_\ell$ (location and speed)

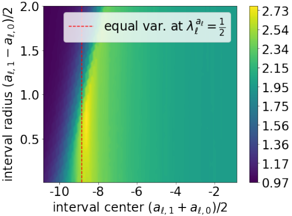
Deriving a new Likelihood (uncensored)

$$\tilde{\pi}(y_{n,\ell}|\mathbf{\Theta}) \quad \propto \quad \pi^{(a)}(y_{n,\ell}|\mathbf{\Theta})^{1-\lambda_{n,\ell}} \quad \pi^{(m)}(y_{n,\ell}|\mathbf{\Theta})^{\lambda_{n,\ell}}$$

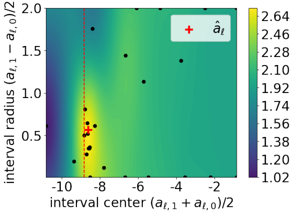
with $\lambda_{n,\ell} = \lambda\left(f_{n,\ell}(\mathbf{\Theta})\right)$: | **controls the mixing of the two approx**
sigmoid parametrized by $\boldsymbol{a}_\ell$ (location and speed)

To evaluate approx: Kolmogorov-Smirnov-based metric $\varphi(\boldsymbol{a}_\ell)$.
To get best approx: Minimize φ with Bayesian Optimization (BO).

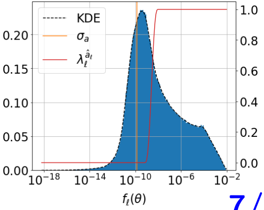
grid eval of $-\log_{10}\varphi$



BO after 25 iterations



best λ

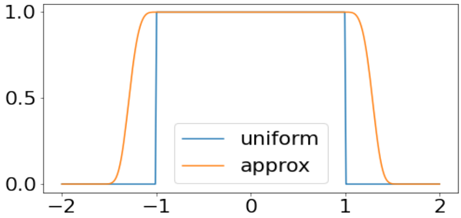


A priori information

a priori information on $\boldsymbol{\Theta} \in \mathbb{R}^{N \times D}$ combines 2 priors:

- **spatial regularization**, e.g.,
  - smoothed Total Variation
  - L_2-norm of image gradient
  - L_2-norm of image Laplacian
  - L_2-norm of image wavelet decomposition

- **Validity domain** for each physical parameter $\theta_{n,d}$
  $\Longrightarrow$ BUT non-smooth
  $\Longrightarrow$ use smooth penalty function when $\theta_{n,d}$ is out of validity domain:



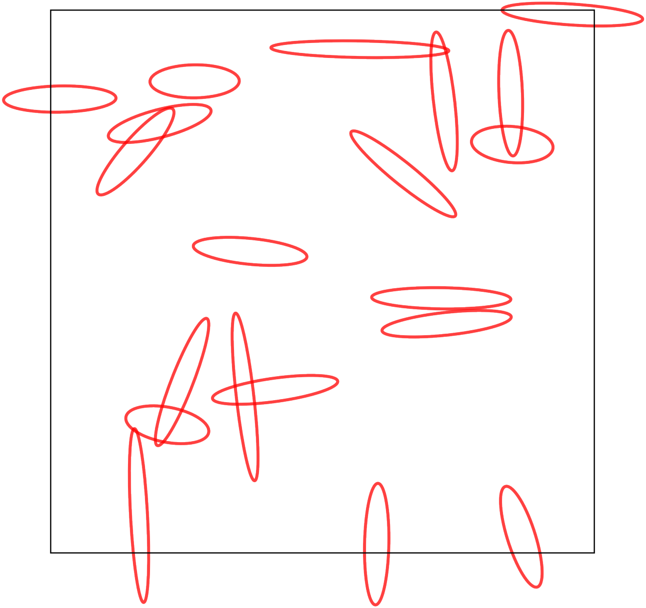
Sampler

Proposed sampler: **two kernels**:

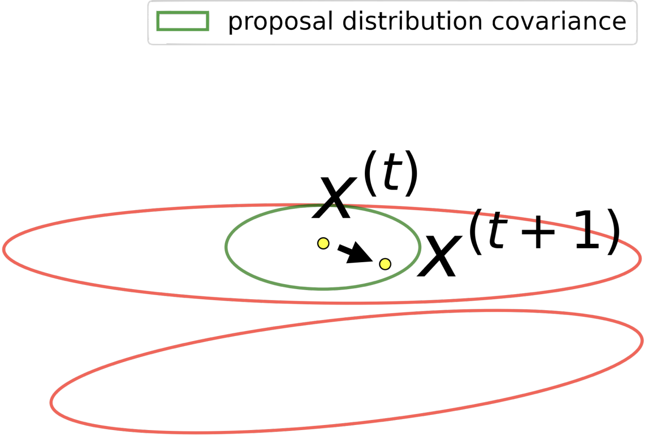
- Forward model covers multiple decades
  $\implies$ non-grad.-Lipschitz log-posterior
  Preconditioned-MALA kernel with RMSProp
  Role: Efficient local exploration
  Limitation: restricted to smooth log-posteriors

- Non-log-concave posterior
  $\implies$ potential multimodality
  Multiple-Try Metropolis (MTM) kernel
  Role: Jumps between modes

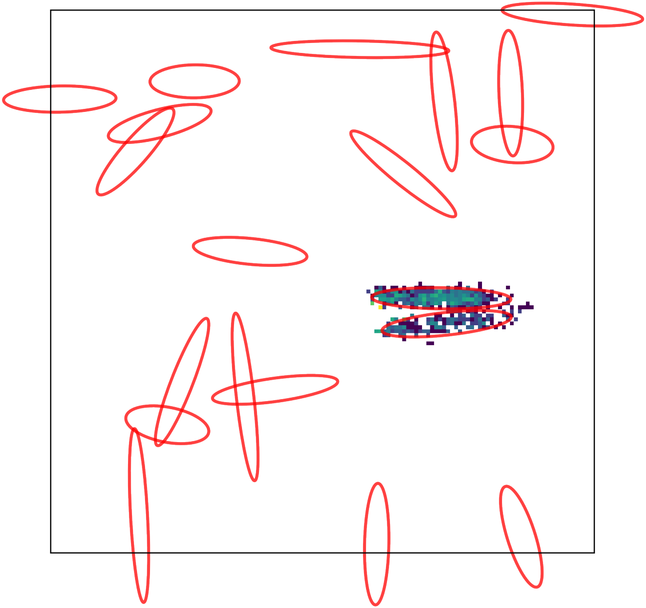
Sampler: illustration on Gaussian mixture model



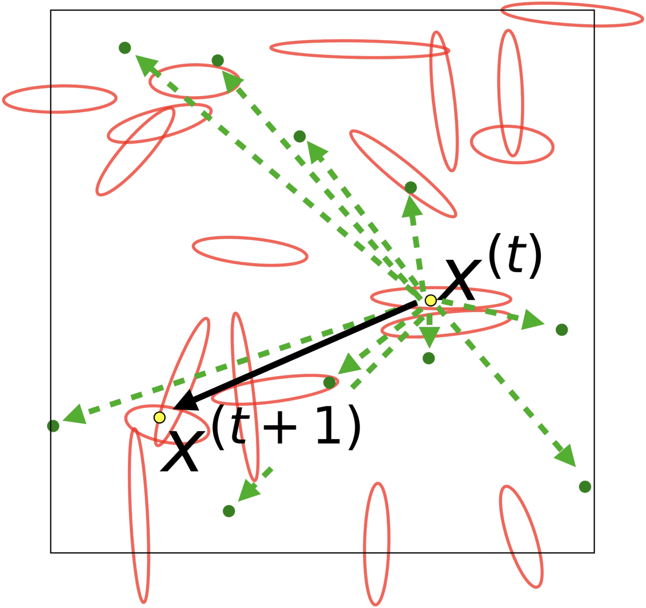
Sampler: illustration on Gaussian mixture model



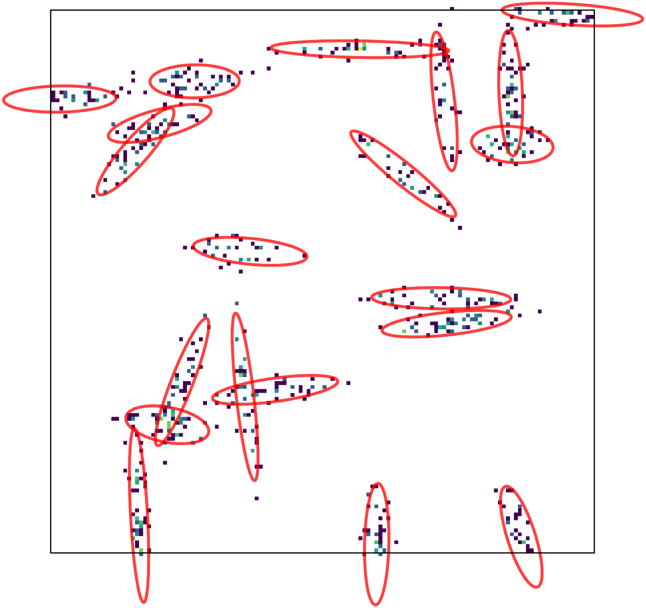
Sampler: illustration on Gaussian mixture model

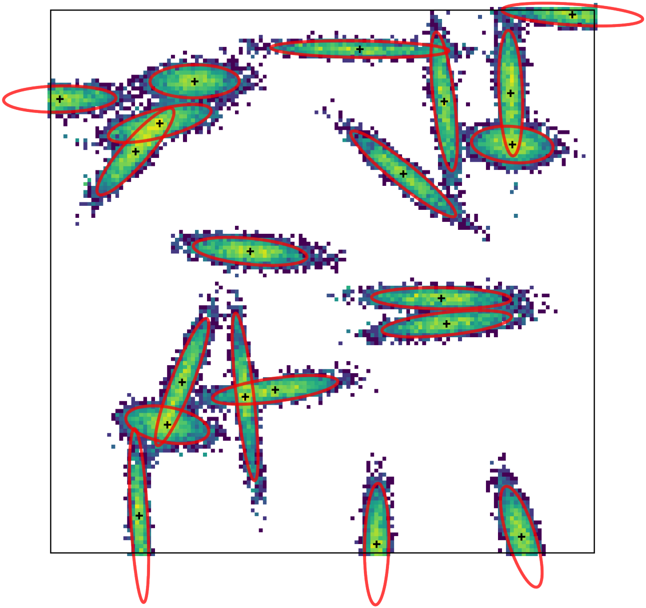


Sampler: illustration on Gaussian mixture model

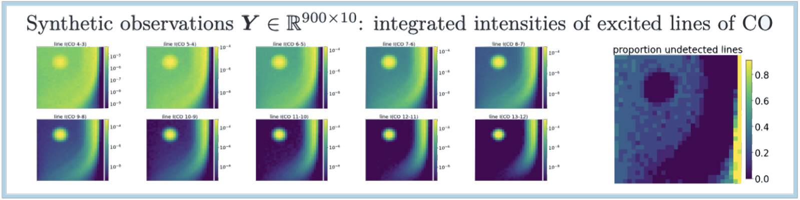


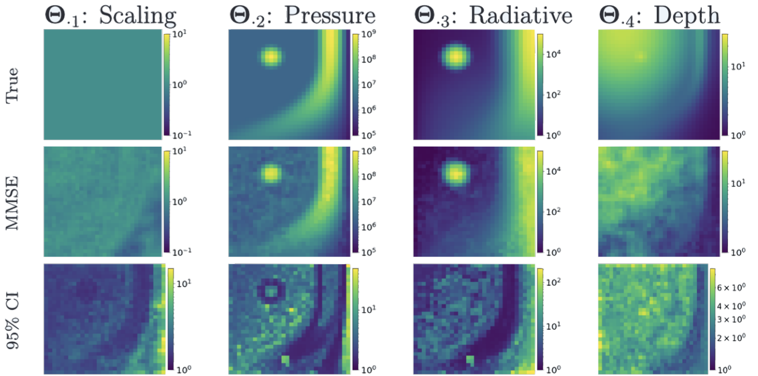
Sampler: illustration on Gaussian mixture model





Application to an astrophysics synthetic dataset





Conclusion

	Log-likelihood	Proposition
Mixture of noises	Untractable	Approx.
Forward model:		
➙ Black-box expensive		Model reduction
➙ Non-linear	Non-concave	MTM kernel
➙ Spans multiple decades	Non-grad-Lipschitz	P-MALA kernel

Applications ✓ Astrophysics synthetic yet realistic dataset
➙ OrionB data and JWST obs. (perspective)