Towards fast and scalable uncertainty quantification for scientific imaging

Tobías I. Liaudat
IRFU, CEA Paris-Saclay

NumPEx workshop Exa-DI

Artificial Intelligence for HPC @ Exscale

2nd October 2024

Inverse problems

General model

$$Y \sim P(A(x)) \xrightarrow{\text{linear case}} y = Ax + n \tag{1}$$

- $Y = y \in \mathbb{R}^m$: Observations/Measurements.
- $x \in \mathcal{X} \subset \mathbb{R}^n$: Signal/image to reconstruct from a given signal set $\mathcal{X}$.
- A : Forward model including the deterministic physical part.
- P : Probabilistic model encompassing stochastic aspects of the observation y, e.g. noise n.
- **Objective:** estimate x from y given the model in Eq (1).
  - An ill-posed problem.

Inverse problem examples

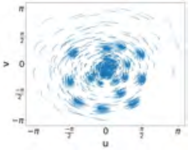
General model

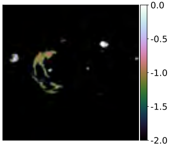
$$Y \sim P(A(x)) \xrightarrow{\text{linear case}} y = Ax + n$$

Radio interferometric imaging

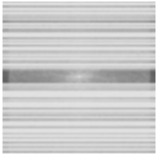
Magnetic Resonance Imaging

y → $\hat{x}$





y → $\hat{x}$



Other inverse problems: cosmological mass-mapping, PSF modelling, computed tomography imaging, deblurring, super-resolution, denoising, among others.

Bayesian inference

Bayes' theorem

$$\underbrace{p(x \mid y, M)}_{\text{posterior}} = \frac{\overbrace{p(y \mid x, M)}^{\text{likelihood}} \overbrace{p(x \mid M)}^{\textit{prior}}}{\underbrace{p(y \mid M)}_{\text{evidence}}} = \frac{\overbrace{\mathcal{L}(x)}^{\text{likelihood}} \overbrace{\pi(x)}^{\text{prior}}}{\underbrace{\mathcal{Z}}_{\textit{evidence}}}$$

for a model M, observation y and signal x.

We often only require the unnormalised probability (disregarding $\mathcal{Z}$) to compute a point estimator or samples from the posterior distribution,

$$\overbrace{p(x \mid y, M)}^{\text{posterior}} \propto \overbrace{p(y \mid x, M)}^{\text{likelihood}} \overbrace{p(x \mid M)}^{\textit{prior}}$$

- We rely on *Markov Chain Monte Carlo* (MCMC) to estimate posterior samples,

Point estimates and priors

We select a point estimate to use as reconstruction, for example:

- MMSE estimation: $\hat{x}_{M,\text{MMSE}} = \mathbb{E}[\, x \mid y, M \,]$ (posterior mean)
- *maximum-a-posteriori* (MAP) estimation: $\hat{x}_{M,\text{MAP}} = \arg\max_{x \in \mathbb{R}^n} \, p(x \mid y, M)$ (posterior mode)

Then,

1. The likelihood is based on the physics of the inverse problem.
2. We choose the prior based on our previous knowledge of $\mathcal{X}$.
3. We usually characterise the high-dimensional posterior through posterior samples.

Point estimates and priors

We select a point estimate to use as reconstruction, for example:

- MMSE estimation: $\hat{x}_{M,\text{MMSE}} = \mathbb{E}[\, x \mid y, M \,]$ (posterior mean)
- *maximum-a-posteriori* (MAP) estimation: $\hat{x}_{M,\text{MAP}} = \arg\max_{x \in \mathbb{R}^n} \; p(x \mid y, M)$ (posterior mode)

Then,

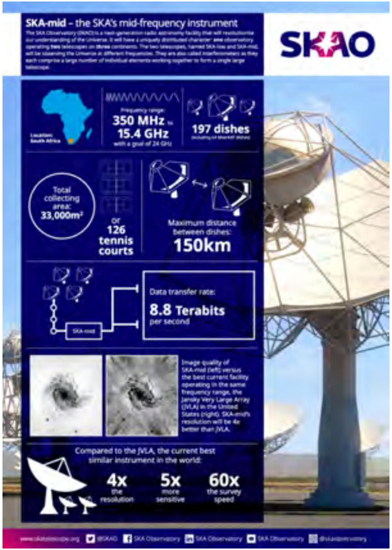
1. The likelihood is based on the physics of the inverse problem.
2. We choose the prior based on our previous knowledge of $\mathcal{X}$.
3. We usually characterise the high-dimensional posterior through posterior samples.

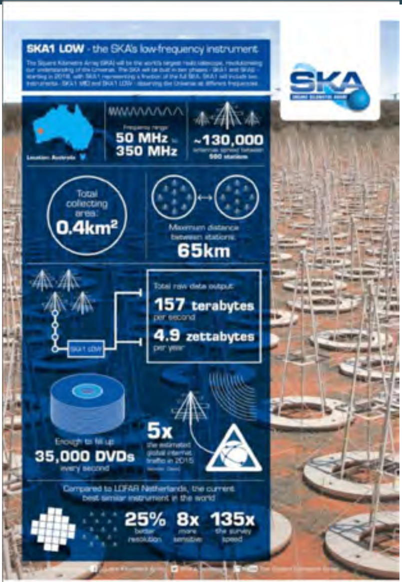
Outline

1. **Fast and scalable UQ for radio interferometric imaging: Towards SKA**
    - QuantifAI: Bayesian model with convex data-driven priors
    - EVIL-Deconv: Fast reconstruction through algorithm unrolling
    - CARB: unsupervised UQ for fast unrolled models
    - Approximate posterior sampling with rcGANs

2. **What data-driven prior should I use for my problem?**
    - Nested sampling for high-dimensional imaging problems

Motivation: SKA's radio interferometer





Radio interferometric imaging

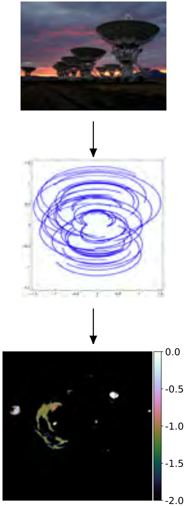
Linear observational model

$$\mathbf{y} = \mathbf{\Phi}\mathbf{x} + \mathbf{n}$$

- $\mathbf{y} \in \mathbb{C}^M$: Observed Fourier coefficients
- $\mathbf{n} \in \mathbb{C}^M$: Observational noise (assumed White and Gaussian)
- $\mathbf{x} \in \mathbb{R}^N$: Sky intensity image
- $\mathbf{\Phi} \in \mathbb{C}^{M \times N}$: Linear measurement operator
  - In its simplest case: FFT and Fourier mask

Due to $\mathbf{n}$ and $\mathbf{\Phi}$ the inverse problem is ill-posed

Goal: Estimate $\hat{\mathbf{x}}$ from $\mathbf{y}$

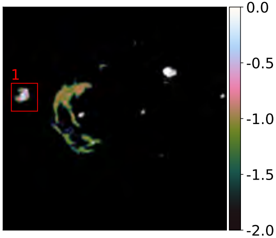


$\hat{\mathbf{x}}$

Uncertainty quantification: more than a point estimate

Based on: **Scalable Bayesian uncertainty quantification with data-driven priors for radio interferometric imaging** (Liaudat, *et al.*, 2024)

Image reconstruction: $\hat{\mathbf{x}}$



Is this blob *physical*?

→ Is it a reconstruction artefact?

→ Is it backed by the data?

→ Can we base a scientific decision on this image?

Several reasons motiaves us to develop uncertainty quantification (UQ) techniques for the reconstruction methods

- Usual UQ techniques from the Bayesian framework rely on interrogating the posterior exploiting Bayes' theorem.

For example, Cai et al. (2018a) applies this for radio imaging using a ℓ_1 regularised wavelet-based prior.

Sample from the posterior which is non-smooth to obtain $\{\mathbf{x}^{(j)}\}_{j=1}^{K}$, $\mathbf{x}^{(j)} \sim p(\mathbf{x}|\mathbf{y})$

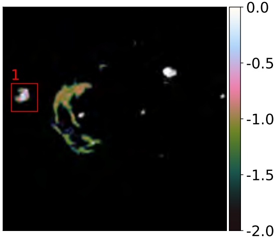
→ **Proximal MCMC algorithm** (Pereyra, 2016) following Langevin dynamics

Is the problem solved?

Uncertainty quantification: more than a point estimate

Based on: **Scalable Bayesian uncertainty quantification with data-driven priors for radio interferometric imaging** (Liaudat, *et al.*, 2024)

Image reconstruction: $\hat{\mathbf{x}}$



Is this blob *physical*?

→ Is it a reconstruction artefact?

→ Is it backed by the data?

→ Can we base a scientific decision on this image?

Several reasons motiaves us to develop uncertainty quantification (UQ) techniques for the reconstruction methods

- Usual UQ techniques from the Bayesian framework rely on interrogating the posterior exploiting Bayes' theorem.

For example, Cai et al. (2018a) applies this for radio imaging using a ℓ_1 regularised wavelet-based prior.

Sample from the posterior which is non-smooth to obtain $\{\mathbf{x}^{(j)}\}_{j=1}^{K}$, $\mathbf{x}^{(j)} \sim p(\mathbf{x}|\mathbf{y})$

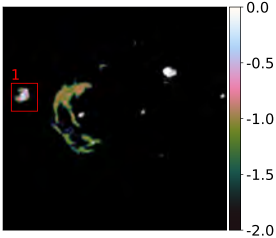
→ **Proximal MCMC algorithm** (Pereyra, 2016) following Langevin dynamics

Is the problem solved?

Uncertainty quantification: more than a point estimate

Based on: **Scalable Bayesian uncertainty quantification with data-driven priors for radio interferometric imaging** (Liaudat, *et al.*, 2024)

Image reconstruction: $\hat{\mathbf{x}}$



Is this blob *physical*?

→ Is it a reconstruction artefact?

→ Is it backed by the data?

→ Can we base a scientific decision on this image?

Several reasons motiaves us to develop uncertainty quantification (UQ) techniques for the reconstruction methods

- Usual UQ techniques from the Bayesian framework rely on interrogating the posterior exploiting Bayes' theorem.

For example, Cai et al. (2018a) applies this for radio imaging using a ℓ_1 regularised wavelet-based prior.

Sample from the posterior which is non-smooth to obtain $\{\mathbf{x}^{(j)}\}_{j=1}^{K}$, $\mathbf{x}^{(j)} \sim p(\mathbf{x}|\mathbf{y})$

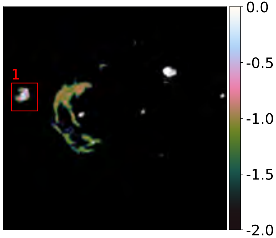
→ **Proximal MCMC algorithm** (Pereyra, 2016) following Langevin dynamics

Is the problem solved?

Uncertainty quantification: more than a point estimate

Based on: **Scalable Bayesian uncertainty quantification with data-driven priors for radio interferometric imaging** (Liaudat, *et al.*, 2024)

Image reconstruction: $\hat{\mathbf{x}}$



Is this blob *physical*?

→ Is it a reconstruction artefact?

→ Is it backed by the data?

→ Can we base a scientific decision on this image?

Several reasons motiaves us to develop uncertainty quantification (UQ) techniques for the reconstruction methods

- Usual UQ techniques from the Bayesian framework rely on interrogating the posterior exploiting Bayes' theorem.

For example, Cai et al. (2018a) applies this for radio imaging using a ℓ_1 regularised wavelet-based prior.

Sample from the posterior which is non-smooth to obtain $\{\mathbf{x}^{(j)}\}_{j=1}^{K}$, $\mathbf{x}^{(j)} \sim p(\mathbf{x}|\mathbf{y})$

→ **Proximal MCMC algorithm** (Pereyra, 2016) following Langevin dynamics

Is the problem solved?

The problem is not solved

Difficulties in the high-dimensional setting:

1. Even if we know the likelihood, applying $\mathbf{\Phi}$ is **computationally expensive**
2. Handcrafted priors like wavelets are **not expressive enough**
3. Sampling-based techniques are **prohibitively expensive** in this setting

How can we obtain information from the high-dimensional posterior $p(\mathbf{x}|\mathbf{y})$ without sampling from it?

If we restrict to log-concave posteriors something beautiful happens!
→ **A concentration phenomenom** (Pereyra, 2017)

log-concave posterior $p(\mathbf{x}|\mathbf{y}) = \exp[-f(\mathbf{x}) - g(\mathbf{x})]/Z \rightarrow$ convex potential $f(\mathbf{x}) + g(\mathbf{x})$

The problem is not solved

Difficulties in the high-dimensional setting:

1. Even if we know the likelihood, applying $\mathbf{\Phi}$ is **computationally expensive**
2. Handcrafted priors like wavelets are **not expressive enough**
3. Sampling-based techniques are **prohibitively expensive** in this setting

How can we obtain information from the high-dimensional posterior $p(\mathbf{x}|\mathbf{y})$ without sampling from it?

If we restrict to log-concave posteriors something beautiful happens!
$\rightarrow$ **A concentration phenomenom** (Pereyra, 2017)

log-concave posterior $p(\mathbf{x}|\mathbf{y}) = \exp[-f(\mathbf{x}) - g(\mathbf{x})]/Z \rightarrow$ convex potential $f(\mathbf{x}) + g(\mathbf{x})$

The problem is not solved

Difficulties in the high-dimensional setting:

1. Even if we know the likelihood, applying $\boldsymbol{\Phi}$ is **computationally expensive**
2. Handcrafted priors like wavelets are **not expressive enough**
3. Sampling-based techniques are **prohibitively expensive** in this setting

How can we obtain information from the high-dimensional posterior $p(\mathbf{x}|\mathbf{y})$ without sampling from it?

If we restrict to log-concave posteriors something beautiful happens!
$\rightarrow$ **A concentration phenomenom** (Pereyra, 2017)

log-concave posterior $p(\mathbf{x}|\mathbf{y}) = \exp[-f(\mathbf{x}) - g(\mathbf{x})]/Z \rightarrow$ convex potential $f(\mathbf{x}) + g(\mathbf{x})$

Highest posterior density region

Posterior credible region:

$$p(\mathbf{x} \in C_\alpha | \mathbf{y}) = \int_{\mathbf{x} \in \mathbb{R}^N} p(\mathbf{x}|\mathbf{y}) \mathbb{1}_{C_\alpha} \mathrm{d}\mathbf{x} = 1 - \alpha,$$

We consider the highest posterior density (HPD) region

$$C^*_\alpha = \{\mathbf{x} : \underbrace{f(\mathbf{x}) + g(\mathbf{x})}_{\text{potential}} \leq \gamma_\alpha\}, \quad \text{with } \gamma_\alpha \in \mathbb{R}, \quad \text{and } p(\mathbf{x} \in C^*_\alpha | \mathbf{y}) = 1 - \alpha \text{ holds},$$

Theorem 3.1 (Pereyra, 2017)

Suppose the posterior $p(\mathbf{x}|\mathbf{y}) = \exp[-f(\mathbf{x}) - g(\mathbf{x})]/Z$ is log-concave on $\mathbb{R}^N$. Then, for any $\alpha \in (4\exp[(-N/3)], 1)$, the HPD region C^*_α is contained by

$$\hat{C}_\alpha = \left\{\mathbf{x} : f(\mathbf{x}) + g(\mathbf{x}) \leq \hat{\gamma}_\alpha = f(\hat{\mathbf{x}}_{\text{MAP}}) + g(\hat{\mathbf{x}}_{\text{MAP}}) + \sqrt{N}\tau_\alpha + N\right\},$$

with a positive constant $\tau_\alpha = \sqrt{16 \log(3/\alpha)}$ independent of $p(\mathbf{x}|\mathbf{y})$.

We only need to evaluate $f + g$ on the MAP estimation $\hat{\mathbf{x}}_{\text{MAP}}$!

Highest posterior density region

Posterior credible region:

$$p(\mathbf{x} \in C_\alpha | \mathbf{y}) = \int_{\mathbf{x} \in \mathbb{R}^N} p(\mathbf{x}|\mathbf{y}) \mathbb{1}_{C_\alpha} \mathrm{d}\mathbf{x} = 1 - \alpha,$$

We consider the highest posterior density (HPD) region

$$C^*_\alpha = \big\{ \mathbf{x} : \underbrace{f(\mathbf{x}) + g(\mathbf{x})}_{\text{potential}} \leq \gamma_\alpha \big\}, \quad \text{with } \gamma_\alpha \in \mathbb{R}, \quad \text{and } p(\mathbf{x} \in C^*_\alpha | \mathbf{y}) = 1 - \alpha \text{ holds},$$

Theorem 3.1 (Pereyra, 2017)

Suppose the posterior $p(\mathbf{x}|\mathbf{y}) = \exp[-f(\mathbf{x}) - g(\mathbf{x})]/Z$ is log-concave on $\mathbb{R}^N$. Then, for any $\alpha \in (4\exp[(-N/3)], 1)$, the HPD region C^*_α is contained by

$$\hat{C}_\alpha = \Big\{ \mathbf{x} : f(\mathbf{x}) + g(\mathbf{x}) \leq \hat{\gamma}_\alpha = f(\hat{\mathbf{x}}_{\text{MAP}}) + g(\hat{\mathbf{x}}_{\text{MAP}}) + \sqrt{N}\tau_\alpha + N \Big\},$$

with a positive constant $\tau_\alpha = \sqrt{16 \log(3/\alpha)}$ independent of $p(\mathbf{x}|\mathbf{y})$.

We only need to evaluate $f + g$ on the MAP estimation $\hat{\mathbf{x}}_{\text{MAP}}$!

Highest posterior density region

Posterior credible region:

$$p(\mathbf{x} \in C_\alpha | \mathbf{y}) = \int_{\mathbf{x} \in \mathbb{R}^N} p(\mathbf{x}|\mathbf{y}) \mathbb{1}_{C_\alpha} \mathrm{d}\mathbf{x} = 1 - \alpha,$$

We consider the highest posterior density (HPD) region

$$C^*_\alpha = \{\mathbf{x} : \underbrace{f(\mathbf{x}) + g(\mathbf{x})}_{\text{potential}} \leq \gamma_\alpha\}, \quad \text{with } \gamma_\alpha \in \mathbb{R}, \quad \text{and } p(\mathbf{x} \in C^*_\alpha|\mathbf{y}) = 1 - \alpha \text{ holds},$$

Theorem 3.1 (Pereyra, 2017)

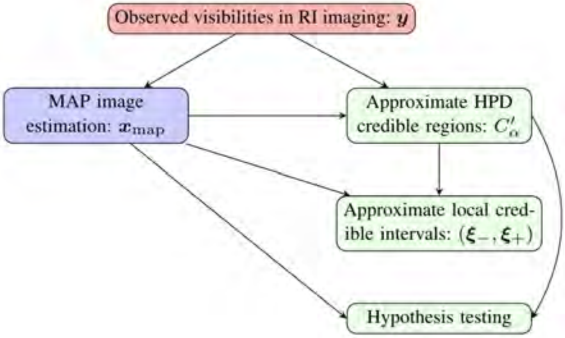
Suppose the posterior $p(\mathbf{x}|\mathbf{y}) = \exp[-f(\mathbf{x}) - g(\mathbf{x})]/Z$ is log-concave on $\mathbb{R}^N$. Then, for any $\alpha \in (4\exp[(-N/3)], 1)$, the HPD region C^*_α is contained by

$$\hat{C}_\alpha = \left\{\mathbf{x} : f(\mathbf{x}) + g(\mathbf{x}) \leq \hat{\gamma}_\alpha = f(\hat{\mathbf{x}}_{\text{MAP}}) + g(\hat{\mathbf{x}}_{\text{MAP}}) + \sqrt{N}\tau_\alpha + N\right\},$$

with a positive constant $\tau_\alpha = \sqrt{16\log(3/\alpha)}$ independent of $p(\mathbf{x}|\mathbf{y})$.

We only need to evaluate $f + g$ on the MAP estimation $\hat{\mathbf{x}}_{\text{MAP}}$!

MAP-based uncertainty quantification



Cai et al. (2018b)

UQ techinques:

- Hypothesis test with significance α
  - e.g. with respect to a surrogate image with an inpainted structure.
- Local credible intervals (LCI)
  - Test the approx HPD region for each pixel or super-pixel in the image.
- Fast pixel-wise errors at different scales
  - Test the approx HPD region from the coefficients of a multi-resolution decomposition of the image.

Scalable Bayesian uncertainty quantification

1. **Scalability** $\rightarrow$ Need to rely on optimisation ~~sampling~~, use the MAP estimator
2. **Uncertainty quantification** $\rightarrow$ Need the potential to be convex and explicit
3. **Good reconstruction** $\rightarrow$ Need to use data-driven (learned) approaches

The approach requires our prior to be convex and with an explicit potential

We constrain our prior to be convex, but we gain an effortless UQ!

Scalable Bayesian uncertainty quantification

1. **Scalability** $\rightarrow$ Need to rely on optimisation ~~sampling~~, use the MAP estimator
2. **Uncertainty quantification** $\rightarrow$ Need the potential to be convex and explicit
3. **Good reconstruction** $\rightarrow$ Need to use data-driven (learned) approaches

The approach requires our prior to be convex and with an explicit potential

We constrain our prior to be convex, but we gain an effortless UQ!

Learned convex regulariser

We use the neural-network-based convex regulariser R from Goujon et al. (2023), where

$$R : \mathbb{R}^N \mapsto \mathbb{R}, \quad R(\mathbf{x}) = \sum_{n=1}^{N_C} \sum_{k} \psi_n \left((\mathbf{h}_n * \mathbf{x}) [k] \right),$$

- ψ_n are learned convex profile functions with Lipschitz continuous derivate
    - Learnable 2nd degree splines
- There are N_C learned convolutional filters $\mathbf{h}_n$
- R is trained as a (multi-)gradient step denoiser

Properties:

1. **Explicit cost**
2. **Convex**
3. **Smooth regulariser with known Lipschitz constant**

Learned convex regulariser

We use the neural-network-based convex regulariser R from Goujon et al. (2023), where

$$R : \mathbb{R}^N \mapsto \mathbb{R}, \quad R(\mathbf{x}) = \sum_{n=1}^{N_C} \sum_{k} \psi_n \left((\mathbf{h}_n * \mathbf{x}) [k] \right),$$

- ψ_n are learned convex profile functions with Lipschitz continuous derivate
  - Learnable 2nd degree splines
- There are N_C learned convolutional filters $\mathbf{h}_n$
- R is trained as a (multi-)gradient step denoiser

Properties:

1. **Explicit cost**
2. **Convex**
3. **Smooth regulariser with known Lipschitz constant**

Numerical experiments

RI imaging models:

Model from Cai et al. (2018a): $$\hat{\mathbf{x}}_{\text{MAP}} = \arg\min_{\mathbf{x}\in\mathbb{R}^N} \|\mathbf{y} - \mathbf{\Phi}\mathbf{x}\|_2^2/2\sigma^2 + \lambda_1 \|\mathbf{\Psi}^\dagger \mathbf{x}\|_1 + \iota_{\mathbb{R}^N}(\mathbf{x}),$$

Proposed model: $$\hat{\mathbf{x}}_{\text{MAP}} = \arg\min_{\mathbf{x}\in\mathbb{R}^N} \|\mathbf{y} - \mathbf{\Phi}\mathbf{x}\|_2^2/2\sigma^2 + \lambda_2/\mu R_{\boldsymbol{\theta}}(\mu\mathbf{x}) + \iota_{\mathbb{R}^N}(\mathbf{x}),$$

MAP estimations are computed using the FISTA algorithm

Validation of the UQ is done by sampling both posterior distributions using a proximal MCMC algorithm, SK-ROCK (Pereyra et al., 2020)

Experiment settings:

- Image size 256×256
- Input SNR of 30dB
- Gridded Fourier sampling: 10% coverage from a Gaussian distribution ($M \approx 6.5 \times 10^3$)
- Wavelets used: Daubechies 8

Numerical experiments

RI imaging models:

Model from Cai et al. (2018a): $\hat{\mathbf{x}}_{\mathrm{MAP}} = \arg\min_{\mathbf{x}\in\mathbb{R}^N} \|\mathbf{y} - \mathbf{\Phi}\mathbf{x}\|_2^2/2\sigma^2 + \lambda_1\|\mathbf{\Psi}^\dagger\mathbf{x}\|_1 + \iota_{\mathbb{R}^N}(\mathbf{x})\,,$

Proposed model: $\hat{\mathbf{x}}_{\mathrm{MAP}} = \arg\min_{\mathbf{x}\in\mathbb{R}^N} \|\mathbf{y} - \mathbf{\Phi}\mathbf{x}\|_2^2/2\sigma^2 + \lambda_2/\mu R_{\boldsymbol{\theta}}(\mu\mathbf{x}) + \iota_{\mathbb{R}^N}(\mathbf{x})\,,$

MAP estimations are computed using the FISTA algorithm

Validation of the UQ is done by sampling both posterior distributions using a proximal MCMC algorithm, SK-ROCK (Pereyra et al., 2020)

Experiment settings:

- Image size 256×256
- Input SNR of 30dB
- Gridded Fourier sampling: 10% coverage from a Gaussian distribution ($M \approx 6.5 \times 10^3$)
- Wavelets used: Daubechies 8

Numerical experiments

RI imaging models:

Model from Cai et al. (2018a): $\hat{\mathbf{x}}_{\text{MAP}} = \underset{\mathbf{x} \in \mathbb{R}^N}{\arg\min} \|\mathbf{y} - \boldsymbol{\Phi}\mathbf{x}\|_2^2 / 2\sigma^2 + \lambda_1 \|\boldsymbol{\Psi}^\dagger \mathbf{x}\|_1 + \iota_{\mathbb{R}^N}(\mathbf{x})$,

Proposed model: $\hat{\mathbf{x}}_{\text{MAP}} = \underset{\mathbf{x} \in \mathbb{R}^N}{\arg\min} \|\mathbf{y} - \boldsymbol{\Phi}\mathbf{x}\|_2^2 / 2\sigma^2 + \lambda_2 / \mu R_{\boldsymbol{\theta}}(\mu \mathbf{x}) + \iota_{\mathbb{R}^N}(\mathbf{x})$,

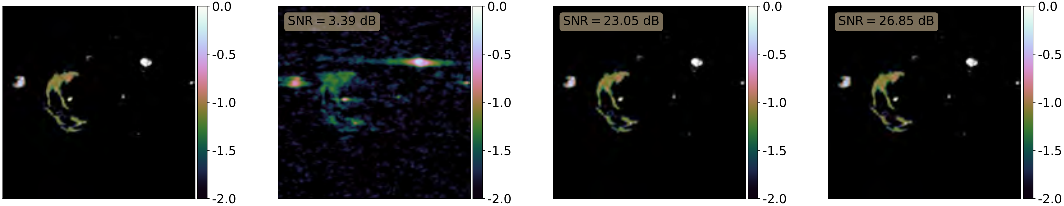
MAP estimations are computed using the FISTA algorithm

Validation of the UQ is done by sampling both posterior distributions using a proximal MCMC algorithm, SK-ROCK (Pereyra et al., 2020)

Experiment settings:

- Image size 256×256
- Input SNR of 30dB
- Gridded Fourier sampling: 10% coverage from a Gaussian distribution ($M \approx 6.5 \times 10^3$)
- Wavelets used: Daubechies 8

MAP reconstructions

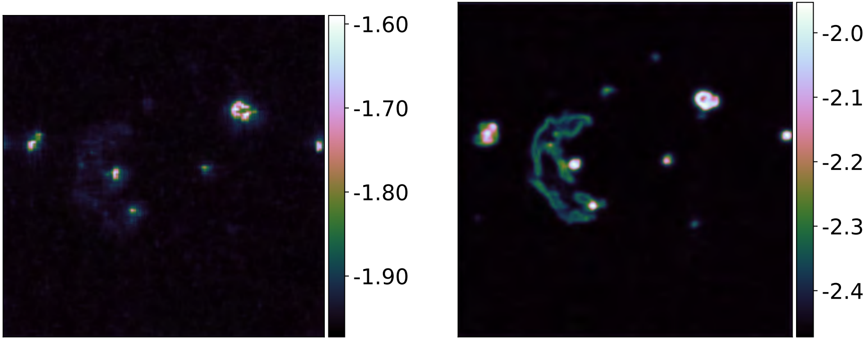


Ground truth	Dirty image	Wavelet prior	Learned prior
	SNR = 3.39 dB	SNR = 23.05 dB	**SNR = 26.85 dB**

Improved the reconstruction by 3.8 dB

Posterior standard deviation

Computed using 10^4 samples obtained from the sampling algorithm SK-ROCK (Pereyra et al., 2020)



Wavelet | Learned regulariser

More meaningful uncertainties in the posterior Std Dev

The learned convex regulariser was trained on natural images, not RI images

Computing time and likelihood evaluations

Computation wall-clock times for the W28 image in seconds.

Models	MAP optim.	Posterior sampling	LCIs 8×8	Fast pixel UQ
Wavelet-based	0.94	36.0×10^3	149.7	—
QuantifAI	0.64	6.44×10^3	108.2	**0.17**

The number of measurement operator evaluations used by QuantifAI for the W28 image.

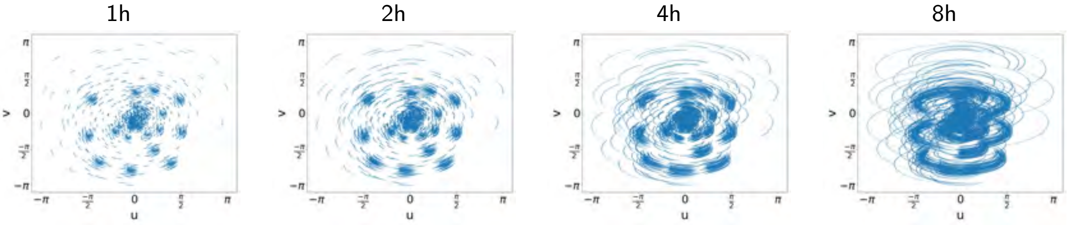
MCMC sampling	LCIs 8×8	LCIs 16×16	Fast pixel UQ
11×10^6	81.5×10^3	21.2×10^3	**28**

The fast pixel UQ is 10^6 and 10^3 times faster than the MCMC sampling and LCIs, respectively.

A more realistic experiment

Simulate single frequency MeerKAT ungridded visibility patterns

- Start frequency of 1400MHz with a channel width of 10MHz
- Pointing: J2000, RA=13h18m54.86s, DEC=-15d36m04.25s

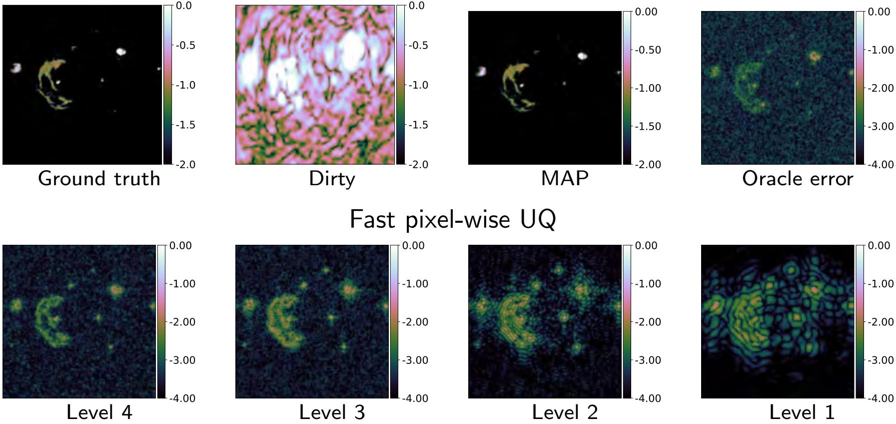


$M \approx 3 \times 10^4$ $M \approx 6 \times 10^4$ $M \approx 1.2 \times 10^5$ $M \approx 2.4 \times 10^5$

We use forward operator based on a torch-based 2D NUFFT with Kaisser-Bessel gridding.

A more realistic experiment

Results for **8h of observation time** ($M \approx 2.4 \times 10^5$). MAP reconstruction SNR: **28.56dB**



Computation wall-clock time: MAP estimation $\rightarrow$ 137.0s, **fast pixel UQ $\rightarrow$ 1.84s**

Conclusions

- **Scalable uncertainty quantification**
  - We exploit a concentration phenomenon of log-concave posteriors
  - Focus on hypothesis test and pixel-wise errors at different scales
- Only rely on **optimisation** to compute the MAP and avoid sampling
- We used **learned convex regularisers**
  - Decreased reconstruction errors and improved quality of the posterior Std Dev

Conclusions

Ongoing work with colleagues at UCL (UK) to

- Interface QuantifAI with PURIFY (realistic RI forward operator) and SOPT (performant and parallel convex optimisation routines),
- Implement & benchmark QuantifAI on a massively parallelised computing env.

Codes:

- QuantifAI github.com/astro-informatics/quantifai
- PURIFY github.com/astro-informatics/purify
- SOPT github.com/astro-informatics/sopt

We rely on an iterative algorithm for the optimisation, which can be **computationally expensive**

- We need to decrease the number of iteration to further accelerate

Conclusions

Ongoing work with colleagues at UCL (UK) to

- Interface QuantifAI with PURIFY (realistic RI forward operator) and SOPT (performant and parallel convex optimisation routines),
- Implement & benchmark QuantifAI on a massively parallelised computing env.

Codes:

- QuantifAI github.com/astro-informatics/quantifai
- PURIFY github.com/astro-informatics/purify
- SOPT github.com/astro-informatics/sopt

We rely on an iterative algorithm for the optimisation, which can be **computationally expensive**

- We need to decrease the number of iteration to further accelerate

Outline

1. **Fast and scalable UQ for radio interferometric imaging: Towards SKA**
    - QuantifAI: Bayesian model with convex data-driven priors
    - EVIL-Deconv: Fast reconstruction through algorithm unrolling
    - CARB: unsupervised UQ for fast unrolled models
    - Approximate posterior sampling with rcGANs

2. **What data-driven prior should I use for my problem?**
    - Nested sampling for high-dimensional imaging problems

Faster reconstruction: algorithm unrolling

Based on: **EVIL-Deconv: Efficient Variability-Informed Learned Deconvolution using Algorithm Unrolling** (Kern, Kervazo & Bobin, 2024 (submitted))

Main motivation:

1. **Reduce the number of iterations!**
2. Improve reconstruction performance

The main algorithm step which is unrolled for L steps

$$x_{l+1} = g_l(x_l + \Phi_l(M)(y - M * x_l))$$

- $\Phi_l(M)$: Learned preconditioning step based on CNNs with M being the PSF
- g_l : Learned proximal operator (denoiser) based on DRUNets

Everything trained on a supervised manner end-to-end for th L unrolled steps.

Faster reconstruction: algorithm unrolling

EVIL-Deconv results:

- Greatly reduced computation budget
- Great reconstruction quality (for in-distribution data)

EVIL-Deconv drawbacks:

- Lost interpretation of the reconstruction
  - Is it the fixed point of an equation?
  - Is the reconstruction related to a posterior probability distribution?
- UQ is missing

These drawbacks limit its scientific application

Faster reconstruction: algorithm unrolling

EVIL-Deconv results:

- Greatly reduced computation budget
- Great reconstruction quality (for in-distribution data)

EVIL-Deconv drawbacks:

- Lost interpretation of the reconstruction
  - Is it the fixed point of an equation?
  - Is the reconstruction related to a posterior probability distribution?
- UQ is missing

These drawbacks limit its scientific application

CARB: Conformalized Augmented Radio Bootstrap

Based on: **Uncertainty quantification for fast reconstruction methods using augmented equivariant bootstrap: Application to radio interferometry**
(Cherif, Liaudat, Kern, Kervazo & Bobin, 2024 (submitted))

Based on the equivariant Bootstrap framework of Tachella & Pereyra (2024)

Given an observation model $y = Ax + n$ (e.g. RI imaging), group actions $\{T_g\}_{g \in \mathcal{G}}$ such that $T_g x \in \mathcal{X}$ and a reconstruction method $\hat{x}(y) = f(y)$ (e.g. EVIL-Deconv):

For $i = 1, \ldots, N$:

1. Draw transform g_i from $\mathcal{G}$ and sample noise $n_i \sim \mathcal{N}(0, \sigma^2 I)$
2. Build bootsrap measurement $\tilde{y}_i = AT_{g_i}\hat{x}(y) + n_i = A_{g_i}\hat{x}(y) + n_i$
3. Reconstruct $\tilde{x}_i = T_{g_i}^{-1}\hat{x}(\tilde{y}_i)$
4. Collect error estimate $e_i = \|\hat{x}(y) - \tilde{x}_i\|^2$

CARB: Conformalized Augmented Radio Bootstrap

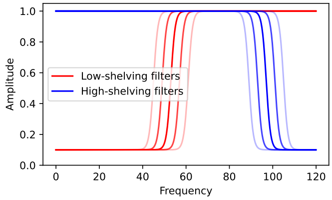
Motivation:

- Unsupervised method → **No ground truth required**
- Independent of the reconstruction method and **each sample can run in parallel**
- Well-suited to ultra-fast reconstruction methods, e.g. unrolled algorithms
- **Carefully selected group transforms allow us to explore the big nullspace of the RI imaging forward operator and better characterise the errors**

CARB method consists of:

1. Fast reconstruction algorithm (EVIL-Deconv)
2. Equivariant bootstrap framework
3. Adapted group actions for the RI imaging problem
4. Conformalisation procedure to guarantee coverage from Angelopoulos and Bates, 2023

Examples of filter transformations:

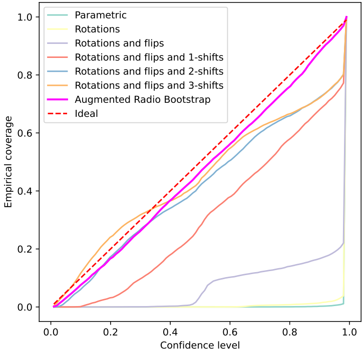


CARB: Conformalized Augmented Radio Bootstrap

Uncertainty quantification performance comparison (90% confidence interval)

Method	Length	Coverage
Quantile Regression (QR)	0.15	14%
Conformalized QR	204.08	92%
Parametric Bootstrap	0.07	0%
Equivariant Bootstrap	0.13	7%
Augmented Radio Bootstrap	0.29	87%
CARB	0.34	91%

Coverage plots for equivariant bootstrap methods with different group actions.



Tight intervals and very good coverage! Results showcase:

- the importance of selecting adapted group actions,
- the conformalisation is useful once the intervals are already good.

We still need to validate the method on higher dimensions.

Bonus: Can we go even faster?

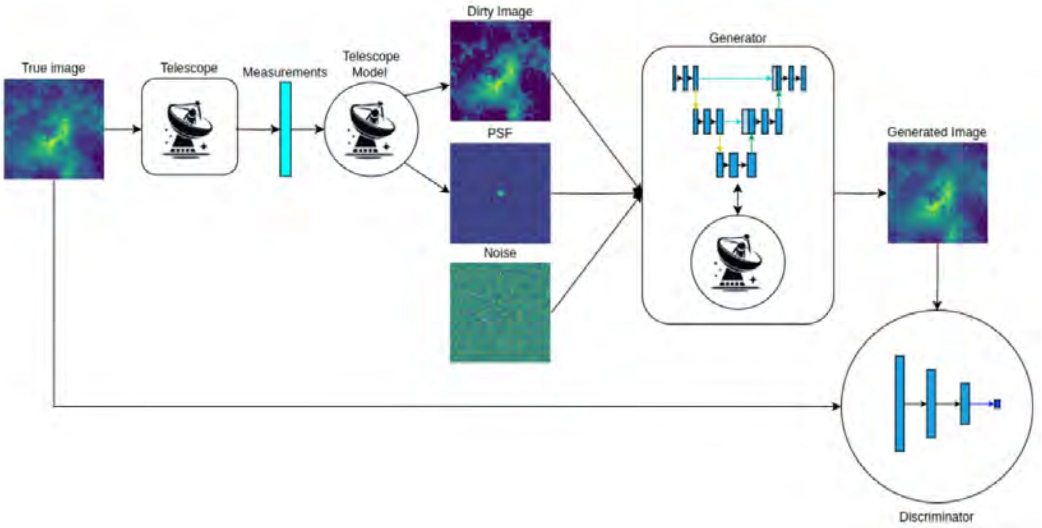
Based on: **Generative imaging for radio interferometry with fast UQ** (Mars, Liaudat, Whitney, Betcke & McEwen, 2024 (in prep.))

Based on the regularised conditional GAN (rcGAN) proposed in Bendel et al., 2023 that is able to generate approximate posterior samples

Main points of the proposed approach:

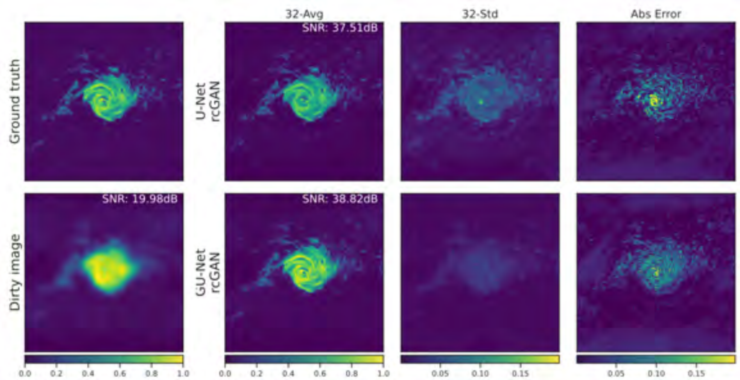
- Builds from the Wasserstein conditional GAN (Adler and Öktem, 2018)
- Regularisation to avoid mode collapse and reward sample diversity.
- Under simplifying assumptions, the first two moments of the approximated posterior (mean and covariance) match the true posterior.
- We condition on the dirty image and the PSF.
- Extremely-fast reconstruction and sampling.

Regularised conditional GAN for RI imaging and fast UQ



Regularised conditional GAN for RI imaging and fast UQ

Reconstruction of simulated MeerKAT observation of galaxies from Illustris TNG simulations.



High reconstruction PSNR and good correlation between the oracle error and the Std Dev.

A deeper validation of the produced samples is yet to be done.

Conclusions

Explored different reconstruction methods **with UQ** for radio interferometric imaging exploiting different ML/AI tools:

1. In a Bayesian framework, favour optimisation and avoid sampling by approximating the HPD region while using learned data-driven priors
2. Accelerate reconstruction with algorithm unrolling but lose interpretability
3. The CARP method picks up the unrolled method and provides UQ in an unsupervised framework based on equivariant bootstrap
4. The regularised conditional GAN trained on a supervised manner allows us to do instant (approximate) posterior sampling

Outline

1. **Fast and scalable UQ for radio interferometric imaging: Towards SKA**
    - QuantifAI: Bayesian model with convex data-driven priors
    - EVIL-Deconv: Fast reconstruction through algorithm unrolling
    - CARB: unsupervised UQ for fast unrolled models
    - Approximate posterior sampling with rcGANs

2. **What data-driven prior should I use for my problem?**
    - Nested sampling for high-dimensional imaging problems

Point estimates and priors

Which prior should we use?

- For some low-dimensional problems it can be simple to select the prior:
  - Physics informed priors (e.g. mass constrained to be positive)
  - Uninformative priors (e.g. invariance to certain symmetry)
  - Data-informed priors (e.g. old data as prior and likelihood on new data)

- For **high-dimensional** imaging problems it is **hard**:
  - For example, encode that $x \in \mathcal{X}$ with $\mathcal{X}$ being a large and complex set of images, i.e. galaxy images, natural images, or MRI brain scans.
  - How to describe such a set?

$\rightarrow$ Informative priors: e.g. sparse in a given wavelet dictinoary

$\rightarrow$ Data-driven priors: e.g. machine learning models, generative AI

On priors

More recently, data-driven priors encoded by deep neural networks (NN) have emerged like Plug-and-Play (based on deep learning-based denoisers), or diffusion models.

Train the NN on samples from the true $X \sim p(x)$ (i.e. dataset of examples of $x \in \mathcal{X}$)

In scientific settings,

- We do not have access to ground truth.
- Which NN prior to use for our scientific scenario? We base the choice on which metric?

Work based on:
Proximal nested sampling for high-dimensional Bayesian model selection
(Cai, McEwen, & Pereyra, 2022)
Proximal nested sampling with data-driven priors for physical scientists
(McEwen, Liaudat, Price, Cai & Pereyra, 2023)

Bayesian model selection

Using Bayes theorem for model M_j:

$$p(M_j \mid y) = \frac{p(y \mid M_j)p(M_j)}{\sum_j p(y \mid M_j)p(M_j)}.$$

For **model selection**, consider the posterior model odds :

$$\underbrace{\frac{p(M_1 \mid y)}{p(M_2 \mid y)}}_{\text{posterior odds}} = \underbrace{\frac{p(y \mid M_1)}{p(y \mid M_2)}}_{\text{Bayes factor}} \times \underbrace{\frac{p(M_1)}{p(M_2)}}_{\text{prior odds}}$$

To compute the bayes factor we need to comute the **Bayesian model evidence** or **marginal likelihood** given by

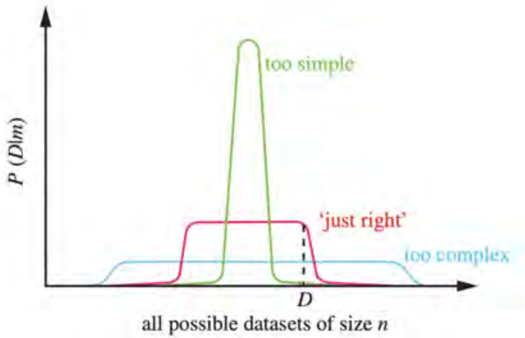
$$\mathcal{Z} = p(y \mid M) = \int \mathrm{d}x\, \mathcal{L}(x)\, \pi(x)$$

Extremely challenging computational problem in high-dimensions.

Occam’s razor

The Bayesian model evidence **naturally incorporates Occam’s razor**, trading off model complexity and goodness of fit.

- In Bayesian formalism models specified as probability distributions over datasets.
- Each model has limited “probability budget”.
- Complex models can represent a wide range of datasets well, but spreads predictive probability widely.
- In doing so, model evidence of complex models penalised if complexity not required.



Ghahramani (2013); MacKay (1991)

Nested sampling is ingenious approach to evaluate the evidence (Skilling 2006).

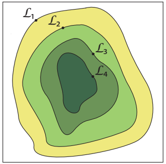
Consider $\Omega_{L^*} = \{x | \mathcal{L}(x) \geq L^*\}$, which groups the parameter space Ω into a series of **nested subspaces**.

Define the prior volume ξ within Ω_{L^*} by $\xi(L^*) = \int_{\Omega_{L^*}} \pi(x) \mathrm{d}x$.

The marginal likelihood integral can then be rewritten as

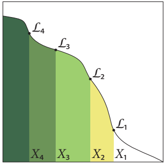
$$\mathcal{Z} = \int_0^1 \mathcal{L}(\xi) \mathrm{d}\xi,$$

which is a **one-dimensional integral** over the prior volume ξ.



Feroz *et al.* (2013)

Nested subspaces



Feroz *et al.* (2013)

Reparameterised likelihood

Nested sampling: constrained sampling

Require strategy to compute likelihood level-sets (iso-contours) L_i and corresponding prior volumes $0 < \xi_i \leq 1$.

Nested sampling (Skilling 2006)

1. Draw N_{live} *live* samples from prior, with prior volume $\xi_0 = 1$.
2. Remove sample with smallest likelihood, say L_i.
3. Replace removed sample with new **sample from the prior but constrained to a higher likelihood** than L_i.
4. Estimate (stochastically) prior volume ξ_i enclosed by likelihood level-set L_i.
5. Repeat 2–5.

Nested sampling: constrained sampling

Require strategy to compute likelihood level-sets (iso-contours) L_i and corresponding prior volumes $0 < \xi_i \leq 1$.

Nested sampling (Skilling 2006)

1. Draw N_{live} *live* samples from prior, with prior volume $\xi_0 = 1$.
2. Remove sample with smallest likelihood, say L_i.
3. Replace removed sample with new **sample from the prior but constrained to a higher likelihood** than L_i.
4. Estimate (stochastically) prior volume ξ_i enclosed by likelihood level-set L_i.
5. Repeat 2–5.

Nested sampling: constrained sampling

Require strategy to compute likelihood level-sets (iso-contours) L_i and corresponding prior volumes $0 < \xi_i \leq 1$.

Nested sampling (Skilling 2006)

1. Draw N_{live} *live* samples from prior, with prior volume $\xi_0 = 1$.
2. Remove sample with smallest likelihood, say L_i.
3. Replace removed sample with new **sample from the prior but constrained to a higher likelihood** than L_i.
4. Estimate (stochastically) prior volume ξ_i enclosed by likelihood level-set L_i.
5. Repeat 2–5.

Nested sampling: constrained sampling

Require strategy to compute likelihood level-sets (iso-contours) L_i and corresponding prior volumes $0 < \xi_i \leq 1$.

Nested sampling (Skilling 2006)

1. Draw N_{live} *live* samples from prior, with prior volume $\xi_0 = 1$.
2. Remove sample with smallest likelihood, say L_i.
3. Replace removed sample with new **sample from the prior but constrained to a higher likelihood** than L_i.
4. Estimate (stochastically) prior volume ξ_i enclosed by likelihood level-set L_i.
5. Repeat 2–5.

Nested sampling: constrained sampling

Require strategy to compute likelihood level-sets (iso-contours) L_i and corresponding prior volumes $0 < \xi_i \leq 1$.

Nested sampling (Skilling 2006)

1. Draw N_{live} *live* samples from prior, with prior volume $\xi_0 = 1$.
2. Remove sample with smallest likelihood, say L_i.
3. Replace removed sample with new **sample from the prior but constrained to a higher likelihood** than L_i.
4. Estimate (stochastically) prior volume ξ_i enclosed by likelihood level-set L_i.
5. Repeat 2–5.

Nested sampling: constrained sampling

Require strategy to compute likelihood level-sets (iso-contours) L_i and corresponding prior volumes $0 < \xi_i \leq 1$.

Nested sampling (Skilling 2006)

1. Draw N_{live} *live* samples from prior, with prior volume $\xi_0 = 1$.
2. Remove sample with smallest likelihood, say L_i.
3. Replace removed sample with new **sample from the prior but constrained to a higher likelihood** than L_i.
4. Estimate (stochastically) prior volume ξ_i enclosed by likelihood level-set L_i.
5. Repeat 2–5.

Proximal nested sampling

Main difficulty: Sampling from the prior, subject the likelihood iso-contour constraint.

Advantage: Apart from the evidence we can do posterior inferences by assigning importance weights.

Proximal nested sampling

- Constrained sampling formulation
- Langevin MCMC sampling
- Moreau-Yosida approximation of constraint (and any non-differentiable prior)
- Limited to classic non-smooth priors

Proximal nested sampling

Main difficulty: Sampling from the prior, subject the likelihood iso-contour constraint.

Advantage: Apart from the evidence we can do posterior inferences by assigning importance weights.

Proximal nested sampling

- Constrained sampling formulation
- Langevin MCMC sampling
- Moreau-Yosida approximation of constraint (and any non-differentiable prior)
- Limited to classic non-smooth priors

Proximal nested sampling Markov chain:

$$\boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k)} + \frac{\delta}{2}\nabla \log \pi(\boldsymbol{x}^{(k)}) - \frac{\delta}{2\lambda}\left[\boldsymbol{x}^{(k)} - \mathrm{prox}_{\chi_{\mathcal{B}_\tau}}(\boldsymbol{x}^{(k)})\right] + \sqrt{\delta}\,\boldsymbol{w}^{(k+1)}.$$

Proximal nested sampling intuition

Recall proximal nested sampling Markov chain (from previous slide):

$$\boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k)} + \frac{\delta}{2}\nabla \log \pi(\boldsymbol{x}^{(k)}) - \frac{\delta}{2\lambda} \left[\boldsymbol{x}^{(k)} - \mathrm{prox}_{\chi_{\mathcal{B}_\tau}}(\boldsymbol{x}^{(k)})\right] + \sqrt{\delta}\,\boldsymbol{w}^{(k+1)}.$$

1. **$\boldsymbol{x}^{(k)}$ is already in $\mathcal{B}_\tau$:** term $\left[\boldsymbol{x}^{(k)} - \mathrm{prox}^{\lambda}_{\chi_{\mathcal{B}_\tau}}(\boldsymbol{x}^{(k)})\right]$ disappears and recover usual Langevin MCMC.

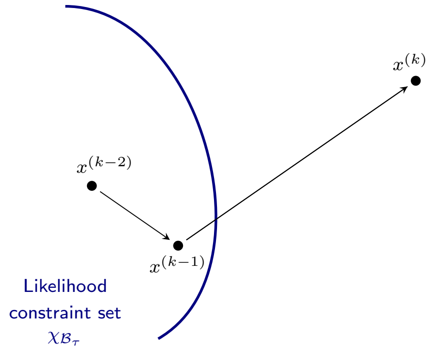
2. **$\boldsymbol{x}^{(k)}$ is not in $\mathcal{B}_\tau$:** a step is also taken in the direction $-\left[\boldsymbol{x}^{(k)} - \mathrm{prox}^{\lambda}_{\chi_{\mathcal{B}_\tau}}(\boldsymbol{x}^{(k)})\right]$, which moves the next iteration in the direction of the projection of $\boldsymbol{x}^{(k)}$ onto the convex set $\mathcal{B}_\tau$. Acts to push the

Proximal nested sampling intuition

Recall proximal nested sampling Markov chain (from previous slide):

$$\boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k)} + \frac{\delta}{2}\nabla \log \pi(\boldsymbol{x}^{(k)}) - \frac{\delta}{2\lambda}\left[\boldsymbol{x}^{(k)} - \text{prox}_{\chi_{\mathcal{B}_\tau}}(\boldsymbol{x}^{(k)})\right] + \sqrt{\delta}\boldsymbol{w}^{(k+1)}.$$

1. **$\boldsymbol{x}^{(k)}$ is already in $\mathcal{B}_\tau$:** term $\left[\boldsymbol{x}^{(k)} - \text{prox}^{\lambda}_{\chi_{\mathcal{B}_\tau}}(\boldsymbol{x}^{(k)})\right]$ disappears and recover usual Langevin MCMC.
2. **$\boldsymbol{x}^{(k)}$ is not in $\mathcal{B}_\tau$:** a step is also taken in the direction $-\left[\boldsymbol{x}^{(k)} - \text{prox}^{\lambda}_{\chi_{\mathcal{B}_\tau}}(\boldsymbol{x}^{(k)})\right]$, which moves the next iteration in the direction of the projection of $\boldsymbol{x}^{(k)}$ onto the convex set $\mathcal{B}_\tau$. Acts to push the Markov chain back into the constraint set $\mathcal{B}_\tau$ if it wanders outside of it.

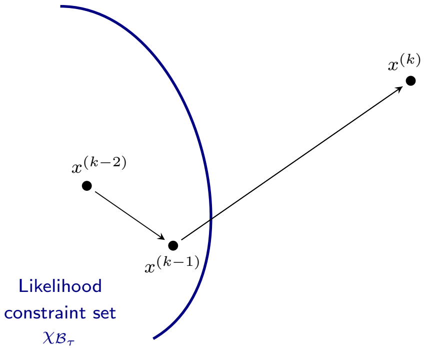


Proximal nested sampling intuition

Recall proximal nested sampling Markov chain (from previous slide):

$$\boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k)} + \frac{\delta}{2}\nabla \log \pi(\boldsymbol{x}^{(k)}) - \frac{\delta}{2\lambda}\left[\boldsymbol{x}^{(k)} - \mathrm{prox}_{\chi_{\mathcal{B}_\tau}}(\boldsymbol{x}^{(k)})\right] + \sqrt{\delta}\,\boldsymbol{w}^{(k+1)}.$$

1. $\boldsymbol{x}^{(k)}$ **is already in** $\mathcal{B}_\tau$: term $\left[\boldsymbol{x}^{(k)} - \mathrm{prox}^{\lambda}_{\chi_{\mathcal{B}_\tau}}(\boldsymbol{x}^{(k)})\right]$ disappears and recover usual Langevin MCMC.
2. $\boldsymbol{x}^{(k)}$ **is not in** $\mathcal{B}_\tau$: a step is also taken in the direction $-\left[\boldsymbol{x}^{(k)} - \mathrm{prox}^{\lambda}_{\chi_{\mathcal{B}_\tau}}(\boldsymbol{x}^{(k)})\right]$, which moves the next iteration in the direction of the projection of $\boldsymbol{x}^{(k)}$ onto the convex set $\mathcal{B}_\tau$. Acts to push the Markov chain back into the constraint set $\mathcal{B}_\tau$ if it wanders outside of it.



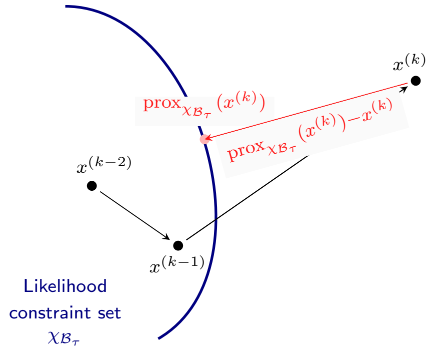
Proximal nested sampling intuition

Recall proximal nested sampling Markov chain (from previous slide):

$$\boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k)} + \frac{\delta}{2}\nabla \log \pi(\boldsymbol{x}^{(k)}) - \frac{\delta}{2\lambda}\left[\boldsymbol{x}^{(k)} - \mathrm{prox}_{\chi_{\mathcal{B}_\tau}}(\boldsymbol{x}^{(k)})\right] + \sqrt{\delta}\,\boldsymbol{w}^{(k+1)}.$$

1. $\boldsymbol{x}^{(k)}$ **is already in** $\mathcal{B}_\tau$: term $\left[\boldsymbol{x}^{(k)} - \mathrm{prox}^{\lambda}_{\chi_{\mathcal{B}_\tau}}(\boldsymbol{x}^{(k)})\right]$ disappears and recover usual Langevin MCMC.

2. $\boldsymbol{x}^{(k)}$ **is not in** $\mathcal{B}_\tau$: a step is also taken in the direction $-\left[\boldsymbol{x}^{(k)} - \mathrm{prox}^{\lambda}_{\chi_{\mathcal{B}_\tau}}(\boldsymbol{x}^{(k)})\right]$, which moves the next iteration in the direction of the projection of $\boldsymbol{x}^{(k)}$ onto the convex set $\mathcal{B}_\tau$. Acts to push the Markov chain back into the constraint set $\mathcal{B}_\tau$ if it wanders outside of it.



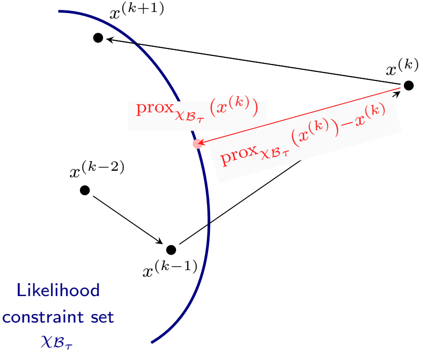
Proximal nested sampling intuition

Recall proximal nested sampling Markov chain (from previous slide):

$$\boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k)} + \frac{\delta}{2}\nabla \log \pi(\boldsymbol{x}^{(k)}) - \frac{\delta}{2\lambda}\left[\boldsymbol{x}^{(k)} - \operatorname{prox}_{\chi_{\mathcal{B}_\tau}}(\boldsymbol{x}^{(k)})\right] + \sqrt{\delta}\,\boldsymbol{w}^{(k+1)}.$$

1. **$\boldsymbol{x}^{(k)}$ is already in $\mathcal{B}_\tau$:** term $\left[\boldsymbol{x}^{(k)} - \operatorname{prox}^{\lambda}_{\chi_{\mathcal{B}_\tau}}(\boldsymbol{x}^{(k)})\right]$ disappears and recover usual Langevin MCMC.

2. **$\boldsymbol{x}^{(k)}$ is not in $\mathcal{B}_\tau$:** a step is also taken in the direction $-\left[\boldsymbol{x}^{(k)} - \operatorname{prox}^{\lambda}_{\chi_{\mathcal{B}_\tau}}(\boldsymbol{x}^{(k)})\right]$, which moves the next iteration in the direction of the projection of $\boldsymbol{x}^{(k)}$ onto the convex set $\mathcal{B}_\tau$. Acts to push the Markov chain back into the constraint set $\mathcal{B}_\tau$ if it wanders outside of it.



Tweedie's formula

Score matching and **denoising diffusion models** achieve state-of-the-art performance in deep generative modelling.

Tweedie's formula

Consider noisy observations $\boldsymbol{z} \sim \mathcal{N}(\boldsymbol{x}, \sigma^2 I)$ of $\boldsymbol{x}$ sampled from some underlying prior $\pi(x)$. **Tweedie's** formula gives the posterior expectation of $\boldsymbol{x}$ given $\boldsymbol{z}$ as

$$\mathbb{E}(\boldsymbol{x} \mid \boldsymbol{z}) = \boldsymbol{z} + \sigma^2 \nabla \log p(\boldsymbol{z}),$$

where $p(z)$ is the marginal distribution of z.

- Can be used to relate a learned denoiser D_σ (MMSE estimator) to the score $\nabla \log p(z)$.
- $p(z)$ is a regularised version of the target prior probability $\pi(x)$, i.e. $\pi_\sigma(x)$

Tweedie's formula

Score matching and **denoising diffusion models** achieve state-of-the-art performance in deep generative modelling.

Tweedie's formula

Consider noisy observations $\boldsymbol{z} \sim \mathcal{N}(\boldsymbol{x}, \sigma^2 I)$ of $\boldsymbol{x}$ sampled from some underlying prior $\pi(x)$. **Tweedie's** formula gives the posterior expectation of $\boldsymbol{x}$ given $\boldsymbol{z}$ as

$$\mathbb{E}(\boldsymbol{x} \mid \boldsymbol{z}) = \boldsymbol{z} + \sigma^2 \nabla \log p(\boldsymbol{z}),$$

where $p(z)$ is the marginal distribution of z.

- Can be used to relate a learned denoiser D_σ (MMSE estimator) to the score $\nabla \log p(z)$.
- $p(z)$ is a regularised version of the target prior probability $\pi(x)$, i.e. $\pi_\sigma(x)$

Tweedie's formula

Score matching and **denoising diffusion models** achieve state-of-the-art performance in deep generative modelling.

Tweedie's formula

Consider noisy observations $\boldsymbol{z} \sim \mathcal{N}(\boldsymbol{x}, \sigma^2 \boldsymbol{I})$ of $\boldsymbol{x}$ sampled from some underlying prior $\pi(x)$. **Tweedie's** formula gives the posterior expectation of $\boldsymbol{x}$ given $\boldsymbol{z}$ as

$$\mathbb{E}(\boldsymbol{x} \,|\, \boldsymbol{z}) = \boldsymbol{z} + \sigma^2 \nabla \log p(\boldsymbol{z}),$$

where $p(z)$ is the marginal distribution of z.

- Can be used to relate a learned denoiser D_σ (MMSE estimator) to the score $\nabla \log p(z)$.
- $p(z)$ is a regularised version of the target prior probability $\pi(x)$, i.e. $\pi_\sigma(x)$

By Tweedie's formula the score of the **regualised prior related to the learned denoiser** by

$$\nabla \log \pi_\sigma(\boldsymbol{x}) = \sigma^{-1}(D_\sigma(\boldsymbol{x}) - \boldsymbol{x}).$$

Proximal nested sampling with learned data-driven priors

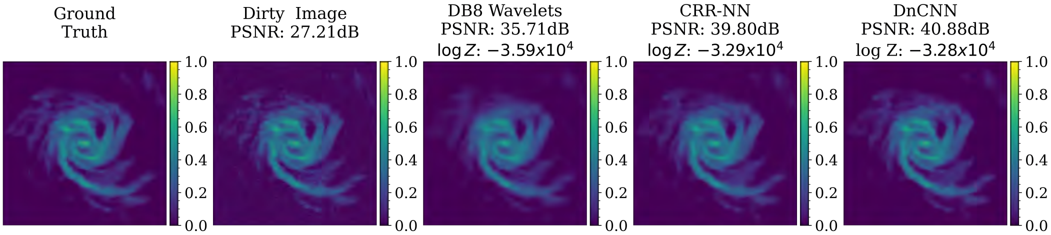
Substituting the denoiser $\nabla \log \pi_\epsilon(\boldsymbol{x}) = \epsilon^{-1}(D_\epsilon(\boldsymbol{x}) - \boldsymbol{x})$ into the proximal nested sampling Markov chain update:

$$\boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k)} - \underbrace{\frac{\delta\alpha}{2\epsilon}\left[\boldsymbol{x}^{(k)} - D_\epsilon(\boldsymbol{x}^{(k)})\right]}_{\text{Prior term}} - \underbrace{\frac{\delta}{2\lambda}\left[\boldsymbol{x}^{(k)} - \text{prox}_{\chi_{\mathcal{B}_\tau}}(\boldsymbol{x}^{(k)})\right]}_{\text{Likelihood constraint term}} + \sqrt{\delta}\boldsymbol{w}^{(k+1)} .$$

Hand-crafted vs data-driven priors

Consider simple galaxy denoising inverse problem with:

- **hand-crafted prior** based on sparsity-promoting wavelet representation;
- **data-driven priors** based on a deep neural networks (Goujon et al., 2023 Ryu et al., 2019).



Which model best?

- PSNR $\Rightarrow$ **data-driven priors best** but **require ground-truth**;
- Bayesian evidence $\Rightarrow$ **data-driven priors best** (**no ground-truth knowledge**).

Summary

- **Proximal nested sampling** (arXiv:2106.03646) framework scales to **high-dimensions**, opening up Bayesian model comparison for, e.g., imaging problems.
- Constrained to **log-convex likelihoods**, which are ubiquitous in imaging sciences (e.g. Gaussian likelihood).
- Prior not constrained to be log-convex so can be a deep neural network.
- Recently developed **learned proximal nested sampling** (arXiv:2307.00056) approach to support data-driven priors exploiting Tweedie's formula.
- We can now compare different data-driven priors for our high-dimensional imaging inverse problems.

References I

Adler, J., & Öktem, O. (2018). Deep Bayesian Inversion. arXiv e-prints, Article arXiv:1811.05910, arXiv:1811.05910.

Angelopoulos, A. N., & Bates, S. (2023). Conformal prediction: A gentle introduction. Foundations and Trends® in Machine Learning, 16(4), 494–591.

Bendel, M., Ahmad, R., & Schniter, P. (2023). A regularized conditional gan for posterior sampling in image recovery problems. In A. Oh, T. Naumann, A. Globerson, K. Saenko, M. Hardt, & S. Levine (Eds.), Advances in neural information processing systems (pp. 68673–68684, Vol. 36). Curran Associates, Inc.

Cai, X., McEwen, J. D., & Pereyra, M. (2022). Proximal nested sampling for high-dimensional bayesian model selection. Statistics and Computing, 32(5), 87.

Cai, X., Pereyra, M., & McEwen, J. D. (2018a). Uncertainty quantification for radio interferometric imaging – I. Proximal MCMC methods. Monthly Notices of the Royal Astronomical Society, 480(3), 4154–4169.

Cai, X., Pereyra, M., & McEwen, J. D. (2018b). Uncertainty quantification for radio interferometric imaging: II. MAP estimation. Monthly Notices of the Royal Astronomical Society, 480(3), 4170–4182.

Goujon, A., Neumayer, S., Bohra, P., Ducotterd, S., & Unser, M. (2023). A neural-network-based convex regularizer for inverse problems. IEEE Transactions on Computational Imaging, 9, 781–795.

Kern, J., Bobin, J., & Kervazo, C. (2024). Evil-deconv: Efficient variability-informed learned deconvolution using algorithm unrolling. Submitted.

References II

Liaudat, T. I., Mars, M., Price, M. A., Pereyra, M., Betcke, M. M., & McEwen, J. D. (2023). Scalable Bayesian uncertainty quantification with data-driven priors for radio interferometric imaging. arXiv e-prints, Article arXiv:2312.00125, arXiv:2312.00125.

McEwen, J. D., Liaudat, T. I., Price, M. A., Cai, X., & Pereyra, M. (2023). Proximal nested sampling with data-driven priors for physical scientists. arXiv e-prints, Article arXiv:2307.00056, arXiv:2307.00056.

Pereyra, M. (2016). Proximal markov chain monte carlo algorithms. Statistics and Computing, 26(4), 745–760.

Pereyra, M. (2017). Maximum-a-posteriori estimation with bayesian confidence regions. SIAM Journal on Imaging Sciences, 10(1), 285–302.

Pereyra, M., Mieles, L. V., & Zygalakis, K. C. (2020). Accelerating proximal markov chain monte carlo by using an explicit stabilized method. SIAM Journal on Imaging Sciences, 13(2), 905–935.

Ryu, E., Liu, J., Wang, S., Chen, X., Wang, Z., & Yin, W. (2019, September). Plug-and-play methods provably converge with properly trained denoisers. In K. Chaudhuri & R. Salakhutdinov (Eds.), Proceedings of the 36th international conference on machine learning (pp. 5546–5557, Vol. 97). PMLR.

Tachella, J., & Pereyra, M. (2023). Equivariant Bootstrapping for Uncertainty Quantification in Imaging Inverse Problems. arXiv e-prints, Article arXiv:2310.11838, arXiv:2310.11838.

Questions?