Proximal Nested Sampling with Data-Driven AI Priors

Tobias I. Liaudat[1], Henry J. Aldridge[2], Matthew A. Price[2], Marcelo Pereyra[3], Xiaohao Cai[4], Jason D. McEwen[2]
[1]CEA, [2]MSSL UCL, [3]HWU, [4]Southhampton.
Correspondence: tobias.liaudat@cea.fr, henry.aldridge.23@ucl.ac.uk, jason.mcewen@ucl.ac.uk

Overview

Problem Nested sampling computes the Bayesian evidence for model selection by sampling from a prior constrained to likelihood-level sets. At high dimensions, existing nested sampling algorithms fail given the computationally challenging task.

Solution Proximal nested sampling utilizes proximal calculus and Moreau-Yosida regularisation to support non-differentiable, log-convex likelihoods, such as those found in imaging applications, to compute the Bayesian evidence at high dimensions.

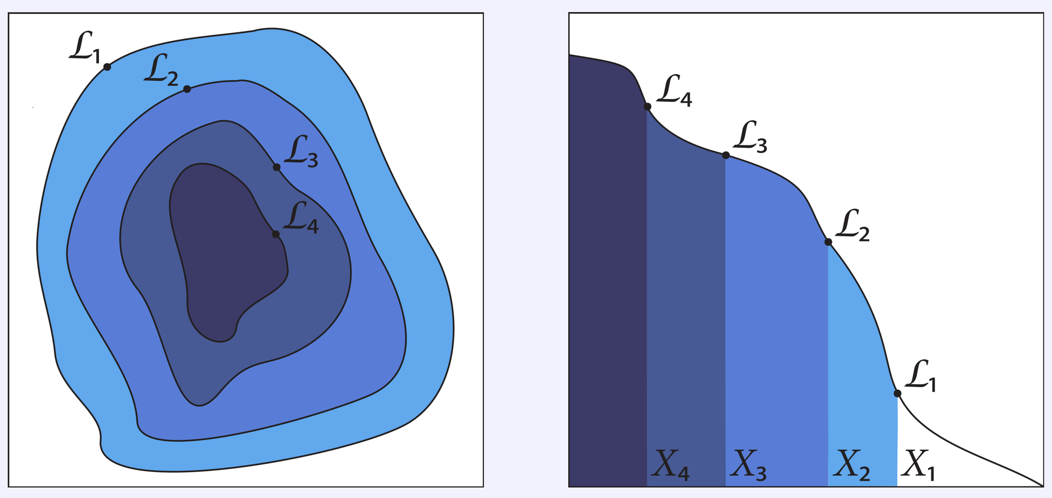
Nested Sampling

▶ **Bayesian model selection** for a set of parameters $\boldsymbol{x} \in \Omega$ and data $\boldsymbol{y}$ requires computation of Bayesian evidence Z for model M

$$Z = P(\boldsymbol{y}|M) = \int_{\Omega} d\boldsymbol{x}\ \mathcal{L}(\boldsymbol{x})\pi(\boldsymbol{x}).$$

where likelihood $\mathcal{L}(\boldsymbol{x}) = P(\boldsymbol{y}|\boldsymbol{x}, M)$ and prior $\pi(\boldsymbol{x}) = P(\boldsymbol{x}|M)$.

▶ **Nested Sampling**[1] computes prior volumes X_i nested within likelihood-level sets L_i to compute Z as a 1-D integral.

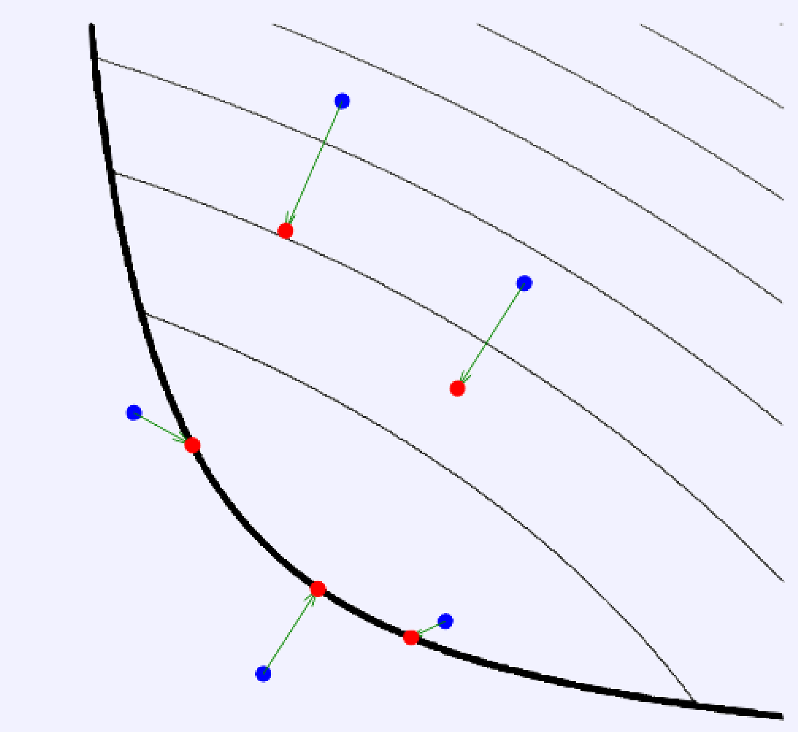


Nested subspaces[2] — Reparameterised likelihood[2]

Proximal Calculus

▶ Proximity operator of a convex function $f : \mathbb{R}^n \mapsto \mathbb{R}$ given by

$$\text{prox}_f^{\lambda}(\boldsymbol{x}) = \underset{\boldsymbol{u}}{\text{argmin}}[f(\boldsymbol{u}) + \|\boldsymbol{u} - \boldsymbol{x}\|^2/2\lambda].$$

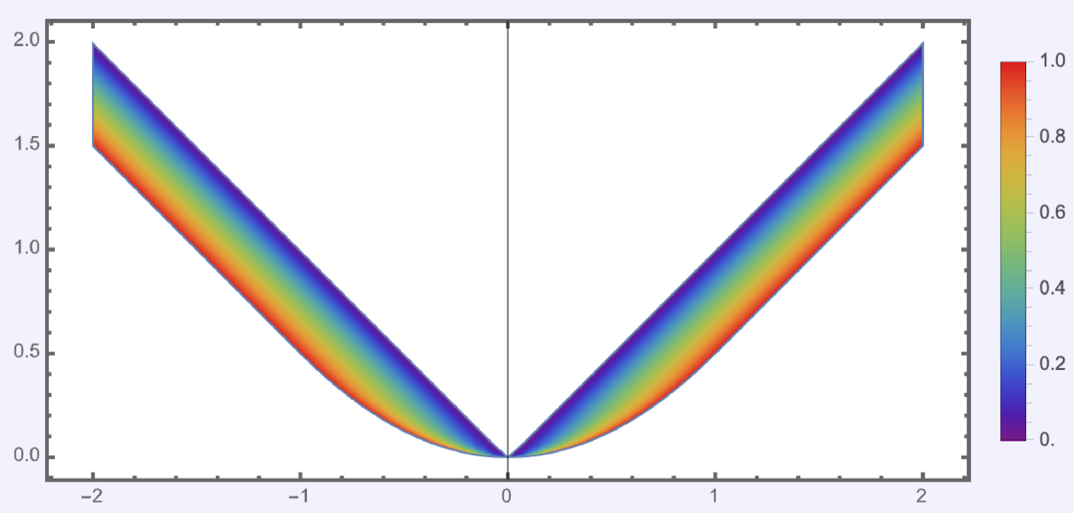


prox_f mapping blue to red points[3]. Domain boundary (thick black) and level-sets (thin black) of f.

▶ Moreau-Yosida envelope of f given by

$$f^{\lambda}(\boldsymbol{x}) = \inf_{\boldsymbol{u} \in \mathbb{R}^n} f(\boldsymbol{u}) + \|\boldsymbol{u} - \boldsymbol{x}\|^2/2\lambda.$$

- As $\lambda \to 0, f^{\lambda}(\boldsymbol{x}) \to f(\boldsymbol{x})$.
- $\nabla f^{\lambda}(\boldsymbol{x}) = (\boldsymbol{x} - \text{prox}_f^{\lambda}(\boldsymbol{x}))/\lambda$.



Moreau-Yosida of $|x|$ for varying λ

Proximal Nested Sampling

▶ **Nested sampling requires sampling from a prior constrained to likelihood level-sets**, which is very difficult at high dimensions e.g. imaging applications.

▶ **Proximal nested sampling**[4] utilizes proximal calculus to alleviate this challenge where likelihoods are log-convex and lower semicontinuous, e.g. in imaging, by **applying the Moreau-Yosida regularisation to likelihood level-set constraints**.

▶ Langevin Markov chain Monte Carlo (MCMC) efficiently samples high dimensions using gradient information; can adopt our method to produce the proximal nested sampling Markov chain

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

with characteristic function $\chi_{\mathcal{B}_\tau}$ corresponding to constraint $\mathcal{L}(\boldsymbol{x}) \geq L^*$ and Brownian motion $\boldsymbol{w}$.

▶ Log-prior $\log \pi(\boldsymbol{x})$ may be non-differentiable, in which case we similarly apply Moreau-Yosida regularisation.

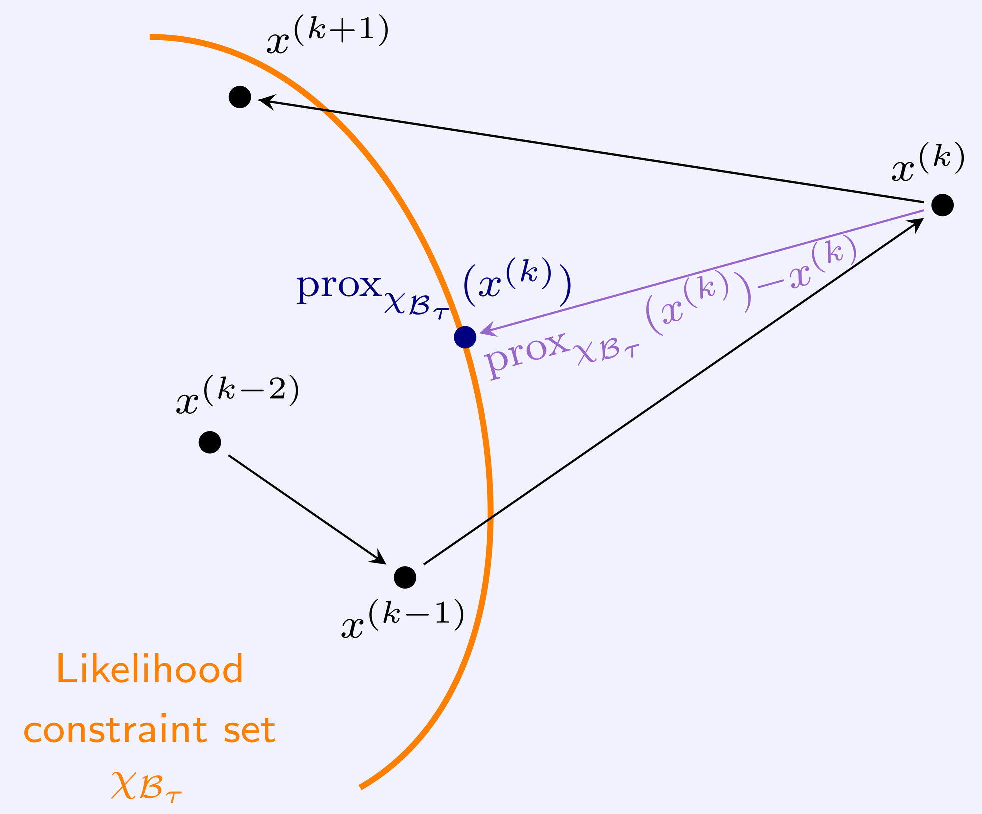
▶ **Tweedie's formula**[5] links noisy observations $z \sim \mathcal{N}(x, \sigma^2 I)$ and samples $x \sim q(x)$ without knowledge of $q(x)$:

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

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

▶ Leveraging Tweedie's formula, can relate score of Gaussian-smoothed prior to denoiser D_ϵ trained to recover x from $x_\epsilon \sim \mathcal{N}(x, \epsilon I)$

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



ProxNest Markov chain (in black). When outside the constrained likelihood level-set $\chi_{\mathcal{B}_\tau}$, $\text{prox}_{\chi_{\mathcal{B}_\tau}}$ pushes the chain back into the constraint set.

▶ Denoisers D_ϵ can include deep neural networks leading to **data-driven AI priors** that can be implemented in a **plug-and-play (PnP)**[6,7] approach with proximal nested sampling[8].

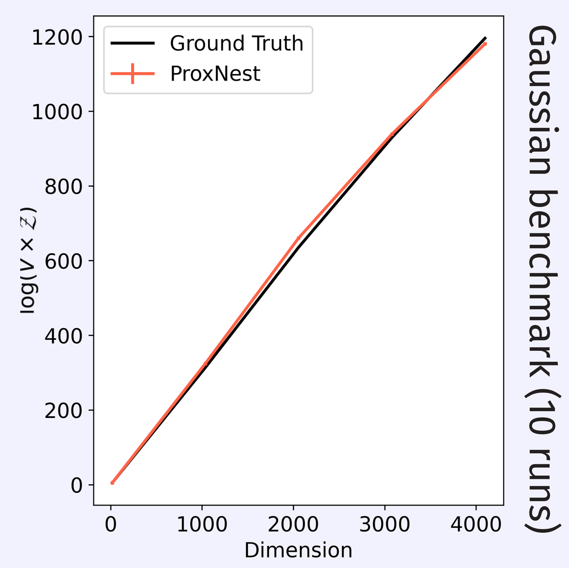
ProxNest with Data-Driven AI Priors

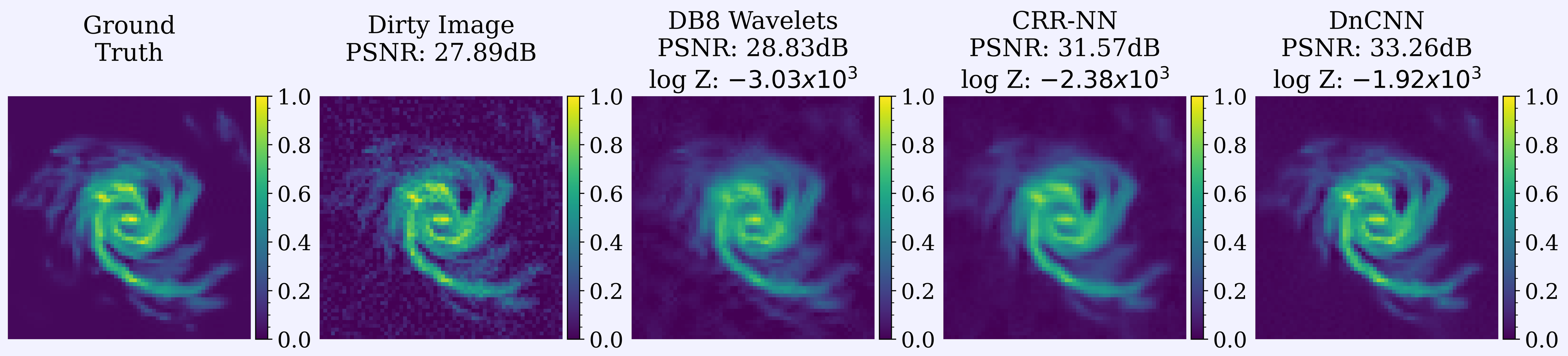
▶ Validated the method with a Gaussian benchmark.

▶ Evaluated denoising an image with selection of denoisers: hand-crafted priors (DB8 wavelets) and data driven AI priors trained on natural images (CRR-NN[9] and DnCNN[7]).

▶ Bayesian evidence favours the DnCNN model, which the ground truth also demonstrates is the best reconstruction.

▶ Method implemented in `ProxNest` Python package. New and updated codebase soon to be available at https://github.com/astro-informatics.





Denoising with DB6 wavelets, CRR-NN[9], DnCNN[7]. Note: Peak signal-to-noise ratio (PSNR) requires the ground truth, which is not available in realistic settings.

References

[1] Skilling J., 2006, Nested sampling for general Bayesian computation.
[2] Feroz F. et al., 2013, Importance nested sampling and the MultiNest algorithm.
[3] Parikh N. and Boyd S., 2013, Proximal algorithms.
[4] Cai X. et al., 2022, Proximal nested sampling for high-dimensional Bayesian model selection.
[5] Robbins H., 1956, An empirical Bayes approach to statistics.
[6] Venkatakrishnan S.V. et al., 2013, Plug-and-play priors for model based reconstruction.
[7] Ryu E. et al., 2019, Plug-and-Play methods provably converge with properly trained denoisers.
[8] McEwen J. D. et al., 2023, Proximal nested sampling with data-driven priors for physical scientists.
[9] Goujon A. et al., 2022, A neural-network-based convex regularizer for inverse problems.