Clustering Considerations for Nested Sampling

5.3.1 Cluster recognition

Any cluster recognition algorithm can be substituted at this point. One must take care that this is not run too often, or one runs the risk of adding a large overhead to the calculation. In practice, checking for clustering every $\sim \mathcal{O}(n_{\text{live}})$ iterations is sufficient, since the prior will have only compressed by a factor e. We encourage users of PolyChord to experiment with their own preferred cluster recognition, in addition to that provided and described below.

"PolyChord: next-generation nested sampling"
arXiv:1506.00171

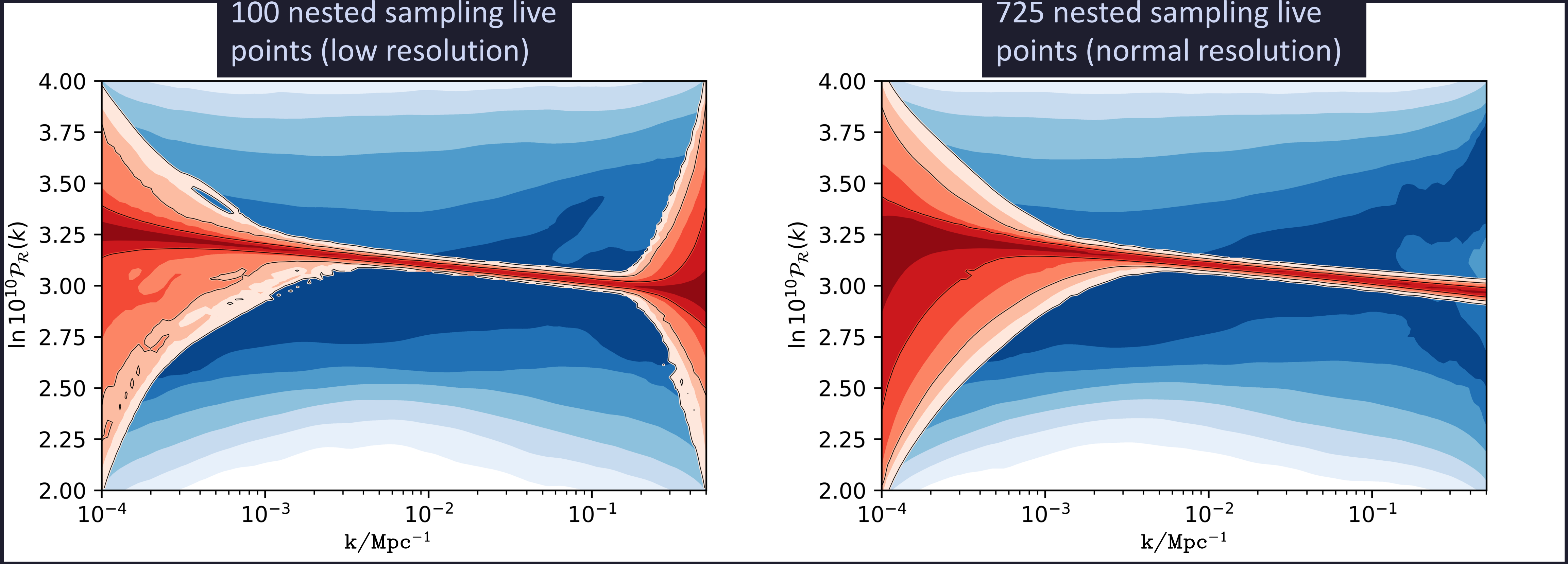
Adam Ormondroyd

University of Cambridge

Kavli Institute for Cosmology

Motive: primordial matter power spectrum flex-knot reconstruction

Primordial matter power spectrum



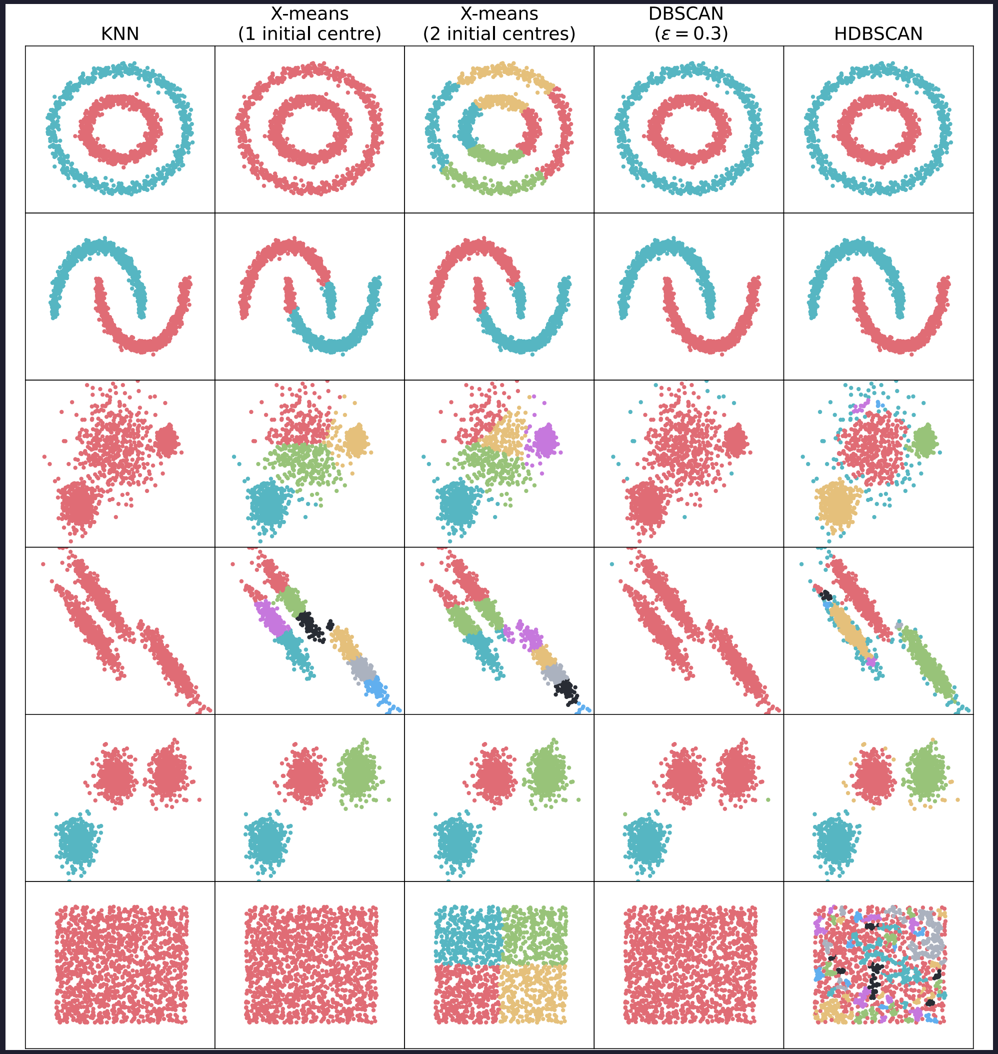
large scales
(cosmic variance)

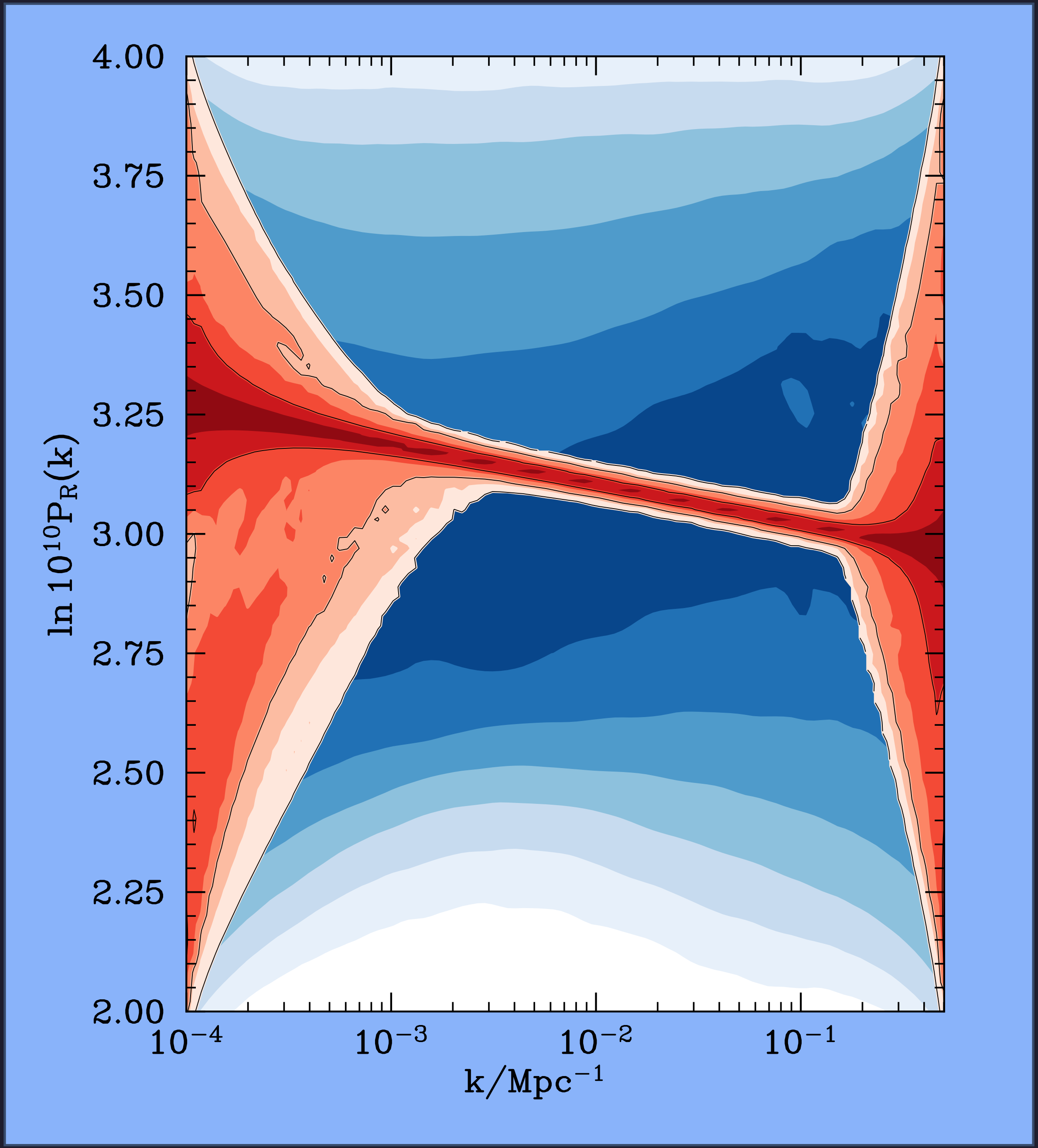
small scales
(detector resolution)

large scales
(cosmic variance)

small scales
(detector resolution)

Clustering choices are abundant!





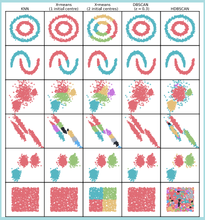
Clustering Considerations for Nested Sampling

Adam Ormondroyd ano23@cam.ac.uk

Kavli Institute for Cosmology · Cavendish Laboratory · University of Cambridge

Clustering algorithms are integral to multi-modal nested sampling, for both region-based samplers such as `MultiNest`, and chain-based samplers such as `PolyChord`. Robust identification clusters of live points is crucial for effective spawning of new live points, prior volume estimation and therefore the total evidence calculation. Reliable cluster detection also allows the calculation of the sub-evidences of each cluster, which may correspond to different physical phenomena. We have explored extensions to the clustering approach within `PolyChord`, and found that including correlation between the volume estimates of clusters increases the accuracy of evidence calculations. We show how different clustering methods affect a reconstruction of the cosmological primordial matter power spectrum $\mathcal{P}_\mathcal{R}(k)$.

Clustering choice



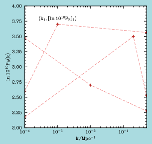
2D demonstration of `PolyChord`'s K-Nearest Neighbours, `pyclustering`'s X-means and `scikit-learn`'s (H)DBSCAN.

The nested sampling [1, 2] algorithm `PolyChord` was originally advertised suggesting that the user should experiment with their favourite cluster-identifying methods, but provides no guidance on how to do so [3, 4]. We have added an interface which allows the user to substitute any clustering strategy at the Python level, allowing for easier experimentation with alternatives such as the selection provided by `scikit-learn` and `pyclustering` [5, 6].

Some algorithms are better suited than others to identifying posterior modes of nested sampling live points, for example K-means and spectral clustering need to be told the number of clusters to look for, others may not assign every point to a cluster. Some approaches find clusters where there are none!

Application to cosmology $\mathcal{P}_\mathcal{R}(k)$

This investigation was initially motivated by a reconstruction of the primordial matter power spectrum $\mathcal{P}_\mathcal{R}(k)$ using flexknots [7, 8], and the Planck 2018 likelihoods [9, 10]. Planck measured the C_ℓ multipole range $1 \leq \ell \leq 7000$, corresponding to $10^{-4} \leq k/\mathrm{Mpc}^{-1} \leq 10^{-0.3}$ [11]. Flexknots are parameterisations of 1D functions, consisting of a series of splines (in this case linear) joined at knots. The number and positions of knots are determined by the data, which can be performed by either combining several runs with fixed number of knots, or the number of knots being itself a parameter. In the former case, we noticed that with three knots only the mode with the central knot towards the left was being fully explored. `PolyChord`'s native K-nearest-neighbours clustering was unable to separate the positions of the central knot into two distinct clusters, so we explored both off-the-shelf clustering algorithms and an approach which includes likelihood information, $\mathcal{L}$-means.



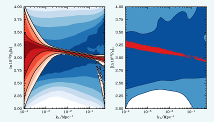
Examples of $\mathcal{P}_\mathcal{R}(k)$ flexknots with three knots each.

Prototype $\mathcal{L}$-means algorithm

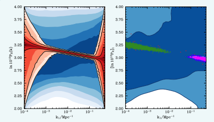
```
Perform K-means with K=2
Calculate the likelihood at the centres of
    the two clusters
find the midpoint of the two centres
calculate the likelihood at the midpoint

if (the midpoint has greater likelihood
   than either centre):
    all points are within the same cluster
else:
    recurively apply L-means to each cluster
```

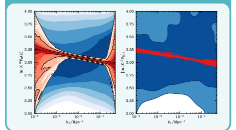
K-nearest-neighbours



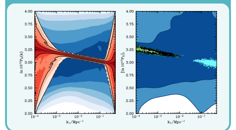
X-means



DBSCAN ($\epsilon = 0.5$)



$\mathcal{L}$-means

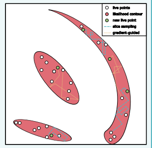


K-nearest-neighbours and DBSCAN fail to separate the central knot into two clusters, while X-means does so reliably. $\mathcal{L}$-means is able to separate the central knot into multiple clusters, but also tends to over-cluster. Functional posterior plots were created using `fgivenx` [12], and scatterplots showing clusters were created using a development branch of `anesthetic` [13].

Classes of Nested Sampler

Nested sampling algorithms have two main strategies for sampling new points from the prior:

Chain-based

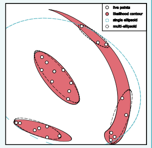


A sufficiently long random walk within the likelihood contour from an existing live point will generate a new live point.

`PolyChord` first uses the covariance matrix of the live points to whiten the space, then performs Neal's slice sampling along orthogonal directions in that space [3, 4]. `GGNS` (gradient-guided nested sampling) also implements both Neal's and Hamiltonian slice sampling, along with uniform sampling and random walks [14].

Multiple-modal problems render contour whitening ineffective, and a strategy is required to decide from which mode to sample since a random walk cannot pass through the likelihood contour.

Region-based



Region-based samplers construct regions around the live points to approximate the likelihood contour.

`MultiNest` constructs a series of ellipsoids [15–17]. These regions are usually expanded by a numerical factor to improve their chances of fully enclosing the likelihood contour, then a point is rejection-sampled from them. The curse of dimensionality means that these techniques are only effective up to $\mathcal{O}(10)$ dimensions, as either rejection sampling becomes inefficient, or the expansion factor would have to be so low that significant regions of the likelihood contour would be missed. Clustering can be used to separate the live points into discrete regions, rather than a single sparse region.

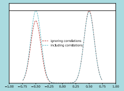
Hybrid methods combine the two approaches in an attempt to alleviate the dimensionality scaling of region-based methods, while reducing the number of likelihood evaluations made outside the contour [14, 18–20].

Correlated cluster volumes (in progress!)

When a cluster p is divided, the remaining prior volume X_p is divided among its subclusters X_i. However, in nested sampling, we do not know the precise prior volumes, only expectation values and errors. This is divided according to the proportion of live points in each cluster n_i:

$$\overline{X_i} = \frac{n_i}{n_p}\overline{X_p}, \quad \overline{X_i^2} = \frac{n_i(n_i+1)}{n_p(n_p+1)}\overline{X_p^2},$$

$$\overline{X_iX_j} = \frac{n_in_j}{n_p(n_p+1)}\overline{X_p^2}.$$



Accounting for prior volume correlations may provide more accurate posteriors.

Since $\overline{X_iX_j} \neq \overline{X_i}\,\overline{X_j}$, the error on the prior volumes estimates of each cluster are correlated. This is important when deciding from which cluster to sample; currently `PolyChord` chooses a cluster propotionally to its prior volume $\overline{X_i}$, but neglects their correlation. We are experimenting with drawing a set of X_i from their joint distribution before each live point is generated, which has shown promise with symmetric multi-modal likelihoods.

References

[1] John Skilling. Nested Sampling. In Rainer Fischer, Roland Preuss, and Udo Von Toussaint, editors, *Bayesian Inference and Maximum Entropy Methods in Science and Engineering: 24th International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering*, volume 735 of *American Institute of Physics Conference Series*, pages 395–405. AIP, November 2004.

[2] Greg Ashton, Noam Bernstein, Johannes Buchner, Xi Chen, Gábor Csányi, Andrew Fowlie, Farhan Feroz, Matthew Griffiths, Will Handley, Michael Habeck, Edward Higson, Michael Hobson, Anthony Lasenby, David Parkinson, Livia B. Pártay, Matthew Pitkin, Doris Schneider, Joshua S. Speagle, Leah South, John Veitch, Philipp Wacker, David J. Wales, and David Yallup. Nested sampling for physical scientists. *Nature Reviews Methods Primers*, 2:39, May 2022.

[3] W. J. Handley, M. P. Hobson, and A. N. Lasenby. polychord: nested sampling for cosmology. *MNRAS*, 450:L61–L65, June 2015.

[4] W. J. Handley, M. P. Hobson, and A. N. Lasenby. POLYCHORD: next-generation nested sampling. *MNRAS*, 453(4):4384–4398, November 2015.

[5] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[6] Andrei Novikov. Pyclustering: Data mining library. *Journal of Open Source Software*, 4(36):1230, apr 2019.

[7] Stefan Heimersheim. What it takes to measure Reionization with Fast Radio Bursts. *arXiv e-prints*, page arXiv:2203.12645, March 2022.

[8] Stefan Heimersheim, Leiv Rønneberg, Henry Linton, Filippo Pagani, and Anastasia Fialkov. FlexKnot and Gaussian Process for 21 cm global signal analysis and foreground separation. *MNRAS*, 527(4):11404–11421, February 2024.

[9] Planck Collaboration. Planck 2018 results. V. CMB power spectra and likelihoods. *A&A*, 641:A5, September 2020.

[10] Planck Collaboration. Planck 2018 results. VIII. Gravitational lensing. *A&A*, 641:A8, September 2020.

[11] Will J. Handley, Anthony N. Lasenby, Hiranya V. Peiris, and Michael P. Hobson. Bayesian inflationary reconstructions from Planck 2018 data. *PRD*, 100(10):103511, November 2019.

[12] Will Handley. fgivenx: A Python package for functional posterior plotting. *JOSS*, 3(28):849, August 2018.

[13] Will Handley. anesthetic: nested sampling visualisation. *JOSS*, 4:1414, May 2019.

[14] Pablo Lemos, Nikolay Malkin, Will Handley, Yoshua Bengio, Yashar Hezaveh, and Laurence Perreault-Levasseur. Improving Gradient-guided Nested Sampling for Posterior Inference. *arXiv e-prints*, page arXiv:2312.03911, December 2023.

[15] F. Feroz and M. P. Hobson. Multimodal nested sampling: an efficient and robust alternative to Markov Chain Monte Carlo methods for astronomical data analyses. *MNRAS*, 384(2):449–463, February 2008.

[16] F. Feroz, M. P. Hobson, and M. Bridges. MULTINEST: an efficient and robust Bayesian inference tool for cosmology and particle physics. *MNRAS*, 398(4):1601–1614, October 2009.

[17] Farhan Feroz, Michael P. Hobson, Ewan Cameron, and Anthony N. Pettitt. Importance Nested Sampling and the MultiNest Algorithm. *The Open Journal of Astrophysics*, 2(1):10, November 2019.

[18] Johannes Buchner. Nested Sampling Methods. *Statistics Surveys*, 17:169–215, January 2023.

[19] Martino Trassinelli. The Nested_fit data analysis program. *arXiv e-prints*, page arXiv:1907.12259, July 2019.

[20] Johannes Buchner. UltraNest - a robust, general purpose Bayesian inference engine. *The Journal of Open Source Software*, 6(60):3001, April 2021.