Diffusion meets Nested Sampling

Hills Coffee Talks

David Yallup - 21/05

SAMPLING

Suppose you are tasked with estimating the volume of a body of water, all I give you is a punt and a punting pole. How do you do it?

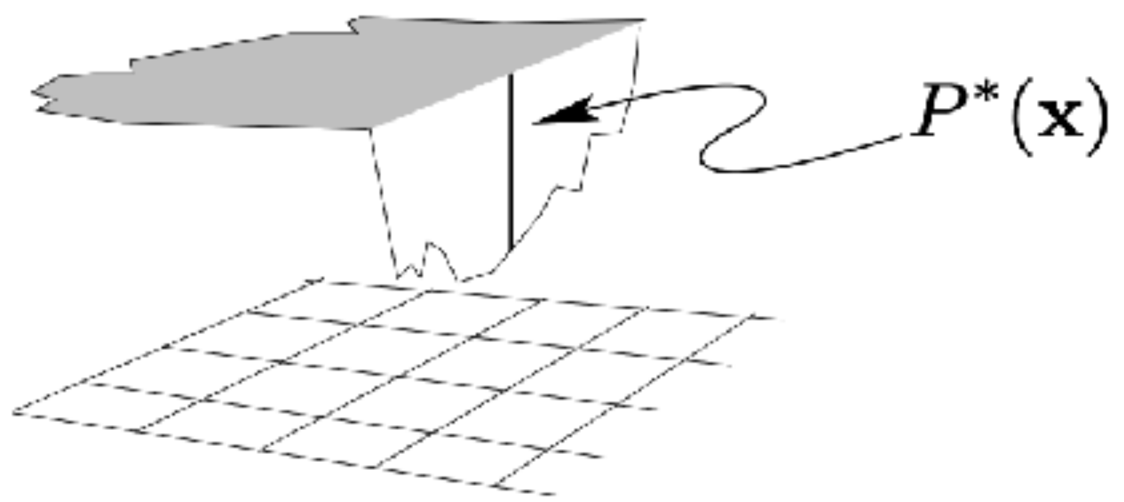
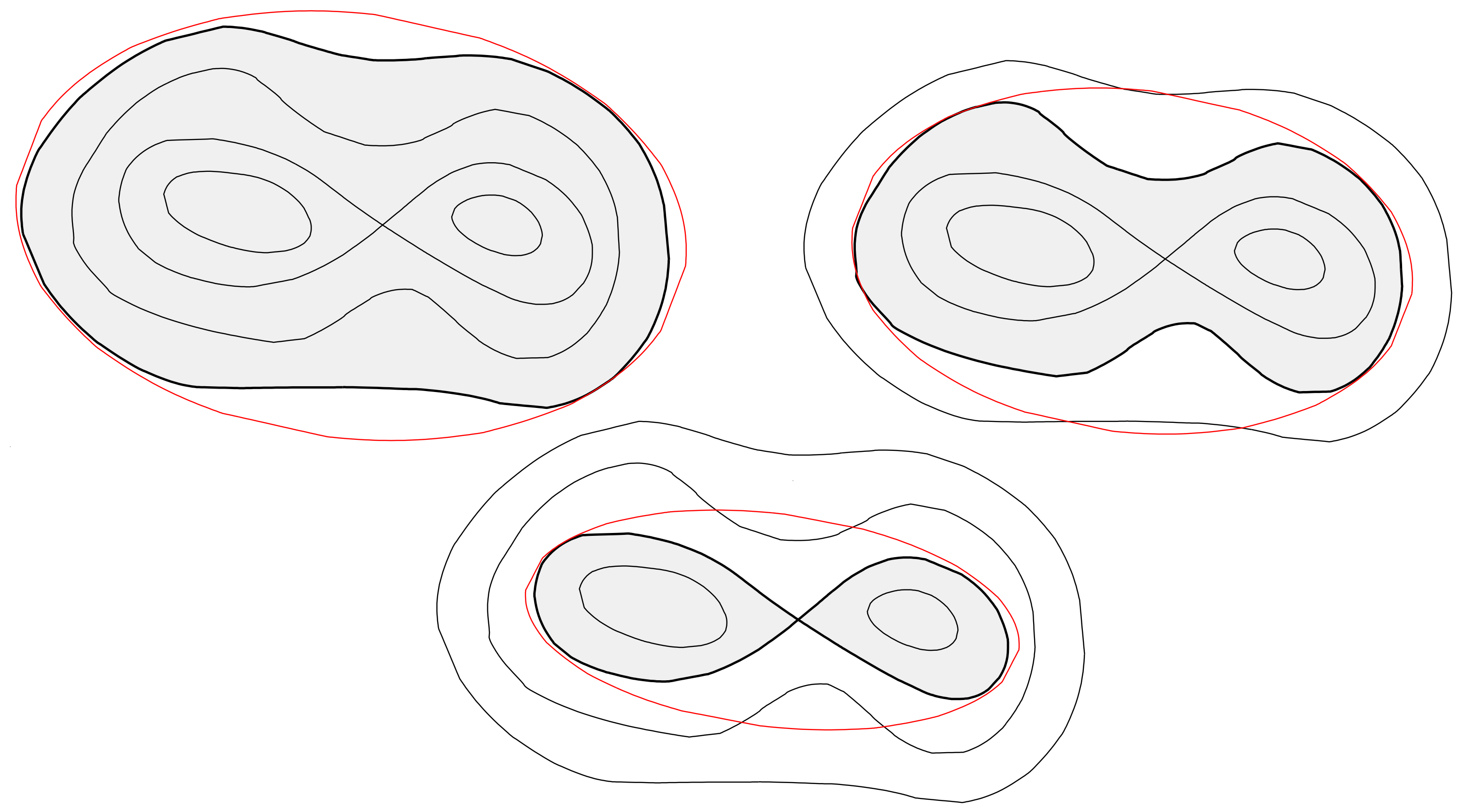


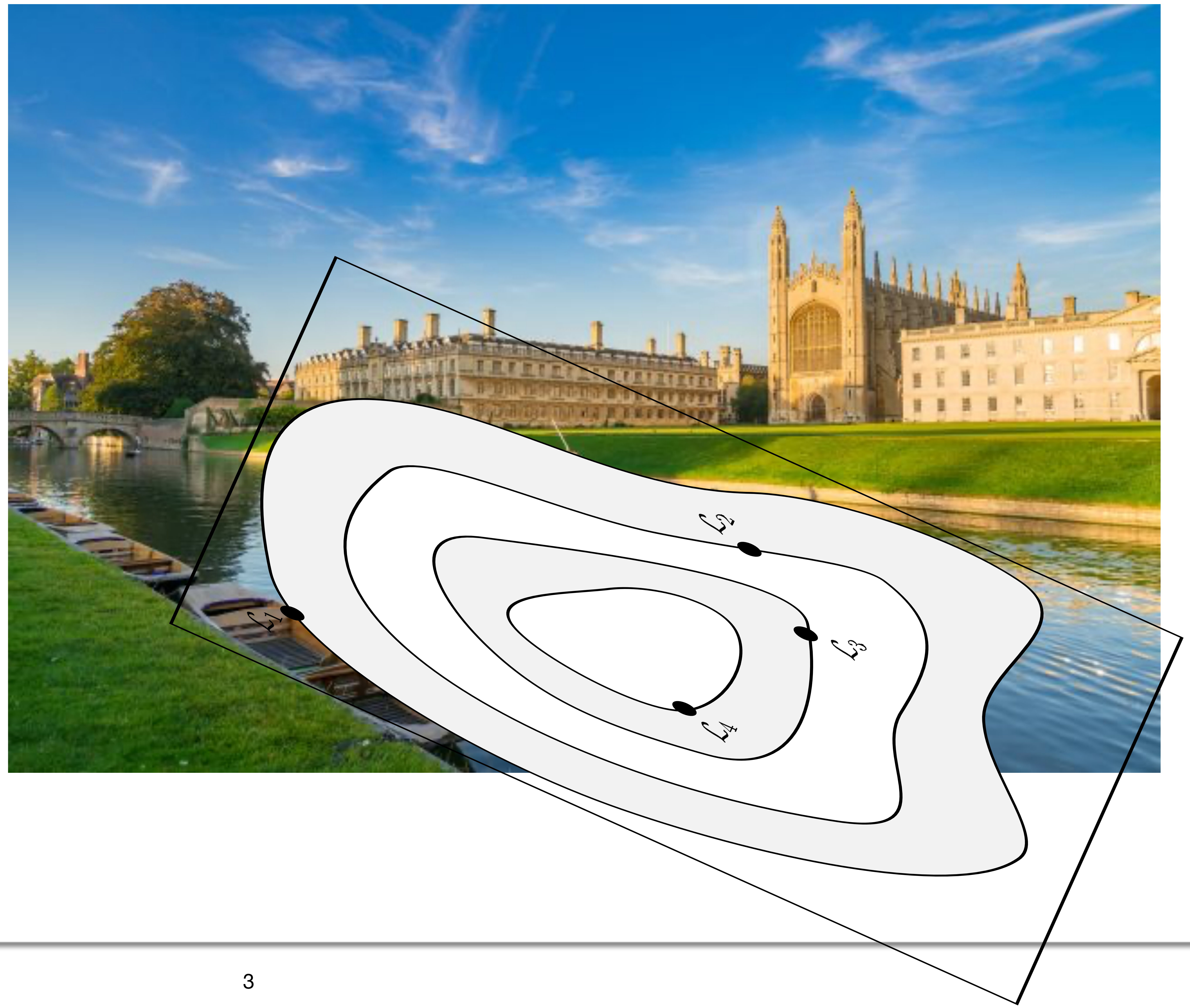
Figure 29.2. A lake whose depth at $\mathbf{x} = (x, y)$ is $P^*(\mathbf{x})$.

Information Theory, Inference, and Learning Algorithms - MacKay

NESTED SAMPLING



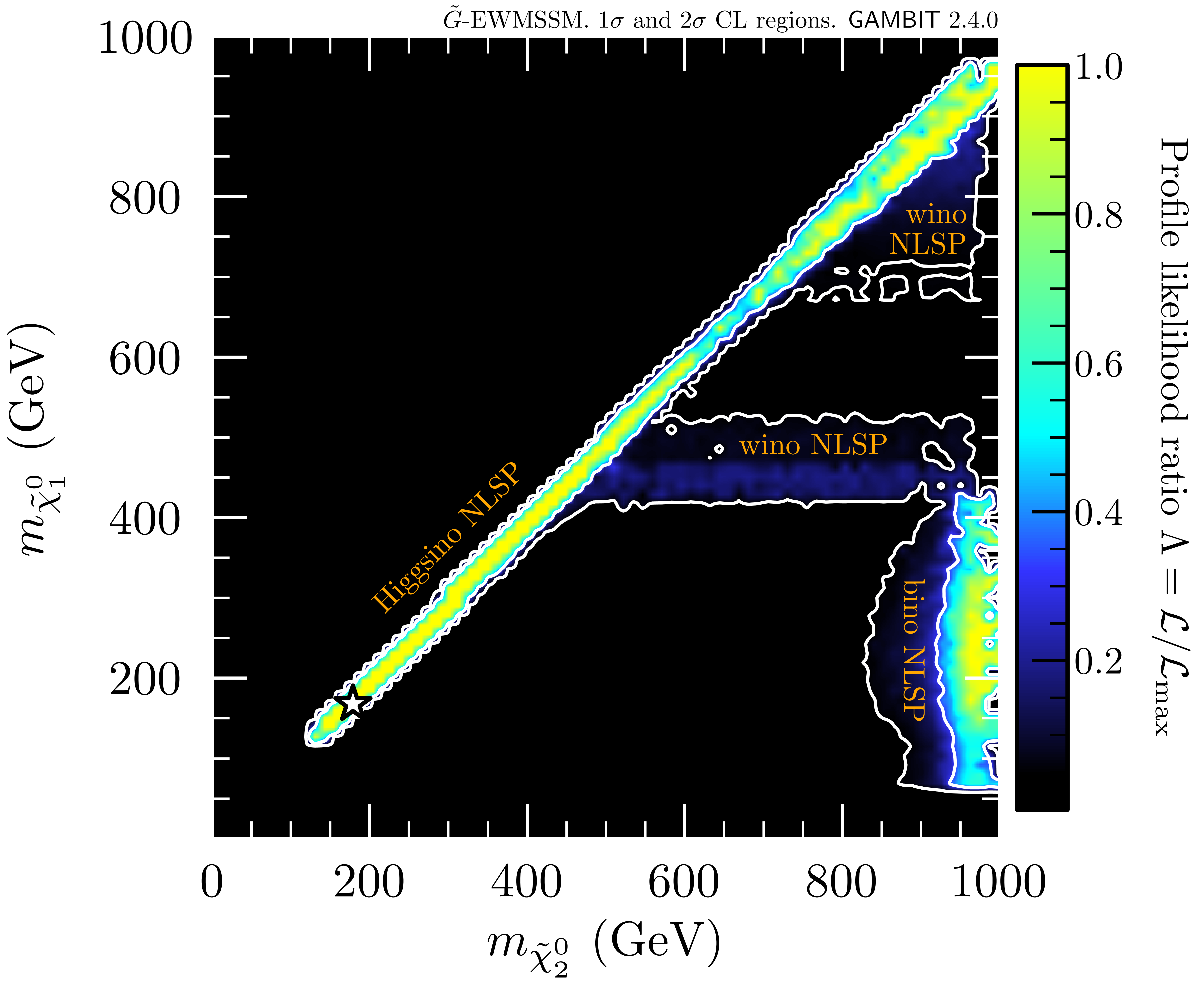
Illustrations from MultiNest paper[13]



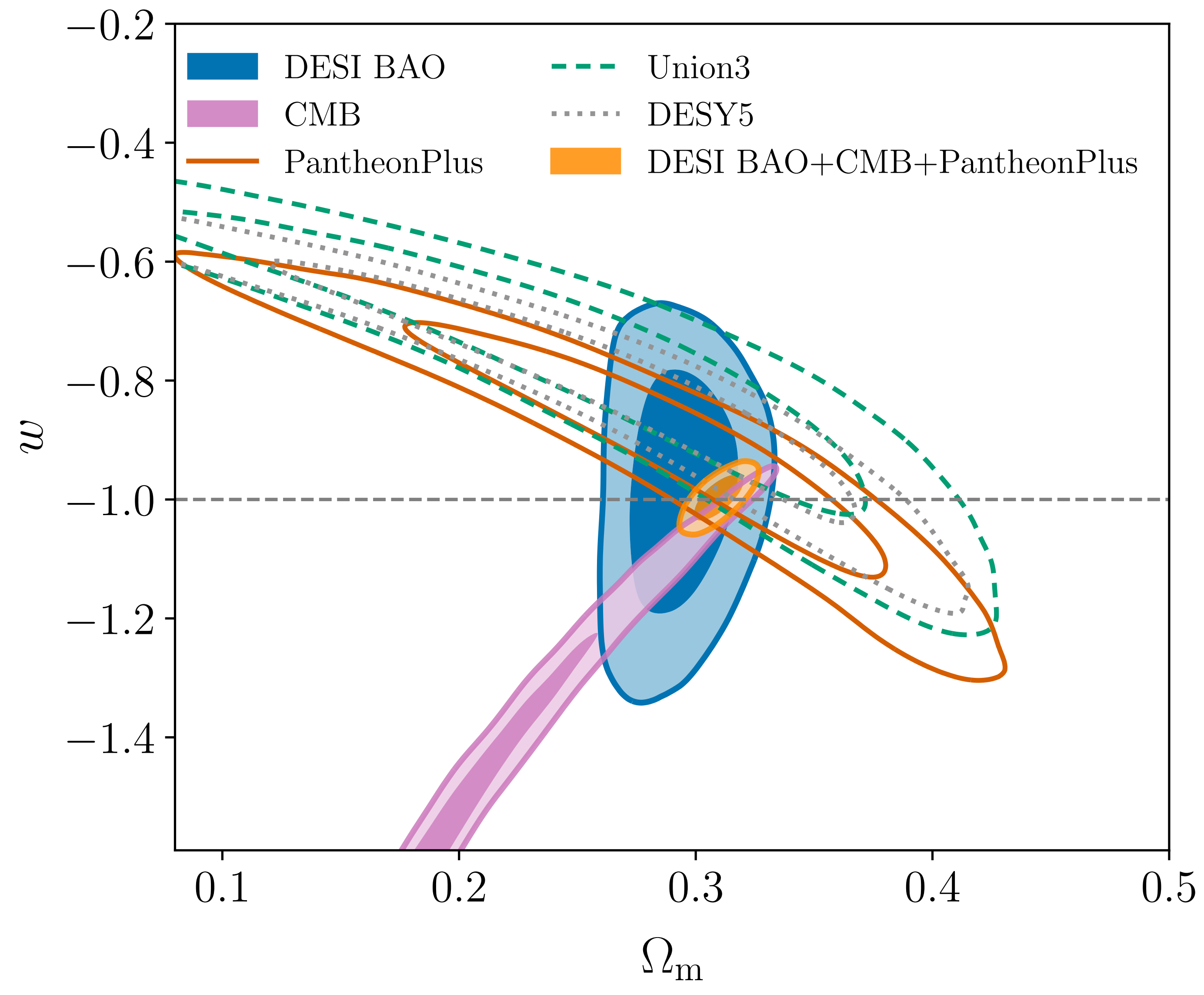
USEFUL TOOLS

INFERENCE

Fundamental physics is full of hard inference problems. Our optimization or sampling algorithms have to be able to navigate complex geometry



Gambit Collaboration SUSY profile likelihood scan[1]

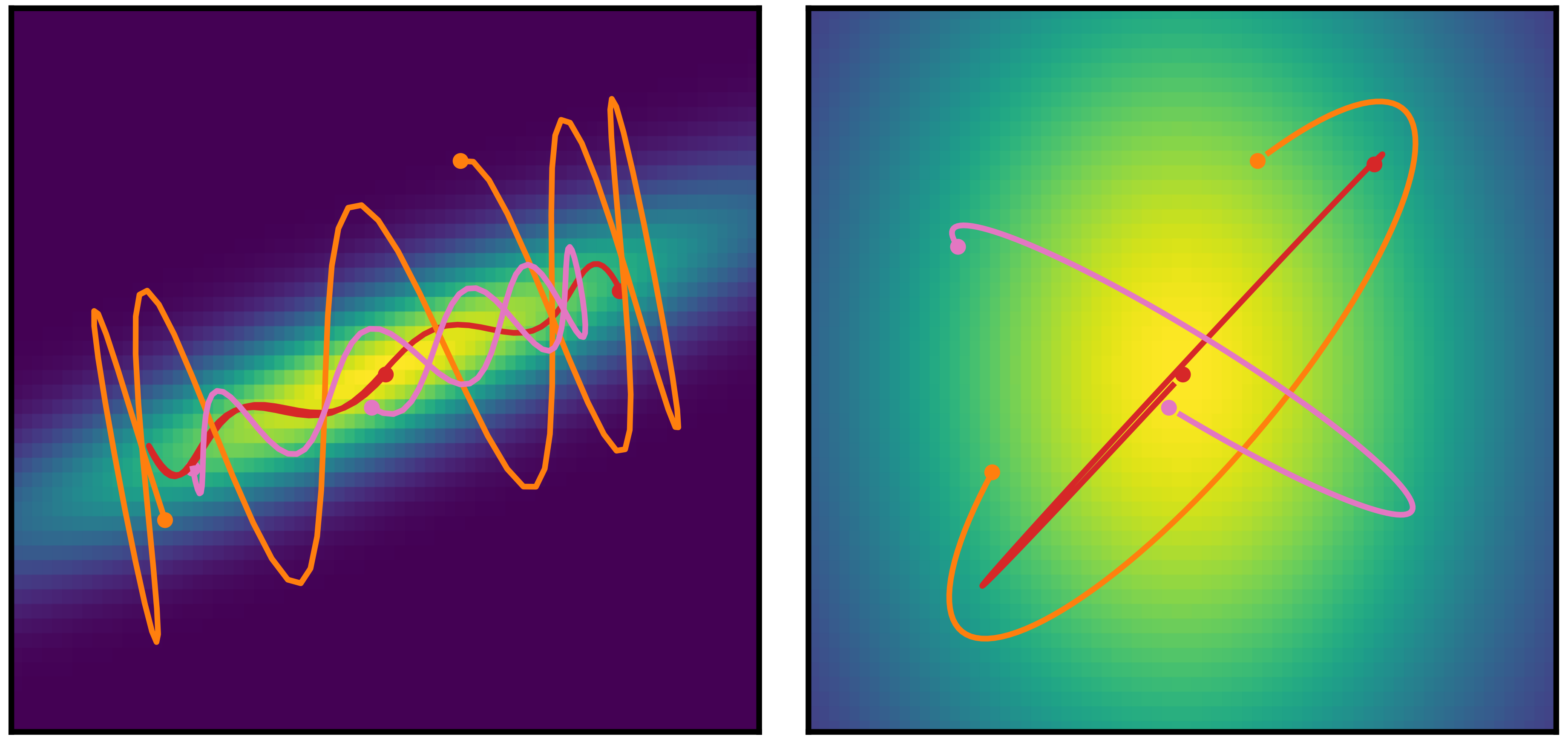


DESI collaboration LCDM samples[2]

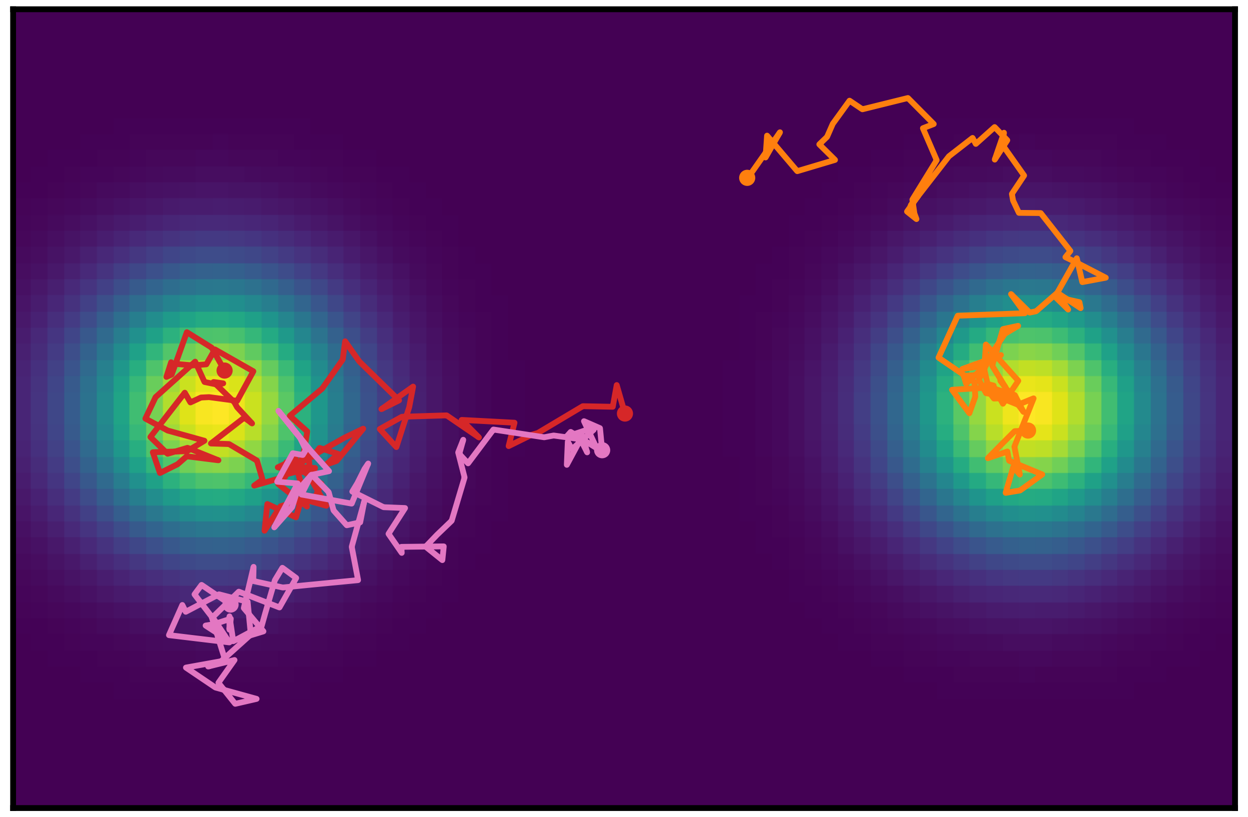
GEOMETRY

Bad geometry[3] in inference problems comes in many guises, and intuition gets progressively less clear in high dimension. Machine learnt neural mappings offer us a new tool to approach this.

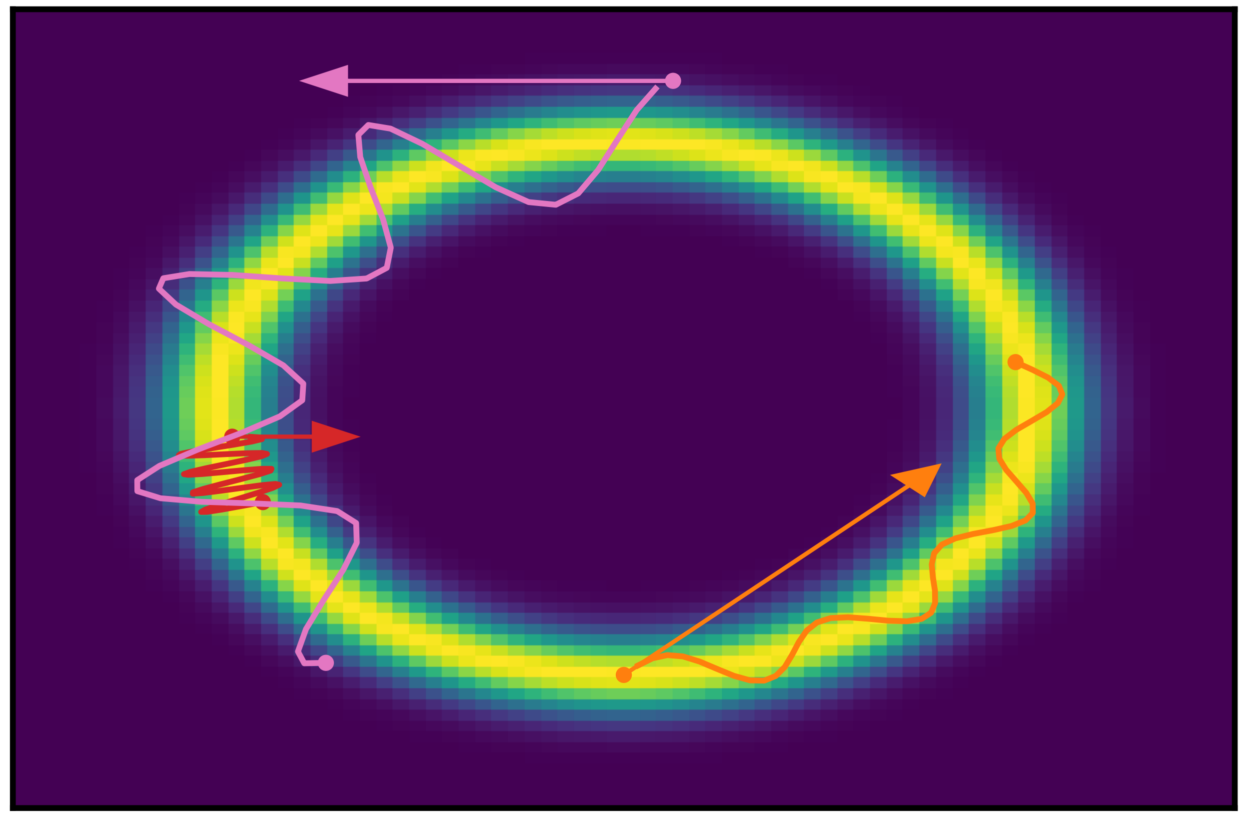
In the context of bridging distributions see `pocoMC`[4], `nessai`[5]



Whitening transforms to regularize the metric.



Clustering/ensembling to deal with multimodalities.



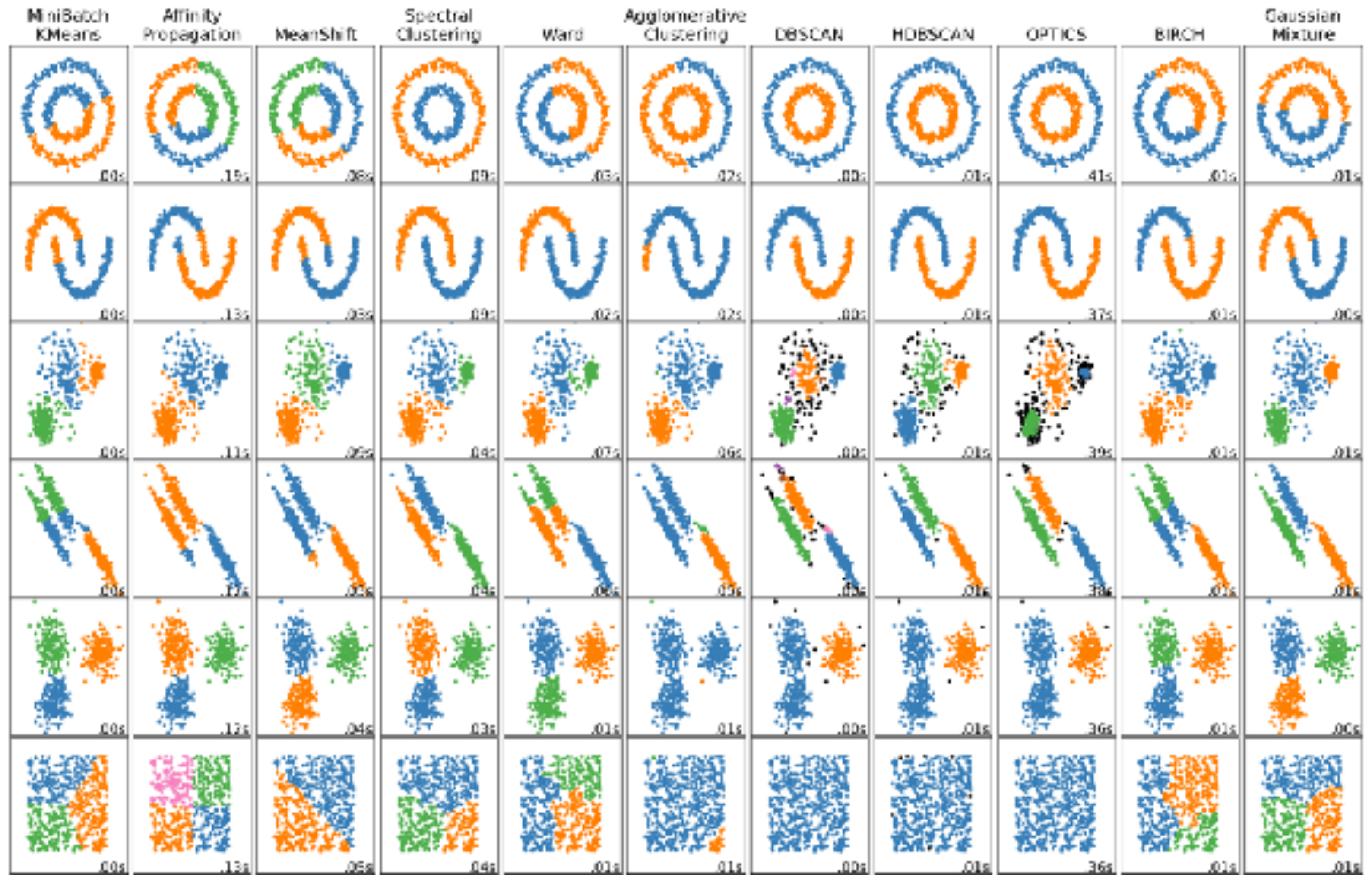
Gradients efficiently explore sweeping degeneracies.

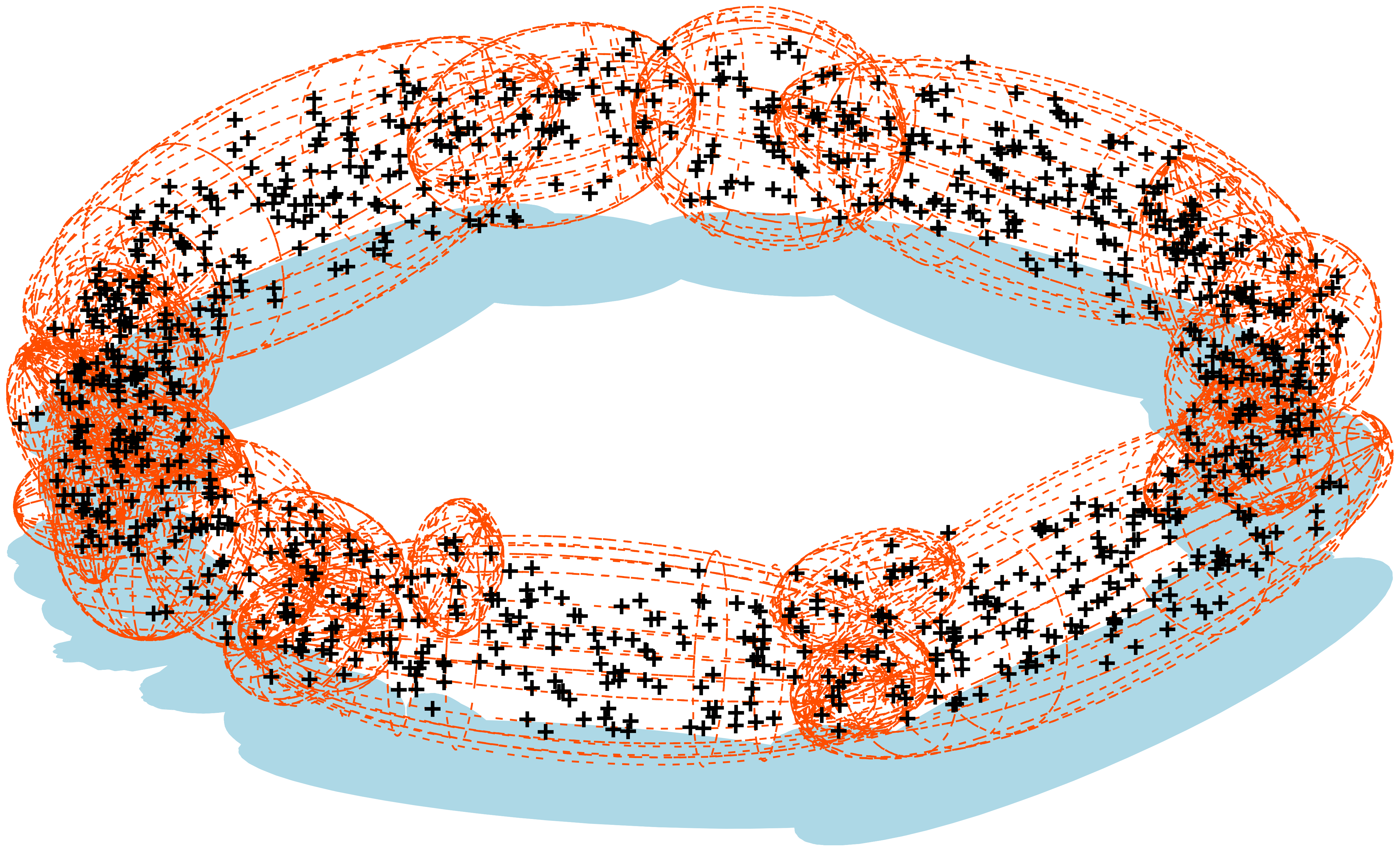
GEOMETRY

Complex geometry represents a hidden source of exponential scaling with dimension.

How many clustering algorithms even work in 1000D?

Mostly rely on dimension reduction techniques to work

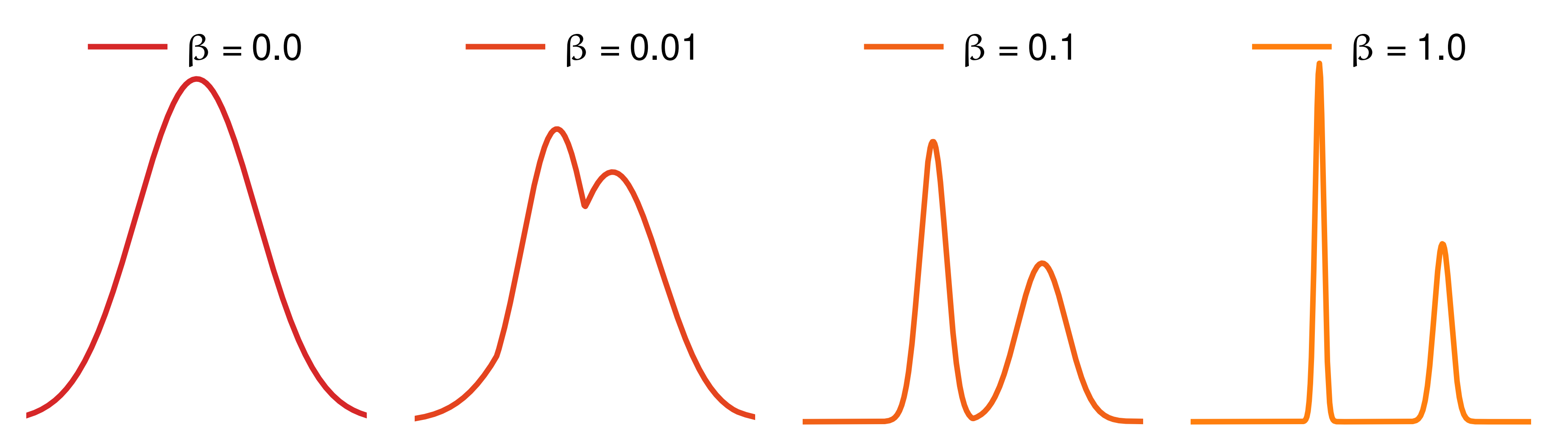




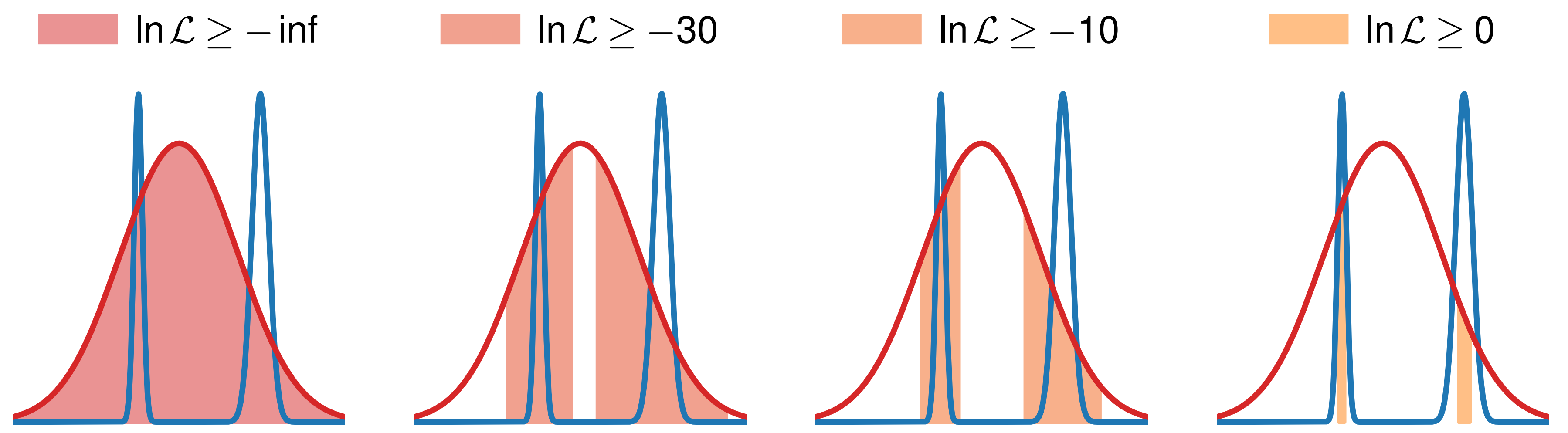
MultiNest clustering

BRIDGING DISTRIBUTIONS

Population Monte Carlo methods — particle filters — form bridges from known (prior) to complex unknown (posterior) distributions. Sequential Monte Carlo (SMC) and Nested Sampling (NS) are two variants evolving populations of points[6]. Both give us access to the normalizing constant Z.

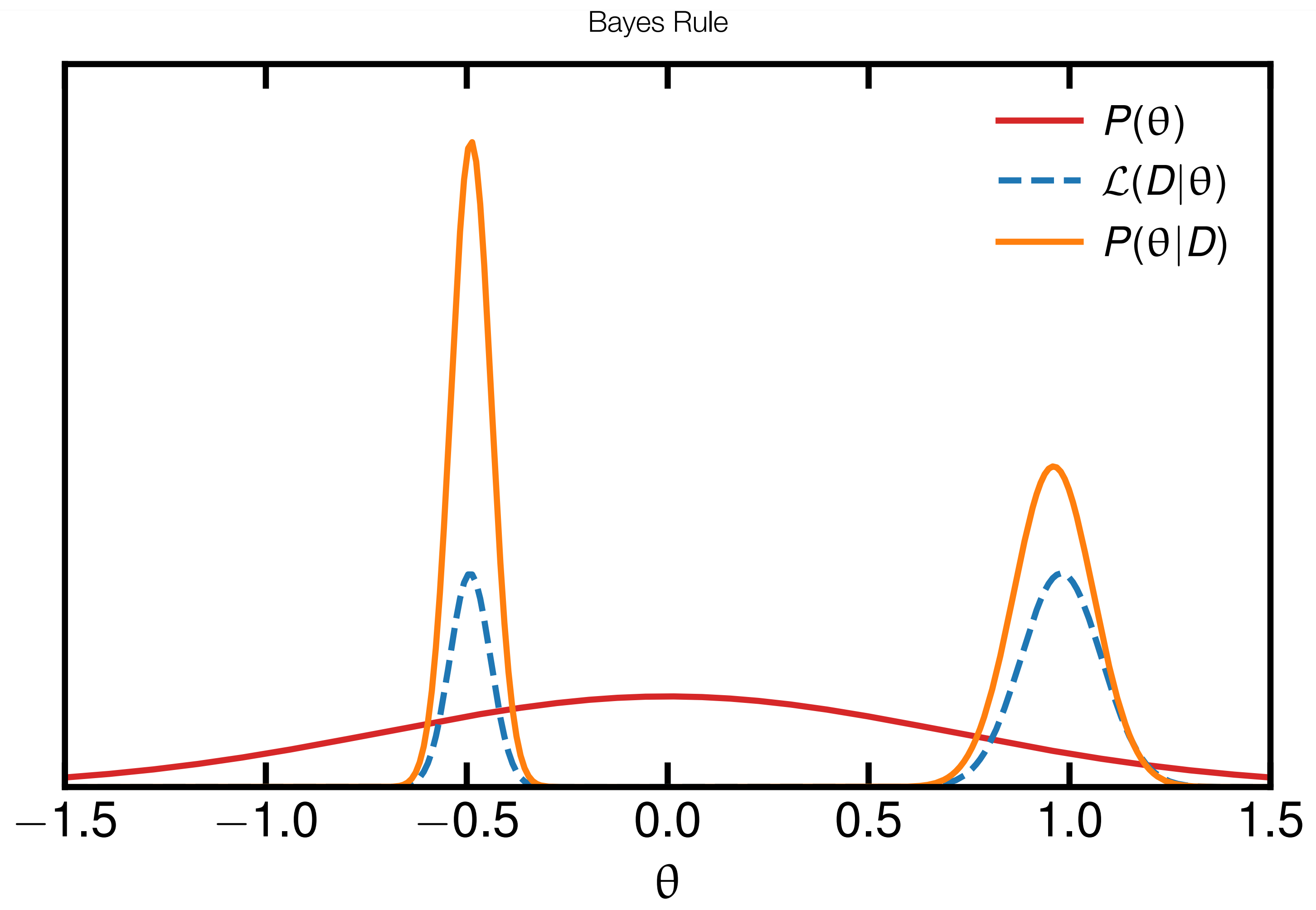


SMC

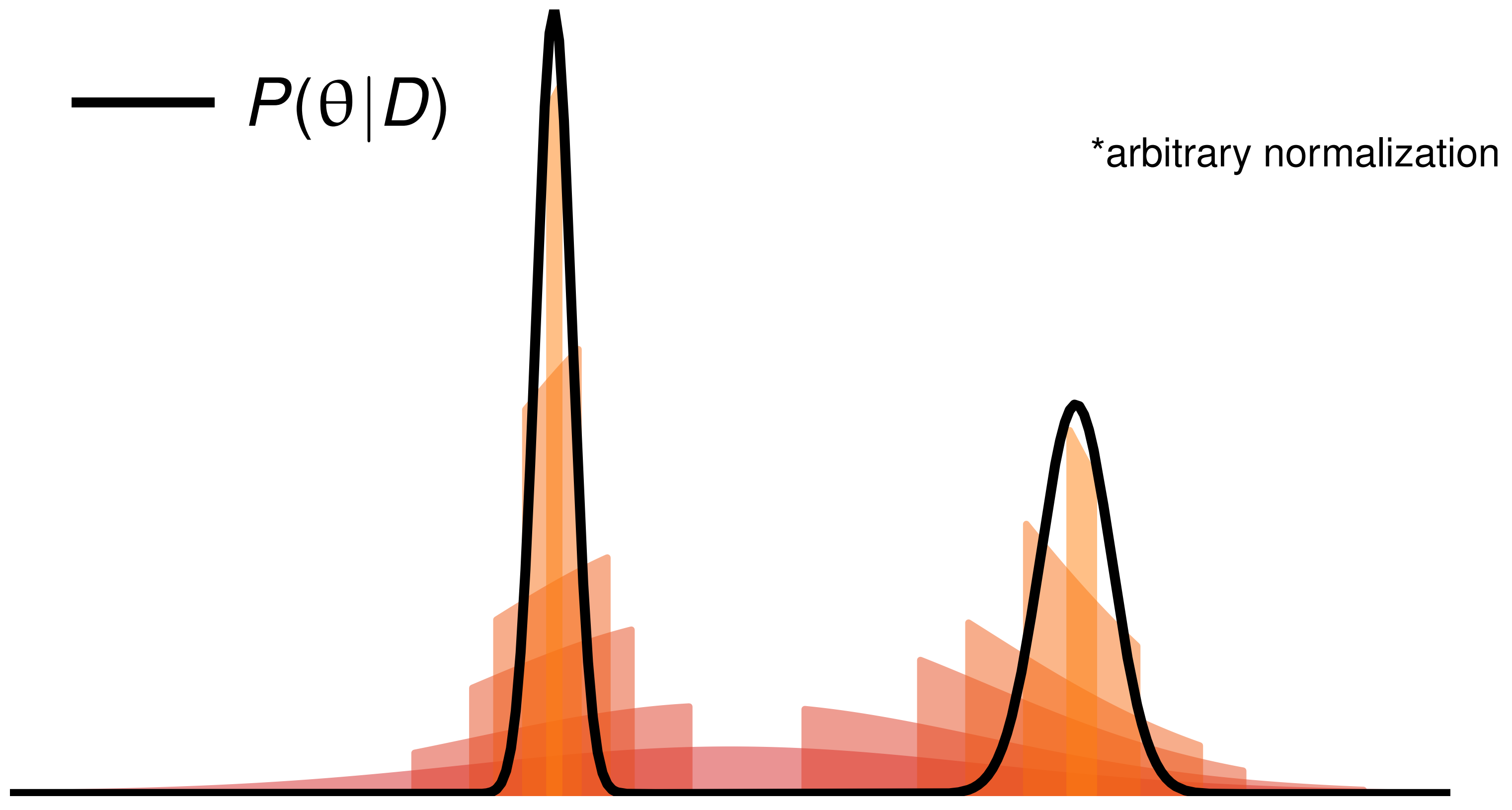


NS

$$P(\theta|D) = \frac{\mathscr{L}(D|\theta)P(\theta)}{Z}$$

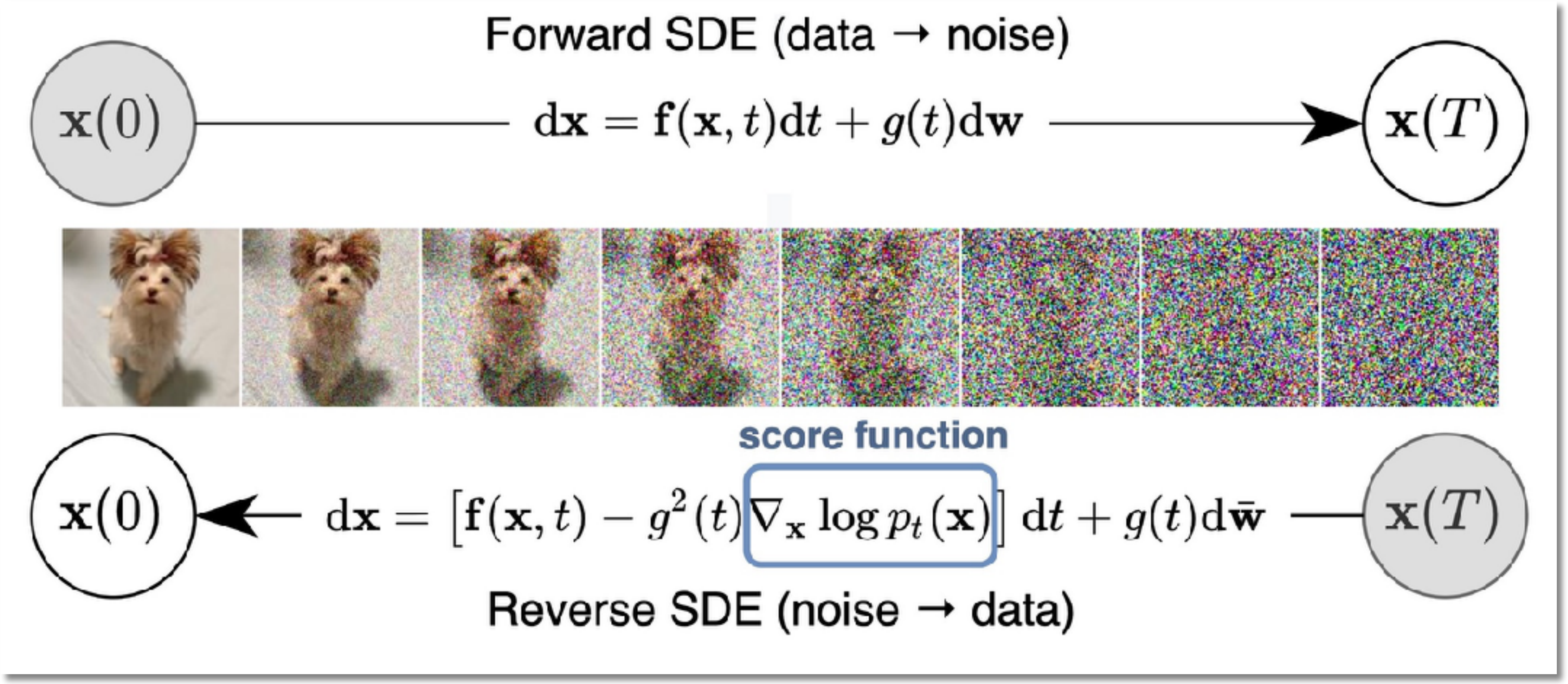


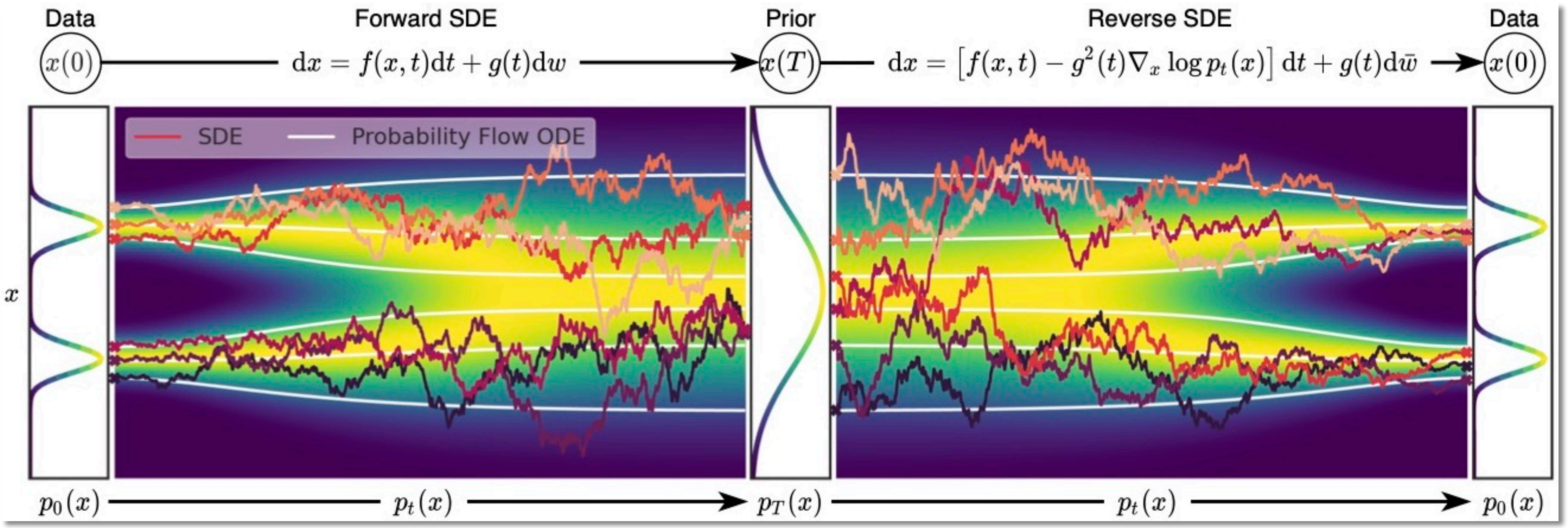
Toy 1D inference problem.



Schematic reconstructing posterior from NS shells.

DIFFUSION



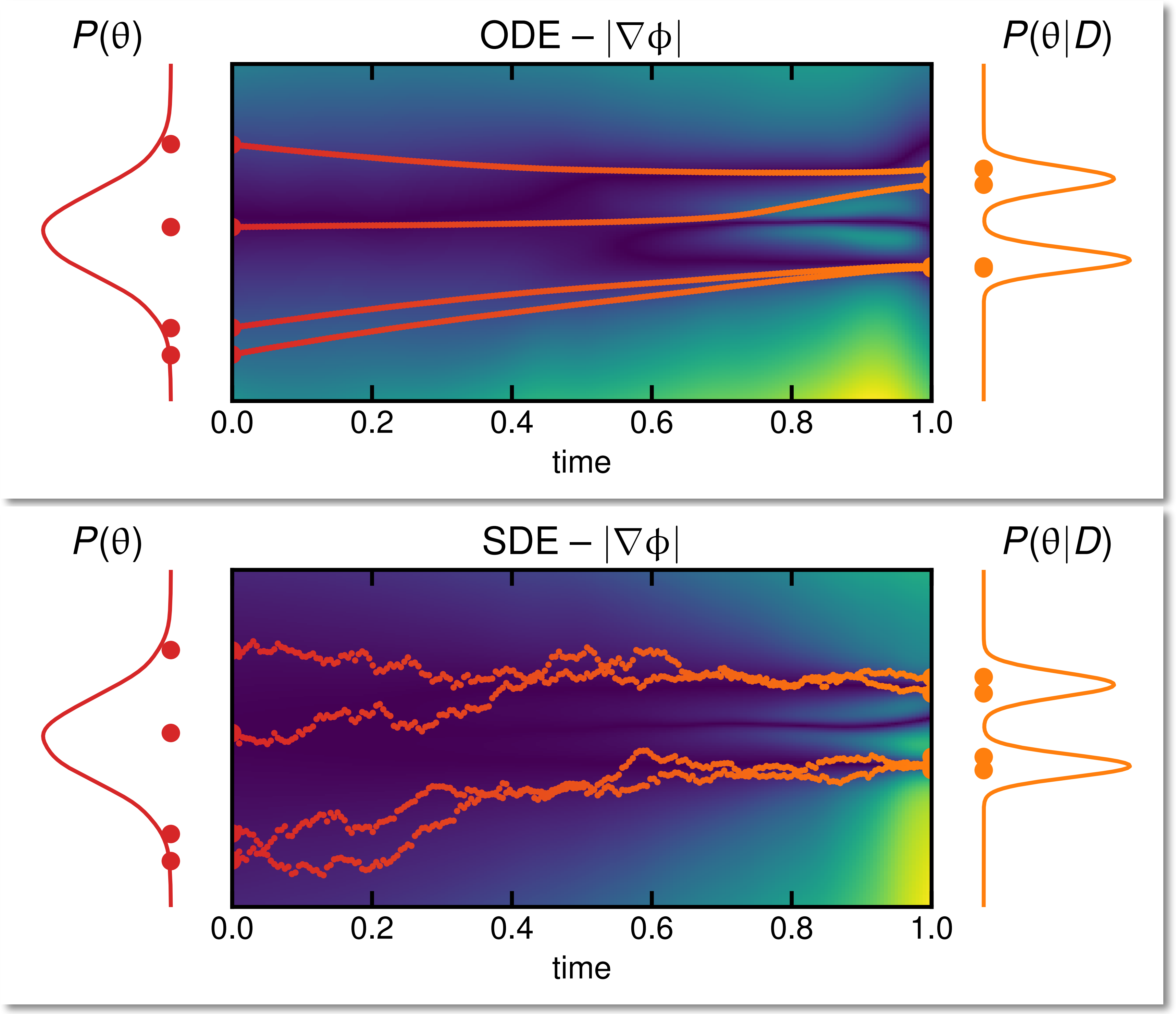


Yang Song - https://yang-song.net/blog/2021/score/

DIFFUSION

Diffusion models learn the gradient of the implicit density of a point cloud. Solving evolution through this field with Stochastic Differential Equation (SDE) or Ordinary Differential Equation (ODE) solvers yields Diffusion[7] or Continuous flows[8].

Neural learnt maps can transport any known distribution to an implicit target, no strict requirement on latent/prior!



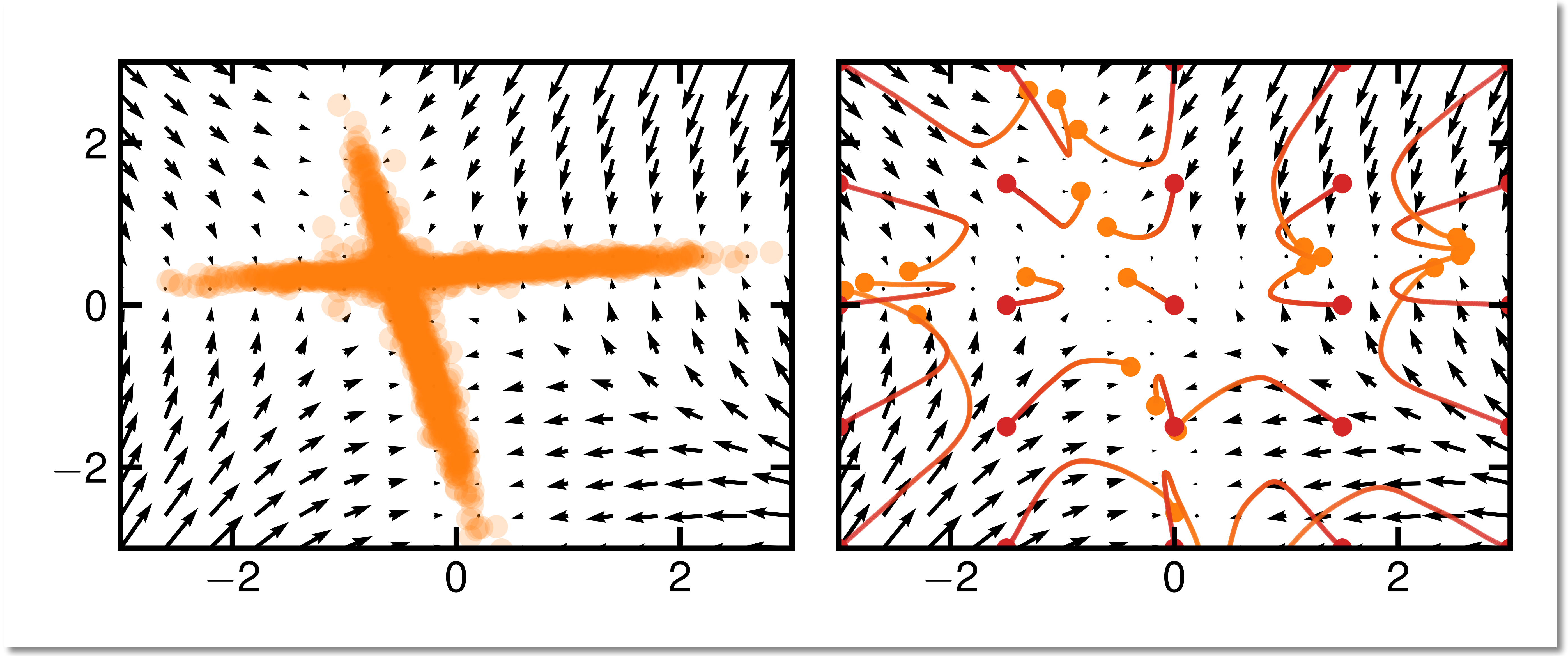
Learnt gradient vector field mapping prior to posterior.

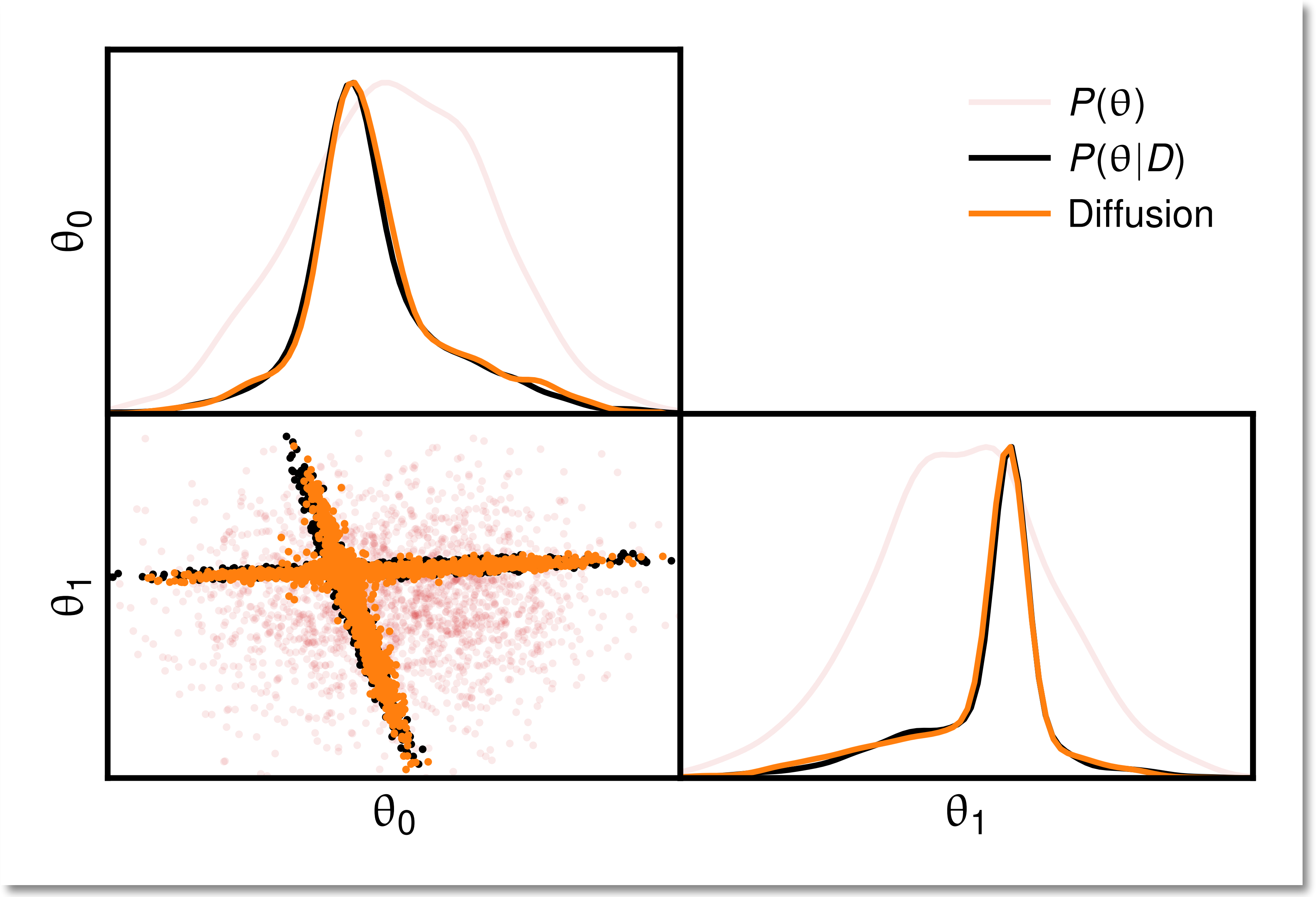
Discretize in time steps and iteratively add noise to solve as SDE

$$\frac{dy}{dt} = \nabla_\theta \phi(t, y(t)), \quad y(0) = y_0$$

Integrate to get continuous flow ODE solutions

DIFFUSION





2D representation of learnt vector fields.

My “images”:

$[x_1^1, x_2^1]$

$[x_1^2, x_2^2]$

$[\ldots]$

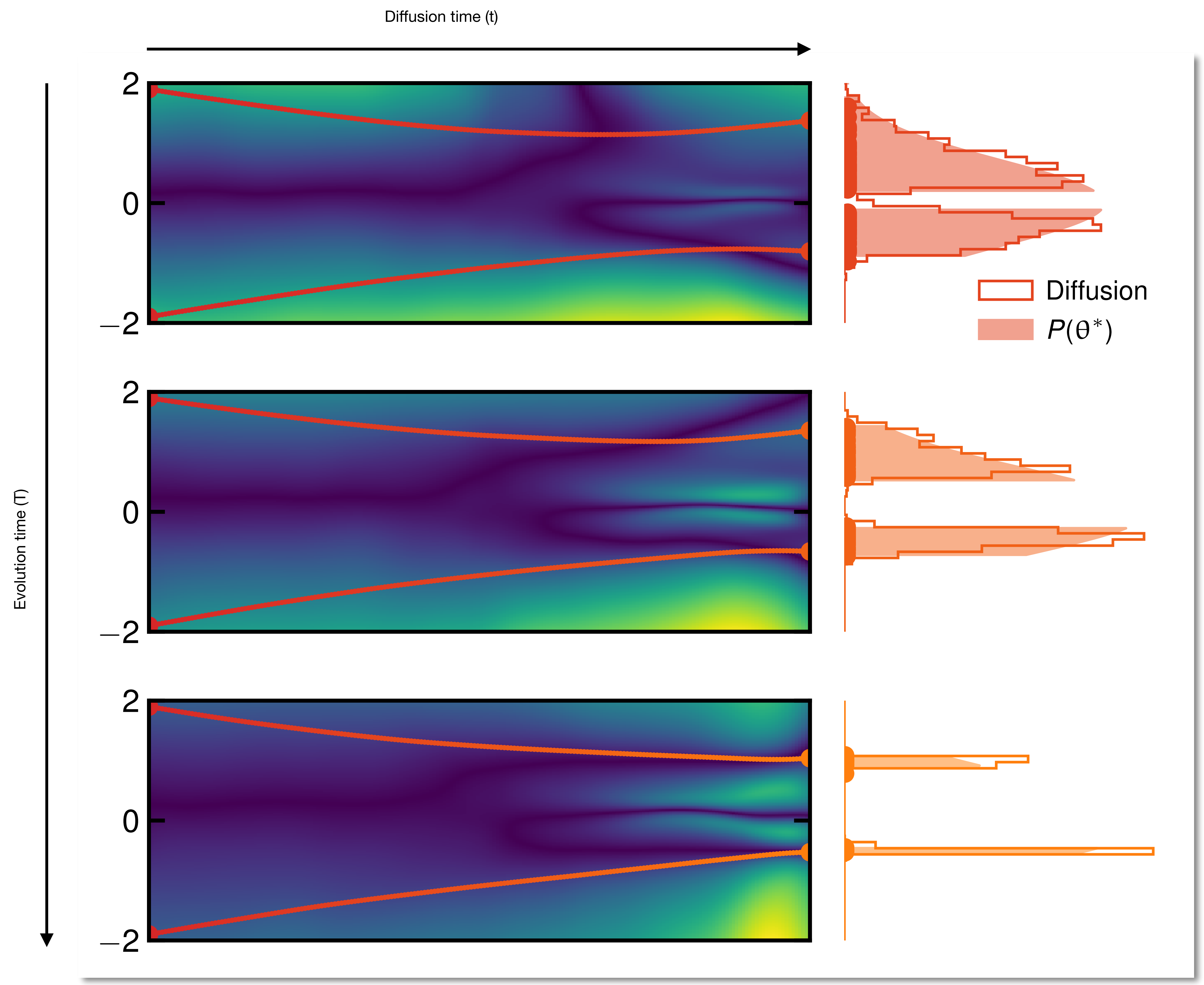
$[x_1^n, x_2^n]$

Which are the locations of my friends on the water:

Generate me a new “image" that looks like an existing “image”
This is what diffusion models do!

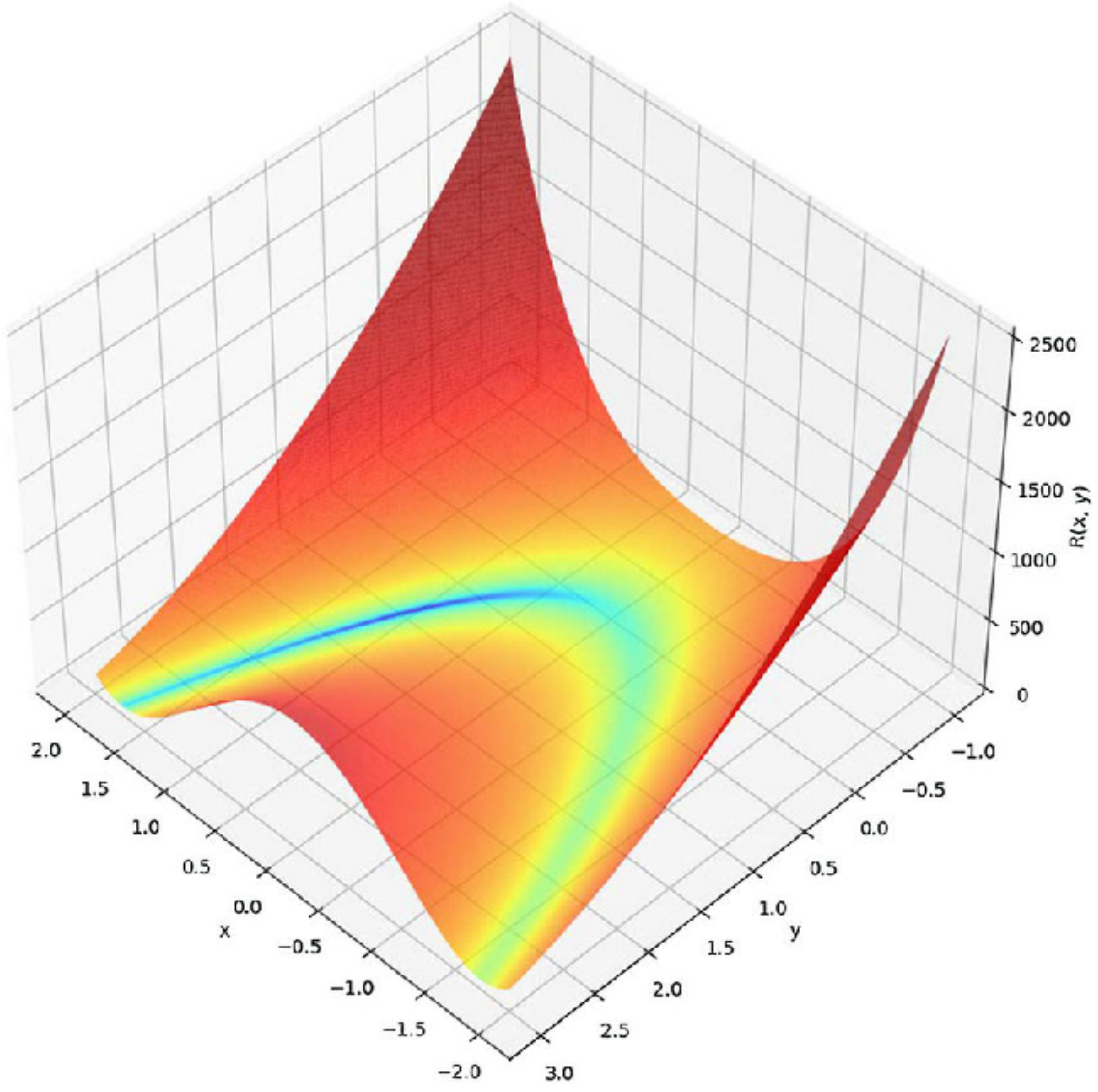
RESULTS

Diffusion models introduce time axis to the problem, bridging algorithms have another time axis we can efficiently evolve by fine tuning the score estimate.

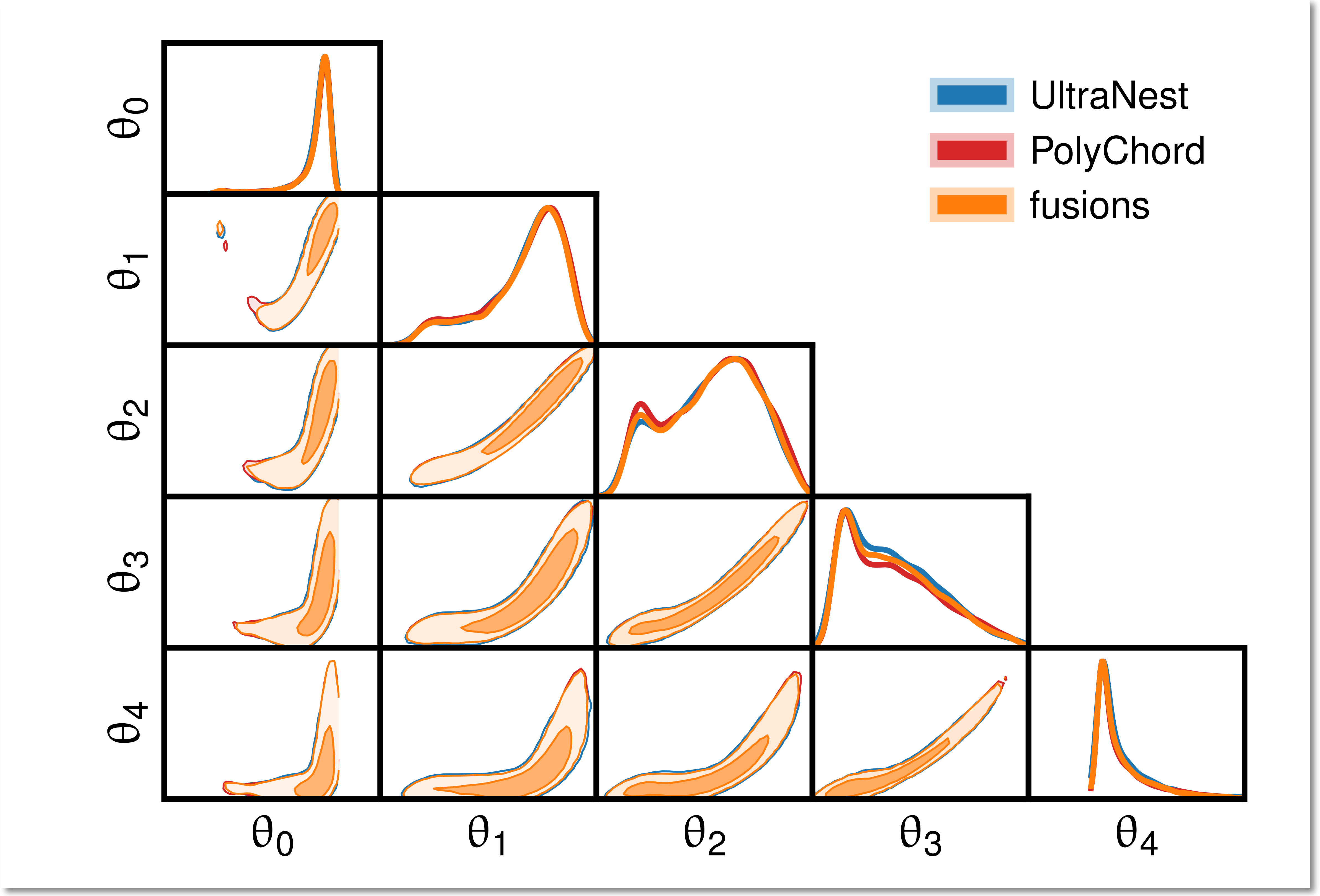


Evolution of likelihood constrained prior through a NS run.

RESULTS

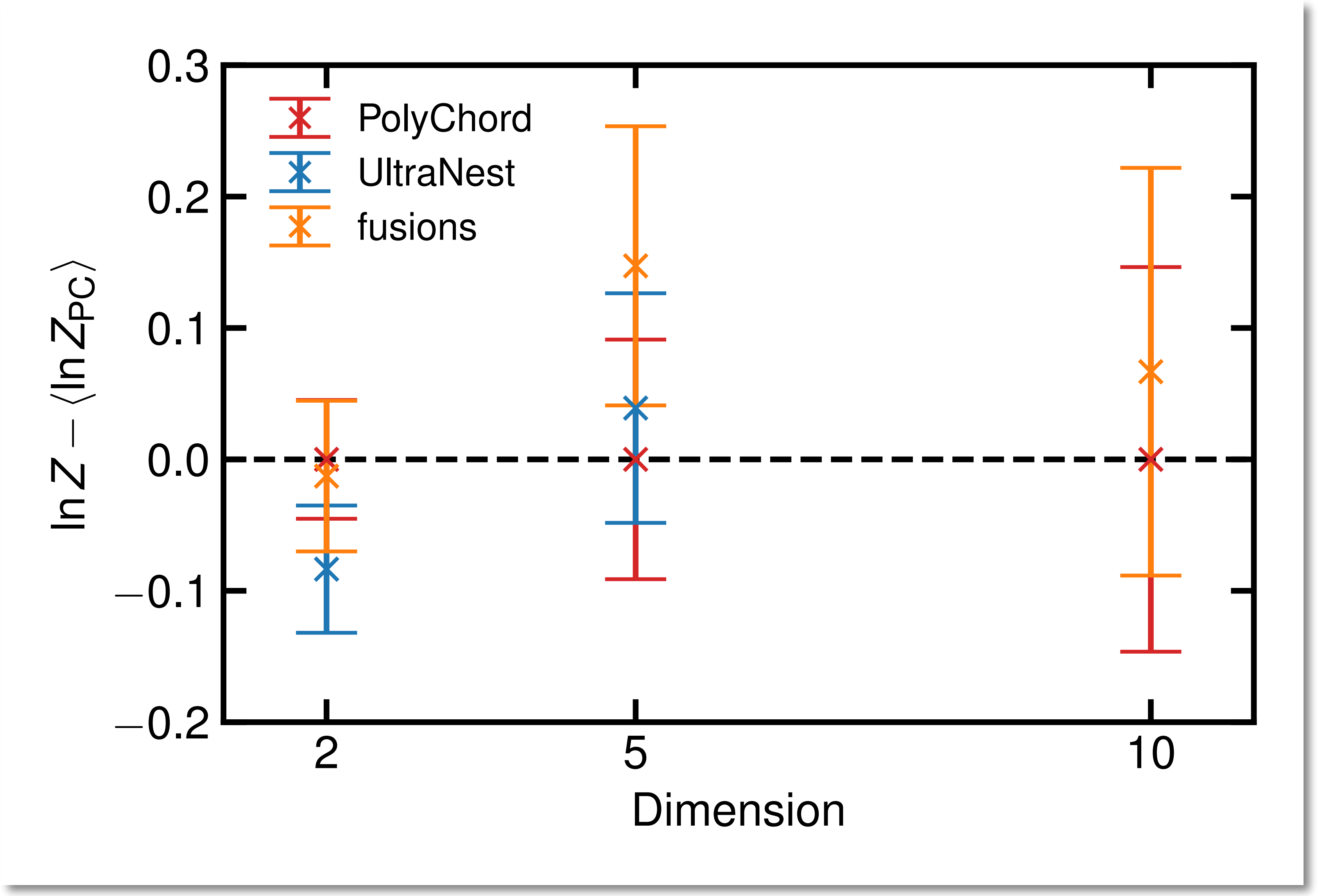


Rosenbrock function in 2D

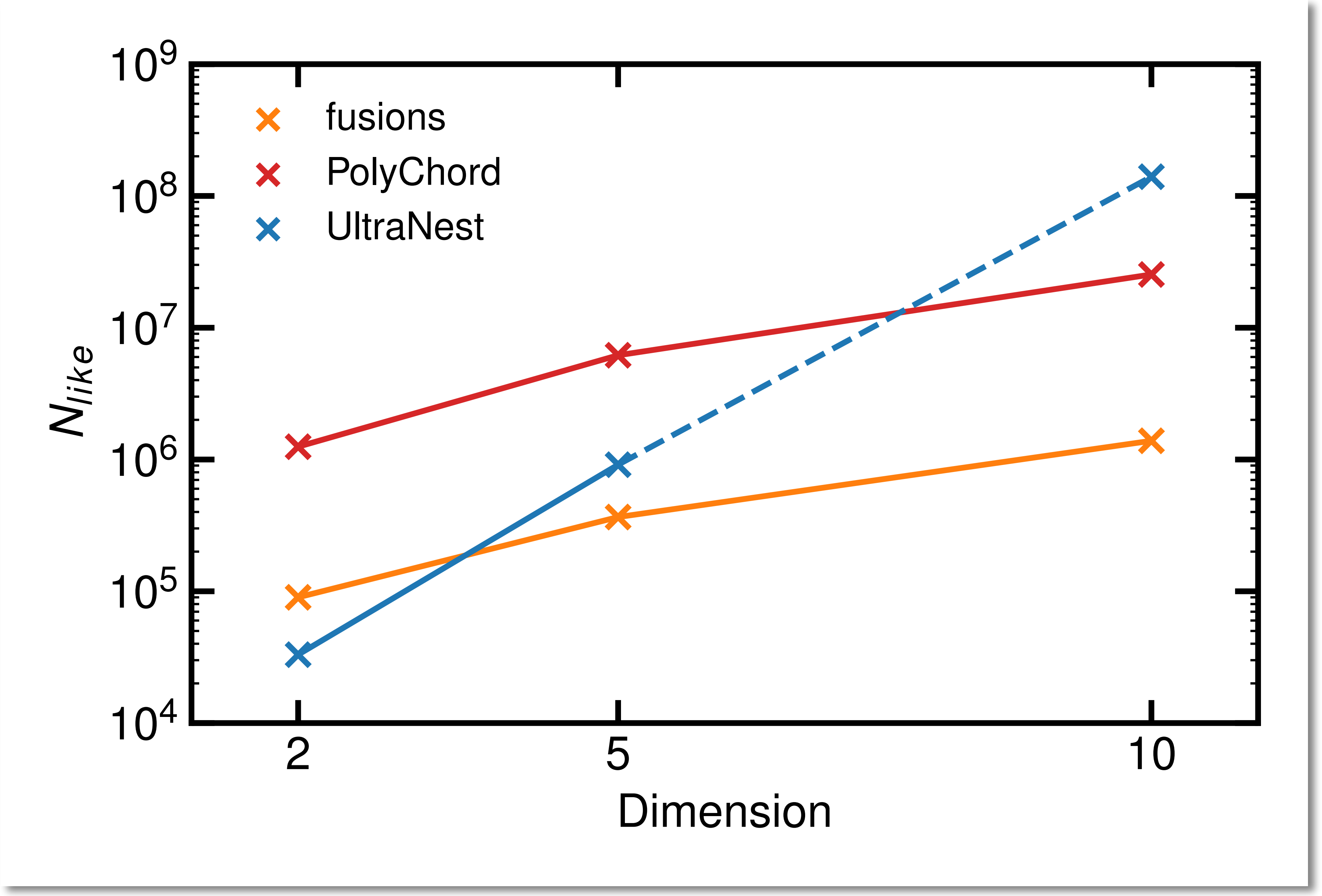


Corner plot of 5D Rosenbrock function

RESULTS



Calculated log integral for Rosenbrock function in various dimensions



Number of function evaluations required for Rosenbrock function in various dimensions

$N_{\rm live} = 2000$ for all, otherwise following defaults.

UltraNest in 10D $N_{\rm like}$ projected after early termination due to exceeding walltime.

Comparison to standard (non-neural) tools[9,10] shows promising scaling, comparable to step samplers despite using rejection sampling, whilst maintaining accurate predictions on benchmark challenging problems.

Algorithm demonstrated uses *zero* classical methods, treating the geometry of the problem solely with neural networks and score based models.

Work in progress, comparison to other neural methods[4,5,11,12], plenty left on the table to tune in the algorithm.

DIFFUSION MEETS NESTED SAMPLING

NEUTRALISING BAD GEOMETRY IN BRIDGING INFERENCE PROBLEMS

DAVID YALLUP

yallup/fusions

dy297@cam.ac.uk

yallup@github.io

References:

1.[2303.09082] The Gambit collaboration
2.[2404.03002] DESI Collaboration
3.[1903.03704] Hoffman et al.
4.[2207.05660] Karamanis et al.
5.[2102.11056] Williams et al.
6.[2205.15570] Ashton et al.
7.[2011.13456] Song et al.
8.[2302.00482] Tong et al.
9.[2101.09604] Buchner
10.[1506.00171] Handley et al.
11.[2306.16923] Lange
12.[1903.10860] Moss
13.[0809.3437] Feroz et al.

Technical references:

- github.com/patrick-kidger/diffrax
- github.com/handley-lab/anesthetic
- github.com/yallup/fusions
- github.com/google/flax
- github.com/google/jax

If you are interested in Diffusion models and applications in science, I am too! Come speak to me

Still plenty of work to be done on this project, tuning, testing before moving onto the next phase.