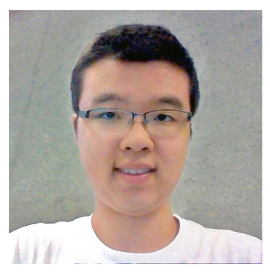
Normalizing Constant Estimation with Optimal Bridge Sampling and Normalizing Flows

He Jia[1,2] Uroš Seljak[2,3]

[1]Physics Department, Peking University [2]Physics Department, UC Berkeley [3]LBNL Berkeley

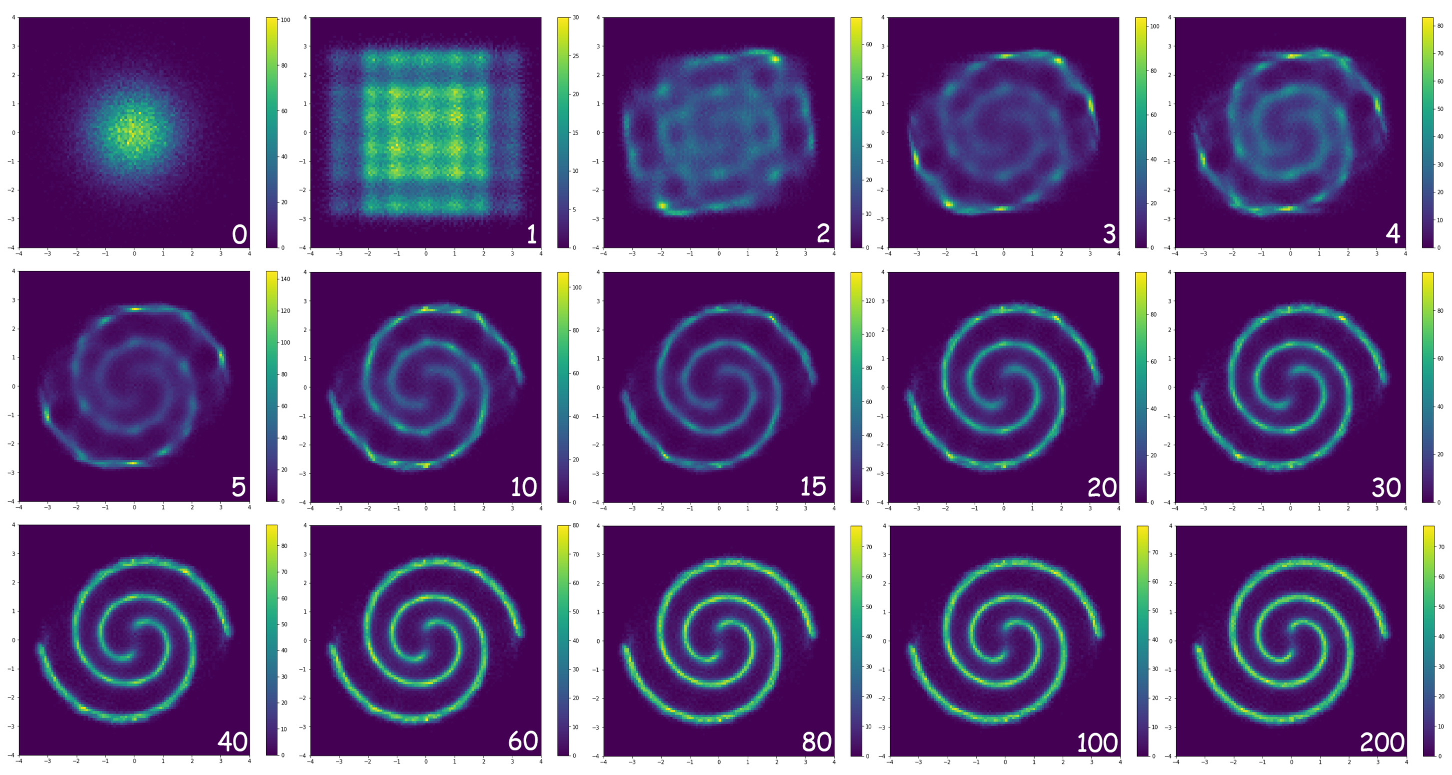


Overview

- Normalizing constant (Bayesian evidence) is one of the central goals of Bayesian analysis, yet its evaluation is tricky in high dimensions.
- We present Gaussianized Bridge Sampling (GBS), which combines Optimal Bridge Sampling (OBS) and Normalizing Flows (NF).
- We develop Fast Optimal Transport (FastOT), a novel NF based on greedy optimization, iteratively finding the most interesting directions along which we apply 1-d transforms.
- We leverage OBS with the proposal given by FastOT, and an adaptive sample allocation strategy allowing users to prioritize between posterior sampling or evidence estimation.

FastOT Normalizing Flows

- We exploit a simplified version of the general FastOT approach, details of which will be presented elsewhere.
- In each iteration, we
  1. use Wasserstein-1 distance to find the most non-Gaussian directions of the data,
  2. apply 1-d spline-based transforms to Gaussianize the data, which match the KDE-smoothed 1-d CDF of the data to standard Gaussian CDF.
- For evidence estimation, usually 5-10 iterations are sufficient.
- Example: Density estimation for the 2-d spiral using FastOT.



Optimal Bridge Sampling

- Based on the following identity, which requires n_p samples from the target distribution $p(\boldsymbol{x})$ and n_q samples from the proposal distribution $q(\boldsymbol{x})$, where $\alpha(\boldsymbol{x})$ is the bridge function.

$$\int_\Omega \alpha(\mathbf{x})p(\mathbf{x})q(\mathbf{x})d\mathbf{x} = \mathcal{Z}_p\left\langle\alpha(\mathbf{x})q(\mathbf{x})\right\rangle_p = \mathcal{Z}_q\left\langle\alpha(\mathbf{x})p(\mathbf{x})\right\rangle_q$$

- Optimal bridge function $\alpha(\boldsymbol{x})$ can be constructed, such that the ratio $r = \mathcal{Z}_p/\mathcal{Z}_q$ is given by the root of the following score function $S(r)$, which asymptotically minimizes the estimate error.

$$S(r)=\sum_{i=1}^{n_p}\frac{n_q r q(\mathbf{x}_{p,i})}{n_p p(\mathbf{x}_{p,i})+n_q r q(\mathbf{x}_{p,i})}-\sum_{i=1}^{n_q}\frac{n_p p(\mathbf{x}_{q,i})}{n_p p(\mathbf{x}_{q,i})+n_q r q(\mathbf{x}_{q,i})}=0$$

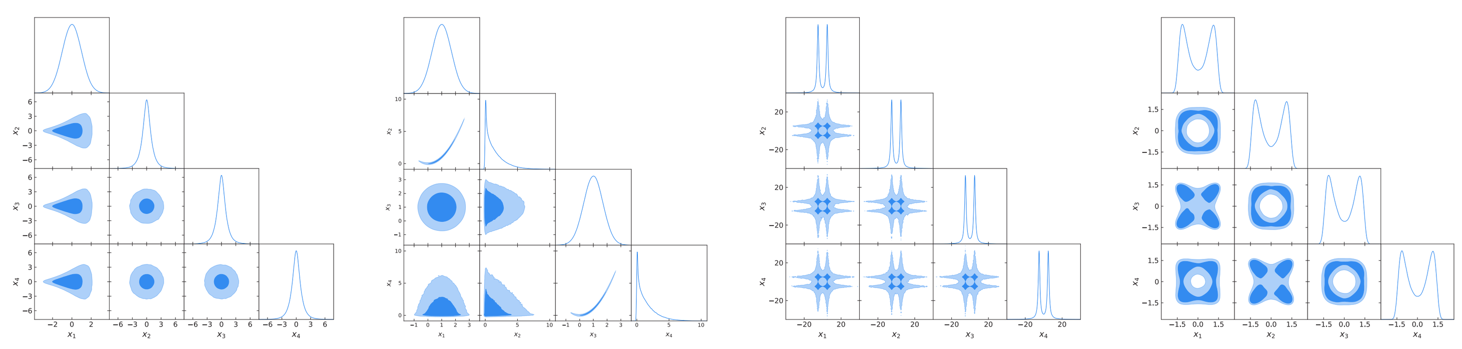
- The relative mean-square error of OBS can be estimated by a sum of two terms, which are proportional to $1/n_p$ and $1/n_q$ respectively. Given n_p samples from $p(\boldsymbol{x})$, we decide n_q such that the fraction of $q(\boldsymbol{x})$ contributions in the estimate error is no larger than 10% by default.

Gaussianized Bridge Sampling

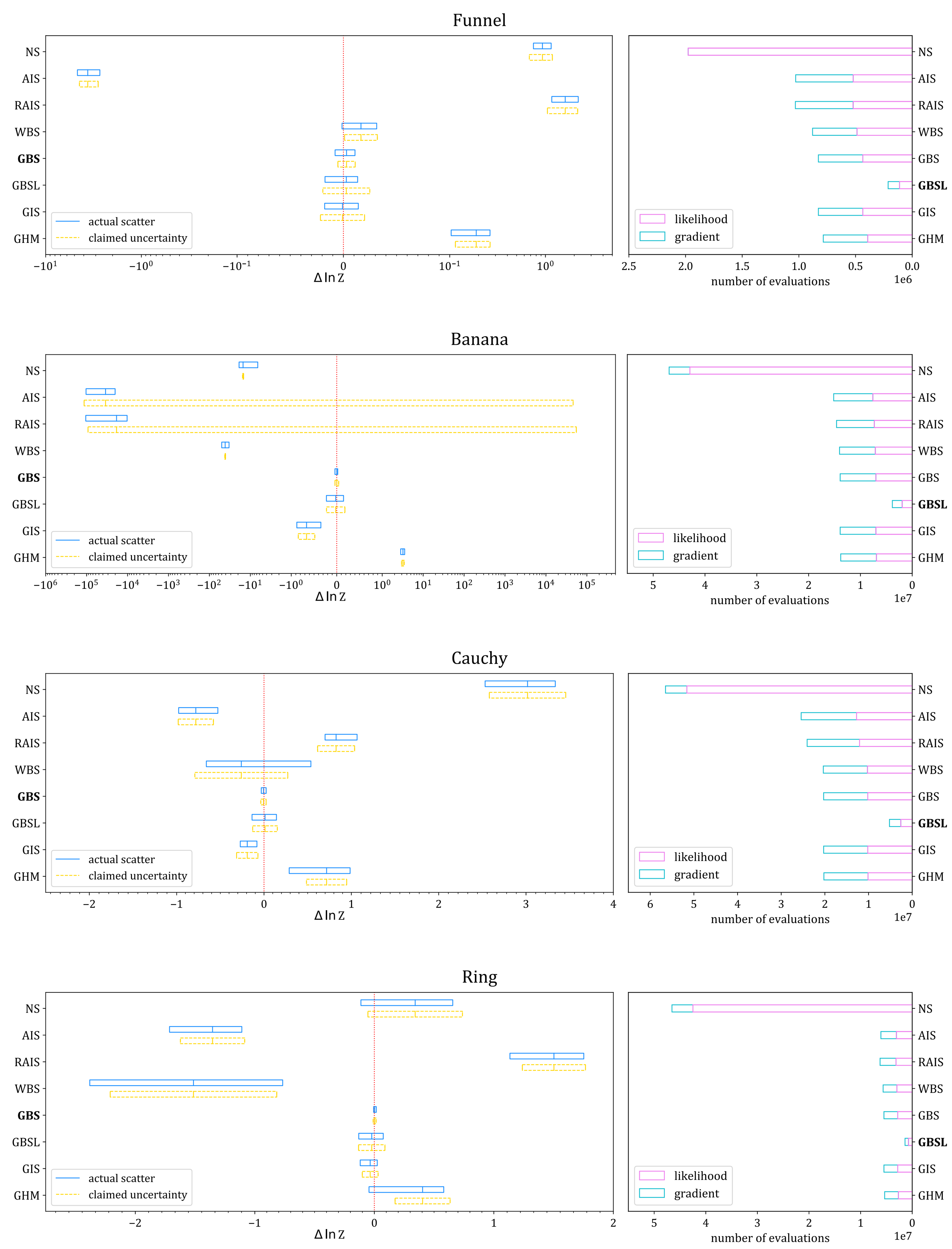
- Given the (unnormalized) target distribution $p(\boldsymbol{x})$ and n_p samples drawn from it (usually using MCMC), we
  1. train FastOT using the n_p samples to obtain the proposal $q(\boldsymbol{x})$,
  2. draw n_q samples from $q(\boldsymbol{x})$ and estimate the normalizing constant of $p(\boldsymbol{x})$ with OBS, while n_q is determined adaptively.
- The Python implementation will be available in the BayesFast package.

Experiments

- We compare GBS (our proposed method) and GBSL (GBS at 25% of computational cost) to the existing methods such as Annealed Importance Sampling (AIS), Reversed AIS (RAIS) and Nested Sampling (NS) on several challenging distributions in 16-64 dimensions.
- In these comparisons, the proposed method is at least 2-3 orders of magnitude faster at equal accuracy. When compared at equal computational time, the method can be several orders of magnitude more accurate. In contrast to existing methods, the method delivers a reliable error estimate.



16-d Funnel 32-d Banana 48-d Cauchy 64-d Ring



Abbreviation Keys: NS, Nested Sampling; **AIS**, Annealed Importance Sampling; **RAIS**, Reversed Annealed Importance Sampling; **WBS**, Warp Bridge Sampling; **GBS**, Gaussianized Bridge Sampling; **GBSL**, Gaussianized Bridge Sampling Lite; **GIS**, Gaussianized Importance Sampling; **GHM**, Gaussianized Harmonic Mean.