Normalizing flows and sequential Bayesian analysis

Hendrik Roch
WSU
17.09.2025

In collaboration with Chun Shen

Learning Informed Prior Distributions with Normalizing Flows for Bayesian Analysis:
Work in progress [HR, Shen]

Learning Informed Prior Distributions with Normalizing Flows for Bayesian Analysis

Hendrik Roch[1,*] and Chun Shen[1,†]

[1] *Department of Physics and Astronomy, Wayne State University, Detroit, Michigan 48201, USA*

We investigate the use of normalizing flow (NF) models as flexible priors in Bayesian inference with Markov Chain Monte Carlo (MCMC) sampling. Trained on posteriors from previous analyses, these models can be used as informative priors, capturing non-trivial distributions and correlations, in subsequent inference tasks. We compare different training strategies and loss functions, finding that training based on Kullback–Leibler (KL) divergence and unsupervised learning consistently yield the most accurate reproductions of reference distributions. Applied in sequential Bayesian workflows, MCMC with the NF-based priors reproduces the results of one-shot joint inferences well, provided the target distributions are unimodal. In cases with pronounced multi-modality or dataset tension, distortions may arise, underscoring the need for caution in multi-stage Bayesian inference. A comparison between the `pocoMC` MCMC sampler and the standard `emcee` sampler further demonstrates the importance of advanced and robust algorithms for exploring the posterior space. Overall, our results establish NF-based priors as a practical and efficient tool for sequential Bayesian inference in high-dimensional parameter spaces.

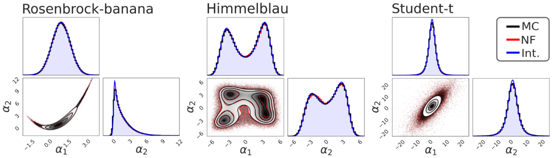
BUQ Meeting

WAYNE STATE UNIVERSITY

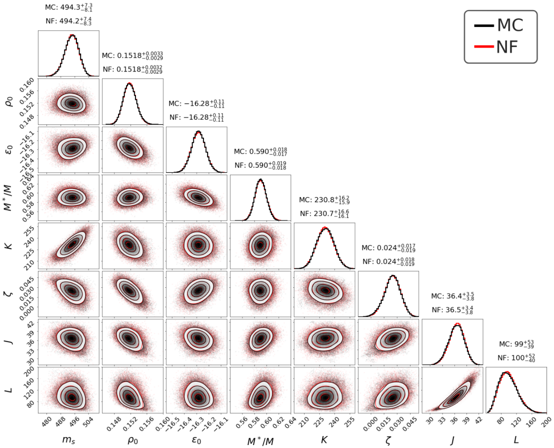
Motivation

- Choice of prior distribution is critical for Bayesian inference:
  - Uniform or Gaussian priors are simple but limited
  - Informative priors improve efficiency and enable sequential analyses
- Directly reusing posteriors as priors is challenging due to:
  - Multimodality and non-Gaussian shapes
  - Non-trivial parameter correlations
  - Inefficient sampling in high dimensions
- Normalizing flows (NFs) offer a flexible generative approach:
  - Capture complex distributions and correlations
  - Provide efficient continuous sampling
  - Enable informative priors for sequential Bayesian analysis
- We extend NF-based priors from 2310.04635 to:
  - Unsupervised learning cases without explicit posterior densities
  - Sequential Bayesian inference in high-energy nuclear physics
  - Systematic tests of posterior consistency

Applications in 2310.04635



- NF seems to be successfull for "simple" distributions peaked not too close to the prior boundary
- What happens when there is tension in the data (multiple modes) or peaks at the boundary?

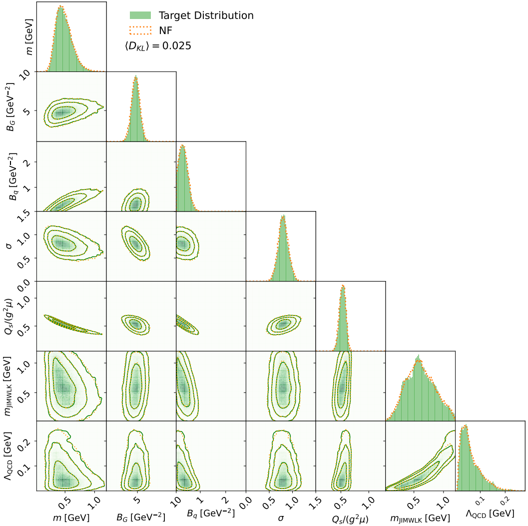


NF Training

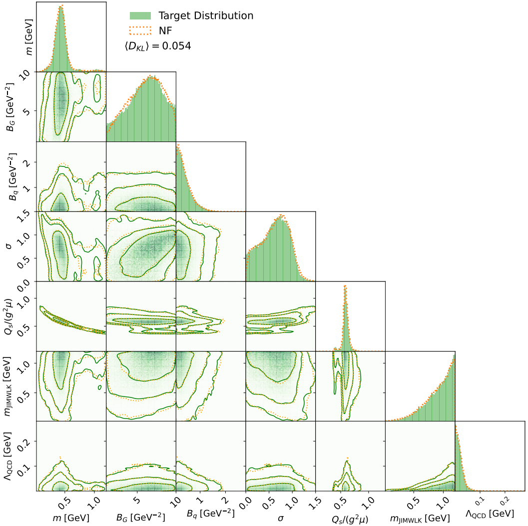
- Batch sizes $\{500, 1000, 2000, 5000\}$, coupling layers $\mathcal{N} \in \{2, 6, 8, 10, 12\}$, learning rates $\{1 \times 10^{-3}, 5 \times 10^{-4}, 1 \times 10^{-4}, 5 \times 10^{-5}\}$
- Select the NF model with the smallest average KL divergence
- We continue with the KL loss function

TABLE II. Best NF model configurations and corresponding average KL divergence $\langle D_{\mathrm{KL}} \rangle$ for the γ+p and γ+Pb datasets.

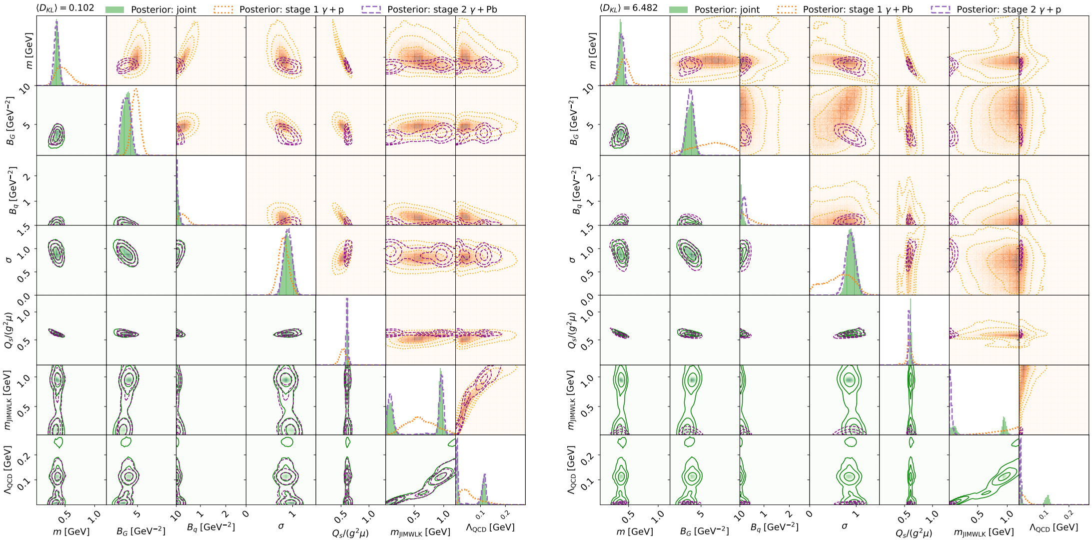
Dataset	Batch Size	Layers	Learning Rate	Loss	$\langle D_{\mathrm{KL}} \rangle$
$\gamma + p$	500	4	1×10^{-4}	Jeffreys'	0.047
$\gamma + p$	5000	6	1×10^{-3}	KL	0.025
$\gamma + p$	1000	10	1×10^{-3}	log-$\mathcal{L}$	0.024
$\gamma + \mathrm{Pb}$	500	12	1×10^{-3}	Jeffreys'	0.061
$\gamma + \mathrm{Pb}$	1000	12	1×10^{-3}	KL	0.052
$\gamma + \mathrm{Pb}$	2000	10	1×10^{-3}	log-$\mathcal{L}$	0.052



$\gamma + p$ dataset

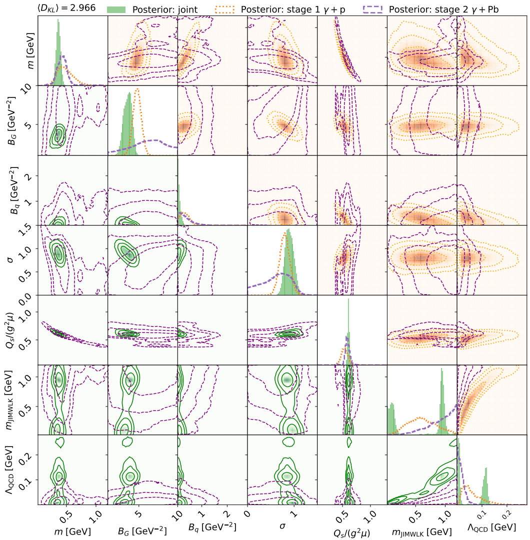


$\gamma + \text{Pb}$ dataset



$\langle D_{KL} \rangle = 0.102$
Posterior: joint
Posterior: stage 1 $\gamma + p$
Posterior: stage 2 $\gamma + Pb$
$\langle D_{KL} \rangle = 6.482$
Posterior: joint
Posterior: stage 1 $\gamma + Pb$
Posterior: stage 2 $\gamma + p$
m [GeV]
B_G [GeV^{-2}]
B_q [GeV^{-2}]
σ
$Q_s/(g^2\mu)$
m_{JIMWLK} [GeV]
Λ_{QCD} [GeV]

- Example using **`emcee`** (standard) sampler
- The quality of the MCMC sampler is crucial for sequential Bayesian inference!



Does this mean we can not do sequential Bayesian inference?

Does this mean we can not do sequential Bayesian inference?

NO!

We just have to be more careful where we apply it!

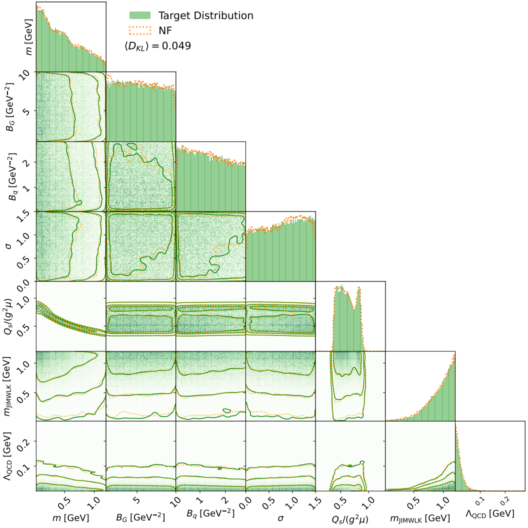
Let's look at a simpler example: divide $\gamma + \mathrm{Pb}$ dataset into integrated and differential cross sections $\rightarrow$ unimodal final posterior

NF Training

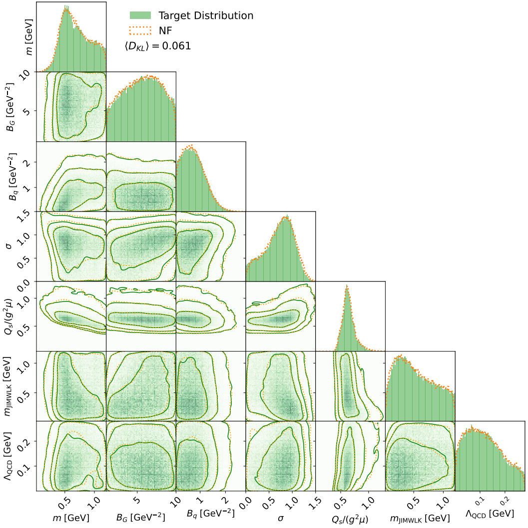
- Batch sizes $\{500, 1000, 2000, 5000\}$, coupling layers $\mathcal{N} \in \{2, 6, 8, 10, 12\}$, learning rates $\{1 \times 10^{-3}, 5 \times 10^{-4}, 1 \times 10^{-4}, 5 \times 10^{-5}\}$
- Select the NF model with the smallest average KL divergence
- We continue with the KL loss function

TABLE III. Best NF model configurations and corresponding average KL divergence $\langle D_{\mathrm{KL}} \rangle$ for the integrated and differential $\gamma + \mathrm{Pb}$ datasets.

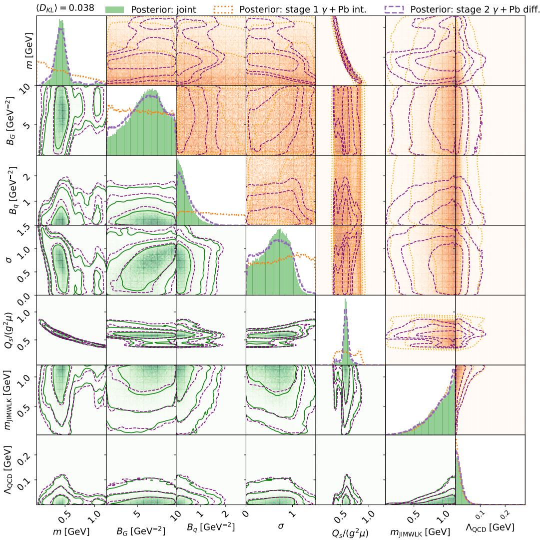
Dataset	Batch Size	Layers	Learning Rate	Loss	$\langle D_{\mathrm{KL}} \rangle$
int.	2000	12	1×10^{-3}	Jeffreys'	0.054
int.	5000	8	1×10^{-3}	KL	0.049
int.	5000	12	1×10^{-3}	log-$\mathcal{L}$	0.048
diff.	2000	12	1×10^{-3}	Jeffreys'	0.063
diff.	2000	10	1×10^{-3}	KL	0.061
diff.	5000	10	1×10^{-3}	log-$\mathcal{L}$	0.059



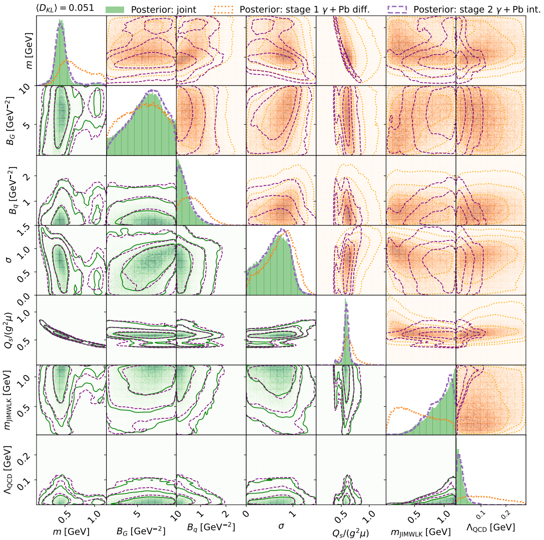
$\gamma + \mathrm{Pb}$ int. dataset



$\gamma + \mathrm{Pb}$ diff. dataset

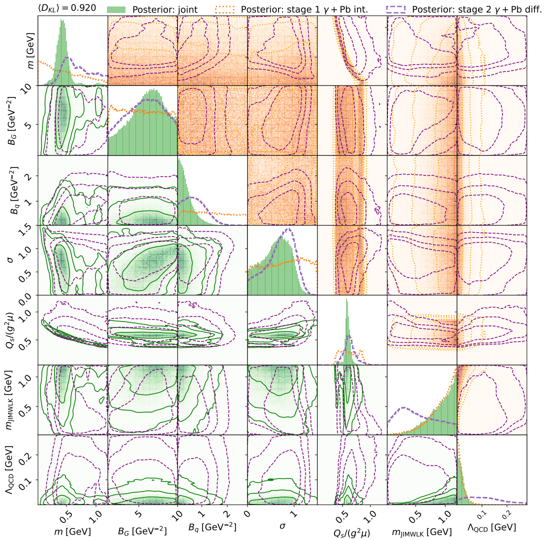


$\langle D_{KL} \rangle = 0.038$
Posterior: joint
Posterior: stage 1 γ + Pb int.
Posterior: stage 2 γ + Pb diff.
m [GeV]
B_G [GeV^{-2}]
B_q [GeV^{-2}]
σ
$Q_s/(g^2\mu)$
$m_{\rm JIMWLK}$ [GeV]
$\Lambda_{\rm QCD}$ [GeV]



$\langle D_{KL} \rangle = 0.051$
Posterior: joint
Posterior: stage 1 γ + Pb diff.
Posterior: stage 2 γ + Pb int.
m [GeV]
B_G [GeV^{-2}]
B_q [GeV^{-2}]
σ
$Q_s/(g^2\mu)$
$m_{\rm JIMWLK}$ [GeV]
$\Lambda_{\rm QCD}$ [GeV]

- Example using `emcee` (standard) sampler
- Also failing for a more "simple" distribution



Conclusion

- Normalizing flows provide powerful, flexible priors
- Capture complex posterior features and correlations beyond simple priors
- Enable efficient and consistent sequential analyses across datasets (if the first stage constraints are not too restrictive)
- Open path to scalable Bayesian studies in nuclear and particle physics using knowledge from earlier studies