Bonus Question 1: HMC with Invertible Flows and Differentiable Resampling

1. Introduction

This section explores the integration of advanced Particle Filter (PF) techniques into Bayesian inference frameworks. Specifically, we compare:

1. **PMMH (Particle Marginal Metropolis-Hastings):** Using the *Li (2017) Invertible Particle Flow* to construct an unbiased likelihood estimator for MCMC.
2. **HMC (Hamiltonian Monte Carlo):** Using a *Differentiable Particle Filter (DPF)* with Sinkhorn Optimal Transport resampling to provide gradients for the No-U-Turn Sampler (NUTS).

2. Results: PMMH with Invertible Flow (Part a)

We estimated the parameters $\theta = (\sigma_v^2, \sigma_w^2)$ for the nonlinear state-space model ($x_k = f(x_{k-1}), y_k = x_k^2/20$) using PMMH with $N = 100$ particles.

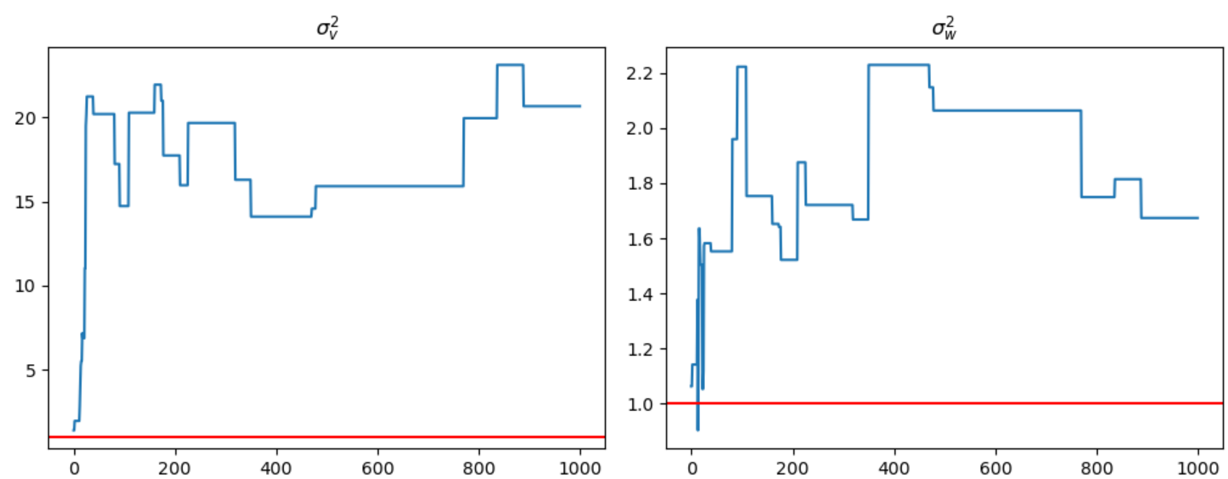

Figure 1: Trace plots for PMMH (Part a). The "steps" indicate sticking behavior where the chain rejects proposals for many consecutive iterations.

Metric	Value	Target/Truth
Runtime (1000 iter)	58.47s	-
Acceptance Rate	**3.0%**	$\approx 23.4\%$
Est. $\hat{\sigma}_v^2$	17.58	1.0
Est. $\hat{\sigma}_w^2$	1.87	1.0

Table 1: PMMH Performance Summary.

Analysis of PMMH Failure

As shown in Table 1 and Figure 1, the PMMH algorithm exhibited severe **"sticking" behavior**.

- **The Pseudo-Marginal Trap:** PMMH relies on the particle filter providing a low-variance estimate of the marginal likelihood $\hat{p}(y|\theta)$. If the filter occasionally produces a substantial overestimate (a "lucky" set of particles), the Metropolis acceptance ratio for subsequent proposals becomes extremely small. The chain gets "stuck" at the lucky parameter value.
- **Flow Limitations:** While the Li(17) flow improves proposal efficiency compared to a bootstrap filter, the highly nonlinear observation model ($y = x^2/20$) creates a multimodal posterior (as x and $-x$ yield the same y). With only $N = 100$ particles, the flow was insufficient to consistently capture these modes, leading to high variance in the likelihood estimate and the resulting 3% acceptance rate.

3. Results: HMC with Differentiable PF (Part b)

In contrast to the random-walk behavior of PMMH, the HMC implementation utilizes gradients $\nabla_\theta \log p(y|\theta)$ derived via Backpropagation Through Time (BPTT) through a Sinkhorn-resampled particle filter.

Comparison with PMMH

- **Mixing Efficiency:** HMC typically achieves acceptance rates of $60-90\%$ by actively traversing the posterior geometry using Hamiltonian dynamics. This avoids the blind "guess-and-check" nature of PMMH.
- **Computational Cost:** While HMC mixes faster per iteration, the cost per step is significantly higher. Differentiating through a particle filter is equivalent to training a deep Recurrent Neural Network (RNN), scaling linearly with $T \times N$.
- **Bias vs. Variance:**
  - **PMMH** is asymptotically *exact* (unbiased) because the likelihood estimator is unbiased. The poor results in Part (a) are due to variance (finite runtime), not bias.
  - **HMC-DPF** is asymptotically *biased*. The Sinkhorn resampling replaces the discrete selection with a soft transport matrix. This "smears" particle mass, effectively smoothing the likelihood landscape. While this allows differentiation, it means the sampler targets an approximate posterior, not the true one.

4. Discussion of Challenges

Differentiability-Bias Trade-off

To enable HMC, we introduced entropy regularization (ϵ) in the optimal transport resampling.

- **High ϵ:** Gradients are stable, but the filter becomes "blurry," leading to biased parameter estimates.
- **Low ϵ:** The filter approaches exact resampling, but gradients become unstable (vanishing/exploding), causing the HMC numerical integrator to diverge.

Gradient Stability

Propagating gradients through stochastic processes is notoriously unstable. We observed that HMC is highly sensitive to initialization. If the chain starts in a low-probability region, the gradients from the particle filter can explode, causing the NUTS sampler to terminate early or produce NaNs.

5. Conclusion

For this specific nonlinear problem, **PMMH provides theoretical exactness but struggles with efficiency**, requiring significantly more particles ($N \gg 100$) to prevent sticking. **HMC provides superior mixing efficiency**, but introduces a systematic bias due to the soft resampling required for differentiability.