Parallel tempering with a variational reference

THE UNIVERSITY OF BRITISH COLUMBIA

Nikola Surjanovic[1], Saifuddin Syed[2], Alexandre Bouchard-Côté[1], Trevor Campbell[1]

Summary

Parallel tempering (PT) is a class of MCMC algorithms that constructs a path of distributions annealing between a reference, π_0, and intractable target, π_1. States along the path are swapped to improve mixing in the target.

Problem: Past work on PT has only used linear paths with a **fixed reference** that is often different than the target. PT swapping is inefficient and does not improve much on MCMC.

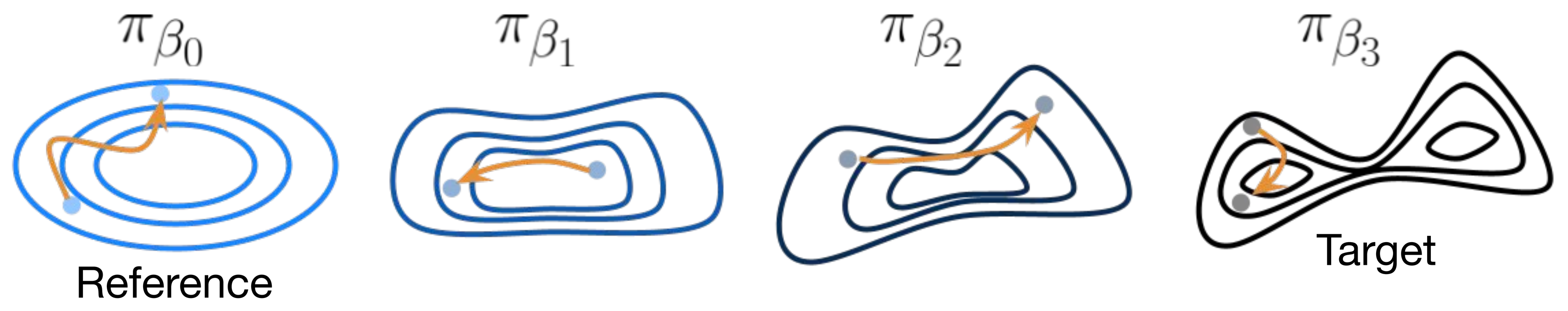
Contribution: We extend to annealing paths with a **variational reference** and optimize the choice of reference. This improves PT swapping significantly.

Parallel tempering

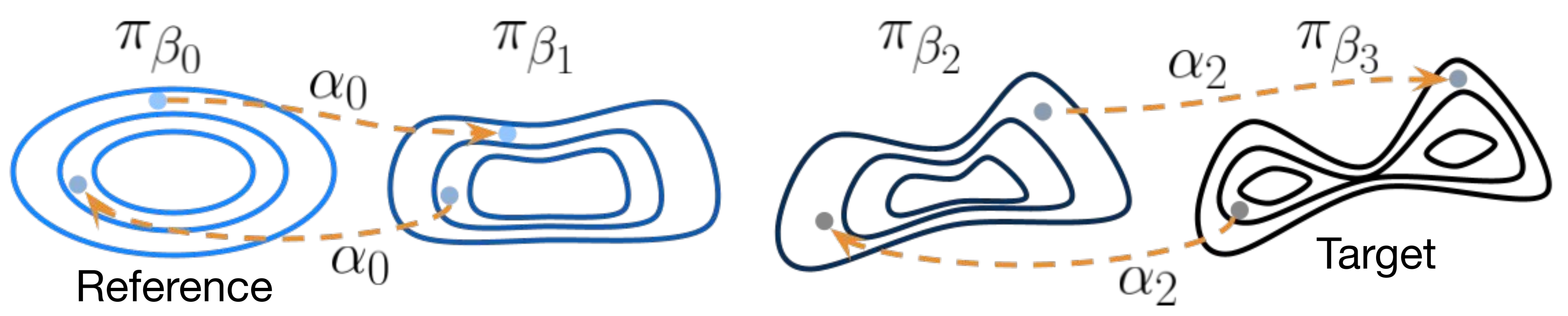
Annealing path: π_β is a path of distributions between π_0 and π_1. A linear path with a **fixed reference** is typically used: $\pi_\beta(x) \propto \pi_0^{1-\beta}(x) \cdot \pi_1^{\beta}(x)$

Run $N+1$ chains targeting π_{β_n}. Alternate local exploration and communication.

Local exploration: Update each chain according to an MCMC algorithm.



Communication: Swap states between chains n and $n+1$ with probability α_n.



Objective: Minimize the **global communication barrier** (GCB)

$$\Lambda(\pi_0, \pi_1) \approx \sum_{n=0}^{N-1} r_n, \quad r_n = 1 - \mathbb{E}[\alpha_n]$$

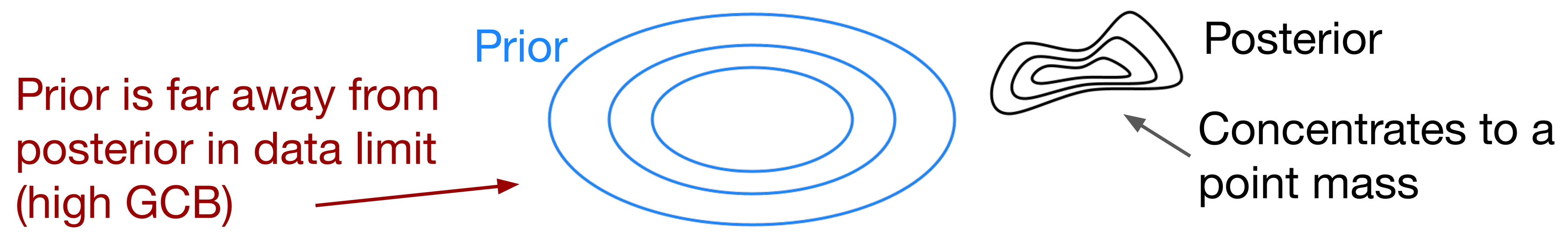
The GCB characterizes the efficiency of PT [Syed et al. 2019].

Suboptimality of a fixed reference

When the reference and target do not overlap much, PT will often reject communication swaps between chains.

Proposition: Suppose $\pi_1(x) \propto \pi_0(x) \cdot \prod_{i=1}^m f(Y_i; x)$, where $Y_1, \ldots, Y_m \overset{iid}{\sim} f(\cdot; x_0)$
Under some regularity conditions, with a fixed reference

$$\lim_{m\to\infty} \mathbb{E}[\Lambda(\pi_0, \pi_{1,m})] = \infty$$



Annealing paths with a variational reference

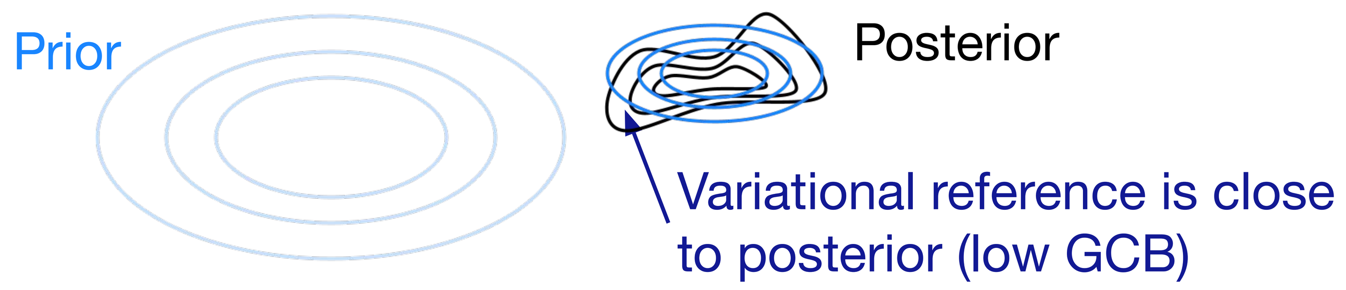
We introduce a **variational reference** family, $\{q_\phi : \phi \in \Phi\}$. The linear annealing path with the modified reference is

$$\pi_{\phi,\beta}(x) \propto q_\phi^{1-\beta}(x) \cdot \pi_1^{\beta}(x)$$

We consider exponential family reference distributions

$$q_\phi(x) = c(\phi) h(x) \exp(\phi^\top \eta(x))$$

and choose a reference distribution close to the target.



Proposition: Suppose $\pi_1(x) \propto \pi_0(x) \cdot \prod_{i=1}^m f(Y_i; x)$ where $Y_1, \ldots, Y_m \overset{iid}{\sim} f(\cdot; x_0)$
Then, there exists a sequence of multivariate normal reference distributions, q_{ϕ_m}, such that

$$\lim_{m\to\infty} \mathbb{E}[\Lambda(q_{\phi_m}, \pi_1)] = 0$$

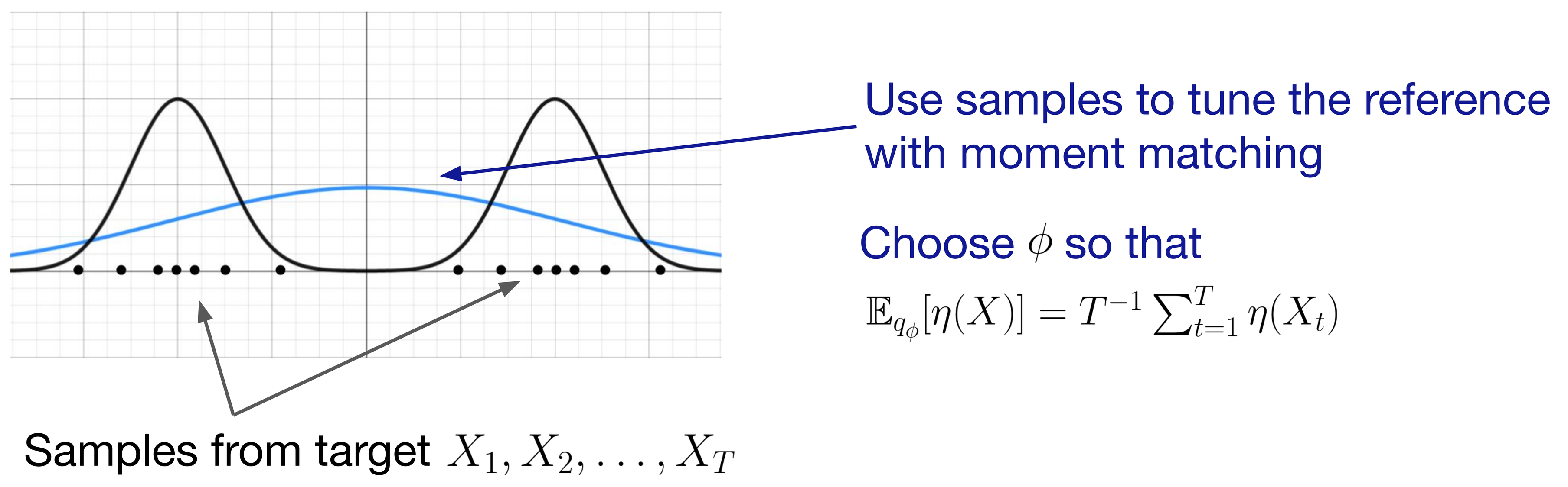
Variational reference tuning

We minimize the forward KL divergence,

$$\mathrm{KL}(\pi_1 || q_\phi) = \mathbb{E}_{\pi_1}[\log \pi_1(X)] - \mathbb{E}_{\pi_1}[\log q_\phi(X)]$$

and use a gradient-free procedure to tune the reference.

Variational reference tuning (...)



Use samples to tune the reference with moment matching

Choose ϕ so that

$$\mathbb{E}_{q_\phi}[\eta(X)] = T^{-1} \sum_{t=1}^T \eta(X_t)$$

Algorithm:

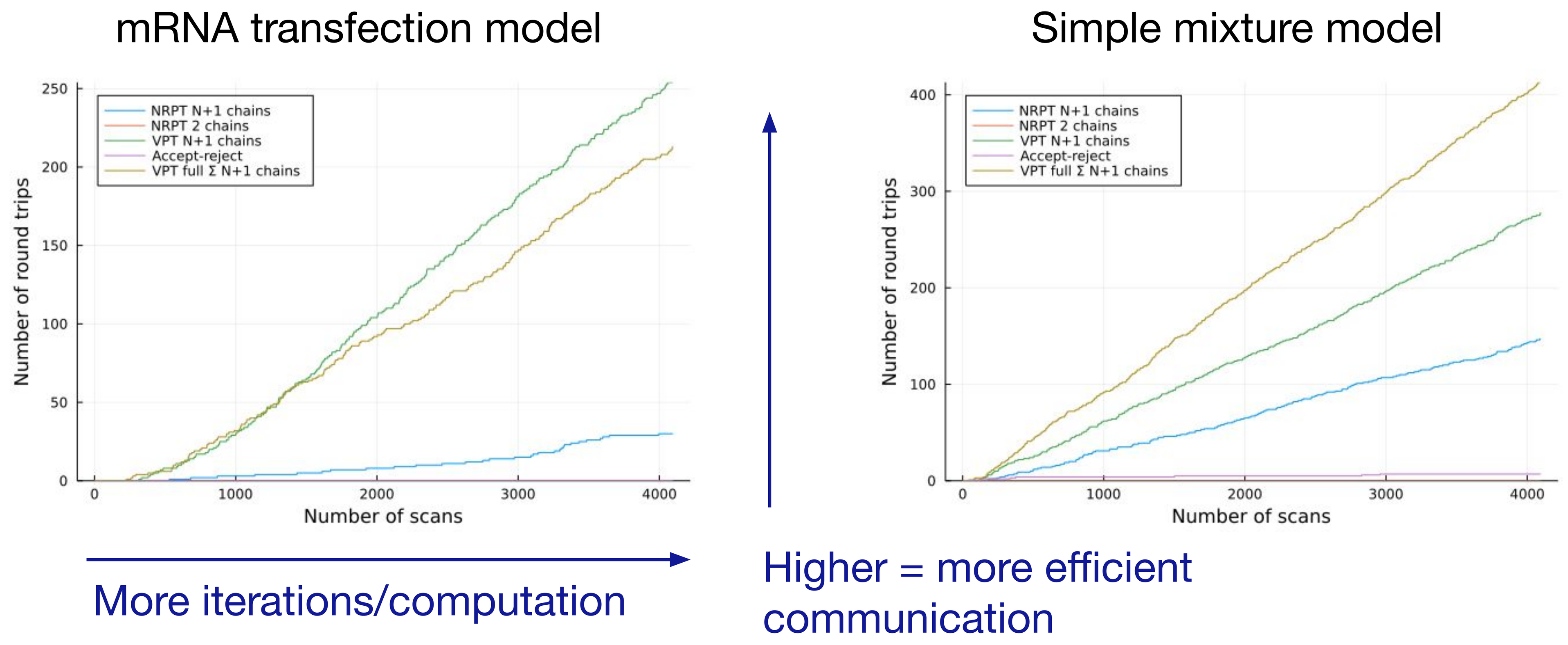
1) Run the non-reversible PT algorithm (**NRPT**) [Syed et al. 2019]
2) Use obtained samples $X_1, X_2, \ldots, X_T$ from the target chain to update the annealing schedule using the procedure in [Syed et al. 2019]
3) Update ϕ so that $\mathbb{E}_{q_\phi}[\eta(X)] = T^{-1} \sum_{t=1}^T \eta(X_t)$
4) Repeat 1-3 until the computational budget is depleted

Theorem (sketch): Let ϕ_{KL}^* minimize $\mathrm{KL}(\pi_1 || q_\phi)$ and let $\hat{\phi}_r$ be the variational parameter during the r-th tuning round. Then, $\hat{\phi}_r \to \phi_{\mathrm{KL}}^*$ almost surely.

We also offer a result that bounds the GCB at the forward KL minimum, $\Lambda(q_{\phi_{\mathrm{KL}}^*}, \pi_1)$, in terms of the flexibility of the variational family. (More details in paper.)

Experiments

PT with a variational reference empirically outperforms PT with a fixed reference. Green and gold: variational PT with a normal reference (mean-field approximation and full covariance). Blue: NRPT [Syed et al. 2019].



[1]Department of Statistics, UBC [2]Department of Statistics, University of Oxford

Email: nikola.surjanovic@stat.ubc.ca