State-Informed Temperature Swaps in Simulated Tempering

Hugo Queniat [1] Gareth O. Roberts [1] Nicholas G. Tawn [1]
[1]Department of Statistics, University of Warwick

Tempering in MCMC

- For a **multimodal** target π, standard **MCMC** algorithms with localized proposals typically **fail to move between modes**.
- One solution is **tempering**: raise π to a power $\beta \in (0,1]$ to **flatten** the target and enable **global exploration**.

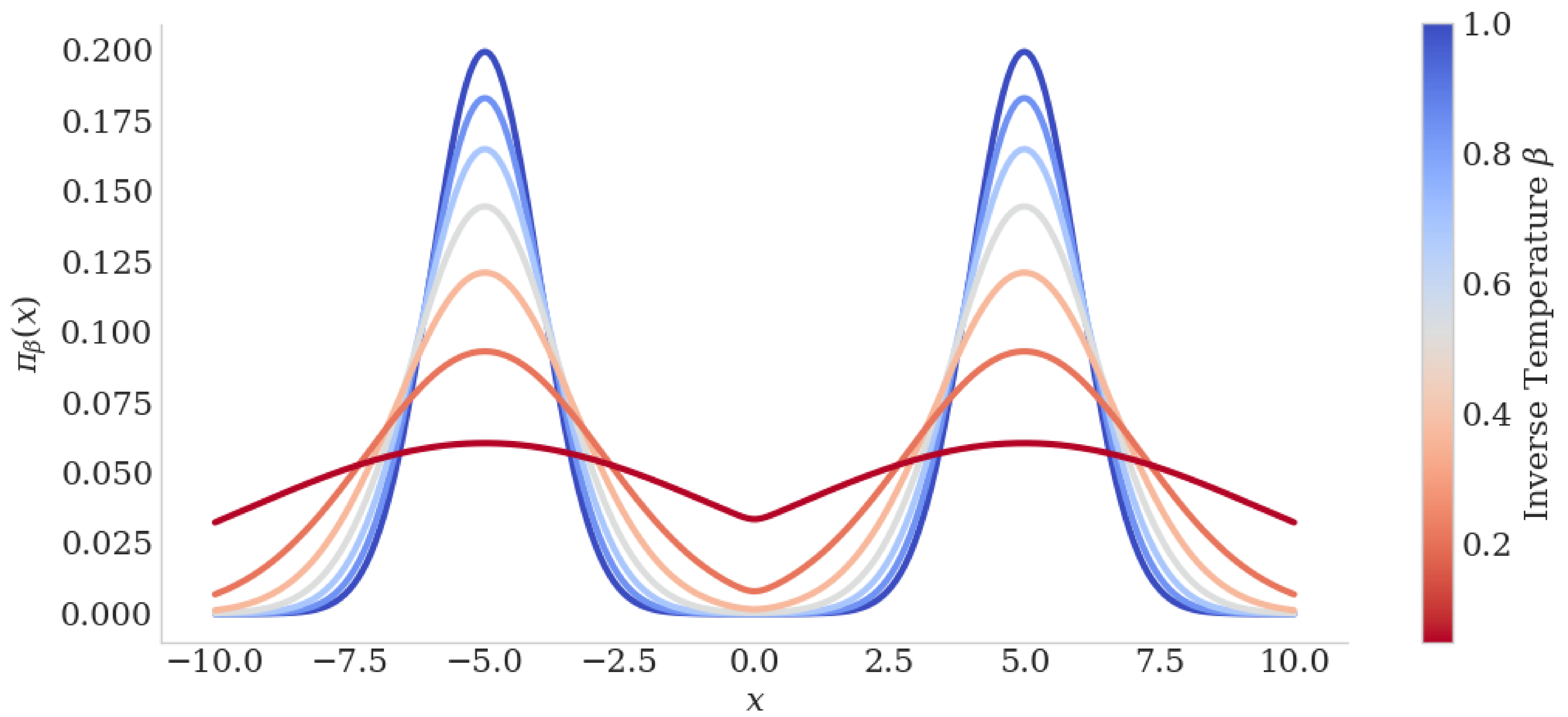

Figure 1. Bimodal mixture at decreasing inverse temperature β.

Simulated Tempering (ST)

Given a **temperature schedule** $1=\beta_0>\ldots>\beta_N>0$, **Simulated Tempering** [1] augments the target with a temperature index $j\in\{0,\ldots,N\}$,

$$\pi(x,j)\;\propto\;\pi(x)^{\beta_j}. \tag{1}$$

The ST procedure then alternates between **two** types of **moves**,

- **Within-temperature exploration**
  State-space exploration through a conventional transition kernel (RWM, HMC, *etc.*).
- **Temperature Swap moves**
  Temperature-space update to a neighbouring level; accepted according to a Metropolis ratio.

Standard Swap Proposal Schemes

- **Random-walk temperature moves** [2]
  The chain proposes to move to a neighbouring temperature chosen uniformly at random, creating a random walk over temperature levels.
- **Momentum-based temperature moves** [3]
  A momentum variable determines the direction of travel along the temperature schedule; moves continue in that direction until a rejection flips the momentum, producing a guided walk [4].

Towards efficient state-informed swap proposals?

Both random walk and guided walk temperature proposals are **blind** to the chain's position in the **state space**. However, the Metropolis ratio,

$$\alpha(i\to j)=\min\left\{1,\frac{\pi(\mathbf{x})^{\beta_j}}{\pi(\mathbf{x})^{\beta_i}}\right\}, \tag{2}$$

clearly depends on the current state $\mathbf{x}$.

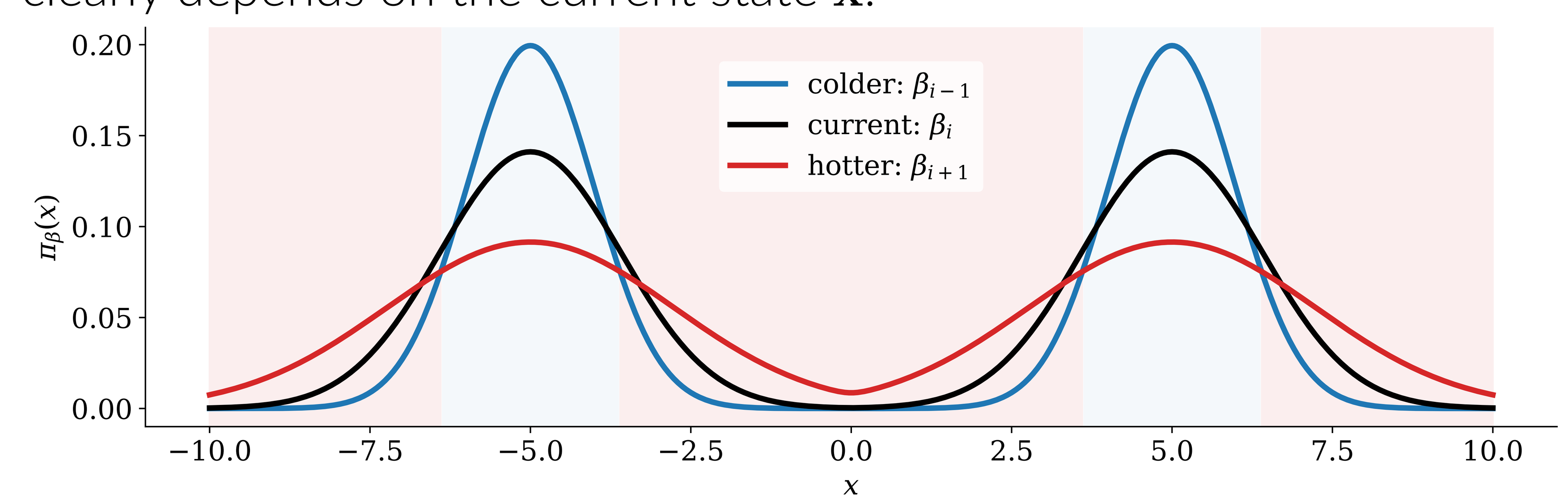

Figure 2. Locations where hotter or colder temperature moves have higher acceptance.

Locally Balanced Proposals

- To swap β_i and β_j, we use **locally balanced (LB) proposals** [5],

$$Q(i\to j)\propto g\left(\frac{\pi(x)^{\beta_j}}{\pi(x)^{\beta_i}}\right), \tag{3}$$

where g is a *balancing function*, i.e. satisfies $g(t)=t\,g(1/t)$.
Typical choices: $g_{\sqrt{}}(t)=\sqrt{t}$, $g_{\mathrm{Barker}}(t)=\frac{t}{1+t}$.

- The *asymptotic* acceptance probability of a **locally balanced square–root proposal** with *temperature scaling* $\ell>0$ is

$$\mathrm{ACC}_{\mathrm{LB}\text{-}\sqrt{}}(\ell)=2\,\Phi\Big(-\tfrac{\ell}{2}\sqrt{I}\Big)+2\int_{\frac{\ell^2 I}{2}}^{\infty}\tanh\Big(\tfrac{w}{2}\Big)\,\mathcal{N}(w;0,\ell^2 I)\,dw, \tag{4}$$

where I is problem–specific. We recognize the Random Walk Metropolis acceptance [6], which gives

$$\mathrm{ACC}_{\mathrm{LB}\text{-}\sqrt{}}(\ell)=\mathrm{ACC}_{RW}(\ell)+2\int_{\frac{\ell^2 I}{2}}^{\infty}\tanh\Big(\tfrac{w}{2}\Big)\,\mathcal{N}(w;0,\ell^2 I)\,dw. \tag{5}$$

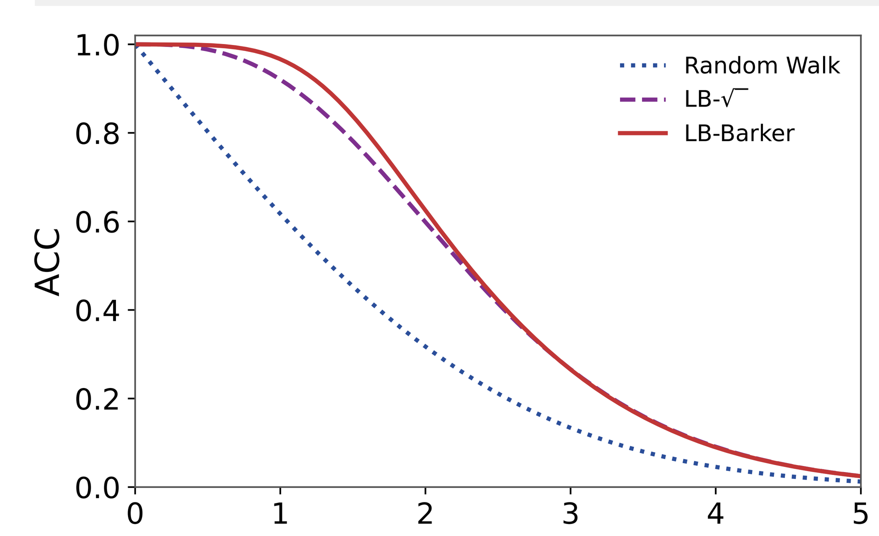

Figure 3. Acceptance versus scaling ℓ.

- Theory (5) and the plot show that the **Square–root Balancing** yields **higher acceptance rates** than Random Walk.
- **Barker Balancing** provides the **best performance we observed**, and will be used in the subsequent experiments and plots.

Efficiency Guarantees

- We quantify mixing by the **volatility of the scaled diffusion limit**, a measure of **how fast** the chain **moves across temperatures** [6, 7].

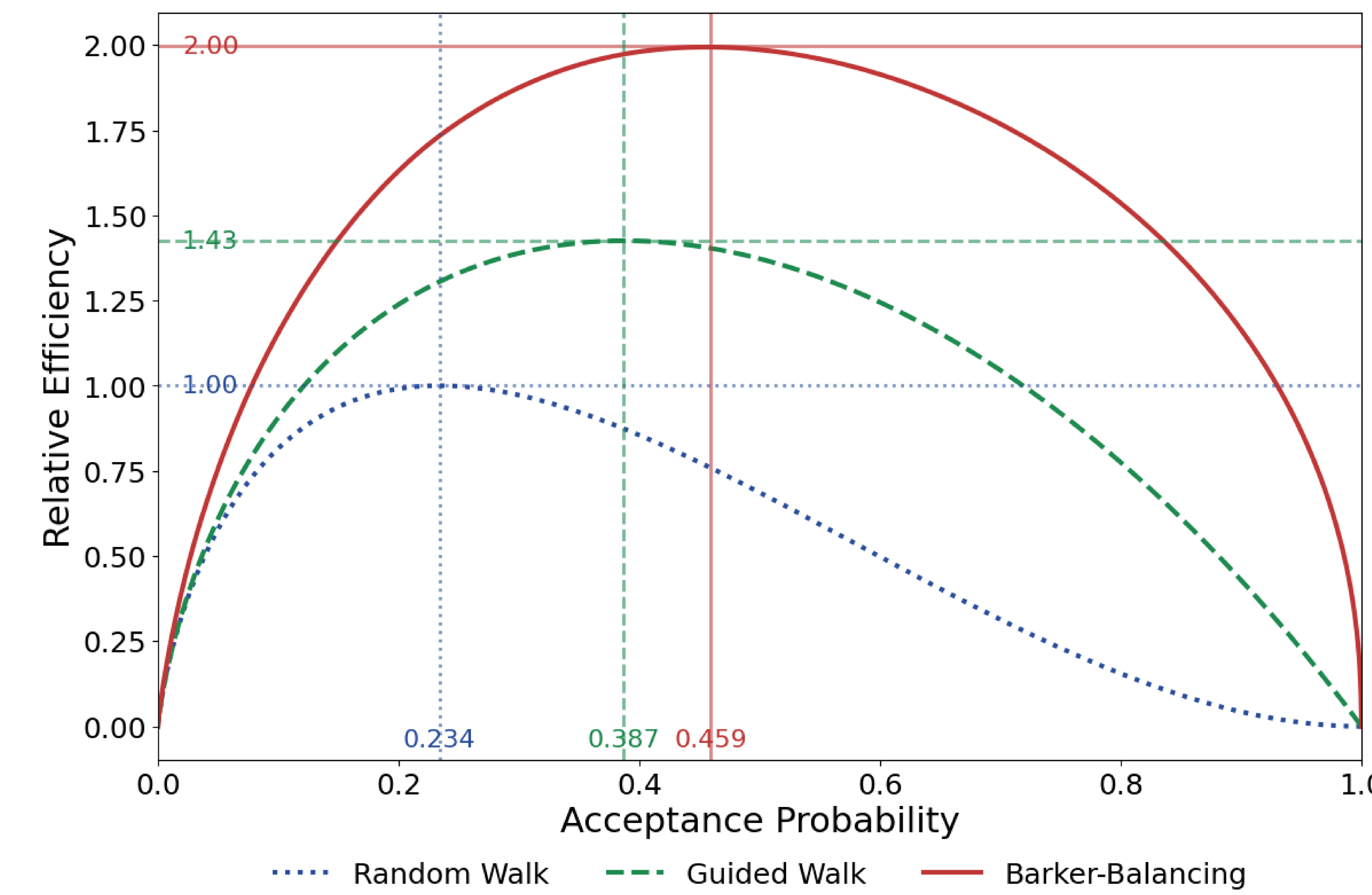

Figure 4. **Efficiency curves**, rescaled to the maximum efficiency of the random walk scheme.

Empirical Validation

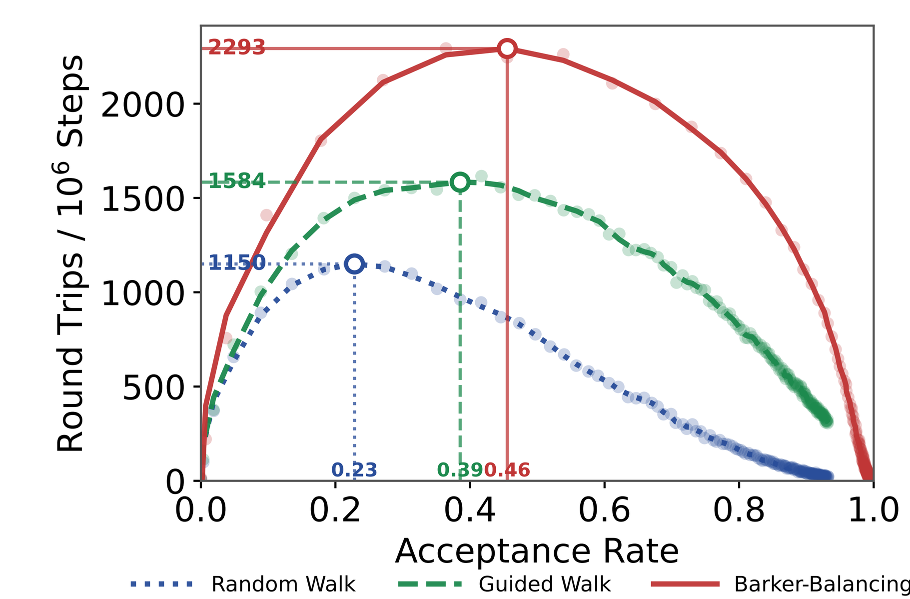

Figure 5. **Round trips** plotted against acceptance rate when targeting an isotropic Gaussian ($d=50$).

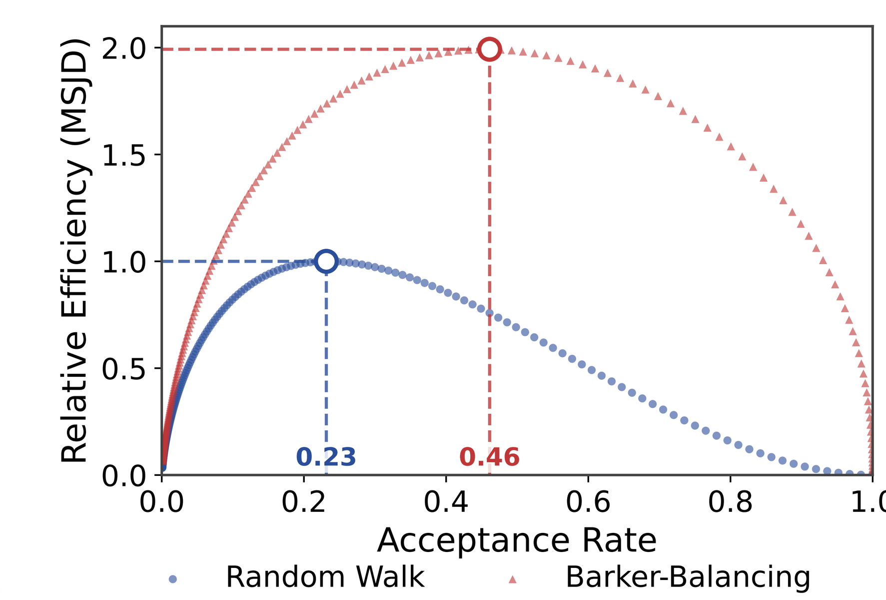

Figure 6. **Squared jumping distance**, simulating swaps for an anisotropic Gaussian ($d=300$).

References

[1] E Marinari and G Parisi. Simulated Tempering: A New Monte Carlo Scheme. *Europhysics Letters (EPL)*, 19(6):451–458, July 1992.
[2] Charles J. Geyer and Elizabeth A. Thompson. Annealing Markov Chain Monte Carlo with Applications to Ancestral Inference. *Journal of the American Statistical Association*, 90(431):909–920, September 1995.
[3] Miguel Biron-Lattes, Trevor Campbell, and Alexandre Bouchard-Côté. Automatic Regenerative Simulation via Non-Reversible Simulated Tempering. *Journal of the American Statistical Association*, 120(549):318–330, January 2025.
[4] Paul Gustafson. A guided walk Metropolis algorithm. *Statistics and Computing*, 8(4):357–364, December 1998.
[5] Giacomo Zanella. Informed Proposals for Local MCMC in Discrete Spaces. *Journal of the American Statistical Association*, 115(530):852–865, April 2020.
[6] Gareth O. Roberts and Jeffrey S. Rosenthal. Minimising MCMC variance via diffusion limits, with an application to simulated tempering. *The Annals of Applied Probability*, 24(1):131–149, February 2014.
[7] Gareth O. Roberts and Jeffrey S. Rosenthal. Quantifying the Speed-Up from Non-Reversibility in MCMC Tempering Algorithms, January 2025.