# MSci Project Literature Survey and Outline : Irreversible Markov Samplers

Eoghan Lemoine. SN: 21010005
Supervisor: Professor Edina Rosta
University College London

November 2024

---

## INDEX:

## Contents

# Abstract

Markov Chain Monte Carlo (MCMC) methods are widely used to sample from complex probability distributions. Traditional MCMC algorithms, such as the Metropolis-Hastings algorithm, rely on the detailed balance condition to ensure convergence to the target distribution. However, this condition can limit the efficiency of sampling, especially in high-dimensional spaces. Irreversible MCMC methods, which relax the detailed balance condition, have gained increasing attention as a means to accelerate sampling. A particularly promising approach involves the skewed detailed balance condition, which allows for asymmetric proposal distributions, leading to faster convergence and better mixing properties. This project extends the application of the skewed detailed balance condition to higher-dimensional spaces. By exploiting the flexibility afforded by irreversible dynamics, efficient sampling algorithms will be developed for a broad range of complex distributions. Additionally several new auxiliary optimisation techniques will be deployed with an irreversible sampler, with the aim of gaining improvements in efficiency.


## Background and theory

Markov chains are a class of stochastic processes where the future state of a system depends only on its present state [1], not its past history. This property of 'memorylessness' is known as the Markov property, and it allows us to model a wide range of phenomena, from weather patterns to financial markets and statistical physics [2]. Monte Carlo methods, on the other hand, are a class of computational techniques that rely on repeated random sampling to obtain a target distribution [3]. By generating a large number of random samples and analysing their statistical properties, these methods can approximate a 'target' distribution that is otherwise difficult or impossible to solve analytically. Markov Chain Monte Carlo (MCMC) methods amalgamate the concepts of Markov chains with algorithmic techniques, such as the Metropolis-Hastings algorithm [4], which generates a sequence of samples by proposing a candidate sample based on the current state of the Markov chain and then accepting or rejecting the candidate based on an acceptance criterion derived from the target distribution.

In essence we construct a Markov chain whose stationary distribution converges to a desired target distribution. By simulating this chain for an adequate duration - the mixing time, one can recover the underlying distribution being studied.

A Markov matrix $P$ is a stochastic matrix satisfying two important conditions. For a given transition matrix $P$ over a discrete state space $\Omega = \{1, \ldots, N\}$:

1. **Non-Negativity:**

$$P_{ij} \geq 0, \quad \forall i, j \in \Omega, \tag{1}$$

   Every entry represents a probability and must be non-negative.

2. **Row-Stochasticity:**

$$\sum_j P_{ij} = 1, \quad \forall i, j \in \Omega, \tag{2}$$

   The sum of transition probabilities for any state $i$ equals 1.

The conditions and assumptions allowed, and the strictness with which they are imposed can significantly influence the success or failure of a sampling method across different contexts. Aperiodicity removes the possibility of cyclical behaviour (see figure 1). Ergodicity or irreducibility ensures that the chain can eventually reach any state from any other state in a finite time [1].

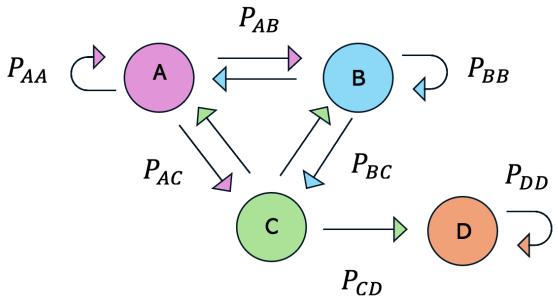


*Figure 1: Diagram showing an example Markov chain which is inherently periodic. Once state D is reached, there is zero probability of escape.*

$$P^n(i,j) > 0 \tag{3}$$

Our first two assumptions allow us to assume that in the transition probabilities converge to the stationary distribution as the number of steps taken approaches infinity.
Overall we find an invariant target probability distribution $\pi = (\pi_1, \ldots, \pi_N)$ with $\pi_i > 0$ and $\sum_{i=1}^{N} \pi_i = 1$
The global balance condition is the most basic and fundamental of conditions, ensuring that, in a stationary state, the rate at which probability mass flows into a state i is equal to the rate at which it flows out. For a static probability distribution [5]:

$$\pi_i = \sum_{j=1}^{N} \pi_j P_{ji}, \quad \forall i, j \in \Omega, \tag{4}$$

where:

- $\pi_i$: The stationary probability of being in state $i$.
- $P_{ji}$: The transition probability from state $j$ to state $i$.

Samplers in the past have long relied on satisfaction of the detailed balance condition (DBC) to ensure convergence to the target distribution. The DBC is expressed as:

$$\pi_i P_{ij} = \pi_j P_{ji} \tag{5}$$

Detail balance (equation(5)) is a more stringent condition[4], ensuring the probability flux between any two given states is equal. While always guaranteeing convergence, adhering to DBC will often limit the efficiency of sampling. Potential barriers and local minima pose a threat to the reversible sampler often leading to topological freezing as seen in several recent studies [6-10]. In practice these conditions lead to the effective breaking of ergodicity. While the chain is still irreducible the time and computational cost associated with sheer number of calculations required make convergence to the static target distribution impossible in practice. More recently several algorithms have been proposed which break the detailed balance condition [11-16]. This is discussed further in the literature overview section.

We focus on the model proposed by Turitsyn et al [17], which introduces a lifting framework combined with a skewed detailed balance condition (SDBC) to break the symmetry of the system and create an irreversible Markov chain. The asymmetry introduced is controlled, as shown in figure 2. By introducing an additive variable $\epsilon \in \{+, -\}$ into the state space $I = \{1, 2, \ldots, S\}$, and lifting this space, the system's state space is effectively doubled (or expanded further, depending on the number of replicas). Under these conditions, a new balance condition can be derived from the global balance condition (equation (4)):

$$\pi_i P_{ij}^{+\varepsilon} = \pi_j P_{ji}^{-\varepsilon}, \tag{6}$$

The skewed detailed balance (equation (6)) ensures the overall conservation of probability flux within the system. The inter-replica transition probabilities $\Lambda$ are adjusted to maintain the stochasticity condition allowing convergence to a steady-state distribution. The proposal algorithm must be changed to reflect these new transitions. An irreversible variation of the metropolis hastings algorithm must be used, handling the creation of random variables and the rejection/acceptance of new states.

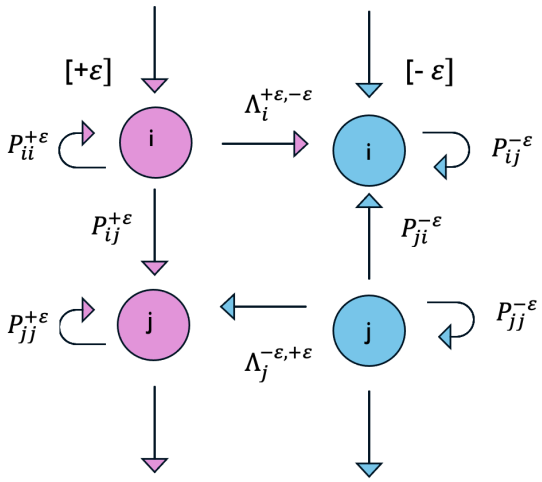


*Figure 2: Diagram showing lifting variable and extension of the phase space*

The newly allowed transitions between replicas, represented by $\Lambda^{(+,-)}$ and $\Lambda^{(-,+)}$, facilitate faster exploration of the state space. This allows the system to switch between different replicas, helping it avoid getting trapped in loops and local minima. The updated transition matrix, incorporating these inter-replica transitions, is given by:

$$\hat{P} = \begin{pmatrix} P^{(+)} & \Lambda^{(+,-)} \\ \Lambda^{(-,+)} & P^{(-)} \end{pmatrix}$$

Relaxation time is one of the primary metrics by which the performance of a Markov sampler can be quantified. Relaxation time describes the characteristic time it takes for the system to converge from an arbitrary initial distribution to its stationary distribution $\pi$ [18]. A shorter relaxation time indicates faster convergence and thus a more efficient sampler, assessing the ability a Markov chain explores its state space.

Relaxation time is often associated with mixing time, which measures how quickly the Markov chain reaches a state that is close enough to its stationary distribution, typically quantified in terms of the total variation distance (TVD). The TVD measures the difference between the current distribution of the Markov chain and the stationary distribution [19]:

$$\text{TVD}(P_t, \pi) = \frac{1}{2} \sum_i |P_t(i) - \pi(i)| \tag{7}$$

The mixing time depends on the threshold chosen to determine when the total variation distance between the current and stationary distributions is sufficiently small (often below a certain threshold, e.g., $\epsilon = 1/2$). At this point, the matrix is deemed to be 'mixed'. This threshold can be adjusted depending on the system and the desired accuracy.

$$t_{\text{mix}}(\epsilon) = \min \left\{ t : \max_x \left| P^t(x, \cdot) - \pi \right|_{\text{TV}} \leq \epsilon \right\} \tag{8}$$

The relaxation time is influenced by the spectral properties of the transition matrix, particularly the second-largest eigenvalue of the transition matrix $\lambda_2$. The larger the spectral gap, defined as $1 - |\lambda_2|$, the faster the chain mixes and thus the shorter the relaxation time [20]. The mixing time, $T_{\text{mix}}$, is typically proportional to

$$T_{\text{mix}} \sim \frac{1}{1 - |\lambda_2|} \tag{9}$$

## Literature Overview and Aims

Indeed, traditional sampling methods in this field encounter significant issues due to their reliance on satisfying detailed balance, which often halts efficient sampling of the phase space. There is no shortage of examples of these problems. Schaefer et al.[8] employed reversible MCMC to sample Quantum Chromodynamics (QCD) effects in lattice simulations and encountered severe topological freezing. They observed extremely large autocorrelation times at fine lattice spacings. Similarly, Bonati et al. [10] investigated a quantum mechanical particle under various constraints and faced critical slowdown as the continuum limit was approached, resulting in a loss of ergodicity. More recent work [11-17] explores innovative approaches to breaking the detailed balance condition, creative approaches are often taken to create irreversible chains which converge and satisfy the balance condition. These approaches have been shown to achieve significant improvements

in sampling efficiency and convergence speed. Bernard et al.[12] broke detail balance by developing the event-chain Monte Carlo algorithms, which were applied to model hard-sphere systems. Contrary to the metropolis hastings, these algorithms are rejection-free and involve moving particles in a deterministic chain until a specified total displacement is reached. The irreversible versions of their algorithms showed even greater performance improvements, significantly outperforming the traditional Metropolis algorithm and achieving faster equilibration than molecular dynamics simulations. Suwa and Todo [13] developed an algorithm this time minimising rejections using a "weight landfill" method, they create a directed flow in configuration space, reducing random walk behaviour. This approach significantly shortens autocorrelation times, achieving up to six times faster convergence than the Metropolis algorithm in models like the Potts model and quantum Heisenberg spin chain.

Higher-dimensional irreversible samplers would be particularly useful in complex systems such as molecular dynamics, bayesian inference in high-dimensional models, and large-scale network analysis. These applications often involve intricate landscapes and require efficient sampling methods to obtain accurate results. There is a noticeable lack of literature investigating irreversible samplers in higher dimensions, in fact, irreversible samplers in general have received surprisingly little attention. This may be simply due to the fact that the majority of real word systems which are being studied are by nature more complex and multivariate in nature, unlike their hypothetical counterparts.

Irreversible MCMC samplers are inherently more complex to design and implement compared to reversible ones. Ensuring that these samplers maintain the correct stationary distribution without detailed balance requires sophisticated mathematical and computational techniques. This complexity increases significantly in higher dimensions.

Other authors propose auxiliary techniques which aim to optimise reversible chains. Geyer et al [21] implements a simulated annealing optimisation technique to help identify a global minimum, this model includes a temperature parameter to aid with escaping local minima. Koskin et al. [22] also take a somewhat different approach projecting the network nodes onto a one-dimensional reaction coordinate and performing a variational boundary search using a parallel tempering algorithm. Parallel tempering involves running multiple simulations (replicas) at different temperatures and allowing exchanges between neighbouring temperatures. This helps the system escape local minima and explore the configuration space more effectively. Naturally seeing these auxiliary techniques find success on reversible algorithms makes the idea of applying them to an irreversible chain in pursuit of even greater performance a very attractive one.

The primary goal of this project is to develop a practical framework for applying the modified version of the metropolis hastings algorithm with the lifting variable in higher dimensions. As an extension several auxiliary optimisation techniques which have previously only been applied in conjunction with a traditional reversible MCMC will be applied in the irreversible case.

This project specifically focuses on the skewed detailed balance condition which enables asymmetric transitions in the Markov chain, breaking reversibility while still preserving the stationary distribution. By leveraging this framework, the project aims to develop higher-dimensional irreversible samplers by extending the SDBC framework to higher-dimensional spaces, addressing the scarcity of practical studies in this area. This project is mainly concerned on the optimisation of the sampling via increased efficiency. This project also hopes to extend the scope of existing sampling techniques to higher dimensions. Another way this can be achieved is through the incorporation of auxiliary optimisation techniques. This project aims to investigate whether optimisation methods traditionally applied to reversible samplers, such as simulated annealing and parallel tempering can further enhance the performance of irreversible samplers.

This project aims to apply these methods in various different environments. Initially a potential energy landscape of various dimensionality will

be used as a sample data for algorithm validation. This simple system mimics biomolecular dynamics and is widely applicable in Monte Carlo simulations. Tests will be extended to larger and more complex state spaces to study and document the difficulty of expanding the sampling to higher systems of higher dimensions.

## Project Timeline

- **October:** Create literature survey and draft project outline.
- **November:** Begin coding, create tools to evaluate performance, integrate with HPC cluster.
- **December – January:**
  - Test and quantify performance of irreversible samplers in $N \geq 2$ dimensions.
  - Extend the system to include a larger number of states to analyse scalability. Demonstrate improved efficiency in convergence and reduced sampling error for high-dimensional systems.
- **February Onward:**
  - Testing algorithms on practical real world applications (e.g: molecular simulations, statistical physics).
  - Implement and test auxiliary techniques without detail balance (e.g. Temperature annealing and parallel tempering).

# References


[1] Kemeny, J. G., & Snell, J. L. (1976). *Finite Markov Chains*. Springer-Verlag.

[2] Brooks, S., Gelman, A., Jones, G. L., & Meng, X.-L. (2011). *Handbook of Markov Chain Monte Carlo*. CRC Press.

[3] Rubinstein, R. Y., & Kroese, D. P. (2017). *Simulation and the Monte Carlo Method* (3rd ed.). Wiley.

[4] Metropolis, F., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., & Teller, E. (1953). Equation of state calculations by fast computing machines. *J. Chem. Phys.*, 21, 1087.

[5] Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57(1), 97–109.

[6] Alles, B., Boyd, G., D'Elia, M., Di Giacomo, A., & Vicari, E. (1996). Hybrid Monte Carlo and topological modes of full QCD. *Physics Letters B*, 389, 107.

[7] Del Debbio, L., Manca, G. M., & Vicari, E. (2004). Critical slowing down of topological modes. *Physics Letters B*, 594, 315.

[8] Schaefer, S., Sommer, R., Virotta, F., & the A. Collaboration. (2011). Critical slowing down and error analysis in lattice QCD simulations. *Nuclear Physics B*, 845, 93.

[9] McGlynn, G., & Mawhinney, R. D. (2014). Diffusion of topological charge in lattice QCD simulations. *Physical Review D*, 90, 074502.

[10] Bonati, C., & D'Elia, M. (2018). Topological critical slowing down: variations on a toy model. *Physical Review E*, 98, 013308.

[11] Diaconis, P., Holmes, S., & Neal, R. M. (2000). Analysis of a nonreversible Markov chain sampler. *Annals of Applied Probability*, 726.

[12] Bernard, E. P., Krauth, W., & Wilson, D. B. (2009). Event-chain Monte Carlo algorithms for hard-sphere systems. *Physical Review E*, 80, 056704.

[13] Suwa, H., & Todo, S. (2010). Markov chain Monte Carlo method without detailed balance. *Physical Review Letters*, 105, 120603.

[14] Schram, R. D., & Barkema, G. T. (2015). Monte Carlo methods beyond detailed balance. *Physica A: Statistical Mechanics and its Applications*, 418, 88.

[15] Kaiser, M., Jack, R. L., & Zimmer, J. (2017). Acceleration of convergence to equilibrium in Markov chains by breaking detailed balance. *Journal of Statistical Physics*, 168, 259.

[16] Faizi, F., Deligiannidis, G., & Rosta, E. (2020). Efficient irreversible Monte Carlo samplers. *Journal of Chemical Theory and Computation*, 16, 2124.

[17] Turitsyn, K. S., Chertkov, M., & Vucelja, M. (2011). Irreversible Monte Carlo algorithms for efficient sampling. *Physica D: Nonlinear Phenomena*, 240(4), 410–414.

[18] Lovász, L., & Simonovits, M. (1993). Random walks in a convex body and an improved volume algorithm. *Journal of Computer and System Sciences*, 47(3), 384–403.

[19] Levin, D. A., Peres, Y., & Wilmer, E. L. (2009). *Markov Chains and Mixing Times*. American Mathematical Society.

[20] Sinclair, A. (1992). Improved bounds for mixing rates of Markov chains and multicommodity flow. *SIAM Journal on Computing*, 21(4), 721–746.

[21] Geyer, C. J., & Thompson, E. A. (1995). Annealing Markov chain Monte Carlo with applications to ancestral inference. *Journal of the American Statistical Association*, 90(431), 909.

[22] Koskin, V., Kells, A., Clayton, J., Hartmann, A. K., Annibale, A., & Rosta, E. (2023). Variational kinetic clustering of complex networks. *The Journal of Chemical Physics*, 158(12), 124112.