Founding papers on non-reversible MCMC

Werner Krauth
Laboratoire de Physique, Ecole normale supérieure, Paris, France

05 August 2021
Non-reversible Markovian Monte Carlo @ Lorentz Center

Diaconis P, Holmes S, Neal RM *Analysis of a nonreversible Markov chain sampler*.
Ann Appl Probab. (2000) 10:726–52. doi: 10.1214/aoap/1019487508

Chen F, Lovász L, Pak I *Lifting Markov chains to speed up mixing.*
In: Proceedings of the 17th Annual ACM Symposium on Theory of Computing. Providence, RI (1999). p. 275. doi: 10.1145/301250.301315

Baik J, Liu Z. *TASEP on a ring in sub-relaxation time scale.*
J Stat Phys. (2016)165:1051–85. doi: 10.1007/s10955-016-1665-y

---

Kapfer SC, Krauth W. *Irreversible local Markov chains with rapid convergence towards equilibrium*
Phys Rev Lett. (2017) 119:240603. doi: 10.1103/PhysRevLett.119.240603

Work supported by A. v. Humboldt Foundation

Département de Physique
École normale supérieure

Diaconis–Holmes–Neal (2000) 1/2

The Annals of Applied Probability
2000, Vol. 10, No. 3, 726–752

ANALYSIS OF A NONREVERSIBLE MARKOV CHAIN SAMPLER

BY PERSI DIACONIS,[1] SUSAN HOLMES AND RADFORD M. NEAL[2]

Stanford University, Stanford University and INRA and University of Toronto

We analyze the convergence to stationarity of a simple nonreversible Markov chain that serves as a model for several nonreversible Markov chain sampling methods that are used in practice. Our theoretical and numerical results show that nonreversibility can indeed lead to improvements over the diffusive behavior of simple Markov chain sampling schemes. The analysis uses both probabilistic techniques and an explicit diagonalization.

1. Introduction. Markov chain sampling methods are commonly used in statistics [33, 32], computer science [31], statistical mechanics [3] and quantum field theory [34, 23]. In all these fields, distributions are encountered that are difficult to sample from directly, but for which a Markov chain that converges to the distribution can easily be constructed. For many such methods (e.g., the Metropolis algorithm [25, 13], and the Gibbs sampler [17, 16] with a random scan) the Markov chain constructed is reversible. Some of these methods explore the distribution by means of a diffusive random walk. We use the term "diffusive" for processes like the ordinary random walk on a d-dimensional lattice which require time of order T^2 to travel distance T. Some other common methods, such as the Gibbs sampler with a systematic scan, use a Markov chain that is not reversible, but have diffusive behavior resembling that of a related reversible chain [30].

Diaconis–Holmes–Neal (2000) 1/2

The Annals of Applied Probability
2000, Vol. 10, No. 3, 726–752

ANALYSIS OF A NONREVERSIBLE MARKOV CHAIN SAMPLER

BY PERSI DIACONIS,[1] SUSAN HOLMES AND RADFORD M. NEAL[2]

Stanford University, Stanford University and INRA and University of Toronto

We analyze the convergence to stationarity of a simple nonreversible Markov chain that serves as a model for several nonreversible Markov chain sampling methods that are used in practice. Our theoretical and numerical results show that nonreversibility can indeed lead to improvements over the diffusive behavior of simple Markov chain sampling schemes. The analysis uses both probabilistic techniques and an explicit diagonalization.

1. Introduction. Markov chain sampling methods are commonly used in statistics [33, 32], computer science [31], statistical mechanics [3] and quantum field theory [34, 23]. In all these fields, distributions are encountered that are difficult to sample from directly, but for which a Markov chain that converges to the distribution can easily be constructed. For many such methods (e.g., the Metropolis algorithm [25, 13], and the Gibbs sampler [17, 16] with a random scan) the Markov chain constructed is reversible. Some of these methods explore the distribution by means of a diffusive random walk. We use the term "diffusive" for processes like the ordinary random walk on a d-dimensional lattice which require time of order T^2 to travel distance T. Some other common methods, such as the Gibbs sampler with a systematic scan, use a Markov chain that is not reversible, but have diffusive behavior resembling that of a related reversible chain [30].

The Annals of Applied Probability
2000, Vol. 10, No. 3, 726–752

ANALYSIS OF A NONREVERSIBLE MARKOV CHAIN SAMPLER

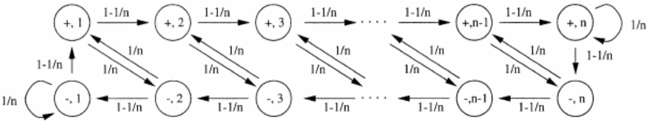
BY PERSI DIACONIS,[1] SUSAN HOLMES AND RADFORD M. NEAL[2]

Stanford University, Stanford University and INRA and University of Toronto

We analyze the convergence to stationarity of a simple nonreversible Markov chain that serves as a model for several nonreversible Markov chain sampling methods that are used in practice. Our theoretical and numerical results show that nonreversibility can indeed lead to improvements over the diffusive behavior of simple Markov chain sampling schemes. The analysis uses both probabilistic techniques and an explicit diagonalization.

1. Introduction. Markov chain sampling methods are commonly used in statistics [33, 32], computer science [31], statistical mechanics [3] and quantum field theory [34, 23]. In all these fields, distributions are encountered that are difficult to sample from directly, but for which a Markov chain that converges to the distribution can easily be constructed. For many such methods (e.g., the Metropolis algorithm [25, 13], and the Gibbs sampler [17, 16] with a random scan) the Markov chain constructed is reversible. Some of these methods explore the distribution by means of a diffusive random walk. We use the term "diffusive" for processes like the ordinary random walk on a d-dimensional lattice which require time of order T^2 to travel distance T. Some other common methods, such as the Gibbs sampler with a systematic scan, use a Markov chain that is not reversible, but have diffusive behavior resembling that of a related reversible chain [30].

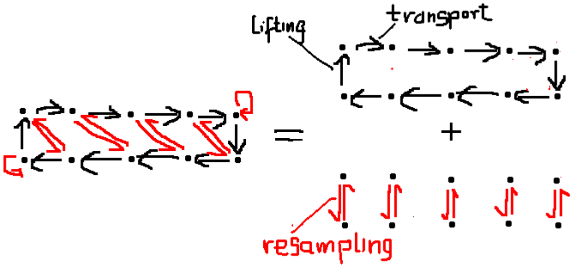
Diaconis–Holmes–Neal (2000) 2/2



The transition matrix is doubly stochastic, and thus the stationary distri-

Département de Physique
École normale supérieure

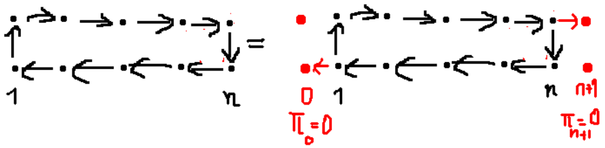
Diaconis–Holmes–Neal (2000) Comments 1/4



Glossary:

- 'lifting move': required by global balance.
- 'resampling': special lifting move good for irreducibility, aperiodicity or for speed.

Diaconis–Holmes–Neal (2000) Comments 2/4



- Single particle on a path graph $\mathcal{P}_n$, not on a ring.
- Phantom vertices 0 and $n+1$ on $\mathcal{P}_n$ illustrate the 'rejections $\rightarrow$ liftings' mystery.
- NB: Lifting moves $\neq$ resamplings

Metropolis et al (1953) (1/2)

URNAL OF CHEMICAL PHYSICS VOLUME 21, NUMBER 6 JUN

Equation of State Calculations by Fast Computing Machines

NICHOLAS METROPOLIS, ARIANNA W. ROSENBLUTH, MARSHALL N. ROSENBLUTH, AND AUGUSTA H. TELLER,
Los Alamos Scientific Laboratory, Los Alamos, New Mexico

AND

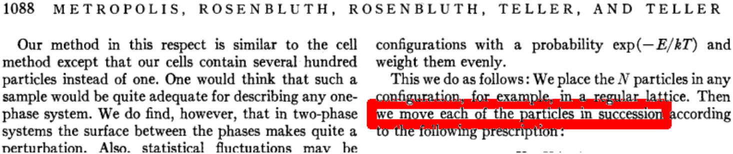
EDWARD TELLER,* *Department of Physics, University of Chicago, Chicago, Illinois*

(Received March 6, 1953)

A general method, suitable for fast computing machines, for investigating such properties as equations of state for substances consisting of interacting individual molecules is described. The method consists of a modified Monte Carlo integration over configuration space. Results for the two-dimensional rigid-sphere system have been obtained on the Los Alamos MANIAC and are presented here. These results are compared to the free volume equation of state and to a four-term virial coefficient expansion.

Département de Physique
École normale supérieure

Metropolis et al (1953) (2/2)



1088 METROPOLIS, ROSENBLUTH, ROSENBLUTH, TELLER, AND TELLER

Our method in this respect is similar to the cell method except that our cells contain several hundred particles instead of one. One would think that such a sample would be quite adequate for describing any one-phase system. We do find, however, that in two-phase systems the surface between the phases makes quite a perturbation. Also, statistical fluctuations may be

configurations with a probability $\exp(-E/kT)$ and weight them evenly.

This we do as follows: We place the N particles in any configuration, for example, in a regular lattice. Then we move each of the particles in succession according to the following prescription:

- The 'Sweep' variant of any reversible single-particle-move MCMC satisfies global balance (same for single spin flips)

Diaconis–Holmes–Neal (2000) Comments 3/4

- Metropolis flows on $\mathcal{P}_n$ (with 'phantom' vertices):

$$\begin{tikzcd}[column sep=huge]
\boxed{\text{i-1}} \arrow[r, shift left, "\frac{1}{2}\min(\pi_{i-1},\pi_i)"] & \boxed{\text{i}} \arrow[l, shift left, "\frac{1}{2}\min(\pi_i,\pi_{i-1})"] \arrow[r, shift left, "\frac{1}{2}\min(\pi_i,\pi_{i+1})"] \arrow[loop above, "\underbrace{\pi_i - \frac{1}{2}\min(\pi_i, \pi_{i-1}) - \frac{1}{2}\min(\pi_i, \pi_{i+1})}"] & \boxed{\text{i+1}} \arrow[l, shift left, "\frac{1}{2}\min(\pi_{i+1},\pi_i)"]
\end{tikzcd}$$

- Lifted-Metropolis flows on $\mathcal{P}_n$ (with 'phantom' vertices):

$$\begin{tikzcd}[column sep=huge]
\boxed{\text{(i-1,+1)}} \arrow[r, "\frac{1}{2}\min(\pi_{i-1},\pi_i)"] & \boxed{\text{i,+1}} \arrow[r, "\frac{1}{2}\min(\pi_i,\pi_{i+1})"] \arrow[d, shift right, "\frac{1}{2}[\pi_i-\min(\pi_i,\pi_{i+1})]"'] & \boxed{\text{i+1,+1}} \\
\boxed{\text{i-1,-1}} & \boxed{\text{i,-1}} \arrow[l, "\frac{1}{2}\min(\pi_{i-1},\pi_i)"'] \arrow[u, shift right, "\frac{1}{2}[\pi_i-\min(\pi_{i-1},\pi_i)]"'] & \boxed{\text{i+1,-1}} \arrow[l, "\frac{1}{2}\min(\pi_i,\pi_{i+1})"']
\end{tikzcd}$$

Diaconis–Holmes–Neal (2000) Comments 4/4

- The ‘rejections $\rightarrow$ liftings’ idea works best for constant π_i.
- see Hildebrand (2004) for V-shaped. Hayes–Janes (2013) for general case.
- Appears impossible to generalize, but is not.

Lifting Markov Chains to Speed up Mixing

Fang Chen
Department of Mathematics
Yale University
fchen@math.yale.edu

László Lovász *
Department of Computer Science
Yale University
lovasz@cs.yale.edu

Igor Pak †
Department of Mathematics
Yale University
paki@math.yale.edu

Abstract

There are several examples where the mixing time of a Markov chain can be reduced substantially, often to about its square root, by "lifting", i.e., by splitting each state into several states. In several examples of random walks on groups, the lifted chain not only mixes better, but is easier to analyze.

We characterize the best mixing time achievable through lifting in terms of multicommodity flows. We show that the reduction to square root is best possible. If the lifted chain is time-reversible, then the gain is smaller, at most a factor of $\log(1/\pi_0)$, where π_0 is the smallest stationary probability of any state. We give an example showing that a gain of a factor of $\log(1/\pi_0)/\log\log(1/\pi_0)$ is possible.

1 Introduction

The estimation of the mixing time of finite Markov chains (the time needed for the chain to become approximately stationary) has emerged as a major issue in the design and analysis of various algorithms for sampling, enumeration, optimization, integration etc.

The research presented in this paper was motivated by the work of Diaconis, Holmes and Neal [5], who observed that certain non-reversible chains mix substantially faster than closely related reversible chains. We view their example in a different way: we represent a given chain as the "projection" of another chain, and analyze how this improves the mixing time.

Chen–Lovásc–Pak (1999) 2/3

2.4 Conductance and mixing

Standard results, first obtained by Jerrum and Sinclair [6], use conductance to bound mixing parameters. For reversible chains, the following bounds are well known. Let $1-\lambda_2$ be the eigenvalue gap, and define the *relaxation time* to be $\mathcal{L} = \frac{1}{1-\lambda_2}$. Then

$$\frac{1}{\Phi} \leq \mathcal{L} \leq \frac{8}{\Phi^2}, \tag{3}$$

and

$$\frac{1}{\Phi} \leq \mathcal{H} \leq \log \frac{1}{\pi_0} \frac{10}{\Phi^2}. \tag{4}$$

For general chains, there does not seem to be a standard way to define the eigenvalue gap, but similar bounds can be proved for the set-time $\mathcal{A}$.

Lemma 2.1 *For every Markov chain,*

$$\frac{1}{4\Phi} \leq \mathcal{A} \leq \frac{20}{\Phi^2} \tag{5}$$

and

$$\frac{1}{\Phi} \leq \mathcal{H} \leq \frac{3000}{\Phi^2} \log \frac{1}{\pi_0} \tag{6}$$

The proofs of these lemmas are omitted. The connection between multicommodity flows and mixing, in one direction, is established by the following lemma:

Département de Physique
École normale supérieure

3 Lifting and Collapsing

Let M and $\widehat{M}$ be two finite Markov chains with underlying sets V and $\widehat{V}$, respectively. We denote by $\widehat{\pi}$, $\widehat{p}$ etc. the stationary distributions, transition probabilities etc. in $\widehat{M}$.

We say that M is a *collapsing* of $\widehat{M}$, if there is a mapping $\widehat{V} \to V$ such that

$$\pi_v = \widehat{\pi}(f^{-1}(v)) = \sum_{i \in f^{-1}(v)} \widehat{\pi}_i$$

for every $v \in V(G)$, and

$$p_{vu} = \sum_{i \in f^{-1}(v), j \in f^{-1}(u)} \frac{\widehat{\pi}_i}{\widehat{\pi}(f^{-1}(v))} \widehat{p}_{ij}$$

for every pair $v, u \in V(G)$. We also say that $\widehat{M}$ is a *lifting* of M. For the random walk on an undirected graph, collapsing

Chen–Lovásc–Pak (1999) Comments 1/2

- Required: Mapping from $\hat{\Omega}$ (lifted sample space) to Ω that preserves stationary probability distribution.
- Required: Lifted transition matrix $\hat{P}$ that preserves flow.
- Optional: $\hat{\Omega} = \Omega \times \mathcal{L}$ (with $\mathcal{L}$: set of lifting variables).
- Optional:

$$\frac{\hat{\pi}(u,\sigma)}{\pi(u)} = \frac{\hat{\pi}(v,\sigma)}{\pi(v)} \quad \forall\ u, v \in \Omega; \forall\ \sigma \in \mathcal{L}. \tag{1}$$

- There are many liftings $\hat{P}$ for a given lifted sample space $\hat{\Omega}$.

Chen–Lovásc–Pak (1999) Comments 2/2

- Conductance = bottleneck ratio = Cheeger constant.
- Conductance lower bounds miraculous.
- Inequalities apply only to finite Markov chains.
- In event-driven algorithms, mixing and correlation times may not reflect computational effort.

Baik–Liu (2016) 1/3

J Stat Phys (2016) 165:1051–1085
DOI 10.1007/s10955-016-1665-y

CrossMark

TASEP on a Ring in Sub-relaxation Time Scale

Jinho Baik[1] · **Zhipeng Liu**[2]

Abstract Interacting particle systems in the KPZ universality class on a ring of size L with $O(L)$ number of particles are expected to change from KPZ dynamics to equilibrium dynamics at the so-called relaxation time scale $t = O(L^{3/2})$. In particular the system size is expected to have little effect to the particle fluctuations in the sub-relaxation time scale $1 \ll t \ll L^{3/2}$. We prove that this is indeed the case for the totally asymmetric simple exclusion process (TASEP) with two types of initial conditions. For flat initial condition, we show that the particle fluctuations are given by the Airy_1 process as in the infinite TASEP with flat initial condition. On the other hand, the TASEP on a ring with step initial condition is equivalent to the periodic TASEP with a certain shock initial condition. We compute the fluctuations explicitly both away from and near the shocks for the infinite TASEP with same initial condition, and then show that the periodic TASEP has same fluctuations in the sub-relaxation time scale.

Département de Physique
École normale supérieure

Baik–Liu (2016) 2/3

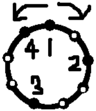
J Stat Phys (2016) 165:1051–1085
DOI 10.1007/s10955-016-1665-y

CrossMark

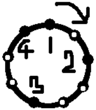
TASEP on a Ring in Sub-relaxation Time Scale

Jinho Baik[1] · **Zhipeng Liu**[2]

Abstract Interacting particle systems in the KPZ universality class on a ring of size L with $O(L)$ number of particles are expected to change from KPZ dynamics to equilibrium dynamics at the so-called relaxation time scale $t = O(L^{3/2})$. In particular the system size is expected to have little effect to the particle fluctuations in the sub-relaxation time scale $1 \ll t \ll L^{3/2}$. We prove that this is indeed the case for the totally asymmetric simple exclusion process (TASEP) with two types of initial conditions. For flat initial condition, we show that the particle fluctuations are given by the Airy$_1$ process as in the infinite TASEP with flat initial condition. On the other hand, the TASEP on a ring with step initial condition is equivalent to the periodic TASEP with a certain shock initial condition. We compute the fluctuations explicitly both away from and near the shocks for the infinite TASEP with same initial condition, and then show that the periodic TASEP has same fluctuations in the sub-relaxation time scale.



SSEP



TASEP

Baik–Liu (2016) 3/3

J Stat Phys (2016) 165:1051–1085
DOI 10.1007/s10955-016-1665-y

CrossMark

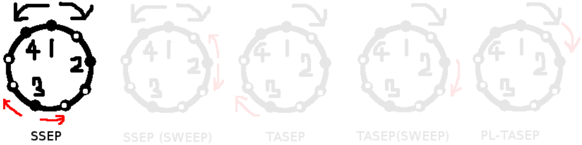
TASEP on a Ring in Sub-relaxation Time Scale

Jinho Baik[1] · Zhipeng Liu[2]

Abstract Interacting particle systems in the KPZ universality class on a ring of size L with $O(L)$ number of particles are expected to change from KPZ dynamics to equilibrium dynamics at the so-called relaxation time scale $t = O(L^{3/2})$. In particular the system size is expected to have little effect to the particle fluctuations in the sub-relaxation time scale $1 \ll t \ll L^{3/2}$. We prove that this is indeed the case for the totally asymmetric simple exclusion process (TASEP) with two types of initial conditions. For flat initial condition, we show that the particle fluctuations are given by the Airy$_1$ process as in the infinite TASEP with flat initial condition. On the other hand, the TASEP on a ring with step initial condition is equivalent to the periodic TASEP with a certain shock initial condition. We compute the fluctuations explicitly both away from and near the shocks for the infinite TASEP with same initial condition, and then show that the periodic TASEP has same fluctuations in the sub-relaxation time scale.

SSEP TASEP

Baik–Liu (2016) Comments 1/5



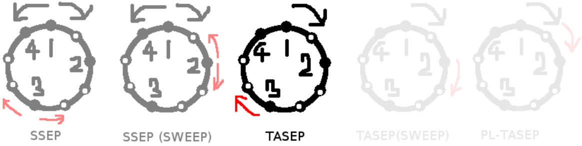
- Symmetric simple exclusion process (alias: 'Local Metropolis')
- $\Omega^{\mathrm{SSEP}} = \{x_1 < x_2, \ldots, < x_N\}$ with $x_i \in \{1, \ldots, n\}$. PBC.
- P (transition matrix): nearest-neighbor, one per unit of time.
- Mixing time: $\mathcal{O}\left(N^3 \log N\right)$ (rigorous: Lacoin 2014).
- Continuous versions exist.

Département de Physique
École normale supérieure

Baik–Liu (2016) Comments 2/5

- Particle-lifted SSEP (sweep)
- $\hat{\Omega}^{\text{SSEP-Sweep}} = \hat{\Omega}^{\text{SSEP}} \times \mathcal{N}$ with $\mathcal{N} = \{1, \ldots, N\}$.
- Mixing time: $\mathcal{O}\left(N^3 \log N\right)$ (numerics: Kapfer, Krauth 2017).
- OK, as any 'particle lifting' of a reversible Markov chain.

Baik–Liu (2016) Comments 3/5



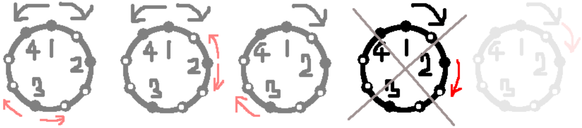
- Totally asymmetric simple exclusion process.
- $\hat{\Omega}^{\text{TASEP}} = \Omega^{\text{SSEP}} \times \mathcal{D}$ with $\mathcal{D} = \{-1, +1\}$.
- Displacement-lifted SSEP.
- Lifted-sample-space halving applies.
- Mixing time: $\mathcal{O}\left(N^{5/2}\right)$ (Baik–Liu 2016)

Département de Physique
École normale supérieure

Baik–Liu (2016) Comments 4/5

- Particle-lifted TASEP (Sweep).
- $\hat{\Omega}^{\text{TASEP-Sweep}} = \Omega^{\text{SSEP}} \times \mathcal{D} \times \mathcal{N}$.
- Lifted-sample-space halving applies.
- Violates global balance (Kapfer–Krauth 2017).

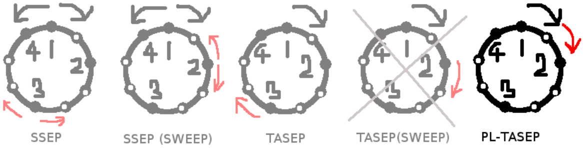
Baik–Liu (2016) Comments 4/5



- Particle-lifted TASEP (Sweep).
- $\hat{\Omega}^{\text{TASEP-Sweep}} = \Omega^{\text{SSEP}} \times \mathcal{D} \times \mathcal{N}$.
- Lifted-sample-space halving applies.
- Violates global balance (Kapfer–Krauth 2017).

Département de Physique
École normale supérieure

Baik–Liu (2016) Comments 5/5



- Particle-lifted TASEP.
- Particle-lifted displacement-lifted SSEP.
- $\hat{\Omega}^{\text{PL-TASEP}} = \Omega^{\text{SSEP}} \times \mathcal{D} \times \mathcal{N}$.
- Sample-space halving applies. Resampling essential.
- Mixing time $\mathcal{O}\left(N^2 \log N\right)$ (Kapfer–Krauth 2017, rigorous: Lei–Krauth 2018)
- The coupon-collector log can be eliminated, and $\mathcal{O}\left(N^2\right)$ mixing time reached.

Conclusions

- The 'particle-on-a-path-graph lifting' (Diaconis–Holmes–Neal 2000) illustrates the 'rejections $\to$ liftings' miracle.
- There's more to liftings than 'momenta'.
- Infinite speedups for $N \to \infty$ can be carried over to interactive-particle systems (TASEP and ECMC).
- Must be super careful. Design principles would be useful.
- Conductance arguments must be extended for event-driven continuous MCMC.

Département de Physique
École normale supérieure