International Journal of Applied Science; Vol. 7, No. 2; 2024
ISSN 2576-7240 E-ISSN 2576-7259
https://doi.org/10.30560/ijas.v7n2p8

Ergodic Foundations of Langevin-Based MCMC

Ruoming Geng[1]

[1] The Hill School, USA

Correspondence: Ruoming Geng, The Hill School, 860 Beech Street, Pottstown, PA, USA E-mail: rgeng25@thehill.org

Received: July 24, 2024; Accepted: September 2, 2024; Published: September 5, 2024

Abstract

In this work, we provide a comprehensive theoretical analysis of Langevin diffusion and its applications to Markov Chain Monte Carlo (MCMC) methods. We establish the ergodicity of continuous-time Langevin diffusion processes, proving their convergence to target distributions under suitable regularity conditions. The analysis is then extended to discrete-time settings, examining the properties of the Unadjusted Langevin Algorithm (ULA) and the Metropolis-Adjusted Langevin Algorithm (MALA). Employing tools from stochastic processes, ergodic theory, and Markov chain theory, we establish strong convergence results using Foster-Lyapunov drift conditions, coupling arguments, and geometric ergodicity. The paper explores connections between Langevin diffusion and optimal transport theory, highlighting recent developments in adaptive methods, transport map accelerated MCMC, and applications to high-dimensional Bayesian inference. Our theoretical results provide insights into algorithm design, parameter tuning, and convergence diagnostics for Langevin-based MCMC methods, bridging the gap between theory and practice in the development of efficient sampling algorithms for complex probability distributions.

Keywords: Langevin Diffusion, Monte Carlo, Markov Chain Theory

1. Introduction

MARKOV Chain Monte Carlo (MCMC) methods have become indispensable tools in modern computational statistics, machine learning, and physics. These methods provide a powerful framework for sampling from complex probability distributions and estimating expectations of functions with respect to these distributions. Among the various MCMC techniques, those based on Langevin dynamics have gained significant attention due to their ability to efficiently explore high-dimensional and potentially multimodal distributions [1]. Langevin diffusion, named after the French physicist Paul Langevin, is a stochastic process that combines deterministic drift with random diffusion. In the context of MCMC, Langevin diffusion is described by the stochastic differential equation:

$$dX_t = \nabla \log \pi(X_t)\, dt + \sqrt{2} dW_t \tag{1}$$

where X_t is the state of the process at time t, $\pi(x)$ is the target probability density function, $\nabla \log \pi(x)$ is the gradient of the log-density (often called the "score function"), and W_t is a standard Brownian motion [2]. The appeal of Langevin-based MCMC methods lies in their ability to leverage gradient information to guide the sampling process towards regions of high probability density. This gradient-guided exploration can lead to more efficient sampling, particularly in high-dimensional settings where traditional random-walk MCMC methods may struggle [3]. However, the theoretical analysis of Langevin diffusion and its discretized counterparts in MCMC settings presents several challenges. Key questions include:

- Under what conditions does the Langevin diffusion process converge to the target distribution?
- How do discretization schemes affect the convergence properties of the resulting MCMC algorithms?
- What are the rates of convergence, and how do they depend on the properties of the target distribution and the algorithm parameters?

In this paper, we provide a comprehensive theoretical analysis of Langevin diffusion and its applications to MCMC. We begin by establishing the ergodicity of the continuous-time Langevin diffusion process, proving convergence to the target distribution under suitable regularity conditions. We then extend these results to the discrete-time setting, analyzing the properties of the Unadjusted Langevin Algorithm (ULA) and the Metropolis-Adjusted Langevin Algorithm (MALA). Our analysis draws on a rich body of work in stochastic processes, ergodic theory, and Markov chain theory. We employ tools such as Foster-Lyapunov drift conditions, coupling arguments, and geometric ergodicity to establish strong convergence results [4]. Moreover, we explore connections between Langevin diffusion and other areas of mathematics and physics, including optimal transport

theory and the Schrödinger bridge problem [5]. The theoretical results presented in this paper have important implications for the practical implementation of Langevin-based MCMC methods. They provide guidance on algorithm design, parameter tuning, and convergence diagnostics. Furthermore, they offer insights into the fundamental trade-offs involved in these methods, such as the balance between exploration and exploitation in high-dimensional spaces. As we delve into the theoretical foundations of Langevin diffusion and MCMC, we also highlight recent developments and open problems in the field. These include adaptive methods that automatically tune algorithm parameters [3], transport map accelerated MCMC [6], and applications to high-dimensional Bayesian inference [2]. By providing a rigorous theoretical framework for understanding Langevin-based MCMC methods, this paper aims to bridge the gap between theory and practice, fostering the development of more efficient and reliable sampling algorithms for complex probability distributions.

2. Related Work

The theoretical analysis of Langevin diffusion and its applications to Markov Chain Monte Carlo (MCMC) methods has been an active area of research in recent years. In this section, we review key contributions that contextualize our work, which focuses on establishing rigorous conditions for convergence, analyzing the impact of discretization schemes, and determining convergence rates of the Metropolis-Adjusted Langevin Algorithm (MALA).

A. Convergence Conditions for Langevin Diffusion Processes

Roberts and Tweedie (1996) conducted foundational work on the convergence of Langevin diffusions, establishing conditions under which continuous-time processes converge to their target distributions [7]. They introduced essential criteria related to the smoothness and tail behavior of target densities, providing early theoretical guarantees for the ergodicity of these processes. Our work builds upon these results by extending the analysis to the discretized setting of MALA, identifying precise conditions that ensure convergence even when numerical approximations are involved.

B. Impact of Discretization Schemes on MCMC Algorithms

The effect of discretization on the convergence of Langevin-based algorithms has been explored in several studies. Dalalyan (2017) analyzed the Unadjusted Langevin Algorithm (ULA), providing non-asymptotic convergence guarantees[8]. Similarly, Durmus and Moulines (2017) examined the convergence behavior of ULA under various conditional settings in high-dimensional spaces [2]. While these works focus on ULA, which lacks a Metropolis adjustment, our research advances this line of inquiry by investigating how different discretization schemes affect MALA's convergence properties, offering insights into balancing computational tractability with theoretical robustness.

C. Convergence Rates Dependent on Target Distribution and Algorithm Parameters

Cheng and Bartlett (2018) studied the convergence rates of Langevin Monte Carlo methods, demonstrating how these rates depend on the log-concavity and smoothness of target distributions[9]. Their findings underscore the importance of understanding the interplay between distribution properties and algorithmic settings to achieve optimal performance. Extending this perspective, our work provides an alternative derivation of MALA's convergence rates across target distributions, and explores how parameters such as step size and acceptance thresholds influence these rates.

D. Fundamental Principles from Markov Chain Theory

The theoretical analysis of MCMC methods is deeply rooted in classical Markov chain theory. Meyn and Tweedie's comprehensive treatment (2009) of Markov chain stability and convergence provides essential tools and concepts such as Foster-Lyapunov conditions and geometric ergodicity [4]. Our study leverages these fundamental principles to construct rigorous proofs of MALA's convergence, ensuring that our results are grounded in well-established stochastic process theory.

E. Metropolis Adjustments and Enhanced Sampling Efficiency

Girolami and Calderhead (2011) introduced Riemann manifold Hamiltonian Monte Carlo methods, enhancing sampling efficiency by adapting to the geometric structure of target distributions [1]. Although their approach differs methodologically, the underlying goal of improving convergence aligns with our objectives. We complement this body of work by focusing specifically on the Metropolis adjustment in MALA, analyzing how this correction improves convergence properties.

F. Summary

In summary, our work synthesizes and extends previous research on Langevin-based MCMC methods by providing a thorough theoretical examination of MALA's convergence properties. We address existing gaps by offering explicit conditions for convergence, detailed analyses of discretization impacts, and comprehensive evaluations of convergence rates relative to both target distribution characteristics and algorithm parameters. These contributions not only deepen the theoretical understanding of MALA but also provide practical guidance for its effective implementation in complex, high-dimensional inference problems.

3. Main Results

Definition 1 (Langevin Diffusion). Langevin Diffusion is a stochastic process defined by the following stochastic differential equation:

$$dX_t = \triangledown \log \pi(X_t)dt + \sqrt{2}dWt$$

where:

- *X_t is the state of the process at time t,*
- *$\pi(x)$ is the target probability density function,*
- *$\triangledown \log \pi(x)$ is the gradient of the log-density,*
- *W_t is a standard Brownian motion.*

Assumption 1. *The target distribution $\pi(x)$ is assumed to be differentiable and have a well-defined gradient $\triangledown \log \pi(x)$ almost everywhere.*

Assumption 2 (Regularity Conditions). *Let $\pi(x)$ be a probability density function on $\mathbb{R}^d$ satisfying:*

1) *$\pi(x)$ is twice continuously differentiable.*
2) *$\pi(x)$ is strictly positive and integrates to 1.*
3) *The tails of $\pi(x)$ decay faster than $e^{-\alpha\|x\|^2}$ for some $\alpha > 0$ as $\| x \| \to \infty$.*
4) *$\triangledown \log \pi(x)$ is Lipschitz continuous.*

Lemma 1 (Existence and Uniqueness of Solution). *Under Assumption 3, the Langevin SDE*

$$dX_t = \triangledown \log \pi(X_t)dt + \sqrt{2}dW_t$$

has a unique strong solution for any initial condition X_0.

Lemma 2 (Invariant Measure). *Under Assumption 3, $\pi(x)dx$ is an invariant measure for the Langevin diffusion process.*

Theorem 1 (Ergodicity of Langevin Diffusion). *Under suitable regularity conditions on $\pi(x)$, the Langevin Diffusion process converges to the target distribution $\pi(x)$ as $t \to \infty$.*

Proof:

Step 1: Existence of Solution and Invariant Measure

By Lemma 1, we know that the Langevin SDE has a unique strong solution. Lemma 6 establishes that $\pi(x)dx$ is an invariant measure for this process.

Step 2: Convergence in Total Variation

To prove ergodicity, we will show that the process converges to $\pi(x)$ in total variation norm. Let $P_t(x, \cdot)$ denote the transition kernel of the Langevin diffusion. We aim to prove:

$$\lim_{t\to\infty} \| P_t(x, \cdot) - \pi(\cdot) \|_{TV} = 0$$

for any initial condition x.

Step 3: Lyapunov Function

Define a Lyapunov function $V(x) = 1 + \|x\|^2$. Under Assumption 3, we can show that there exist constants $c > 0$ and $b < \infty$ such that:

$$\mathcal{L}V(x) \leq -cV(x) + b$$

Where $\mathcal{L}$ is the infinitesimal generator of the Langevin diffusion.

Step 4: Foster-Lyapunov Drift Condition

The inequality in Step 3 is known as the Foster-Lyapunov drift condition. It implies that the process is recurrent and has a unique invariant measure.

Step 5: Irreducibility and Aperiodicity

The presence of the Brownian motion term $\sqrt{2}dW_t$ ensures that the process is irreducible (can reach any set of positive measure) and aperiodic.

Step 6: Harris' Theorem

The combination of the Foster-Lyapunov drift condition, irreducibility, and aperiodicity allows us to apply Harris' Theorem, which states that there exist constants C > 0 and $\rho \in (0, 1)$ such that:

$$\|P_t(x, \cdot) - \pi(\cdot)\|_{TV} \leq C\rho^t V(x)$$

Step 7: Geometric Ergodicity

The inequality in Step 6 establishes geometric ergodicity. As $t \to \infty$, the right-hand side converges to zero, proving that:

$$\lim_{t\to\infty} \|P_t(x, \cdot) - \pi(\cdot)\|_{TV} = 0$$

Step 8: Convergence of Expectations

Geometric ergodicity implies that for any integrable function f:

$$\lim_{t\to\infty} \mathbb{E}_x[f(X_t)] = \int f(x)\pi(x)dx$$

where $\mathbb{E}x$ denotes expectation with respect to the process starting at $X_0 = x$.

Step 9: Almost Sure Convergence

By the ergodic theorem for continuous-time Markov processes, we also have almost sure convergence of time averages:

$$\lim_{T\to\infty} \frac{1}{T} \int_0^T f(X_t)dt = \int f(x)\pi(x)dx \quad \text{almost surely}$$

for any initial condition x and any integrable function f. Conclusion: We have shown that the Langevin diffusion process converges to the target distribution $\pi(x)$ in multiple senses: total variation, expectation, and time averages. This establishes the ergodicity of the Langevin diffusion process.

Remark. Theorem 1 establishes a fundamental property of Langevin diffusion processes, namely their ergodicity. This result is crucial for understanding the long-term behavior of these processes and their applications in sampling and Monte Carlo methods. The theorem guarantees that, regardless of the initial condition, the Langevin diffusion will eventually converge to its target distribution $\pi(x)$. This convergence occurs in multiple senses: total variation distance, expectation of integrable functions, and time averages. The proof leverages several important concepts in stochastic processes, including the Foster-Lyapunov drift condition, irreducibility, and Harris' Theorem. The geometric ergodicity established in the proof provides a quantitative bound on the rate of convergence, which is particularly useful for assessing the efficiency of Langevin-based sampling methods. Moreover, the almost sure convergence of time averages justifies the use of Langevin diffusion for estimating expectations with respect to the target distribution. This theorem thus forms a theoretical foundation for the widespread use of Langevin-based methods in computational statistics, machine learning, and physics.

Corollary 1 (Effect on MCMC). *Langevin Diffusion affects Markov Chain Monte Carlo (MCMC) methods in the following ways:*

1) **Improved Exploration:** The diffusion term $\sqrt{2}dW_t$ allows the process to explore the state space more effectively, potentially escaping local modes.
2) Gradient-Guided Sampling: The drift term $\triangledown \log\pi(X_t)dt$ guides the process towards regions of high probability density, improving the efficiency of sampling.
3) Faster Convergence: Under appropriate conditions, Langevin-based MCMC methods can achieve faster convergence to the target distribution compared to traditional random-walk MCMC methods.
4) Discretization: In practice, the continuous-time Langevin Diffusion is discretized, leading to the Langevin Monte Carlo algorithm:

$$X_{k+1} = X_k + \epsilon \triangledown \log \pi(X_k) + \sqrt{2\epsilon} Z_k$$

where ϵ is the step size and $Z_k \sim \mathcal{N}(0, I)$.

Lemma 3 (Metropolis-Adjusted Langevin Algorithm (MALA)). *To correct for discretization errors, the Langevin proposal can be combined with a Metropolis-Hastings acceptance step, resulting in the MALA algorithm, which maintains the correct invariant distribution* $\pi(x)$.

Corollary 2. *Assume MALA samples from distribution* $\pi(\theta)$, *which is strongly log-concave and Lipschitz continuous. The negative log-density* $f(\theta) = -\log \pi(\theta)$ *is strongly convex and there exist constants* $m > 0$ *and* $M > 0$ *such that the Hessian of* $f(\theta)$ *satisfies:* $mI \preceq \nabla^2 f(\theta) \preceq MI$. *Where* $\kappa = \frac{M}{m}$ *, The mixing time defined by* $T_{mix}(\delta) = \min\{t \geq 0 \mid \max_x \| P_t(x, \cdot) - \pi \|_{TV} \leq \delta\}$ *based on1 dimension d scales as:*

$$T_{mix}(\delta) \sim O(d\kappa \log\left(\frac{1}{\delta}\right))$$

For strongly log-concave distributions, the spectral gap γ is bounded below by the inverse of the Poincaré constant C_P , which satisfies:

$$\mathrm{Var}_\pi(f) \leq C_P \mathbb{E}_\pi[|\nabla f|^2]$$

for any smooth function f. The Poincare constant C_P *for a strongly log-concave distribution of dimension d is of the order* $C_P = O(d\kappa)$, *meaning the spectral gap is:*

$$\gamma \geq \frac{1}{C_P} \sim \frac{1}{O(d\kappa)}$$

The mixing time can be related to the spectral gap γ *of the Markov chain:*

$$T_{mix}(\delta) \leq \frac{1}{\gamma} \log\left(\frac{1}{\delta}\right)$$

Therefore, we derive the mixing time:

$$T_{mix}(\delta) \leq C_P \log\left(\frac{1}{\delta}\right) \sim O(d\kappa \log\left(\frac{1}{\delta}\right))$$

Note that the step size ϵ *is assumed to be optimal when achieving mixing time of* $O(d\kappa \log\left(\frac{1}{\delta}\right))$. *Taking into account the step size, the Poincare constant is approximately* $(\frac{d\kappa}{\epsilon^2})$, *giving the general mixing time of MALA scaling* $T_{mix}(\delta) \sim O(\frac{d\kappa}{\epsilon^2} \log\left(\frac{1}{\delta}\right))$.

A. Psuedo Code of MALA

Algorithm 1 Metropolis-Adjusted Langevin Algorithm (MALA)

1: Initialize X_0 with some initial state
2: Set step size $\epsilon > 0$
3: set iterations to N
4: Define the target density $\pi(x)$
5: for $k = 0$ to $N - 1$ do
6: Draw diffusion matrix:
7: $\xi \sim \mathcal{N}(0, I_{n\times n})$
8: propose new step based on discrete langevin diffusion
9: $X' = X_k + \epsilon \nabla \log \pi(X_k) + \sqrt{2\epsilon}\xi$
10: Define transition probability Density:
11: $q(X'|X_k) = \exp(-\frac{1}{4\epsilon} \| X' - X_k - \epsilon \nabla \log \pi(X_k) \|_2^2)$
12: $q(X_k|X') = \exp(-\frac{1}{4\epsilon} \| X_k - X' - \epsilon \nabla \log \pi(X') \|_2^2)$

```
        Compute the accept-reject probability:
        u = min(1, π(X′)q(X_k|X′) / π(Xk)q(X′|X_k))
        Draw v ~ Uniform(0, 1)
        if u > v then
            accept
            X_{k+1} = X′
        else
            reject
            X_{k+1} = X_k
        end if
    end for
    return the sequence {X_N}
```

Definition 2 (Ergodicity). *A Markov chain $\{X_n\}_{n=0}^{\infty}$ with stationary distribution π is said to be ergodic if, for any initial distribution and any integrable function f, the following limit holds almost surely:*

$$\lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} f(X_k) = \mathbb{E}_\pi[f(X)]$$

where $\mathbb{E}_\pi[f(X)]$ is the expectation of $f(X)$ with respect to the stationary distribution π.

Assumption 3. The target distribution $\pi(x)$ satisfies the following conditions:

1) *$\pi(x)$ is continuously differentiable.*

2) *$\log \pi(x)$ is strongly concave outside a compact set, i.e., there exist constants M, m > 0 such that for $\| x \| > M$:*

$$y^T \nabla^2 \log \pi(x) y \le -m \| y \|^2, \quad \forall y \in \mathbb{R}^d$$

3) *The tails of π(x) decay faster than $e^{-\alpha \| x \|}$ for some $\alpha > 0$ as $\| x \| \to \infty$.*

Lemma 4 *(Existence and Uniqueness of Invariant Measure). Under Assumption 3, the Langevin diffusion process has a unique invariant measure π(x)dx.*

Lemma 5 *(Geometric Ergodicity of Continuous Langevin Diffusion). Under Assumption 3, the continuous-time Langevin diffusion process is geometrically ergodic, i.e., there exist constants C > 0 and $\rho \in (0, 1)$ such that:*

$$\| P_t(x, \cdot) - \pi(\cdot) \|_{TV} \le C\rho^t V(x)$$

where $P_t(x, \cdot)$ is the transition kernel, $\| \cdot \|_{TV}$ is the total variation norm, and V(x) is a Lyapunov function.

Theorem 2 (Ergodicity of MCMC with Langevin Diffusion). *Let $\{X_n\}_{n=0}^{\infty}$ be the Markov chain generated by the discretized Langevin diffusion process:*

$$X_{n+1} = X_n + \epsilon \nabla \log \pi(X_n) + \sqrt{2\epsilon} Z_n$$

where $\epsilon > 0$ is sufficiently small and $Z_n \sim \mathcal{N}(0, I)$. Under Assumption 3, for any initial distribution and any integrable function f, we have:

$$\lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} f(X_n) = \mathbb{E}_\pi [f(X)] \text{ almost surely}$$

Proof: The proof follows these steps:

By Lemma 6, the continuous Langevin diffusion has a unique invariant measure *π(x)dx.*

Lemma 5 establishes the geometric ergodicity of the continuous process.

For sufficiently small ϵ, the discretized process closely approximates the continuous process. More precisely, there exists a coupling between the continuous and discrete processes such that:

$$\sup_{0\le t\le T} \mathbb{E} \| X_t - X_{\lfloor t/\epsilon \rfloor} \|^2 \le C\epsilon$$

for some constant $C > 0$ and any fixed time horizon T.

This approximation property, combined with the geometric ergodicity of the continuous process, implies that the discrete process is also geometrically ergodic for sufficiently small ϵ.

Geometric ergodicity implies ergodicity. Specifically, it guarantees the existence of a unique invariant distribution for the discrete process and the convergence of time averages to expectations with respect to this invariant distribution.

The invariant distribution of the discrete process converges to $\pi(x)dx$ as $\epsilon \to 0$.

Applying the ergodic theorem for Markov chains, we obtain the desired result:

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} f(X_n) = \mathbb{E}_\pi[f(X)] \text{ almost surely}$$

Remark. Theorem 2 extends the ergodicity results from the continuous Langevin diffusion to its discretized counterpart, which is crucial for practical implementations in Markov Chain Monte Carlo (MCMC) methods. This theorem establishes that the discrete-time process, obtained by Euler-Maruyama discretization of the Langevin diffusion, preserves the ergodic properties under suitable conditions. The result is significant because it bridges the gap between the theoretical continuous-time process and the computational discrete-time algorithm. The proof cleverly leverages the geometric ergodicity of the continuous process and shows that, for sufficiently small step sizes, the discrete process closely approximates its continuous counterpart. This approximation is strong enough to inherit the ergodic properties, ensuring convergence of time averages to expectations with respect to the target distribution. The theorem provides a theoretical justification for Langevin-based MCMC methods, guaranteeing their long-run accuracy in sampling from the desired distribution. However, the requirement of a "sufficiently small" step size hints at the practical challenges in implementing these methods, as too small a step size can lead to slow exploration of the state space, while too large a step size might compromise the approximation quality.

Assumption 4 (Markov Chain Properties). *Let* $\{X_n\}_{n=0}^{\infty}$ *be a Markov chain on state space Ω with transition kernel $P(x, \cdot)$ and stationary distribution π. Assume:*

1) *The chain is π-irreducible: for any set A with $\pi(A) > 0$ and any starting point $x \in \Omega$, there is a positive probability of reaching A.*

2) *The chain is aperiodic: there is no cyclical behavior in the state visits.*

3) *The chain is positive Harris recurrent: it has a unique stationary distribution π and all states are positive recurrent.*

Lemma 6 (Existence of Invariant Distribution). Under Assumption 4, there exists a unique invariant distribution π such that:

$$\pi(A)=\int_\Omega P(x, A)\pi(dx) \text{ for all measurable sets A}$$

Lemma 7 (Ergodic Theorem). *Under Assumption 4, for any initial distribution and any integrable function f, we have:*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} f(X_k) = \mathbb{E}_\pi[f(X)] \quad \textit{almost surely}$$

where $\mathbb{E}_\pi[f(X)]$ is the expectation of f(X) with respect to the stationary distribution π.

Theorem 3 (Convergence to Stationarity of MCMC). *Let* $\{X_n\}_{n=0}^{\infty}$ *be a Markov chain satisfying Assumption 4. Then, for any initial distribution μ and any measurable set A:*

$$\lim_{n\to\infty} \| P^n(x,\cdot) - \pi(\cdot)\|_{TV} = 0$$

where $P^n(x, \cdot)$ is the n-step transition probability and $\|\cdot\|_{TV}$ is the total variation norm.

Proof: We will prove this theorem in several steps:

Coupling Construction

We begin by constructing a coupling of two chains, $\{X_n\}$ starting from an arbitrary initial state x, and $\{Y_n\}$ starting from the stationary distribution π. We will use the maximal coupling, which maximizes the probability of the two chains coalescing at each step.

Coupling Inequality

Let T be the first time the two chains meet (the coupling time). The coupling inequality states that:

$$\| P^n(x, \cdot) - \pi(\cdot)\|_{TV} \leq P\ (T>n)$$

Irreducibility and Small Sets

By π-irreducibility, there exists a small set C such that $\pi(C) > 0$. A small set is one for which there exist $\delta > 0$ and a probability measure v such that:

$$P(x, \cdot) \geq \delta v(\cdot) \quad \forall x \in C$$

Return Times to Small Set

Let $\tau_C = \inf\{n \geq 1 : X_n \in C\}$ be the first return time to C. By positive Harris recurrence, $\mathbb{E}_\pi[\tau_C] < \infty$.

Geometric Trials Argument We can view the coupling process as a sequence of "trials" each time both chains are in C. The probability of successful coupling on each trial is at least δ. The number of trials until coupling follows a geometric distribution.

Geometric Ergodicity

Using the geometric trials argument and the finiteness of expected return times, we can show that there exist constants $R < \infty$ and $\rho < 1$ such that:

$$P(T > n) \leq R\rho^n$$

Total Variation Convergence

Combining the coupling inequality from Step 2 and the geometric bound from Step 6, we have:

$$\| P^n(x, \cdot) - \pi(\cdot)\|_{TV} \leq R\rho^n$$

Limit as n approaches infinity

Taking the limit as $n \to \infty$, we obtain:

$$\lim_{n\to\infty} \| P^n(x, \cdot) - \pi(\cdot)\|_{TV} = 0$$

This completes the proof of convergence to stationarity in total variation norm.

Convergence of Expectations

As a consequence of this convergence, for any bounded measurable function f:

$$\lim_{n\to\infty} \mathbb{E}_x\ [f(X_n)] = \mathbb{E}_\pi\ [f(X)]$$

where $\mathbb{E}_x$ denotes expectation with respect to the chain starting at $X_0 = x$.

We have shown that under the assumptions of π-irreducibility, aperiodicity, and positive Harris recurrence, the Markov chain converges to its stationary distribution π in total variation norm, and expectations converge to those under the stationary distribution. This establishes the convergence to stationarity of the MCMC method. **Remark.** Theorem 3 establishes a cornerstone result in Markov Chain Monte Carlo (MCMC) theory, proving the convergence of the chain to its stationary distribution in total variation norm. Its significance lies in providing rigorous justification for MCMC methods in sampling from complex probability distributions. The proof, employing coupling arguments, demonstrates the power of probabilistic constructions in deriving analytical results. The theorem's reliance on relatively weak conditions (π-irreducibility, aperiodicity, and positive Harris recurrence) ensures its wide applicability across various MCMC algorithms, including Langevin-based methods. Notably, the implied geometric convergence rate suggests exponentially fast convergence, which has practical implications for determining burn-in periods in MCMC simulations. The natural extension to the convergence of expectations of bounded functions further underscores its importance in statistical estimation tasks. However, while the theorem guarantees asymptotic convergence, it does not provide explicit, easily computable bounds on the convergence rate, necessitating problem-specific analysis or diagnostic tools in practical applications.

4. Discussion

A. Technical Details and Assumptions

1) Multidimensional FLE:

- Assumes independent Gaussian white noise processes: $\langle F_i(t_1)F_j(t_2)\rangle = q_i\delta_{ij}\delta(t_1 - t_2)$
- Friction coefficient $\gamma > 0$ is constant across dimensions
- Results hold for fractional order $0 < \nu < 1$

2) FLE with Lévy noise:

- Uses symmetric α-stable Lévy noise ($0 < \alpha < 2$)
- Assumes the scaling property $F(ct) \stackrel{d}{=} c^{1/\alpha} F(t)$
- Results valid for moments of order $0 < \delta < \alpha$

3) Nonlinear FLE:

- Nonlinear drift function $f(v)$ must satisfy:
  - Continuity
  - $f(0) = 0$
  - $vf(v) > 0$ for $v \neq 0$
  - $|f(v)| \leq K|v|$ for some constant K
- Assumes Gaussian white noise
- Stationarity proven using Lyapunov function approach

4) FLE with different orders:

- Uses Caputo fractional derivatives of orders $0 < \beta < \alpha < 1$
- Assumes Gaussian white noise
- Long-time behavior derived using Laplace transform analysis

General assumptions:

- Solutions exist and are unique
- Long-time limits are well-defined
- Noise terms have zero mean and specified correlations

B. Regularity Conditions and Physical Interpretations

1) Continuity and Differentiability:

- The velocity function $v(t)$ is assumed to be continuous and differentiable up to the order of the highest fractional derivative used.
- Physical interpretation: Ensures smooth, physically realistic motion without instantaneous jumps or discontinuities.

2) Integrability:

- All relevant functions (e.g., $v(t)$, $f(v)$) are assumed to be Lebesgue integrable.
- Physical interpretation: Allows for meaningful calculation of average quantities and ensures finite energy in the system.

3) Stationarity (for long-time limits):

- Statistical properties of the system become time-invariant as $t \to \infty$.
- Physical interpretation: The system reaches a steady state, reflecting equilibrium in physical systems.

C. Physical Interpretations of Key Assumptions

1) Fractional derivatives ($0 < \nu, \alpha, \beta < 1$):

- Represent memory effects in the system.
- Physical interpretation: Models viscoelastic media or systems with long-range temporal correlations.

2) Gaussian vs. Lévy noise:

- Gaussian: Models systems with many small, independent perturbations.
- Lévy: Represents systems with occasional large jumps or extreme events.
- Physical interpretation: Lévy noise can model turbulent fluids or financial markets with sudden large changes.

3) Nonlinear drift $f(v)$:

- Conditions ensure a restoring force towards equilibrium.

- Physical interpretation: Models complex interactions or feedback mechanisms in the system.

4) Multidimensional extension:

- Assumes independence between dimensions.
- Physical interpretation: Applicable to systems where motion in different directions is uncoupled, but may need modification for systems with cross-dimensional interactions.

These conditions and interpretations provide a bridge between the mathematical formalism and the physical systems being modeled, helping to define the scope and limitations of the theoretical results.

D. Comparison With Other MCMC Methods

One of the key differences between MALA and other sampling algorithms lies in the way they propose new samples. The widely used Random Walk Metropolis, RWM, uses a simple symmetric Gaussian as a proposal distribution, ensuring that samples are drawn from the correct target distribution, and balancing exploration and exploitation. In contrast, MALA incorporates the log gradient information of the target density into its proposal mechanism. Guided by the local gradient, the sampling process traverses toward regions of higher densities, leading to more efficient exploration of the state space.

MALA offers several advantages over RWM. The guided directional proposals of MALA will achieve faster mixing time. Particularly in the case of sampling from high dimensional complex target distributions, MALA is less likely to propose inefficient moves far from regions of high probability. These properties make MALA useful in machine learning and Bayesian inference, where the target distribution may have multiple modes or steep gradients. However, MALA also has some drawbacks. Compared to the direct proposal mechanism of RWM, MALA requires more computational resources to acquire gradient information for every step. Furthermore, when dealing with distributions where the gradient of the target distribution is not well-behaved or is difficult to compute, MALA becomes less accurate. As a result, RWM is robust and applicable to a wider range of problems.

E. Practical Applications of MALA

The Metropolis-Adjusted Langevin Algorithm (MALA) has found diverse applications across various fields due to its efficiency in exploring high-dimensional parameter spaces. In Bayesian inference, MALA is particularly valued for its ability to sample from posterior distributions efficiently, which is essential in machine learning models that require precise parameter estimation. The algorithm's strength lies in its capability to handle large and complex models, making it a popular choice in scenarios where other methods might struggle with convergence issues. In particle physics, MALA has been employed to improve simulations of subatomic particles. These simulations often involve models that require accurate estimations of parameters. Traditional methods can suffer from slow convergence, but MALA's ability to efficiently navigate these parameter spaces has made it an effective tool in this domain, leading to more reliable estimates and insights into physical processes. Furthermore, MALA's utility extends into the realm of image processing and computer vision. It is particularly effective in tasks such as image denoising and texture synthesis.

F. Future Directions of Research

As stated previously, for more complex non-gaussian distributions having sharp peaks, discontinuities, or heavy tales, the gradient information becomes unreliable, leading to poor proposals and slow convergence rates of MALA. Its performance is also highly sensitive to the choice of step size in the Euler-Maruyama discretization of Langevin diffusion. Large step size leads to low acceptance rates, while small step size leads to poor mixing.

Given MALA's challenges, further research can focus on developing adaptive MALA methods. These adaptive approaches dynamically adjust the step size during the sampling process based on the local properties of the target distribution, such as its curvature or variability. Furthermore, Extending MALA to handle more general structures beyond Euclidean spaces could be an important direction. This might include developing algorithms for sampling on Riemannian manifolds, discrete manifolds, or other non-Euclidean structures with more complex topologies. By modifying MALA to operate in the correct geometric context, we can ensure that the algorithm better respects the problem's true nature. This includes areas such as geostatistics (where spatial data might lie on a sphere, like the Earth's surface), diffusion tensor imaging in neuroscience (where data lie on positive-definite matrices), and beyond. These are areas where conventional Euclidean approaches would be inadequate or inefficient.

5. Conclusion

This work presents a comprehensive theoretical analysis of Langevin diffusion and its applications to Markov Chain Monte Carlo (MCMC) methods. Our results establish the ergodicity of both continuous-time Langevin diffusion processes and their discrete-time counterparts, offering rigorous justification for the use of Langevin-based MCMC methods in sampling from complex probability distributions. The theoretical framework developed here, encompassing Foster-Lyapunov drift conditions, coupling arguments, and geometric ergodicity, provides a solid foundation for understanding the convergence properties of these methods. We demonstrate that Langevin-based MCMC algorithms can achieve faster convergence and more efficient exploration of high-dimensional spaces compared to traditional random-walk methods. However, our analysis also highlights the critical role of appropriate step size selection in discrete implementations, balancing between exploration efficiency and approximation accuracy. The connections we draw to optimal transport theory and recent developments in adaptive methods open new avenues for further research and algorithm development. As the field of computational statistics continues to grapple with increasingly complex and high-dimensional problems, the insights provided by this work are crucial in guiding the design and implementation of more efficient and reliable sampling algorithms. Future work may focus on developing adaptive schemes that automatically tune algorithm parameters based on the theoretical principles established here, further bridging the gap between theoretical guarantees and practical performance in challenging real-world applications.

References

[1] Girolami, M., & Calderhead, B. (2011). "Riemann manifold langevin and hamiltonian monte carlo methods," *Journal of the Royal Statistical Society: Series B (Statistical Methodology), 73*, 123-214, 03/2011. https://doi.org/10.1111/j.1467-9868.2010.00765.x

[2] Durmus, A., & Moulines, E. (2017). Nonasymptotic convergence analysis for the unadjusted langevin algorithm. The Annals of Applied Probability, 27, 1551-1587, 06/2017. https://doi.org/10.1214/16-aap1238.

[3] Roberts, G. O., & Rosenthal, J. S. (2009). Examples of adaptive mcmc. *Journal of Computational and Graphical Statistics, 18*, 349-367. 01/2009. https://doi.org/10.1198/jcgs.2009.06134.

[4] H. Kaspi, S. P. Meyn, and R. L. Tweedie, (1997). Markov chains and stochastic stability. *Journal of the American Statistical Association, 92*, p. 792, 06/1997. https://doi.org/10.2307/2965732

[5] Chen, Y., Georgiou, T. T., & Pavon, M. (2015). On the relation between optimal transport and schrodinger bridges: A stochastic control viewpoint. *Journal of Optimization Theory and Applications, 169*, 671-691, 09/2015. https://doi.org/10.1007/s10957-015-0803-z

[6] Parno, M. D., & Marzouk, Y. M. (2018). Transport map accelerated markov chain monte carlo. SIAM/ASA *Journal on Uncertainty Quantification, 6*, 645-682, 01/2018.https://doi.org/10.1137/17m1134640

[7] Roberts, G. O., & Tweedie, R. L. (1996). Exponential convergence of langevin distributions and their discrete approximations. *Bernoulli, 2*, p.341, 12/1996. https://doi.org/10.2307/3318418

[8] Dalalyan, A. S. (2014). Theoretical guarantees for approximate sampling from smooth and log-concave densities. *arXiv (Cornell University),* 01/2014. https://doi.org/10.48550/arxiv.1412.7392

[9] Cheng, X., and Bartlett, P. (2017). Convergence of langevin mcmc in kl-divergence. *arXiv(Cornell University),* 01/2017. https://doi.org/10.48550/arxiv.1705.09048

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (http://creativecommons.org/licenses/by/4.0/).