Journal of Machine Learning **ISSN**: 2790-2048(e), 2790-203X(p)

Convergence Analysis of Discrete Diffusion Model: Exact Implementation through Uniformization

Hongrui Chen [*1] and Lexing Ying [†1]

[1]Department of Mathematics, Stanford University, Stanford, CA 94305, USA.

Abstract. Diffusion models have achieved huge empirical success in data generation tasks. Recently, some efforts have been made to adapt the framework of diffusion models to discrete state space, providing a more natural approach for modeling intrinsically discrete data, such as language and graphs. This is achieved by formulating both the forward noising process and the corresponding reversed process as continuous time Markov chains. In this paper, we investigate the theoretical properties of the discrete diffusion model. Specifically, we introduce an algorithm leveraging the uniformization of continuous Markov chains, implementing transitions on random time points. Under reasonable assumptions on the learning of the discrete score function, we derive total variation distance and Kullback–Leibler divergence guarantees for sampling from any distribution on a hypercube. Our results align with state-of-the-art achievements for diffusion models in $\mathbb{R}^d$ and further underscore the advantages of discrete diffusion models in comparison to the $\mathbb{R}^d$ setting.

Keywords:
Diffusion model,
Sampling,
Machine learning theory.

Article Info.:
Volume: 4
Number: 2
Pages: 108 - 127
Date: June/2025
doi.org/10.4208/jml.240812

Article History:
Received: 12/08/2024
Accepted: 08/02/2025

Communicated by:
Jianfeng Lu

1 Introduction

Generative modeling is one of the central tasks in machine learning, which aims to learn a probability distribution from data and generate data from the learned distribution. The diffusion model has emerged as a powerful and versatile framework in generative modeling, achieving state-of-the-art performance in a variety of data generation tasks, including image generation [3, 28], audio generation [33], video generation [17, 43], text-to-image synthesis [30, 31], and computational biology [15]. The general framework of the score-based generative model involves:

1) defining a forward noising process to gradually diffuse the data distribution to some simple distribution (like standard Gaussian);

2) learning a reversed process to denoising the simple distribution to the data distribution by estimating the score functions of the forward diffusion process.

Works on the diffusion model focus on the forward processes defined in the Euclidean state space $\mathbb{R}^d$. In such scenarios, an ideal choice of the forward process is the Ornstein-Uhlenbeck (OU) process, which is driven by a stochastic differential equation (SDE) on $\mathbb{R}^d$,

*Corresponding author. hongrui@stanford.edu
†lexing@stanford.edu

and the corresponding reversed process is also given by an SDE. Nevertheless, certain data generation tasks present an intrinsic characteristic of discrete data. For example, natural language processing operates within a discrete token space, computer vision involves discrete representations of images, and molecular graph modeling engages with graph data in a discrete structure [18,19,44]. Thus, it is more natural to use diffusion processes on the discrete state space to model these discrete data distributions.

To this end, some recent works [5,7,25,27,32] have introduced a framework for diffusion models in discrete state spaces. This framework notably utilizes a continuous-time Markov chain (CTMC) in the discrete state space for the forward process, and the corresponding reverse process is also a CTMC. Moreover, mirroring the concept of score estimation in diffusion models on $\mathbb{R}^d$, they proposed a discrete score function given by the ratios of probability mass on different states, and the score entropy loss as a new score matching objective that is derived from Kullback–Leibler (KL) divergence divergence between the path measures of the forward and the reversed process. Combining the learning of the discrete score function through minimizing the score entropy and the sampling from the learned reversed process, a completed procedure for the diffusion model on discrete state space has been established.

However, despite the potential advantage of the discrete diffusion model, unlike the extensively studied SDE framework, the theoretical understanding of the CTMC framework has not been built. A line of works [4,8,10,21–23] concentrated on the theory of diffusion model on $\mathbb{R}^d$. Generally speaking, the established theoretical results can be summarized as follows:

- Sampling is as easy as learning the score: for arbitrary data distribution, suppose one can estimate the score function at multiple noise levels, then one can approximately sample from the data distribution.
- Quantitatively, under an L^2 accurate score estimator on the forward process, $\mathcal{O}(d\log(1/\delta)/\epsilon^2)$ iterations suffices to output a distribution that is ϵ^2-close in KL divergence to a distribution p_δ, where p_δ is a variance-δ Gaussian perturbation of the data distribution.
- There are three sources of error in the diffusion model: 1) the error from the inexact score estimator, 2) the error from insufficient mixing of the forward process, and 3) the discretization error. The discretization error causes the key challenges in the analysis due to the error propagation in the numerical simulation of a non-contractive dynamic.

In this paper, we take a step toward the theory of diffusion model in the CTMC framework and aim to understand how the theoretical property of discrete diffusion compares to the established theory for diffusion model on $\mathbb{R}^d$. Our results suggest that:

- One can implement the reversed CTMC in an exact way, i.e. without discretization error, through an algorithm based on the uniformization technique [12,14,40]. This presents a surprising advantage of the CTMC framework compared to the SDE framework, where discrete errors are significant in the analysis.