East Asian Journal on Applied Mathematics
doi: 10.4208/eajam.2024-158.280924

Vol. **15**, No. 4, pp. 669-700
November 2025

Mathematical Analysis of Singularities in the Diffusion Model Under the Submanifold Assumption

Yubin Lu[1], Zhongjian Wang[2,*] and Guillaume Bal[3]

[1]*Department of Applied Mathematics, Illinois Institute of Technology, Chicago, IL 60616, USA.*
[2]*Division of Mathematical Sciences, School of Physical and Mathematical Sciences, Nanyang Technological University, Singapore 639798.*
[3]*Departments of Mathematics, Statistics and CCAM, University of Chicago, Chicago, IL 60637, USA.*

Received 18 September 2024; Accepted (in revised version) 28 September 2024.

Abstract. This paper concerns the mathematical analyses of the diffusion model in machine learning. The drift term of the backward sampling process is represented as a conditional expectation involving the data distribution and the forward diffusion. The training process aims to find such a drift function by minimizing the mean-squared residue related to the conditional expectation. Using small-time approximations of the Green's function of the forward diffusion, we show that the analytical mean drift function in DDPM and the score function in SGM asymptotically blow up in the final stages of the sampling process for singular data distributions such as those concentrated on lower-dimensional manifolds, and are therefore difficult to approximate by a network. To overcome this difficulty, we derive a new target function and associated loss, which remains bounded even for singular data distributions. We validate the theoretical findings with several numerical examples.

AMS subject classifications: 68T07, 62G07, 60H10, 35B40, 35Q84

Key words: Generative model, singularities, Green's kernel, adaptive time-stepping, low-dimensional manifold.

1. Introduction

The field of generative models has emerged as a powerful tool for building continuous probabilistic models that generate new samples from given discrete datasets. Furthermore, by accounting for the joint distribution of observable (condition, query) and target variables, these models offer a flexible and efficient way to generate samples based on queries. Such generative models have been applied across a wide range of disciplines, including

*Corresponding author. *Email addresses:* ylu117@iit.edu (Y. Lu), zhongjian.wang@ntu.edu.sg (Z. Wang), guillaumebal@uchicago.edu (G. Bal)

computer vision [13], speech signal processing [43], natural language processing [20], and natural sciences [41]. Recent advances in generative models, including the popular variational autoencoder [23], generative adversarial network [17], flow-based model [2], and DeepParticle model [45], have demonstrated their ability to solve diverse problems across different domains. These models share a common feature: they directly model a push-forward map from easy-to-sample distribution to an unknown distribution driven by data through neural networks (including the composition of invertible functions).

In contrast to these direct constructions, another type of generative model links distributions through one-parameter continuous deformations (parameter $t \in [0,T]$ or $[0,\infty)$ in the following up various models). This approach has a long history in the mathematical literature. A straightforward example is the Langevin Monte Carlo (LMC), where we start from arbitrary distribution solving the following SDE:

$$dX_t = v(t,X_t)dt + \sqrt{2D(t,X_t)}dW_t, \tag{1.1}$$

where $(v,D) = (-\nabla \log p_{data}, I)$. From arbitrary initial distribution $X_0 \sim p_0$, LMC (1.1) solves until t is sufficiently large. Then the distribution of X_t continuously deforms from arbitrary distribution p_0 to p_{data} as $t \to \infty$.

Without analytic expression to p_{data}, a number of constructive data-driven approaches to (v,D) have been considered in the literature — cf. Refs. [4,8,31,39,44]. Once a model of (v,D) is derived either analytically or from data, the stochastic differential equation (SDE) integrator can be used to numerically solve the SDE from initial to terminal time (some $T > 0$ or ∞), thereby interpreting it as a generative model.

Among these constructions, the diffusion models have attracted huge attention due to their well-known performance in practical applications. Inspired by non-equilibrium thermodynamics [38], Ho *et al.* [19] proposed denoising diffusion probabilistic models (DDPMs), a class of latent variable models, as an early diffusion model. Later, Song *et al.* [40] unified several earlier models through the lens of stochastic differential equations and proposed score-based generative models (SGMs). The backward process (generation of new samples) can be interpreted as solving Eq. (1.1) with a tweak that reverts the notation of time and the initial distribution after the tweak is assumed to be a standard normal distribution [1]. Luo [28] and Yang *et al.* [46] provided literature reviews from different perspectives.

Despite its success, the sampling process for diffusion models is extremely slow and the computational cost is high. In DDPMs [19], for instance, 1000 steps are typically needed to generate samples. Several works have attempted to accelerate the sampling process [21, 25–27, 29, 37, 47, 48]. In addition, Zhang and Chen [47] pointed out that there were dramatically different performances in terms of discretization error and training error when they trained the score function of SGM on different datasets. On theoretical side, several recent works have shown the convergence bound for diffusion models, for instance [6,9,11, 24]. It is then natural to seek justifications for the gap between practical costly performance and theoretical convergence guarantee.

In fact, the aforementioned theoretical results require the score function to be approximated well by a neural network, in the L^2 or L^∞ sense. While in practice, when