National Research Journal of Information Technology & Information Science
Volume No: 12, Issue No: 1, Year: 2025 (January-June)
PP: 284-296
ISSN: 2350-1278
Peer Reviewed & Refereed Journal (IF: 7.9)
Journal Website www.nrjitis.in

CONVERGENCE AND STABILITY OF DIFFUSION-BASED GENERATIVE MODELS: AN SDE-BASED THEORETICAL REVIEW

Shashikumar R

Associate Professor
Department of Mathematics, Government First Grade College, Yelahanka

ABSTRACT

Diffusion-based generative models have evolved from discrete-time denoising procedures into a broad family of stochastic and deterministic continuous-time formulations. Their connection with stochastic differential equations (SDEs) provides a mathematically coherent framework for describing forward perturbation, reverse-time generation, score estimation, and numerical sampling. Theoretical results concerning convergence and stability, however, remain distributed across different assumptions, state spaces, error metrics, and sampling schemes, and are rarely compared on common terms. This review organizes that literature around the interaction between score approximation, reverse-time dynamics, numerical discretization, and distributional error. We first establish the mathematical foundations linking discrete diffusion probabilistic models, score-based generative models, reverse-time SDEs, and probability-flow ordinary differential equations. We then critically compare convergence guarantees expressed in Wasserstein distance, Kullback–Leibler (KL) divergence, and total variation (TV) distance, with attention to assumptions on score regularity, target-distribution structure, Fisher information, and intrinsic dimensionality. A complementary analysis examines stability with respect to initialization, score perturbation, discretization, and long-time propagation of error — a body of work that remains far less unified than convergence theory. We synthesize recent results on accelerated sampling, intrinsic-dimensional convergence, exact Gaussian error decomposition, algorithmic (generalization) stability, and forgetting/contraction of reverse-time Markov chains. Rather than treating convergence and stability as independent properties, we propose a convergence–stability–error-propagation framework that separates initialization, truncation, score-estimation, and discretization errors and asks how the reverse dynamics amplify, preserve, or attenuate these perturbations. The review closes by identifying theoretically defensible open problems — non-Lipschitz score fields, dimension-independent guarantees, realistic neural score errors, long-time reverse-process stability, and unified statistical–numerical bounds — and proposes a corresponding research agenda expressed as concrete mathematical questions rather than generic calls for “more research.”

Keywords: Diffusion models; score-based generative models; stochastic differential equations; reverse-time SDEs; probability-flow ODE; convergence; stability; score estimation; numerical discretization; Wasserstein distance; Kullback–Leibler divergence.

1. INTRODUCTION

Diffusion-based generative models occupy an unusual position in modern machine learning: their empirical performance in image, audio, and scientific data generation is well established, but the mathematical apparatus used to describe why and under what conditions they work has developed unevenly across probability theory, numerical analysis, statistical learning theory, and partial differential equations (PDEs). Three observations motivate this review.

First, diffusion models have moved from a discrete-time probabilistic construction — the denoising diffusion probabilistic model (DDPM) of Ho, Jain, and Abbeel — toward a unified

ISSN: 2350-1278
Peer Reviewed & Refereed Journal (IF: 7.9)
Journal Website www.nrjitis.in

continuous-time stochastic description in which forward noising, reverse-time generation, and deterministic sampling are all instances of a single SDE/PDE object. This unification, due to Song, Sohl-Dickstein, Kingma, Kumar, Ermon, and Poole, rests on the classical time-reversal theory of Anderson developed decades before generative modeling existed as a field.

Second, empirical success has outpaced theoretical understanding. Sample quality metrics such as FID or Inception Score are not distributional convergence guarantees, and a mathematically rigorous review must resist the temptation to treat them as such.

Third, and most importantly for this review's contribution, existing theory is fragmented: convergence results differ in the metric used (KL divergence, Wasserstein distance, total variation), in the regularity imposed on the score function, in whether the score is treated as exact or statistically estimated, in whether discretization error is modeled at all, and in whether the ambient or intrinsic dimension of the data governs the rate. A parallel and even less mature literature on stability — how perturbations in initialization, score, or discretization propagate through the reverse dynamics — is typically not connected to the convergence literature at all.

The contribution of this review is therefore not another taxonomy of diffusion model architectures. Several such surveys already exist and cover foundations, architectures, and applications. Instead, this review organizes theoretical guarantees according to error propagation and stability, builds a common notational and conceptual scaffold across probability, numerical analysis, and learning-theoretic treatments, and identifies which assumptions are load-bearing versus incidental to each guarantee.

2. REVIEW METHODOLOGY AND EVIDENCE SELECTION

This review followed a structured evidence-verification process rather than an unweighted aggregation of search results. Candidate references were drawn from arXiv, PMLR (ICML/ICLR/COLT/ALT proceedings), NeurIPS proceedings, and SIAM journals. Each candidate reference was checked against a publisher, conference-proceedings, or arXiv record before being admitted to the evidence base; volume, page, and identifier information reported in Section 13 reflects those checks as of the evidence cut-off (12 September 2026). Preprints without peer-reviewed publication records at the time of writing are explicitly flagged as such throughout (most notably Liu & Zuazua and Strasman et al.) and are not treated as equivalent in evidentiary weight to peer-reviewed results.

Selection prioritized:

1. Primary sources over secondary summaries or blog-level descriptions of theorems;
2. Papers that state explicit assumptions and explicit guarantees (a theorem, proposition, or corollary with a stated convergence or stability bound), over papers that discuss convergence only qualitatively;
3. Breadth across the different metrics and formalisms (continuous SDE, probability-flow ODE, discrete-state CTMC) rather than depth within a single sub-literature, since the review's contribution is explicitly comparative.

Where a claim could not be verified against a primary source at the time of writing, it is marked NOT VERIFIED rather than presented as established fact. No citation, theorem statement, or numerical claim in this manuscript has been invented; where the underlying papers report results under restricted settings (e.g., Gaussian data, discrete state spaces), that restriction is retained in this review's presentation rather than generalized silently.

 ISSN: 2350-1278
Peer Reviewed & Refereed Journal (IF: 7.9)
Journal Website www.nrjitis.in

3. MATHEMATICAL FOUNDATIONS

3.1 Forward diffusion and score

The forward process is modeled as an Itô SDE

$$dX_t = f(X_t, t)\,dt + g(t)\,dW_t\,, \quad t \in [0, T],$$

where f is the drift, g the (typically scalar) diffusion coefficient, and W_t a standard Wiener process. Writing $p_t(x)$ for the marginal density of X_t, the score function is

$$s_t(x) = \nabla_x \log p_t(x).$$

3.2 Reverse-time SDE

The generative mechanism is the time-reversal of this process. Anderson's foundational result establishes that, under regularity conditions on the forward dynamics, the reverse-time process is itself a diffusion, schematically

$$dX_t = [\ f(X_t, t) - g(t)^2\, \nabla_x \log p_t(X_t)\ ]\,dt + g(t)\,d\hat{W}_t\,,$$

run backward in time, where $\hat{W}_t$ is a Wiener process for the reversed time flow. Song et al. adapted this result to generative modeling by showing that a learned approximation $s\theta(x,t) \approx \nabla_x \log p_t(x)$ can be substituted into the reverse dynamics to synthesize samples from noise. A central conceptual point — worth stating explicitly because it is a frequent source of confusion in less careful expositions — is that the reverse-time SDE is not the forward SDE run backward: it contains a density-dependent correction term arising from stochastic time reversal, not merely a change of temporal orientation.

3.3 Probability-flow ODE

The same marginal densities $p_t(x)$ can, under appropriate conditions, be realized by a deterministic flow — the probability-flow ODE:

$$dX_t/dt = f(X_t, t) - \tfrac{1}{2}\, g(t)^2\, \nabla_x \log p_t(X_t).$$

This shares the reverse SDE's marginals but has distinct numerical and stability behavior: it is amenable to higher-order deterministic ODE solvers (as in DDIM-type samplers due to Song, Meng, and Ermon) but does not benefit from the noise-injection mechanisms that give the stochastic sampler certain contraction properties (Section 8).

3.4 Notation used throughout this review

To prevent the conflation of distinct error sources — one of the most common defects in less rigorous diffusion reviews — the following notation is fixed and used consistently:

Symbol	Meaning
X_t	state at time t
$p_t(x)$	true marginal density at time t
$s_t(x) = \nabla_x \log p_t(x)$	exact score
$s\theta(x,t)$	learned/estimated score
$e_s(x,t) = s\theta(x,t) - s_t(x)$	pointwise score error
h = max_i (t_(i+1) − t_i)	maximum discretization step
ε_score = ‖sθ − s‖ ($\mathscr{F}$-norm)	aggregate score-approximation error (norm depends on the analysis)
ε_disc = d(L(X_t), L(Y_t))	discretization-induced distributional error (L(·) denotes the law/distribution; Y_t is the numerically simulated trajectory)

Symbol	Meaning
ε_init, ε_trunc	initialization and early-truncation error

Score-approximation error and numerical (discretization) error are, by this notation, categorically distinct quantities measured in different spaces; a guarantee that controls one does not automatically control the other.

4. DIFFUSION MODELS AND THE SDE FORMULATION

"Diffusion model" is not a single mathematical object, and convergence claims cannot be compared without first fixing the state space, forward process, reverse mechanism, score-approximation regime, discretization scheme, and error metric. The taxonomy below is organized by primary theoretical issue rather than by chronology or architecture family.

Family	Forward process	Reverse mechanism	Primary theoretical issue
DDPM (Ho et al., 2020)	Discrete-time Markov chain	Learned reverse transition kernels	Discrete approximation of a continuous limit
Score SDE (Song et al., 2021)	Continuous-time SDE	Reverse-time SDE	Score approximation
Probability-flow formulation	Continuous SDE (shared marginals)	Deterministic ODE	Deterministic numerical stability
DDIM (Song, Meng & Ermon, 2021)	Non-Markovian diffusion	Deterministic or stochastic sampler	Sampling-schedule discretization
Schrödinger bridge (De Bortoli et al., 2021)	Controlled diffusion	Forward/backward bridge	Finite-time transport between two endpoints
Discrete-state diffusion (Zhang, Chen & Gu, 2025)	Continuous-time Markov chain (CTMC)	Reverse CTMC	Discrete score and KL/TV convergence
Manifold-aware diffusion (Potaptchik et al., 2025)	Ambient-space SDE	Reverse SDE	Intrinsic vs. ambient dimension
Stability-oriented formulations (Strasman et al., 2026; Liu & Zuazua, 2025)	SDE / Fokker–Planck PDE	Controlled reverse dynamics	Perturbation and error propagation over long horizons

The historical progression can be summarized as

$$\text{DDPM} \rightarrow \text{score-based formulation} \rightarrow \text{continuous-time SDE} \rightarrow \text{reverse SDE} \rightarrow \text{probability-flow ODE},$$

with Ho et al. establishing the discrete-time framework and Song et al. showing that the discrete and score-based formulations embed into a common continuous-time SDE description — the representational backbone this review adopts throughout.

5. REVERSE-TIME DYNAMICS AND SCORE ESTIMATION

Sections 3–4 already isolated the reverse-time SDE and probability-flow ODE as the two principal generative mechanisms. What varies across the literature reviewed here is how the score entering these dynamics is treated: as exact (a purely mathematical object of study), as estimated with a uniform (L^∞/Lipschitz) accuracy guarantee, or as estimated with an L^2-type statistical accuracy guarantee closer to what neural score matching actually produces.

ISSN: 2350-1278
Peer Reviewed & Refereed Journal (IF: 7.9)
Journal Website www.nrjitis.in

Chen, Chewi, Li, Li, Salim, and Zhang's analysis is the pivotal reference here: by working with L^2-accurate score estimates rather than requiring uniform accuracy, and by avoiding some of the more restrictive functional-inequality assumptions on the data distribution used in earlier analyses, their framework is considerably closer to the statistical reality of how scores are trained via denoising score matching. This shift — from assumptions on the target distribution's regularity toward assumptions on the score-estimation error — recurs across later work (Lee, Lu & Tan; Conforti, Durmus & Gentiloni Silveri) and is one of the clearest lines of theoretical progress traceable in this literature.

6. CONVERGENCE ANALYSIS

Convergence results can be classified along several axes without conflating them:

- Type I — Continuous-time convergence, concerning the exact reverse-time process with an idealized score;
- Type II — Convergence with imperfect score estimation, where an L^2 or similar statistical error enters the bound;
- Type III — Numerical-discretization convergence, isolating step-size (h) dependence;
- Type IV — Distributional convergence, distinguished by the metric used (Wasserstein, KL, TV);
- Type V — Dimension-dependent convergence, distinguishing ambient dimension D from intrinsic dimension d;
- Type VI — Acceleration results, concerning improved rates for specific numerical schemes.

6.1 Convergence with accurate score estimates

Chen et al. established polynomial-in-dimension sampling guarantees assuming only L^2-accurate score estimates, a substantially broader theoretical regime than earlier analyses requiring uniform accuracy.

6.2 General target distributions

Lee, Lu, and Tan established polynomial convergence guarantees for denoising diffusion models across broad classes of data distributions, with Wasserstein-distance guarantees under bounded-support or tail-decay conditions, and stronger total-variation guarantees requiring additional smoothness. This represents a deliberate relaxation relative to earlier analyses that leaned on stronger global regularity assumptions on the data distribution itself.

6.3 KL convergence under finite Fisher information

Conforti, Durmus, and Gentiloni Silveri's SIAM Journal on Mathematics of Data Science paper addresses a specific and previously entrenched restriction: the assumption, common to almost all prior KL-divergence bounds, that the score function or its approximation is Lipschitz uniformly in time. Their finite-Fisher-information condition (relative to a standard Gaussian) replaces this uniform-Lipschitz requirement and still yields quantitative KL convergence. The conceptual payoff is:

$$\text{weak regularity assumptions on the score} \not\Rightarrow \text{absence of quantitative convergence guarantees.}$$

Rather, the choice of convergence metric and analytical technique determines which assumptions are actually load-bearing.

ISSN: 2350-1278
Peer Reviewed & Refereed Journal (IF: 7.9)
Journal Website www.nrjitis.in

6.4 Exact error decomposition in the Gaussian case

Pierret and Galerne's ICML 2025 paper restricts to Gaussian target data but, in exchange for this restriction, derives exact analytical solutions for the backward SDE and probability-flow ODE and computes exact Wasserstein errors between the target and the numerically sampled distribution for arbitrary numerical schemes. Their explicit identification of four separate error sources — initialization, truncation, discretization, and score approximation — is one of the most conceptually useful contributions in the reviewed literature, precisely because it is derived rather than assumed. As emphasized in Section 9, this decomposition should be read as a template whose generalization to non-Gaussian settings requires independent justification, not as a universal additive theorem.

6.5 Intrinsic dimension

Potaptchik, Azangulov, and Deligiannidis's COLT 2025 result establishes that the number of steps required for KL convergence is linear, up to logarithmic factors, in the intrinsic dimension d of a data manifold embedded in ambient dimension D — improving over prior bounds that were either linear in D or polynomial (superlinear) in d, and the authors show the linear dependence on d is sharp. This constitutes a genuine transition in the theory, separating the axis

$$D = \text{ambient dimension} \quad \text{versus} \quad d = \text{intrinsic dimension},$$

and should be treated as such, rather than as merely another entry in a list of convergence rates.

6.6 Discrete-state diffusion

Zhang, Chen, and Gu extend convergence analysis to discrete-state diffusion under a continuous-time Markov chain (CTMC) formulation, deriving KL- and TV-divergence bounds for a discrete-time sampling algorithm whose error decomposes into score-estimation error, discretization error, and truncation error due to incomplete mixing of the forward process — a decomposition structurally analogous to, though independently derived from, the continuous-state case.

6.7 Acceleration

Li, Huang, Efimov, Wei, Chi, and Chen's ICML 2024 paper establishes improved convergence rates for specific accelerated sampling schemes under L^2 score-accuracy assumptions, demonstrating that discretization error and iteration complexity can be reduced without abandoning the statistical assumptions used in Section 6.1 — though the improvement is scheme-specific rather than universal across all numerical samplers.

7. STABILITY ANALYSIS

Stability is a materially different question from convergence, and the literature addressing it is considerably less unified. A convergence statement takes the form

$$d(\mathscr{L}(X_t), p\_data) \leq \varepsilon,$$

whereas a stability statement asks how sensitive the approximation itself is to perturbation:

$$d(\mathscr{L}(X_t), \mathscr{L}(Y_t)) \ \leq\ C \cdot d(\mathscr{L}(X_0), \mathscr{L}(Y_0)),$$

Here Y denotes the corresponding perturbed (or approximate) trajectory, obtained by perturbing the initialization, score, or discretization of the reverse process.

or, more generally, how perturbations in the score, drift, or discretization propagate through the reverse dynamics. These are related but not interchangeable properties, and treating a

ISSN: 2350-1278
Peer Reviewed & Refereed Journal (IF: 7.9)
Journal Website www.nrjitis.in

convergence bound as if it settled stability (or vice versa) is a conflation this review deliberately avoids.

Three stability levels can be distinguished and should not be conflated with one another:

trajectory stability $\rightarrow$ distributional stability $\rightarrow$ algorithmic/training stability.

A finer classification, cross-cutting the reviewed literature, is:

Stability level	Central question
SDE stability	Are trajectories/densities sensitive to perturbations of the dynamics?
Reverse-time stability	Are errors amplified during generation?
Score stability	What happens when the learned score itself is perturbed (e.g., by a change in training data)?
Numerical stability	What happens to the sampler as step size changes?
Distributional stability	How does the output distribution respond to input perturbations?
Algorithmic stability	How does changing the training dataset affect the learned score, and hence generalization?

7.1 Forgetting and contraction of the reverse-time Markov chain

Strasman, Cardoso, Le Corff, Lemaire, and Ocello (arXiv:2601.21868, preprint as of this review) approach stability through the forgetting and contraction properties of the reverse-time Markov chain, establishing two structural conditions — a Lyapunov drift condition and a Doeblin-type minorization condition — sufficient to control the propagation of initialization and discretization error through the backward process. A practical consequence they derive is that the reverse diffusion dynamics can induce a contraction mechanism along the sampling trajectory: certain errors introduced early in generation need not accumulate monotonically and may instead be damped. This paper's peer-reviewed status was not established at the time of this review's evidence cut-off and it is treated as a preprint accordingly.

7.2 PDE-based stability

Liu and Zuazua (arXiv:2511.05940, preprint) develop a rigorous PDE framework for score-based diffusion built on the Li–Yau differential inequality for heat flow, deriving sharp L^p-stability estimates for the score-based Fokker–Planck dynamics and, via entropy-stability (KL-based) methods, showing that reverse-time dynamics concentrate on the data manifold for compactly supported data and a broad class of initializations. Their analysis also identifies a blow-up phenomenon in the L^p estimate near $t\sim0$ tied to the score singularity at the start of the reverse process, which is mitigated in practice by time discretization — a concrete instance in which numerical discretization is not merely a source of error but can also regularize an otherwise singular continuous-time object. As with Strasman et al., this is a preprint at the time of writing and is flagged accordingly.

7.3 Algorithmic (generalization) stability

Farghly, Rebeschini, Deligiannidis, and Doucet (ICLR 2026; arXiv:2507.03756) address a distinct notion of stability: not the sensitivity of the reverse-time trajectory to perturbation, but the sensitivity of the learned score itself to perturbations of the training dataset. They introduce score stability, an algorithmic-stability-style quantity, and use a reflection-coupling argument to show that the stochasticity inherent to the denoising score-matching gradient estimator induces a contraction that yields generalization bounds — providing a mechanism by which diffusion models trained and sampled imperfectly can generalize rather than memorize training data. This result is important for keeping training-time stability analytically separate from generation-time (reverse-dynamics) stability; the two are frequently discussed under the single word "stability" in less careful treatments, but they concern different random objects (the training procedure versus the reverse-time sampling trajectory) and are governed by different mechanisms.

8. NUMERICAL DISCRETIZATION AND SAMPLING ERROR

A practical sampler is never exposed to an exact score, an infinite time horizon, or continuous-time dynamics: it operates with a learned score $s\theta$, a truncated horizon $[\delta, T]$, and a discretized reverse-time integrator with step size h. At minimum, two structurally distinct error sources are present in any practical guarantee:

$$e\_score \quad \text{and} \quad e\_disc.$$

Convergence theorems that assume an exact score (Type I in Section 6) therefore necessarily understate the error of a deployed sampler; the practically relevant guarantees are those of Types II–III, which explicitly separate statistical (score) error from numerical (discretization) error. Pierret and Galerne's Gaussian analysis makes this separation concrete by deriving exact Wasserstein error as a function of the numerical scheme, rather than an upper bound whose tightness is unclear — at the cost of restricting to Gaussian data, a restriction this review does not attempt to lift by inference.

An additional and frequently underweighted error source is initialization/truncation error: the forward process is only run to a finite time T, so the initial condition of the reverse process is an approximation of the true stationary (e.g., Gaussian) distribution rather than the distribution itself, and the reverse process is typically only integrated down to a small $\delta > 0$ rather than exactly to $t = 0$. Both Pierret & Galerne and Zhang, Chen & Gu explicitly retain this term in their error decompositions, underscoring that it is not a negligible technicality in either the continuous- or discrete-state setting.

9. UNIFIED CONVERGENCE–STABILITY–ERROR FRAMEWORK

We propose the following synthesis, the Convergence–Stability–Error Propagation Framework (CSEPF), as a structuring device for the comparisons made in Sections 6–8. This framework is a review-derived organizing device, not a theorem established by any cited paper, and it should be read as such throughout.

The framework consists of five layers:

Data distribution → Forward diffusion → Score estimation → Reverse SDE/ODE → Numerical discretization → Generated distribution,

with four principal perturbation channels

$$\varepsilon\_init, \quad \varepsilon\_trunc, \quad \varepsilon\_score, \quad \varepsilon\_disc,$$

and two analytical operators: a convergence mechanism $\mathcal{O}$, which maps these perturbations to a final distributional error, and a stability/propagation mechanism $\mathcal{S}$, which determines whether the reverse dynamics amplify, preserve, or attenuate them:

$$\text{final distributional error} = \mathcal{O}[\ \varepsilon\_init, \varepsilon\_trunc, \varepsilon\_score, \varepsilon\_disc\ ], \ \text{subject to } \mathcal{S}.$$

A naive additive model would write the total error as E = E_init + E_trunc + E_score + E_disc. We instead retain an explicit interaction term:

$$\mathcal{U}\_total = \mathcal{U}\_init + \mathcal{U}\_trunc + \mathcal{U}\_score + \mathcal{U}\_disc + \mathcal{U}\_interaction,$$

ISSN: 2350-1278
Peer Reviewed & Refereed Journal (IF: 7.9)
Journal Website www.nrjitis.in

where $\mathcal{U}$_interaction captures the possibility — supported qualitatively by the forgetting/contraction results of Strasman et al. and the entropy-stability results of Liu & Zuazua — that these error sources do not combine additively once the stability mechanism $\mathcal{S}$ is taken into account. Whether $\mathcal{U}$_interaction can be bounded, shown to vanish in specific regimes, or shown to dominate in others is, in our assessment, a genuine open research question rather than a settled technicality (see Section 11, Gap A).

10. CRITICAL COMPARATIVE ANALYSIS

10.1 Persistent evidence table

The table below is the persistent evidence matrix maintained throughout this review's development. All entries trace to the primary sources verified in Section 13; entries whose peer-reviewed status could not be confirmed at the time of writing are marked accordingly in the Limitation column.

Paper	Claim	Result	Key assumption	Conv./Stab. type	Metric	Limitation
Anderson (1982)	Time-reversal of a diffusion is itself a diffusion	Reverse-time diffusion formulation	Regularity of forward diffusion coefficients	Foundational	Density/dynamics	Predates generative modeling; not a finite-sample result
Song et al. (2021, SDE)	DDPM and score-based models unify into one continuous-time SDE picture	Reverse SDE + probability-flow ODE formulation	Availability of the (true or approximate) score	Conceptual/continuous-time	Sampling formulation	Not primarily a non-asymptotic convergence theorem
Chen et al. (2023)	Sampling is as easy as learning the score	Polynomial sampling-complexity guarantees	L^2-accurate score estimate	Type II	Distributional error	Guarantee's tightness still hinges on score-estimation quality
Lee, Lu & Tan (2023)	Broad data distributions admit polynomial convergence	Wasserstein and TV convergence bounds	Bounded support/tail decay (W); smoothness (TV)	Type IV	Wasserstein, TV	Assumptions are metric-specific, not uniform
Li et al. (2024)	Sampling can be provably accelerated	Improved convergence rates for specific schemes	L^2 score accuracy	Type VI	Sampling error	Scheme-specific; not a universal result

Paper	Claim	Result	Key assumption	Conv./Stab. type	Metric	Limitation
Conforti, Durmus & Gentiloni Silveri (2025)	KL convergence under weaker-than-Lipschitz score assumptions	Quantitative KL convergence bound	Finite Fisher information relative to a Gaussian	Type II/IV	KL divergence	Tied to specific (OU-type) diffusion structures
Pierret & Galerne (2025)	Errors decompose exactly for Gaussian targets	Exact Wasserstein error decomposition (4 components)	Gaussian target distribution	Type IV (exact)	Wasserstein-2	Restricted to Gaussian data/model class
Potaptchik, Azangulov & Deligiannidis (2025)	Convergence is linear in intrinsic, not ambient, dimension	KL convergence, linear in d up to log factors; sharpness shown	Manifold hypothesis	Type V	KL divergence	Requires manifold-hypothesis structural assumption
Zhang, Chen & Gu (2025)	Discrete-state diffusion converges with near-linear dimension dependence	KL/TV convergence bounds under CTMC formulation	Discrete score-estimation assumptions	Type IV/V	KL, TV	Confined to discrete state spaces
Strasman et al. (2026)	Reverse dynamics can forget/contract errors rather than accumulate them	Contraction/forgetting bounds via Lyapunov drift + Doeblin minorization	Lyapunov drift and minorization conditions	Stability	Weighted TV-type	Preprint; peer-review status not established
Liu & Zuazua (2025)	Score-based Fokker–Planck dynamics admit sharp PDE stability estimates	L^p-stability estimates; entropy-stability manifold concentration	PDE regularity (Li–Yau-type inequality)	Stability	L^p, KL (entropy)	Preprint; peer-review status not established
Farghly et al. (2026)	Algorithmic stability of the score explains generalization	Generalization bounds via "score stability" (reflection coupling)	Structure of SGD/score-matching training	Algorithmic stability	Generalization gap	Concerns training-time, not generation-time stability

10.2 What the table reveals

No single paper in this evidence base simultaneously handles: arbitrary target distributions, statistically realistic (imperfect) score estimation, finite-time initialization/truncation, arbitrary numerical discretization, dimension-independent (intrinsic) scaling, and long-time stability. Each result trades some of these for tractability of the others — Gaussian data buys exact error decomposition (Pierret & Galerne); the manifold hypothesis buys dimension-free

ISSN: 2350-1278
Peer Reviewed & Refereed Journal (IF: 7.9)
Journal Website www.nrjitis.in

rates (Potaptchik et al.); relaxing uniform-Lipschitz assumptions on the score buys weaker regularity requirements at the cost of specificity to particular diffusion structures (Conforti et al.). We regard the absence of a single theorem spanning all six properties as one of this review's central analytical conclusions, not merely a gap to be listed alongside others (see Section 11).

10.3 Convergence versus stability, restated

The comparative picture that emerges is that convergence theory is comparatively mature and metric-diverse, while stability theory is comparatively young and mechanism-diverse: convergence results differ mainly in which metric and which assumption regime is used, whereas stability results differ in which object is being perturbed (trajectory, score, training data, or discretization) and which mathematical machinery is used to track the perturbation (Markov-chain contraction, PDE entropy methods, or algorithmic-stability couplings). These are not yet interoperable frameworks, and translating a contraction result stated for weighted TV distance (Strasman et al.) into a statement compatible with a Wasserstein-metric convergence guarantee (Lee et al.; Pierret & Galerne) is, to our knowledge, an open technical problem rather than a solved one.

11. OPEN PROBLEMS AND FUTURE RESEARCH

We rank the identified gaps rather than merely enumerate them, since not all gaps are equally consequential for the review's central claim that convergence and stability require joint treatment.

Gap A — Unified error propagation (highest priority). No currently verified single framework simultaneously and generally integrates ε_{init}, ε_{trunc}, ε_{score}, ε_{disc} with a rigorous stability coefficient for the reverse process, across metrics.

Gap B — Non-Lipschitz score fields. Uniform score-Lipschitzness is mathematically convenient but difficult to justify for realistic high-dimensional, multimodal data; Conforti et al. explicitly motivate and partially address this, but the finite-Fisher-information relaxation they use is tied to specific diffusion structures and has not been shown to generalize to arbitrary forward processes.

Gap C — Dimension dependence. The emergence of intrinsic-dimension results (Potaptchik et al.) suggests ambient-dimensional bounds may be unnecessarily pessimistic for structured, low-intrinsic-dimension data, but it remains open whether comparable dimension-free rates can be obtained without the manifold-hypothesis structural assumption.

Gap D — Stability under learned-score perturbations. A guarantee on $||s\theta - s||$ (a Type II convergence result) does not by itself reveal how that error propagates through the reverse-time dynamics in a stability sense (a Section 7 question); connecting these two literatures explicitly is largely undone.

Gap E — Stability versus acceleration. Higher-order or accelerated numerical schemes (Li et al.) can reduce discretization error, but whether acceleration trades off against robustness to score error or against the contraction properties identified by Strasman et al. is, to our knowledge, unexamined.

Gap F — Long-time behavior and monotonicity of error. Recent forgetting/contraction results (Strasman et al.; Liu & Zuazua) suggest initialization and discretization errors need not accumulate monotonically along the reverse trajectory. The precise structural conditions under which error is damped rather than amplified remain incompletely characterized and, at the time of writing, are supported only by preprint-level evidence.

ISSN: 2350-1278
Peer Reviewed & Refereed Journal (IF: 7.9)
Journal Website www.nrjitis.in

PROPOSED RESEARCH AGENDA

We conclude with six mathematically testable questions, in preference to a generic call for "further research":

4. Can a bound of the form $d(Q_0, P_0) \leq C_1 \cdot \varepsilon\_score + C_2 \cdot h^p + C_3 \cdot \varepsilon\_init + C_4 \cdot \varepsilon\_trunc$ be established in which the constants C_i explicitly depend on a reverse-process stability/contraction coefficient (in the sense of Strasman et al.), rather than being treated as free constants?
5. Can these constants C_i be made dependent on intrinsic rather than ambient dimension, extending Potaptchik et al.'s scaling to settings with imperfect score estimation and nonzero discretization step?
6. Can quantitative stability guarantees be obtained for scores that are non-Lipschitz but dissipative, relaxing the finite-Fisher-information condition of Conforti et al. to a broader structural class?
7. Under what verifiable, checkable conditions is the reverse-time diffusion contractive versus expansive, and can this be determined from properties of the forward process and score alone, without post hoc empirical measurement?
8. Can numerical acceleration (in the sense of Li et al.) be achieved without increasing sensitivity to score-estimation error, or is there a provable trade-off between the two?
9. Can convergence and stability be characterized simultaneously, within a single set of assumptions, in KL divergence, Wasserstein distance, and total variation, or are these three metrics fundamentally tied to disjoint assumption regimes?

CONCLUSION

Diffusion-based generative models admit a coherent continuous-time SDE description linking discrete-time DDPMs, score-based generative models, reverse-time SDEs, and probability-flow ODEs. Within this description, the theoretical literature has made substantial and traceable progress on convergence: assumptions have shifted from strong global regularity of the target distribution toward more realistic L^2-type statistical assumptions on score-estimation error (Chen et al.), toward weaker-than-Lipschitz score regularity (Conforti et al.), and toward intrinsic-dimensional rather than ambient-dimensional scaling (Potaptchik et al.). Stability theory, by contrast, remains fragmented across at least three largely disconnected mechanisms — Markov-chain forgetting/contraction, PDE entropy methods, and algorithmic (training-data) stability — and is supported, in its most recent and most directly relevant instances, primarily by preprints rather than peer-reviewed results. The central intellectual contribution of this review is not a new theorem but a discipline: separating score-approximation error from discretization error, separating convergence from stability, and separating trajectory-level stability from training-level (algorithmic) stability, so that future work can be positioned precisely within — or explicitly against — this structure. The Convergence–Stability–Error Propagation Framework proposed in Section 9 is offered as exactly such a positioning device, not as a claim that the field's fragmentation has already been resolved.

REFERENCES

1. Anderson, B. D. O. (1982). Reverse-time diffusion equation models. Stochastic Processes and Their Applications, 12(3), 313–326. https://doi.org/10.1016/0304-4149(82)90051-5

2. Chen, S., Chewi, S., Li, J., Li, Y., Salim, A., & Zhang, A. (2023). Sampling is as easy as learning the score: Theory for diffusion models with minimal data assumptions. International Conference on Learning Representations. (arXiv:2209.11215)
3. Conforti, G., Durmus, A., & Gentiloni Silveri, M. (2025). KL convergence guarantees for score diffusion models under minimal data assumptions. SIAM Journal on Mathematics of Data Science, 7(1), 86–109. https://doi.org/10.1137/23M1613670
4. De Bortoli, V., Thornton, J., Heng, J., & Doucet, A. (2021). Diffusion Schrödinger bridge with applications to score-based generative modeling. Advances in Neural Information Processing Systems, 34, 17695–17709.
5. Farghly, T., Rebeschini, P., Deligiannidis, G., & Doucet, A. (2026). Implicit regularisation in diffusion models: An algorithm-dependent generalisation analysis. International Conference on Learning Representations. (arXiv:2507.03756)
6. Ho, J., Jain, A., & Abbeel, P. (2020). Denoising diffusion probabilistic models. Advances in Neural Information Processing Systems, 33, 6840–6851.
7. Lee, H., Lu, J., & Tan, Y. (2023). Convergence of score-based generative modeling for general data distributions. Proceedings of the 34th International Conference on Algorithmic Learning Theory, 201, 946–985.
8. Li, G., Huang, Y., Efimov, T., Wei, Y., Chi, Y., & Chen, Y. (2024). Accelerating convergence of score-based diffusion models, provably. Proceedings of the 41st International Conference on Machine Learning, 235, 27942–27954. (arXiv:2403.03852)
9. Liu, K., & Zuazua, E. (2025). A PDE perspective on generative diffusion models (arXiv:2511.05940) [Preprint; peer-reviewed status not verified as of this review].
10. Pierret, E., & Galerne, B. (2025). Diffusion models for Gaussian distributions: Exact solutions and Wasserstein errors. Proceedings of the 42nd International Conference on Machine Learning, 267, 49355–49381.
11. Potaptchik, P., Azangulov, I., & Deligiannidis, G. (2025). Linear convergence of diffusion models under the manifold hypothesis. Proceedings of the 38th Annual Conference on Learning Theory, 291, 4668–4685.
12. Song, J., Meng, C., & Ermon, S. (2021). Denoising diffusion implicit models. International Conference on Learning Representations.
13. Song, Y., Sohl-Dickstein, J., Kingma, D. P., Kumar, A., Ermon, S., & Poole, B. (2021). Score-based generative modeling through stochastic differential equations. International Conference on Learning Representations.
14. Strasman, S., Cardoso, G., Le Corff, S., Lemaire, V., & Ocello, A. (2026). On forgetting and stability of score-based generative models (arXiv:2601.21868) [Preprint; peer-reviewed status not verified as of this review].
15. Zhang, Z., Chen, Z., & Gu, Q. (2025). Convergence of score-based discrete diffusion models: A discrete-time analysis. International Conference on Learning Representations. (arXiv:2410.02321)