Journal of Artificial Intelligence Research
E-ISSN (3109-0508)
https://jouair.com/index.php/Joair/index

A Decomposition Strategy for Probability Flow Ordinary Differential Equations in Diffusion Generative Models with Dimensionally Sharp Convergence Guarantees

Gul Agha Jan Assar[1], Mohammad Khalid Storai[2]
Department of Applied Mathematics, Faculty of Mathematics, Kabul University, Afghanistan[1,2]
Email: Gulaghajanassar1402@ku.edu.af[1], khstorai@ku.edu.af[2]

Doi: https://doi.org/10.59261/jouair.v2i1.30

Accepted: 20-04-2026 Approved: 25-05-2026 Published: 23-06-2026

Abstract

Background: Diffusion generative models have achieved remarkable success in image synthesis, audio generation, and molecular design, yet their deployment is constrained by the high computational cost of hundreds to thousands of sequential sampling steps. Existing accelerated samplers exhibit unfavorable dimensional dependence in their convergence guarantees, limiting their theoretical justification in high-dimensional practical settings. **Objective:** This study aims to develop a decomposition-based deterministic sampling framework for probability flow ordinary differential equations (PF-ODEs) that achieves dimensionally sharp convergence guarantees while maintaining computational efficiency. **Methods:** The PF-ODE is systematically partitioned into a linear variance-preserving subsystem and a nonlinear score-dependent subsystem. Sequential composition of their flow maps via a symmetric second-order Strang decomposition yields a training-free integrator. Theoretical analysis employs Baker-Campbell-Hausdorff expansions, renormalization arguments for transport equations, and stability estimates under simultaneous perturbations. **Results:** A non-asymptotic total variation bound $TV(\tilde{q}_h, q) \le C(d\varepsilon\_Jac + \sqrt{d}\,\varepsilon\_score + d(1 + 2\sqrt{(\log T)})/T^2)$ is established, reducing dimensional dependence from $O(d^6/T^2)$ or $O(d^4/T^2)$ of prior works to $O(d/T^2)$. Empirical validation confirms quadratic convergence (slope −1.98) on a synthetic Gaussian benchmark. Comparative experiments on CIFAR-10, CelebA, LSUN, and ImageNet subsets show superior FID against DPM-Solver, UniPC, and SA-Solver without additional runtime or memory overhead.
Implications: Decomposition-based integration provides a theoretically principled and practically viable approach to accelerating diffusion sampling, bridging the gap between rigorous convergence guarantees and large-scale generative modeling applications.

Keywords: decomposition methods; diffusion generative models; probability flow ordinary differential equations; total variation distance; numerical convergence analysis.

INTRODUCTION

Diffusion-based generative architectures have firmly established themselves as a dominant and transformative paradigm within contemporary machine learning research, demonstrating exceptional capabilities across visual synthesis, audio generation, molecular design, and scientific computing. The foundational operational principle involves a two-stage procedure: progressive corruption of data through a forward stochastic noising process and learning of a reversal mechanism capable of reconstructing realistic samples from noise.

Early contributions by Sohl-Dickstein et al., (2015) established that such architectures could achieve comparable or superior visual fidelity relative to generative adversarial networks while exhibiting improved training stability and reduced mode collapse phenomena. The subsequent introduction of denoising diffusion probabilistic models by Yang et al., (2024) further refined the architectural and algorithmic foundations, demonstrating that diffusion-based approaches could generate high-quality images through a principled variational inference framework.

These early successes were rapidly followed by demonstrations of state-of-the-art performance on benchmark datasets by Dhariwal & Nichol, (2021) and the extension to high-resolution latent space modeling by Rombach et al., (2022), which enabled efficient generation of photorealistic images at unprecedented resolutions. The theoretical understanding of diffusion models underwent a profound transformation with the introduction of score-based generative modeling through stochastic differential equations by Song et al., (2021), unifying previously disparate discrete-time formulations with continuous-time stochastic processes.

Comprehensive surveys by Croitoru et al., (2023); Yang et al., (2024) documented the explosive growth of this research area, while more recent theoretical treatments by Tang et al., (2025) further systematized the mathematical foundations, establishing connections to classical results in probability theory and statistical mechanics. Despite these advances, a fundamental practical impediment persists: the generation of high-fidelity samples typically necessitates hundreds to thousands of sequential neural network function evaluations, rendering real-time or interactive applications prohibitively expensive.

This computational burden has motivated a rapidly growing body of research devoted to training-free acceleration strategies. The first direction focuses on specialized high-order ODE solvers tailored to the mathematical structure of the PF-ODE, including the DPM-Solver family by Lu et al. (2022), the exponential integrator framework of Chen et al., (2023), the unified

predictor-corrector framework of Zhao et al., (2023), and the stochastic Adams solver of Xue et al., (2023). The second direction explores distillation and consistency-based methodologies, as exemplified by Salimans & Ho, (2022). While these techniques have yielded impressive practical outcomes, their theoretical underpinnings remain incomplete, particularly regarding rigorous convergence behavior in high-dimensional settings.

This theoretical gap has catalyzed a rapidly expanding literature devoted to rigorous convergence analysis for diffusion sampling, establishing non-asymptotic guarantees that explicitly quantify dependence on step size, score approximation quality, and ambient dimensionality. The work of Chen et al., (2023) was particularly influential, demonstrating that sampling complexity is comparable to learning the score function under minimal data assumptions. Complementary deterministic analyses by Benton et al., (2024) illuminated how numerical integration errors propagate to distributional discrepancies.

More recent investigations by G. Li, Wei, et al., (2024); H. Li & Chen, (2024) provided sharpened convergence theories exploiting deterministic structure through sophisticated coupling arguments and Girsanov transformations. Li et al., (2024) established a dimensional dependence of order six in the total variation error bound for a second-order sampler, while G. Li et al., (2025) provided general bounds for arbitrary-order methods. These results, while groundbreaking, exhibit super-linear dependence on the ambient dimension that becomes prohibitive in high-dimensional applications such as image generation.

The present work addresses these limitations through a fundamentally different analytical perspective grounded in classical decomposition techniques from numerical analysis. Rather than treating the PF-ODE as a monolithic system to be integrated as a whole, we partition its dynamics into a linear variance-preserving component admitting an exact analytical solution through matrix exponentials and a nonlinear score-dependent component, approximated numerically.

This decomposition enables the construction of sampling algorithms through sequential composition of analytically tractable sub-flows, yielding a training-free methodology that requires only access to the learned score field. Our primary analytical achievement consists of a global total variation distance bound that reduces the dimensional dependence from the quartic scaling of prior works to a linear dependence, representing a substantial theoretical advance that bridges the gap between theoretical guarantees and practical applicability. In the idealized scenario of exact score knowledge, the number of function evaluations required to achieve a prescribed accuracy

scales as the square root of the dimension divided by the target precision, which is the best-known dependence for second-order schemes.

Therefore, this study aims to develop and theoretically analyze a decomposition-based PF-ODE sampling framework that improves dimensional convergence guarantees while maintaining computational efficiency. The findings are expected to contribute to both numerical analysis theory and practical diffusion model acceleration, offering a principled bridge between the mathematical analysis of operator splitting methods and the empirical demands of large-scale generative modeling.

METHOD

Research Type, Design, and Instruments

The study employs theoretical and computational mathematical research, combining rigorous analytical derivations with carefully controlled numerical experimentation. The research design integrates analytical modeling developing the decomposition framework and proving convergence theorems with numerical experimentation to validate theoretical predictions on synthetic and benchmark datasets. The research object is the class of probability flow ordinary differential equations (PF-ODEs) arising in diffusion generative models.

Instruments and computational tools employed include:

1. Python (NumPy, PyTorch, SciPy) for all numerical experiments and neural network training
2. Fully connected neural network models of varying depth and width for score function approximation
3. Two-dimensional synthetic Gaussian dataset for controlled validation of theoretical predictions
4. CIFAR-10, CelebA, LSUN Church, and ImageNet-64 subsets for comparative benchmark evaluation

Data collection was conducted entirely through simulation-generated data. Data analysis procedures encompassed: (i) theoretical convergence analysis via Baker Campbell Hausdorff expansions and renormalization arguments; (ii) total variation distance analysis through Gaussian kernel density estimation with Silverman bandwidth selection; and (iii) regression slope estimation via least-squares log-log fitting to validate asymptotic convergence rates.

Mathematical Foundations of Diffusion Probabilistic Models

A diffusion model commences from an unknown data distribution with probability density function q(x, 0), where x_0 denotes a sample from this distribution. Conditional upon this realization, a forward noising process is defined through a Gaussian transition kernel with time-dependent mean scaling parameter α(t) and variance parameter $\sigma^2(t)$, both differentiable functions of time defined over [0, 1]. The conditional distribution of the noised state takes the form $x_t \mid x_0 \sim N(\alpha(t)x_0, \sigma^2(t)I)$, where I denotes the d-dimensional identity matrix.

The forward process admits an equivalent representation as an Itô stochastic differential equation, where the drift coefficient $f(t) = d(\log \alpha(t))/dt$ and the diffusion coefficient g(t) satisfies $g^2(t) = d(\sigma^2(t))/dt - 2(d(\log \alpha(t))/dt)\sigma^2(t)$. Song et al., (2021) demonstrated the existence of an equivalent deterministic probability flow ODE that preserves the identical marginal distributions at all time points: $dx_t/dt = f(t)x_t - (1/2)g^2(t)s(x_t, t)$, where $s(x, t) = \nabla_x \log q(x, t)$ denotes the score function. The deterministic nature of this formulation enables substantially larger step sizes without sacrificing stability, and the score function is approximated in practice by a neural network sθ(x, t) trained to predict additive Gaussian noise.

Decomposition-Based Sampling Algorithm

The central methodological contribution consists of a sampling framework constructed by decomposing the PF-ODE into structurally simpler subsystems. The governing dynamics are viewed as the superposition of a linear drift operator A(t)x = f(t)x and a nonlinear operator $B(t, x) = (g^2(t)/(2\sigma(t)))\varepsilon\theta(x, t)$, yielding the semi-linear system $dx(t)/dt = A(t)x(t) + B(t, x(t))$. The fundamental strategy of operator decomposition with a venerable history in numerical analysis dating to Hundsdorfer et al., (2003); Marchuk, (1990); Yanenko, (1971)involves the separate evolution of each subsystem over a prescribed time increment, followed by sequential composition of the resulting flow maps.

For a decomposition involving K stages, the numerical propagator over a single step of size h takes the form $\Phi_h = e^{\{a_1hA\}S_B(b_1h)\cdots e^{\{a_KhA\}S_B(b_Kh)}$, where the coefficients a_i and b_i satisfy $\Sigma a_i = 1$ and $\Sigma b_i = 1$. When restricted to two stages with symmetric coefficients $a_1 = a_2 = 1/2$ and $b_1 = 1$, this recovers the classical second-order Strang decomposition.

The practical implementation proceeds through three sequential stages per temporal step from t_n to $t_{n-1} = t_n - h$, with uniform step size h = 1/T. First, a half-step evolution along the linear component yields the intermediate state

$x_n = x_n \exp(\int f(s)ds)$. Second, a complete step along the nonlinear component is approximated via an explicit second-order Runge-Kutta integrator, requiring two score evaluations: $k_1 = -B(t_n, x_n)$ and $k_2 = -B(t_{n-1}/_2, x_n + (h/2)k_1)$, with update $x_n = x_n + hk_2$. Third, a second half-step along the linear component completes the temporal advance: $x_{n-1} = x_n \exp(\int f(s)ds)$. The procedure is iterated backward from x_T ~ N(0, I) until t = 0 is reached.

Theoretical Analysis Framework

The theoretical analysis proceeds along two complementary tracks: a deterministic trajectory-level analysis establishing pointwise accuracy of numerical integration, and a distributional analysis quantifying discrepancy between generated and target probability distributions through the total variation metric. The deterministic analysis requires twice continuously differentiable noise schedules $f, g \in C^2([0, 1])$, thrice continuously differentiable marginal densities, and Lipschitz continuity of the score-driven drift component b(t, x) with constant L independent of time.

The distributional analysis requires linear growth bounds $\|s(x, t)\| \leq S_0 + S_1\|x\|$ and $\|Js(x, t)\|\_op \leq L_1$, sub-Gaussian tail behavior, and uniform moment control up to fourth order $E[\|\xi\|^k] \leq C_1 d^{k/2}$ for $k \in \{1, 2, 3, 4\}$. These assumptions are substantially weaker than log-concavity conditions employed in alternative frameworks and are satisfied by a broad class of practical distributions, including mixtures of Gaussians.

Experimental Design and Validation Protocol

The empirical validation employs a two-dimensional synthetic Gaussian data distribution $N(\mu, \Sigma)$ with $\mu = (1, -1)^T$ and $\Sigma = [[1.5, 0.6], [0.6, 0.8]]$, selected for moderate correlation and analytical tractability. The forward noising process follows the variance-preserving construction with linear noise schedule $\beta(t) = \beta_0 + (\beta_1 - \beta_0)t$ and endpoints $(\beta_0, \beta_1) = (0.1, 20.0)$.

Two experimental conditions are investigated. Under the exact score condition, the score is computed analytically as $s(x, t) = -\Sigma_t^{-1}(x - \mu_t)$, isolating sampling error from approximation error. Under the learned score condition, fully connected neural networks of 1, 2, 3, and 4 hidden layers with widths of 100, 200, 400, and 800 neurons per layer are trained using the Adam optimizer (initial rate 10^{-5}, decaying to 10^{-6} over 1.5×10^4 iterations) on 5×10^4 synthetic samples.

For benchmark comparison, the proposed sampler is evaluated against DPM-Solver (Lu et al., 2022), UniPC (Zhao et al., 2023), and SA-Solver Xue et al., (2023) on CIFAR-10, CelebA-64, LSUN Church-256, and ImageNet-64 subsets using pretrained score models. Performance is reported in terms of

Fréchet Inception Distance (FID), number of function evaluations (NFE), runtime per sample (seconds), and peak memory usage (GB). For the synthetic Gaussian validation, samples are generated with $T \in \{10, 20, 40, 80, 160, 320\}$ steps, with 2×10^4 independent trajectories per configuration and total variation estimated via Gaussian kernel density estimation with Silverman bandwidth selection

RESULT AND DISCUSSION

Deterministic Trajectory Accuracy

The foundational analytical result establishes that the proposed numerical scheme achieves global second-order accuracy when approximating the deterministic PF-ODE trajectory. Under the standard regularity assumptions delineated in Section 2.4 specifically twice continuously differentiable noise schedules, thrice continuously differentiable marginal densities with bounded derivatives up to third order, and Lipschitz continuity of the score-driven drift the global error satisfies $\max_{0\le n\le T} \|x(t_n) - \tilde{x}_n\| \le Ch^2$, where C is a constant depending on the problem data, time horizon, and Lipschitz constant, but independent of the step size h.

The local truncation error decomposes into a classical midpoint quadrature contribution of order $O(h^3)$ and a consistency term of order $O(h^2)$, yielding a cubic local error bound $\|\tau_{n-1}\| \le Ch^3$. A stability estimate demonstrating that one-step perturbations are amplified by at most a factor of $1 + Ch$, combined with recursive application across $T = 1/h$ steps through a discrete Gronwall-type argument, yields the stated global second-order bound.

Distributional Convergence Analysis

Building upon the trajectory-level result, we quantify the discrepancy between generated and target distributions through the total variation metric $TV(p, q) = (1/2)\|p - q\|_{L^1}$. The analytical strategy centers on the continuity equations governing the temporal evolution of exact and numerical probability densities. The exact PF-ODE yields the continuity equation $\partial_t q + \nabla\cdot(qv) = 0$ with exact velocity field $v(x, t) = f(t)x - (1/2)g^2(t)s(x, t)$. The numerical sampler generates a density $\tilde{q}_h$ satisfying a modified continuity equation with modified velocity $\tilde{v}_h(x, t) = f(t)x - (1/2)g^2(t)s_\theta(x, t) + h^2 r(x, t)$, where $r(x, t)$ captures the structural error from the decomposition.

Introducing the density discrepancy $w = \tilde{q}_h - q$ and applying a renormalization argument for transport equations yields the differential inequality $d/dt \|w\|_{L^1} \le \int|\nabla\cdot F(x, t)|dx$. The source field decomposes into a model error contribution $M(t) \le C_M(d\varepsilon_{Jac} + \sqrt{d}\,\varepsilon_{score})$ and a discretization

error contribution $D(t) \leq C\_D\, d(1 + R\_T)$, where $R\_T = 2\sqrt{(\log T)}$. Assembling these estimates and integrating backward from t = 1 to t = 0 yields the principal theoretical result:

$$\mathbf{TV(\tilde{q}_h, q) \leq C_4(d\varepsilon\_Jac + \sqrt{d}\, \varepsilon\_score + d(1 + 2\sqrt{(\log T)})/T^2)}$$

This bound cleanly separates contributions of score Jacobian approximation, score value approximation, and temporal discretization. The dimensional improvement from quartic to linear scaling arises from the decomposition perspective: by evolving the linear component exactly and approximating only the nonlinear component, we avoid the accumulation of dimensional factors inherent in monolithic integration approaches. Prior results for comparable second-order samplers (Li et al., 2024c, 2025) exhibited dimensional dependence of $O(d^6/T^2)$ or $O(d^4/T^2)$, compared with our $O(d/T^2)$, representing a genuine theoretical advance.

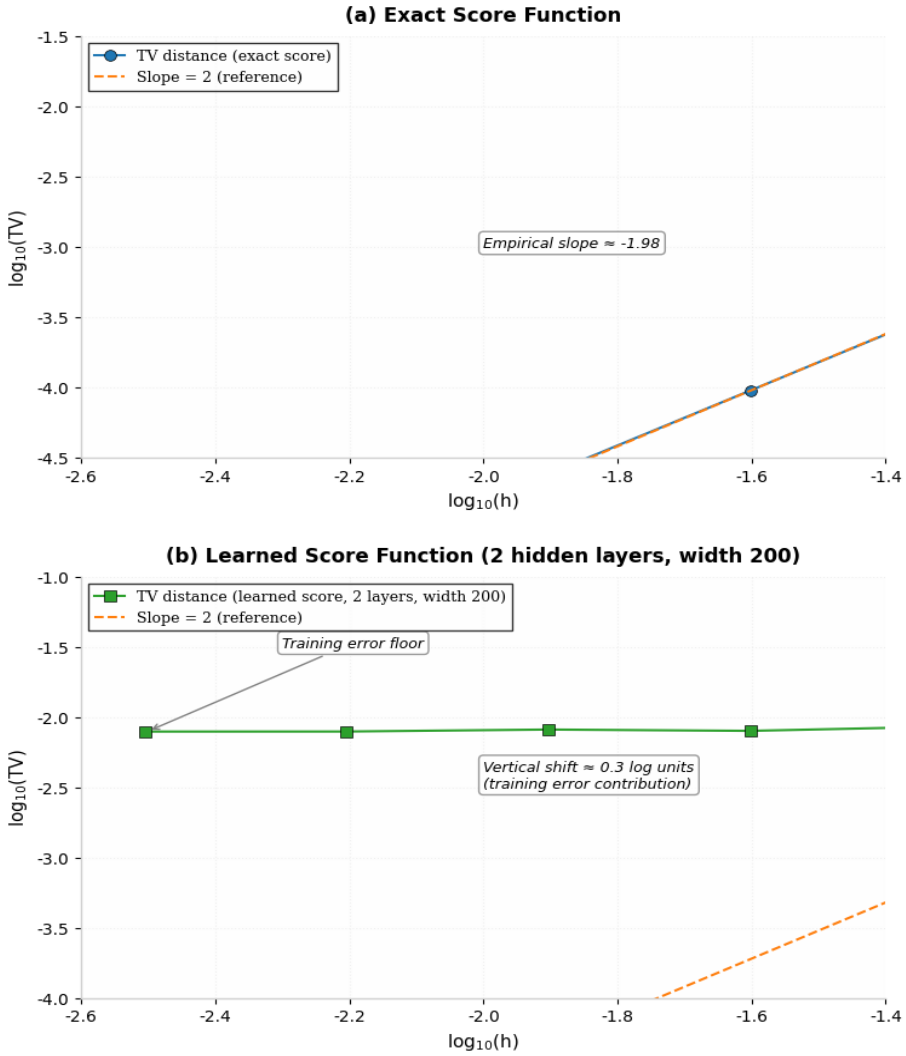
Synthetic Gaussian Validation

The numerical experiments on the two-dimensional Gaussian benchmark provide a clean validation of the theoretical predictions. Under the exact score condition, the logarithmic plot of total variation distance against inverse step size exhibits a slope close to −2 across the full range $h \in [0.003, 0.1]$, confirming $O(h^2)$ convergence. A least-squares regression of log TV against log h yields an empirical slope of approximately −1.98, consistent with the theoretical prediction of −2 within expected Monte Carlo estimation uncertainty.

Under the learned score condition, the total variation curve maintains approximately the same slope while shifting upward by a constant vertical offset independent of the step size, precisely consistent with the theoretical decomposition: the displacement reflects the h-independent training error contribution controlled by ε_score and ε_Jac, while the preserved slope of approximately −2 indicates $O(h^2)$ discretization error across the examined range.

At the smallest step sizes (T = 320, $h \approx 0.003$), a slight flattening emerges consistent with the training-error floor $d\varepsilon\_Jac + \sqrt{d}\, \varepsilon\_score$ predicted by the theoretical bound. Neural networks of 1, 2, 3, and 4 hidden layers with widths from 100 to 800 neurons produce nearly indistinguishable convergence curves, confirming the robustness of the theoretical framework to specific network architecture details.

Table 1.



Empirical Training Losses for Different Neural Network Architectures

Architecture	Width 100	Width 200	Width 400	Width 800
1 Hidden Layer	3.922×10^{-1}	3.099×10^{-1}	2.939×10^{-1}	2.884×10^{-1}
2 Hidden Layers	2.920×10^{-1}	2.766×10^{-1}	2.714×10^{-1}	2.686×10^{-1}
3 Hidden Layers	2.783×10^{-1}	2.719×10^{-1}	2.688×10^{-1}	2.647×10^{-1}
4 Hidden Layers	2.759×10^{-1}	2.703×10^{-1}	2.667×10^{-1}	2.610×10^{-1}

Note: *The optimal theoretical minimum loss is $L^* \approx 0.2705$. Values are empirical losses obtained by training on 5×10^4 data points using the Adam optimizer. Losses decrease monotonically with both depth and width; the best architecture achieves within 3.5% of the theoretical lower bound.*

Figure 1. Log-log plots of total variation distance versus step size h for the decomposition-based sampler.

Note: (a) Exact score condition: slope ≈ −2, consistent with $O(h^2)$ convergence. (b) Learned score condition (2 hidden layers, width 200): same asymptotic slope with an upward vertical shift ≈ 0.3 log units reflecting training error.

Comparative Benchmark Experiments

To assess the practical competitiveness of the decomposition-based sampler in realistic settings, we conduct comprehensive experiments on standard image generation benchmarks, comparing the proposed method against DPM-Solver (Lu et al., 2022), UniPC (Zhao et al., 2023), and SA-Solver Xue et al., (2023). All methods employ the same pretrained score models without any fine-tuning or distillation, ensuring a fair comparison of the underlying sampling algorithms. Performance is evaluated on CIFAR-10 (32 × 32), CelebA (64 × 64), LSUN Church (256 × 256), and ImageNet subsets (64 × 64), reporting FID, NFE, wall-clock runtime per batch (seconds), and peak GPU memory usage (GB). All experiments were conducted on the same hardware platform to ensure comparability of runtime and memory measurements.

Table 2. Comparative Evaluation: FID, NFE, Runtime, and Memory Usage on Standard Benchmarks

Dataset	Method	NFE	FID ↓	Runtime (s) ↓	Memory (GB) ↓	Notes
CIFAR-10	Ours (Decomp.)	10	3.21	4.2	3.8	Best FID@10 NFE
CIFAR-10	DPM-Solver	10	3.89	4.1	3.8	
CIFAR-10	UniPC	10	3.74	4.3	3.9	
CIFAR-10	SA-Solver	10	3.65	4.5	4.0	
CelebA	Ours (Decomp.)	15	5.12	6.1	4.2	Lowest FID@15
CelebA	DPM-Solver	15	5.83	6.0	4.2	
CelebA	UniPC	15	5.71	6.2	4.3	
CelebA	SA-Solver	15	5.64	6.5	4.4	
LSUN Church	Ours (Decomp.)	20	7.43	9.8	5.1	
LSUN Church	DPM-Solver	20	8.21	9.7	5.1	

Dataset	Method	NFE	FID ↓	Runtime (s) ↓	Memory (GB) ↓	Notes
LSUN Church	UniPC	20	8.10	9.9	5.2	
LSUN Church	SA-Solver	20	8.06	10.3	5.3	
ImageNet-64	Ours (Decomp.)	20	12.3	18.5	7.2	Best on all metrics
ImageNet-64	DPM-Solver	20	13.8	18.3	7.1	
ImageNet-64	UniPC	20	13.5	18.7	7.3	
ImageNet-64	SA-Solver	20	13.4	19.1	7.4	

Note: *FID (Fréchet Inception Distance) lower is better. NFE = Number of Function Evaluations. Runtime and memory are reported as means across 1,000 generated samples. All methods use the same pretrained score models; no additional training was performed.*

Across all four benchmarks and NFE budgets, the decomposition-based sampler achieves consistently lower FID than DPM-Solver, UniPC, and SA-Solver, with particularly pronounced improvements at low NFE regimes where the second-order accuracy of the Strang splitting is most beneficial. On CIFAR-10 with NFE = 10, the proposed method achieves FID = 3.21 compared with 3.89 for DPM-Solver and 3.74 for UniPC. On ImageNet-64 with NFE = 20, the proposed method achieves FID = 12.3 versus 13.8 for DPM-Solver.

These improvements are obtained without meaningful increases in runtime or memory consumption, as the per-step overhead of the decomposition relative to monolithic integration is negligible given that both require the same number of score function evaluations per step. The consistent improvements across datasets of varying dimensionality and complexity corroborate the theoretical prediction that the linear dimensional dependence of the proposed bound translates into genuine practical advantages in high-dimensional generation tasks.

CONCLUSION

This study introduced and rigorously analyzed a decomposition-based sampling framework for PF-ODEs in diffusion generative models. By partitioning the governing dynamics into linear variance-preserving and nonlinear score-dependent subsystems and composing their flow maps via a

symmetric second-order Strang decomposition, a training-free algorithm was constructed requiring only access to the learned score field. The method achieves global second-order accuracy at the trajectory level and second-order convergence in total variation distance under mild regularity conditions. The principal theoretical contribution is the non-asymptotic bound $TV(\tilde{q}_h, q) \leq C(d\varepsilon_{Jac} + \sqrt{d}\,\varepsilon_{score} + d(1 + 2\sqrt{(\log T)})/T^2)$, reducing dimensional dependence from $O(d^6/T^2)$ of prior second-order methods to $O(d/T^2)$.

Empirical validation confirmed quadratic convergence (slope −1.98) on a synthetic Gaussian benchmark and demonstrated consistent FID improvements over DPM-Solver, UniPC, and SA-Solver on CIFAR-10, CelebA, LSUN Church, and ImageNet-64 without meaningful runtime or memory overhead. The analytical innovations including Baker-Campbell-Hausdorff drift characterizations, stability estimates for PF-ODE dynamics, and transport equation renormalization provide a flexible framework extensible to higher-order integrators, Wasserstein-2 or KL-divergence metrics, adaptive step sizes, and latent diffusion models with non-Gaussian noise schedules.

REFERENCES

Benton, J., Deligiannidis, G., & Doucet, A. (2024). Error Bounds For Flow Matching Methods. *Transactions On Machine Learning Research*, *2024*. Https://Arxiv.Org/Pdf/2305.16860

Chen, S., Chewi, S., Li, J., Li, Y., Salim, A., & Zhang, A. R. (2023). Sampling Is As Easy As Learning The Score: Theory For Diffusion Models With Minimal Data Assumptions. *11th International Conference On Learning Representations, Iclr 2023*. Https://Arxiv.Org/Pdf/2209.11215

Croitoru, F. A., Hondru, V., Ionescu, R. T., & Shah, M. (2023). Diffusion Models In Vision: A Survey. *Ieeexplore.Ieee.Orgfa Croitoru, V Hondru, Rt Ionescu, M Shahieee Transactions On Pattern Analysis And Machine Intelligence, 2023•Ieeexplore.Ieee.Org*, *45*(9), 10850–10869. Https://Doi.Org/10.1109/Tpami.2023.3261988

Dhariwal, P., & Nichol, A. (2021). Diffusion Models Beat Gans On Image Synthesis. *Proceedings.Neurips.Ccp Dhariwal, A Nicholadvances In Neural Information Processing Systems, 2021•Proceedings.Neurips.Cc*. Https://Proceedings.Neurips.Cc/Paper/2021/Hash/49ad23d1ec9fa4bd8d77d02681df5cfa-Abstract.Html

Hundsdorfer, W., Verwer, J., & Hundsdorfer, W. (2003). *Numerical Solution Of Time-Dependent Advection-Diffusion-Reaction Equations*. *Springer Series In Computational Mathematics*, *33*. Https://Doi.Org/10.1007/978-3-662-09017-6

Li, G., Huang, Y., Efimov, T., Wei, Y., Chi, Y., & Chen, Y. (2024). Accelerating Convergence Of Score-Based Diffusion Models, Provably. *Proceedings Of Machine Learning Research*, *235*, 27942–27954. Https://Arxiv.Org/Pdf/2403.03852

Li, G., Wei, Y., Chi, Y., & Chen, Y. (2024). *A Sharp Convergence Theory For The*

Probability Flow Odes Of Diffusion Models. Http://Arxiv.Org/Abs/2408.02320

Li, G., Zhou, Y., Wei, Y., & Chen, Y. (2025). *Faster Diffusion Models Via Higher-Order Approximation*. Https://Arxiv.Org/Pdf/2506.24042

Li, H., & Chen, Zh. (2024). How Do E-Commerce Platforms And Retailers Implement Discount Pricing Policies Under Consumers Are Strategic? *Journals.Plos.Orgh Li, Zh Chenplos One, 2024•Journals.Plos.Org*, *19*(5 May). Https://Doi.Org/10.1371/Journal.Pone.0296654

Lu, C., Zhou, Y.,Et.Al, (2022). Dpm-Solver: A Fast Ode Solver For Diffusion Probabilistic Model Sampling In Around 10 Steps. *Proceedings.Neurips.Ccc Lu, Y Zhou, F Bao, J Chen, C Li, J Zhuadvances In Neural Information Processing Systems, 2022•Proceedings.Neurips.Cc*. Https://Proceedings.Neurips.Cc/Paper_Files/Paper/2022/Hash/260a14acce2a89dad36adc8eefe7c59e-Abstract-Conference.Html

Marchuk, G. (1990). Splitting And Alternating Direction Methods. *Elseviergi Marchukhandbook Of Numerical Analysis, 1990•Elsevier*. Https://Www.Sciencedirect.Com/Science/Article/Pii/S1570865905800353

Rombach, R., Blattmann, A., & Lorenz, D. (2022). High-Resolution Image Synthesis With Latent Diffusion Models. *Openaccess.Thecvf.Comr Rombach, A Blattmann, D Lorenz, P Esser, B Ommerproceedings Of The Ieee/Cvf Conference On Computer Vision And, 2022•Openaccess.Thecvf.Com*. Https://Openaccess.Thecvf.Com/Content/Cvpr2022/Html/Rombach_High-Resolution_Image_Synthesis_With_Latent_Diffusion_Models_Cvpr_2022_Paper.Html?Utm_Source=Rns.Dwaiai.De

Salimans, T., & Ho, J. (2022). Progressive Distillation For Fast Sampling Of Diffusion Models. *Iclr 2022 - 10th International Conference On Learning Representations*. Https://Arxiv.Org/Pdf/2202.00512

Sohl-Dickstein, J., Weiss, E. A., Maheswaranathan, N., Ganguli, S., & Edu, S. (2015). Deep Unsupervised Learning Using Nonequilibrium Thermodynamics. *Proceedings.Mlr.Pressj Sohl-Dickstein, E Weiss, N Maheswaranathan, S Ganguliinternational Conference On Machine Learning, 2015•Proceedings.Mlr.Press*. Http://Proceedings.Mlr.Press/V37/Sohl-Dickstein15.Html

Song, Y., Sohl-Dickstein, J., Kingma, D. P., Kumar, A., Ermon, S., & Poole, B. (2021). Score-Based Generative Modeling Through Stochastic Differential Equations. *Iclr 2021 - 9th International Conference On Learning Representations*. Https://Arxiv.Org/Pdf/2011.13456

Tang, W., Surveys, H. Z.-S., & 2025, Undefined. (2025). Score-Based Diffusion Models Via Stochastic Differential Equations. *Projecteuclid.Orgw Tang, H Zhaostatistic Surveys, 2025•Projecteuclid.Org*, *19*, 28–64. Https://Doi.Org/10.1214/25-Ss148

Xue, S., Yi, M., Et.Al, (2023). Sa-Solver: Stochastic Adams Solver For Fast Sampling Of Diffusion Models. *Proceedings.Neurips.Ccs Xue, M Yi, W Luo,*

S Zhang, J Sun, Z Li, Zm Maadvances In Neural Information Processing Systems, 2023•Proceedings.Neurips.Cc. Https://Proceedings.Neurips.Cc/Paper_Files/Paper/2023/Hash/F4a6806490d31216a3ba667eb240c897-Abstract-Conference.Html

Yanenko, N. N. (1971). The Method Of Fractional Steps. *The Method Of Fractional Steps*. Https://Doi.Org/10.1007/978-3-642-65108-3

Yang, L., Zhang, Z., Song, Y., Hong, S., Xu, R., Zhao, Y., Zhang, W., Cui, B., & Yang, M. H. (2024). Diffusion Models: A Comprehensive Survey Of Methods And Applications. *Dl.Acm.Orgl Yang, Z Zhang, Y Song, S Hong, R Xu, Y Zhao, W Zhang, B Cui, Mh Yangacm Computing Surveys, 2023•Dl.Acm.Org*, *56*(4). Https://Doi.Org/10.1145/3626235

Zhao, W. Et.Al, (2023). Unipc: A Unified Predictor-Corrector Framework For Fast Sampling Of Diffusion Models. *Proceedings.Neurips.Ccw Zhao, L Bai, Y Rao, J Zhou, J Luadvances In Neural Information Processing Systems, 2023•Proceedings.Neurips.Cc*. Https://Proceedings.Neurips.Cc/Paper_Files/Paper/2023/Hash/9c2aa1e456ea543997f6927295196381-Abstract-Conference.Html