Resolving Score Singularity to Enhance Diffusion Model Performance

Weizhong Wang

This version: December 13, 2025

Abstract

We present a refined decomposition of the score function in diffusion models into singular and non-singular components with respect to diffusion time. This separation elucidates the small-time behavior of the score, a regime where existing methods often exhibit instability and approximation errors. We show that the non-singular term is determined by intrinsic parameters that have been largely neglected. Motivated by this finding, we introduce a two-network architecture that learns the two components separately under a theory-driven optimization framework. Experiments demonstrate consistent improvements over the standard DDPM baseline, highlighting the advantages of disentangling the score into its fundamental parts.

Keywords: diffusion model, score singularity, score decomposition, manifold learning

[This manuscript is a working paper for our project on the analysis of score singularities in diffusion models. Although the presentation is still being refined, the theoretical analysis in Section 3 is fully developed and forms a firm theoretical backbone of the work.]

1 Introduction

Diffusion models (Song and Ermon (2019); Ho et al. (2020); Song et al. (2021)) have achieved state-of-the-art performance in image synthesis (Dhariwal and Nichol (2021); Rombach et al. (2022a)) and video generation (Ho et al. (2022); Melnik et al. (2024)), underpinning advanced image generation systems such as DALL-E-2 (Ramesh et al. (2022)) and Stable Diffusion (Rombach et al. (2022b)). More recently, Diffusion Transformers (DiTs) (Peebles and Xie (2022); Zhu et al. (2024); Ma et al. (2024); Chen et al. (2025)) have attracted increasing attention, extending traditional transformers by integrating diffusion processes with autoregressive generation of text and images.

Beyond Euclidean spaces, many scientific fields focus on data lying on low-dimensional Riemannian manifolds, valued for their reduced complexity and broad applicability. Algorithms tailored to such manifolds have advanced diverse areas of machine learning, including dimensionality reduction (Coifman and Lafon (2006)), generative modeling (Kingma and Welling (2013)), manifold learning (Meilă and Zhang (2024)), and robot learning (Liu et al. (2022)). Recent work further shows that exploiting these low-dimensional structures can significantly enhance diffusion models (De Bortoli (2022), Tang and Yang (2024), Azangulov et al. (2024)).

Realizing these benefits is often constrained by singularities in the score function, a challenge extensively documented Chen et al. (2023); Yang et al. (2024); Lu et al. (2023); Liu et al. (2025a). While prior work has characterized the singular term, the structure of the non-singular component remains largely unexplored. Current understanding is limited to its boundedness at $O(1)$ Liu et al. (2025a); Lu et al. (2023). In diffusion models, the score function further diverges at small diffusion scales, hindering neural network training and degrading denoising performance. Existing remedies include heat kernel formulations Lou et al. (2023), consistent time embeddings Yang et al. (2024), and the use of intrinsic manifold features Liu et al. (2025b). However, none leverage the intrinsic geometry of the unknown data manifold.

Main contributions. We study diffusion models governed by a drift-free stochastic differential equation (SDE), $d\mathbf{x}_t = \sigma(t)\, d\mathbf{w}_t$. Our contributions are:

1. We refine prior score decomposition analyses (Chen et al. (2023); Lu et al. (2023); Stanczuk et al. (2024)) by deriving an explicit formulation separating the score into singular and non-singular components w.r.t. diffusion time t, providing the first characterization of the non-singular term. The singular component diverges as $t \to 0$, while the non-singular component remains bounded by a constant determined by the data distribution, manifold geometry, diffusion time, and the sample's distance from the manifold.

2. We propose a two-stage algorithm leveraging this decomposition: for large t, training and sampling follow a standard diffusion model; for small t, two neural networks separately estimate the singular and non-singular components. Experiments show improved performance across diverse manifolds.

Related works. Diffusion models on Riemannian manifolds have been studied to address the inefficiency and inaccuracy of early Euclidean approaches (Song and Ermon (2019); Ho et al. (2020)), which neglect low-dimensional manifold structure. The first attempt, Huang et al. (2022), introduced a variational framework and a Riemannian continuous-time ELBO for training. Lou et al. (2023) proposed efficient heat kernel approximations on Riemannian symmetric spaces, while Jo and Hwang (2024) avoided heat kernel computation via two-way bridge matching based on drift regression.

Theoretical analyses show that manifold geometry strongly influences score estimation (Block et al. (2022); Liu et al. (2025b)). De Bortoli (2022) proved that the Wasserstein distance between the forward process and the data distribution can be bounded by intrinsic manifold properties, assuming L_2 score estimation error. Tang and Yang (2024); Azangulov et al. (2024) established that convergence rates depend on intrinsic dimension.

Applications further exploit this geometry–score relationship. Stanczuk et al. (2024) found that for small diffusion time t, the score function is orthogonal to the tangent space, revealing a score singularity that encodes intrinsic dimensionality. Chung et al. (2022) showed that manifold-constrained gradients improve performance. Azeglio and Di Bernardo (2025); Saito and Matsubara (2025) designed Riemannian metrics from diffusion scores, highlighting their intrinsic link to manifold geometry.

Score singularity remains a practical challenge. Lu et al. (2023); Chen et al. (2023) first identified the issue and proposed mitigation strategies, though Chen et al. (2023) addressed only hyperplanes. Liu et al. (2025a,b) proposed manifold-specific solutions requiring prior geometric knowledge, limiting applicability. Here, we address score singularities without assuming prior knowledge of the underlying manifold.

Outline. The remainder of the paper is organized as follows. Section 2 reviews the diffusion model and presents the problem setup. Section 3 provides the theoretical analysis underlying our main result on score decomposition. Section 4 introduces a method to address score singularity based on the preceding theory. Section 5 demonstrates the approach on real datasets. Section 6 concludes the paper.

Notations. Let $\mathcal{N}(\mu, \Sigma)$ denote the normal distribution with mean μ and variance Σ. We write $a = \mathcal{O}(b)$ if there exists $C > 0$ such that $a \leq Cb$, and $f(x) = o(g(x))$ if $f(x)/g(x) \to 0$ as $x \to 0$.

Unless otherwise stated, $\|\mathbf{x}\|$ denotes the Euclidean (2-)norm $\|\mathbf{x}\|_2$. Let $S \subset \mathbb{R}^d$ be a set. Denote $S - \mathbf{z}^* := \{\mathbf{z} - \mathbf{z}^* : \mathbf{z} \in S\}$, $kS := \{k\mathbf{z} : \mathbf{z} \in S\}$. $[\cdot]$ denotes the floor function.

2 Set-up and assumptions

In this section, we provide necessary backgrounds on diffusion models and elementary assumptions of data manifold and distribution.

2.1 Forward and backward SDEs

The forward process in diffusion models incrementally perturbs data with noise. We consider the drift-free Brownian motion, $d\mathbf{X}_t = \sigma(t)\, d\mathbf{W}_t, \quad \sigma(t) > 0$, where $\mathbf{X}_0 \sim P_{\text{data}}$, $(\mathbf{W}_t)_{t\geq 0}$ is a standard d-dimensional Wiener process, and $\sigma(t)$ is a time-dependent diffusion coefficient. The marginal at time t is denoted P_t. This formulation perturbs data solely via Gaussian noise with variance $\sigma^2(t)$. Given $\mathbf{X}_0$, the conditional distribution of $\mathbf{X}_t$ is $\mathcal{N}\Big(\mathbf{X}_0, \Big(\int_0^t \sigma^2(s)\, ds\Big)\, \mathbf{I}_D\Big)$. Under mild conditions, P_t converges to a Gaussian as $t \to \infty$.

In practice, the process stops at a finite T, where $P_T \approx \mathcal{N}(0, T\,\mathbf{I}_D)$. Sampling proceeds by reversing the dynamics, yielding the backward SDE, $d\mathbf{X}_t^{\leftarrow} = \sigma^2(T-t)\nabla_{\mathbf{x}} \log p_{T-t}(\mathbf{X}_t^{\leftarrow})\, dt + \sigma(T-t)\, d\overline{\mathbf{W}}_t$, where $\nabla \log p_t$ is the score function and $\overline{\mathbf{W}}_t$ is a reversed Wiener process. When initialized with P_T, the backward process has the same law as the time reversal of the forward process (Anderson (1982)).

Direct simulation is infeasible since P_T and $\nabla \log p_t$ are unknown. In practice, P_T is replaced by $\mathcal{N}(0, \mathbf{I}_D)$ and the score is approximated by a neural estimator $\hat{\mathbf{s}}$, giving $d\tilde{\mathbf{X}}_t^{\leftarrow} = \sigma^2(T-t)\, \hat{\mathbf{s}}(\tilde{\mathbf{X}}_t^{\leftarrow}, T-t)\, dt + \sigma(T-t)\, d\overline{\mathbf{W}}_t$, $\tilde{\mathbf{X}}_0^{\leftarrow} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_D)$. A discretized version with step size η is used in generation: $d\tilde{\mathbf{X}}_t^{\leftarrow} = \sigma^2(T-t)\, \hat{\mathbf{s}}(\tilde{\mathbf{X}}_{k\eta}^{\leftarrow}, T-k\eta)\, dt + \sigma(T-t)\, d\overline{\mathbf{W}}_t$, $t \in [k\eta, (k+1)\eta]$.

Our analysis considers the general drift-free Brownian motion presented at the beginning of this section. Experiments employ the Variance Exploding (VE) SDE, $d\mathbf{X}_t = \sqrt{\frac{d\sigma_t^2}{dt}}\, d\mathbf{W}_t$, $\sigma_t = \sigma_{\min}\left(\frac{\sigma_{\max}}{\sigma_{\min}}\right)^{t/T}$, for which $p_t(\mathbf{X}_t) = \mathcal{N}(\mathbf{X}_0, \sigma_t^2 \mathbf{I})$.

2.2 Data manifold and distribution

Let $\mathcal{M} \subset \mathbb{R}^D$ be a compact, d-dimensional ($d \geq 2$) Riemannian manifold without boundary, endowed with metric g and volume form dV. The data distribution p_{data} has compact support on $\mathcal{M}$, and $\mathbf{x}_0 \sim p_{\text{data}}$. Let p_t denote the marginal distribution of the diffusion process at time $t > 0$.

To simplify theoretical analysis on the manifold, we impose the following assumptions.

Assumption 2.1. *Each $\mathbf{x} \in \mathcal{M}$ can be expressed as $\mathbf{x} = \psi(\mathbf{z})$, where $\psi : S \subset \mathbb{R}^d \to \mathcal{M}$ is an isometric embedding, S is open in $\mathbb{R}^d$, and $\mathbf{z} \in S$ is distributed according to P_z with a continuously differentiable density p_z. The density satisfies $p_z \leq p_{\max} < \infty$ and $|\nabla p_z| \leq D_U$ for constants $p_{\max}, D_U > 0$.*

Assumption 2.2. *For a given $\mathbf{z} \in S$, the Jacobian J_ψ of ψ is L-Lipschitz in a neighborhood of any $\mathbf{z}$: $\|J_\psi(\mathbf{z}) - J_\psi(\mathbf{z}')\|_2 \leq L\|\mathbf{z} - \mathbf{z}'\|_2$, $\forall \mathbf{z}' \in B_\delta(\mathbf{z}) \cap S$, where $B_\delta(\mathbf{z}) = \{\mathbf{z}' \in \mathbb{R}^d : \|\mathbf{z} - \mathbf{z}'\|_2 \leq \delta\}$.*

Our objective is to compute the score function $\nabla \log p_t(\mathbf{x}_t)$.

3 Score decomposition

We show that for a low-dimensional data distribution, the score function can be decomposed, with one term of closed form and one term can be bounded by constant depending on diffusion time t, distance from the manifold, intrinsic quantity of manifold and property of original distribution. Such property will be utilized in establishing the algorithm to tackle singularity of score when $t \to 0$. Such decomposition is summarized into the following theorem.

Theorem 3.1 (Score decomposition). *Assume Assumptions 2.1 and 2.2 hold, and let* $\mathbf{z}^* = \arg\min_{\mathbf{z}\in S} \|\mathbf{x} - \psi(\mathbf{z})\|_2^2$ *with Lipschitz radius* r. *Define* $s^2 = \int_0^t \sigma^2(\tau)\, d\tau$, $\mathbf{d} = \psi(\mathbf{z}^*) - \mathbf{x}$, *and* $P = \frac{p_z(\mathbf{z}^*)}{p_{\max}}$. *Suppose* $p_z(\mathbf{z}^*) > 0$. *Choose constants* A, B *such that* $P \geq A\left(B + \frac{2}{\sqrt{\pi}}\right) d$ *and* $A < \min\left\{\frac{1}{2}, \frac{1}{4(d-1)}, \frac{Lr}{4}\right\}$. *Assume* $L\|\mathbf{d}\| < A$ *and* $s < \min\left\{\frac{r}{4\sqrt{\frac{d-1}{2}}}, B, \frac{p_z(\mathbf{z}^*)}{4\sqrt{2\pi}\left(\frac{p_z(\mathbf{z}^*)}{r} + D_U\right)}\right\}$. *Then the score function satisfies*

$$\left\|\nabla \log p_t(\mathbf{x}) - \frac{1}{s^2}(\psi(\mathbf{z}^*) - \mathbf{x})\right\|_2 \leq C(d)\left(1 + \frac{1}{r^2} + \frac{1}{r^3}\right)\left(\frac{1}{P} + \frac{1}{P^2}\right) \times \left[(s^2 + A + 1)\left(\frac{Lr + 1 + \|J_\psi(\mathbf{z}^*)\|_2}{Pr} + \frac{s}{r^2} + \frac{D_U}{p_z(\mathbf{z}^*)}\right) + s + \frac{A^2}{s}\right], \tag{3.1}$$

where $C(d)$ *depends only on the intrinsic dimension* d.

The proof of Theorem 3.1 is divided into three parts, each corresponding to one of the following lemmas. The theorem follows directly by combining these lemmas. Proofs of the three lemmas are in Appendix A.1. From the three lemmas, we can also derive that the non-singular component is composed of three components: one term in the tangent space of the projection point, one term in the normal space and one diminishing term w.r.t. time. This provides an further insight into the components of the non-singular term, further than Lu et al. (2023); Liu et al. (2025a).

Lemma 3.1. *Under the setting of Theorem 3.1, let* $\phi_t(\cdot \mid \mathbf{z}) = (2\pi s^2)^{-d/2} \exp\left(-\frac{\|\cdot-\mathbf{z}\|_2^2}{2s^2}\right)$ *denote the Gaussian density, and define*

$$\hat{\mathbf{s}}_{\|}(\mathbf{x}, t) = \frac{\int \frac{\psi(\mathbf{z}) - \psi(\mathbf{z}^*)}{s^2}\, \phi_t(\mathbf{z}^* \mid \mathbf{z})\, p_z(\mathbf{z})\, d\mathbf{z}}{\int \phi_t(\mathbf{z}^* \mid \mathbf{z})\, p_z(\mathbf{z})\, d\mathbf{z}}.$$

Then

$$\left\|\nabla \log p_t(\mathbf{x}) - \hat{\mathbf{s}}_{\|}(\mathbf{x}, t) - \frac{\psi(\mathbf{z}^*) - \mathbf{x}}{s^2}\right\|_2 \leq C(d)\left(1 + \frac{1}{r^2} + \frac{1}{r^3}\right)\left(\frac{1}{P} + \frac{1}{P^2}\right)\left[(s^2 + A)\, \|\hat{\mathbf{s}}_{\|}(\mathbf{x}, t)\|_2 + s + \frac{A^2}{s}\right],$$

where $C(d)$ *depends only on the intrinsic dimension* d.

Lemma 3.1 shows that the score function can be decomposed into a term of the similar form as a score function on the data manifold, plus a singular term, with the approximation error vanishing as $t \to 0$. The former can thus be interpreted as a manifold-level score.

Lemma 3.2. *Under the conditions of Theorem 3.1, the difference between* $\hat{\mathbf{s}}_{\|}(\mathbf{x}, t)$ *and the manifold score function satisfies* $\left\|\hat{\mathbf{s}}_{\|}(\mathbf{x}, t) - J_\psi(\mathbf{z}^*)\, \nabla \log p_t^{LD}(\mathbf{z}^*)\right\|_2 \leq \frac{1}{P}\left[\frac{1}{2}Ld + 4(1 + \|J_\psi(\mathbf{z}^*)\|)\, \frac{e^{-d/2}\left(\frac{d}{2}\right)^{d/2}}{r\,\Gamma\left(\frac{d}{2}\right)}\right]$. *where*

$$\nabla \log p_t^{LD}(\mathbf{z}^*) = \frac{\int_S (\mathbf{z} - \mathbf{z}^*)\, \phi_t(\mathbf{z}^* \mid \mathbf{z}) p_z(\mathbf{z})\, d\mathbf{z}}{s^2 \int_S \phi_t(\mathbf{z}^* \mid \mathbf{z}) p_z(\mathbf{z})\, d\mathbf{z}}, \tag{3.2}$$

From Lemma 3.2, the discrepancy between the true parallel component and the estimated score function arises from the curvature of the data manifold. For a hypersurface $\mathcal{M} \subset \mathbb{R}^D$, we have $L = 0$ and $r = +\infty$, giving $\hat{\mathbf{s}}_{\|}(\mathbf{x}, t) = J_\psi(\mathbf{z}^*)\nabla \log p_t^{LD}(\mathbf{z}^*)$. This recovers the linear case of Lemma 1 in Chen et al. (2023).

Lemma 3.3. *Using the notation of Theorem 3.1, take $\varepsilon > r$ such that $B_{\mathbf{z}^*}(\varepsilon) \subset S$. The score function (A.1) satisfies $\left\|\nabla \log p_t^{LD}(\mathbf{z}^*)\right\|_2 \le C(d)\left(\frac{s}{\varepsilon^2} + \frac{D_U}{p_z(\mathbf{z}^*)}\right)$, where $C(d)$ depends only on d.*

Lemma 3.3 shows that the on-support score Chen et al. (2023) can be bounded by a constant, a point left unresolved in the original work. Moreover, the lemma establishes that this bound is determined primarily by the data density and the intrinsic dimension of the manifold, with little dependence on other geometric features.

Theorem 3.1 reveals the underlying decomposition of the score function's parallel component, which has not been fully characterized in Chen et al. (2023); Liu et al. (2025a). As $t \to 0$, both s and $\|\mathbf{d}\|$ vanish, ensuring the right-hand side of (A.9) remains bounded. In contrast, the orthogonal term $\frac{1}{s^2}(\psi(\mathbf{z}^*) - \mathbf{x})$ diverges since $\psi(\mathbf{z}^*) - \mathbf{x} = \mathcal{O}(s)$. This behavior aligns with observations in De Bortoli (2022); Lu et al. (2023); Stanczuk et al. (2024), which further corroborate and refine the precision of our result.

Using Lemma 3.1, an asymptotic estimation of score decomposition can be given.

Corollary 3.1 (Asymptotic estimation)**.** *Let the conditions of Theorem 3.1 hold. For fixed L, r, as $t \to 0$ (equivalently $s \to 0$) and $L\|\psi(\mathbf{z}^*) - \mathbf{x}\|_2 \to 0$, the score function admits the asymptotic expansion*

$$\nabla \log p_t(\mathbf{x}) = \hat{\mathbf{s}}_{\|}(\mathbf{x}, t)\big[1 + \mathcal{O}\big(L\|\mathbf{d}\| + s^2\big)\big] - \frac{\psi(\mathbf{z}^*) - \mathbf{x}}{s^2} + \mathcal{O}(L\|\mathbf{d}\|\, s) + \frac{1}{s}\tilde{\mathcal{O}}\left(e^{-\frac{1}{2(L\|\mathbf{d}\|)^{2l}}}\right) + \tilde{\mathcal{O}}\left(e^{-\frac{r^2}{2s^2}}\right),$$

where $\tilde{\mathcal{O}}\left(e^{-\frac{1}{(L\|\mathbf{d}\|)^{2l}}}\right) = \mathcal{O}\left(\frac{1}{(L\|\mathbf{d}\|)^{l(d-1)-1}} e^{-\frac{1}{(L\|\mathbf{d}\|)^{2l}}}\right), \tilde{\mathcal{O}}\left(e^{-\frac{r^2}{2s^2}}\right) = \mathcal{O}\left(\frac{1}{s^{d+1}} e^{-\frac{r^2}{2s^2}}\right)$.

The above discussion considers manifolds isometric to an open subset of the ambient space. We extend the result to a more general setting by removing the isometry constraint, with details provided in Appendix A.2. In the multi-covering case, we assume only that $\mathcal{M}$ is a C^1 manifold, ensuring that the Jacobian of the coordinate map is uniformly bounded. This property allows us to handle coverings where the target point does not lie in the initial chart.

We conclude this section by analyzing perturbed data. Real-world datasets are often noisy and high-dimensional, yet their distributions typically concentrate near low-dimensional manifolds. We consider the case where $p(\mathbf{x})$ is not exactly supported on a manifold but is ϵ-close, in the Wasserstein-∞ sense, to a distribution $\overline{p}(\mathbf{x})$ supported on a compact manifold $\mathcal{M} \subset \mathbb{R}^D$. Formally, there exist $\overline{p}$, $\mathcal{M}$, and $\epsilon > 0$ such that $\operatorname{supp}(\overline{p}) \subset \mathcal{M}$ and

$$W_\infty(p, \overline{p}) := \inf_{\gamma \in \Pi(p, \overline{p})} \operatorname{ess\,sup}_{(\mathbf{x},\mathbf{y})\sim\gamma} \|\mathbf{x} - \mathbf{y}\|_2 \le \epsilon,$$

where $\Pi(p, \overline{p})$ is the set of couplings with marginals p and $\overline{p}$. For sufficiently small $W_\infty(p, \overline{p})$, the score function admits a decomposition. The following corollary strengthens Lemma 3.1 by quantifying this proximity. Proof is postponed in Appendix A.3.

Corollary 3.2. *Let $\overline{p}$ be supported on a compact d-dimensional manifold $\mathcal{M} \subset \mathbb{R}^D$, and assume $W_\infty(p, \overline{p}) \le \epsilon < \frac{1}{2} p_t(\mathbf{x})^2$. Suppose Assumptions 2.1 and 2.2 hold for $(\mathcal{M}, \overline{p})$. For $\overline{p}$, let*

$\mathbf{z}^*, \mathbf{d}, s^2, P, \phi_t(\cdot \mid \mathbf{z}), p_t^{LD}(\mathbf{z}^*), \hat{\mathbf{s}}_{\|}(\mathbf{x}, t)$ *be as defined in Lemmas 3.1 and 3.2. Fix constants* A, B *with* $A < \min\{\frac{1}{2}, \frac{1}{4(d-1)}, \frac{Lr}{4}\}$ *and* $P \geq A(B + \frac{2}{\sqrt{\pi}})d$*. Assume* $L\|\mathbf{d}\| < A$ *and* $s < \min\{\frac{r}{4}\sqrt{\frac{2}{d-1}}, B\}$*. Then*

$$\left\|\nabla \log p_t(\mathbf{x}) - \hat{\mathbf{s}}_{\|}(\mathbf{x}, t) + \tfrac{\mathbf{d}}{s^2}\right\|_2 \leq C(d)\big(1 + \tfrac{1}{r^2} + \tfrac{1}{r^3}\big)\big(\tfrac{1}{P} + \tfrac{1}{P^2}\big)\left[(s^2 + A)\,\|\hat{\mathbf{s}}_{\|}(\mathbf{x}, t)\|_2 + s + \tfrac{A^2}{s}\right] + \frac{2\sqrt{2\pi}\epsilon}{s\, p_t(\mathbf{x})^2}\big(p_t(\mathbf{x}) + p_{\max}\big),$$

where $C(d)$ *depends only on* d*.*

4 Small-time refinement via twin networks

Let $\{\mathbf{X}_i\}_{i=1}^N$ be points sampled from an unknown manifold. We employ a diffusion model to recover latent representations on the manifold. The diffusion process spans a total time T and is discretized into M steps, with samples generated at times $\frac{i}{M}T$ for $i = 1, \ldots, M$. Points at step j are denoted $\{\mathbf{X}_i^{(j)}\}_{i=1}^N$.

We propose an algorithm to address the singularity at $t = 0$, an inherent feature of classical diffusion models (Zhang et al. (2024)). From Theorem 3.1, the score function decomposes into an orthogonal term $\frac{1}{s^2}\big(\psi(\mathbf{z}^*) - \mathbf{x}\big)$ and a residual term $\nabla \log p_t(\mathbf{x}) - \frac{1}{s^2}\big(\psi(\mathbf{z}^*) - \mathbf{x}\big)$. Lemmas 3.1 and 3.2 show that the residual term has a bounded orthogonal component; we therefore refer to it as the parallel term.

For the classical VESDE, Tweedie's formula gives

$$\nabla \log p_t(\mathbf{X}_t) = \mathbb{E}\left(\frac{\mathbf{X}_0 - \mathbf{X}_t}{\sigma_t^2}\,\middle|\,\mathbf{X}_t\right) = -\frac{\epsilon_t}{\sigma_t}, \quad \mathbf{X}_t = \mathbf{X}_0 + \sigma_t \epsilon_t.$$

Conventional models train a single network M_{sin} via

$$\min_{s \in \mathcal{S}} \int_0^T \mathbb{E}_{\epsilon_t \sim \mathcal{N}(0,I)}\left[\|s(\mathbf{X}_t, t) + \epsilon_t\|_2^2\right] dt.$$

As $t \to 0$, the orthogonal term diverges while the parallel term remains bounded, impairing accurate learning of the latter. We refine training on the first m steps (time cutoff $\tilde{t}$) using two networks dedicated to these components. For training, we sample points from timestep $\frac{i}{M}T, i = 1, 2, \cdots, m$. Let $\mathbf{X}_\perp$ be the projection of $\mathbf{X}_t$ onto the data manifold. Then

$$\nabla \log p_t(\mathbf{X}_t) = \underbrace{\frac{\mathbf{X}_\perp - \mathbf{X}_t}{\sigma_t^2}}_{\text{orthogonal}} + \underbrace{\mathbb{E}\left(\frac{\mathbf{X}_0 - \mathbf{X}_\perp}{\sigma_t^2}\,\middle|\,\mathbf{X}_t\right)}_{\text{parallel}}.$$

The orthogonal network M_{ort} is targeted to

$$\min_{s \in \mathcal{S}} \int_0^{\tilde{t}} \mathbb{E}_{\mathbf{X}_t \sim p_t}\left[\left\|s(\mathbf{X}_t, t) - \tfrac{\mathbf{X}_\perp - \mathbf{X}_t}{\sigma_t}\right\|_2^2\right] dt,$$

and the parallel network M_{par} is targeted to

$$\min_{s \in \mathcal{S}} \int_0^{\tilde{t}} \mathbb{E}_{\mathbf{X}_t \sim p_t}\left[\left\|s(\mathbf{X}_t, t) - \tfrac{\mathbf{X}_0 - \mathbf{X}_\perp}{\sigma_t^2}\right\|_2^2\right] dt.$$

Denoting the learned functions by $s_\perp$ and $s_\|$, the score is estimated as $\nabla \log p_t(\mathbf{X}_t) = \frac{s_\perp(\mathbf{X}_t, t)}{\sigma_t} + s_\|(\mathbf{X}_t, t)$.

The task is to project a point $\mathbf{X}_t$ onto an unknown manifold $\mathcal{M}$ sampled by $\{\mathbf{X}_i\}_{i=1}^N$. We first identify the k nearest neighbors $\{\mathbf{X}_{i_j}\}_{j=1}^k$ of $\mathbf{X}_t$ using kNN. A local projector is then constructed via PCA, where the effective rank of the SVD of the distance matrix determines the embedding dimension. The projection yields the approximation $\mathbf{X}_{\text{sol}} \approx \mathbf{X}_\perp$, as summarized in Algorithm 1.

Algorithm 1 Twin-network training for small-t refinement

Input: Timestep cutoff $m \in \mathbb{N}_+$; training data $\{\mathbf{X}_i\}_{i=1}^N$; number of nearest neighbors $k \in \mathbb{N}_+$.
for $\underline{\text{step}_i \in \{1, 2, \ldots, m\}}$ **do**
  Sample $\mathbf{X}_t$ from $\mathbf{X}_i$ via the forward SDE;
  Estimate $\mathbf{X}_\perp$ using kNN+PCA (repeated for stability);
  Update $s_\perp$ by minimizing $\left\| s_\perp(\mathbf{X}_t, t) - \frac{\mathbf{X}_\perp - \mathbf{X}_t}{\sigma_t} \right\|_2^2$;
  Update $s_\parallel$ by minimizing $\left\| s_\parallel(\mathbf{X}_t, t) - \frac{\mathbf{X}_0 - \mathbf{X}_\perp}{\sigma_t^2} \right\|_2^2$;
end
Output: Orthogonal network $s_\perp$, parallel network $s_\parallel$.

This algorithm addresses the small-t regime, while the standard diffusion model is applied for larger t.

5 Numerical Experiments

We present numerical results based on the algorithm in Section 4 and conduct experiments to validate the model.

5.1 Toy model: 3D GMM embedded in $\mathbb{R}^{15}$

We constructed a three-cluster Gaussian mixture model (GMM) embedded in the first three coordinates of $\mathbb{R}^{15}$, with each cluster having variance $[0.05, 0.05, 0.05]$, and generated 10,000 samples. Diffusion time was set to $T = 1$ and discretized into 1,000 steps for both forward and reverse processes. The VESDE (Section 2.1) was parameterized with $\sigma_{\min} = 0.01$ and $\sigma_{\max} = 5.0$.

As a baseline, we trained a diffusion model using a four-layer multilayer perceptron with SiLU activations and a learning rate of 10^{-3}. Our approach instead learns two score terms during the first 100 noising steps, trained for 100 epochs in Table 1 and 200 epochs in Table 2. Baseline results were obtained under the same training conditions. In Algorithm 1, we fixed $k = 20$ and repeated the projection search for two cycles. All models were optimized with Adam.

Performance was evaluated using maximum mean discrepancy (MMD) between generated and ground-truth samples (Tables 1, 2) across varying baseline training epochs. Relative improvement was computed as Improvement = (baseline − ours)/baseline. The results show that, at well-trained baselines (e.g. 2000 epochs for both 100- and 200-epoch Twin settings), the Twin network achieves the largest gains, demonstrating that our algorithm consistently outperforms the baseline in the VESDE setting.

To address concerns that our algorithm may incur higher runtime than the baseline, we compare cases with comparable computational cost. Table 3 shows that selecting an appropriate number of epochs for DDPM at large t is critical for performance. As the number of DDPM epochs increases, our method consistently outperforms the traditional DDPM baseline.

	100	200	500	1000	2000
DDPM (baseline)	2.28e-2	1.86e-2	1.94e-2	1.52e-3	2.15e-4
Twin (ours)	2.19e-2	1.90e-2	1.72e-2	1.20e-3	6.00e-5
Improvement (%)	3.9%	-2.2%	11.3%	21.1%	72.1%

Table 1: MMD results for different methods (100 epochs Twin).

	200	500	1000	2000	4000
DDPM (baseline)	6.07e-3	4.96e-3	4.94e-4	2.15e-4	2.84e-3
Twin (ours)	4.16e-3	4.24e-3	3.29e-4	5.20e-5	1.02e-3
Improvement (%)	31.5%	14.5%	33.4%	75.8%	64.1%

Table 2: MMD results for different methods (200 epochs Twin).

6 Discussion

In this work, we present a refined decomposition of the score function in diffusion models. In particular, we analyze in detail the structure of the non-singular term in the score expression, thereby providing a clearer understanding of its composition. Building on this theoretical insight, we introduce a two-network approach to learn the score function in regimes where singularities arise at small diffusion times. Our method surpasses the baseline and achieves more accurate score estimation.

This contribution opens several directions for future work. First, extending our analysis from drift-free Brownian motion to processes with drift (Lu et al. (2023)) may advance understanding of the Ornstein–Uhlenbeck process. Second, projecting a point onto an unknown manifold remains unresolved; current optimization-based approaches are computationally expensive and prone to redundant calculations (e.g., on identical points). More efficient solutions are needed. Finally, selecting the truncated time for twin-network training poses an open challenge: small times risk overfitting due to limited data, while large times conflict with our theoretical framework and degrade performance.

Acknowledgment

The author thanks Kaizheng Wang for valuable advice and generous assistance.

References

Anderson, B. D. O. (1982). Reverse-time diffusion equation models. *Stochastic Processes and their Applications* **12** 313–326.

Azangulov, I., Deligiannidis, G. and Rousseau, J. (2024). Convergence of diffusion models under the manifold hypothesis in high-dimensions. *arXiv preprint arXiv:2409.18804* . URL https://arxiv.org/abs/2409.18804

Azeglio, S. and Di Bernardo, A. (2025). What's inside your diffusion model? a score–based riemannian metric to explore the data manifold. In *arXiv preprint arXiv:2505.11128*. V2, 19 May 2025.

Training scheme	**MMD**
1000 epochs DDPM	2.50e-3
200 epochs DDPM + 200 epochs Twin	1.13e-2
1200 epochs DDPM	1.95e-3
400 epochs DDPM + 200 epochs Twin	2.33e-3
1600 epochs DDPM	3.41e-3
800 epochs DDPM + 200 epochs Twin	1.08e-3
2000 epochs DDPM	8.28e-4
1200 epochs DDPM + 200 epochs Twin	6.70e-4

Table 3: **MMD performance.** Results compare DDPM trained for different numbers of epochs with the proposed Twin-augmented approach at matched computational budgets. Appropriate selection of DDPM epochs at large t substantially improves performance.

Block, A., Mroueh, Y. and Rakhlin, A. (2022). Generative modeling with denoising auto-encoders and langevin sampling.

Chen, C., Qian, R., Hu, W., Fu, T.-J., Li, L., Zhang, B., Schwing, A., Liu, W. and Yang, Y. (2025). Dit-air: Revisiting the efficiency of diffusion model architecture design in text to image generation. *arXiv preprint arXiv:2503.10618* .

Chen, M., Huang, K., Zhao, T. and Wang, M. (2023). Score approximation, estimation and distribution recovery of diffusion models on low-dimensional data. In *Proceedings of the 40th International Conference on Machine Learning*. PMLR.
URL https://proceedings.mlr.press/v202/chen23o.html

Chung, H., Sim, B., Ryu, D. and Ye, J. C. (2022). Improving diffusion models for inverse problems using manifold constraints. In *Advances in Neural Information Processing Systems*, vol. 35.

Coifman, R. R. and Lafon, S. (2006). Diffusion maps. *Applied and Computational Harmonic Analysis* **21** 5–30.

De Bortoli, V. (2022). Convergence of denoising diffusion models under the manifold hypothesis. *arXiv preprint arXiv:2208.05314* .
URL https://arxiv.org/abs/2208.05314

Dhariwal, P. and Nichol, A. (2021). Diffusion models beat gans on image synthesis. *arXiv preprint arXiv:2105.05233* .

Ho, J., Chan, W., Saharia, C., Whang, J., Gao, R., Gritsenko, A., Kingma, D. P., Poole, B., Norouzi, M., Fleet, D. J. and Salimans, T. (2022). Imagen video: High definition video generation with diffusion models. *arXiv preprint arXiv:2210.02303* .

Ho, J., Jain, A. and Abbeel, P. (2020). Denoising diffusion probabilistic models. In *Advances in Neural Information Processing Systems*, vol. 33.

Huang, C.-W., Aghajohari, M., Bose, A. J., Panangaden, P. and Courville, A. (2022). Riemannian diffusion models. In *Advances in Neural Information Processing Systems (NeurIPS)*. NeurIPS 2022.

Jo, J. and Hwang, S. J. (2024). Generative modeling on manifolds through mixture of riemannian diffusion processes. In *Proceedings of the International Conference on Machine Learning (ICML)*.

Kingma, D. P. and Welling, M. (2013). Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114* .

Liu, P., Tateo, D., Ammar, H. B. and Peters, J. (2022). Robot reinforcement learning on the constraint manifold. In *Proceedings of the 5th Conference on Robot Learning*, vol. 164 of *Proceedings of Machine Learning Research*. PMLR.

Liu, Z., Zhang, W. and Li, T. (2025a). Improving the euclidean diffusion generation of manifold data by mitigating score function singularity. In *arXiv preprint arXiv:2505.09922*.

Liu, Z., Zhang, W., SchÃŒtte, C. and Li, T. (2025b). Riemannian denoising diffusion probabilistic models. In *arXiv preprint arXiv:2505.04338*.

Lou, A., Xu, M. and Ermon, S. (2023). Scaling riemannian diffusion models. In *Neural Information Processing Systems*.

Lu, Y., Wang, Z. and Bal, G. (2023). Mathematical analysis of singularities in the diffusion model under the submanifold assumption. *arXiv preprint arXiv:2301.07882* .

Ma, N., Goldstein, M., Albergo, M. S., Boffi, N. M., Vanden-Eijnden, E. and Xie, S. (2024). Sit: Exploring flow and diffusion-based generative models with scalable interpolant transformers. *arXiv preprint arXiv:2401.08740* .

Meilă, M. and Zhang, H. (2024). Manifold learning: What, how, and why. *Annual Review of Statistics and Its Application* **11** 393–417.

Melnik, A., Ljubljanac, M., Lu, C., Yan, Q., Ren, W. and Ritter, H. (2024). Video diffusion models: A survey. *arXiv preprint arXiv:2405.03150* .

Peebles, W. and Xie, S. (2022). Scalable diffusion models with transformers. In *International Conference on Computer Vision (ICCV)*.

Ramesh, A., Dhariwal, P., Nichol, A., Chu, C. and Chen, M. (2022). Hierarchical text-conditional image generation with clip latents (dallÂ·e 2). *arXiv preprint arXiv:2204.06125* .

Rombach, R., Blattmann, A., Lorenz, D., Esser, P. and Ommer, B. (2022a). High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.

Rombach, R., Blattmann, A., Lorenz, D., Esser, P. and Ommer, B. (2022b). High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.

Saito, S. and Matsubara, T. (2025). Image interpolation with score-based riemannian metrics of diffusion models. In *ICLR 2025 Workshop on Deep Generative Model in Machine Learning: Theory, Principle and Efficacy*. Workshop, Singapore.

Song, Y. and Ermon, S. (2019). Generative modeling by estimating gradients of the data distribution. In *Advances in Neural Information Processing Systems (NeurIPS)*.

Song, Y., Sohl-Dickstein, J., Kingma, D. P., Kumar, A., Ermon, S. and Poole, B. (2021). Score-based generative modeling through stochastic differential equations. In *International Conference on Learning Representations (ICLR)*.

Stanczuk, J. P., Batzolis, G., Deveney, T. and Schönlieb, C.-B. (2024). Diffusion models encode the intrinsic dimension of data manifolds. *arXiv preprint arXiv:2406.06877* .
URL https://openreview.net/forum?id=a0XiA6v256

Tang, R. and Yang, Y. (2024). Adaptivity of diffusion models to manifold structures. In *Proceedings of Machine Learning Research*, vol. 238.
URL https://arxiv.org/abs/2409.12345

Yang, Z., Feng, R., Zhang, H., Shen, Y., Zhu, K., Huang, L., Zhang, Y., Liu, Y., Zhao, D., Zhou, J. and Cheng, F. (2024). Lipschitz singularities in diffusion models. In *Proceedings of the Twelfth International Conference on Learning Representations (ICLR)*.

Zhang, P., Yin, H., Li, C. and Xie, X. (2024). Tackling the singularities at the endpoints of time intervals in diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.

Zhu, R., Pan, Y., Li, Y., Yao, T., Sun, Z., Mei, T. and Chen, C. W. (2024). Sd-dit: Unleashing the power of self-supervised discrimination in diffusion transformer. *arXiv preprint arXiv:2403.17004* .

A Proofs in Section 3

A.1 Proofs of isometric manifold

First, we put forward a lemma, which will be used all through the three lemmas. It clarifies that $p_t^{LD}(\mathbf{z}^*)$ is lower bounded by a constant if t is small enough.

Lemma A.1. *The latent disturbed distribution is defined as*

$$p_t^{LD}(\mathbf{z}^*) = \int_S \phi_t(\mathbf{z}^* \mid \mathbf{z}) p_z(\mathbf{z})\, d\mathbf{z}, \tag{A.1}$$

if $s < \frac{p_z(\mathbf{z}^*)}{4\sqrt{2\pi}\left(\frac{p_z(\mathbf{z}^*)}{\varepsilon}+D_U\right)}$, *we have*

$$p_t^{LD}(\mathbf{z}^*) > \frac{1}{2} p_z(\mathbf{z}^*).$$

Proof of Lemma A.1. By identical used in simplification of score function for $p_t^{LD}(\mathbf{z}^*)$, we simplify the expression

$$p_t^{LD}(\mathbf{z}^*) = \frac{1}{\pi^{d/2}} \int_{\frac{(S-\mathbf{z}^*)}{\sqrt{2}s}} e^{-\|\mathbf{v}\|^2} p_z(\mathbf{z}^* + \sqrt{2}s\mathbf{v})\, d\mathbf{v},$$

when p_z is continuously differentiable and ∇p_z can be uniformly upper bounded by D_U, since $s < \frac{r}{4\sqrt{\frac{d-1}{2}}} < \frac{\varepsilon}{4\sqrt{\frac{d-1}{2}}}$,

$$\begin{aligned}
p_t^{LD}(\mathbf{z}^*) &= \frac{1}{\pi^{d/2}} \int_{\frac{(S-\mathbf{z}^*)}{\sqrt{2}s}} e^{-\|\mathbf{v}\|^2} \left(p_z(\mathbf{z}^*) + \sqrt{2}s\mathbf{v}^\top \nabla p_z(\mathbf{z}^* + t_{\mathbf{v}}\mathbf{v})\right) d\mathbf{v} \\
&\geq \frac{1}{\pi^{d/2}} \left(p_z(\mathbf{z}^*) \int_{\frac{(S-\mathbf{z}^*)}{\sqrt{2}s}} e^{-\|\mathbf{v}\|^2}\, d\mathbf{v} - \sqrt{2}sD_U \int_{\frac{(S-\mathbf{z}^*)}{\sqrt{2}s}} \|\mathbf{v}\| e^{-\|\mathbf{v}\|^2}\, d\mathbf{v} \right) \\
&\geq \frac{1}{\pi^{d/2}} \left(p_z(\mathbf{z}^*)\mathrm{Area}(S^{d-1}) \int_0^{\frac{\varepsilon}{\sqrt{2}s}} r^{d-1} e^{-r^2}\, dr - \sqrt{2}sD_U \mathrm{Area}(S^{d-1}) \int_0^\infty r^d e^{-r^2}\, dr \right) \\
&\geq \frac{1}{\pi^{d/2}} p_z(\mathbf{z}^*)\mathrm{Area}(S^{d-1}) \int_0^{\frac{\varepsilon}{\sqrt{2}s}} r^{d-1} e^{-r^2}\, dr - \sqrt{2}sD_U \frac{\Gamma\left(\frac{d+1}{2}\right)}{\Gamma\left(\frac{d}{2}\right)} \\
&= \frac{p_z(\mathbf{z}^*)}{\Gamma\left(\frac{d}{2}\right)} \int_0^{\frac{\varepsilon^2}{2s^2}} x^{\frac{d}{2}-1} e^{-x} dx - \sqrt{2}sD_U \frac{\Gamma\left(\frac{d+1}{2}\right)}{\Gamma\left(\frac{d}{2}\right)} \\
&\geq \frac{p_z(\mathbf{z}^*)}{\Gamma\left(\frac{d}{2}\right)} \left(\Gamma\left(\frac{d}{2}\right) - \frac{\left(\frac{\varepsilon^2}{2s^2}\right)^{\frac{d}{2}} e^{-\frac{\varepsilon^2}{2s^2}}}{\frac{\varepsilon^2}{2s^2} - \frac{d}{2} + 1} \right) - \sqrt{2}sD_U \frac{\Gamma\left(\frac{d+1}{2}\right)}{\Gamma\left(\frac{d}{2}\right)} \\
&\geq \frac{p_z(\mathbf{z}^*)}{\Gamma\left(\frac{d}{2}\right)} \left(\Gamma\left(\frac{d}{2}\right) - 2\left(\frac{\varepsilon^2}{2s^2}\right)^{\frac{d}{2}-1} e^{-\frac{\varepsilon^2}{2s^2}} \right) - \sqrt{2}sD_U \frac{\Gamma\left(\frac{d+1}{2}\right)}{\Gamma\left(\frac{d}{2}\right)} \\
&\geq p_z(\mathbf{z}^*) - 2\sqrt{2\pi}\frac{s}{\varepsilon} p_z(\mathbf{z}^*) - \sqrt{2\pi}sD_U \\
&\geq p_z(\mathbf{z}^*) - 2\sqrt{2\pi}\left(\frac{p_z(\mathbf{z}^*)}{\varepsilon} + D_U\right)s
\end{aligned}$$

$$\geq \frac{1}{2} p_z(\mathbf{z}^*)$$

The last but one inequality is because of the analysis of minima of function and $\frac{\Gamma\left(\frac{d+1}{2}\right)}{\Gamma\left(\frac{d}{2}\right)} \leq \sqrt{\pi}$:

$$\left(\frac{\varepsilon^2}{2s^2}\right)^{\frac{d}{2}-1} e^{-\frac{\varepsilon^2}{2s^2}} \leq \frac{\sqrt{2}s}{\varepsilon} \frac{\left(\frac{d-1}{2}\right)^{\frac{d-1}{2}} e^{-\frac{d-1}{2}}}{\Gamma\left(\frac{d}{2}\right)} \leq \frac{\sqrt{2}s}{\varepsilon} \frac{\Gamma\left(\frac{d+1}{2}, \frac{d-1}{2}\right)}{\Gamma\left(\frac{d}{2}\right)} \leq \sqrt{2\pi} \frac{s}{\varepsilon}.$$

□

Proof of Theorem 3.1. Due to the noising process and Assumption 2.1, the probability function of $\mathbf{x}$ can be expressed by latent variable $\mathbf{z}$:

$$p_t(\mathbf{x}) = \int \phi_t(\mathbf{x}|\psi(\mathbf{z})) p_z(\mathbf{z})\, d\mathbf{z},$$

where

$$\phi_t(\mathbf{x}|\psi(\mathbf{z})) = (2\pi)^{-D/2} \left(\int_0^t \sigma^2(\tau) d\tau\right)^{-D/2} \exp\left(-\frac{1}{2\int_0^t \sigma^2(\tau)d\tau} \|\psi(\mathbf{z}) - \mathbf{x}\|_2^2\right).$$

Then the score function can be written as

$$\nabla \log p_t(\mathbf{x}) = \frac{\nabla \int \phi_t(\mathbf{x}|\psi(\mathbf{z})) p_z(\mathbf{z})\, d\mathbf{z}}{\int \phi_t(\mathbf{x}|\psi(\mathbf{z})) p_z(\mathbf{z})\, d\mathbf{z}} = \frac{\int \nabla \phi_t(\mathbf{x}|\psi(\mathbf{z})) p_z(\mathbf{z})\, d\mathbf{z}}{\int \phi_t(\mathbf{x}|\psi(\mathbf{z})) p_z(\mathbf{z})\, d\mathbf{z}}, \tag{A.2}$$

where the last equality holds since $\phi_t(\mathbf{x}|\psi(\mathbf{z}))$ is continuously differentiable in $\mathbf{x}$. Let $\mathbf{z}^*$ to be the nearest point to $\mathbf{x}$ on the manifold. Substituting $\phi_t(\mathbf{x}|\psi(\mathbf{z}))$ into (A.2) gives rise to

$$\begin{aligned}
&\nabla \log p_t(\mathbf{x}) \\
&= \frac{(2\pi)^{-D/2}\left(\int_0^t \sigma^2(\tau)d\tau\right)^{-D/2} \int \frac{1}{\int_0^t \sigma^2(\tau)d\tau} (\psi(\mathbf{z}) - \mathbf{x}) \exp\left(-\frac{1}{2\int_0^t \sigma^2(\tau)d\tau}\|\psi(\mathbf{z}) - \mathbf{x}\|_2^2\right) p_z(\mathbf{z})\, d\mathbf{z}}{\int \phi_t(\mathbf{x}|\psi(\mathbf{z})) p_z(\mathbf{z})\, d\mathbf{z}} \\
&= \frac{(2\pi)^{-D/2}\left(\int_0^t \sigma^2(\tau)d\tau\right)^{-D/2} \int \frac{1}{\int_0^t \sigma^2(\tau)d\tau} (\psi(\mathbf{z}) - \psi(\mathbf{z}^*)) \exp\left(-\frac{1}{2\int_0^t \sigma^2(\tau)d\tau}\|\psi(\mathbf{z}) - \mathbf{x}\|_2^2\right) p_z(\mathbf{z})\, d\mathbf{z}}{\int \phi_t(\mathbf{x}|\psi(\mathbf{z})) p_z(\mathbf{z})\, d\mathbf{z}} \\
&\quad + \frac{(2\pi)^{-D/2}\left(\int_0^t \sigma^2(\tau)d\tau\right)^{-D/2} \int \frac{1}{\int_0^t \sigma^2(\tau)d\tau} (\psi(\mathbf{z}^*) - \mathbf{x}) \cdot \exp\left(-\frac{1}{2\int_0^t \sigma^2(\tau)d\tau}\|\psi(\mathbf{z}) - \mathbf{x}\|_2^2\right) p_z(\mathbf{z})\, d\mathbf{z}}{\int \phi_t(\mathbf{x}|\psi(\mathbf{z})) p_z(\mathbf{z})\, d\mathbf{z}} \\
&= \underbrace{\frac{1}{\int \phi_t(\mathbf{x}|\psi(\mathbf{z})) p_z(\mathbf{z})\, d\mathbf{z}} \int \frac{1}{\int_0^t \sigma^2(\tau)d\tau} (\psi(\mathbf{z}) - \psi(\mathbf{z}^*)) \phi_t(\mathbf{x}|\psi(\mathbf{z})) p_z(\mathbf{z})\, d\mathbf{z}}_{\mathbf{s}_\parallel} + \underbrace{\frac{1}{\int_0^t \sigma^2(\tau)d\tau} (\psi(\mathbf{z}^*) - \mathbf{x})}_{\mathbf{s}_\perp}.
\end{aligned}$$

We can further simplify $\mathbf{s}_\parallel$. We decompose $\phi_t(\mathbf{x} \mid \psi(\mathbf{z}))$ as

$$\begin{aligned}
&\phi_t(\mathbf{x} \mid \psi(\mathbf{z})) \\
&= (2\pi)^{-D/2} \left(\int_0^t \sigma^2(\tau) d\tau\right)^{-D/2} \exp\left(-\frac{1}{2\int_0^t \sigma^2(\tau)d\tau} \|\psi(\mathbf{z}) - \psi(\mathbf{z}^*) + \psi(\mathbf{z}^*) - \mathbf{x}\|_2^2\right)
\end{aligned}$$

$$
\begin{aligned}
&= (2\pi)^{-D/2} \left(\int_0^t \sigma^2(\tau)d\tau\right)^{-D/2} \exp\left(-\frac{1}{2\int_0^t \sigma^2(\tau)d\tau} \left(\|\psi(\mathbf{z}) - \psi(\mathbf{z}^*)\|_2^2 + \|\psi(\mathbf{z}^*) - \mathbf{x}\|_2^2\right.\right. \\
&\quad \left.\left. +2\left(\psi(\mathbf{z}) - \psi(\mathbf{z}^*)\right)^\top \left(\psi(\mathbf{z}^*) - \mathbf{x}\right)\right)\right) \\
&= (2\pi)^{-d/2} \left(\int_0^t \sigma^2(\tau)d\tau\right)^{-d/2} \exp\left(-\frac{1}{2\int_0^t \sigma^2(\tau)d\tau} \left(\|\psi(\mathbf{z}) - \psi(\mathbf{z}^*)\|_2^2 + 2\left(\psi(\mathbf{z}) - \psi(\mathbf{z}^*)\right)^\top \left(\psi(\mathbf{z}^*) - \mathbf{x}\right)\right)\right) \\
&\quad \times (2\pi)^{-(D-d)/2} \left(\int_0^t \sigma^2(\tau)d\tau\right)^{-(D-d)/2} \exp\left(-\frac{1}{2\int_0^t \sigma^2(\tau)d\tau} \|\psi(\mathbf{z}^*) - \mathbf{x}\|_2^2\right).
\end{aligned}
$$

We denote

$$
\phi_t(\mathbf{z}^* \mid \mathbf{z}) = (2\pi)^{-d/2} \left(\int_0^t \sigma^2(\tau)d\tau\right)^{-d/2} \exp\left(-\frac{1}{2\int_0^t \sigma^2(\tau)d\tau} \|\mathbf{z} - \mathbf{z}^*\|_2^2\right)
$$

and

$$
\phi_t(\mathbf{x}) = (2\pi)^{-(D-d)/2} \left(\int_0^t \sigma^2(\tau)d\tau\right)^{-(D-d)/2} \exp\left(-\frac{1}{2\int_0^t \sigma^2(\tau)d\tau} \|\psi(\mathbf{z}^*) - \mathbf{x}\|_2^2\right)
$$

being both Gaussian densities. Substituting

$$
\phi_t(\mathbf{x} \mid \psi(\mathbf{z})) = \phi_t(\mathbf{z}^* \mid \mathbf{z}) \cdot \phi_t(\mathbf{x}) \cdot \exp\left(-\frac{1}{\int_0^t \sigma^2(\tau)d\tau} \left(\psi(\mathbf{z}) - \psi(\mathbf{z}^*)\right)^\top \left(\psi(\mathbf{z}^*) - \mathbf{x}\right)\right)
$$

into $\mathbf{s}_\parallel$, we obtain:

$$
\mathbf{s}_\parallel(\mathbf{x}, t) = \frac{\int \frac{1}{\int_0^t \sigma^2(\tau)d\tau} \left(\psi(\mathbf{z}) - \psi(\mathbf{z}^*)\right) \phi_t(\mathbf{z}^* \mid \mathbf{z}) \exp\left(-\frac{1}{\int_0^t \sigma^2(\tau)d\tau} \left(\psi(\mathbf{z}) - \psi(\mathbf{z}^*)\right)^\top \left(\psi(\mathbf{z}^*) - \mathbf{x}\right)\right) p_z(\mathbf{z})\, d\mathbf{z}}{\int \phi_t(\mathbf{z}^* \mid \mathbf{z}) \exp\left(-\frac{1}{\int_0^t \sigma^2(\tau)d\tau} \left(\psi(\mathbf{z}) - \psi(\mathbf{z}^*)\right)^\top \left(\psi(\mathbf{z}^*) - \mathbf{x}\right)\right) p_z(\mathbf{z})\, d\mathbf{z}}
$$

We aim to compare the parallel term $\mathbf{s}_\parallel(\mathbf{x}, t)$ with

$$
\hat{s}_\parallel(\mathbf{x}, t) = \frac{\int \frac{1}{\int_0^t \sigma^2(\tau)d\tau} \left(\psi(\mathbf{z}) - \psi(\mathbf{z}^*)\right) \phi_t(\mathbf{z}^* \mid \mathbf{z}) p_z(\mathbf{z})\, d\mathbf{z}}{\int \phi_t(\mathbf{z}^* \mid \mathbf{z}) p_z(\mathbf{z})\, d\mathbf{z}}.
$$

Take $\mathbf{u} := \psi(\mathbf{z}) - \psi(\mathbf{z}^*)$, $\Delta\mathbf{z} = \mathbf{z} - \mathbf{z}^*$, $\mathbf{d} = \psi(\mathbf{z}^*) - \mathbf{x}$ and $\theta = -\frac{\mathbf{d}}{s^2}$. The above expressions can be transformed into

$$
\begin{aligned}
\mathbf{s}_\parallel(\mathbf{x}, t) &= \frac{1}{s^2} \frac{\int_S \mathbf{u}\, e^{\theta^\top \mathbf{u}}\, \phi_t(\mathbf{z}^* \mid \mathbf{z})\, p_z(\mathbf{z})\, d\mathbf{z}}{\int_S e^{\theta^\top \mathbf{u}}\, \phi_t(\mathbf{z}^* \mid \mathbf{z})\, p_z(\mathbf{z})\, d\mathbf{z}} =: \frac{I_{\text{num}}}{s^2 I_{\text{den}}}, \\
\hat{s}_\parallel(\mathbf{x}, t) &= \frac{1}{s^2} \frac{\int_S \mathbf{u}\, \phi_t(\mathbf{z}^* \mid \mathbf{z})\, p_z(\mathbf{z})\, d\mathbf{z}}{\int_S \phi_t(\mathbf{z}^* \mid \mathbf{z})\, p_z(\mathbf{z})\, d\mathbf{z}} =: \frac{s_{\text{num}}}{s^2 s_{\text{den}}}.
\end{aligned}
$$

We divide the integral into three parts:

$$
\begin{aligned}
I_{\text{num}} &= \int_S \mathbf{u}\, e^{\theta^\top \mathbf{u}}\, \phi_t p_z\, d\mathbf{z} = I_{\text{num}}^{\text{loc}} + I_{\text{num}}^{\text{mid}} + I_{\text{num}}^{\text{far}}, \\
I_{\text{den}} &= \int_S e^{\theta^\top \mathbf{u}}\, \phi_t p_z\, d\mathbf{z} = I_{\text{den}}^{\text{loc}} + I_{\text{den}}^{\text{mid}} + I_{\text{den}}^{\text{far}},
\end{aligned}
$$

where "loc" represents the region of $S \cap \{\|\mathbf{z}-\mathbf{z}^*\| \le M\}$, "mid" represents the region of $S \cap \{M \le \|\mathbf{z}-\mathbf{z}^*\| \le r\}$ and "far" represents their complement. Take $M = \frac{s}{(L\|\mathbf{d}\|)^{1/2}}$. If $M > r$, we define $I_{\text{num}}^{\text{mid}} = I_{\text{den}}^{\text{mid}} = 0$.

Similarly, divide $\hat{s}_{\|}(\mathbf{x}, t) = \frac{s_{\text{num}}}{s^2 s_{\text{den}}}$ into two parts:

$$
\begin{aligned}
s_{\text{num}} &= \int_S \mathbf{u}\, \phi_t p_z \, d\mathbf{z} = s_{\text{num}}^{\text{loc}} + s_{\text{num}}^{\text{far}}, \\
s_{\text{den}} &= \int_S \phi_t p_z \, d\mathbf{z} = s_{\text{den}}^{\text{loc}} + s_{\text{den}}^{\text{far}},
\end{aligned}
$$

where "loc" represents the region of $S \cap \{\|\mathbf{z}-\mathbf{z}^*\| \le M\}$, and "far" represents its complement. The following $C_i(d)$ all refer to constants depending on d.

① **The medium case.** When $M \le \|\mathbf{z}-\mathbf{z}^*\| \le r$, due to Lemma B.2, we have

$$
\|\mathbf{d}^\top \mathbf{u}\| = \|\left(\psi(\mathbf{z}^*) - \mathbf{x}\right)^\top \left(\psi(\mathbf{z}) - \psi(\mathbf{z}^*)\right)\| \le \frac{L}{2}\|\psi(\mathbf{z}^*) - \mathbf{x}\|\|\mathbf{z}-\mathbf{z}^*\|^2. \tag{A.3}
$$

thus the Gaussian kernel satisfies

$$
e^{\theta^\top \mathbf{u}}\, \phi_t(\mathbf{z}^* \mid \mathbf{z}) \le \frac{1}{(2\pi s^2)^{\frac{d}{2}}} e^{-\frac{\|\Delta \mathbf{z}\|^2}{2s^2} + \frac{L\|\mathbf{d}\|\|\Delta \mathbf{z}\|^2}{2s^2}}
$$

we can estimate the term $M \le \|\mathbf{z}-\mathbf{z}^*\| \le r$:

$$
\begin{aligned}
&\int_{S\cap\{M\le\|\mathbf{z}-\mathbf{z}^*\|\le r\}} e^{\theta^\top \mathbf{u}}\, \phi_t p_z \, d\mathbf{z} \\
\le\; & \frac{p_{max}}{(2\pi s^2)^{\frac{d}{2}}} \int_{\{M\le\|\mathbf{z}-\mathbf{z}^*\|\le r\}} \exp\left(\tfrac{(L\|\mathbf{d}\|-1)\|\Delta \mathbf{z}\|^2}{2s^2}\right) d\mathbf{z} \\
=\; & \frac{p_{max} Area(S^{d-1})}{(2\pi s^2)^{\frac{d}{2}}} \int_M^r w^{d-1} \exp\left(\tfrac{(L\|\mathbf{d}\|-1)w^2}{2s^2}\right) dw \quad (S^{d-1} \text{ denotes the unit sphere in } \mathbb{R}^d) \\
=\; & \frac{p_{max}}{\Gamma\left(\frac{d}{2}\right)(1-L\|\mathbf{d}\|)^{d/2}} \int_{\alpha M^2}^{\alpha r^2} u^{\frac{d}{2}-1} e^{-u} du \quad \left(\alpha = \frac{1-L\|\mathbf{d}\|}{2s^2}\right)
\end{aligned}
$$

As $L\|\mathbf{d}\| < \min\{\frac{1}{2}, \frac{1}{4(d-2)}\}$, this yields that

$$
\alpha M^2 = \frac{1-L\|\mathbf{d}\|}{2L\|\mathbf{d}\|} > \frac{1}{4L\|\mathbf{d}\|} > d-2,
$$

then the middle term becomes

$$
\begin{aligned}
\frac{p_{max}}{\Gamma\left(\frac{d}{2}\right)(1-L\|\mathbf{d}\|)^{d/2}} \int_{\alpha M^2}^{\alpha r^2} u^{\frac{d}{2}-1} e^{-u} du &\le \frac{p_{max}}{\Gamma\left(\frac{d}{2}\right)(1-L\|\mathbf{d}\|)^{d/2}} \int_{\alpha M^2}^{+\infty} u^{\frac{d}{2}-1} e^{-u} du \\
&\le \frac{p_{max}}{\Gamma\left(\frac{d}{2}\right)(1-L\|\mathbf{d}\|)^{d/2}} \frac{\left(\frac{1-L\|\mathbf{d}\|}{2L\|\mathbf{d}\|}\right)^{\frac{d}{2}} e^{-\frac{1-L\|\mathbf{d}\|}{2L\|\mathbf{d}\|}}}{\frac{1-L\|\mathbf{d}\|}{2L\|\mathbf{d}\|} - \frac{d}{2} + 1} \\
&\le \frac{p_{max}}{2^{\frac{d}{2}-2}\Gamma\left(\frac{d}{2}\right)} \frac{e^{-\frac{1-L\|\mathbf{d}\|}{2L\|\mathbf{d}\|}}}{(1-L\|\mathbf{d}\|)(L\|\mathbf{d}\|)^{\frac{d-2}{2}}}
\end{aligned}
$$

$$\le p_{max}\, C_2(d) \cdot A =: b_2$$

where the last inequality comes from $L\|\mathbf{d}\| \le A < \frac{1}{2}$ and the exponential decay of the function of $L\|\mathbf{d}\|$.

Similarly, by taking $L\|\mathbf{d}\| < \min\{\frac{1}{2}, \frac{1}{4(d-2)}\}$, we have

$$\|I_{\text{num}}^{\text{mid}}\|_2 \le p_{max}\, C_1(d) \cdot sA^2 =: b_1,$$

② **The faraway case.** When $\|\mathbf{z} - \mathbf{z}^*\| \ge r$, the Gaussian kernel satisfies

$$e^{\theta^\top \mathbf{u}} \phi_t(\mathbf{z}^* \mid \mathbf{z}) \le \frac{1}{(2\pi s^2)^{\frac{d}{2}}} e^{-\frac{\|\Delta \mathbf{z}\|^2}{2s^2} + \frac{\|\mathbf{d}\|\|\Delta \mathbf{z}\|}{s^2}}$$

we can estimate the term $\|\mathbf{z} - \mathbf{z}^*\| \ge r$:

$$\begin{aligned}
&\int_{S \cap \{\|\mathbf{z}-\mathbf{z}^*\| \ge r\}} e^{\theta^\top \mathbf{u}} \phi_t p_z \, d\mathbf{z} \\
&\le \frac{e^{\frac{\|\mathbf{d}\|^2}{2s^2}} p_{max}}{(2\pi s^2)^{\frac{d}{2}}} \int_{\{\|\Delta \mathbf{z}\| \ge r\}} \exp\left(-\tfrac{(\|\Delta \mathbf{z}\| - \|\mathbf{d}\|)^2}{2s^2}\right) d\mathbf{z} \\
&\le e^{\frac{\|\mathbf{d}\|^2}{2s^2}} p_{max} \frac{Area(S^{d-1})}{(2\pi s^2)^{\frac{d}{2}}} \int_r^\infty w^{d-1} \exp\left(-\tfrac{(w - \|\mathbf{d}\|)^2}{2s^2}\right) dw \quad (S^{d-1} \text{ denotes the unit sphere in } \mathbb{R}^d) \\
&= e^{\frac{\|\mathbf{d}\|^2}{2s^2}} p_{max} \frac{Area(S^{d-1})}{(2\pi s^2)^{\frac{d}{2}}} \sum_{k=0}^{d-1} \binom{d-1}{k} \|\mathbf{d}\|^{d-1-k} \int_{r-\|\mathbf{d}\|}^\infty w^k e^{-\frac{w^2}{2s^2}} dw \\
&= \frac{e^{\frac{\|\mathbf{d}\|^2}{2s^2}} p_{max}}{\Gamma\left(\frac{d}{2}\right)} \sum_{k=0}^{d-1} \binom{d-1}{k} \left(\frac{\|\mathbf{d}\|^2}{2s^2}\right)^{\frac{d-1-k}{2}} \int_{\frac{(r-\|\mathbf{d}\|)^2}{2s^2}}^\infty v^{\frac{k-1}{2}} e^{-v} dv
\end{aligned}$$

When $r > 3\|\mathbf{d}\|$ and $\frac{(r-\|\mathbf{d}\|)^2}{2s^2} > 2\left(\frac{d}{2} - 1\right)$, i.e., $s < \frac{r}{4\sqrt{\frac{d}{2}-1}}$, we obtain

$$\begin{aligned}
I_{\text{den}}^{\text{far}} &\le \frac{e^{\frac{\|\mathbf{d}\|^2}{2s^2}} p_{max}}{\Gamma\left(\frac{d}{2}\right)} \sum_{k=0}^{d-1} \binom{d-1}{k} \left(\frac{\|\mathbf{d}\|^2}{2s^2}\right)^{\frac{d-1-k}{2}} \frac{\left(\frac{(r-\|\mathbf{d}\|)^2}{2s^2}\right)^{\frac{k+1}{2}} e^{-\frac{(r-\|\mathbf{d}\|)^2}{2s^2}}}{\frac{(r-\|\mathbf{d}\|)^2}{2s^2} - \frac{k-1}{2}} \\
&\le \frac{2e^{\frac{\|\mathbf{d}\|^2}{2s^2}} p_{max}}{\Gamma\left(\frac{d}{2}\right)} \sum_{k=0}^{d-1} \binom{d-1}{k} \left(\frac{\|\mathbf{d}\|^2}{2s^2}\right)^{\frac{d-1-k}{2}} \left(\frac{(r-\|\mathbf{d}\|)^2}{2s^2}\right)^{\frac{k-1}{2}} e^{-\frac{(r-\|\mathbf{d}\|)^2}{2s^2}} \\
&= \frac{2e^{\frac{\|\mathbf{d}\|^2}{2s^2}} p_{max}}{\Gamma\left(\frac{d}{2}\right)} \left(\frac{\|\mathbf{d}\|}{\sqrt{2}s}\right)^{d-1} \sum_{k=0}^{d-1} \binom{d-1}{k} \left(\frac{r-\|\mathbf{d}\|}{\|\mathbf{d}\|}\right)^k e^{-\frac{(r-\|\mathbf{d}\|)^2}{2s^2}} \frac{\sqrt{2}s}{r-\|\mathbf{d}\|} \\
&= \frac{2e^{\frac{\|\mathbf{d}\|^2}{2s^2}} p_{max}}{\Gamma\left(\frac{d}{2}\right)} \left(\frac{\|\mathbf{d}\|}{\sqrt{2}s}\right)^{d-1} \left(\frac{r}{\|\mathbf{d}\|}\right)^{d-1} e^{-\frac{(r-\|\mathbf{d}\|)^2}{2s^2}} \frac{\sqrt{2}s}{r-\|\mathbf{d}\|} \\
&= p_{max} C_4(d) \frac{1}{r^2} s^2 =: c_2
\end{aligned}$$

the last equality uses $A < \frac{Lr}{4}$ and exponential decay of the function regarding $\frac{r}{s}$.

For the numerator, when $r > 3\|\mathbf{d}\|$ and $\frac{(r-\|\mathbf{d}\|)^2}{2s^2} > 2\left(\frac{d+1}{2} - 1\right)$, i.e., $s < \frac{r}{4\sqrt{\frac{d-1}{2}}}$,

$$\begin{aligned}
&\left\|\int_{S\cap\{\|\mathbf{z}-\mathbf{z}^*\|\geq r\}} \mathbf{u} e^{\theta^\top \mathbf{u}}\, \phi_t p_z \, d\mathbf{z}\right\| \\
&\leq e^{\frac{\|\mathbf{d}\|^2}{2s^2}} p_{max} \frac{\operatorname{Area}\left(S^{d-1}\right)}{(2\pi s^2)^{\frac{d}{2}}} \int_r^\infty r^d \exp\left(-\frac{(r-\|\mathbf{d}\|)^2}{2s^2}\right) dr \\
&= e^{\frac{\|\mathbf{d}\|^2}{2s^2}} p_{max} \frac{\operatorname{Area}\left(S^{d-1}\right)}{(2\pi s^2)^{\frac{d}{2}}} \sum_{k=0}^{d} \binom{d}{k} \|\mathbf{d}\|^{d-k} \int_{r-\|\mathbf{d}\|}^\infty w^k e^{-\frac{w^2}{2s^2}} dw \\
&= e^{\frac{\|\mathbf{d}\|^2}{2s^2}} p_{max} \frac{\operatorname{Area}\left(S^{d-1}\right)}{(2\pi s^2)^{\frac{d}{2}}} \sum_{k=0}^{d} \binom{d}{k} \|\mathbf{d}\|^{d-k} \left(2s^2\right)^{\frac{k-1}{2}} s \int_{\frac{(r-\|\mathbf{d}\|)^2}{2s^2}}^\infty v^{\frac{k-1}{2}} e^{-v} dv \\
&= \frac{\sqrt{2}\, s\, e^{\frac{\|\mathbf{d}\|^2}{2s^2}} p_{max}}{\Gamma\left(\frac{d}{2}\right)} \sum_{k=0}^{d} \binom{d}{k} \left(\frac{\|\mathbf{d}\|^2}{2s^2}\right)^{\frac{d-k}{2}} \int_{\frac{(r-\|\mathbf{d}\|)^2}{2s^2}}^\infty v^{\frac{k-1}{2}} e^{-v} dv.
\end{aligned}$$

Thus

$$\|I_{\text{num}}^{\text{far}}\|_2 \leq p_{max} C_3(d) \frac{1}{r^3} s^3 =: c_1.$$

③ The local case. Note that $\forall x < \frac{1}{2}$, we have $|e^x - 1| \leq 2|x|$. Thus when $L\|\mathbf{d}\| < 1$, we have $\|\mathbf{z} - \mathbf{z}^*\| \leq \frac{s}{(L\|\mathbf{d}\|)^{1/2}}$, (A.3) still holds. It yields that

$$|e^{\theta^\top \mathbf{u}} - 1| \leq L\|\mathbf{d}\| \|\Delta \mathbf{z}\|^2.$$

therefore,

$$\begin{aligned}
\left| I_{\text{den}}^{\text{loc}} - \int_{S\cap\{\|\mathbf{z}-\mathbf{z}^*\|\leq M\}} \phi_t p_z \, d\mathbf{z} \right| &\leq \int_{\{\|\mathbf{z}-\mathbf{z}^*\|\leq M\}} L\|\mathbf{d}\| \|\Delta \mathbf{z}\|^2 \phi_t p_z \, d\mathbf{z} \\
&\leq \frac{L\|\mathbf{d}\| p_{max} Area(S^{d-1})}{(2\pi s^2)^{\frac{d}{2}}} \int_0^M u^{d+1} e^{-\frac{u^2}{2s^2}} du \\
&= \frac{2 p_{max} L\|\mathbf{d}\| s^2}{\Gamma\left(\frac{d}{2}\right)} \int_0^{\frac{M^2}{2s^2}} w^{\frac{d}{2}} e^{-w} dw \\
&\leq p_{max} d \cdot L\|\mathbf{d}\| s^2 =: a_2.
\end{aligned}$$

and similarly,

$$\left| I_{\text{num}}^{\text{loc}} - \int_{S\cap\{\|\mathbf{z}-\mathbf{z}^*\|\leq M\}} \mathbf{u}\, \phi_t p_z \, d\mathbf{z} \right| \leq \int_{\{\|\mathbf{z}-\mathbf{z}^*\|\leq M\}} L\|\mathbf{d}\| \|\Delta \mathbf{z}\|^3\, \phi_t p_z \, d\mathbf{z} \leq \sqrt{2\pi}\, p_{max} (d+1) s^3 =: a_1.$$

We further need to give a upper bound of $\int_{\{\|\mathbf{z}-\mathbf{z}^*\|\geq M\}} \phi_t p_z \, d\mathbf{z}$. Apply Lemma B.1, if $\frac{1}{2L\|\mathbf{d}\|} > 2\left(\frac{d}{2} - 1\right)$ i.e. $L\|\mathbf{d}\| < (2d-4)^{-1}$, we observe

$$s_{\text{den}}^{\text{far}} = \int_{S\cap\{\|\mathbf{z}-\mathbf{z}^*\|\geq M\}} \phi_t p_z \, d\mathbf{z} \leq \frac{p_{max}}{\Gamma\left(\frac{d}{2}\right)} \int_{\frac{M^2}{2s^2}}^{+\infty} u^{\frac{d}{2}-1} e^{-u} du \leq \frac{p_{max}}{\Gamma\left(\frac{d}{2}\right)} \frac{\left(\frac{M^2}{2s^2}\right)^{\frac{d}{2}} e^{-\frac{M^2}{2s^2}}}{\frac{M^2}{2s^2} - \frac{d}{2} + 1}$$

$$\le \frac{2p_{max}}{\Gamma\left(\frac{d}{2}\right)}\left(\frac{M^2}{2s^2}\right)^{\frac{d}{2}-1} e^{-\frac{M^2}{2s^2}} = \frac{2p_{max}}{\Gamma\left(\frac{d}{2}\right)}\left(\frac{1}{2L\|\mathbf{d}\|}\right)^{\frac{d}{2}-1} e^{-\frac{1}{2L\|\mathbf{d}\|}}$$
$$\le p_{max}C_6(d)A =: d_2.$$

By utilizing the maximum of function $f(x) = x^{-\frac{d}{2}}e^{-\frac{1}{x}}$, we derive that $\left(\frac{1}{2L\|\mathbf{d}\|}\right)^{\frac{d}{2}-1} e^{-\frac{1}{2L\|\mathbf{d}\|}} \le 2\left(\frac{d}{2}\right)^{d/2} e^{-d/2}A$. Combining with Lemma B.3 which indicates $\Gamma(\frac{d}{2}) > \sqrt{2\pi}\left(\frac{d}{2}\right)^{\frac{d-1}{2}} e^{-d/2}$, we obtain that

$$\frac{2p_{max}}{\Gamma\left(\frac{d}{2}\right)}\left(\frac{1}{2L\|\mathbf{d}\|}\right)^{\frac{d}{2}-1} e^{-\frac{1}{2L\|\mathbf{d}\|}} \le \frac{4p_{max}A}{\sqrt{2\pi}\left(\frac{d}{2}\right)^{\frac{d-1}{2}} e^{-d/2}}\left(\frac{d}{2}\right)^{d/2} e^{-d/2} = \frac{2}{\sqrt{\pi}}p_{max}d^{1/2}A.$$

Thus we can denote $C_6(d) = \frac{2}{\sqrt{\pi}}d^{1/2}$.

For $\frac{1}{2L\|\mathbf{d}\|} > d-1$, i.e. $L\|\mathbf{d}\| < (2d-2)^{-1}$,

$$\begin{aligned}\left\|s^{\text{far}}_{\text{num}}\right\|_2 = \left\|\int_{S\cap\{\|\mathbf{z}-\mathbf{z}^*\|\ge M\}} \mathbf{u}\,\phi_t p_z\,d\mathbf{z}\right\|_2 &\le \frac{p_{max}s}{\Gamma\left(\frac{d}{2}\right)}\int_{\frac{M^2}{2s^2}}^{+\infty} u^{\frac{d-1}{2}}e^{-u}du \\ &\le \frac{p_{max}s}{\Gamma\left(\frac{d}{2}\right)}\frac{\left(\frac{1}{2L\|\mathbf{d}\|}\right)^{\frac{d+1}{2}} e^{-\frac{1}{2L\|\mathbf{d}\|}}}{\frac{1}{2L\|\mathbf{d}\|}-\frac{d-1}{2}} \\ &\le \frac{2p_{max}s}{\Gamma\left(\frac{d}{2}\right)}\left(\frac{1}{2L\|\mathbf{d}\|}\right)^{\frac{d-1}{2}} e^{-\frac{1}{2L\|\mathbf{d}\|}} \\ &\le p_{max}C_5(d)sA =: d_1.\end{aligned}$$

Due to the above definition from a_1 to d_2, and simultaneously $\frac{p_t^{LD}(\mathbf{z}^*)}{p_{max}}$ is lowered bounded by $dE_1\left(E_2+\frac{2}{\sqrt{\pi}}\right) \ge E_1\left(E_2 d+\frac{2}{\sqrt{\pi}}d^{1/2}\right) \ge \frac{2(a_2+d_2)}{p_{max}}$, thus $p_t^{LD}(\mathbf{z}^*) - a_2 - d_2 \ge \frac{1}{2}p_t^{LD}(\mathbf{z}^*)$. Then we are able to bound the difference between $s_{\|}(\mathbf{x},t)$ and $\hat{s}_{\|}(\mathbf{x},t)$:

$$\begin{aligned}&\left\|s_{\|}(\mathbf{x},t)-\hat{s}_{\|}(\mathbf{x},t)\right\|_2 \\ =\ &\frac{1}{s^2}\left\|\frac{I^{\text{loc}}_{\text{num}}+I^{\text{mid}}_{\text{num}}+I^{\text{far}}_{\text{num}}}{I^{\text{loc}}_{\text{den}}+I^{\text{mid}}_{\text{den}}+I^{\text{far}}_{\text{den}}}-\frac{s^{\text{loc}}_{\text{num}}+s^{\text{far}}_{\text{num}}}{s^{\text{loc}}_{\text{den}}+s^{\text{far}}_{\text{den}}}\right\|_2 \\ \le\ &\frac{\left\|I^{\text{loc}}_{\text{num}}s^{\text{loc}}_{\text{den}}-I^{\text{loc}}_{\text{den}}s^{\text{loc}}_{\text{num}}\right\|_2+\left\|(I^{\text{mid}}_{\text{num}}+I^{\text{far}}_{\text{num}})s_{\text{den}}\right\|_2+\left\|s_{\text{num}}(I^{\text{mid}}_{\text{den}}+I^{\text{far}}_{\text{den}})\right\|_2+\left\|I^{\text{loc}}_{\text{num}}s^{\text{far}}_{\text{den}}-I^{\text{loc}}_{\text{den}}s^{\text{far}}_{\text{num}}\right\|_2}{s^2p_t^{LD}(\mathbf{z}^*)\left(p_t^{LD}(\mathbf{z}^*)-a_2-d_2\right)} \\ \le\ &\frac{2}{s^2\left(p_t^{LD}(\mathbf{z}^*)\right)^2}\Big(a_1(p_t^{LD}(\mathbf{z}^*)-d_2)+a_2s^2p_t^{LD}(\mathbf{z}^*)\left\|\hat{s}_{\|}(\mathbf{x},t)\right\|_2+a_2d_1+(b_1+c_1)p_t^{LD}(\mathbf{z}^*) \\ &+(b_2+c_2)s^2p_t^{LD}(\mathbf{z}^*)\left\|\hat{s}_{\|}(\mathbf{x},t)\right\|_2+a_1d_2+a_2d_1\Big) \\ =\ &\frac{2\Lambda_1}{p_t^{LD}(\mathbf{z}^*)}+\frac{2\Lambda_2}{p_t^{LD}(\mathbf{z}^*)}\left\|\hat{s}_{\|}(\mathbf{x},t)\right\|_2+\frac{4a_2d_1}{s^2\left(p_t^{LD}(\mathbf{z}^*)\right)^2}\end{aligned} \tag{A.4}$$

where

$$\Lambda_1 = \frac{a_1+b_1+c_1}{s^2} = p_{max}\cdot C_7(d)\left(1+\frac{1}{r^3}\right)\left(s+\frac{A^2}{s}\right),$$

$$\Lambda_2 = a_2 + b_2 + c_2 = p_{max} \cdot C_8(d) \left(1 + \frac{1}{r^2}\right) \left(A + s^2\right).$$

and $C_7(d) = \max\{\sqrt{2\pi}(d+1), C_1(d), C_3(d)\}, C_8(d) = \max\{\frac{d}{2}, C_2(d), C_4(d)\}$. The second inequality of (A.4) is due to ① for the first term,

$$\begin{aligned}
\left\|I^{\text{loc}}_{\text{num}} s^{\text{loc}}_{\text{den}} - I^{\text{loc}}_{\text{den}} s^{\text{loc}}_{\text{num}}\right\|_2 &\leq \left\|(I^{\text{loc}}_{\text{num}} - s^{\text{loc}}_{\text{num}}) s^{\text{loc}}_{\text{den}}\right\|_2 + \left\|s^{\text{loc}}_{\text{num}}(I^{\text{loc}}_{\text{den}} - s^{\text{loc}}_{\text{den}})\right\|_2 \\
&\leq a_1 s^{\text{loc}}_{\text{den}} + a_2 s^{\text{loc}}_{\text{num}} \\
&\leq a_1 (p_t^{LD}(\mathbf{z}^*) - d_2) + a_2 \left(s^2 p_t^{LD}(\mathbf{z}^*) \left\|\hat{s}_\|(\mathbf{x}, t)\right\|_2 + d_1\right).
\end{aligned}$$

where $\|s^{\text{loc}}_{\text{num}}\|_2$ can be bounded by

$$\|s^{\text{loc}}_{\text{num}}\|_2 \leq \left\|\int_{S\cap\{\|\mathbf{z}-\mathbf{z}^*\|\leq M\}} \mathbf{u}\phi_t p_z \, d\mathbf{z}\right\|_2 \leq \|s_{\text{num}}\|_2 + d_1 = s^2 p_t^{LD}(\mathbf{z}^*) \left\|\hat{s}_\|(\mathbf{x}, t)\right\|_2 + d_1.$$

and ② for the last term,

$$\left\|I^{\text{loc}}_{\text{num}} s^{\text{far}}_{\text{den}} - I^{\text{loc}}_{\text{den}} s^{\text{far}}_{\text{num}}\right\|_2 \leq \left\|(I^{\text{loc}}_{\text{num}} - s^{\text{loc}}_{\text{num}}) s^{\text{far}}_{\text{den}}\right\|_2 + \left\|s^{\text{loc}}_{\text{num}}(I^{\text{loc}}_{\text{den}} - s^{\text{far}}_{\text{den}})\right\|_2 \leq a_1 d_2 + a_2 d_1.$$

We can further simplify the final expression of (A.4) into the form of $K_1 \left\|\hat{s}_\|(\mathbf{x}, t)\right\|_2 + K_2$, where

$$\begin{aligned}
K_1 &= \frac{2\Lambda_1}{p_t^{LD}(\mathbf{z}^*)} + \frac{4a_2 d_1}{s^2 \left(p_t^{LD}(\mathbf{z}^*)\right)^2} \leq C_9(d) \left(1 + \frac{1}{r^3}\right) \left(s + \frac{A^2}{s}\right) \left(\frac{1}{P} + \frac{1}{P^2}\right), \\
K_2 &= \frac{2\Lambda_2}{p_t^{LD}(\mathbf{z}^*)} \leq C_{10}(d) \left(1 + \frac{1}{r^2}\right) \left(A + s^2\right) \frac{1}{P},
\end{aligned}$$

where $C_9(d) = \max\{2C_7(d), C_5(d)\}, C_{10}(d) = 2C_8(d)$, and that we have

$$\begin{aligned}
\left\|s_\|(\mathbf{x}, t) - \hat{s}_\|(\mathbf{x}, t)\right\|_2 &\leq K_1 \left\|\hat{s}_\|(\mathbf{x}, t)\right\|_2 + K_2 \\
&\leq C(d) \left(1 + \frac{1}{r^2} + \frac{1}{r^3}\right) \left(\frac{1}{P} + \frac{1}{P^2}\right) \left(\left(A + s^2\right) \left\|\hat{s}_\|(\mathbf{x}, t)\right\|_2 + \left(s + \frac{A^2}{s}\right)\right).
\end{aligned}$$

where $C(d) = \max\{C_9(d), C_{10}(d)\}$. The proof is complete. □

Proof of Lemma 3.2. Define

$$R(\mathbf{z}^*) = \hat{s}_\|(\mathbf{x}, t) - J_\psi(\mathbf{z}^*) \nabla \log p_t^{LD}(\mathbf{z}^*) = \frac{\int \left(\psi(\mathbf{z}) - \psi(\mathbf{z}^*) - J_\psi(\mathbf{z}^*)\left(\mathbf{z} - \mathbf{z}^*\right)\right) \phi_t(\mathbf{z}^* \mid \mathbf{z}) p_z(\mathbf{z}) \, d\mathbf{z}}{s^2 \int \phi_t(\mathbf{z}^* \mid \mathbf{z}) p_z(\mathbf{z}) \, d\mathbf{z}}$$

Fix r. The residue can be separated into two terms:

$$\begin{aligned}
&\int_S \left(\psi(\mathbf{z}) - \psi(\mathbf{z}^*) - J_\psi(\mathbf{z}^*)\left(\mathbf{z} - \mathbf{z}^*\right)\right) \phi_t(\mathbf{z}^* \mid \mathbf{z}) p_z(\mathbf{z}) \, d\mathbf{z} \\
= &\int_{B_r(\mathbf{z}^*)\cap S} \left(\psi(\mathbf{z}) - \psi(\mathbf{z}^*) - J_\psi(\mathbf{z}^*)\left(\mathbf{z} - \mathbf{z}^*\right)\right) \phi_t(\mathbf{z}^* \mid \mathbf{z}) p_z(\mathbf{z}) \, d\mathbf{z} \\
&+ \int_{S\setminus B_r(\mathbf{z}^*)} \left(\psi(\mathbf{z}) - \psi(\mathbf{z}^*) - J_\psi(\mathbf{z}^*)\left(\mathbf{z} - \mathbf{z}^*\right)\right) \phi_t(\mathbf{z}^* \mid \mathbf{z}) p_z(\mathbf{z}) \, d\mathbf{z}
\end{aligned}$$

By Lemma B.2, the first term can be simplified as:

$$
\begin{aligned}
&\left\|\int_{B_r(\mathbf{z}^*)\cap S}\left(\psi(\mathbf{z})-\psi(\mathbf{z}^*)-J_\psi(\mathbf{z}^*)\left(\mathbf{z}-\mathbf{z}^*\right)\right)\phi_t(\mathbf{z}^*\mid\mathbf{z})p_z(\mathbf{z})\,d\mathbf{z}\right\|_2\\
&\leq \frac{L}{2}\int_{B_r(\mathbf{z}^*)\cap S}\|\mathbf{z}-\mathbf{z}^*\|_2^2\,\phi_t(\mathbf{z}^*\mid\mathbf{z})\,p_z(\mathbf{z})\,d\mathbf{z}\\
&\leq \frac{Lp_{max}}{2}\int_{B_r(\mathbf{z}^*)}\|\mathbf{z}-\mathbf{z}^*\|_2^2\,\phi_t(\mathbf{z}^*\mid\mathbf{z})\,d\mathbf{z}\\
&=\frac{Lp_{max}}{2}\cdot\frac{2s^2}{\Gamma(\frac{d}{2})}\int_0^{\frac{r^2}{2s^2}}u^{\frac{d}{2}}e^{-u}\,du=\frac{\gamma\left(\frac{d}{2}+1,\frac{r^2}{2s^2}\right)}{\Gamma(\frac{d}{2})}Ls^2\leq\frac{1}{2}p_{max}Lds^2
\end{aligned}
$$

The second term can be bounded as:

$$
\begin{aligned}
&\left\|\int_{S\setminus B_r(\mathbf{z}^*)}\left(\psi(\mathbf{z})-\psi(\mathbf{z}^*)-J_\psi(\mathbf{z}^*)\left(\mathbf{z}-\mathbf{z}^*\right)\right)\phi_t(\mathbf{z}^*\mid\mathbf{z})p_z(\mathbf{z})\,d\mathbf{z}\right\|\\
&\leq p_{max}\int_{S\setminus B_r(\mathbf{z}^*)}\|\psi(\mathbf{z})-\psi(\mathbf{z}^*)-J_\psi(\mathbf{z}^*)\left(\mathbf{z}-\mathbf{z}^*\right)\|\,\phi_t(\mathbf{z}^*\mid\mathbf{z})\,d\mathbf{z}\\
&\leq p_{max}\int_{S\setminus B_r(\mathbf{z}^*)}\left(\|\psi(\mathbf{z})-\psi(\mathbf{z}^*)\|+\|J_\psi(\mathbf{z}^*)\left(\mathbf{z}-\mathbf{z}^*\right)\|\right)\phi_t(\mathbf{z}^*\mid\mathbf{z})\,d\mathbf{z}\\
&\leq p_{max}\int_{S\setminus B_r(\mathbf{z}^*)}\left(1+\|J_\psi(\mathbf{z}^*)\|\right)\|\mathbf{z}-\mathbf{z}^*\|\,\phi_t(\mathbf{z}^*\mid\mathbf{z})\,d\mathbf{z}\\
&=p_{max}\left(1+\|J_\psi(\mathbf{z}^*)\|\right)\frac{\sqrt{2}s}{\Gamma(\frac{d}{2})}\Gamma\left(\frac{d+1}{2},\frac{r^2}{2s^2}\right)
\end{aligned}
$$

When $\frac{r^2}{2s^2}>d-1$,

$$
\begin{aligned}
\left(1+\|J_\psi(\mathbf{z}^*)\|\right)\frac{\sqrt{2}s}{\Gamma(\frac{d}{2})}\Gamma\left(\frac{d+1}{2},\frac{r^2}{2s^2}\right)&\leq\left(1+\|J_\psi(\mathbf{z}^*)\|\right)\frac{\sqrt{2s^2}}{\Gamma(\frac{d}{2})}\frac{\left(\frac{r^2}{2s^2}\right)^{\frac{d+1}{2}}e^{-\frac{r^2}{2s^2}}}{\frac{r^2}{2s^2}-\frac{d-1}{2}}\\
&\leq 2\left(1+\|J_\psi(\mathbf{z}^*)\|\right)\frac{\sqrt{2s^2}}{\Gamma(\frac{d}{2})}\left(\frac{r^2}{2s^2}\right)^{\frac{d-1}{2}}e^{-\frac{r^2}{2s^2}} \qquad \text{(A.5)}
\end{aligned}
$$

where the first inequality is due to Lemma B.1, and the second is due to $\frac{r^2}{2s^2}>d-1$. By utilizing the minima of $f(s)=\frac{e^{-\frac{r^2}{2s^2}}}{s^d}$, we derive that $\frac{e^{-\frac{r^2}{2s^2}}}{s^{d-2}}\leq\frac{d^{\frac{d}{2}}}{r^de^{\frac{d}{2}}}s^2$, thus

$$
\left(1+\|J_\psi(\mathbf{z}^*)\|\right)\frac{\sqrt{2}s}{\Gamma(\frac{d}{2})}\Gamma\left(\frac{d+1}{2},\frac{r^2}{2s^2}\right)\leq 4\left(1+\|J_\psi(\mathbf{z}^*)\|\right)\frac{1}{r\,\Gamma(\frac{d}{2})}e^{-\frac{d}{2}}\left(\frac{d}{2}\right)^{\frac{d}{2}}s^2 \qquad \text{(A.6)}
$$

By combining (A.5) and (A.6), we obtain

$$
\left(1+\|J_\psi(\mathbf{z}^*)\|\right)\frac{\sqrt{2s^2}}{\Gamma(\frac{d}{2})}\int_{\frac{r^2}{2s^2}}^{\infty}u^{\frac{d-1}{2}}e^{-u}du\leq 4\left(1+\|J_\psi(\mathbf{z}^*)\|\right)\frac{1}{r\,\Gamma(\frac{d}{2})}e^{-\frac{d}{2}}\left(\frac{d}{2}\right)^{\frac{d}{2}}s^2 \qquad \text{(A.7)}
$$

Thus

$$
\left\|\int_S\left(\psi(\mathbf{z})-\psi(\mathbf{z}^*)-J_\psi(\mathbf{z}^*)\left(\mathbf{z}-\mathbf{z}^*\right)\right)\phi_t(\mathbf{z}^*\mid\mathbf{z})p_z(\mathbf{z})\,d\mathbf{z}\right\|
$$

$$
\begin{aligned}
&\leq \left\| \int_{B_r(\mathbf{z}^*)\cap S} \left(\psi(\mathbf{z}) - \psi(\mathbf{z}^*) - J_\varphi(\mathbf{z}^*)\,(\mathbf{z}-\mathbf{z}^*)\right) \phi_t(\mathbf{z}^* \mid \mathbf{z}) p_z(\mathbf{z})\, d\mathbf{z} \right\| \\
&\quad + \left\| \int_{S\setminus B_r(\mathbf{z}^*)} \left(\psi(\mathbf{z}) - \psi(\mathbf{z}^*) - J_\varphi(\mathbf{z}^*)\,(\mathbf{z}-\mathbf{z}^*)\right) \phi_t(\mathbf{z}^* \mid \mathbf{z}) p_z(\mathbf{z})\, d\mathbf{z} \right\| \\
&\leq p_{max} \left(\frac{1}{2} L d + 4\left(1 + \|J_\psi(\mathbf{z}^*)\|\right) \frac{e^{-\frac{d}{2}} \left(\frac{d}{2}\right)^{\frac{d}{2}}}{r\,\Gamma(\frac{d}{2})} \right) s^2
\end{aligned}
$$

Finally, we conclude the above discussion:

$$
\begin{aligned}
\left\|\hat{s}_{\parallel}(\mathbf{x},t) - J_\psi(\mathbf{z}^*)\nabla \log p_t^{LD}(\mathbf{z}^*)\right\|_2 &\leq \frac{\left(\frac{1}{2} L d + 4\left(1 + \|J_\psi(\mathbf{z}^*)\|\right) \frac{e^{-\frac{d}{2}} \left(\frac{d}{2}\right)^{\frac{d}{2}}}{r\,\Gamma(\frac{d}{2})} \right) s^2}{s^2\, p_t^{LD}(\mathbf{z}^*)} \\
&\leq \frac{1}{P}\left(\frac{1}{2} L d + 4\left(1 + \|J_\psi(\mathbf{z}^*)\|\right) \frac{e^{-\frac{d}{2}} \left(\frac{d}{2}\right)^{\frac{d}{2}}}{r\,\Gamma(\frac{d}{2})} \right).
\end{aligned}
$$

The proof is complete. □

Proof of Lemma 3.3. By straight computation and change of variable, we observe

$$
\begin{aligned}
\nabla \log p_t^{LD}(\mathbf{z}^*) &= \frac{\sqrt{2}\int_{\frac{(S-\mathbf{z}^*)}{\sqrt{2}s}} \mathbf{v}\, e^{-\|\mathbf{v}\|^2} p_z(\mathbf{z}^* + \sqrt{2}s\mathbf{v})\, d\mathbf{v}}{s\,\pi^{d/2}\, p_t^{LD}(\mathbf{z}^*)} \\
&= \frac{\sqrt{2}\int_{\frac{(S-\mathbf{z}^*)}{\sqrt{2}s}} \mathbf{v}\, e^{-\|\mathbf{v}\|^2} \left(p_z(\mathbf{z}^*) + \sqrt{2}s\mathbf{v}^\top \nabla p_z(\mathbf{z}^* + t_{\mathbf{v}}\mathbf{v})\right)\, d\mathbf{v}}{s\,\pi^{d/2}\, p_t^{LD}(\mathbf{z}^*)}
\end{aligned}
$$

here $0 < t_{\mathbf{v}} < 1$ is a constant regrading $\mathbf{v}$. Thus,

$$
\left\|\nabla \log p_t^{LD}(\mathbf{z}^*)\right\|_2 \leq \left\| \frac{\sqrt{2}\, p_z(\mathbf{z}^*) \int_{\left(\frac{(S-\mathbf{z}^*)}{\sqrt{2}s}\right)^C} \mathbf{v}\, e^{-\|\mathbf{v}\|^2}\, d\mathbf{v}}{s\,\pi^{d/2}\, p_t^{LD}(\mathbf{z}^*)} \right\|_2 + \frac{2D_U \int_{\mathbb{R}^d} \|\mathbf{v}\|^2\, e^{-\|\mathbf{v}\|^2} d\mathbf{v}}{\pi^{d/2}\, p_t^{LD}(\mathbf{z}^*)} =: ① + ②
$$

Since $B_0(\varepsilon) \subset S - \mathbf{z}^*$, and by Lemma B.1, we deduce that when $s < \frac{r}{4\sqrt{\frac{d-1}{2}}} \leq \frac{\varepsilon}{4\sqrt{\frac{d-1}{2}}}$,

$$
\begin{aligned}
\left\| \int_{\left(\frac{(S-\mathbf{z}^*)}{\sqrt{2}s}\right)^C} \mathbf{v} e^{-\|\mathbf{v}\|^2}\, d\mathbf{v} \right\|_2 &\leq \int_{\left(B_0\left(\frac{\varepsilon}{\sqrt{2}s}\right)\right)^C} \|\mathbf{v}\|\, e^{-\|\mathbf{v}\|^2}\, d\mathbf{v} \\
&\leq \frac{\pi^{d/2}}{\Gamma(d/2)} \int_{\frac{\varepsilon^2}{2s^2}}^{\infty} e^{-x} x^{\frac{d-1}{2}}\, dx \\
&\leq \frac{\pi^{d/2}}{\Gamma(d/2)} \frac{\left(\frac{\varepsilon^2}{2s^2}\right)^{\frac{d+1}{2}} e^{-\frac{\varepsilon^2}{2s^2}}}{\frac{\varepsilon^2}{2s^2} - \frac{d}{2} + 1} \\
&\leq \frac{2\pi^{d/2}}{\Gamma(d/2)} \left(\frac{\varepsilon^2}{2s^2}\right)^{\frac{d-1}{2}} e^{-\frac{\varepsilon^2}{2s^2}}
\end{aligned}
$$

$$\leq C_1(d)\frac{s^2}{\varepsilon^2}.$$

The last inequality is by same technique we used in proof of Lemma 3.1. $C_1(d)$ is a constant depending on d. We then obtain

$$① \leq \frac{C_1(d)p_z(\mathbf{z}^*)}{\varepsilon^2 p_t^{LD}(\mathbf{z}^*)}s, \quad ② \leq \frac{D_U C_2(d)}{p_t^{LD}(\mathbf{z}^*)}.$$

where $C_2(d) = \int_{\mathbb{R}^d} \|\mathbf{v}\|^2 e^{-\|\mathbf{v}\|^2} d\mathbf{v}$.

Using Lemma A.1, we indicates that if $s < \frac{p_z(\mathbf{z}^*)}{4\sqrt{2\pi}\left(\frac{p_z(\mathbf{z}^*)}{\varepsilon}+D_U\right)}$, we obtain $p_t^{LD}(\mathbf{z}^*) > \frac{1}{2}p_z(\mathbf{z}^*)$, thus deducing

$$\left\|\nabla \log p_t^{LD}(\mathbf{z}^*)\right\|_2 \leq C(d)\left(\frac{s}{\varepsilon^2} + \frac{D_U}{p_z(\mathbf{z}^*)}\right).$$

□

A.2 Score decomposition for general manifolds

We refine the assumptions on isometric manifolds as follows:

Assumption A.1. *Let $\mathbf{x}$ be a fixed point. Consider a coordinate chart $\{(U_i, \psi_i)\}_{i\in I}$ of $\mathcal{M}$, where $\psi_i : V_i \subset \mathbb{R}^d \to U_i \subset \mathcal{M} \subset \mathbb{R}^D$, with V_i and U_i open in $\mathbb{R}^d$ and $\mathcal{M}$, respectively. Assume $\mathbf{x} \in U_{i^*}$ and ψ_{i^*} is an isometry. Each U_i is a ball centered at $\mathbf{x}_i$ with radius $r_i < \mathrm{inj}(\mathcal{M})$, and $k = \inf_{\mathbf{y}\in\cup_{i\neq i^*}U_i} \|\mathbf{x}-\mathbf{y}\| > 0$. Require $\mathcal{M}$ to be C^1.*

For $\mathbf{z} \in V_i$, let the latent variable follow a continuous distribution P_i with density p_i. Define soft weights $w_i(\mathbf{z})$ such that

$$w_i(\mathbf{z}) \geq 0, \quad \sum_{i\in I} w_i(\psi_i^{-1}(\mathbf{x})) = 1 \quad (\forall \mathbf{x} \in \mathcal{M}), \quad \sum_{i\in I}\int_{V_i} w_i(\mathbf{z})p_i(\mathbf{z})\, d\mathbf{z} = 1.$$

Assume $p_i \leq p_{\max} < \infty$ and $|\nabla(w_i p_i)| \leq D_U$ for constants $p_{\max}, D_U > 0$.

These restrictions can be satisfied by an appropriate choice of chart. By the definition of a manifold, such a set U_{i^*} exists. Since $\mathbf{x}$ lies in the interior of U_{i^*}, there is $\epsilon > 0$ with $B_\epsilon(\mathbf{x}) \subset U_{i^*}$, where $B_\epsilon(\mathbf{x}) := \{\mathbf{y} : \|\mathbf{x}-\mathbf{y}\| < \epsilon\}$. Setting $k = \epsilon/2$, we cover each $\mathbf{y} \notin B_k(\mathbf{x})$ with a ball of radius less than $\|\mathbf{y}-\mathbf{x}\|/2$. Compactness of $\mathcal{M}$ ensures that finitely many such balls, together with U_{i^*}, cover $\mathcal{M}$.

Since $\mathcal{M}$ is a C^1 manifold, the Jacobians J_{ψ_i} are continuous. On the compact sets $\overline{U_i \cap \mathcal{M}}$, each J_{ψ_i} is bounded, and compactness of $\mathcal{M}$ guarantees a uniform bound across all charts. We denote this bound as S.

Assumption A.2. *Jacobian of embedding (U_{i^*}, ψ_{i^*}) which contains $\mathbf{x}$ is L-Lipschitz: for any $\mathbf{z}, \mathbf{z}' \in S$, $\left\|J_{\psi_{i^*}}(\mathbf{z}) - J_{\psi_{i^*}}(\mathbf{z}')\right\|_2 \leq L\|\mathbf{z}-\mathbf{z}'\|_2$.*

Lemma A.1 can be refined as follows. The proof is almost the same as Lemma A.1. We only need to do what had applied to $p_z(\mathbf{z})$ to $w_{i^*}(\mathbf{z})p_{i^*}(\mathbf{z})$.

Lemma A.2. *The latent disturbed distribution is defined as*

$$p_t^{LD}(\mathbf{z}^*) = \int_{V_{i^*}} \phi_t(\mathbf{z}^* \mid \mathbf{z})w_{i^*}(\mathbf{z})p_{i^*}(\mathbf{z})\, d\mathbf{z}, \tag{A.8}$$

if $s < \frac{p_{i^*}(\mathbf{z}^*)}{4\sqrt{2\pi}\left(\frac{p_{i^*}(\mathbf{z}^*)}{\varepsilon}+D_U\right)}$, *we have*

$$p_t^{LD}(\mathbf{z}^*) > \frac{1}{2} w_{i^*}(\mathbf{z}) p_{i^*}(\mathbf{z}^*).$$

We refine our Theorem as follows:

Theorem A.1 (Score decomposition)**.** *Assume Assumptions A.1 and A.2 hold, and let* $\mathbf{z}^* = \arg\min_{\mathbf{z}\in V_i} \|\mathbf{x} - \psi_{i^*}(\mathbf{z})\|_2^2$ *with Lipschitz radius* r*. Define* $s^2 = \int_0^t \sigma^2(\tau)\, d\tau$*,* $\mathbf{d} = \psi_{i^*}(\mathbf{z}^*) - \mathbf{x}$*, and* $P = \frac{w_{i^*}(\mathbf{z}^*)p_{i^*}(\mathbf{z}^*)}{p_{\max}}$*. Suppose* $w_{i^*}(\mathbf{z}^*)p_{i^*}(\mathbf{z}^*) > 0$*. Choose constants* A, B *such that* $P \geq A\left(B + \frac{2}{\sqrt{\pi}}\right) d$ *and* $A < \min\left\{\frac{1}{2}, \frac{1}{4(d-1)}, \frac{Lr}{4}, \frac{\min_{i\neq i^*}(d_i - r_i)}{4}\right\}$*. Assume* $L\|\mathbf{d}\| < A$ *and*

$$s < \min\left\{\frac{r}{4\sqrt{\frac{d-1}{2}}},\ B,\ \frac{w_{i^*}(\mathbf{z}^*)p_{i^*}(\mathbf{z}^*)}{4\sqrt{2\pi}\left(\frac{w_{i^*}(\mathbf{z}^*)p_{i^*}(\mathbf{z}^*)}{r} + D_U\right)},\ \frac{3\min_{i\neq i^*}(d_i - r_i)}{8\sqrt{\frac{D-1}{2}}}\right\}$$

. Then the score function satisfies

$$\left\|\nabla \log p_t(\mathbf{x}) - \frac{1}{s^2}(\psi_{i^*}(\mathbf{z}^*) - \mathbf{x})\right\|_2 \leq C(d)\left(1 + \frac{1}{r^2} + \frac{1}{r^3}\right)\left(\frac{1}{P} + \frac{1}{P^2}\right)\left(1 + S\left[\sum_{i\neq i^*}(d_i - r_i)^{D-d-2}\right]\right)$$
$$\times\left[(s^2 + A + 1)\left(\frac{Lr + 1 + \|J_{\psi_{i^*}}(\mathbf{z}^*)\|_2}{Pr} + \frac{s}{r^2} + \frac{D_U}{w_{i^*}(\mathbf{z}^*)p_{i^*}(\mathbf{z}^*)}\right) + s + \frac{A^2}{s}\right], \tag{A.9}$$

where $C(d)$ *depends only on the intrinsic dimension* d*.*

The refinement of Lemma 3.2 and 3.3 can be straightforwardly generalized through same technique as Lemma A.1. We only need to show how Lemma 3.1 transforms.

Lemma A.3. *Under the setting of Theorem A.1, let* $\phi_{i^*}(\mathbf{z}^* \mid \mathbf{z}) = (2\pi s^2)^{-d/2} \exp\left(-\frac{1}{2s^2}\|\mathbf{z} - \mathbf{z}^*\|_2^2\right)$ *denote the Gaussian density, and define*

$$\hat{s}_{\|}(\mathbf{x}, t) = \frac{\int_{V_{i^*}} \frac{1}{s^2}\left(\psi_{i^*}(\mathbf{z}) - \psi_{i^*}(\mathbf{z}^*)\right)\phi_{i^*}(\mathbf{z}^* \mid \mathbf{z}) w_{i^*}(\mathbf{z}) p_{i^*}(\mathbf{z})\, d\mathbf{z}}{\int_{V_{i^*}} \phi_{i^*}(\mathbf{z}^* \mid \mathbf{z}) w_{i^*}(\mathbf{z}) p_{i^*}(\mathbf{z})\, d\mathbf{z}},$$

Then

$$\left\|\nabla \log p_t(\mathbf{x}) - \hat{\mathbf{s}}_{\|}(\mathbf{x}, t) - \frac{\psi_{i^*}(\mathbf{z}^*) - \mathbf{x}}{s^2}\right\|_2 \leq C(d)\left(1 + \frac{1}{r^2} + \frac{1}{r^3}\right)\left(\frac{1}{P} + \frac{1}{P^2}\right)$$
$$\times\left(1 + S\left[\sum_{i\neq i^*}(d_i - r_i)^{D-d-2}\right]\right)\left[(s^2 + A)\|\hat{\mathbf{s}}_{\|}(\mathbf{x}, t)\|_2 + s + \frac{A^2}{s}\right],$$

where $C(d)$ *depends only on the intrinsic dimension* d*.*

Proof. Due to the noising process and Assumption A.1, the probability function of $\mathbf{x}_0$ can be expressed by latent variable $\mathbf{z}$:

$$p_t(\mathbf{x}) = \sum_{i \in I} \int_{V_i} w_i(\mathbf{z}) \phi_t(\mathbf{x}|\psi_i(\mathbf{z})) p_i(\mathbf{z}) \, d\mathbf{z},$$

where

$$\phi_t(\mathbf{x}|\psi_i(\mathbf{z})) = (2\pi)^{-D/2} \left(\int_0^t \sigma^2(\tau) d\tau \right)^{-D/2} \exp\left(-\frac{1}{2 \int_0^t \sigma^2(\tau) d\tau} \|\psi_i(\mathbf{z}) - \mathbf{x}\|_2^2 \right).$$

Then the score function can be written as

$$\nabla \log p_t(\mathbf{x}) = \frac{\sum_{i \in I} \nabla \int_{V_i} w_i(\mathbf{z}) \phi_t(\mathbf{x}|\psi(\mathbf{z})) p_i(\mathbf{z}) \, d\mathbf{z}}{\sum_{i \in I} \int_{V_i} w_i(\mathbf{z}) \phi_t(\mathbf{x}|\psi(\mathbf{z})) p_i(\mathbf{z}) \, d\mathbf{z}} = \frac{\sum \int w_i(\mathbf{z}) \nabla \phi_t(\mathbf{x}|\psi(\mathbf{z})) p_i(\mathbf{z}) \, d\mathbf{z}}{\sum \int w_i(\mathbf{z}) \phi_t(\mathbf{x}|\psi(\mathbf{z})) p_i(\mathbf{z}) \, d\mathbf{z}}, \tag{A.10}$$

where the last equality holds since $\phi_t(\mathbf{x}|\psi(\mathbf{z}))$ is continuously differentiable in $\mathbf{x}$. Substituting $\phi_t(\mathbf{x}|\psi_i(\mathbf{z}))$ into (A.10) gives rise to

$$\begin{aligned}
&\nabla \log p_t(\mathbf{x}) \\
&= \frac{(2\pi)^{-D/2} \left(\int_0^t \sigma^2(\tau) d\tau \right)^{-D/2} \sum_{i \in I} \int_{V_i} \frac{1}{\int_0^t \sigma^2(\tau) d\tau} (\psi_i(\mathbf{z}) - \mathbf{x}) \exp\left(-\frac{1}{2 \int_0^t \sigma^2(\tau) d\tau} \|\psi_i(\mathbf{z}) - \mathbf{x}\|_2^2 \right) w_i(\mathbf{z}) p_i(\mathbf{z}) \, d\mathbf{z}}{\sum_{i \in I} \int_{V_i} w_i(\mathbf{z}) \phi_t(\mathbf{x}|\psi_i(\mathbf{z})) p_i(\mathbf{z}) \, d\mathbf{z}} \\
&= \frac{(2\pi)^{-D/2} \left(\int_0^t \sigma^2(\tau) d\tau \right)^{-D/2} \sum_{i \in I} \int_{V_i} \frac{1}{\int_0^t \sigma^2(\tau) d\tau} (\psi_i(\mathbf{z}) - \psi_{i^*}(\mathbf{z}^*)) \exp\left(-\frac{1}{2 \int_0^t \sigma^2(\tau) d\tau} \|\psi_i(\mathbf{z}) - \mathbf{x}\|_2^2 \right) w_i(\mathbf{z}) p_i(\mathbf{z}) \, d\mathbf{z}}{\sum_{i \in I} \int_{V_i} w_i(\mathbf{z}) \phi_t(\mathbf{x}|\psi_i(\mathbf{z})) p_i(\mathbf{z}) \, d\mathbf{z}} \\
&\quad + \frac{(2\pi)^{-D/2} \left(\int_0^t \sigma^2(\tau) d\tau \right)^{-D/2} \sum_{i \in I} \int_{V_i} \frac{1}{\int_0^t \sigma^2(\tau) d\tau} (\psi_{i^*}(\mathbf{z}^*) - \mathbf{x}) \cdot \exp\left(-\frac{1}{2 \int_0^t \sigma^2(\tau) d\tau} \|\psi_i(\mathbf{z}) - \mathbf{x}\|_2^2 \right) w_i(\mathbf{z}) p_i(\mathbf{z}) \, d\mathbf{z}}{\sum_{i \in I} \int_{V_i} w_i(\mathbf{z}) \phi_t(\mathbf{x}|\psi_i(\mathbf{z})) p_i(\mathbf{z}) \, d\mathbf{z}} \\
&= \underbrace{\frac{1}{\sum_{i \in I} \int_{V_i} w_i(\mathbf{z}) \phi_t(\mathbf{x}|\psi_i(\mathbf{z})) p_i(\mathbf{z}) \, d\mathbf{z}} \int \frac{1}{\int_0^t \sigma^2(\tau) d\tau} (\psi_i(\mathbf{z}) - \psi_{i^*}(\mathbf{z}^*)) \phi_t(\mathbf{x}|\psi_i(\mathbf{z})) w_i(\mathbf{z}) p_i(\mathbf{z}) \, d\mathbf{z}}_{\mathbf{s}_\parallel} \\
&\quad + \underbrace{\frac{1}{\int_0^t \sigma^2(\tau) d\tau} (\psi_{i^*}(\mathbf{z}^*) - \mathbf{x})}_{\mathbf{s}_\perp}.
\end{aligned}$$

We can further simplify $\mathbf{s}_\parallel$. We decompose $\phi_t(\mathbf{x} \mid \psi_i(\mathbf{z}))$ as

$$\begin{aligned}
&\phi_t(\mathbf{x} \mid \psi_i(\mathbf{z})) \\
&= (2\pi)^{-D/2} \left(\int_0^t \sigma^2(\tau) d\tau \right)^{-D/2} \exp\left(-\frac{1}{2 \int_0^t \sigma^2(\tau) d\tau} \|\psi_i(\mathbf{z}) - \psi_{i^*}(\mathbf{z}^*) + \psi_{i^*}(\mathbf{z}^*) - \mathbf{x}\|_2^2 \right) \\
&= (2\pi)^{-D/2} \left(\int_0^t \sigma^2(\tau) d\tau \right)^{-D/2} \exp\left(-\frac{1}{2 \int_0^t \sigma^2(\tau) d\tau} \left(\|\psi_i(\mathbf{z}) - \psi_{i^*}(\mathbf{z}^*)\|_2^2 + \|\psi_{i^*}(\mathbf{z}^*) - \mathbf{x}\|_2^2 \right.\right. \\
&\quad \left.\left. + 2 \left(\psi_i(\mathbf{z}) - \psi_{i^*}(\mathbf{z}^*) \right)^\top \left(\psi_{i^*}(\mathbf{z}^*) - \mathbf{x} \right) \right) \right)
\end{aligned}$$

$$= (2\pi)^{-d/2} \left(\int_0^t \sigma^2(\tau) d\tau \right)^{-d/2} \exp\left(-\frac{1}{2\int_0^t \sigma^2(\tau)d\tau} \left(\|\psi_i(\mathbf{z}) - \psi_{i^*}(\mathbf{z}^*)\|_2^2 + 2\left(\psi_i(\mathbf{z}) - \psi_{i^*}(\mathbf{z}^*)\right)^\top \left(\psi_{i^*}(\mathbf{z}^*) - \mathbf{x}\right) \right) \right)$$

$$\times (2\pi)^{-(D-d)/2} \left(\int_0^t \sigma^2(\tau) d\tau \right)^{-(D-d)/2} \exp\left(-\frac{1}{2\int_0^t \sigma^2(\tau)d\tau} \|\psi_{i^*}(\mathbf{z}^*) - \mathbf{x}\|_2^2 \right).$$

We denote

$$\phi_i(\mathbf{z}^* \mid \mathbf{z}) = (2\pi)^{-d/2} \left(\int_0^t \sigma^2(\tau) d\tau \right)^{-d/2} \exp\left(-\frac{1}{2\int_0^t \sigma^2(\tau)d\tau} \|\psi_i(\mathbf{z}) - \psi_{i^*}(\mathbf{z}^*)\|_2^2 \right)$$

and

$$\phi_t(\mathbf{x}) = (2\pi)^{-(D-d)/2} \left(\int_0^t \sigma^2(\tau) d\tau \right)^{-(D-d)/2} \exp\left(-\frac{1}{2\int_0^t \sigma^2(\tau)d\tau} \|\psi_{i^*}(\mathbf{z}^*) - \mathbf{x}\|_2^2 \right)$$

being both Gaussian densities. Substituting

$$\phi_t(\mathbf{x} \mid \psi_i(\mathbf{z})) = \phi_i(\mathbf{z}^* \mid \mathbf{z}) \cdot \phi_t(\mathbf{x}) \cdot \exp\left(-\frac{1}{\int_0^t \sigma^2(\tau)d\tau} \left(\psi_i(\mathbf{z}) - \psi_{i^*}(\mathbf{z}^*)\right)^\top \left(\psi_{i^*}(\mathbf{z}^*) - \mathbf{x}\right) \right)$$

into $\mathbf{s}_{\|}$, we obtain:

$$\mathbf{s}_{\|}(\mathbf{x}, t)$$
$$= \frac{\sum_{i\in I} \int_{V_i} \frac{1}{\int_0^t \sigma^2(\tau)d\tau} \left(\psi_i(\mathbf{z}) - \psi_{i^*}(\mathbf{z}^*)\right) \phi_i(\mathbf{z}^* \mid \mathbf{z}) \exp\left(-\frac{1}{\int_0^t \sigma^2(\tau)d\tau} \left(\psi_i(\mathbf{z}) - \psi_{i^*}(\mathbf{z}^*)\right)^\top \left(\psi_{i^*}(\mathbf{z}^*) - \mathbf{x}\right) \right) w_i(\mathbf{z}) p_i(\mathbf{z})\, d\mathbf{z}}{\sum_{i\in I} \int_{V_i} \phi_i(\mathbf{z}^* \mid \mathbf{z}) \exp\left(-\frac{1}{\int_0^t \sigma^2(\tau)d\tau} \left(\psi_i(\mathbf{z}) - \psi_{i^*}(\mathbf{z}^*)\right)^\top \left(\psi_{i^*}(\mathbf{z}^*) - \mathbf{x}\right) \right) w_i(\mathbf{z}) p_i(\mathbf{z})\, d\mathbf{z}}$$

Fix $i \neq i^*$. Denote $\mathbf{u} := \psi_i(\mathbf{z}) - \psi_{i^*}(\mathbf{z}^*)$, $d_i = \|\psi_{i^*}(\mathbf{z}^*) - \mathbf{x}_i\|_2$ and $\theta := -\frac{\mathbf{d}}{s^2}$. Observe that $|w_i(\mathbf{z})p_i(\mathbf{z})| \leq p_{max}$ for $\forall \mathbf{z} \in V_i$, then

$$e^{\theta^\top \mathbf{u}} \phi_i(\mathbf{z}^* \mid \mathbf{z}) \leq \frac{1}{(2\pi s^2)^{\frac{d}{2}}} e^{-\frac{\|\mathbf{u}\|^2}{2s^2} + \frac{\|\mathbf{d}\|\|\mathbf{u}\|}{s^2}} = \frac{1}{(2\pi s^2)^{\frac{d}{2}}} e^{-\frac{(\|\mathbf{u}\| - \|\mathbf{d}\|)^2}{2s^2}} e^{\frac{\|\mathbf{d}\|^2}{2s^2}}.$$

Then one term in the denominator becomes

$$\begin{aligned}
\left| (2\pi s^2)^{\frac{d}{2}} \int_{V_i} e^{\theta^\top \mathbf{u}} \phi_i(\mathbf{z}^* \mid \mathbf{z}) w_i(\mathbf{z}) p_i(\mathbf{z})\, d\mathbf{z} \right| &\leq p_{max} \int_{V_i} e^{\theta^\top \mathbf{u}} (2\pi s^2)^{\frac{d}{2}} \phi_i(\mathbf{z}^* \mid \mathbf{z})\, d\mathbf{z} \\
&\leq p_{max} e^{\frac{\|\mathbf{d}\|^2}{2s^2}} \int_{V_i} e^{-\frac{(\|\mathbf{u}\| - \|\mathbf{d}\|)^2}{2s^2}}\, d\mathbf{z} \\
&\leq p_{max} e^{\frac{\|\mathbf{d}\|^2}{2s^2}} \int_{\{\|\mathbf{u}\| \geq d_i - r_i\}} e^{-\frac{(\|\mathbf{u}\| - \|\mathbf{d}\|)^2}{2s^2}}\, d\mathbf{z} \qquad \text{(A.11)}
\end{aligned}$$

As $s < \frac{3(d_i - r_i)}{8\sqrt{\frac{D-1}{2}}}$ and by Lemma B.1, the above expression becomes

$$\text{r.h.s. of (A.11)} \leq p_{max} e^{\frac{\|\mathbf{d}\|^2}{2s^2}} S \int_{\{\|\mathbf{u}\| \geq d_i - r_i\}} e^{-\frac{(\|\mathbf{u}\| - \|\mathbf{d}\|)^2}{2s^2}}\, d\mathbf{u}$$

$$
\begin{aligned}
&\leq p_{max} e^{\frac{\|\mathbf{d}\|^2}{2s^2}} S \int_{\{\|\mathbf{u}\| \geq \frac{3(d_i - r_i)}{4}\}} e^{-\frac{\|\mathbf{u}\|^2}{2s^2}} \, d\mathbf{u} \\
&= p_{max} e^{\frac{\|\mathbf{d}\|^2}{2s^2}} S (2\pi s^2)^{D/2} \frac{\Gamma\left(\frac{D}{2}, \frac{9(d_i - r_i)^2}{32 s^2}\right)}{\Gamma\left(\frac{D}{2}\right)} \\
&\leq \frac{1}{\Gamma\left(\frac{D}{2}\right)} 2 p_{max} S (2\pi s^2)^{D/2} \left(\frac{9(d_i - r_i)^2}{32 s^2}\right)^{D/2-1} e^{-\frac{(d_i - r_i)^2}{4s^2}} \\
&= C_{11}(d) p_{max} S (d_i - r_i)^{D-d-2} s^{d+2}
\end{aligned}
$$

Thus, we deduce that

$$
\left| \sum_{i \neq i^*} \int_{V_i} e^{\theta^\top \mathbf{u}} \phi_i(\mathbf{z}^* \mid \mathbf{z}) w_i(\mathbf{z}) p_i(\mathbf{z}) \, d\mathbf{z} \right| \leq C_{12}(d) p_{max} S \left[\sum_{i \neq i^*} (d_i - r_i)^{D-d-2} \right] s^2.
$$

Similarly, we estimate the numerator

$$
\left\| \sum_{i \neq i^*} \int_{V_i} \mathbf{u} \, e^{\theta^\top \mathbf{u}} \phi_i(\mathbf{z}^* \mid \mathbf{z}) w_i(\mathbf{z}) p_i(\mathbf{z}) \, d\mathbf{z} \right\|_2 \leq C_{13}(d) p_{max} S \left[\sum_{i \neq i^*} (d_i - r_i)^{D-d-2} \right] s^3.
$$

We only need to refine (A.4). To be precise, denote

$$
\hat{s}_{\|}(\mathbf{x}, t) = \frac{\int_{V_{i^*}} \frac{1}{s^2} \left(\psi_{i^*}(\mathbf{z}) - \psi_{i^*}(\mathbf{z}^*)\right) \phi_{i^*}(\mathbf{z}^* \mid \mathbf{z}) w_{i^*}(\mathbf{z}) p_{i^*}(\mathbf{z}) \, d\mathbf{z}}{\int_{V_{i^*}} \phi_{i^*}(\mathbf{z}^* \mid \mathbf{z}) w_{i^*}(\mathbf{z}) p_{i^*}(\mathbf{z}) \, d\mathbf{z}},
$$

and

$$
p_t^{LD}(\mathbf{z}^*) = \int_{V_{i^*}} \phi_{i^*}(\mathbf{z}^* \mid \mathbf{z}) w_{i^*}(\mathbf{z}) p_{i^*}(\mathbf{z}) \, d\mathbf{z},
$$

(A.4) becomes

$$
\begin{aligned}
\left\| s_{\|}(\mathbf{x}, t) - \hat{s}_{\|}(\mathbf{x}, t) \right\|_2 \leq\ & C(d) \left(1 + \frac{1}{r^2} + \frac{1}{r^3}\right) \left(\frac{1}{P} + \frac{1}{P^2}\right) \left(\left(A + s^2\right) \left\| \hat{s}_{\|}(\mathbf{x}, t) \right\|_2 + \left(s + \frac{A^2}{s}\right) \right) \\
& + \frac{C_{12}(d) p_{max} S \left[\sum_{i \neq i^*} (d_i - r_i)^{D-d-2}\right] s^4 \left\| \hat{s}_{\|}(\mathbf{x}, t) \right\|_2 + C_{13}(d) p_{max} S \left[\sum_{i \neq i^*} (d_i - r_i)^{D-d-2}\right]}{s^2 p_t^{LD}(\mathbf{z}^*)} \\
=\ & C'(d) \left(1 + \frac{1}{r^2} + \frac{1}{r^3}\right) \left(\frac{1}{P} + \frac{1}{P^2}\right) \left(1 + S \left[\sum_{i \neq i^*} (d_i - r_i)^{D-d-2}\right]\right) \\
& \times \left(\left(A + s^2\right) \left\| \hat{s}_{\|}(\mathbf{x}, t) \right\|_2 + \left(s + \frac{A^2}{s}\right) \right)
\end{aligned}
$$

where $C'(d) = \max\{C(d), C_{12}(d), C_{13}(d)\}$. $\square$

A.3 Proofs of perturbed data manifold

We first establish a lemma quantifying the gap between the perturbed score and the manifold-supported score. This lemma then yields Corollary 3.2.

Lemma A.4. *The smoothed score function is stable with respect to perturbations in W_∞. For all $\mathbf{x} \in \mathbb{R}^D$, if $\epsilon < \frac{1}{2}(p_t(\mathbf{x}))^2$,*

$$\|\nabla \log \overline{p}_t(\mathbf{x}) - \nabla \log p_t(\mathbf{x})\|_2 \leq \frac{2\sqrt{2\pi}\epsilon}{s\,(p_t(\mathbf{x}))^2}\,(p_t(\mathbf{x}) + p_{max})\,, \tag{A.12}$$

where $s = \int_0^t \sigma^2(\tau)\,d\tau$ denotes the accumulated diffusion scale.

Proof of Lemma A.4. Using the identity for the difference of two normalized expectations, we write:

$$\begin{aligned}
&\|\nabla \log \overline{p}_t(\mathbf{x}) - \nabla \log p_t(\mathbf{x})\|_2 \\
&= \frac{1}{s^2}\left\| \frac{\int_{\mathbb{R}^D} (\mathbf{x}-\mathbf{y})\overline{p}(\mathbf{y})\phi_t(\mathbf{x}-\mathbf{y})d\mathbf{y}}{\int_{\mathbb{R}^D} \overline{p}(\mathbf{y})\phi_t(\mathbf{x}-\mathbf{y})d\mathbf{y}} - \frac{\int_{\mathbb{R}^D} (\mathbf{x}-\mathbf{y})p(\mathbf{y})\phi_t(\mathbf{x}-\mathbf{y})d\mathbf{y}}{\int_{\mathbb{R}^D} p(\mathbf{y})\phi_t(\mathbf{x}-\mathbf{y})d\mathbf{y}} \right\|_2 \\
&= \frac{1}{s^2 \int_{\mathbb{R}^D} \overline{p}(\mathbf{y})\phi_t(\mathbf{x}-\mathbf{y})d\mathbf{y} \int_{\mathbb{R}^D} p(\mathbf{y})\phi_t(\mathbf{x}-\mathbf{y})d\mathbf{y}} \\
&\quad \left(\left\| \left(\int_{\mathbb{R}^D} (\mathbf{x}-\mathbf{y})\left(\overline{p}(\mathbf{y}) - p(\mathbf{y})\right)\phi_t(\mathbf{x}-\mathbf{y})d\mathbf{y} \right) \int_{\mathbb{R}^D} p(\mathbf{y})\phi_t(\mathbf{x}-\mathbf{y})d\mathbf{y} \right\|_2 \right. \\
&\quad \left. + \left\| \left(\int_{\mathbb{R}^D} \left(\overline{p}(\mathbf{y}) - p(\mathbf{y})\right)\phi_t(\mathbf{x}-\mathbf{y})d\mathbf{y} \right) \int_{\mathbb{R}^D} (\mathbf{x}-\mathbf{y})p(\mathbf{y})\phi_t(\mathbf{x}-\mathbf{y})d\mathbf{y} \right\|_2 \right) \\
&\leq \frac{\epsilon}{s^2\left(\left(\int_{\mathbb{R}^D} p(\mathbf{y})\phi_t(\mathbf{x}-\mathbf{y})d\mathbf{y}\right)^2 - \epsilon \right)} \left(\left\| \int_{\mathbb{R}^D} \|\mathbf{x}-\mathbf{y}\|\phi_t(\mathbf{x}-\mathbf{y})d\mathbf{y} \int_{\mathbb{R}^D} p(\mathbf{y})\phi_t(\mathbf{x}-\mathbf{y})d\mathbf{y} \right\|_2 \right. \\
&\quad \left. + \left\| \int_{\mathbb{R}^D} \phi_t(\mathbf{x}-\mathbf{y})d\mathbf{y} \int_{\mathbb{R}^D} (\mathbf{x}-\mathbf{y})p(\mathbf{y})\phi_t(\mathbf{x}-\mathbf{y})d\mathbf{y} \right\|_2 \right) \\
&\leq \frac{2\sqrt{2\pi}\epsilon}{s\,(p_t(\mathbf{x}))^2}\,(p_t(\mathbf{x}) + p_{max})
\end{aligned}$$

which indicates that the constant only depends on p. □

Proof of Corollary 3.2. Using Lemma A.4 and Lemma 3.1, the result is straightforward by noticing that

$$\begin{aligned}
&\left\| \nabla \log p_t(\mathbf{x}) - \hat{\mathbf{s}}_{\|}(\mathbf{x}, t) - \frac{\psi(\mathbf{z}^*) - \mathbf{x}}{s^2} \right\|_2 \\
&\leq \|\nabla \log \overline{p}_t(\mathbf{x}) - \nabla \log p_t(\mathbf{x})\|_2 + \left\| \nabla \log \overline{p}_t(\mathbf{x}) - \hat{\mathbf{s}}_{\|}(\mathbf{x}, t) - \frac{\psi(\mathbf{z}^*) - \mathbf{x}}{s^2} \right\|_2 .
\end{aligned}$$

□

B Useful lemmas

Lemma B.1 (Inequality of incomplete Gamma function)**.** *Suppose $a > s - 1$ and $s > 0$, denote $\Gamma(s, a) = \int_a^{+\infty} x^{s-1}e^{-x}\,dx$ to be an incomplete Gamma function, we have*

$$a^{s-1}e^{-a} \leq \Gamma(s, a) \leq \frac{a^s e^{-a}}{a - s + 1}.$$

Proof of Lemma B.1. The first inequality is due to

$$\int_a^{+\infty} x^{s-1}e^{-x}\,dx \geq a^{s-1}\int_a^{+\infty} e^{-x}\,dx = a^{s-1}e^{-a}.$$

and the second is due to

$$\int_a^{+\infty} x^{s-1}e^{-x}\,dx = a^{s-1}e^{-a}\int_0^{+\infty}\left(\frac{x}{a}+1\right)^{s-1}e^{-x}\,dx \leq a^{s-1}e^{-a}\int_0^{+\infty}\exp\left(-x+\frac{x(s-1)}{a}\right)\,dx = \frac{a^s e^{-a}}{a-s+1},$$

where the inequality is because

$$\left(\frac{x}{a}+1\right)^{s-1} = \left(\frac{x}{a}+1\right)^{\frac{a}{x}\cdot\frac{x(s-1)}{a}} \leq e^{\frac{x(s-1)}{a}}.$$

$\square$

Lemma B.2. *Suppose $\psi:\mathbb{R}^m\to\mathbb{R}^n$ is continuously differentiable, and its Jacobian is L-Lipschitz continuous in $B_\delta(\mathbf{z})$, the quadratic upper bound inequality holds for $\forall\mathbf{z}'\in B_\delta(\mathbf{z})$:*

$$\left\|\psi(\mathbf{z})-\psi(\mathbf{z}')-J_\psi(\mathbf{z}')\left(\mathbf{z}-\mathbf{z}'\right)\right\|_2 \leq \frac{L}{2}\|\mathbf{z}-\mathbf{z}'\|_2^2.$$

Proof of Lemma B.2. Since the Lipschitz property of Jacobian, we observe that for $\|\mathbf{z}-\mathbf{z}'\|\leq\delta$,

$$\begin{aligned}
&\left\|\psi(\mathbf{z})-\psi(\mathbf{z}')-J_\psi(\mathbf{z}')\left(\mathbf{z}-\mathbf{z}'\right)\right\|_2\\
=&\left\|\int_0^1\left(J_\psi\left(t\mathbf{z}+(1-t)\mathbf{z}'\right)-J_\psi(\mathbf{z}')\right)\left(\mathbf{z}-\mathbf{z}'\right)\,dt\right\|_2\\
\leq&\int_0^1\left\|J_\psi\left(t\mathbf{z}+(1-t)\mathbf{z}'\right)-J_\psi(\mathbf{z}')\right\|_2\|\mathbf{z}-\mathbf{z}'\|_2\,dt\\
\leq&\int_0^1 Lt\|\mathbf{z}-\mathbf{z}'\|_2^2\,dt=\frac{L}{2}\|\mathbf{z}-\mathbf{z}'\|_2^2
\end{aligned}$$

$\square$

Lemma B.3 (Stirling-type bounds for Gamma function)**.** *For any $x>0$, we have*

$$\Gamma(x) > \sqrt{2\pi}x^{x-\frac{1}{2}}e^{-x}.$$