Gaussian Interpolation Flows

Yuan Gao   YUAN0.GAO@CONNECT.POLYU.HK
Department of Applied Mathematics
The Hong Kong Polytechnic University
Hong Kong SAR, China

Jian Huang   J.HUANG@POLYU.EDU.HK
Departments of Data Science and AI, and Applied Mathematics
The Hong Kong Polytechnic University
Hong Kong SAR, China

Yuling Jiao   YULINGJIAOMATH@WHU.EDU.CN
School of Mathematics and Statistics
and Hubei Key Laboratory of Computational Science
Wuhan University, Wuhan, China

Editor: Mingyuan Zhou

Abstract

Gaussian denoising has emerged as a powerful method for constructing simulation-free continuous normalizing flows for generative modeling. Despite their empirical successes, theoretical properties of these flows and the regularizing effect of Gaussian denoising have remained largely unexplored. In this work, we aim to address this gap by investigating the well-posedness of simulation-free continuous normalizing flows built on Gaussian denoising. Through a unified framework termed Gaussian interpolation flow, we establish the Lipschitz regularity of the flow velocity field, the existence and uniqueness of the flow, and the Lipschitz continuity of the flow map and the time-reversed flow map for several rich classes of target distributions. This analysis also sheds light on the auto-encoding and cycle consistency properties of Gaussian interpolation flows. Additionally, we study the stability of these flows in source distributions and perturbations of the velocity field, using the quadratic Wasserstein distance as a metric. Our findings offer valuable insights into the learning techniques employed in Gaussian interpolation flows for generative modeling, providing a solid theoretical foundation for end-to-end error analyses of learning Gaussian interpolation flows with empirical observations.

Keywords: Continuous normalizing flows, Gaussian denoising, generative modeling, Lipschitz transport maps, stochastic interpolation.

1. Introduction

Generative modeling, which aims to learn the underlying data generating distribution from a finite sample, is a fundamental task in the field of machine learning and statistics (Salakhutdinov, 2015). Deep generative models (DGMs) find wide-ranging applications across diverse domains such as computer vision, natural language processing, drug discovery, and recom-

mendation systems. The core objective of DGMs is to learn a nonlinear mapping, either deterministic or stochastic (with outsourcing randomness), which transforms latent samples drawn from a simple reference distribution into samples that closely resemble the target distribution.

Generative adversarial networks (GANs) have emerged as a prominent class of DGMs (Goodfellow et al., 2014; Arjovsky et al., 2017; Goodfellow et al., 2020). Through an adversarial training process, GANs learn to approximately generate samples from the data distribution. Variational auto-encoders (VAEs) are another category of DGMs (Kingma and Welling, 2014; Rezende et al., 2014; Kingma and Welling, 2019). In VAEs, the encoding and decoding procedures produce a compressed and structured latent representation, enabling efficient sampling and interpolation. Score-based diffusion models are a promising approach to deep generative modeling that has evolved rapidly since its emergence (Song and Ermon, 2019, 2020; Ho et al., 2020; Song et al., 2021b,a). The basis of score-based diffusion models lies in the notion of the score function, which characterizes the gradient of the log-density function of a given distribution.

In addition, normalizing flows have gained attention as another powerful class of DGMs (Tabak and Vanden-Eijnden, 2010; Tabak and Turner, 2013; Kobyzev et al., 2020; Papamakarios et al., 2021). In normalizing flows, an invertible mapping is learned to transform a simple source distribution into a more complex target distribution by a composition of a series of parameterized, invertible and differentiable intermediate transformations. This framework allows for efficient sampling and training by maximum likelihood estimation (Dinh et al., 2014; Rezende and Mohamed, 2015). Continuous normalizing flows (CNFs) pursue this idea further by performing the transformation over continuous time, enabling fine-grained modeling of dynamic systems from the source distribution to the target distribution. The essence of CNFs lies in defining ordinary differential equations (ODEs) that govern the evolution of CNFs in terms of continuous trajectories. Inspired by the Gaussian denoising approach, which learns a target distribution by denoising its Gaussian smoothed counterpart, many authors have considered simulation-free estimation methods that have shown great potential in large-scale applications (Song et al., 2021a; Liu et al., 2023; Albergo and Vanden-Eijnden, 2023; Lipman et al., 2023; Neklyudov et al., 2023; Tong et al., 2023; Chen and Lipman, 2023; Albergo et al., 2023b; Shaul et al., 2023; Pooladian et al., 2023). However, despite the empirical success of simulation-free CNFs based on Gaussian denoising, a rigorous theoretical analysis of these CNFs has received limited attention thus far.

In this work, we explore an ODE flow-based approach for generative modeling, which we refer to as Gaussian Interpolation Flows (GIFs). This method is derived from the Gaussian stochastic interpolation detailed in Section 3. GIFs represent a straightforward extension of the stochastic interpolation method (Albergo and Vanden-Eijnden, 2023; Liu et al., 2023; Lipman et al., 2023). They can be considered a class of CNFs and encompass various ODE flows as special cases. According to the classical Cauchy-Lipschitz theorem, also known as the Picard-Lindelöf theorem (Hartman, 2002b, Theorem 1.1), a unique solution to the initial value problem for an ODE flow exists if the velocity field is continuous in the time variable and uniformly Lipschitz continuous in the space variable. In the case of GIFs, the velocity field depends on the score function of the push-forward measure. Therefore, it remains to be shown that this velocity field satisfies the regularity conditions stipulated

by the Cauchy-Lipschitz theorem. These regularity conditions are commonly assumed in the literature when analyzing the convergence properties of CNFs or general neural ODEs (Chen et al., 2018; Biloš et al., 2021; Marion et al., 2023; Marion, 2023; Marzouk et al., 2023). However, there is a theoretical gap in understanding how to translate these regularity conditions on velocity fields into conditions on target distributions.

The main focus of this work is to study and establish the theoretical properties of Gaussian interpolation flow and its corresponding flow map. We show that the regularity conditions of the Cauchy-Lipschitz theorem are satisfied for several rich classes of probability distributions using variance inequalities. Based on the obtained regularity results, we further expose the well-posedness of GIFs, the Lipschitz continuity of flow mappings, and applications to generative modeling. The well-posedness results are crucial for studying the approximation and convergence properties of GIFs learned with the flow or score matching method. When applied to generative modeling, our results further elucidate the auto-encoding and cycle consistency properties exhibited by GIFs.
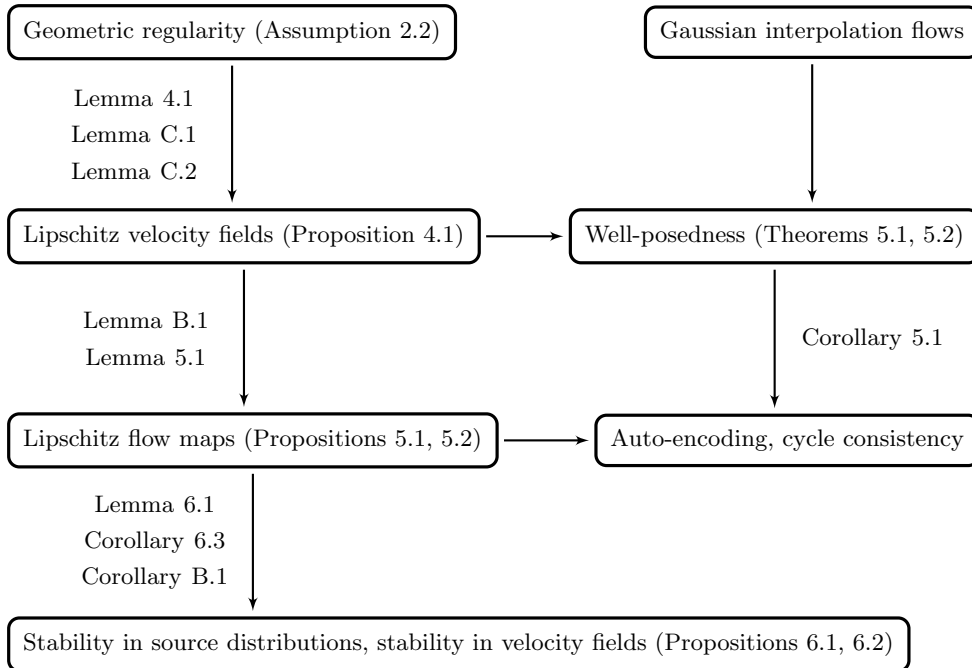

Figure 1: Roadmap of the main results.

1.1 Our main contributions

We provide an overview of the main results in Figure 1, in which we indicate the assumptions used in our analysis and the relationship between the results. We also summarize our main contributions below.

- In Section 3, we extend the framework of stochastic interpolation proposed in Albergo and Vanden-Eijnden (2023). Various ODE flows can be considered special cases of the extended framework. We prove that the marginal distributions of GIFs satisfy

the continuity equation converging to the target distribution in the weak sense. Several explicit formulas of the velocity field and its derivatives are derived, which can facilitate computation and regularity estimation.

- In Sections 4 and 5, we establish the spatial Lipschitz regularity of the velocity field for a range of target measures with rich structures, which is sufficient to guarantee the well-posedness of GIFs. Additionally, we deduce the Lipschitz regularity of both the flow map and its time-reversed counterpart. The well-posedness of GIFs is an essential attribute, serving as a foundational requirement for investigating numerical solutions of GIFs. It is important to note that while the flow maps are demonstrated to be Lipschitz continuous transport maps for generative modeling, the Lipschitz regularity for optimal transport maps has only been partially established to date.

- In Section 6, we show that the auto-encoding and cycle consistency properties of GIFs are inherently satisfied when the flow maps exhibit Lipschitz continuity with respect to the spatial variable. This demonstrates that exact auto-encoding and cycle consistency are intrinsic characteristics of GIFs. Our findings lend theoretical support to the findings made by Su et al. (2023), as illustrated in Figures 3 and 4.

- In Section 6, we conduct the stability analysis of GIFs, examining how they respond to changes in source distributions and to perturbations in the velocity field. This analysis, conducted in terms of the quadratic Wasserstein distance, provides valuable insights that justify the use of learning techniques such as Gaussian initialization and flow or score matching.

2. Preliminaries

In this section, we include several preliminary setups to show notations, basic assumptions, and several useful variance inequalities.

Notation. Here we summarize the notation. The space $\mathbb{R}^d$ is endowed with the Euclidean metric and we denote by $\|\cdot\|$ and $\langle\cdot,\cdot\rangle$ the corresponding norm and inner product. Let $\mathbb{S}^{d-1} := \{x \in \mathbb{R}^d : \|x\| = 1\}$. For a matrix $A \in \mathbb{R}^{k\times d}$, we use $A^\top$ for the transpose, and the spectral norm is denoted by $\|A\|_{2,2} := \sup_{x\in\mathbb{S}^{d-1}} \|Ax\|$. For a square matrix $A \in \mathbb{R}^{d\times d}$, we use $\det(A)$ for the determinant and $\mathrm{Tr}(A)$ for the trace. We use $\mathbf{I}_d$ to denote the $d \times d$ identity matrix. For two symmetric matrices $A, B \in \mathbb{R}^{d\times d}$, we denote $A \succeq B$ or $B \preceq A$ if $A - B$ is positive semi-definite. For two vectors $x, y \in \mathbb{R}^d$, we denote $x \otimes y := xy^\top$. For $\Omega_1 \subset \mathbb{R}^k, \Omega_2 \subset \mathbb{R}^d, n \geq 1$, we denote by $C^n(\Omega_1; \Omega_2)$ the space of continuous functions $f : \Omega_1 \to \Omega_2$ that are n times differentiable and whose partial derivatives of order n are continuous. If $\Omega_2 \subset \mathbb{R}$, we simply write $C^n(\Omega_1)$. For any $f(x) \in C^2(\mathbb{R}^d)$, let $\nabla_x f, \nabla_x^2 f, \nabla_x \cdot f$, and $\Delta_x f$ denote its gradient, Hessian, divergence, and Laplacian, respectively. We use $X \lesssim Y$ to denote $X \leq CY$ for some constant $C > 0$. The function composition operation is marked as $g \circ f := g(f(x))$ for functions f and g.

The Borel σ-algebra of $\mathbb{R}^d$ is denoted by $\mathcal{B}(\mathbb{R}^d)$. The space of probability measures defined on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ is denoted as $\mathcal{P}(\mathbb{R}^d)$. For any $\mathbb{R}^d$-valued random variable X, we use $\mathbb{E}[\mathsf{X}]$ and $\mathrm{Cov}(\mathsf{X})$ to denote its expectation and covariance matrix, respectively. We use $\mu * \nu$ to denote the convolution for any two probability measures μ and ν, and we use $\overset{d}{=}$ to indicate

two random variables have the same probability distribution. For a random variable X, let $\mathrm{Law}(\mathsf{X})$ denote its probability distribution. Let $g : \mathbb{R}^k \to \mathbb{R}^d$ be a measurable mapping and μ be a probability measure on $\mathbb{R}^k$. The push-forward measure $f_\# \mu$ of a measurable set A is defined as $f_\# \mu := \mu(f^{-1}(A))$. Let $N(m, \Sigma)$ denote a d-dimensional Gaussian random variable with mean vector $m \in \mathbb{R}^d$ and covariance matrix $\Sigma \in \mathbb{R}^{d \times d}$. For simplicity, let $\gamma_{d,\sigma^2} := N(0, \sigma^2 \mathbf{I}_d)$, and let $\varphi_{m,\sigma^2}(x)$ denote the probability density function of $N(m, \sigma^2 \mathbf{I}_d)$ with respect to the Lebesgue measure. If $m = 0, \sigma = 1$, we abbreviate these as γ_d and $\varphi(x)$. Let $L^p(\mathbb{R}^d; \mathbb{R}^\ell, \mu)$ denote the L^p space with the L^p norm for $p \in [1, \infty]$ w.r.t. a measure μ. To simplify the notation, we write $L^p(\mathbb{R}^d, \mu)$ if $\ell = 1$, $L^p(\mathbb{R}^d; \mathbb{R}^\ell)$ if the Lebesgue measure is used, and $L^p(\mathbb{R}^d)$ if both hold.

2.1 Assumptions

We focus on the probability distributions satisfying several types of assumptions of weak convexity, which offer a geometric notion of regularity that is dimension-free in the study of high-dimensional distributions (Klartag, 2010). On the one hand, weak-convexity regularity conditions are useful in deriving dimension-free guarantees for generative modeling and sampling from high-dimensional distributions. On the other hand, they accommodate distributions with complex shapes, including those with multiple modes.

Definition 2.1 (Cattiaux and Guillin, 2014) *A probability measure $\mu(\mathrm{d}x) = \exp(-U)\mathrm{d}x$ is κ-semi-log-concave for some $\kappa \in \mathbb{R}$ if its support $\Omega \subseteq \mathbb{R}^d$ is convex and its potential function $U \in C^2(\Omega)$ satisfies*

$$\nabla_x^2 U(x) \succeq \kappa \mathbf{I}_d, \quad \forall x \in \Omega.$$

The κ-semi-log-concavity condition is a relaxed notion of log-concavity, since here $\kappa < 0$ is allowed. When $\kappa \geq 0$, we are considering a log-concave probability measure that is proved to be unimodal (Saumard and Wellner, 2014). However, when $\kappa < 0$, a κ-semi-log-concave probability measure can be multimodal.

Definition 2.2 (Eldan and Lee, 2018) *A probability measure $\mu(\mathrm{d}x) = \exp(-U)\mathrm{d}x$ is β-semi-log-convex for some $\beta > 0$ if its support $\Omega \subseteq \mathbb{R}^d$ is convex and its potential function $U \in C^2(\Omega)$ satisfies*

$$\nabla_x^2 U(x) \preceq \beta \mathbf{I}_d, \quad \forall x \in \Omega.$$

The following definition of L-log-Lipschitz continuity is a variant of L-Lipschitz continuity. It characterizes a first-order condition on the target function rather than a second-order condition such as κ-semi-log-concavity and β-semi-log-convexity in Definitions 2.1 and 2.2.

Definition 2.3 *A function $f : \mathbb{R}^d \to \mathbb{R}_+$ is L-log-Lipschitz continuous if its logarithm is L-Lipschitz continuous for some $L \geq 0$.*

Based on the definitions, we present two assumptions on the target distribution. Assumption 2.1 concerns the absolute continuity and the moment condition. Assumption 2.2 imposes geometric regularity conditions.

Assumption 2.1 *The probability measure ν is absolutely continuous with respect to the Lebesgue measure and has a finite second moment.*

Assumption 2.2 *Let $D := (1/\sqrt{2})\mathrm{diam}(\mathrm{supp}(\nu))$. The probability measure ν satisfies one or more of the following conditions:*

    (i) ν is β-semi-log-convex for some $\beta > 0$ and κ-semi-log-concave for some $\kappa > 0$ with $\mathrm{supp}(\nu) = \mathbb{R}^d$;

   (ii) ν is κ-semi-log-concave for some $\kappa \in \mathbb{R}$ with $D \in (0, \infty)$;

  (iii) $\nu = \gamma_{d,\sigma^2} * \rho$ where ρ is a probability measure supported on a Euclidean ball of radius R on $\mathbb{R}^d$;

  (iv) ν is β-semi-log-convex for some $\beta > 0$, κ-semi-log-concave for some $\kappa \le 0$, and $\frac{\mathrm{d}\nu}{\mathrm{d}\gamma_d}(x)$ is L-log-Lipschitz in x for some $L \ge 0$ with $\mathrm{supp}(\nu) = \mathbb{R}^d$.

Multimodal distributions. Assumption 2.2 enumerates scenarios where probability distributions are endowed with geometric regularity. We examine the scenarios and clarify whether they cover multimodal distributions. Scenario (i) is referred to as the classical strong log-concavity case ($\kappa > 0$), and thus, describes unimodal distributions. Scenario (ii) allows $\kappa \le 0$ and requires that the support is bounded. Mixtures of Gaussian distributions are considered in Scenario (iii), and typically are multimodal distributions. Scenario (iv) also allows $\kappa \le 0$ when considering a log-Lipschitz perturbation of the standard Gaussian distribution. Both Scenario (ii) and Scenario (iv) incorporate multimodal distributions due to the potential negative lower bound κ.

Lipschitz score. Lipschitz continuity of the score function is a basic regularity assumption on target distributions in the study of sampling algorithms based on Langevin and Hamiltonian dynamics. Even for high-dimensional distributions, this assumption endows a great source of regularity. For an L-Lipschitz score function, its corresponding distribution is both L-semi-log-convex and $(-L)$-semi-log-concave for some $L \ge 0$.

2.2 Variance inequalities

Variance inequalities like the Brascamp-Lieb inequality and the Cramér-Rao inequality are fundamental inequalities for explaining the regularizing effect of Gaussian denoising. Combined with κ-semi-log-concavity and β-semi-log-convexity, these inequalities are crucial for deducing the Lipschitz regularity of the velocity fields of GIFs in Proposition 4.1-(b) and (c).

Lemma 2.1 (Brascamp-Lieb inequality) *Let $\mu(\mathrm{d}x) = \exp(-U(x))\mathrm{d}x$ be a probability measure on a convex set $\Omega \subseteq \mathbb{R}^d$ whose potential function $U : \Omega \to \mathbb{R}$ is of class C^2 and strictly convex. Then for every locally Lipschitz function $f \in L^2(\Omega, \mu)$,*

$$\mathrm{Var}_\mu(f) \le \mathbb{E}_\mu\left[\langle \nabla_x f, (\nabla_x^2 U)^{-1} \nabla_x f\rangle\right]. \tag{2.1}$$

When applied to functions of the form $f : x \mapsto \langle x, e\rangle$ for any $e \in \mathbb{S}^{d-1}$, the Brascamp-Lieb inequality yields an upper bound of the covariance matrix

$$\mathrm{Cov}_\mu(\mathsf{X}) \preceq \mathbb{E}_\mu\left[(\nabla_x^2 U(x))^{-1}\right] \tag{2.2}$$

with equality if $\mathsf{X} \sim N(m, \Sigma)$ with Σ positive definite.

Under the strong log-concavity condition, that is, μ is κ-semi-log-concave with $\kappa > 0$, and if the Euclidean Bakry-Émery criterion is satisfied (Bakry and Émery, 1985), the Brascamp-Lieb inequality instantly recovers the Poincaré inequality (see Definition G.2).

The Brascamp-Lieb inequality originally appeared in (Brascamp and Lieb, 1976, Theorem 4.1). Alternative proofs are provided in Bobkov and Ledoux (2000); Bakry et al. (2014); Cordero-Erausquin (2017). The dimension-free inequality (2.1) can be further strengthened to obtain several variants with dimensional improvement.

Lemma 2.2 (Cramér-Rao inequality) *Let $\mu(\mathrm{d}x) = \exp(-U(x))\mathrm{d}x$ be a probability measure on $\mathbb{R}^d$ whose potential function $U : \mathbb{R}^d \to \mathbb{R}$ is of class C^2. Then for every $f \in C^1(\mathbb{R}^d)$,*

$$\mathrm{Var}_\mu(f) \geq \langle \mathbb{E}_\mu[\nabla_x f], \left(\mathbb{E}_\mu[\nabla_x^2 U]\right)^{-1} \mathbb{E}_\mu[\nabla_x f]\rangle. \tag{2.3}$$

When applied to functions of the form $f : x \mapsto \langle x, e \rangle$ for any $e \in \mathbb{S}^{d-1}$, the Cramér-Rao inequality yields a lower bound of the covariance matrix

$$\mathrm{Cov}_\mu(\mathsf{X}) \succeq \left(\mathbb{E}_\mu[\nabla_x^2 U(x)]\right)^{-1} \tag{2.4}$$

with equality as well if $\mathsf{X} \sim N(m, \Sigma)$ with Σ positive definite.

The Cramér-Rao inequality plays a central role in asymptotic statistics as well as in information theory. The inequality (2.4) has an alternative derivation from the Cramér-Rao bound for the location parameter. For detailed proofs of the Cramér-Rao inequality, readers are referred to Chewi and Pooladian (2022); Dai et al. (2023), and the references therein.

3. Gaussian interpolation flows

Simulation-free CNFs represent a potent class of generative models based on ODE flows. Albergo and Vanden-Eijnden (2023) and Albergo et al. (2023b) introduce an innovative CNF that is constructed using stochastic interpolation techniques, such as Gaussian denoising. They conduct a thorough investigation of this flow, particularly examining its applications and effectiveness in generative modeling.

We study the ODE flow and its associated flow map as defined by the Gaussian denoising process. This process has been explored from various perspectives, including diffusion models and stochastic interpolants. Building upon the work of Albergo and Vanden-Eijnden (2023) and Albergo et al. (2023b), we expand the stochastic interpolant framework by relaxing certain conditions on the functions a_t and b_t, offering a more comprehensive perspective on the Gaussian denoising process.

In our generalization, we introduce an adaptive starting point to the stochastic interpolation framework, which allows for greater flexibility in the modeling process. By examining this modified framework, we aim to demonstrate that the Gaussian denoising principle is effectively implemented within the context of stochastic interpolation.

Definition 3.1 (Vector interpolation) *Let $z \in \mathbb{R}^d$, $x_1 \in \mathbb{R}^d$ be two vectors in the Euclidean space and let $x_0 := a_0 z + b_0 x_1$ with $a_0 > 0, b_0 \geq 0$. Then we construct an interpolant*

between x_0 and x_1 over time $t \in [0, 1]$ through $I_t(x_0, x_1)$, defined by

$$I_t(x_0, x_1) = a_t z + b_t x_1, \tag{3.1}$$

where a_t, b_t satisfy

$$\begin{aligned}
&\dot{a}_t \leq 0, \quad \dot{b}_t \geq 0, \quad a_0 > 0, \quad b_0 \geq 0, \quad a_1 = 0, \quad b_1 = 1, \\
&a_t > 0 \; \text{ for any } t \in (0, 1), \quad b_t > 0 \; \text{ for any } t \in (0, 1), \\
&a_t, b_t \in C^2([0, 1)), \quad a_t^2 \in C^1([0, 1]), \quad b_t \in C^1([0, 1]).
\end{aligned} \tag{3.2}$$

Remark 3.1 *Compared with the vector interpolant defined by Albergo and Vanden-Eijnden (2023) (a.k.a. one-sided interpolant in Albergo et al. (2023b)), we extend its definition by relaxing the requirements that $a_0 = 1, b_0 = 0$ with $a_0 > 0, b_0 \geq 0$. This consideration is largely motivated by analyzing the probability flow ODEs of the variance-exploding (VE) SDE and the variance-preserving (VP) SDE (Song et al., 2021b). We illustrate examples of interpolants incorporated by Definition 3.1 in Table 1.*

Remark 3.2 *We have eased the smoothness conditions for the functions a_t and b_t required in Albergo and Vanden-Eijnden (2023). Specifically, we consider the case where $a_t, b_t \in C^2([0, 1))$, $a_t^2 \in C^1([0, 1])$, and $b_t \in C^1([0, 1])$. This relaxation enables us to include the Föllmer flow into our framework, characterized by $a_t = \sqrt{1 - t^2}$ and $b_t = t$. It is evident that $a_t = \sqrt{1 - t^2}$ does not fulfill the condition $a_t \in C^2([0, 1])$, but it does meet the requirements $a_t \in C^2([0, 1))$ and $a_t^2 \in C^1([0, 1])$.*

Remark 3.3 *The C^2 regularity of a_t, b_t is necessary to derive the regularity of the velocity field $v(t, x)$ in Eq. (3.5) concerning the time variable t. In addition, the C^1 regularity of a_t^2, b_t is sufficient to ensure the Lipschitz regularity of the velocity field $v(t, x)$ in Eq. (3.5) concerning the space variable x.*

A natural generalization of the vector interpolant (3.1) is to construct a set interpolant between two convex sets through Minkowski sum, which is common in convex geometry. A set interpolant stimulates the construction of a measure interpolant between a structured source measure and a target measure.

As noted, we can construct a measure interpolation using a Gaussian convolution path. The measure interpolation is particularly relevant to Gaussian denoising and Gaussian channels in information theory as elucidated in Remark 3.8. Because of this connection with Gaussian denoising, we call the measure interpolation a Gaussian stochastic interpolation. The Gaussian stochastic interpolation can be understood as a collection of linear combinations of a standard Gaussian random variable and the target random variable. The coefficients of the linear combinations vary with time $t \in [0, 1]$ as shown in Definition 3.1. Later in this section, we will show this Gaussian stochastic interpolation can be transformed into a deterministic ODE flow.

Gaussian stochastic interpolation has been investigated from several perspectives in the literature. The rectified flow has been proposed in Liu et al. (2023), and its theoretical connection with optimal transport has been investigated in Liu (2022). The formulation of the rectified flow is to learn the ODE flow defined by stochastic interpolation with linear

time coefficients. In Appendix C of Liu et al. (2023), there is a nonlinear extension of the rectified flow in which the linear coefficients are replaced by general nonlinear coefficients. Albergo et al. (2023b) extends the stochastic interpolant framework proposed in (Albergo and Vanden-Eijnden, 2023) by considering a linear combination among three random variables. In Section 3 of Albergo et al. (2023b), the original stochastic interpolant framework is recovered as a one-sided interpolant between the Gaussian distribution and the target distribution. Moreover, Lipman et al. (2023) propose a flow matching method which directly learns a Gaussian conditional probability path with a neural ODE. In Section 4.1 of Lipman et al. (2023), the velocity fields of the variance exploding and variance preserving probability flows are shown as special instances of the flow matching framework. We summarize these formulations as Gaussian stochastic interpolation by slightly extending the original stochastic interpolant framework.

Type	VE	VP	Linear	Föllmer	Trigonometric
a_t	α_t	α_t	$1 - t$	$\sqrt{1 - t^2}$	$\cos(\frac{\pi}{2} t)$
b_t	1	$\sqrt{1 - \alpha_t^2}$	t	t	$\sin(\frac{\pi}{2} t)$
a_0	α_0	α_0	1	1	1
b_0	1	$\sqrt{1 - \alpha_0^2}$	0	0	0
Source	Convolution	Convolution	γ_d	γ_d	γ_d

Table 1: Summary of various measure interpolants including VE interpolant (Song et al., 2021b), VP interpolant (Song et al., 2021b), linear interpolant (Liu et al., 2023), Föllmer interpolant (Dai et al., 2023), and trigonometric interpolant (Albergo and Vanden-Eijnden, 2023). There are two types of source measures including a standard Gaussian distribution γ_d and a convoluted distribution consisting of the target distribution and γ_d.

Definition 3.2 (Measure interpolation) *Let $\mu = \mathrm{Law}(\mathsf{X}_0)$ and $\nu = \mathrm{Law}(\mathsf{X}_1)$ be two probability measures satisfying $\mathsf{X}_0 = a_0 \mathsf{Z} + b_0 \mathsf{X}_1$ where $\mathsf{Z} \sim \gamma_d := N(0, \mathbf{I}_d)$ is independent from X_1. We call $(\mathsf{X}_t)_{t \in [0,1]}$ a Gaussian stochastic interpolation from the source measure μ to the target measure ν, which is defined through I_t over time interval $[0, 1]$ as follows*

$$\mathsf{X}_t = I_t(\mathsf{X}_0, \mathsf{X}_1), \quad \mathsf{X}_0 = a_0 \mathsf{Z} + b_0 \mathsf{X}_1, \quad \mathsf{Z} \sim \gamma_d, \quad \mathsf{X}_1 \sim \nu. \tag{3.3}$$

Remark 3.4 *It is obvious that the marginal distribution of X_t satisfies $\mathsf{X}_t \stackrel{d}{=} a_t \mathsf{Z} + b_t \mathsf{X}_1$ with $\mathsf{Z} \sim \gamma_d, \mathsf{X}_1 \sim \nu$.*

Motivated by the time-varying properties of the Gaussian stochastic interpolation, we derive that its marginal flow satisfies the continuity equation. This result characterizes the dynamics of the marginal density flow of the Gaussian stochastic interpolation.

Theorem 3.1 *Suppose that Assumption 2.1 holds. Then the marginal flow $(p_t)_{t \in [0,1]}$ of the Gaussian stochastic interpolation $(\mathsf{X}_t)_{t \in [0,1]}$ between μ and ν satisfies the continuity equation*

$$\partial_t p_t + \nabla_x \cdot (p_t v(t,x)) = 0, \quad (t,x) \in [0,1] \times \mathbb{R}^d, \quad p_0(x) = \tfrac{\mathrm{d}\mu}{\mathrm{d}x}(x), \quad p_1(x) = \tfrac{\mathrm{d}\nu}{\mathrm{d}x}(x) \quad (3.4)$$

in the weak sense with the velocity field

$$v(t,x) := \mathbb{E}[\dot{a}_t \mathsf{Z} + \dot{b}_t \mathsf{X}_1 | \mathsf{X}_t = x], \quad t \in (0,1), \tag{3.5}$$

$$v(0,x) := \lim_{t \downarrow 0} v(t,x), \qquad v(1,x) := \lim_{t \uparrow 1} v(t,x). \tag{3.6}$$

Remark 3.5 *We notice that $x = a_t \mathbb{E}[\mathsf{Z}|\mathsf{X}_t = x] + b_t \mathbb{E}[\mathsf{X}_1|\mathsf{X}_t = x]$ due to Eq. (3.3). Then it holds that*

$$v(t,x) = \tfrac{\dot{a}_t}{a_t} x + \left(\dot{b}_t - \tfrac{\dot{a}_t}{a_t} b_t \right) \mathbb{E}[\mathsf{X}_1 | \mathsf{X}_t = x], \quad t \in (0,1). \tag{3.7}$$

We also notice that, according to Tweedie's formula (cf. Lemma G.1 in the appendix), it holds that

$$s(t,x) = \tfrac{b_t}{a_t^2} \mathbb{E}[\mathsf{X}_1 | \mathsf{X}_t = x] - \tfrac{1}{a_t^2} x, \quad t \in (0,1), \tag{3.8}$$

where $s(t,x)$ is the score function of the marginal distribution of $\mathsf{X}_t \sim p_t$.

Combining (3.7) and (3.8), it follows that the velocity field is a gradient field and its nonlinear term is the score function $s(t,x)$, namely, for any $t \in (0,1)$,

$$v(t,x) = \tfrac{\dot{b}_t}{b_t} x + \left(\tfrac{\dot{b}_t}{b_t} a_t^2 - \dot{a}_t a_t \right) s(t,x). \tag{3.9}$$

Remark 3.6 *A relevant result has been provided in the proof of (Albergo and Vanden-Eijnden, 2023, Proposition 4) in a restricted case that $a_0 = 1, b_0 = 0$. In this case, if $\dot{a}_0, \dot{a}_1, \dot{b}_0, \dot{b}_1$ are well-defined, the velocity field reads*

$$v(0,x) = \dot{a}_0 x + \dot{b}_0 \mathbb{E}_\nu[\mathsf{X}_1], \quad v(1,x) = \dot{b}_1 x + \dot{a}_1 \mathbb{E}_{\gamma_d}[\mathsf{Z}]$$

at time 0 and 1. Otherwise, if any one of $\dot{a}_0, \dot{a}_1, \dot{b}_0, \dot{b}_1$ is not well-defined, the velocity field $v(0,x)$ or $v(1,x)$ should be considered on a case-by-case basis. In addition, we provide an alternative viewpoint of the relationship between the velocity field associated with stochastic interpolation and the score function of its marginal flow using Tweedie's formula in Lemma G.1.

Remark 3.7 (Diffusion process) *The marginal flow of the Gaussian stochastic interpolation (3.3) coincides with the time-reversed marginal flow of a diffusion process $(\overline{X}_t)_{t \in [0,1)}$ (Albergo et al., 2023b, Theorem 3.5) defined by*

$$\mathrm{d}\overline{X}_t = -\tfrac{\dot{b}_{1-t}}{b_{1-t}} \overline{X}_t + \sqrt{2 \left(\tfrac{\dot{b}_{1-t}}{b_{1-t}} a_{1-t}^2 - \dot{a}_{1-t} a_{1-t} \right)} \mathrm{d}\overline{W}_t.$$

Remark 3.8 (Gaussian denoising) *The Gaussian stochastic interpolation has an information-theoretic interpretation as a time-varying Gaussian channel. Here a_t^2 and b_t^2/a_t^2 stand for the noise level and signal-to-noise ratio (SNR) for time $t \in [0,1]$, respectively. As time $t \to 1$, we are approaching the high-SNR regime, that is, the SNR b_t^2/a_t^2 grows to ∞. Moreover, the SNR b_t^2/a_t^2 is monotonically increasing in time t over $[0,1]$. The Gaussian noise level gets reduced through this Gaussian denoising process.*
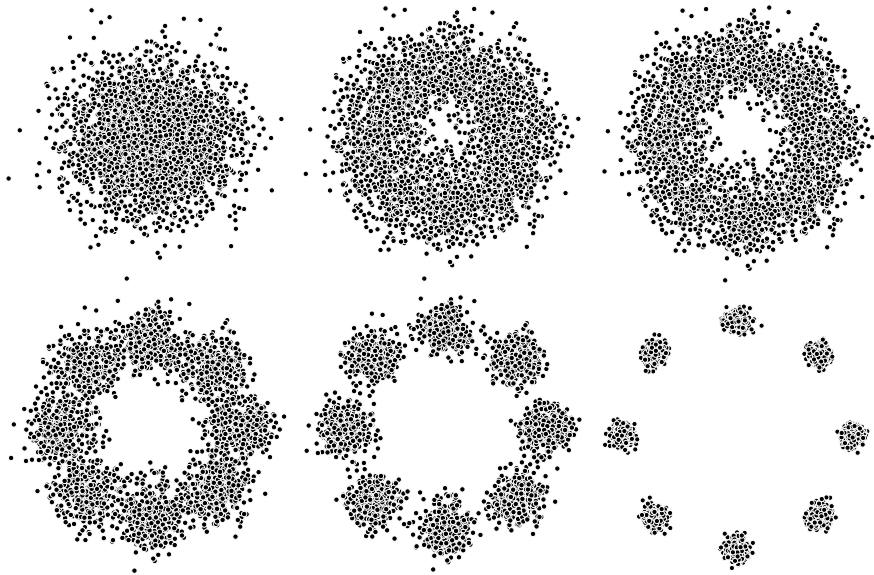
Figure 2: Snapshots of a Gaussian interpolation flow based on the Föllmer interpolant. The source distribution is the standard two-dimensional Gaussian distribution γ_2, and the target distribution is a mixture of six two-dimensional Gaussian distributions as the shape of a circle. The image panels are placed sequentially from time $t = 0$ to time $t = 1$.

We are now ready to define Gaussian interpolation flows by representing the continuity equation (3.4) with Lagrangian coordinates (Ambrosio and Crippa, 2014). A basic observation is that GIFs share the same marginal density flow with Gaussian stochastic interpolations. The continuity equation (3.4) plays a central role in the derandomization procedure from Gaussian stochastic interpolations to GIFs. We additionally illustrate GIFs using a two-dimensional example as in Figure 2.

Definition 3.3 (Gaussian interpolation flow) *Suppose that probability measure ν satisfies Assumption 2.1. If $(X_t)_{t\in[0,1]}$ solves the initial value problem (IVP)*

$$\frac{\mathrm{d}X_t}{\mathrm{d}t}(x) = v(t, X_t(x)), \quad X_0(x) \sim \mu, \quad t \in [0,1], \tag{3.10}$$

where μ is defined in Definition 3.2 and the velocity field v is given by Eq. (3.5) and (3.6), we call $(X_t)_{t\in[0,1]}$ a Gaussian interpolation flow associated with the target measure ν.

4. Spatial Lipschitz estimates for the velocity field

We have explicated the idea of Gaussian denoising with the procedure of Gaussian stochastic interpolation or a Gaussian channel with increasing SNR w.r.t. time. By interpreting the process as an ODE flow, we derive the framework of Gaussian interpolation flows. First and foremost, an intuition is that the regularizing effect of Gaussian denoising would ensure

11

the Lipschitz smoothness of the velocity field. Since the standard Gaussian distribution is both 1-semi-log-concave and 1-semi-log-convex, its convolution with a target distribution will maintain its high regularity as long as the target distribution satisfies the regularity conditions. We rigorously justify this intuition by establishing spatial Lipschitz estimates for the velocity field. These estimates are established based on the upper bounds and lower bounds regarding the Jacobian matrix of the velocity field $v(t, x)$ according to the Cauchy-Lipschitz theorem, which are given in Proposition 4.1 below. To deal with the Jacobian matrix $\nabla_x v(t, x)$, we introduce a covariance expression of it and present the associated upper bounds and lower bounds.

The velocity field $v(t, x)$ is decomposed into a linear term and a nonlinear term, the score function $s(t, x)$. To analyze the Jacobian $\nabla_x v(t, x)$, we only need to focus on $\nabla_x s(t, x)$, that is, $\nabla_x^2 \log p_t(x)$. To ease the notation, we would henceforth use Y for X_1. Correspondingly, we replace $p_1(x)$ with $p_1(y)$ for the density function of Y.

According to Bayes' theorem, the marginal density p_t of X_t satisfies

$$p_t(x) = \int p(t, x|y) p_1(y) \mathrm{d}y$$

where $\mathsf{Y} \sim p_1(y)$ and $p(t, x|y) = \varphi_{b_t y, a_t^2}(x)$ is a conditional distribution induced by the Gaussian noise. Due to the factorization $p_t(x) p(y|t, x) = p(t, x|y) p_1(y)$, the score function $s(t, x)$ and its derivative $\nabla_x s(t, x)$ have the following expressions

$$s(t, x) = -\nabla_x \log p(y|t, x) - \frac{x - b_t y}{a_t^2}, \quad \nabla_x s(t, x) = -\nabla_x^2 \log p(y|t, x) - \frac{1}{a_t^2} \mathbf{I}_d.$$

Thanks to the expressions above, a covariance matrix expression of $\nabla_x s(t, x)$ is endowed by the exponential family property of $p(y|t, x)$.

Lemma 4.1 *The conditional distribution $p(y|t, x)$ is an exponential family distribution and a covariance matrix expression of the log-Hessian matrix $\nabla_x^2 \log p(y|t, x)$ for any $t \in (0, 1)$ is given by*

$$\nabla_x^2 \log p(y|t, x) = -\frac{b_t^2}{a_t^4} \mathrm{Cov}(\mathsf{Y}|\mathsf{X}_t = x), \tag{4.1}$$

where $\mathrm{Cov}(\mathsf{Y}|\mathsf{X}_t = x)$ is the covariance matrix of $\mathsf{Y}|\mathsf{X}_t = x \sim p(y|t, x)$. Moreover, for any $t \in (0, 1)$, it holds that

$$\nabla_x s(t, x) = \frac{b_t^2}{a_t^4} \mathrm{Cov}(\mathsf{Y}|\mathsf{X}_t = x) - \frac{1}{a_t^2} \mathbf{I}_d, \tag{4.2}$$

and that

$$\nabla_x v(t, x) = \frac{b_t^2}{a_t^2} \left(\frac{\dot{b}_t}{b_t} - \frac{\dot{a}_t}{a_t} \right) \mathrm{Cov}(\mathsf{Y}|\mathsf{X}_t = x) + \frac{\dot{a}_t}{a_t} \mathbf{I}_d. \tag{4.3}$$

Remark 4.1 *Since $\partial_t \left(\frac{b_t^2}{a_t^2} \right) = \frac{2b_t^2}{a_t^2} \left(\frac{\dot{b}_t}{b_t} - \frac{\dot{a}_t}{a_t} \right)$, it follows from (4.3) that the derivative of the SNR with respect to time t controls the dependence of $\nabla_x v(t, x)$ on $\mathrm{Cov}(\mathsf{Y}|\mathsf{X}_t = x)$.*

The representation (4.3) can be used to upper bound and lower bound $\nabla_x v(t, x)$. This technique has been widely used to deduce the regularity of the score function concerning the space variable (Mikulincer and Shenfeld, 2021, 2023; Chen et al., 2023b; Lee et al., 2023;

Chen et al., 2023a). The covariance matrix expression (4.2) of the score function has a close connection with the Hatsell-Nolte identity in information theory (Hatsell and Nolte, 1971; Palomar and Verdú, 2005; Wu and Verdú, 2011; Cai and Wu, 2014; Wibisono et al., 2017; Wibisono and Jog, 2018a,b; Dytso et al., 2023a,b).

Employing the covariance expression in Lemma 4.1, we establish several bounds on $\nabla_x v(t, x)$ in the following proposition.

Proposition 4.1 *Let $\nu(\mathrm{d}y) = p_1(y)\mathrm{d}y$ be a probability measure on $\mathbb{R}^d$ with $D := (1/\sqrt{2})\mathrm{diam}(\mathrm{supp}(\nu))$.*

(a) *For any $t \in (0, 1)$,*

$$\frac{\dot{a}_t}{a_t}\mathbf{I}_d \preceq \nabla_x v(t, x) \preceq \left\{\frac{b_t(a_t\dot{b}_t - \dot{a}_t b_t)}{a_t^3}D^2 + \frac{\dot{a}_t}{a_t}\right\}\mathbf{I}_d.$$

(b) *Suppose that p_1 is β-semi-log-convex with $\beta > 0$ and $\mathrm{supp}(p_1) = \mathbb{R}^d$. Then for any $t \in (0, 1]$,*

$$\nabla_x v(t, x) \succeq \frac{\beta a_t \dot{a}_t + b_t \dot{b}_t}{\beta a_t^2 + b_t^2}\mathbf{I}_d.$$

(c) *Suppose that p_1 is κ-semi-log-concave with $\kappa \in \mathbb{R}$. Then for any $t \in (t_0, 1]$,*

$$\nabla_x v(t, x) \preceq \frac{\kappa a_t \dot{a}_t + b_t \dot{b}_t}{\kappa a_t^2 + b_t^2}\mathbf{I}_d,$$

where t_0 is the root of the equation $\kappa + \frac{b_t^2}{a_t^2} = 0$ over $t \in (0, 1)$ if $\kappa < 0$ and $t_0 = 0$ if $\kappa \geq 0$.

(d) *Fix a probability measure ρ on $\mathbb{R}^d$ supported on a Euclidean ball of radius R, and let $\nu := \gamma_{d,\sigma^2} * \rho$ with $\sigma > 0$. Then for any $t \in (0, 1)$,*

$$\frac{\dot{a}_t a_t + \sigma^2 \dot{b}_t b_t}{a_t^2 + \sigma^2 b_t^2}\mathbf{I}_d \preceq \nabla_x v(t, x) \preceq \left\{\frac{a_t b_t(a_t \dot{b}_t - \dot{a}_t b_t)}{(a_t^2 + \sigma^2 b_t^2)^2}R^2 + \frac{\dot{a}_t a_t + \sigma^2 \dot{b}_t b_t}{a_t^2 + \sigma^2 b_t^2}\right\}\mathbf{I}_d.$$

(e) *Suppose that $\frac{\mathrm{d}\nu}{\mathrm{d}\gamma_d}(x)$ is L-log-Lipschitz for some $L \geq 0$. Then for any $t \in (0, 1)$,*

$$\left\{\left(\frac{\dot{b}_t}{b_t}a_t^2 - \dot{a}_t a_t\right)\left(-B_t - L^2\left(\frac{b_t}{a_t^2 + b_t^2}\right)^2\right) + \frac{\dot{a}_t a_t + \dot{b}_t b_t}{a_t^2 + b_t^2}\right\}\mathbf{I}_d$$
$$\preceq \nabla_x v(t, x) \preceq \left\{\left(\frac{\dot{b}_t}{b_t}a_t^2 - \dot{a}_t a_t\right)B_t + \frac{\dot{a}_t a_t + \dot{b}_t b_t}{a_t^2 + b_t^2}\right\}\mathbf{I}_d,$$

where $B_t := 5Lb_t(a_t^2 + b_t^2)^{-\frac{3}{2}}(L + (\log(\sqrt{a_t^2 + b_t^2}/b_t))^{-\frac{1}{2}})$.

Comparing part (a) with part (d) in Proposition 4.1, we can see that the bounds in (a) are consistent with those in (d) in the sense that (a) is a limiting case of part (d) as $\sigma \to 0$. The lower bound in part (a) blows up at time $t = 1$ owing to $a_1 = 0$, while in part (d) it

behaves well since the lower bound in part (d) coincides with a lower bound indicated by the $\frac{1}{\sigma^2}$-semi-log-convex property. It reveals that the regularity of the velocity field $v(t, x)$ with respect to the space variable x improves when the target random variable is bounded and is subject to Gaussian perturbation.

The lower bound in part (b) and the upper bound in part (c) are tight in the sense that both of them are attainable for a Gaussian target distribution, that is,

$$\nabla_x v(t, x) = \frac{\beta a_t \dot{a}_t + b_t \dot{b}_t}{\beta a_t^2 + b_t^2} \mathbf{I}_d \quad \text{if } \nu = \gamma_{d, 1/\beta}.$$

The upper and lower bounds in Proposition 4.1-(a) and (e) become vacuous as they both blow up at time $t = 1$. The intuition behind is that the Jacobian matrix of the velocity field can be both lower and upper bounded at time $t = 1$ only if the score function of the target measure is Lipschitz continuous in the space variable x. Under an additional Lipschitz score assumption (equivalently, β-semi-log-convex and κ-semi-log-concave for some $\beta = -\kappa \geq 0$), the upper and lower bounds in part (a) and part (e) can be strengthened at time $t = 1$ based on the lower bound in (b) and the upper bound in part (c).

According to Proposition 4.1-(a) and (c), there are two upper bounds available that shall be compared with each other. One is the D^2-based bound in part (a), and the other is the κ-based bound in part (c). According to the proof of Proposition 4.1 given in the Appendix, these two upper bounds are equal if and only if the corresponding upper bounds on $\text{Cov}(Y|X_t = x)$ are equal, that is,

$$D^2 = \left(\kappa + \frac{b_t^2}{a_t^2} \right)^{-1}. \tag{4.4}$$

Then the critical case is $\kappa D^2 = 1$ since simplifying Eq. (4.4) reveals that

$$D^{-2} - \kappa = \frac{b_t^2}{a_t^2}. \tag{4.5}$$

We note that b_t^2/a_t^2, ranging over $(0, \infty)$, is monotonically increasing w.r.t. $t \in (0, 1)$. Suppose that $\kappa D^2 > 1$. Then (4.5) has no root over $t \in (0, 1)$, which implies that the κ-based bound is tighter over $[0, 1)$, i.e.,

$$D^2 > \left(\kappa + \frac{b_t^2}{a_t^2} \right)^{-1}, \quad \forall t \in [0, 1).$$

Otherwise, suppose that $\kappa D^2 < 1$. Then (4.5) has a root $t_1 \in (0, 1)$, which implies that the D^2-based bound is tighter over $[0, t_1)$, i.e.,

$$D^2 < \left(\kappa + \frac{b_t^2}{a_t^2} \right)^{-1}, \quad \forall t \in [0, t_1),$$

and that the κ-based bound is tighter over $[t_1, 1)$, i.e.,

$$D^2 \geq \left(\kappa + \frac{b_t^2}{a_t^2} \right)^{-1}, \quad \forall t \in [t_1, 1).$$

Next, we present several upper bounds on the maximum eigenvalue of the Jacobian matrix of the velocity field $\lambda_{\max}(\nabla_x v(t, x))$ and its exponential estimates for studying the Lipschitz regularity of the flow maps as noted in Lemma 5.1.

Corollary 4.1 *Let ν be a probability measure on $\mathbb{R}^d$ with $D := (1/\sqrt{2})\mathrm{diam}(\mathrm{supp}(\nu))$ and suppose that ν is κ-semi-log-concave with $\kappa \geq 0$.*

(a) If $\kappa D^2 \geq 1$, then

$$\lambda_{\max}(\nabla_x v(t, x)) \leq \theta_t := \frac{\kappa a_t \dot{a}_t + b_t \dot{b}_t}{\kappa a_t^2 + b_t^2}, \ t \in [0, 1]. \tag{4.6}$$

(b) If $\kappa D^2 < 1$, then

$$\lambda_{\max}(\nabla_x v(t, x)) \leq \theta_t := \begin{cases} \frac{b_t^2}{a_t^2}\left(\frac{\dot{b}_t}{b_t} - \frac{\dot{a}_t}{a_t}\right) D^2 + \frac{\dot{a}_t}{a_t}, & t \in [0, t_1), \\ \frac{\kappa a_t \dot{a}_t + b_t \dot{b}_t}{\kappa a_t^2 + b_t^2}, & t \in [t_1, 1], \end{cases} \tag{4.7}$$

where t_1 solves (4.5).

Corollary 4.2 *Let ν be a probability measure on $\mathbb{R}^d$ with $D := (1/\sqrt{2})\mathrm{diam}(\mathrm{supp}(\nu)) < \infty$ and suppose that ν is κ-semi-log-concave with $\kappa < 0$. Then*

$$\lambda_{\max}(\nabla_x v(t, x)) \leq \theta_t := \begin{cases} \frac{b_t^2}{a_t^2}\left(\frac{\dot{b}_t}{b_t} - \frac{\dot{a}_t}{a_t}\right) D^2 + \frac{\dot{a}_t}{a_t}, & t \in [0, t_1), \\ \frac{\kappa a_t \dot{a}_t + b_t \dot{b}_t}{\kappa a_t^2 + b_t^2}, & t \in [t_1, 1], \end{cases} \tag{4.8}$$

where t_1 solves (4.5).

Corollary 4.3 *Fix a probability measure ρ on $\mathbb{R}^d$ supported on a Euclidean ball of radius R and let $\nu := \gamma_{d,\sigma^2} * \rho$ with $\sigma > 0$. Then*

$$\lambda_{\max}(\nabla_x v(t, x)) \leq \theta_t := \frac{\dot{a}_t a_t + \sigma^2 \dot{b}_t b_t}{a_t^2 + \sigma^2 b_t^2} + \frac{a_t b_t(a_t \dot{b}_t - \dot{a}_t b_t)}{(a_t^2 + \sigma^2 b_t^2)^2} R^2. \tag{4.9}$$

Corollary 4.4 *Suppose that ν is κ-semi-log-concave for some $\kappa \leq 0$, and $\frac{\mathrm{d}\nu}{\mathrm{d}\gamma_d}(x)$ is L-log-Lipschitz for some $L \geq 0$. Then*

$$\lambda_{\max}(\nabla_x v(t, x)) \leq \theta_t := \begin{cases} \left(\frac{\dot{b}_t}{b_t} a_t^2 - \dot{a}_t a_t\right) B_t + \frac{\dot{a}_t a_t + \dot{b}_t b_t}{a_t^2 + b_t^2}, & t \in [0, t_2), \\ \frac{\kappa a_t \dot{a}_t + b_t \dot{b}_t}{\kappa a_t^2 + b_t^2}, & t \in [t_2, 1], \end{cases} \tag{4.10}$$

where $B_t := 5Lb_t(a_t^2 + b_t^2)^{-\frac{3}{2}}(L + (\log(\sqrt{a_t^2 + b_t^2}/b_t))^{-\frac{1}{2}})$ and $t_2 \in (t_0, 1)$.

5. Well-posedness and Lipschtiz flow maps

In this section, we study the well-posedness of GIFs and the Lipschitz properties of their flow maps. We also show that the marginal distributions of GIFs satisfy the log-Sobolev inequality and the Poincaré inequality if Assumptions 2.1 and 2.2 are satisfied.

Theorem 5.1 (Well-posedness) *Suppose Assumptions 2.1 and 2.2-(i), (iii), or (iv) are satisfied. Then there exists a unique solution $(X_t)_{t\in[0,1]}$ to the IVP (3.10). Moreover, the push-forward measure satisfies $X_{t\#}\mu = \mathrm{Law}(a_t\mathsf{Z} + b_t\mathsf{X}_1)$ with $\mathsf{Z} \sim \gamma_d, \mathsf{X}_1 \sim \nu$.*

Theorem 5.2 *Suppose Assumptions 2.1 and 2.2-(ii) are satisfied. For any $\underline{t} \in (0,1)$, there exists a unique solution $(X_t)_{t\in[0,1-\underline{t}]}$ to the IVP (3.10). Moreover, the push-forward measure satisfies $X_{t\#}\mu = \mathrm{Law}(a_t\mathsf{Z} + b_t\mathsf{X}_1)$ with $\mathsf{Z} \sim \gamma_d, \mathsf{X}_1 \sim \nu$.*

Corollary 5.1 (Time-reversed flow) *Suppose Assumptions 2.1 and 2.2-(i), (iii), or (iv) are satisfied. Then the time-reversed flow $(X_t^*)_{t\in[0,1]}$ associated with ν is a unique solution to the IVP:*

$$\frac{\mathrm{d}X_t^*}{\mathrm{d}t}(x) = -v(1-t, X_t^*(x)), \quad X_0^*(x) \sim \nu, \quad t \in [0,1]. \tag{5.1}$$

The push-forward measure satisfies $X_{t\#}^*\nu = \mathrm{Law}(a_{1-t}\mathsf{Z} + b_{1-t}\mathsf{X}_1)$ where $\mathsf{Z} \sim \gamma_d, \mathsf{X}_1 \sim \nu$. Moreover, the flow map satisfies $X_t^*(x) = X_t^{-1}(x)$.

Corollary 5.2 *Suppose Assumptions 2.1 and 2.2-(ii) are satisfied. For any $\underline{t} \in (0,1)$, the time-reversed flow $(X_t^*)_{t\in[\underline{t},1]}$ associated with ν is a unique solution to the IVP:*

$$\frac{\mathrm{d}X_t^*}{\mathrm{d}t}(x) = -v(1-t, X_t^*(x)), \quad X_{\underline{t}}^*(x) \sim \mathrm{Law}(a_{1-\underline{t}}\mathsf{Z} + b_{1-\underline{t}}\mathsf{X}_1), \quad t \in [\underline{t}, 1], \tag{5.2}$$

where $\mathsf{Z} \sim \gamma_d, \mathsf{X}_1 \sim \nu$. The push-forward measure satisfies $X_{t\#}^*\nu = \mathrm{Law}(a_{1-t}\mathsf{Z} + b_{1-t}\mathsf{X}_1)$. Moreover, the flow map satisfies $X_t^*(x) = X_t^{-1}(x)$.

Based on the well-posedness of the flow, we can provide an upper bound on the Lipschitz constant of the induced flow map.

Lemma 5.1 *Suppose that a flow $(X_t)_{t\in[0,1]}$ is well-posed with a velocity field $v(t,x) : [0,1] \times \mathbb{R}^d \to \mathbb{R}^d$ of class C^1 in x, and that for any $(t,x) \in [0,1] \times \mathbb{R}^d$, it holds $\nabla_x v(t,x) \preceq \theta_t \mathbf{I}_d$. Let the flow map $X_{s,t} : \mathbb{R}^d \to \mathbb{R}^d$ be of class C^1 in x for any $0 \le s \le t \le 1$. Then the flow map $X_{s,t}$ is Lipschitz continuous with an upper bound of its Lipschitz constant given by*

$$\|\nabla_x X_{s,t}(x)\|_{2,2} \le \exp\left(\int_s^t \theta_u \mathrm{d}u\right). \tag{5.3}$$

Using Lemma 5.1, we show that the flow map of a GIF is Lipschitz continuous in the space variable x.

Proposition 5.1 (Lipschitz mappings) *Suppose that Assumptions 2.1 and 2.2-(i) hold.*

16

(i) *If ν is κ-semi-log-concave for some $\kappa > 0$, then the flow map $X_1(x)$ is a Lipschitz mapping, that is,*

$$\|\nabla_x X_1(x)\|_{2,2} \leq \frac{1}{\sqrt{\kappa a_0^2 + b_0^2}}, \quad \forall x \in \mathbb{R}^d.$$

In particular, if $a_0 = 1$ and $b_0 = 0$, then

$$\|\nabla_x X_1(x)\|_{2,2} \leq \frac{1}{\sqrt{\kappa}}, \quad \forall x \in \mathbb{R}^d.$$

(ii) *If ν is β-semi-log-convex for some $\beta > 0$, then the time-reversed flow map $X_1^*(x)$ is a Lipschitz mapping, that is,*

$$\|\nabla_x X_1^*(x)\|_{2,2} \leq \sqrt{\beta a_0^2 + b_0^2}, \quad \forall x \in \mathrm{supp}(\nu).$$

In particular, if $a_0 = 1$ and $b_0 = 0$, then

$$\|\nabla_x X_1^*(x)\|_{2,2} \leq \sqrt{\beta}, \quad \forall x \in \mathrm{supp}(\nu).$$

Proposition 5.2 (Gaussian mixtures) *Suppose that Assumptions 2.1 and 2.2-(iii) hold. Then the flow map $X_1(x)$ is a Lipschitz mapping, that is,*

$$\|\nabla_x X_1(x)\|_{2,2} \leq \frac{\sigma}{\sqrt{a_0^2 + \sigma^2 b_0^2}} \exp\left(\frac{a_0^2}{a_0^2 + \sigma^2 b_0^2} \cdot \frac{R^2}{2\sigma^2}\right), \quad \forall x \in \mathbb{R}^d.$$

In particular, if $a_0 = 1$ and $b_0 = 0$, then

$$\|\nabla_x X_1(x)\|_{2,2} \leq \sigma \exp\left(\frac{R^2}{2\sigma^2}\right), \quad \forall x \in \mathbb{R}^d.$$

Moreover, the time-reversed flow map $X_1^*(x)$ is a Lipschitz mapping, that is,

$$\|\nabla_x X_1^*(x)\|_{2,2} \leq \sqrt{\sigma^{-2} a_0^2 + b_0^2}, \quad \forall x \in \mathrm{supp}(\nu).$$

In particular, if $a_0 = 1$ and $b_0 = 0$, then

$$\|\nabla_x X_1^*(x)\|_{2,2} \leq \frac{1}{\sigma}, \quad \forall x \in \mathrm{supp}(\nu).$$

Remark 5.1 *Well-posed GIFs produce diffeomorphisms that transport the source measure onto the target measure. The diffeomorphism property of the transport maps are relevant to the auto-encoding and cycle consistency properties of their generative modeling applications. We defer a detailed discussion to Section 6.*

Early stopping implicitly mollifies the target measure with a small Gaussian noise. For image generation tasks (with bounded pixel values), the mollified target measure is indeed a Gaussian mixture distribution considered in Theorem 5.2. The regularity of the target measure largely gets enhanced through such mollification, especially when the target measure is supported on a low-dimensional manifold in accordance with the data manifold hypothesis. Therefore, although such a diffeomorphism $X_1(x)$ may not be well-defined for general bounded target measures, an off-the-shelf solution would be to perturb the target measure with a small Gaussian noise or to employ the early stopping technique. Both approaches will smooth the landscape of the target measure.

Proposition 5.3 *Suppose the target measure ν satisfies the log-Sobolev inequality with constant $C_{\mathrm{LS}}(\nu)$. Then the marginal distribution of the GIF $(p_t)_{t\in[0,1]}$ satisfies the log-Sobolev inequality, and its log-Sobolev constant $C_{\mathrm{LS}}(p_t)$ is bounded as*

$$C_{\mathrm{LS}}(p_t) \leq a_t^2 + b_t^2 C_{\mathrm{LS}}(\nu).$$

Moreover, suppose the target measure ν satisfies the Poincaré inequality with constant $C_P(\nu)$. Then the marginal distribution of the GIF $(p_t)_{t\in[0,1]}$ satisfies the Poincaré inequality, and its Poincaré constant $C_{\mathrm{P}}(p_t)$ is bounded as

$$C_{\mathrm{P}}(p_t) \leq a_t^2 + b_t^2 C_{\mathrm{P}}(\nu).$$

The log-Sobolev and Poincaré inequalities (see Definitions G.1 and G.2) are fundamental tools for establishing convergence guarantees for Langevin Monte Carlo algorithms. From an algorithmic viewpoint, the predictor-corrector algorithm in score-based diffusion models and the corresponding probability flow ODEs essentially combine the ODE numerical solver (performing as the predictor) and the overdamped Langevin diffusion (performing as the corrector) to simulate samples from the marginal distributions (Song et al., 2021b). Proposition 5.3 shows that the marginal distributions all satisfy the log-Sobolev and Poincaré inequalities under mild assumptions on the target distribution. This conclusion suggests that Langevin Monte Carlo algorithms are certified to have convergence guarantees for sampling from the marginal distributions of GIFs. Furthermore, the target distributions covered in Assumption 2.2 are shown to satisfy the log-Sobolev and Poincaré inequalities (Mikulincer and Shenfeld, 2021; Dai et al., 2023; Fathi et al., 2023), which suggests that the assumptions of Proposition 5.3 generally hold.

6. Applications to generative modeling

Auto-encoding is a primary principle in learning a latent representation with generative models (Goodfellow et al., 2016, Chapter 14). Meanwhile, the concept of cycle consistency is important to unpaired image-to-image translation between the source and target domains (Zhu et al., 2017). The recent work by Su et al. (2023) propose the dual diffusion implicit bridges (DDIB) for image-to-image translation, which shows a strong pattern of exact auto-encoding and image-to-image translation. DDIBs are built upon the denoising diffusion implicit models (DDIM), which share the same probability flow ODE with VESDE (considered as VE interpolant in Table 1), as pointed out by (Song et al., 2021a, Proposition 1). First, DDIBs attain latent embeddings of source images encoded with one DDIM operating in the source domain. The encoding embeddings are then decoded using another DDIM trained in the target domain to construct target images. The whole process consisting of two DDIMs seems to be cycle consistent up to numerical errors. Several phenomena of auto-encoding and cycle consistency are observed in the unpaired data generation procedure with DDIBs.

We replicate the two-dimensional experiments by Su et al. (2023) in Figures 3 and 4 to show the phenomena of approximate auto-encoding and cycle consistency of GIFs[1]. To elucidate the empirical auto-encoding and cycle consistency for measure transport, we derive

---

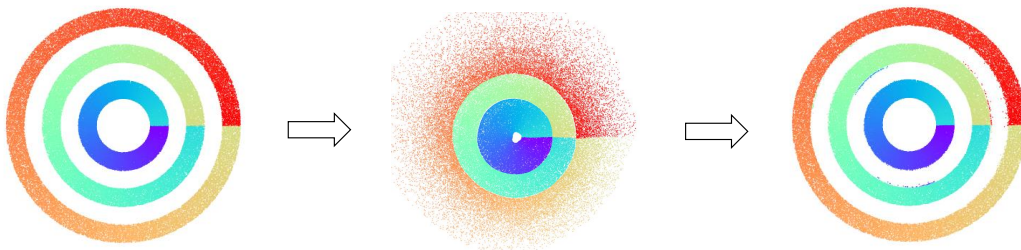1. The implementation is based on the GitHub repository at `https://github.com/suxuann/ddib`.

Figure 3: An illustration of auto-encoding using DDIBs. The Concentric Rings data in the source domain (the first panel) is encoded into the latent domain (the second panel), and then decoded into the source domain (the third panel). According to the consistent color pattern and pointwise correspondences across the domains, both the learned encoder mapping and the learned decoder mapping exhibit approximate Lipschitz continuity with respect to the space variable. One justification of such auto-encoding observation is presented in Corollary 6.1 where we prove that the composition of the encoder map and the decoder map yields an identity map.

Corollaries 6.1 and 6.2 below and analyze the transport maps defined by GIFs (covering the probability flow ODE of VESDE used by DDIBs). We consider the continuous-time framework and the population level, which precludes learning errors including the time discretization errors and velocity field estimation errors, and show that the transport maps naturally possess the exact auto-encoding and cycle consistency properties at the population level.

Corollary 6.1 (Auto-encoding) *Suppose Assumptions 2.1 and 2.2-(i), (iii), or (iv) hold for a target measure ν. The Gaussian interpolation flow $(X_t)_{t\in[0,1]}$ and its time-reversed flow $(X_t^*)_{t\in[0,1]}$ form an auto-encoder with a Lipschitz encoder $X_1^*(x)$ and a Lipschitz decoder $X_1(x)$. The auto-encoding property holds in the sense that*

$$X_1 \circ X_1^* = \mathbf{I}_d. \tag{6.1}$$

Corollary 6.2 (Cycle consistency) *Suppose Assumptions 2.1 and 2.2-(i), (iii), or (iv) hold for the target measures ν_1 and ν_2. For the target measure ν_1, we define the Gaussian interpolation flow $(X_{1,t})_{t\in[0,1]}$ and its time-reversed flow $(X_{1,t}^*)_{t\in[0,1]}$. We also define the Gaussian interpolation flow $(X_{2,t})_{t\in[0,1]}$ and its time-reversed flow $(X_{2,t}^*)_{t\in[0,1]}$ for the target measure ν_2 using the same a_t and b_t. Then the transport maps $X_{1,1}(x)$, $X_{1,1}^*(x)$, $X_{2,1}(x)$, and $X_{2,1}^*(x)$ are Lipschitz continuous in the space variable x. Furthermore, the cycle consistency property holds in the sense that*

$$X_{1,1} \circ X_{2,1}^* \circ X_{2,1} \circ X_{1,1}^* = \mathbf{I}_d. \tag{6.2}$$

Corollaries 6.1 and 6.2 show that the auto-encoding and cycle consistency properties hold for the flows at the population level. These results provide insights to the approximate auto-encoding and cycle consistency properties at the sample level.
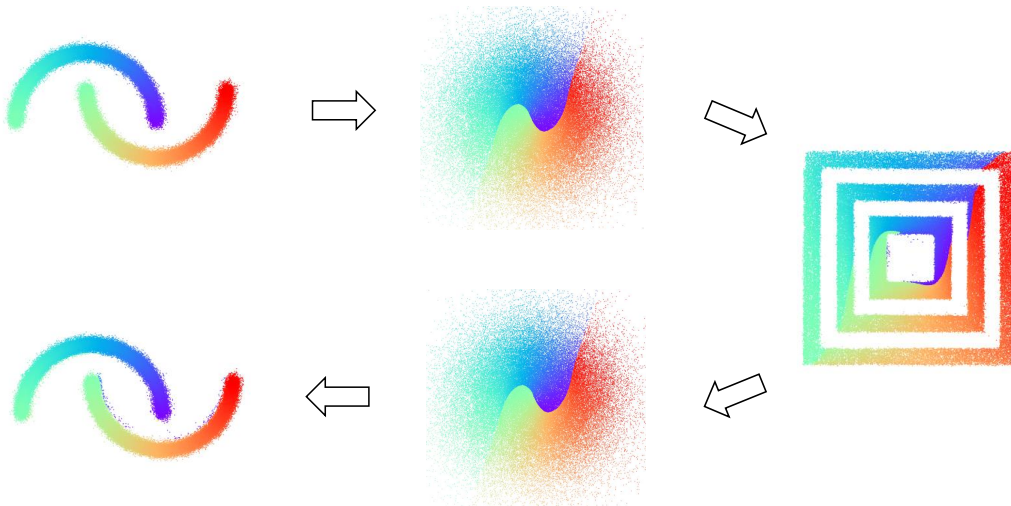
Figure 4: An illustration of cycle consistency using DDIBs. The cycle consistency property is manifested through the consistency of color patterns across the transformations. We transform the Moons data in the source domain onto the Concentric Squares data in the target domain, and then complete the cycle by mapping the target data back to the source domain. The latent spaces play a central role in the bidirectional translation. We provide a proof in Corollary 6.2 accounting for the cycle consistency property.

There are several types of errors introduced in the training of GIFs. On the one hand, the approximation in specifying source measures would exert influence on modeling the distribution. On the other hand, the approximation in the velocity field also affects the distribution learning error. We use the stability analysis method in the differential equations theory to address the potential effects of these errors.

Corollary 6.3 *Suppose Assumptions 2.1 and 2.2-(i), (iii), or (iv) hold. It holds that*

$$C_1 := \sup_{x \in \mathbb{R}^d} \|\nabla_x X_1(x)\|_{2,2} < \infty, \quad C_2 := \sup_{(t,x) \in [0,1] \times \mathbb{R}^d} \|\nabla_x v(t,x)\|_{2,2} < \infty.$$

Proposition 6.1 (Stability in the source distribution) *Suppose Assumptions 2.1 and 2.2-(i), (iii), or (iv) hold. If the source measure $\mu = \mathrm{Law}(a_0 \mathsf{Z} + b_0 \mathsf{X}_1)$ is replaced with the Gaussian measure γ_{d,a_0^2}, then the stability of the transport map X_1 is guaranteed by the W_2 distance between the push-forward measure $X_{1\#}\gamma_{d,a_0^2}$ and the target measure $\nu = \mathrm{Law}(\mathsf{X}_1)$ as follows*

$$W_2(X_{1\#}\gamma_{d,a_0^2}, \nu) \le C_1 b_0 \sqrt{\mathbb{E}_\nu[\|\mathsf{X}_1\|^2]} \exp(C_2 d). \tag{6.3}$$

The stability analysis in Proposition 6.1 provides insights into the selection of source measures for learning probability flow ODEs and GIFs. The error bound (6.3) demonstrates that when the signal intensity is reasonably small in the source measure, that is, $b_0 \ll 1$, the

distribution estimation error, induced by the approximation with a Gaussian source measure, is small as well in the sense of the quadratic Wasserstein distance. Using a Gaussian source measure to replace the true convolution source measure is a common approximation method for learning probability flow ODEs and GIFs. Our analysis shows this replacement is reasonable for the purpose of distribution estimation.

The Alekseev-Gröbner formula and its stochastic variants (Del Moral and Singh, 2022) have been shown effective in quantifying the stability of well-posed ODE and SDE flows against perturbations of its velocity field or drift (Bortoli, 2022; Benton et al., 2023). We state these results below for convenience.

Lemma 6.1 *(Hairer et al., 1993, Theorem 14.5) Let $(X_t)_{t\in[0,1]}$ and $(Y_t)_{t\in[0,1]}$ solve the following IVPs, respectively*

$$\frac{\mathrm{d}X_t}{\mathrm{d}t} = v(t, X_t), \quad X_0 = x_0, \quad t \in [0,1],$$

$$\frac{\mathrm{d}Y_t}{\mathrm{d}t} = \tilde{v}(t, Y_t), \quad Y_0 = x_0, \quad t \in [0,1],$$

where $v(t,x) : [0,1] \times \mathbb{R}^d \to \mathbb{R}^d$ and $\tilde{v}(t,x) : [0,1] \times \mathbb{R}^d \to \mathbb{R}^d$ are the velocity fields.

(i) *Suppose that v is of class C^1 in x. Then the Alekseev-Gröbner formula for the difference $X_t(x_0) - Y_t(x_0)$ is given by*

$$X_t(x_0) - Y_t(x_0) = \int_0^t (\nabla_x X_{s,t})(Y_s(x_0))^\top \left(v(s, Y_s(x_0)) - \tilde{v}(s, Y_s(x_0))\right) \mathrm{d}s \qquad (6.4)$$

where $\nabla_x X_{s,t}(x)$ satisfies the variational equation

$$\partial_t(\nabla_x X_{s,t}(x)) = (\nabla_x v)(t, X_{s,t}(x))\nabla_x X_{s,t}(x), \quad \nabla_x X_{s,s}(x) = \mathbf{I}_d. \qquad (6.5)$$

(ii) *Suppose that $\tilde{v}$ is of class C^1 in x. Then the Alekseev-Gröbner formula for the difference $Y_t(x_0) - X_t(x_0)$ is given by*

$$Y_t(x_0) - X_t(x_0) = \int_0^t (\nabla_x Y_{s,t})(X_s(x_0))^\top \left(\tilde{v}(s, X_s(x_0)) - v(s, X_s(x_0))\right) \mathrm{d}s \qquad (6.6)$$

where $\nabla_x Y_{s,t}(x)$ satisfies the variational equation

$$\partial_t(\nabla_x Y_{s,t}(x)) = (\nabla_x \tilde{v})(t, Y_{s,t}(x))\nabla_x Y_{s,t}(x), \quad \nabla_x Y_{s,s}(x) = \mathbf{I}_d. \qquad (6.7)$$

Exploiting the Alekseev-Gröbner formulas in Lemma 6.1 and uniform Lipschitz properties of the velocity field, we deduce two error bounds in terms of the quadratic Wasserstein (W_2) distance to show the stability of the ODE flow when the velocity field is not accurate.

Proposition 6.2 (Stability in the velocity field) *Suppose Assumptions 2.1 and 2.2 hold. Let $\tilde{q}_t$ denote the density function of $Y_{t\#}\mu$.*

(i) Suppose that

$$\int_0^1 \int_{\mathbb{R}^d} \|v(t,x) - \tilde{v}(t,x)\|^2 \tilde{q}_t(x) \mathrm{d}x \mathrm{d}t \le \varepsilon. \tag{6.8}$$

Then

$$W_2^2(Y_{1\#}\mu, \nu) \le \varepsilon \int_0^1 \exp\left(2 \int_s^1 \theta_u \mathrm{d}u\right) \mathrm{d}s. \tag{6.9}$$

(ii) Suppose that

$$\sup_{(t,x)\in[0,1]\times\mathbb{R}^d} \|\nabla_x \tilde{v}(t,x)\|_{2,2} \le C_3. $$

Then

$$W_2^2(Y_{1\#}\mu, \nu) \le \frac{\exp(2C_3) - 1}{2C_3} \int_0^1 \int_{\mathbb{R}^d} \|v(t,x) - \tilde{v}(t,x)\|^2 p_t(x) \mathrm{d}x \mathrm{d}t. \tag{6.10}$$

Proposition 6.2 provides a stability analysis against the estimation error of the velocity field using the W_2 distance. The estimation error originates from the flow matching or score matching procedures and the approximation error rising from using deep neural networks in estimating the velocity field or the score function. These two W_2 bounds imply that the distribution estimation error is controlled by the L_2 estimation error of flow matching and score matching. Indeed, this point justifies the soundness of the approximation method through flow matching and score matching. The first W_2 bound (6.9) relies on the L_2 control (6.8) of the perturbation error of the velocity field. The second W_2 bound (6.10) is slightly better than that provided in (Albergo and Vanden-Eijnden, 2023, Proposition 3) but still has exponential dependence on the Lipschitz constant of $\tilde{v}(t,x)$.
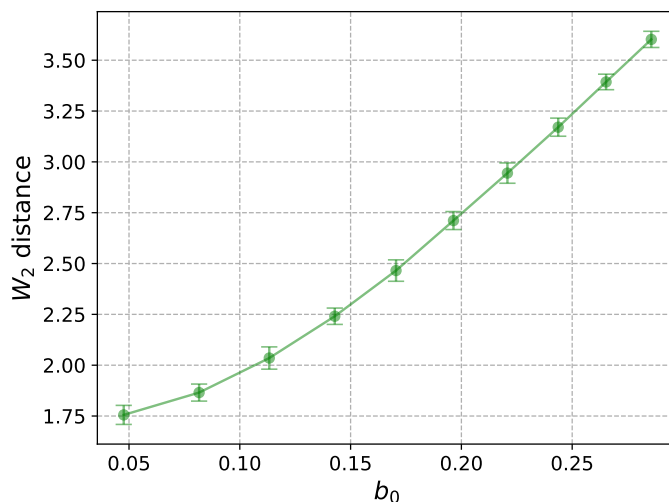

Figure 5: An approximately linear relation between b_0 and the Wasserstein-2 distance.

To demonstrate the bounds presented in Propositions 6.1 and 6.2, we conducted further experiments with a mixture of eight two-dimensional Gaussian distributions. These

propositions provide bounds for the stability of the flow when subjected to perturbations in either the source distribution or the velocity field. Let the target distribution be the following two-dimensional Gaussian mixture

$$p(x) = \sum_{j=1}^{8} \phi(x; \mu_j, \Sigma_j),$$

where $\phi(x; \mu_j, \Sigma_j)$ is the probability density function for the Gaussian distribution with mean $\mu_j = 12(\sin(2(j-1)\pi/8), \cos(2(j-1)\pi/8))^{\top}$ and covariance matrix $\Sigma_j = 0.03^2 \mathbf{I}_2$ for $j = 1, \cdots, 8$. For Gaussian mixtures, the velocity field has an explicit formula, which facilitates the perturbation analysis.

To illustrate the bound in Proposition 6.1, we consider a perturbation of the source distribution for the following model:

$$\mathsf{X}_t = a_t \mathsf{Z} + b_t \mathsf{X} \quad \text{with} \quad a_t = 1 - \frac{t+\zeta}{1+\zeta}, \quad b_t = \frac{t+\zeta}{1+\zeta},$$

where $\zeta \in [0, 0.3]$ is a value controlling the perturbation level. It is easy to see $a_0 = 1/(1+\zeta), b_0 = \zeta/(1+\zeta)$. Thus, the source distribution $\text{Law}(a_0 \mathsf{Z} + b_0 \mathsf{X})$ is a mixture of Gaussian distributions. Practically, we can use a Gaussian distribution γ_{2, a_0^2} to replace this source distribution. In Proposition 6.1, we bound the error between the distributions of generated samples due to the replacement, that is,

$$W_2(X_{1\#}\gamma_{d, a_0^2}, \nu) \leq C b_0,$$

where C is a constant. We illustrate this theoretical bound using the mixture of Gaussian distributions and the Gaussian interpolation flow given above. We consider a mesh for the variable ζ and plot the curve for b_0 and $W_2(X_{1\#}\gamma_{d, a_0^2}, \nu)$ in Figure 5. Through Figure 5, an approximate linear relation between b_0 and $W_2(X_{1\#}\gamma_{d, a_0^2}, \nu)$ is observed, which supports the results of Proposition 6.1.

We now consider perturbing the velocity field v_t by adding random noise. Let $\epsilon \in [0.5, 5.5]$. The random noise is generated using a Bernoulli random variable supported on $\{-\epsilon, \epsilon\}$. Let $\tilde{v}_t$ denote the perturbed velocity field. Then we can compute

$$\Delta v_t := \|v_t - \tilde{v}_t\|^2 = 2\epsilon^2.$$

We use the velocity field v_t and the perturbed velocity field $_t$ to generate samples and compute the squared Wasserstein-2 distance between the sample distributions. According to Proposition 6.2, the squared Wasserstein-2 distance should be linearly upper bounded as $\mathcal{O}(\Delta v_t)$, that is,

$$W_2^2(Y_{1\#}\mu, \nu) \leq \tilde{C} \int_0^1 \int_{\mathbb{R}^2} \epsilon^2 p_t(x) \mathrm{d}x \mathrm{d}t = \tilde{C}\epsilon^2,$$

where $\tilde{C}$ is a constant. This theoretical insight is illustrated in Figure 6, where a linear relationship between these two variables is observed.
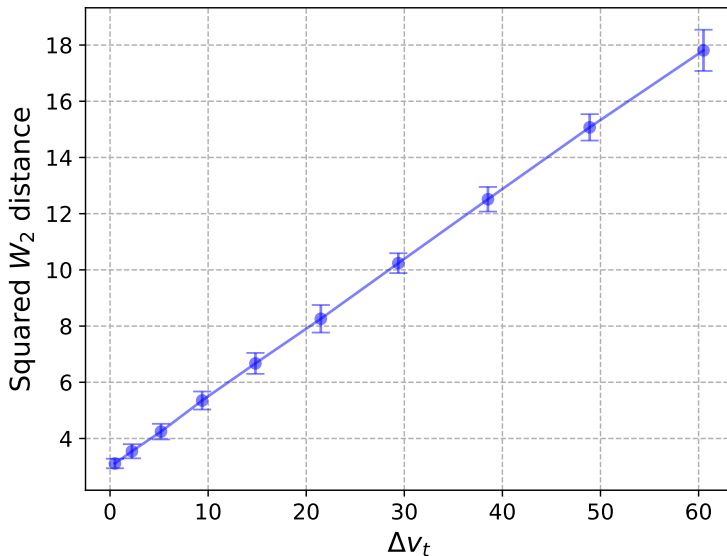
Figure 6: A linear relation between Δv_t and the squared Wasserstein-2 distance.

7. Related work

GIFs and the induced transport maps are related to CNFs and score-based diffusion models. Mathematically, they interrelate with the literature on Lipschitz mass transport and Wasserstein gradient flows. A central question in developing the ODE flow or transport map method for generative modeling is how to construct an ODE flow or transport map that are sufficiently smooth and enable efficient computation. Various approaches have been proposed to answer the question.

CNFs construct invertible mappings between an isotropic Gaussian distribution and a complex target distribution (Chen et al., 2018; Grathwohl et al., 2019). They fall within the broader framework of neural ODEs (Chen et al., 2018; Ruiz-Balet and Zuazua, 2023). A major challenge for CNFs is designing a time-dependent ODE flow whose marginal distribution converges to the target distribution while allowing for efficient estimation of its velocity field. Previous work has explored several principles to construct such flows, including optimal transport, Wasserstein gradient flows, and diffusion processes. Additionally, Gaussian denoising has emerged as an effective principle for constructing simulation-free CNFs in generative modeling.

Liu et al. (2023) propose the rectified flow, which is based on a linear interpolation between a standard Gaussian distribution and the target distribution, mimicking the Gaussian denoising procedure. Albergo and Vanden-Eijnden (2023) study a similar formulation called stochastic interpolation, defining a trigonometric interpolant between a standard Gaussian distribution and the target distribution. Albergo et al. (2023b) extend this idea by proposing a stochastic bridge interpolant between two arbitrary distributions. Under a few regularity assumptions, the velocity field of the ODE flow modeling the stochastic bridge interpolant is proven to be continuous in the time variable and smooth in the space variable.

Lipman et al. (2023) introduce a nonlinear least squares method called flow matching to directly estimate the velocity field of probability flow ODEs. All of these models are encompassed within the framework of simulation-free CNFs, which have been the focus of numerous ongoing research efforts (Neklyudov et al., 2023; Tong et al., 2023; Chen and Lipman, 2023; Albergo et al., 2023b; Shaul et al., 2023; Pooladian et al., 2023; Albergo et al., 2023a,c). Furthermore, Marzouk et al. (2023) provide the first statistical convergence rate for the simulation-based method by placing neural ODEs within the nonparametric estimation framework.

Score-based diffusion models integrate the time reversal of stochastic differential equations (SDEs) with the score matching technique (Sohl-Dickstein et al., 2015; Song and Ermon, 2019; Ho et al., 2020; Song and Ermon, 2020; Song et al., 2021b,a; De Bortoli et al., 2021). These models are capable of modeling highly complex probability distributions and have achieved state-of-the-art performance in image synthesis tasks (Dhariwal and Nichol, 2021; Rombach et al., 2022). The probability flow ODEs of diffusion models can be considered as CNFs, whose velocity field incorporates the nonlinear score function (Song et al., 2021b; Karras et al., 2022; Lu et al., 2022b,a; Zheng et al., 2023). In addition to the score matching method, Lu et al. (2022a) and Zheng et al. (2023) explore maximum likelihood estimation for probability flow ODEs. However, the regularity of these probability flow ODEs has not been studied and their well-posedness properties remain to be established.

A key concept in defining measure transport is Lipschitz mass transport, where the transport maps are required to be Lipschitz continuous. This ensures the smoothness and stability of the measure transport. There is a substantial body of research on the Lipschitz properties of transport maps. The celebrated Caffarelli's contraction theorem (Caffarelli, 2000, Theorem 2) establishes the Lipschitz continuity of optimal transport maps that push the standard Gaussian measure onto a log-concave measure. Colombo et al. (2017) study a Lipschitz transport map between perturbations of log-concave measures using optimal transport theory.

Mikulincer and Shenfeld (2021) demonstrate that the Brownian transport map, defined by the Föllmer process, is Lipschitz continuous when it pushes forward the Wiener measure on the Wiener space to the target measure on the Euclidean space. Additionally, Neeman (2022) and Mikulincer and Shenfeld (2023) prove that the transport map along the reverse heat flow of certain target measures is Lipschitz continuous.

Beyond studying Lipschitz transport maps, significant effort has been devoted to applying optimal transport theory in generative modeling. Zhang et al. (2018) propose the Monge-Ampe're flow for generative modeling by solving the linearized Monge-Ampère equation. Optimal transport theory has been utilized as a general principle to regularize the training of continuous normalizing flows or generators for generative modeling (Finlay et al., 2020; Yang and Karniadakis, 2020; Onken et al., 2021; Makkuva et al., 2020). Liang (2021) leverage the regularity theory of optimal transport to formalize the generator-discriminator-pair regularization of GANs under a minimax rate framework.

In our work, we study the Lipschitz transport maps defined by GIFs, which differ from the optimal transport map. GIFs naturally fit within the framework of continuous normalizing flows, and their flow mappings are examined from the perspective of Lipschitz mass transport.

Wasserstein gradient flows offer another principled approach to constructing ODE flows for generative modeling. A Wasserstein gradient flow is derived from the gradient descent minimization of a certain energy functional over probability measures endowed with the quadratic Wasserstein metric (Ambrosio et al., 2008). The Eulerian formulation of Wasserstein gradient flows produces the continuity equations that govern the evolution of marginal distributions. After transferred into a Lagrangian formulation, Wasserstein gradient flows define ODE flows that have been widely explored for generative modeling (Johnson and Zhang, 2018; Gao et al., 2019; Liutkus et al., 2019; Johnson and Zhang, 2019; Arbel et al., 2019; Mroueh et al., 2019; Ansari et al., 2021; Mroueh and Nguyen, 2021; Fan et al., 2022; Gao et al., 2022; Duncan et al., 2023; Xu et al., 2022). Wasserstein gradient flows are shown to be connected with the forward process of diffusion models. The variance preserving SDE of diffusion models is equivalent to the Langevin dynamics towards the standard Gaussian distribution that can be interpreted as a Wasserstein gradient flow of the Kullback–Leibler divergence for a standard Gaussian distribution (Song et al., 2021b). In the meantime, the probability flow ODE of the variance preserving SDE conforms to the Eulerian formulation of this Wasserstein gradient flow. However, when assigning a general distribution instead of the standard Gaussian distribution, it remains unclear whether the ODE formulation of Wasserstein gradient flows possesses well-posedness.

The main contribution of our work lies in establishing the theoretical properties of GIFs and their associated flow maps in a unified way. Our theoretical results encompass the Lipschitz continuity of both the flow velocity field and the flow map, addressing the existence, uniqueness, and stability of the flow. We also demonstrate that both the flow map and its inverse possess Lipschitz properties.

Our proposed framework for Gaussian interpolation flow builds upon previous research on probability flow methods in diffusion models (Song et al., 2021b,a) and stochastic interpolation methods for generative modeling (Liu et al., 2023; Albergo and Vanden-Eijnden, 2023; Lipman et al., 2023). Rather than adopting a methodological perspective, we focus on elucidating the theoretical aspects of these flows from a unified standpoint, thereby enhancing the understanding of various methodological approaches. Our theoretical results are derived from geometric considerations of the target distribution and from analytic calculations that exploit the Gaussian denoising property.

8. Conclusions and discussion

Gaussian denoising as a framework for constructing continuous normalizing flows holds great promise in generative modeling. Through a unified framework and rigorous analysis, we have established the well-posedness of these flows, shedding light on their capabilities and limitations. We have examined the Lipschitz regularity of the corresponding flow maps for several rich classes of probability measures. When applied to generative modeling based on Gaussian denoising, we have shown that GIFs possess auto-encoding and cycle consistency properties at the population level. Additionally, we have established stability error bounds for the errors accumulated during the process of learning GIFs.

The regularity properties of the velocity field established in this paper provide a solid theoretical basis for end-to-end error analyses of learning GIFs using deep neural networks with empirical data. Another potential application is to perform rigorous analyses of consis-

tency models, a nascent family of ODE-based deep generative models designed for one-step generation (Song et al., 2023; Kim et al., 2023; Song and Dhariwal, 2023). We intend to investigate these intriguing problems in our subsequent work. We expect that our analytical results will facilitate further studies and advancements in applying simulation-free CNFs, including GIFs, to a diverse range of generative modeling tasks.

Acknowledgements

Appendix

In the appendices, we prove the results stated in the paper and provide necessary technical details and discussions.

Appendix A. Proofs of Theorem 3.1 and Lemma 4.1

Dynamical properties of Gaussian interpolation flow $(\mathsf{X}_t)_{t\in[0,1]}$ form the cornerstone of the measure interpolation method. Following Albergo and Vanden-Eijnden (2023); Albergo et al. (2023b), we leverage an argument of characteristic functions to quantify the dynamics of its marginal flow, and in result, to prove Theorem 3.1.

Proof [Proof of Theorem 3.1] Let $\omega \in \mathbb{R}^d$. For the Gaussian stochastic interpolation $(\mathsf{X}_t)_{t\in[0,1]}$, we define the characteristic function of X_t by

$$\Psi(t,\omega) := \mathbb{E}[\exp(i\langle\omega,\mathsf{X}_t\rangle)] = \mathbb{E}[\exp(i\langle\omega,a_t\mathsf{Z}+b_t\mathsf{X}_1\rangle)] = \mathbb{E}[\exp(ia_t\langle\omega,\mathsf{Z}\rangle)]\mathbb{E}[\exp(ib_t\langle\omega,\mathsf{X}_1\rangle)],$$

where the last equality is due to the independence of between $\mathsf{Z} \sim \gamma_d$ and $\mathsf{X}_1 \sim \nu$. Taking the time derivative of $\Psi(t,\omega)$ for $t \in (0,1)$, we derive that

$$\partial_t\Psi(t,\omega) = i\langle\omega,\psi(t,\omega)\rangle$$

where

$$\psi(t,\omega) := \mathbb{E}[\exp(i\langle\omega,\mathsf{X}_t\rangle)(\dot{a}_t\mathsf{Z}+\dot{b}_t\mathsf{X}_1)].$$

We first define

$$v(t,\mathsf{X}_t) := \mathbb{E}[\dot{a}_t\mathsf{Z}+\dot{b}_t\mathsf{X}_1|\mathsf{X}_t]. \tag{A.1}$$

Using the double expectation formula, we deduce that

$$\psi(t,\omega) = \mathbb{E}[\exp(i\langle\omega,\mathsf{X}_t\rangle)\mathbb{E}[\dot{a}_t\mathsf{Z}+\dot{b}_t\mathsf{X}_1|\mathsf{X}_t]] = \mathbb{E}[\exp(i\langle\omega,\mathsf{X}_t\rangle)v(t,\mathsf{X}_t)].$$

Applying the inverse Fourier transform to $\psi(t,\omega)$, it holds that

$$j(t,x) := (2\pi)^{-d}\int_{\mathbb{R}^d}\exp(-i\langle\omega,x\rangle)\psi(t,\omega)\mathrm{d}\omega = p_t(x)v(t,x),$$

where $v(t,x) := \mathbb{E}[\dot{a}_t\mathsf{Z}+\dot{b}_t\mathsf{X}_1|\mathsf{X}_t = x]$. Then it further yields that

$$\partial_t p_t + \nabla_x \cdot j(t,x) = 0,$$

that is,

$$\partial_t p_t + \nabla_x \cdot (p_t v(t,x)) = 0.$$

Next, we study the property of $v(t,x)$ at $t = 0$ and $t = 1$. Notice that

$$x = a_t\mathbb{E}[\mathsf{Z}|\mathsf{X}_t = x] + b_t\mathbb{E}[\mathsf{X}_1|\mathsf{X}_t = x]. \tag{A.2}$$

Combining Eq. (A.1) and (A.2), it implies that

$$v(t,x) = \frac{\dot{a}_t}{a_t}x + \left(\dot{b}_t - \frac{\dot{a}_t}{a_t}b_t\right)\mathbb{E}[\mathsf{X}_1|\mathsf{X}_t = x], \quad t \in (0,1). \tag{A.3}$$

According to Tweedie's formula in Lemma G.1, it holds that

$$s(t, x) = \frac{b_t}{a_t^2} \mathbb{E}\left[\mathsf{X}_1 | \mathsf{X}_t = x\right] - \frac{1}{a_t^2} x, \quad t \in (0, 1), \tag{A.4}$$

where $s(t, x)$ is the score function of the marginal distribution of $\mathsf{X}_t \sim p_t$.

Combining Eq. (A.3), (A.4), it holds that the velocity field is a gradient field and its nonlinear term is the score function $s(t, x)$, namely, for any $t \in (0, 1)$,

$$v(t, x) = \frac{\dot{b}_t}{b_t} x + \left(\frac{\dot{b}_t}{b_t} a_t^2 - \dot{a}_t a_t\right) s(t, x). \tag{A.5}$$

By the regularity properties that $a_t, b_t \in C^2([0, 1)), a_t^2 \in C^1([0, 1]), b_t \in C^1([0, 1])$, we have that $\dot{a}_0, \dot{b}_0, \dot{a}_1 a_1$, and $\dot{b}_1$ are well-defined. Then by Eq. (A.3), we define that

$$v(0, x) := \lim_{t \downarrow 0} v(t, x) = \frac{\dot{a}_0}{a_0} x + \left(\dot{b}_0 - \frac{\dot{a}_0}{a_0} b_0\right) \mathbb{E}[\mathsf{X}_1 | \mathsf{X}_0 = x]$$

Using Eq. (A.5) yields that

$$v(1, x) := \lim_{t \uparrow 1} v(t, x) = \frac{\dot{b}_1}{b_1} x - \dot{a}_1 a_1 s(1, x). \tag{A.6}$$

This completes the proof. ∎

Lemma 4.1 presents several standard properties of Gaussian channels in information theory (Wibisono and Jog, 2018a,b; Dytso et al., 2023b) that will facilitate our proof.
Proof [Proof of Lemma 4.1] By Bayes' rule, $\mathrm{Law}(\mathsf{Y}|\mathsf{X}_t = x) = p(y|t, x)$ can be represented as

$$
\begin{aligned}
p(y|t, x) &= \varphi_{b_t y, a_t^2}(x) p_1(y) / p_t(x) \\
&= (2\pi)^{-d/2} a_t^{-d} \exp\left(-\frac{\|x - b_t y\|^2}{2 a_t^2}\right) p_1(y) / p_t(x) \\
&= (2\pi)^{-d/2} a_t^{-d} \exp\left(-\frac{\|x\|^2}{2 a_t^2} + \frac{b_t \langle x, y \rangle}{a_t^2} - \frac{b_t^2 \|y\|^2}{2 a_t^2}\right) p_1(y) / p_t(x) \\
&= \left\{\exp\left(\frac{b_t \langle x, y \rangle}{a_t^2} - \frac{b_t^2 \|y\|^2}{2 a_t^2}\right) p_1(y)\right\} / \left\{(2\pi)^{d/2} a_t^d \exp\left(\frac{\|x\|^2}{2 a_t^2}\right) p_t(x)\right\}.
\end{aligned}
$$

Let $\theta = \frac{b_t x}{a_t^2}, h(y) = p_1(y) \exp(-\frac{b_t^2 \|y\|^2}{2 a_t^2})$, and the logarithmic partition function

$$A(\theta) = \log \int_{\mathbb{R}^d} h(y) \exp(\langle y, \theta \rangle) \mathrm{d}y,$$

then by the definition of exponential family distributions, we conclude that

$$p(y|t, x) = h(y) \exp(\langle y, \theta \rangle - A(\theta))$$

is an exponential family distribution of y. By simple calculation, it follows that

$$\nabla_x^2 \log p(y|t, x) = -\frac{b_t^2}{a_t^4} \nabla_\theta^2 A(\theta).$$

For an exponential family distribution, a basic equality shows that

$$\nabla_\theta^2 A(\theta) = \text{Cov}(\mathsf{Y}|\mathsf{X}_t = x),$$

which further yields that $\nabla_x^2 \log p(y|t,x) = -\frac{b_t^2}{a_t^4}\text{Cov}(\mathsf{Y}|\mathsf{X}_t = x)$. ∎

Appendix B. Auxiliary lemmas for Lipschitz flow maps

The following lemma, due to G. Peano (Hartman, 2002a, Theorem 3.1), describes several meaningful differential equations associated with well-posed flows and supports the derivation of Lipschitz continuity of their flow maps.

Lemma B.1 *(Ambrosio et al., 2023, Lemma 3.4) Suppose that a flow $(X_t)_{t\in[0,1]}$ is well-posed and its velocity field $v(t,x) : [0,1] \times \mathbb{R}^d \to \mathbb{R}^d$ is of class C^1. Then the flow map $X_{s,t} : \mathbb{R}^d \to \mathbb{R}^d$ is of class C^1 for any $0 \le s \le t \le 1$. Fix $(s,x) \in [0,1] \times \mathbb{R}^d$ and set the following functions defined with $t \in [s,1]$*

$$y(t) := \nabla_x X_{s,t}(x), \qquad\qquad J(t) := (\nabla_x v)(t, X_{s,t}(x)),$$
$$w(t) := \det(\nabla_x X_{s,t}(x)), \qquad b(t) := (\nabla_x \cdot v)(t, X_{s,t}(x)) = \text{Tr}(J(t)).$$

Then $y(t)$ and $w(t)$ are the unique C^1 solutions of the following IVPs

$$\dot{y}(t) = J(t)y(t), \quad y(s) = \mathbf{I}_d, \tag{B.1}$$
$$\dot{w}(t) = b(t)w(t), \quad w(s) = 1. \tag{B.2}$$

We present an upper bound of the Lipschitz constant of its flow map $X_{s,t}(x)$ in Lemma 5.1. The upper bound has been deduced in Mikulincer and Shenfeld (2023); Ambrosio et al. (2023); Dai et al. (2023). For completeness, we derive it as a direct implication of Eq. (B.1) in Lemma B.1 and an upper bound of the Jacobian matrix of the velocity field.
Proof [Proof of Lemma 5.1] Let $y(u) = \nabla_x X_{s,u}(x)$, $J(u) = (\nabla_x v)(u, X_{s,u}(x))$. Owing to Lemma B.1, $y(u)$ is of class C^1, and the function $u \mapsto \|y(u)\|_{2,2}$ is absolutely continuous over $[s,t]$. By Lemma B.1, it follows that

$$\partial_u \|y(u)\|_{2,2}^2 = 2\langle y(u), \dot{y}(u)\rangle = 2\langle y(u), J(u)y(u)\rangle \le 2\theta_u \|y(u)\|_{2,2}^2.$$

Applying Grönwall's inequality yields that $\|y(t)\|_{2,2} \le \exp(\int_s^t \theta_u du)$ which concludes the proof. ∎

Another result is concerning the theorem of instantaneous change of variables that is widely deployed in studying neural ODEs (Chen et al., 2018, Theorem 1). We also exploit the instantaneous change of variables to prove Proposition 6.1. To make the proof self-contained, we show that the instantaneous change of variables directly follows Eq. (B.2) in Lemma B.1. Compared with the original proof in (Chen et al., 2018, Theorem 1), we illustrate that the well-posedness of a flow is sufficient to ensure the instantaneous change of variables property, without a boundedness condition on the flow.

Corollary B.1 (Instantaneous change of variables) *Suppose that a flow $(X_t)_{t\in[0,1]}$ is well-posed with a velocity field $v(t,x) : [0,1] \times \mathbb{R}^d \to \mathbb{R}^d$ of class C^1 in x. Let $X_0(x) \sim \pi_0(X_0(x))$ be a distribution of the initial value. Then the law of $X_t(x)$ satisfies the following differential equation*

$$\partial_t \log \pi_t(X_t(x)) = -\operatorname{Tr}((\nabla_x v)(t, X_t(x))).$$

Proof Let $\delta(t) := \det(\nabla_x X_t(x))$. Thanks to Eq. (B.2) in Lemma B.1, it holds that

$$\dot{\delta}(t) = \operatorname{Tr}((\nabla_x v)(t, X_t(x)))\delta(t), \quad \delta(0) = 1,$$

which implies $\delta(t) > 0$ for $t \in [0,1]$. Notice that $\log \pi_t(X_t(x)) = \log \pi_0(X_0(x)) - \log |\delta(t)|$ by change of variables. Then it follows that $\partial_t \log \pi_t(X_t(x)) = -\operatorname{Tr}((\nabla_x v)(t, X_t(x)))$. ∎

Appendix C. Proofs of spatial Lipschitz estimates for the velocity field

The main results in Section 4 are proved in this appendix. We first present some ancillary lemmas before proceeding to give the proofs.

Lemma C.1 (Fathi et al., 2023) *Suppose that $f : \mathbb{R}^d \to \mathbb{R}_+$ is L-log-Lipschitz for some $L \geq 0$. Let $\mathcal{P}_t$ be the Ornstein–Uhlenbeck semigroup defined by $\mathcal{P}_t h(x) := \mathbb{E}_{Z \sim \gamma_d}[h(e^{-t}x + \sqrt{1 - e^{-2t}}Z)]$ for any $h \in C(\mathbb{R}^d)$ and $t \geq 0$. Then it holds that*

$$\left\{-5Le^{-t}(L + t^{-\frac{1}{2}}) - L^2 e^{-2t}\right\} \mathbf{I}_d \preceq \nabla_x^2 \log \mathcal{P}_t f(x) \preceq \left\{5Le^{-t}(L + t^{-\frac{1}{2}})\right\} \mathbf{I}_d.$$

Proof This is a restatement of known results. See Proposition 2, Proposition 6, Theorem 6, and their proofs in Fathi et al. (2023). ∎

Corollary C.1 *Suppose that $f : \mathbb{R}^d \to \mathbb{R}_+$ is L-log-Lipschitz for some $L \geq 0$. Let $\mathcal{Q}_t$ be an operator defined by*

$$\mathcal{Q}_t h(x) := \mathbb{E}_{Z \sim \gamma_d}[h(\beta_t x + \alpha_t Z)] \tag{C.1}$$

for any $h \in C(\mathbb{R}^d)$ and $t \in [0,1]$ where $0 \leq \alpha_t \leq 1, \beta_t \geq 0$ for any $t \in [0,1]$. Then it holds that

$$\left(-A_t - L^2 \beta_t^2\right) \mathbf{I}_d \preceq \nabla_x^2 \log \mathcal{Q}_t f(x) \preceq A_t \mathbf{I}_d,$$

where $A_t := 5L\beta_t^2(1 - \alpha_t^2)^{-\frac{1}{2}}(L + (-\frac{1}{2}\log(1 - \alpha_t^2))^{-\frac{1}{2}})$.

Proof It is easy to notice that $\mathcal{Q}_t f(x) = \mathcal{P}_s f(\beta_t e^s x)$ where $s = -\frac{1}{2}\log(1 - \alpha_t^2)$. Then it follows that $\nabla_x^2 \log \mathcal{Q}_t f(x) = (\beta_t e^s)^2 (\nabla_x^2 \log \mathcal{P}_s f)(\beta_t e^s x)$ which yields

$$\left(-A_t - L^2 \beta_t^2\right) \mathbf{I}_d \preceq \nabla_x^2 \log \mathcal{Q}_t f(x) \preceq A_t \mathbf{I}_d,$$

where $A_t := 5L\beta_t^2(1 - \alpha_t^2)^{-\frac{1}{2}}(L + (-\frac{1}{2}\log(1 - \alpha_t^2))^{-\frac{1}{2}})$. ∎

Lemma C.2 *The Jacobian matrix of the velocity field (3.5) has an alternative expression over time $t \in (0,1)$, that is,*

$$\nabla_x v(t,x) = \left(\frac{\dot{b}_t}{b_t} a_t^2 - \dot{a}_t a_t \right) \left(\nabla_x^2 \log \widetilde{\mathcal{Q}}_t f(x) - \frac{1}{a_t^2 + b_t^2} \mathbf{I}_d \right) + \frac{\dot{b}_t}{b_t} \mathbf{I}_d,$$

where $f(x) := \frac{\mathrm{d}\nu}{\mathrm{d}\gamma_d}(x)$ and $\widetilde{\mathcal{Q}}_t f(x) := \mathbb{E}_{\mathsf{Z} \sim \gamma_d}[f(\frac{b_t}{a_t^2 + b_t^2} x + \frac{a_t}{\sqrt{a_t^2 + b_t^2}} \mathsf{Z})]$.

Proof By direct calculations, it holds that

$$p_t(x) = a_t^{-d} \int_{\mathbb{R}^d} p_1(y) \varphi\left(\frac{x - b_t y}{a_t} \right) \mathrm{d}y = a_t^{-d} \int_{\mathbb{R}^d} f(y) \varphi(y) \varphi\left(\frac{x - b_t y}{a_t} \right) \mathrm{d}y$$

$$= a_t^{-d} \varphi\left((a_t^2 + b_t^2)^{-\frac{1}{2}} x \right) \int_{\mathbb{R}^d} f(y) \varphi\left(\left(\frac{a_t}{\sqrt{a_t^2 + b_t^2}} \right)^{-1} \left(y - \frac{b_t}{a_t^2 + b_t^2} x \right) \right) \mathrm{d}y$$

$$= a_t^{-d} \varphi\left((a_t^2 + b_t^2)^{-\frac{1}{2}} x \right) \left(\frac{a_t}{\sqrt{a_t^2 + b_t^2}} \right)^d \int_{\mathbb{R}^d} f\left(\frac{b_t}{a_t^2 + b_t^2} x + \frac{a_t}{\sqrt{a_t^2 + b_t^2}} z \right) \mathrm{d}\gamma_d(z)$$

$$= (a_t^2 + b_t^2)^{-d/2} \varphi\left((a_t^2 + b_t^2)^{-\frac{1}{2}} x \right) \widetilde{\mathcal{Q}}_t f(x).$$

Taking the logarithm and then the second-order derivative of the equation above, it yields

$$\nabla_x s(t,x) = \nabla_x^2 \log \widetilde{\mathcal{Q}}_t f(x) - \frac{1}{a_t^2 + b_t^2} \mathbf{I}_d.$$

Recalling that $\nabla_x v(t,x) = \left(\frac{\dot{b}_t}{b_t} a_t^2 - \dot{a}_t a_t \right) \nabla_x s(t,x) + \frac{\dot{b}_t}{b_t} \mathbf{I}_d$, it further yields that

$$\nabla_x v(t,x) = \left(\frac{\dot{b}_t}{b_t} a_t^2 - \dot{a}_t a_t \right) \nabla_x^2 \log \widetilde{\mathcal{Q}}_t f(x) + \frac{\dot{a}_t a_t + \dot{b}_t b_t}{a_t^2 + b_t^2} \mathbf{I}_d,$$

which completes the proof. $\blacksquare$

Corollary C.2 *Suppose that $f(x) := \frac{\mathrm{d}\nu}{\mathrm{d}\gamma_d}(x)$ is L-log-Lipschitz for some $L \geq 0$. Then for $t \in (0,1)$, it holds that*

$$\left\{ \left(\frac{\dot{b}_t}{b_t} a_t^2 - \dot{a}_t a_t \right) \left(-B_t - L^2 \left(\frac{b_t}{a_t^2 + b_t^2} \right)^2 \right) + \frac{\dot{a}_t a_t + \dot{b}_t b_t}{a_t^2 + b_t^2} \right\} \mathbf{I}_d$$

$$\preceq \nabla_x v(t,x) \preceq \left\{ \left(\frac{\dot{b}_t}{b_t} a_t^2 - \dot{a}_t a_t \right) B_t + \frac{\dot{a}_t a_t + \dot{b}_t b_t}{a_t^2 + b_t^2} \right\} \mathbf{I}_d,$$

where $B_t := 5Lb_t(a_t^2 + b_t^2)^{-\frac{3}{2}}(L + (\log(\sqrt{a_t^2 + b_t^2}/b_t))^{-\frac{1}{2}})$.

Proof Let $\alpha_t = \frac{a_t}{\sqrt{a_t^2 + b_t^2}}$ and $\beta_t = \frac{b_t}{a_t^2 + b_t^2}$. Then these bounds hold according to Corollary C.1 and Lemma C.2. $\blacksquare$

Then we are prepared to prove Proposition 4.1. The proof is mainly based on the techniques for bounding conditional covariance matrices that are developed in a series of work (Wibisono and Jog, 2018a,b; Mikulincer and Shenfeld, 2021, 2023; Chewi and Pooladian, 2022; Dai et al., 2023).

Proof [Proof of Proposition 4.1]

(a) By Jung's theorem (Danzer et al., 1963, Theorem 2.6), there exists a closed Euclidean ball with radius less than $D := (1/\sqrt{2})\mathrm{diam}(\mathrm{supp}(\nu))$ that contains $\mathrm{supp}(\nu)$ in $\mathbb{R}^d$. Then the desired bounds hold due to $0\mathbf{I}_d \preceq \mathrm{Cov}(\mathsf{Y}|\mathsf{X}_t = x) \preceq D^2\mathbf{I}_d$ and Eq. (4.3).

(b) Let p_1 be β-semi-log-convex for some $\beta > 0$ on $\mathbb{R}^d$. Then for any $t \in [0,1)$, the conditional distribution $p(y|t,x)$ is $\left(\beta + \frac{b_t^2}{a_t^2}\right)$-semi-log-convex because

$$-\nabla_y^2 \log p(y|t,x) = -\nabla_y^2 \log p_1(y) - \nabla_y^2 \log p(t,x|y) \preceq \left(\beta + \frac{b_t^2}{a_t^2}\right)\mathbf{I}_d.$$

By the Cramér-Rao inequality (2.4), we obtain

$$\mathrm{Cov}(\mathsf{Y}|\mathsf{X}_t = x) \succeq \left(\beta + \frac{b_t^2}{a_t^2}\right)^{-1}\mathbf{I}_d.$$

Therefore, by Eq. (4.3), we obtain

$$\nabla_x v(t,x) \succeq \left\{\left(\frac{\dot{b}_t}{b_t} - \frac{\dot{a}_t}{a_t}\right)\frac{b_t^2}{\beta a_t^2 + b_t^2} + \frac{\dot{a}_t}{a_t}\right\}\mathbf{I}_d,$$

which implies

$$\nabla_x v(t,x) \succeq \frac{\beta a_t \dot{a}_t + b_t \dot{b}_t}{\beta a_t^2 + b_t^2}\mathbf{I}_d.$$

In addition, the bound above can be verified at time $t = 1$ by the definition (A.6).

(c) Let p_1 be κ-semi-log-concave for some $\kappa \in \mathbb{R}$. Then for any $t \in [0,1)$, the conditional distribution $p(y|t,x)$ is $\left(\kappa + \frac{b_t^2}{a_t^2}\right)$-semi-log-concave because

$$-\nabla_y^2 \log p(y|t,x) = -\nabla_y^2 \log p_1(y) - \nabla_y^2 \log p(t,x|y) \succeq \left(\kappa + \frac{b_t^2}{a_t^2}\right)\mathbf{I}_d.$$

When $t \in \left\{t : \kappa + \frac{b_t^2}{a_t^2} > 0, t \in (0,1)\right\}$, by the Brascamp-Lieb inequality (2.2), we obtain

$$\mathrm{Cov}(\mathsf{Y}|\mathsf{X}_t = x) \preceq \left(\kappa + \frac{b_t^2}{a_t^2}\right)^{-1}\mathbf{I}_d.$$

Therefore, by Eq. (4.3), we obtain

$$\nabla_x v(t,x) \preceq \left\{\left(\frac{\dot{b}_t}{b_t} - \frac{\dot{a}_t}{a_t}\right)\frac{b_t^2}{\kappa a_t^2 + b_t^2} + \frac{\dot{a}_t}{a_t}\right\}\mathbf{I}_d,$$

which implies

$$\nabla_x v(t,x) \preceq \frac{\kappa a_t \dot{a}_t + b_t \dot{b}_t}{\kappa a_t^2 + b_t^2}\mathbf{I}_d.$$

Moreover, the bound above can be verified at time $t = 1$ by the definition (A.6).

(d) Notice that

$$p(y|t,x) = \frac{p(t,x|y)}{p_t(x)} \frac{\mathrm{d}(\gamma_{d,\sigma^2} * \rho)}{\mathrm{d}y}$$

$$= A_{x,t} \int_{\mathbb{R}^d} \varphi_{z,\sigma^2}(y) \varphi_{\frac{x}{b_t}, \frac{a_t^2}{b_t^2}}(y) \rho(\mathrm{d}z),$$

where the prefactor $A_{x,t}$ only depends on x and t. Then it follows that

$$p(y|t,x) = \int_{\mathbb{R}^d} \varphi_{\frac{a_t^2 z + \sigma^2 b_t x}{a_t^2 + \sigma^2 b_t^2}, \frac{\sigma^2 a_t^2}{a_t^2 + \sigma^2 b_t^2}}(y) \tilde{\rho}(\mathrm{d}z)$$

where $\tilde{\rho}$ is a probability measure on $\mathbb{R}^d$ whose density function is a multiple of ρ by a positive function. It also indicates that $\tilde{\rho}$ is supported on the same Euclidean ball as ρ. To further illustrate $p(y|t,x)$, let $\mathsf{Q} \sim \tilde{\rho}$ and $\mathsf{Z} \sim \gamma_d$ be independent. Then it holds that

$$\frac{a_t^2}{a_t^2 + \sigma^2 b_t^2} \mathsf{Q} + \sqrt{\frac{\sigma^2 a_t^2}{a_t^2 + \sigma^2 b_t^2}} \mathsf{Z} + \frac{\sigma^2 b_t}{a_t^2 + \sigma^2 b_t^2} x \sim p(y|t,x).$$

Thus, it holds that

$$\mathrm{Cov}(\mathsf{Y}|\mathsf{X}_t = x) = \left(\frac{a_t^2}{a_t^2 + \sigma^2 b_t^2}\right)^2 \mathrm{Cov}(\mathsf{Q}) + \frac{\sigma^2 a_t^2}{a_t^2 + \sigma^2 b_t^2}\mathbf{I}_d$$

$$\preceq \left\{ \left(\frac{a_t^2}{a_t^2 + \sigma^2 b_t^2}\right)^2 R^2 + \frac{\sigma^2 a_t^2}{a_t^2 + \sigma^2 b_t^2} \right\} \mathbf{I}_d.$$

By Eq. (4.3), it holds that

$$\nabla_x v(t,x) \preceq \frac{b_t^2}{a_t^2}\left(\frac{\dot{b}_t}{b_t} - \frac{\dot{a}_t}{a_t}\right)\left(\left(\frac{a_t^2}{a_t^2 + \sigma^2 b_t^2}\right)^2 R^2 + \frac{\sigma^2 a_t^2}{a_t^2 + \sigma^2 b_t^2}\right)\mathbf{I}_d + \frac{\dot{a}_t}{a_t}\mathbf{I}_d,$$

which implies

$$\nabla_x v(t,x) \preceq \left\{ \frac{a_t b_t(a_t \dot{b}_t - \dot{a}_t b_t)}{(a_t^2 + \sigma^2 b_t^2)^2}R^2 + \frac{\dot{a}_t a_t + \sigma^2 \dot{b}_t b_t}{a_t^2 + \sigma^2 b_t^2} \right\} \mathbf{I}_d.$$

Analogously, due to $\mathrm{Cov}(\mathsf{Q}) \succeq 0\mathbf{I}_d$, a lower bound would be yielded as follows

$$\nabla_x v(t,x) \succeq \frac{\dot{a}_t a_t + \sigma^2 \dot{b}_t b_t}{a_t^2 + \sigma^2 b_t^2}\mathbf{I}_d.$$

Then the results follow by combining the upper and lower bounds.

(e) The result follows from Corollary C.2.

We complete the proof. ■

Proof [Proof of Corollary 4.1] Let us consider that $\kappa > 0$ which is divided into two cases where $\kappa D^2 \geq 1$ and $\kappa D^2 < 1$. On the one hand, suppose that the first case $\kappa D^2 \geq 1$ holds. By Proposition 4.1, the κ-based upper bound is tighter, that is,

$$\lambda_{\max}(\nabla_x v(t,x)) \leq \theta_t := \frac{\kappa a_t \dot{a}_t + b_t \dot{b}_t}{\kappa a_t^2 + b_t^2}.$$

On the other hand, suppose that the second case $\kappa D^2 < 1$ holds. Let t_1 be defined in Eq. (4.5). Again, by Proposition 4.1, the D^2-based upper bound is tighter over $[0, t_1)$ and the κ-based upper bound is tighter over $[t_1, 1]$, which is denoted by

$$\lambda_{\max}(\nabla_x v(t,x)) \leq \theta_t := \begin{cases} \frac{b_t^2}{a_t^2}\left(\frac{\dot{b}_t}{b_t} - \frac{\dot{a}_t}{a_t}\right) D^2 + \frac{\dot{a}_t}{a_t}, & t \in [0, t_1), \\ \frac{\kappa a_t \dot{a}_t + b_t \dot{b}_t}{\kappa a_t^2 + b_t^2}, & t \in [t_1, 1]. \end{cases}$$

This completes the proof. ■

Proof [Proof of Corollary 4.2] Let $\kappa < 0, D < \infty$ such that $\kappa D^2 < 1$ is fulfilled. Then an argument similar to the proof of Corollary 4.1 yields the desired bounds. ■

Proof [Proof of Corollary 4.3] The result follows from Proposition 4.1-(d). ■

Proof [Proof of Corollary 4.4] The L-based upper and lower bounds in Proposition 4.1-(e) would blow up at time $t = 1$ because the term $(\log(\sqrt{a_t^2 + b_t^2}/b_t))^{-\frac{1}{2}}$ in B_t goes to ∞ as $t \to 1$. To ensure the spatial derivative of the velocity field $v(t,x)$ is upper bounded at time $t = 1$, we additionally require the target measure is κ-semi-log-concave with $\kappa \leq 0$. Hence, a κ-based upper bound is available for $t \in (t_0, 1]$ as shown in Proposition 4.1-(c). Next, these two upper bounds are combined by choosing any $t_2 \in (t_0, 1)$ first. Then we exploit the L-based bound over $[0, t_2)$ and κ-based bound over $[t_0, 1]$. This completes the proof. ■

Appendix D. Proofs of well-posedness and Lipschitz flow maps

The proofs of main results in Section 5 are offered in the following. Before proceeding, let us introduce some definitions and notations about function spaces that are collected in (Evans, 2010, Chapter 5). Let $L^1_{\text{loc}}(\mathbb{R}^d; \mathbb{R}^\ell) := \{\text{locally integrable function } u : \mathbb{R}^d \to \mathbb{R}^\ell\}$. For integers $k \geq 0$ and $1 \leq p \leq \infty$, we define the Sobolev space $W^{k,p}(\mathbb{R}^d) := \{u \in L^1_{\text{loc}}(\mathbb{R}^d) | D^\alpha u \text{ exists and } D^\alpha u \in L^p(\mathbb{R}^d) \text{ for } |\alpha| \leq k\}$, where $D^\alpha u$ is the weak derivative of u. Then the local Sobolev space $W^{k,p}_{\text{loc}}(\mathbb{R}^d)$ is defined as the function space such that for any $u \in W^{k,p}_{\text{loc}}(\mathbb{R}^d)$ and any compact set $\Omega \subset \mathbb{R}^d$, $u \in W^{k,p}(\Omega)$. As a result, we denote the vector-valued local Sobolev space by $W^{k,p}_{\text{loc}}(\mathbb{R}^d; \mathbb{R}^\ell)$. Provided that $v(t,x) : [0,1] \times \mathbb{R}^d \to \mathbb{R}^d$, we use $v \in L^1([0,1]; W^{1,\infty}_{\text{loc}}(\mathbb{R}^d; \mathbb{R}^d))$ to indicate that v has a finite L^1 norm over $(t,x) \in [0,1] \times \mathbb{R}^d$

and $v(t, \cdot) \in W^{1,\infty}_{\text{loc}}(\mathbb{R}^d; \mathbb{R}^d)$ for any $t \in [0, 1]$. Similarly, we say $v \in L^1([0, 1]; L^\infty(\mathbb{R}^d; \mathbb{R}^d))$ when v has a finite L^1 norm over $(t, x) \in [0, 1] \times \mathbb{R}^d$ and $v(t, \cdot) \in L^\infty(\mathbb{R}^d; \mathbb{R}^d)$ for every $t \in [0, 1]$. We will use the definitions and notations in the following proof.

Proof [Proof of Theorem 5.1] Under Assumptions 1 and 2, we claim that the velocity field $v(t, x)$ satisfies

$$v \in L^1([0, 1]; W^{1,\infty}_{\text{loc}}(\mathbb{R}^d; \mathbb{R}^d)), \quad \frac{\|v\|_2}{1 + \|x\|_2} \in L^1([0, 1]; L^\infty(\mathbb{R}^d; \mathbb{R}^d)).$$

where the first condition indicates the velocity field v is locally bounded and locally Lipschitz continuous in x, and the second condition is a growth condition on v. According to the Cauchy-Lipschitz theorem (Ambrosio and Crippa, 2014, Remark 2.4), we have the representation formulae for solutions of the continuity equation. As a result, there exists a flow $(X_t)_{t \in [0,1]}$ uniquely solves the IVP (3.10). Furthermore, the marginal flow of $(X_t)_{t \in [0,1]}$ satisfies the continuity equation (3.4) in the weak sense. Then it remains to show the velocity field v is locally bounded and locally Lipschitz continuous in x, and satisfies the growth condition. By the lower and upper bounds given in Proposition 4.1, we know that v is globally Lipschitz continuous in x under Assumptions 1 and 2. Indeed, the global Lipschitz continuity leads to local boundedness and linear growth properties by simple arguments. More concretely, for any $t \in (0, 1)$, it holds that

$$v(t, 0) = \left(\dot{b}_t - \frac{\dot{a}_t}{a_t} b_t \right) \mathbb{E}[X_1 | X_t = 0] = \left(\dot{b}_t - \frac{\dot{a}_t}{a_t} b_t \right) \int_{\mathbb{R}^d} y p(y | t, 0) \mathrm{d}y$$

$$\lesssim \left(\dot{b}_t - \frac{\dot{a}_t}{a_t} b_t \right) \int_{\mathbb{R}^d} y p_1(y) a_t^{-d} \exp \left(-\frac{b_t^2 \|y\|_2^2}{2 a_t^2} \right) \mathrm{d}y,$$

which implies $\|v(t, 0)\|_2 < \infty$ due to fast growth of the exponential function. Besides, it holds that $v(0, 0) = (\dot{b}_0 - \frac{\dot{a}_0}{a_0} b_0) \mathbb{E}[X_1 | X_0 = x] < \infty, v(1, 0) = \dot{a}_1 a_1 s(1, 0) < \infty$. Then by the boundedness of $\|v(t, 0)\|_2$ and the global Lipschitz continuity in x over $t \in [0, 1]$, we bound $v(t, x)$ as follows

$$\|v(t, x)\|_2 \leq \|v(t, 0)\|_2 + \|v(t, x) - v(t, 0)\|_2$$

$$\leq \|v(t, 0)\|_2 + \left\{ \sup_{(t,y) \in [0,1] \times \mathbb{R}^d} \|\nabla_y v(t, y)\|_{2,2} \right\} \|x\|_2$$

$$\lesssim \max\{\|x\|_2, 1\}.$$

Hence, the local boundedness and linear growth properties of v are proved. This completes the proof. ∎

Proof [Proof of Theorem 5.2] The proof is similar to that of Theorem 5.1. ∎

Proof [Proof of Corollary 5.1] A well-posed ODE flow has the time-reversal symmetry (Lamb and Roberts, 1998). By Theorem 5.1, the desired results are proved. ∎

Proof [Proof of Corollary 5.2] The proof is similar to that of Corollary 5.1. ∎

Proof [Proof of Proposition 5.1] Combining Proposition 4.1-(b), (c), and Lemma 5.1, we complete the proof. ∎

Proof [Proof of Proposition 5.2] Combining Proposition 4.1-(d) and Lemma 5.1, we complete the proof. ∎

Proof [Proof of Corollary 6.1] By Theorem 5.1 and Corollary 5.1, it holds that

$$X_1 \circ X_1^* = X_1 \circ X_1^{-1} = \mathbf{I}_d.$$

This completes the proof. ∎

Proof [Proof of Corollary 6.2] By Theorem 5.1 and Corollary 5.1, it holds that

$$X_{1,1} \circ X_{2,1}^* \circ X_{2,1} \circ X_{1,1}^* = X_{1,1} \circ X_{2,1}^{-1} \circ X_{2,1} \circ X_{1,1}^{-1} = \mathbf{I}_d.$$

This completes the proof. ∎

Proof [Proof of Corollary 6.3] Let Assumptions 2.1 and 2.2 hold. According to Propositions 5.1 and 5.2, $\|\nabla_x X_1(x)\|_{2,2}$ is uniformly bounded for Case (i)-(iii) in Assumption 2.2. For Case (iv), the boundedness of $\|\nabla_x X_1(x)\|_{2,2}$ holds by combining Corollary 4.4 and Lemma 5.1. Using Proposition 4.1, we know that $\|\nabla_x v(t,x)\|_{2,2}$ is uniformly bounded. ∎

Proof [Proof of Proposition 5.3] The proof idea is similar to those of (Ball et al., 2003, Proposition 1) and (Cattiaux and Guillin, 2014, Proposition 18). Let $f : \Omega \to \mathbb{R}$ be of class C^1 and $\mathsf{X}_t \sim p_t$. First, we consider the case of log-Sobolev inequalities. Using that $\mathsf{Z} \sim \gamma_d$ and $\mathsf{X}_1 \sim \nu$ both satisfy the log-Sobolev inequalities in Definition G.1, we have

$$\mathbb{E}[(f^2 \log f^2)(\mathsf{X}_t)] = \mathbb{E}[(f^2 \log f^2)(a_t \mathsf{Z} + b_t \mathsf{X}_1)]$$

$$\leq \int \left(\int f^2(a_t z + b_t x) \mathrm{d}\gamma_d(z) \right) \log \left(\int f^2(a_t z + b_t x) \mathrm{d}\gamma_d(z) \right) \mathrm{d}\nu(x)$$

$$+ \int \left(2C_{\mathrm{LS}}(\gamma_d) \int a_t^2 (\|\nabla f\|_2^2)(a_t z + b_t x) \mathrm{d}\gamma_d(z) \right) \mathrm{d}\nu(x)$$

$$\leq \left(\int \int f^2(a_t z + b_t x) \mathrm{d}\gamma_d(z) \mathrm{d}\nu(x) \right) \log \left(\int \int f^2(a_t z + b_t x) \mathrm{d}\gamma_d(z) \mathrm{d}\nu(x) \right)$$

$$+ 2C_{\mathrm{LS}}(\nu) \int \left\| \nabla_x \left(\int f^2(a_t z + b_t x) \mathrm{d}\gamma_d(z) \right)^{\frac{1}{2}} \right\|_2^2 \mathrm{d}\nu(x)$$

$$+ 2a_t^2 C_{\mathrm{LS}}(\gamma_d) \int \int (\|\nabla f\|_2^2)(a_t z + b_t x) \mathrm{d}\gamma_d(z) \mathrm{d}\nu(x)$$

$$\leq \mathbb{E}[f^2(\mathsf{X}_t)] \log \left(\mathbb{E}[f^2(\mathsf{X}_t)] \right) + 2a_t^2 C_{\mathrm{LS}}(\gamma_d) \mathbb{E}[\|\nabla f(\mathsf{X}_t)\|_2^2]$$

$$+ 2C_{\mathrm{LS}}(\nu) \int \left\| \nabla_x \left(\int f^2(a_t z + b_t x) \mathrm{d}\gamma_d(z) \right)^{\frac{1}{2}} \right\|_2^2 \mathrm{d}\nu(x).$$

By Jensen's inequality and the Cauchy-Schwartz inequality, it holds that

$$\int \left\| \nabla_x \left(\int f^2(a_t z + b_t x) \mathrm{d}\gamma_d(z) \right)^{\frac{1}{2}} \right\|_2^2 \mathrm{d}\nu(x)$$

$$\leq b_t^2 \frac{\int \left(\int (\|f \nabla f\|_2)(a_t z + b_t x) \mathrm{d}\gamma_d(z) \right)^2 \mathrm{d}\nu(x)}{\int \int f^2(a_t z + b_t x) \mathrm{d}\gamma_d(z) \mathrm{d}\nu(x)}$$

$$\leq b_t^2 \int \int (\|\nabla f\|_2^2)(a_t z + b_t x) \mathrm{d}\gamma_d(z) \mathrm{d}\nu(x)$$

$$\leq b_t^2 \mathbb{E}[\|\nabla f(\mathsf{X}_t)\|_2^2].$$

Hence, combining the equations above and the fact that $C_{\mathrm{LS}}(\gamma_d) \leq 1$ (Gross, 1975), it implies that

$$\mathbb{E}[(f^2 \log f^2)(\mathsf{X}_t)] - \mathbb{E}[f^2(\mathsf{X}_t)] \log \left(\mathbb{E}[f^2(\mathsf{X}_t)] \right) \leq 2 \left[a_t^2 + b_t^2 C_{\mathrm{LS}}(\nu) \right] \mathbb{E}[\|\nabla f(\mathsf{X}_t)\|_2^2],$$

that is, $C_{\mathrm{LS}}(p_t) \leq a_t^2 + b_t^2 C_{\mathrm{LS}}(\nu)$.

Next, we tackle the case of Poincaré inequalities by similar calculations. Using that $\mathsf{Z} \sim \gamma_d$ and $\mathsf{X}_1 \sim \nu$ both satisfy the Poincaré inequalities in Definition G.2, we have

$$\mathbb{E}[f^2(\mathsf{X}_t)] = \mathbb{E}[f^2(a_t \mathsf{Z} + b_t \mathsf{X}_1)]$$

$$\leq \int \left(\int f(a_t z + b_t x) \mathrm{d}\gamma_d(z) \right)^2 \mathrm{d}\nu(x)$$

$$+ \int \left(C_{\mathrm{P}}(\gamma_d) \int a_t^2 (\|\nabla f\|_2^2)(a_t z + b_t x) \mathrm{d}\gamma_d(z) \right) \mathrm{d}\nu(x)$$

$$\leq \left(\int \int f(a_t z + b_t x) \mathrm{d}\gamma_d(z) \mathrm{d}\nu(x) \right)^2$$

$$+ C_{\mathrm{P}}(\nu) \int \left\| \nabla_x \left(\int f(a_t z + b_t x) \mathrm{d}\gamma_d(z) \right) \right\|_2^2 \mathrm{d}\nu(x)$$

$$+ a_t^2 C_{\mathrm{P}}(\gamma_d) \int \int (\|\nabla f\|_2^2)(a_t z + b_t x) \mathrm{d}\gamma_d(z) \mathrm{d}\nu(x)$$

$$\leq (\mathbb{E}[f(\mathsf{X}_t)])^2 + \left[a_t^2 C_{\mathrm{P}}(\gamma_d) + b_t^2 C_{\mathrm{P}}(\nu) \right] \mathbb{E}[\|\nabla f(\mathsf{X}_t)\|_2^2].$$

Combining the expression above and $C_{\mathrm{P}}(\gamma_d) \leq 1$, it implies that

$$\mathbb{E}[f^2(\mathsf{X}_t)] - (\mathbb{E}[f(\mathsf{X}_t)])^2 \leq \left[a_t^2 + b_t^2 C_{\mathrm{P}}(\nu) \right] \mathbb{E}[\|\nabla f(\mathsf{X}_t)\|_2^2],$$

that is, $C_{\mathrm{P}}(p_t) \leq a_t^2 + b_t^2 C_{\mathrm{P}}(\nu)$. This completes the proof. ∎

Appendix E. Proofs of the stability results

We provide the proofs of the stability results in Section 6.

Proof [Proof of Proposition 6.1] Let $x_0 = a_0 z + b_0 x_1$ and suppose $X_0(x_0) \sim \mu, X_0(a_0 z) \sim \gamma_{d,a_0^2}$. According to Corollary 6.3, the Lipschitz property of $X_1(x)$ implies that $\|X_1(x_0) - X_1(a_0 z)\| \leq C_1 \|x_0 - a_0 z\|$. We consider an integral defined by

$$I_t := \int \|x_0 - a_0 z\|^2 \mathrm{d}\pi_t(X_t(x_0), X_t(a_0 z)),$$

where π_t is a coupling made of the joint distribution of $(X_t(x_0), X_t(a_0 z))$. In particular, the initial value I_0 is computed by

$$I_0 = \int \|x_0 - a_0 z\|^2 p_0(x_0) \varphi(z) \mathrm{d}x_0 \mathrm{d}z = \int \|b_0 x_1\|^2 p_1(x_1) \mathrm{d}x_1 = b_0^2 \mathbb{E}_\nu[\|\mathsf{X}_1\|^2].$$

Since $(X_t)_{t \in [0,1]}$ is well-posed with $X_0(x_0) \sim \mu$ or $X_0(a_0 z) \sim \gamma_{d,a_0^2}$, according to Corollary B.1, the coupling π_t satisfies the following differential equation

$$\partial_t \log \pi_t(X_t(x_0), X_t(a_0 z)) = -\operatorname{Tr}((\nabla_x v)(t, X_t(x_0))) - \operatorname{Tr}((\nabla_x v)(t, X_t(a_0 z))). \tag{E.1}$$

Taking the derivative of I_t and using Eq. (E.1), it implies that

$$\frac{\mathrm{d}I_t}{\mathrm{d}t} \leq 2 \left(\sup_{(s,x) \in [0,1] \times \mathbb{R}^d} \|\operatorname{Tr}(\nabla_x v(s,x))\| \right) I_t.$$

Thanks to $\|\operatorname{Tr}(\nabla_x v(s,x))\| \leq d \|\nabla_x v(s,x)\|_{2,2}$, it follows that

$$\frac{\mathrm{d}I_t}{\mathrm{d}t} \leq 2 C_2 d I_t, \quad I_0 = b_0^2 \mathbb{E}_\nu[\|\mathsf{X}_1\|^2].$$

By Grönwall's inequality, it holds that $I_t \leq b_0^2 \mathbb{E}_\nu[\|\mathsf{X}_1\|^2] \exp(2 C_2 d t)$. Therefore, we obtain the following W_2 bound

$$W_2(X_{1\#}\gamma_{d,a_0^2}, \nu) = W_2(X_{1\#}\gamma_{d,a_0^2}, X_{1\#}\mu) \leq C_1 \sqrt{I_1} \leq C_1 b_0 \sqrt{\mathbb{E}_\nu[\|\mathsf{X}_1\|^2]} \exp(C_2 d),$$

which completes the proof. $\blacksquare$

Proof [Proof of Proposition 6.2]

(i) On the one hand, by Corollary 6.3, $v(t,x)$ is Lipschitz continuous in x uniformly over $(t,x) \in [0,1] \times \mathbb{R}^d$ with Lipschitz constant C_2. By the variational equation (6.5) and Lemma 5.1, it follows that

$$\|\nabla_x X_{s,t}(x)\|_{2,2}^2 \leq \exp\left(2 \int_s^t \theta_u \mathrm{d}u \right).$$

Due to the equality (6.4), we deduce that

$$\begin{aligned}
&\|X_1(x_0) - Y_1(x_0)\|^2 \\
&\leq \left(\int_0^1 \|(\nabla_x X_{s,1})(Y_s(x_0))\|_{2,2} \|v(s, Y_s(x_0)) - \tilde{v}(s, Y_s(x_0))\| \mathrm{d}s \right)^2 \\
&\leq \left(\int_0^1 \|(\nabla_x X_{s,1})(Y_s(x_0))\|_{2,2}^2 \mathrm{d}s \right) \left(\int_0^1 \|v(s, Y_s(x_0)) - \tilde{v}(s, Y_s(x_0))\|^2 \mathrm{d}s \right) \\
&\leq \int_0^1 \exp\left(2 \int_s^1 \theta_u \mathrm{d}u \right) \mathrm{d}s \int_0^1 \|v(s, Y_s(x_0)) - \tilde{v}(s, Y_s(x_0))\|^2 \mathrm{d}s.
\end{aligned}$$

39

Take expectation and it follows that

$$
\begin{aligned}
W_2^2(Y_{1\#}\mu, \nu) &\leq \mathbb{E}_{x_0 \sim \mu}\left[\|Y_1(x_0) - X_1(x_0)\|^2\right] \\
&\leq \int_0^1 \exp\left(2\int_s^1 \theta_u \mathrm{d}u\right)\mathrm{d}s \int_0^1 \int_{\mathbb{R}^d} \|v(t,x) - \tilde{v}(t,x)\|^2 \tilde{q}_t(x)\mathrm{d}x\mathrm{d}t \\
&\leq \varepsilon \int_0^1 \exp\left(2\int_s^1 \theta_u \mathrm{d}u\right)\mathrm{d}s
\end{aligned}
$$

where $\tilde{q}_t$ denotes the density function of $Y_{t\#}\mu$, and we use the assumption that

$$
\int_0^1 \int_{\mathbb{R}^d} \|v(t,x) - \tilde{v}(t,x)\|^2 \tilde{q}_t(x)\mathrm{d}x\mathrm{d}t \leq \varepsilon
$$

in the last inequality.

(ii) On the other hand, suppose that $\tilde{v}(t,x)$ is Lipschitz continuous in x uniformly over $(t,x) \in [0,1] \times \mathbb{R}^d$ with Lipschitz constant C_3. Applying Grönwall's inequality to the variational equation (6.7), it follows that

$$
\|\nabla_x Y_{s,t}(x)\|_{2,2}^2 \leq \exp(2C_3(t-s)).
$$

By the equality (6.6), it holds that

$$
\begin{aligned}
&\|Y_1(x_0) - X_1(x_0)\|^2 \\
&\leq \left(\int_0^1 \|(\nabla_x Y_{s,1})(X_s(x_0))\|_{2,2}\|v(s, X_s(x_0)) - \tilde{v}(s, X_s(x_0))\|\mathrm{d}s\right)^2 \\
&\leq \left(\int_0^1 \|(\nabla_x Y_{s,1})(X_s(x_0))\|_{2,2}^2\mathrm{d}s\right)\left(\int_0^1 \|v(s, X_s(x_0)) - \tilde{v}(s, X_s(x_0))\|^2\mathrm{d}s\right) \\
&\leq \frac{\exp(2C_3) - 1}{2C_3}\int_0^1 \|v(s, X_s(x_0)) - \tilde{v}(s, X_s(x_0))\|^2\mathrm{d}s.
\end{aligned}
$$

Taking expectations, it further yields that

$$
\begin{aligned}
W_2^2(Y_{1\#}\mu, \nu) &\leq \mathbb{E}_{x_0 \sim \mu}\left[\|Y_1(x_0) - X_1(x_0)\|^2\right] \\
&\leq \frac{\exp(2C_3) - 1}{2C_3}\int_0^1 \int_{\mathbb{R}^d} \|v(t,x) - \tilde{v}(t,x)\|^2 p_t(x)\mathrm{d}x\mathrm{d}t
\end{aligned}
$$

where $X_t(x_0) \sim p_t$.

$\blacksquare$

Appendix F. Time derivative of the velocity field

In this appendix, we are interested in representing the time derivative of the velocity field via moments of $Y|X_t = x$. The result is efficacious for controlling the time derivative with moment estimates, though the computation is somehow tedious.

Proposition F.1 *The time derivative of the velocity field $v(t, x)$ has an expression with moments of $X_1|X_t$ for any $t \in (0, 1)$ as follows*

$$\partial_t v(t, x) = \left(\frac{\ddot{a}_t}{a_t} - \frac{\dot{a}_t^2}{a_t^2} \right) x + \left(a_t^2 \frac{\ddot{b}_t}{b_t} - \dot{a}_t a_t \frac{\dot{b}_t}{b_t} - \ddot{a}_t a_t + \dot{a}_t^2 \right) \frac{b_t}{a_t^2} M_1$$

$$+ \frac{b_t^2}{a_t^2} \left(\frac{\dot{b}_t}{b_t} - \frac{\dot{a}_t}{a_t} \right) \left(\frac{\dot{b}_t}{b_t} - 2\frac{\dot{a}_t}{a_t} \right) M_2^c x - \frac{b_t^3}{a_t^2} \left(\frac{\dot{b}_t}{b_t} - \frac{\dot{a}_t}{a_t} \right)^2 (M_3 - M_2 M_1),$$

where $M_1 := \mathbb{E}[X_1|X_t = x], M_2 := \mathbb{E}[X_1^\top X_1|X_t = x], M_2^c := \mathrm{Cov}(X_1|X_t = x), M_3 := \mathbb{E}[X_1 X_1^\top X_1|X_t = x].$

Proof By direct differentiation, it implies that

$$\partial_t v(t, x) = \partial_t \left(\frac{\dot{b}_t}{b_t} \right) x + \partial_t \left(\frac{\dot{b}_t}{b_t} a_t^2 - \dot{a}_t a_t \right) s(t, x) + \left(\frac{\dot{b}_t}{b_t} a_t^2 - \dot{a}_t a_t \right) \partial_t s(t, x)$$

$$= \frac{\ddot{b}_t b_t - \dot{b}_t^2}{b_t^2} x + \left(\frac{\ddot{b}_t b_t - \dot{b}_t^2}{b_t^2} a_t^2 + \frac{\dot{b}_t}{b_t} 2\dot{a}_t a_t - \ddot{a}_t a_t - \dot{a}_t^2 \right) s(t, x) + \left(\frac{\dot{b}_t}{b_t} a_t^2 - \dot{a}_t a_t \right) \partial_t s(t, x).$$

We first focus on $\partial_t s(t, x)$. Since p_t satisfies the continuity equation (3.4), it holds that

$$\partial_t s(t, x) = \nabla_x (\partial_t \log p_t(x))$$

$$= -\nabla_x \left(\frac{\nabla_x \cdot (p_t(x) v(t, x))}{p_t(x)} \right)$$

$$= -\nabla_x \left(\frac{(\nabla_x p_t(x))^\top v(t, x) + p_t(x)(\nabla_x \cdot v(t, x))}{p_t(x)} \right)$$

$$= -\nabla_x \left(s(t, x)^\top v(t, x) + \nabla_x \cdot v(t, x) \right)$$

$$= -\left((\nabla_x s(t, x))^\top v(t, x) + (\nabla_x v(t, x))^\top s(t, x) + \nabla_x (\nabla_x \cdot v(t, x)) \right)$$

$$= -\left(\nabla_x s(t, x) v(t, x) + \nabla_x v(t, x) s(t, x) + \nabla_x \mathrm{Tr}(\nabla_x v(t, x)) \right).$$

By direct computation, it holds that

$$\nabla_x s(t, x) v(t, x) + \nabla_x v(t, x) s(t, x)$$

$$= \nabla_x s(t, x) \left(\frac{\dot{b}_t}{b_t} x + \left(\frac{\dot{b}_t}{b_t} a_t^2 - \dot{a}_t a_t \right) s(t, x) \right) + \nabla_x \left(\frac{\dot{b}_t}{b_t} x + \left(\frac{\dot{b}_t}{b_t} a_t^2 - \dot{a}_t a_t \right) s(t, x) \right) s(t, x)$$

$$= \frac{\dot{b}_t}{b_t} \nabla_x s(t, x) x + \left(\frac{\dot{b}_t}{b_t} a_t^2 - \dot{a}_t a_t \right) \nabla_x s(t, x) s(t, x) + \frac{\dot{b}_t}{b_t} s(t, x) + \left(\frac{\dot{b}_t}{b_t} a_t^2 - \dot{a}_t a_t \right) \nabla_x s(t, x) s(t, x)$$

$$= \frac{\dot{b}_t}{b_t} s(t, x) + \frac{\dot{b}_t}{b_t} \nabla_x s(t, x) x + 2 \left(\frac{\dot{b}_t}{b_t} a_t^2 - \dot{a}_t a_t \right) \nabla_x s(t, x) s(t, x).$$

41

Then we focus on the trace term

$$\nabla_x \operatorname{Tr}(\nabla_x v(t,x))$$

$$= \nabla_x \operatorname{Tr}\left(\left(\frac{\dot{b}_t}{b_t} - \frac{\dot{a}_t}{a_t}\right)\frac{b_t^2}{a_t^2}\operatorname{Cov}(\mathsf{Y}|\mathsf{X}_t = x) + \frac{\dot{a}_t}{a_t}\mathbf{I}_d\right)$$

$$= \left(\frac{\dot{b}_t}{b_t} - \frac{\dot{a}_t}{a_t}\right)\frac{b_t^2}{a_t^2}\nabla_x\operatorname{Tr}(\operatorname{Cov}(\mathsf{Y}|\mathsf{X}_t = x))$$

$$= \left(\frac{\dot{b}_t}{b_t} - \frac{\dot{a}_t}{a_t}\right)\frac{b_t^2}{a_t^2}\nabla_x\left(\int \|y\|^2 p(y|t,x)\mathrm{d}y - \left\|\int yp(y|t,x)\mathrm{d}y\right\|^2\right)$$

$$= \left(\frac{\dot{b}_t}{b_t} - \frac{\dot{a}_t}{a_t}\right)\frac{b_t^2}{a_t^2}\left(\int \|y\|^2\nabla_x p(y|t,x)\mathrm{d}y - 2\left(\int \nabla_x p(y|t,x)\otimes y\mathrm{d}y\right)\left(\int yp(y|t,x)\mathrm{d}y\right)\right),$$

where we notice that

$$\nabla_x p(y|t,x) = \nabla_x\left(\frac{p(t,x|y)p_1(y)}{p_t(x)}\right)$$

$$= \frac{\nabla_x p(t,x|y)p_1(y)}{p_t(x)} - \frac{p(t,x|y)p_1(y)}{p_t(x)}s(t,x)$$

$$= p(y|t,x)\left(\frac{b_t y - x}{a_t^2} - s(t,x)\right).$$

For ease of presentation, we introduce the following notations to denote several moments of $\mathsf{Y}|\mathsf{X}_t = x$

$$M_1 := \mathbb{E}[\mathsf{Y}|\mathsf{X}_t = x], \qquad\qquad M_2 := \mathbb{E}[\mathsf{Y}^\top\mathsf{Y}|\mathsf{X}_t = x],$$

$$M_2^c := \operatorname{Cov}(\mathsf{Y}|\mathsf{X}_t = x), \qquad\qquad M_3 := \mathbb{E}[\mathsf{Y}\mathsf{Y}^\top\mathsf{Y}|\mathsf{X}_t = x].$$

By Tweedie's formula in Lemma G.1, it yields $s(t,x) = \frac{b_t}{a_t^2}M_1 - \frac{1}{a_t^2}x$. By this expression of $s(t,x)$, it yields

$$\nabla_x s(t,x)v(t,x) + \nabla_x v(t,x)s(t,x)$$

$$= \frac{\dot{b}_t}{b_t}s(t,x) + \frac{\dot{b}_t}{b_t}\nabla_x s(t,x)x + 2\left(\frac{\dot{b}_t}{b_t}a_t^2 - \dot{a}_t a_t\right)\nabla_x s(t,x)s(t,x)$$

$$= \frac{\dot{b}_t}{b_t}\left(\frac{b_t}{a_t^2}M_1 - \frac{1}{a_t^2}x\right) + \frac{\dot{b}_t}{b_t}\left(\frac{b_t^2}{a_t^4}M_2^c - \frac{1}{a_t^2}\mathbf{I}_d\right)x$$

$$\quad + 2\left(\frac{\dot{b}_t}{b_t}a_t^2 - \dot{a}_t a_t\right)\left(\frac{b_t^2}{a_t^4}M_2^c - \frac{1}{a_t^2}\mathbf{I}_d\right)\left(\frac{b_t}{a_t^2}M_1 - \frac{1}{a_t^2}x\right)$$

$$= -2\frac{\dot{a}_t}{a_t^3}x + \frac{b_t}{a_t^2}\left(2\frac{\dot{a}_t}{a_t} - \frac{\dot{b}_t}{b_t}\right)M_1 + \frac{b_t^2}{a_t^4}\left(2\frac{\dot{a}_t}{a_t} - \frac{\dot{b}_t}{b_t}\right)M_2^c x + 2\frac{b_t^3}{a_t^4}\left(\frac{\dot{b}_t}{b_t} - \frac{\dot{a}_t}{a_t}\right)M_2^c M_1$$

42

and $\nabla_x p(y|t,x) = \frac{b_t}{a_t^2}(y - M_1)\, p(y|t,x)$. Therefore, we obtain

$$\int \|y\|^2 \nabla_x p(y|t,x)\mathrm{d}y - 2\left(\int \nabla_x p(y|t,x) \otimes y \mathrm{d}y\right)\left(\int y p(y|t,x)\mathrm{d}y\right)$$

$$= \int \|y\|^2 \frac{b_t}{a_t^2}(y - M_1)\, p(y|t,x)\mathrm{d}y - 2\left(\int \frac{b_t}{a_t^2}(y - M_1) \otimes y p(y|t,x)\mathrm{d}y\right)\left(\int y p(y|t,x)\mathrm{d}y\right)$$

$$= \frac{b_t}{a_t^2}\left[\int \|y\|^2 y p(y|t,x)\mathrm{d}y - \left(\int \|y\|^2 p(y|t,x)\mathrm{d}y\right)M_1\right.$$

$$\left. - 2\left(\int y \otimes y p(y|t,x)\mathrm{d}y - M_1 \otimes \int y p(y|t,x)\mathrm{d}y\right)\left(\int y p(y|t,x)\mathrm{d}y\right)\right]$$

$$= \frac{b_t}{a_t^2}\left(M_3 - M_2 M_1 - 2M_2^c M_1\right).$$

Combining the equations above, we obtain

$$\partial_t v(t,x) = \frac{\ddot{b}_t b_t - \dot{b}_t^2}{b_t^2}x + \left(\frac{\ddot{b}_t b_t - \dot{b}_t^2}{b_t^2}a_t^2 + \frac{\dot{b}_t}{b_t}2\dot{a}_t a_t - \ddot{a}_t a_t - \dot{a}_t^2\right)\left(\frac{b_t}{a_t^2}M_1 - \frac{1}{a_t^2}x\right)$$

$$- \left(\frac{\dot{b}_t}{b_t}a_t^2 - \dot{a}_t a_t\right)\left[-2\frac{\dot{a}_t}{a_t^3}x + \frac{b_t}{a_t^2}\left(2\frac{\dot{a}_t}{a_t} - \frac{\dot{b}_t}{b_t}\right)M_1\right.$$

$$\left. + \frac{b_t^2}{a_t^4}\left(2\frac{\dot{a}_t}{a_t} - \frac{\dot{b}_t}{b_t}\right)M_2^c x + 2\frac{b_t^3}{a_t^4}\left(\frac{\dot{b}_t}{b_t} - \frac{\dot{a}_t}{a_t}\right)M_2^c M_1\right]$$

$$- \left(\frac{\dot{b}_t}{b_t}a_t^2 - \dot{a}_t a_t\right)\left(\frac{\dot{b}_t}{b_t} - \frac{\dot{a}_t}{a_t}\right)\frac{b_t^3}{a_t^4}(M_3 - M_2 M_1 - 2M_2^c M_1)$$

$$= \left(\frac{\ddot{a}_t}{a_t} - \frac{\dot{a}_t^2}{a_t^2}\right)x + \left(a_t^2 \frac{\ddot{b}_t}{b_t} - \dot{a}_t a_t \frac{\dot{b}_t}{b_t} - \ddot{a}_t a_t + \dot{a}_t^2\right)\frac{b_t}{a_t^2}M_1$$

$$+ \frac{b_t^2}{a_t^2}\left(\frac{\dot{b}_t}{b_t} - \frac{\dot{a}_t}{a_t}\right)\left(\frac{\dot{b}_t}{b_t} - 2\frac{\dot{a}_t}{a_t}\right)M_2^c x - \frac{b_t^3}{a_t^2}\left(\frac{\dot{b}_t}{b_t} - \frac{\dot{a}_t}{a_t}\right)^2 (M_3 - M_2 M_1).$$

Then we complete the proof. ∎

Appendix G. Functional inequalities and Tweedie's formula

This appendix is devoted to an exposition of functional inequalities and Tweedie's formula that would assist in our proof.

For a probability measure μ on a compact set $\Omega \subset \mathbb{R}^d$, we define the variance of a function $f \in L^2(\Omega, \mu)$ as

$$\mathrm{Var}_\mu(f) := \int_\Omega f^2 \mathrm{d}\mu - \left(\int_\Omega f \mathrm{d}\mu\right)^2.$$

Moreover, for a probability measure μ on a compact set $\Omega \subset \mathbb{R}^d$ and any positive integrable function $f : \Omega \to \mathbb{R}$ such that $\int_\Omega f \| \log f \| \mathrm{d}\nu < \infty$, we define the entropy of f as

$$\mathrm{Ent}_\mu(f) := \int_\Omega f \log f \mathrm{d}\mu - \int_\Omega f \mathrm{d}\mu \log \left(\int_\Omega f \mathrm{d}\mu \right).$$

Definition G.1 (Log-Sobolev inequality) *A probability measure $\mu \in \mathcal{P}(\Omega)$ is said to satisfy a log-Sobolev inequality with constant $C > 0$, if for all functions $f : \Omega \to \mathbb{R}$, it holds that*

$$\mathrm{Ent}_\mu(f^2) \le 2C \int_\Omega \|\nabla f\|_2^2 \mathrm{d}\mu.$$

The best constant $C > 0$ for which such an inequality holds is referred to as the log-Sobolev constant $C_{\mathrm{LS}}(\mu)$.

Definition G.2 (Poincaré inequality) *A probability measure $\mu \in \mathcal{P}(\Omega)$ is said to satisfy a Poincaré inequality with constant $C > 0$, if for all functions $f : \Omega \to \mathbb{R}$, it holds that*

$$\mathrm{Var}_\mu(f) \le C \int_\Omega \|\nabla f\|_2^2 \mathrm{d}\mu.$$

The best constant $C > 0$ for which such an inequality holds is referred to as the Poincaré constant $C_{\mathrm{P}}(\mu)$.

Finally, for ease of reference, we present Tweedie's formula that was first reported in Robbins (1956), and then was used as a simple empirical Bayes approach for correcting selection bias (Efron, 2011). Here, we use Tweedie's formula to link the score function with the expectation conditioned on an observation with Gaussian noise.

Lemma G.1 (Tweedie's formula) *Suppose that $\mathsf{X} \sim \mu$ and $\epsilon \sim \gamma_{d,\sigma^2}$. Let $\mathsf{Y} = \mathsf{X} + \epsilon$ and $p(y)$ be the marginal density of Y. Then $\mathbb{E}[\mathsf{X}|\mathsf{Y} = y] = y + \sigma^2 \nabla_y \log p(y)$.*

References

Michael S. Albergo and Eric Vanden-Eijnden. Building normalizing flows with stochastic interpolants. In *The Eleventh International Conference on Learning Representations*, 2023.

Michael S. Albergo, Nicholas M. Boffi, Michael Lindsey, and Eric Vanden-Eijnden. Multimarginal generative modeling with stochastic interpolants. *arXiv preprint arXiv:2310.03695*, 2023a.

Michael S. Albergo, Nicholas M. Boffi, and Eric Vanden-Eijnden. Stochastic interpolants: A unifying framework for flows and diffusions. *arXiv preprint arXiv:2303.08797*, 2023b.

Michael S. Albergo, Mark Goldstein, Nicholas M. Boffi, Rajesh Ranganath, and Eric Vanden-Eijnden. Stochastic interpolants with data-dependent couplings. *arXiv preprint arXiv:2310.03725*, 2023c.

Luigi Ambrosio and Gianluca Crippa. Continuity equations and ODE flows with non-smooth velocity. *Proceedings of the Royal Society of Edinburgh Section A: Mathematics*, 144(6):1191–1244, 2014.

Luigi Ambrosio, Nicola Gigli, and Giuseppe Savaré. *Gradient flows: In metric spaces and in the space of probability measures*. Springer Science & Business Media, 2008.

Luigi Ambrosio, Sebastiano N. Golo, and Francesco S. Cassano. Classical flows of vector fields with exponential or sub-exponential summability. *Journal of Differential Equations*, 372:458–504, 2023.

Abdul Fatir Ansari, Ming Liang Ang, and Harold Soh. Refining deep generative models via discriminator gradient flow. In *International Conference on Learning Representations*, 2021.

Michael Arbel, Anna Korba, Adil Salim, and Arthur Gretton. Maximum mean discrepancy gradient flow. In *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019.

Martin Arjovsky, Soumith Chintala, and Léon Bottou. Wasserstein generative adversarial networks. In *Proceedings of the 34th International Conference on Machine Learning*, volume 70, pages 214–223. PMLR, 2017.

Dominique Bakry and Michel Émery. Diffusions hypercontractives. In *Seminaire de probabilités XIX 1983/84*, pages 177–206. Springer, 1985.

Dominique Bakry, Ivan Gentil, and Michel Ledoux. *Analysis and geometry of Markov diffusion operators*, volume 103. Springer, 2014.

Keith Ball, Franck Barthe, and Assaf Naor. Entropy jumps in the presence of a spectral gap. *Duke Mathematical Journal*, 119(1):41 – 63, 2003.

Joe Benton, George Deligiannidis, and Arnaud Doucet. Error bounds for flow matching methods. *arXiv preprint arXiv:2305.16860*, 2023.

Marin Biloš, Johanna Sommer, Syama Sundar Rangapuram, Tim Januschowski, and Stephan Günnemann. Neural flows: Efficient alternative to neural ODEs. In *Advances in Neural Information Processing Systems*, volume 34, pages 21325–21337. Curran Associates, Inc., 2021.

Sergey G. Bobkov and Michel Ledoux. From Brunn-Minkowski to Brascamp-Lieb and to logarithmic Sobolev inequalities. *Geometric and Functional Analysis*, 10(5):1028–1052, 2000.

Valentin De Bortoli. Convergence of denoising diffusion models under the manifold hypothesis. *Transactions on Machine Learning Research*, 2022.

Herm J. Brascamp and Elliott H. Lieb. On extensions of the Brunn-Minkowski and Prékopa-Leindler theorems, including inequalities for log concave functions, and with an application to the diffusion equation. *Journal of Functional Analysis*, 22(4):366–389, 1976.

Luis A. Caffarelli. Monotonicity properties of optimal transportation and the FKG and related inequalities. *Communications in Mathematical Physics*, 214(3):547–563, 2000.

Tony T. Cai and Yihong Wu. Optimal detection of sparse mixtures against a given null distribution. *IEEE Transactions on Information Theory*, 60(4):2217–2232, 2014.

Patrick Cattiaux and Arnaud Guillin. Semi log-concave Markov diffusions. In Catherine Donati-Martin, Antoine Lejay, and Alain Rouault, editors, *Séminaire de probabilités XLVI*, pages 231–292. Springer International Publishing, Cham, 2014.

Hongrui Chen, Holden Lee, and Jianfeng Lu. Improved analysis of score-based generative modeling: User-friendly bounds under minimal smoothness assumptions. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202, pages 4735–4763. PMLR, 2023a.

Ricky T.Q. Chen and Yaron Lipman. Riemannian flow matching on general geometries. *arXiv preprint arXiv:2302.03660*, 2023.

Ricky T.Q. Chen, Yulia Rubanova, Jesse Bettencourt, and David K. Duvenaud. Neural ordinary differential equations. In *Advances in Neural Information Processing Systems*, volume 31. Curran Associates, Inc., 2018.

Sitan Chen, Sinho Chewi, Jerry Li, Yuanzhi Li, Adil Salim, and Anru Zhang. Sampling is as easy as learning the score: Theory for diffusion models with minimal data assumptions. In *The Eleventh International Conference on Learning Representations*, 2023b.

Sinho Chewi and Aram-Alexandre Pooladian. An entropic generalization of Caffarelli's contraction theorem via covariance inequalities. *arXiv preprint arXiv:2203.04954*, 2022.

Maria Colombo, Alessio Figalli, and Yash Jhaveri. Lipschitz changes of variables between perturbations of log-concave measures. *Annali della Scuola Normale Superiore di Pisa. Classe di scienze*, 17(4):1491–1519, 2017.

Dario Cordero-Erausquin. Transport inequalities for log-concave measures, quantitative forms, and applications. *Canadian Journal of Mathematics*, 69(3):481–501, 2017.

Yin Dai, Yuan Gao, Jian Huang, Yuling Jiao, Lican Kang, and Jin Liu. Lipschitz transport maps via the Föllmer flow. *arXiv preprint arXiv:2309.03490*, 2023.

Ludwig Danzer, Branko Grünbaum, and Victor Klee. Helly's theorem and its relatives. In *Proceedings of Symposia in Pure Mathematics: Convexity*, volume VII, pages 101–180, Providence, RI, 1963. American Mathematical Society.

Valentin De Bortoli, James Thornton, Jeremy Heng, and Arnaud Doucet. Diffusion Schrödinger bridge with applications to score-based generative modeling. In *Advances in Neural Information Processing Systems*, volume 34, pages 17695–17709. Curran Associates, Inc., 2021.

Pierre Del Moral and Sumeetpal S. Singh. Backward Itô–Ventzell and stochastic interpolation formulae. *Stochastic Processes and their Applications*, 154:197–250, 2022.

Prafulla Dhariwal and Alexander Nichol. Diffusion models beat GANs on image synthesis. In *Advances in Neural Information Processing Systems*, volume 34, pages 8780–8794, 2021.

Laurent Dinh, David Krueger, and Yoshua Bengio. NICE: Non-linear independent components estimation. *arXiv preprint arXiv:1410.8516*, 2014.

Andrew Duncan, Nikolas Nüsken, and Lukasz Szpruch. On the geometry of Stein variational gradient descent. *Journal of Machine Learning Research*, 24(56):1–39, 2023.

Alex Dytso, Martina Cardone, and Ian Zieder. Meta derivative identity for the conditional expectation. *IEEE Transactions on Information Theory*, 69(7):4284–4302, 2023a.

Alex Dytso, H. Vincent Poor, and Shlomo Shamai Shitz. Conditional mean estimation in Gaussian noise: A meta derivative identity with applications. *IEEE Transactions on Information Theory*, 69(3):1883–1898, 2023b.

Bradley Efron. Tweedie's formula and selection bias. *Journal of the American Statistical Association*, 106(496):1602–1614, 2011.

Ronen Eldan and James R. Lee. Regularization under diffusion and anticoncentration of the information content. *Duke Mathematical Journal*, 167(5):969–993, 2018.

Lawrence C. Evans. *Partial differential equations*, volume 19 of *Graduate studies in mathematics*. American Mathematical Society, Providence, Rhode Island, second edition edition, 2010.

Jiaojiao Fan, Qinsheng Zhang, Amirhossein Taghvaei, and Yongxin Chen. Variational Wasserstein gradient flow. In *Proceedings of the 39th International Conference on Machine Learning*, volume 162, pages 6185–6215. PMLR, 2022.

Max Fathi, Dan Mikulincer, and Yair Shenfeld. Transportation onto log-Lipschitz perturbations. *arXiv preprint arXiv:2305.03786*, 2023.

Chris Finlay, Jörn-Henrik Jacobsen, Levon Nurbekyan, and Adam Oberman. How to train your neural ODE: The world of Jacobian and kinetic regularization. In *Proceedings of the 37th International Conference on Machine Learning*, volume 119, pages 3154–3164. PMLR, 2020.

Yuan Gao, Yuling Jiao, Yang Wang, Yao Wang, Can Yang, and Shunkang Zhang. Deep generative learning via variational gradient flow. In *Proceedings of the 36th International Conference on Machine Learning*, volume 97, pages 2093–2101. PMLR, 2019.

Yuan Gao, Jian Huang, Yuling Jiao, Jin Liu, Xiliang Lu, and Zhijian Yang. Deep generative learning via Euler particle transport. In *Proceedings of the 2nd Mathematical and Scientific Machine Learning Conference*, volume 145, pages 336–368. PMLR, 2022.

Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *Advances in Neural Information Processing Systems 27*, pages 2672–2680. Curran Associates, Inc., 2014.

Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016. http://www.deeplearningbook.org.

Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial networks. *Communications of the ACM*, 63(11):139–144, 2020.

Will Grathwohl, Ricky T.Q. Chen, Jesse Bettencourt, and David Duvenaud. FFJORD: Free-form continuous dynamics for scalable reversible generative models. In *International Conference on Learning Representations*, 2019.

Leonard Gross. Logarithmic Sobolev inequalities. *American Journal of Mathematics*, 97 (4):1061–1083, 1975.

Ernst Hairer, Gerhard Wanner, and Syvert P. Nørsett. *Classical Mathematical Theory*, chapter I, pages 1–128. Springer Berlin Heidelberg, Berlin, Heidelberg, 1993.

Philip Hartman. *Dependence on Initial Conditions and Parameters*, chapter V, pages 93–116. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2002a.

Philip Hartman. *Existence*, chapter II, pages 8–23. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2002b.

Charles P. Hatsell and Loren W. Nolte. Some geometric properties of the likelihood ratio (corresp.). *IEEE Transactions on Information Theory*, 17(5):616–618, 1971.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. In *Advances in Neural Information Processing Systems*, volume 33, pages 6840–6851, 2020.

Rie Johnson and Tong Zhang. Composite functional gradient learning of generative adversarial models. In *Proceedings of the 35th International Conference on Machine Learning*, volume 80, pages 2371–2379. PMLR, 2018.

Rie Johnson and Tong Zhang. A framework of composite functional gradient methods for generative adversarial models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(1):17–32, 2019.

Tero Karras, Miika Aittala, Timo Aila, and Samuli Laine. Elucidating the design space of diffusion-based generative models. In *Advances in Neural Information Processing Systems*, volume 35, pages 26565–26577. Curran Associates, Inc., 2022.

Dongjun Kim, Chieh-Hsin Lai, Wei-Hsiang Liao, Naoki Murata, Yuhta Takida, Toshimitsu Uesaka, Yutong He, Yuki Mitsufuji, and Stefano Ermon. Consistency trajectory models: Learning probability flow ODE trajectory of diffusion. *arXiv preprint arXiv:2310.02279*, 2023.

Diederik P. Kingma and Max Welling. Auto-encoding variational bayes. In *International Conference on Learning Representations*, 2014.

Diederik P. Kingma and Max Welling. An introduction to variational autoencoders. *Foundations and Trends® in Machine Learning*, 12(4):307–392, 2019.

Bo'az Klartag. High-dimensional distributions with convexity properties. In *Proceedings of the Fifth European Congress of Mathematics*, pages 401–417, Amsterdam, 14 July–18 July 2010. European Mathematical Society Publishing House.

Ivan Kobyzev, Simon J.D. Prince, and Marcus A. Brubaker. Normalizing flows: An introduction and review of current methods. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(11):3964–3979, 2020.

Jeroen S.W. Lamb and John A.G. Roberts. Time-reversal symmetry in dynamical systems: A survey. *Physica D: Nonlinear Phenomena*, 112(1-2):1–39, 1998.

Holden Lee, Jianfeng Lu, and Yixin Tan. Convergence of score-based generative modeling for general data distributions. In *Proceedings of The 34th International Conference on Algorithmic Learning Theory*, volume 201, pages 946–985. PMLR, 2023.

Tengyuan Liang. How well generative adversarial networks learn distributions. *Journal of Machine Learning Research*, 22(228):1–41, 2021.

Yaron Lipman, Ricky T.Q. Chen, Heli Ben-Hamu, Maximilian Nickel, and Matthew Le. Flow matching for generative modeling. In *The Eleventh International Conference on Learning Representations*, 2023.

Qiang Liu. Rectified flow: A marginal preserving approach to optimal transport. *arXiv preprint arXiv:2209.14577*, 2022.

Xingchao Liu, Chengyue Gong, and Qiang Liu. Flow straight and fast: Learning to generate and transfer data with rectified flow. In *The Eleventh International Conference on Learning Representations*, 2023.

Antoine Liutkus, Umut Simsekli, Szymon Majewski, Alain Durmus, and Fabian-Robert Stöter. Sliced-Wasserstein flows: Nonparametric generative modeling via optimal transport and diffusions. In *Proceedings of the 36th International Conference on Machine Learning*, volume 97, pages 4104–4113. PMLR, 2019.

Cheng Lu, Kaiwen Zheng, Fan Bao, Jianfei Chen, Chongxuan Li, and Jun Zhu. Maximum likelihood training for score-based diffusion ODEs by high order denoising score matching. In *Proceedings of the 39th International Conference on Machine Learning*, pages 14429–14460. PMLR, 2022a.

Cheng Lu, Yuhao Zhou, Fan Bao, Jianfei Chen, Chongxuan Li, and Jun Zhu. DPM-Solver: A fast ODE solver for diffusion probabilistic model sampling in around 10 steps. In *Advances in Neural Information Processing Systems*, volume 35, pages 5775–5787. Curran Associates, Inc., 2022b.

Ashok Makkuva, Amirhossein Taghvaei, Sewoong Oh, and Jason Lee. Optimal transport mapping via input convex neural networks. In *Proceedings of the 37th International Conference on Machine Learning*, volume 119, pages 6672–6681. PMLR, 2020.

Pierre Marion. Generalization bounds for neural ordinary differential equations and deep residual networks. *arXiv preprint arXiv:2305.06648*, 2023.

Pierre Marion, Yu-Han Wu, Michael E Sander, and Gérard Biau. Implicit regularization of deep residual networks towards neural ODEs. *arXiv preprint arXiv:2309.01213*, 2023.

Youssef Marzouk, Zhi Ren, Sven Wang, and Jakob Zech. Distribution learning via neural differential equations: A nonparametric statistical perspective. *arXiv preprint arXiv:2309.01043*, 2023.

Dan Mikulincer and Yair Shenfeld. The Brownian transport map. *arXiv preprint arXiv:2111.11521*, 2021.

Dan Mikulincer and Yair Shenfeld. On the Lipschitz properties of transportation along heat flows. In *Geometric Aspects of Functional Analysis: Israel Seminar (GAFA) 2020-2022*, pages 269–290. Springer, 2023.

Youssef Mroueh and Truyen Nguyen. On the convergence of gradient descent in GANs: MMD GAN as a gradient flow. In *Proceedings of The 24th International Conference on Artificial Intelligence and Statistics*, volume 130, pages 1720–1728. PMLR, 2021.

Youssef Mroueh, Tom Sercu, and Anant Raj. Sobolev descent. In *Proceedings of the 22nd International Conference on Artificial Intelligence and Statistics*, volume 89, pages 2976–2985. PMLR, 2019.

Joe Neeman. Lipschitz changes of variables via heat flow. *arXiv preprint arXiv:2201.03403*, 2022.

Kirill Neklyudov, Rob Brekelmans, Daniel Severo, and Alireza Makhzani. Action matching: Learning stochastic dynamics from samples. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202, pages 25858–25889. PMLR, 2023.

Derek Onken, Samy Wu Fung, Xingjian Li, and Lars Ruthotto. OT-flow: Fast and accurate continuous normalizing flows via optimal transport. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 9223–9232, 2021.

Daniel P. Palomar and Sergio Verdú. Gradient of mutual information in linear vector Gaussian channels. *IEEE Transactions on Information Theory*, 52(1):141–154, 2005.

George Papamakarios, Eric Nalisnick, Danilo Jimenez Rezende, Shakir Mohamed, and Balaji Lakshminarayanan. Normalizing flows for probabilistic modeling and inference. *Journal of Machine Learning Research*, 22(57):1–64, 2021.

Aram-Alexandre Pooladian, Heli Ben-Hamu, Carles Domingo-Enrich, Brandon Amos, Yaron Lipman, and Ricky T.Q. Chen. Multisample flow matching: Straightening flows with minibatch couplings. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202, pages 28100–28127. PMLR, 2023.

Danilo J. Rezende and Shakir Mohamed. Variational inference with normalizing flows. In *Proceedings of the 32nd International Conference on Machine Learning*, volume 37, pages 1530–1538. PMLR, 2015.

Danilo J. Rezende, Shakir Mohamed, and Daan Wierstra. Stochastic backpropagation and approximate inference in deep generative models. In *Proceedings of the 31st International Conference on Machine Learning*, pages 1278–1286. PMLR, 2014.

Herbert E. Robbins. An empirical Bayes approach to statistics. In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, 1954–1955*, volume I, page 157–163, Berkeley and Los Angeles, 1956. University of California Press.

Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10684–10695, 2022.

Domenec Ruiz-Balet and Enrique Zuazua. Neural ODE control for classification, approximation, and transport. *SIAM Review*, 65(3):735–773, 2023.

Ruslan Salakhutdinov. Learning deep generative models. *Annual Review of Statistics and Its Application*, 2(1):361–385, 2015.

Adrien Saumard and Jon A. Wellner. Log-concavity and strong log-concavity: A review. *Statistics Surveys*, 8:45 – 114, 2014.

Neta Shaul, Ricky T.Q. Chen, Maximilian Nickel, Matthew Le, and Yaron Lipman. On kinetic optimal probability paths for generative models. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202, pages 30883–30907. PMLR, 2023.

Jascha Sohl-Dickstein, Eric Weiss, Niru Maheswaranathan, and Surya Ganguli. Deep unsupervised learning using nonequilibrium thermodynamics. In *Proceedings of the 32nd International Conference on Machine Learning*, volume 37, pages 2256–2265. PMLR, 2015.

Jiaming Song, Chenlin Meng, and Stefano Ermon. Denoising diffusion implicit models. In *International Conference on Learning Representations*, 2021a.

Yang Song and Prafulla Dhariwal. Improved techniques for training consistency models. *arXiv preprint arXiv:2310.14189*, 2023.

Yang Song and Stefano Ermon. Generative modeling by estimating gradients of the data distribution. In *Advances in Neural Information Processing Systems*, volume 32, pages 11895–11907. Curran Associates, Inc., 2019.

Yang Song and Stefano Ermon. Improved techniques for training score-based generative models. In *Advances in Neural Information Processing Systems*, volume 33, pages 12438–12448. Curran Associates, Inc., 2020.

Yang Song, Jascha Sohl-Dickstein, Diederik P. Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. In *International Conference on Learning Representations*, 2021b.

Yang Song, Prafulla Dhariwal, Mark Chen, and Ilya Sutskever. Consistency models. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pages 32211–32252. PMLR, 23–29 Jul 2023.

Xuan Su, Jiaming Song, Chenlin Meng, and Stefano Ermon. Dual diffusion implicit bridges for image-to-image translation. In *The Eleventh International Conference on Learning Representations*, 2023.

Esteban G. Tabak and Cristina V. Turner. A family of nonparametric density estimation algorithms. *Communications on Pure and Applied Mathematics*, 66(2):145–164, 2013.

Esteban G. Tabak and Eric Vanden-Eijnden. Density estimation by dual ascent of the log-likelihood. *Communications in Mathematical Sciences*, 8(1):217–233, 2010.

Alexander Tong, Nikolay Malkin, Guillaume Huguet, Yanlei Zhang, Jarrid Rector-Brooks, Kilian Fatras, Guy Wolf, and Yoshua Bengio. Conditional flow matching: Simulation-free dynamic optimal transport. *arXiv preprint arXiv:2302.00482*, 2023.

Andre Wibisono and Varun Jog. Convexity of mutual information along the heat flow. In *2018 IEEE International Symposium on Information Theory (ISIT)*, pages 1615–1619. IEEE, 2018a.

Andre Wibisono and Varun Jog. Convexity of mutual information along the Ornstein-Uhlenbeck flow. In *2018 International Symposium on Information Theory and Its Applications (ISITA)*, pages 55–59. IEEE, 2018b.

Andre Wibisono, Varun Jog, and Po-Ling Loh. Information and estimation in Fokker-Planck channels. In *2017 IEEE International Symposium on Information Theory (ISIT)*, pages 2673–2677. IEEE, 2017.

Yihong Wu and Sergio Verdú. Functional properties of minimum mean-square error and mutual information. *IEEE Transactions on Information Theory*, 58(3):1289–1301, 2011.

Chen Xu, Xiuyuan Cheng, and Yao Xie. Invertible normalizing flow neural networks by JKO scheme. *arXiv preprint arXiv:2212.14424*, 2022.

Liu Yang and George E. Karniadakis. Potential flow generator with L_2 optimal transport regularity for generative models. *IEEE Transactions on Neural Networks and Learning Systems*, 33(2):528–538, 2020.

Linfeng Zhang, Weinan E, and Lei Wang. Monge-Ampère flow for generative modeling. *arXiv preprint arXiv:1809.10188*, 2018.

Kaiwen Zheng, Cheng Lu, Jianfei Chen, and Jun Zhu. Improved techniques for maximum likelihood estimation for diffusion ODEs. *arXiv preprint arXiv:2305.03935*, 2023.

Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A. Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 2223–2232, 2017.