Proceedings of Machine Learning Research vol ???:1–20, 2025 38th Annual Conference on Learning Theory

On the Contractivity of Stochastic Interpolation Flow

author names withheld

Editor: Under Review for COLT 2025

Abstract

We investigate stochastic interpolation, a recently introduced framework for high dimensional sampling which bears many similarities to diffusion models. Stochastic interpolation generates a data sample by first randomly initializing a particle drawn from a simple base distribution, then simulating deterministic or stochastic dynamics such that in finite time the particle's distribution converges to the target. We show that for a Gaussian base distribution and a strongly log-concave target distribution, the stochastic interpolation flow map is Lipschitz with a sharp constant which matches that of Cafarelli's theorem for optimal transport maps. We are further able to construct Lipschitz transport maps between non-Gaussian distributions, generalizing some recent constructions in the literature on dimension-free functional inequalities. We discuss the practical implications of our theorem for the sampling and estimation problems required by stochastic interpolation.

Keywords: Lipschitz transportation, Stochastic interpolation, Diffusion models

1. Introduction

Recently in the high-dimensional sampling literature there has been significant interest in so-called *dynamical sampling algorithms*. Roughly speaking, these algorithms generate a sample by approximately simulating the dynamics of a particle whose law converges to the target distribution. These dynamics may be stochastic or deterministic and they are typically driven by a velocity field that can be estimated from samples via least squares regression. The most famous example of this methodology is *diffusion modeling*, where the idealized particle trajectory is a time-reversal of the Ornstein-Uhlenbeck process (Song et al., 2021), which is now well understood and which enjoys end-to-end theoretical guarantees on the estimation error of its drift function (Oko et al., 2023; Wibisono et al., 2024; Koehler et al., 2023; Pabbaraju et al., 2023) and the discretization error of various score-based dynamical samplers (Chen et al., 2023b; Benton et al., 2024; Li and Yan, 2024).

Our focus in this work is on a more recent approach to dynamical sampling known as *stochastic interpolation*[1]. Given a base distribution $\mu_0 \in \mathcal{P}_2(\mathbb{R}^d)$ and a target distribution $\mu_1 \in \mathcal{P}_2(\mathbb{R}^d)$, a stochastic interpolant from μ_0 to μ_1 is any member of a family of interpolations $(\mu_t)_{t\in[0,1]}$ constructed in the following manner.

Definition 1 (Definition 2.1, Albergo et al. (2023)) *Let $X_0 \sim \mu_0$ and $X_1 \sim \mu_1$ be independent random variables. Let $I(t, x, x') : [0, 1] \times \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^d$ be such that for each $x, x' \in \mathbb{R}^d$,*

$$t \mapsto I_t(\cdot,\cdot) \in C^2([0,1], (C^2(\mathbb{R}^d \times \mathbb{R}^d))^d) \qquad I_0(x,x') = x \qquad I_1(x,x') = x' \tag{1}$$

and additionally $|\partial_t I(t, x, x')| \le C|x - x'|$ for some $C < \infty$. The I_t-stochastic interpolation between μ_0, μ_1 is the trajectory of densities $(\mu_t)_{t\in[0,1]}$ given by $\mu_t = \mathsf{Law}(I_t(X_0, X_1))$.

1. Multiple authors concurrently proposed this approach under the names *stochastic interpolation* (Albergo and Vanden-Eijnden, 2023), *flow matching* (Lipman et al., 2023), and *flow rectification* (Liu et al., 2023).

Here we call $\mathsf{Law}(X)$ the distribution of the random variable X and $\mathcal{P}_2(\Omega)$ is the class of absolutely continuous densities with finite second moment on a subset Ω of a Euclidean space. For a function $\phi : \mathbb{R}^d \to \mathbb{R}$ we denote by $\nabla\phi(x)$ its gradient, $\nabla^2\phi(x)$ its Hessian, and for $f(x) : \mathbb{R}^d \to \mathbb{R}^d$ we call $D_x f(x)$ its Jacobian and $\nabla \cdot f(x) = \sum_{i=1}^d \partial_{x_i} f^{(i)}(x)$ its divergence. For time dependent functions, we occasionally write $f_t(x)$ in place of $f(t,x)$ and $\dot{f}_t(x)$ in place of $\partial_t f(t,x)$, and we use process notation $(f_t)_{t\in\Lambda}$ or, when the time domain Λ is clear from context, just (f_t).

The key property of a stochastic interpolant is that $(\mu_t)_{t\in[0,1]}$ can easily be realized by a dynamical sampler through a process known colloquially as '*markovianization*.' Recall that the flow map $f_t(x)$ corresponding to a (lipschitz in space, uniformly in time) velocity field $v_t(x) : [0,1] \times \mathbb{R}^d \to \mathbb{R}^d$ is the unique solution to the ODE $\partial_t f_t(x) = v_t(f_t(x))$ with initial condition $f_0(x) = x$.

Proposition 2 (Theorem 2.6, Albergo et al. (2023)) *Let $f^v_t(x)$ be the flow map corresponding to the velocity field,*

$$v_t(x) \coloneqq \mathbb{E}[\partial_t I_t(X_0, X_1) \mid I_t(X_0, X_1) = x]$$

The I_t-stochastic interpolant satisfies the continuity PDE with respect to $(v_t)_{t\in[0,1]}$,

$$\partial_t \mu_t + \nabla \cdot (\mu_t v_t) = 0$$

and consequently $\mu_t = (f^v_t)_\# \mu_0$.

Proof sketch. For an arbitrary test function $\phi \in C^\infty_c(\mathbb{R}^d)$, differentiate $\mu_t(\phi) \coloneqq \mathbb{E}[\phi(I_t(X_0, X_1))]$ in time:

$$\begin{aligned}
\partial_t \mu_t(\phi) &= \mathbb{E}_{X_0,X_1}[\langle \nabla\phi(I_t(X_0, X_1)), \partial_t I_t(X_0, X_1)\rangle] \\
&= \mathbb{E}_{X_t\sim\mu_t}[\langle \nabla\phi(X_t), \mathbb{E}[\partial_t I_t(X_0, X_1) \mid I_t(X_0, X_1) = X_t]\rangle] \\
&= \mathbb{E}_{X_t\sim\mu_t}[\langle \nabla\phi(X_t), v_t(X_t)\rangle],
\end{aligned}$$

which is precisely the weak form of the continuity PDE. ■

Because $I_t(X_0, X_1)$ depends on its endpoints, the trajectory $t \mapsto I_t(X_0, X_1)$ may not be a Markov process. But by replacing the random variable $\partial_t I_t(X_0, X_1)$ with its conditional average $v_t(x)$, it is possible to construct a Markov process $t \mapsto f^v_t(X_0)$ whose marginal distribution at each $t \in [0,1]$ coincides with μ_t. It is also easy to estimate $v_t(x)$ from i.i.d. samples of μ_0, μ_1 by minimizing an empirical approximation to the risk

$$\mathcal{L}[\hat{v}] = \mathbb{E}_{X_0,X_1}\left[\|\partial_t I_t(X_0, X_1) - \hat{v}(I_t(X_0, X_1))\|^2\right]$$

over an appropriate class of models $\hat{v} \in \mathcal{V}$, typically a class of neural networks. These observations together form the basis of a pipeline for dynamical sampling.

Owing to the flexibility of this simple construction, it is possible to define stochastic interpolants in vastly more general settings than Definition 1. Some notable examples include: X_0, X_1 may have an arbitrary coupling (Pooladian et al., 2023; Albergo et al., 2024), $I_t(X_0, X_1)$ may be a random process satisfying (1) almost surely (Albergo et al., 2023, Section 3.1), or X_0, X_1 may be sampled from a non-Euclidean space such as the probability simplex (Stark et al., 2024) or a Riemannian manifold (Kapusniak et al., 2024). This motivates the investigation of the properties of this vast new family of parametrizations of dynamical samplers.

Orthogonal to recent work on dynamical sampling, it is also valuable to understand the relationship between stochastic interpolation flow and other techniques for *transportation of measures*. Perhaps the most famous example is the theory of optimal transportation, which furnishes for any two $\mu_0, \mu_1 \in \mathcal{P}_2(\mathbb{R}^d)$ a unique map $T : \mathbb{R}^d \to \mathbb{R}^d$ which pushes forward $T_\# \mu_0 = \mu_1$ and with minimum *transportation cost*,

$$T = \arg\inf \mathcal{J}[T] \coloneqq \mathbb{E}_{X\sim\mu_0}[\|X - T(X)\|^2] \qquad \text{subject to} \qquad T_\#\mu_0 = \mu_1. \tag{2}$$

The value of this infimum is equal to the squared Wasserstein-2 distance $W_2^2(\mu_0, \mu_1)$. We recall two seminal results in the theory of optimal transport which will serve as points of comparison for our main result. The celebrated *Benamou-Brenier theorem* provides a dynamical perspective on optimal transportation.

Proposition 3 (Benamou and Brenier (2000)) *Given two probability measures μ_0, μ_1, it holds that*

$$W_2^2(\mu_0, \mu_1) = \inf_{\tilde{v}_t} \left\{ \int_0^1 \|\tilde{v}_t\|^2_{\mathcal{L}^2(\mu_t)}\, dt : \ \partial_t \mu_t + \nabla \cdot (\mu_t \tilde{v}_t) = 0, \mu_{t=0} = \mu_0, \mu_{t=1} = \mu_1 \right\}. \tag{3}$$

Moreover, the optimizer (v_t) induces a curve (μ_t) with the following properties.

1. *Conservative dynamics: there exists $\phi \in C^1([0,1], C^1(\mathbb{R}^d))$ such that $v_t(x) = \nabla\phi_t(x)$.*
2. *Zero acceleration: the flow $f_t^v(x)$ satisfies $\partial_t f_t^v(x) = T(x) - x$ where T is the optimal transport map from μ_0 to μ_1.*

The zero acceleration property implies that $f_1^v(x) = x + (T(x) - x) = T(x)$ is the optimal transport map. It turns out that Lipschitz constant of T is closely related to distributional properties of μ_0 and μ_1. Intuitively, if μ_1 is more concentrated than μ_0, then one might expect a bound on the contractivity of T to hold, and this intuition can be made quantitative in terms of uniform bounds on log-concavity of μ_0, μ_1.

Definition 4 (Log-concavity) *A density $\mu(x) = e^{-V(x)}$ is log-concave if $V(x)$ is a convex function. If $V \in C^2(\mathbb{R}^d)$, we say that μ is κ-log-concave if $\nabla^2 V(x) \succcurlyeq \kappa I_d$, and it is η-log-convex if $\nabla^2 V(x) \preccurlyeq \eta I_d$, uniformly in $x \in \mathbb{R}^d$.*

Proposition 5 (Caffarelli (2000)) *Let μ_0 be a smooth and η_0-log-convex density, let μ_1 be a smooth and κ_1-log-concave density, for $0 < \kappa_1 \leq \eta_0$. Then the optimal transport map from μ_0 to μ_1 is $\sqrt{\eta_0/\kappa_1}$ Lipschitz.*

Cafarelli originally proved this result by studying the regularity of solutions to a nonlinear second-order PDE known as the *Monge–Ampère equation*. Later, Chewi and Pooladian (2023) gave a direct proof by studying an entropic regularization of (2) and expressing the Jacobian of (an entropic analogue of) the transport map as a covariance matrix, which is bounded by well-known covariance inequalities that make use of log-concavity and convexity. Our results are in much the same spirit. As we discuss further in Section 2.3, analogues of Caffarelli's theorem have also been shown for the so-called Brownian transport map (Mikulincer and Shenfeld, 2024) and for a construction called 'Föllmer Flow' (Ding et al., 2023; Dai et al., 2023), which turns out to be a special case of Definition 1 corresponding to $I_t(X_0, X_1) = (1-t)X_0 + tX_1$.

We make some additional assumptions that reflect the typical instantiations of stochastic interpolation in practice. First, we specialize to the class of *isotropic linear* stochastic interpolants, where I_t is of the form

$$I_t(X_0, X_1) = \alpha_t X_0 + \beta_t X_1 \qquad \alpha_0 = \beta_1 = 1,\ \alpha_1 = \beta_0 = 0$$

which is a common parametrization in practice (Ma et al., 2024; Albergo and Vanden-Eijnden, 2023; Klein et al., 2023; Shaul et al., 2023). We assume that α_t, β_t are continuously differentiable as functions from $[0,1]$ to $\mathbb{R}$. Our second assumption is that $\mu_0 = \mathcal{N}(0, I_d)$, which is also typical in practice and which plays a similar role to that of the Gaussian distribution in diffusion modeling. In Section 4 we discuss results for non-Gaussian μ_0, μ_1. The simplest version of our main result is as follows.

Theorem 6 *Suppose that $\mu_0 = \mathcal{N}(0, I_d)$ and that $\mu_1(x) = e^{-V_1(x)}$ is a κ-log-concave density for $\kappa \geq 0$. Let $(\mu_t)_{t\in[0,1]}$ be the (isotropic) stochastic interpolant with coefficients $(\alpha_t)_{t\in[0,1]}, (\beta_t)_{t\in[0,1]}$ and let v_t be its drift. Then,*

1. *There exists $\phi_t \in C^1([0,1], C^1(\mathbb{R}^d))$ so that $v_t(x) = \nabla \phi_t(x)$.*
2. *The velocity Jacobian is a bounded symmetric matrix,*

$$D_x v_t(x) = \nabla^2 \phi_t(x) \preccurlyeq \frac{\kappa \alpha_t \dot{\alpha}_t + \beta_t \dot{\beta}_t}{\kappa \alpha_t^2 + \beta_t^2} I_d$$

3. *The flow $f_t^v(x)$ admits the bound,*

$$\|D_x f_t^v(x)\|_{op} \leq \sqrt{\alpha_t^2 + \beta_t^2/\kappa}$$

Since $\alpha_1 = 0$, $\beta_1 = 1$, Theorem 6 implies that the endpoint of *any* isotropic stochastic interpolant flow map f_1^v satisfies a bound that is analogous to Caffarelli's theorem. It is easy to check that the bound in Theorem 6 is tight by considering $\mu_1 = \mathcal{N}(0, \kappa I_d)$ in which case the stochastic interpolant flow is $f_1^v(x) = \sqrt{\kappa} x$. In fact, for any isotropic stochastic interpolation between Gaussian measures $\mu_0 = \mathcal{N}(0, \Sigma_0)$, $\mu_1 = \mathcal{N}(0, \Sigma_1)$ with commuting Σ_0, Σ_1, the flow map $f_1^v(x)$ is equal to the optimal transport map. We prove this fact in Proposition 15 by evaluating the closed form solution of Gaussian SI identified by Albergo et al. (2024), Appendix A.

The proof of Theorem 6 relies on the following definition, which expresses the conditional density of $X_1 \mid I_t(X_0, X_1) = x$ as an exponential family with parameter x.

Definition 7 *Let $X_t \sim \mu_t$ be the random variable $X_t = I_t(X_0, X_1)$ when $X_0 \sim \mathcal{N}(0, I_d)$ and $X_1 \sim \mu_1$ independently. Define the conditional density $\mu_{1,t}(x_1 \mid X_t = x)$ by,*

$$\mu_{1,t}(x_1 \mid X_t = x) := \exp\left(-V_1(x_1) - \frac{1}{2}\frac{\beta_t^2}{\alpha_t^2}\|x_1\|^2 + \frac{\beta_t}{\alpha_t^2}\langle x_1, x\rangle - b_t(x)\right)$$

where $b_t(x)$ is the cumulant generating function,

$$b_t(x) := \log \int \exp\left(-V_1(x_1) - \frac{1}{2}\frac{\beta_t^2}{\alpha_t^2}\|x_1\|^2 + \frac{\beta_t}{\alpha_t^2}\langle x_1, x\rangle\right) dx_1$$

Since $X_0 \sim \mathcal{N}(0, I_d)$, the joint density of (X_1, X_t) is proportional to $e^{-V_1(X_1) - \frac{1}{2}\|\alpha_t^{-1}(X_t - \beta_t X_1)\|^2}$, so that $\mu_{1,t}(\cdot \mid X_t = x)$ is indeed the conditional density of $X_1 \mid I_t(X_0, X_1) = x$. For the isotropic linear stochastic interpolant, the substitution $X_0 = \alpha_t^{-1}(I_t(X_0, X_1) - \beta_t X_1)$ yields

$$\begin{aligned} v_t(x) &= \mathbb{E}_{\mu_{1,t}}\left[\frac{\dot{\alpha}_t}{\alpha_t} X_t + \left(\dot{\beta}_t - \frac{\dot{\alpha}_t \beta_t}{\alpha_t}\right) X_1 \,\middle|\, X_t = x\right] \\ &= \frac{\dot{\alpha}_t}{\alpha_t} x + \left(\dot{\beta}_t - \frac{\dot{\alpha}_t \beta_t}{\alpha_t}\right)\left(\frac{\alpha_t^2}{\beta_t}\right) \nabla b_t(x) \end{aligned} \tag{4}$$

from which it is clear that $v_t(x) = \nabla \phi_t(x)$ for $\phi_t(x) = \frac{\dot{\alpha}_t}{2\alpha_t}\|x\|^2 + \left(\dot{\beta}_t - \frac{\dot{\alpha}_t \beta_t}{\alpha_t}\right)\left(\frac{\alpha_t^2}{\beta_t}\right) b_t(x)$. To prove Theorem (6), we bound the hessian of $b_t(x)$ and then transfer to a bound on $f_t^v(x)$ by Grönwall's lemma.

2. Related Work

Before discussing the proof of Theorem 6 and its generalizations, we consider the relationship between our result and some of the existing literature.

2.1. Dynamical Sampling Methods

Given a base measure μ_0, typically a Gaussian $\mu_0 = \mathcal{N}(0, I_d)$, and a target measure μ_1, there is wide variety of techniques developed to construct an interpolating trajectory of measures $(\mu_t)_{t\geq 0}$ along with a velocity field $(v_t)_{t\geq 0}$ which can be used to implement a dynamical sampler for the trajectory. Some methods prefer to simulate stochastic dynamics, where $(v_t)_{t\geq 0}$ is the drift of an SDE, while other methods prefer to simulate deterministic dynamics as in Proposition 2. Some examples include diffusion models Song et al. (2021), normalizing flows Papamakarios et al. (2021), stochastic interpolation Albergo et al. (2024), and Poisson flow models Xu et al. (2022), to name a few. In the literature on entropic optimal transport, dynamical sampling methods have also been developed for the purpose of sampling $(X_0, X_1) \sim \pi^\varepsilon$ the ε-regularized entropic optimal transport coupling of (μ_0, μ_1) Kassraie et al. (2024); Pooladian and Niles-Weed (2024).

If alongside (v_t) one also has access to the *score* of $s_t(x) := \nabla \log \mu_t(x)$, then it is possible to arbitrarily trade off the amount of stochasticity in the discretization of a dynamical sampler.

Proposition 8 *Suppose $(X_t)_{t\geq 0}$ is a stochastic process such that $X_t \sim \mu_t$ at each $t \geq 0$, and X_t solves the SDE*

$$dX_t = v_t(X_t)\,dt + \sqrt{2}\,dW_t$$

where $(W_t)_{t\geq 0}$ is a Wiener process on $\mathbb{R}^d$. Then for $\varepsilon \in [0,1]$, we have also $X_t' \sim \mu_t$, where $(X_t')_{t\geq 0}$ solves the SDE,

$$dX_t' = [v_t(X_t') - \varepsilon s_t(X_t')]\,dt + \sqrt{2(1-\varepsilon)}dW_t.$$

In the context of diffusion modeling, this observation is used by so-called 'probability flow ODE' models to convert SDE discretizations of diffusion to ODE discretizations (Song et al., 2021). As discussed in Chen et al. (2023a), given oracle access to $v_t(x)$ it is preferrable to discretize deterministic dynamics, since sampling error due to discretization scales as $O(\sqrt{d})$ in dimension rather than

$O(d)$ scaling of SDE-based discretizations. Fully deterministic dynamics are not robust to errors in estimating $v_t(x)$ or $s_t(x)$ and require a stochastic *corrector step* to alleviate these issues.

For the present work, it is important to highlight that stochastic interpolation with $\mu_0 = \mathcal{N}(0, I_d)$ bears many similarities to diffusion modeling. As highlighted by Albergo et al. (2023), the law of an Ornstein-Uhlenbeck process with initial condition $X_0 \sim \mu_1$ can be realized as a stochastic interpolant with adjusted time domain, $(\tilde{\mu}_t)_{t\in[0,\infty]}$, and with $\alpha_t = e^{-t}$, $\beta_t = \sqrt{1-e^{-2t}}$, $\lim_{t\to\infty} \tilde{\mu}_t = \mathcal{N}(0, I_d)$. The benefit of the stochastic interpolant framework is that optimizing over (α_t, β_t) can in fact lead to significant performance improvements Ma et al. (2024); Shaul et al. (2023). In this case Theorem 6 gives a sharp characterization of the contractivity of $v_t(x)$ in terms of α_t, β_t, which we hope will be useful for analyzing discretizations in practice.

2.2. Score estimation

Similar to the stochastic interpolant drift, the score $s_h(x) := \nabla \log \mu_1 * \mathcal{N}(0, h)(x)$ can be estimated by minimizing an empirical approximation to the risk

$$\mathcal{L}_{\text{sgm},h}[\hat{s}_h] = \mathbb{E}_{X_0,X_1}\left[\|\hat{s}(X_1 + \sqrt{h}X_0) - s_h(X_1 + \sqrt{h}X_0)\|^2\right] \tag{5}$$

$$\cong \mathbb{E}_{X_0,X_1}\left[\left\|h\hat{s}(X_1 + \sqrt{h}X_0) - \sqrt{h}X_0\right\|^2\right] \tag{6}$$

where $\cong$ means that the (5) shares the same minimizers as (6). This equivalence follows from *Tweedie's identity*, $hs_h(x) = x - \mathbb{E}[X_1 \mid X_1 + \sqrt{h}X_0 = x]$.

For diffusion models, it was shown in the seminal work (Chen et al., 2023b) that the risk (5) controls the discretization error for diffusion modeling with stochastic sampling. Oko et al. (2023) construct an empirical risk minimization procedure for (6) that achieves the rate $\tilde{O}(n^{-\frac{2s}{2s+d}})$ when the density μ_1 is a Besov function $\mu_1 \in B^s_{p,q}(\Omega)$ with uniformly upper and lower bounded density on its support $\Omega = [-1, 1]^d$. For an appropriate discretization, this leads to the rate $\mathbb{E}[\mathrm{TV}(\mu_1, \hat{\mu}_1)] = \tilde{O}(n^{-\frac{s}{2s+d}})$ where $\hat{\mu}_1$ is the the law of output samples, which matches the minimax rate up to log factors.

More recently, Wibisono et al. (2024) develop an estimator for $s_0(x) = \nabla \log \mu_1(x)$ that is compatible with densities which have unbounded support. Instead they require that μ_1 is α-sub-Gaussian and that $s_0(x)$ is L-Lipschitz. Analyzing the risk (5) is challenging with classical techniques and the authors make use of deep results on the estimation of Empirical Bayes denoising functions developed in (Saha and Guntuboyina, 2020; Jiang and Zhang, 2009). The equivalence between (5) and (6) shows the correspondence between score estimation and estimating an empirical bayes denoising function, and the literature suggests using a regularized estimator of the form

$$\hat{s}^\varepsilon_h(x) := \frac{\nabla(\hat{\mu}_1 * \mathcal{N}(0, h))(x)}{\max(\varepsilon, (\hat{\mu}_1 * \mathcal{N}(0, h))(x))}$$

where $\hat{\mu}_1$ is an empirical measure supported on $n \gg 1$ samples. For an appropriate choice of bandwidth $h \sim n^{-2/(d+4)}$ and thresholding $\varepsilon \sim n^{-2}$, the estimator $\hat{s}^\varepsilon_h$ approximates s_0 at the rate $\mathbb{E}\left[\|\hat{s}^\varepsilon_h - s_0\|^2_{\mathcal{L}^2(\mu_1)}\right] \leq \tilde{O}(n^{-2/(d+4)})$ which is minimax optimal under the assumptions on μ_1.

In our work, (4) provides an analogue of Tweedie's identity that also suggests an analogous estimator for the stochastic interpolant drift,

$$\hat{v}^h_{t,\varepsilon}(x) = \frac{\nabla \hat{\mu}^h_t(x)}{\max(\varepsilon, \hat{\mu}^h_t(x))} \qquad \hat{\mu}^h_t = \frac{1}{n}\sum_{i=1}^n \mathcal{N}(\beta_t X_1^{(i)}, \alpha_t^2)$$

where $X_1^{(i)}$, $i = 1 \ldots n$ are i.i.d. samples from μ_1. Analyzing the risk of this estimator is outside the scope of our work, but it is reasonable to make comparisons to the corresponding estimator for diffusion models, i.e. that the relevant parameter for uniformly estimating $v_t(x)$ is its Lipschitz constant. Applied to diffusion models,Wibisono et al. (2024) are able to bound $\mathbb{E}\mathrm{TV}(\hat{\mu}^h_{1,\varepsilon}, \mu_1) \leq \tilde{O}(n^{-1/(d+4)})$ where $\hat{\mu}^h_{1,\varepsilon}$ is the output of an appropriate diffusion model, under the assumption that $(s_t)_{t\geq 0}$ are uniformly L-lipschitz. Theorem 6 suggests choosing (α_t, β_t) so that $\kappa\alpha_t^2 + \beta_t^2 = \kappa^{1-t}$, which leads to the uniform bound

$$\|Dv_t(x)\|_{\text{op}} \leq \left| \frac{d}{dt} \log\left(\kappa\alpha_t^2 + \beta_t^2\right) \right| = \kappa.$$

and which is compatible with the endpoint constraints on $(\alpha_t), (\beta_t)$.

The Empirical bayes techniques used by (Wibisono et al., 2024) depend heavily on Gaussianity of the noise, which appears to be an obstacle in extending their analysis to have improved rates if $s_0(x)$ is very smooth. As we will discuss further in Section 2.3, the bounds we show in Theorem 11 for non-Gaussian μ_0, μ_1 are worse than the corresponding bounds on the Lipschitz constant of optimal transport maps. We suspect that Theorem 11 may be sub-optimal due to crude bounds used in our proof, which we can avoid when μ_0 is Gaussian. It may be interesting to understand the sharp constant on the contractivity of the stochastic interpolant flow map, given that the corresponding constant is known for optimal transport Caffarelli (2000); Chewi and Pooladian (2023).

2.3. Lipschitz Transport

Lipschitz transportation maps play an important role in the theory of functional inequalities for probability measures. This is because if $\mu_1 = (T)_{\#}\mu_0$ where T is L-lipschitz, then μ_1 inherits functional inequalities which may be satisfied by μ_0 with a constant depending on L but not on the ambient dimension. One can therefore reduce the problem of proving dimension independent functional inequalities for a family of measures to the problem of constructing a Lipschitz transport between any member and a target measure μ_0 satisfying the desired inequalities.

Caffarelli's theorem establishes such a transport map between a Gaussian measure and any log concave measure. It was recently explored the possibility of constructing alternative transportation maps from finite dimensional Mikulincer and Shenfeld (2023) and infinite dimensional Mikulincer and Shenfeld (2024) Gaussian measures to μ_1 which is either log-concave, or which has bounded support, leading to new inequalities on the Poincaré and log-Sobolev constants of (for example) convolutions of Gaussian measures with bounded mixing weights, as well as to drastically simplified proofs of existing inequalities (Mikulincer and Shenfeld, 2024, Table 1). In (Dai et al., 2023; Ding et al., 2023), the authors construct a deterministic version of the Brownian transport map Mikulincer and Shenfeld (2024), and prove it is Lipschitz using techniques similar to Theorem 6. The novelty of the latter theorem is that it holds for *any* isotropic linear stochastic interpolant, allowing to understand how different choices of coefficients affect the contractivity of the velocity field.

3. Proof of Theorem 6

We will now give a simple proof of Theorem 6 that is based on the classical Brascamp-Lieb inequality (Bobkov and Ledoux, 2000; Bakry et al., 2014).

Theorem 9 (Brascamp-Lieb) *If $\mu(x) = e^{-V(x)}$ is strictly log-concave, then for $f : \mathbb{R}^d \to \mathbb{R}$,*

$$\operatorname*{var}_{X\sim\mu}[f] \leq \mathbb{E}_{X\sim\mu}\left[\langle \nabla f(X), (\nabla^2 V(X))^{-1} \nabla f(X)\rangle\right].$$

By Lemma 7, the velocity Jacobian reads

$$Dv_t(x) = \frac{\dot{\alpha}_t}{\alpha_t} I_d + \left(\dot{\beta}_t - \frac{\dot{\alpha}_t \beta_t}{\alpha_t}\right)\left(\frac{\alpha_t^2}{\beta_t}\right) \nabla^2 b_t(x)$$

where, using the fact that $b_t(x)$ is a cumulant generating function,

$$\nabla^2 b_t(x) = \left(\frac{\beta_t^2}{\alpha_t^4}\right) \mathrm{cov}_t[\mathsf{id}, \mathsf{id}]$$

for the covariance operator

$$\mathrm{cov}_t[f, g] = \mathbb{E}[(f(\hat{X}_1) - \mathbb{E}[f(\hat{X}_1)])(g(\hat{X}_1) - \mathbb{E}[g(\hat{X}_1)])^T] \qquad \hat{X}_1 \sim \mu_{1,t}(\cdot \mid X_t = x).$$

Now fix $w \in \mathbb{R}^d$ an arbitrary vector with $\|w\|_2 \leq 1$. Applying Brascamp-Lieb with test function $f_w(x) = \langle w, x\rangle$ yields

$$w^T \mathrm{cov}_t[\mathsf{id}, \mathsf{id}] w = \mathrm{cov}_t[f_w, f_w] \leq \mathbb{E}\left[\langle w, (\nabla^2 V_t(\hat{X}_1))^{-1} w\rangle\right] \qquad \hat{X}_1 \sim \mu_{1,t}(\cdot \mid X_t = x)$$

for the stochastic interpolant potential $V_t(x_1) := V_1(x_1) + \frac{1}{2}\frac{\beta_t^2}{\alpha_t^2}\|x_1\|^2 - \frac{\beta_t}{\alpha_t^2}\langle x_1, x\rangle$ with,

$$(\nabla^2 V_t(x))^{-1} \preccurlyeq \left(\kappa + \frac{\beta_t^2}{\alpha_t^2}\right)^{-1} I_d,$$

so that

$$w^T Dv_t(x) w \leq \frac{\dot{\alpha}_t}{\alpha_t} + \left(\dot{\beta}_t - \frac{\dot{\alpha}_t \beta_t}{\alpha_t}\right)\left(\frac{\beta_t}{\alpha_t^2}\right)\left(\kappa + \frac{\beta_t^2}{\alpha_t^2}\right)^{-1} = \frac{\kappa \dot{\alpha}_t \alpha_t + \dot{\beta}_t \beta_t}{\kappa \alpha_t^2 + \beta_t^2}.$$

The bound on $\|Df_t^v(x)\|_{\mathrm{op}}$ follows from Grönwall's argument which we provide for completeness in Lemma 14.

4. Non-Gaussian Endpoints

Although the assumption $\mu_0 = \mathcal{N}(0, I_d)$ is typical in practical applications, it is of interest to check whether the techniques in the previous section can be applied to non-Gaussian μ_0. In this section we work under the assumptions that $\mu_0(x) = e^{-V_0(x)}$ and $\mu_1(x) = e^{-V_1(x)}$ satisfy the log-concavity bounds,

$$0 \preccurlyeq \nabla^2 V_0(x) \preccurlyeq A \qquad 0 \prec B \preccurlyeq \nabla^2 V_1(x) \qquad B \preccurlyeq A$$

for $x \in \mathbb{R}^d$, $\kappa_0^- > 0$, and where $A, B \in \mathbb{R}^{d\times d}$ are commuting, symmetric positive semidefinite matrices. Define $\kappa_0^+ = \lambda_{\max}(A)$ and $\kappa_1^- = \lambda_{\min}(B)$ be the upper and lower eigenvalues of A, B, respectively, and define also $\kappa_0^- = \inf_{x\in\mathbb{R}^d} \lambda_{\min}(\nabla^2 V_0(x))$ and $\kappa_1^+ = \sup_{x\in\mathbb{R}^d} \lambda_{\max}(\nabla^2 V_1(x))$. As a point of comparison, Chewi and Pooladian (2023) prove the following generalized form of Caffarelli's theorem, which we state in our notation.

Proposition 10 (Theorem 5, (Chewi and Pooladian, 2023)) *Let $P(x) = e^{-V(x)}, Q = e^{-W(x)}$ be probability measures satisfying $\nabla^2 V(x) \preccurlyeq A$, $\nabla^2 W(x) \succcurlyeq B$ for all $x \in \mathbb{R}^d$, where A, B are commuting positive definite matrices. Then the Brenier potential ϕ which transports $P = (\nabla\phi)_\# Q$ satisfies*

$$\nabla^2\phi(x) \preccurlyeq A^{1/2}B^{-1/2}.$$

For Gaussians $\mu_0 = \mathcal{N}(0, A^{-1})$ and $\mu_1 = \mathcal{N}(0, B^{-1})$, the Brenier potential is equal to $\nabla\phi(x) = A^{1/2}B^{-1/2}x$, so the bound cannot be improved in general. If A, B are multiples of the identity, it reduces to $\nabla^2\phi(x) \preccurlyeq \sqrt{\kappa_0^+/\kappa_1^-}\, I_d$. For stochastic interpolants, we are able to show the following bound.

Theorem 11 *Let (μ_t) be the stochastic interpolant between μ_0 and μ_1 with velocity field (v_t), for isotropic coefficients (α_t, β_t) such that $t \mapsto \alpha_t^2 + \beta_t^2$ is constant and $\alpha_t\dot\beta_t - \beta_t\dot\alpha_t \neq 0$ for $t \in (0, 1)$. Then,*

$$\|Df_t^v(x)\|_{op} \le \exp\left(\int_0^t \left(\frac{\dot\alpha_t\alpha_t\kappa_1^- + \dot\beta_t\beta_t\kappa_0^+}{(\alpha_t^2\kappa_1^- + \beta_t^2\kappa_0^+)^{1/2}(\alpha_t^2\kappa_1^- + \beta_t^2\kappa_0^-)^{1/2}}\right) dt\right)$$

To the best of our knowledge, this is the first example of a general purpose alternative to Cafarelli's theorem for constructing Lipschitz transport maps between log-concave μ_1 and μ_0 which is not a Gaussian measure. We can interpret this bound by considering a few choices of (α_t, β_t). First, if α_t decreases monotonically, we have a bound that can be compared directly to Proposition 10.

Corollary 12 *In the setting of Theorem 11, if α_t is decreasing monotonically and $\kappa_0^- > 0$, then*

$$\|Df_t^v(x)\|_{op} \le \left(\frac{\kappa_0^+}{\kappa_1^-}\right)^{\sqrt{\kappa_0^+/\kappa_0^-}}.$$

This follows by observing that if $\beta_t = \sqrt{1-\alpha_t^2}$ increases monotonically,

$$\sup_{t\in[0,1]} \frac{\alpha_t^2\kappa_1^- + \beta_t^2\kappa_0^+}{\alpha_t^2\kappa_1^- + \beta_t^2\kappa_0^-} = \frac{\kappa_0^+}{\kappa_0^-}$$

so the integrand of Theorem 11 has the upper bound

$$\frac{\dot\alpha_t\alpha_t\kappa_1^- + \dot\beta_t\beta_t\kappa_0^+}{(\alpha_t^2\kappa_1^- + \beta_t^2\kappa_0^+)^{1/2}(\alpha_t^2\kappa_1^- + \beta_t^2\kappa_0^-)^{1/2}} \le \left(\frac{d}{dt}\log(\kappa_1^-\alpha_t^2 + \kappa_0^+\beta_t^2)\right)\cdot\left(\frac{\kappa_0^+}{\kappa_0^-}\right)^{1/2}.$$

It is also possible to use Theorem 11 to bound the Lipschitz constant of f_t^v when μ_0 is strictly log concave but with vanishing Hessian, i.e. $\kappa_0^- = 0$. We must however select an α_t that decays sufficiently quickly as $t \to 0$, which leads an integrable singularity in the bound on $Dv_t(x)$ that we extract from Brascamp-Lieb.

Corollary 13 *In the setting of Theorem 11, if $\kappa_0^- = 0$ and if $\alpha_t = O(t^{1/2+\varepsilon})$, then $\|Df_1^v\| < \infty$.*

Proof This follows by applying Cauchy-Schwartz inequality to the integral,

$$\int_0^1 \left(\frac{d}{dt}\log(\kappa_1^-\alpha_t^2 + \kappa_0^+\beta_t^2)\right)\left(1 + \frac{\beta_t^2\kappa_0^+}{\alpha_t^2\kappa_1^-}\right) dt$$
$$= \left(\int_0^1 \left(\frac{d}{dt}\log(\kappa_1^-\alpha_t^2 + \kappa_0^+\beta_t^2)\right)^2 dt\right)^{1/2} \left(1 + \int_0^1 \frac{\beta_t^2\kappa_0^+}{\alpha_t^2\kappa_1^-}\, dt\right)^{1/2} < \infty.$$

■

5. Conclusion and future directions

Our approach to proving Theorem 6 is flexible enough to extend to the cases when μ_0 is non-Gaussian, but it is unclear if the gap in Theorem 11 can be improved or if it is intrinsic to stochastic interpolation. We nevertheless believe that it is a valuable and flexible tool for constructing Lipschitz transport maps. One especially interesting direction for future work is to try to generalize to the case where μ_0, μ_1 are supported on a generic Riemannian manifold. In this setting one must take into account the curvature of the ambient space, and constructing Lipschitz transport maps generically between concentrating measures is significantly more challenging, whereas the Brascamp-Lieb inequality still holds and it is easy to construct $I_t(X_0, X_1)$ using geodesics.

Orthogonal to this, it may also be interesting to identify a choice of $I_t(X_0, X_1)$ such that $v_t(x)$ is *non-conservative*, i.e., it is not the gradient of a function. We believe that it may be possible to improve the Lipschitz constant using a flow that rotates. To gain intuition, consider the distributions

$$\mu_0 = \mathcal{N}\left(0, \begin{bmatrix}\kappa^{-1} & 0 \\ 0 & 1\end{bmatrix}\right) \qquad \mu_1 = \mathcal{N}\left(0, \begin{bmatrix}1 & 0 \\ 0 & \kappa^{-1}\end{bmatrix}\right)$$

for which the both optimal transport map and the stochastic interpolation transport map are equal to

$$T(x) = \begin{bmatrix}\sqrt{\kappa} & 0 \\ 0 & \sqrt{1/\kappa}\end{bmatrix} x$$

whose Lipschitz constant diverges with the log-concavity parameter $\kappa \to 0$. The transport map $U(x)$ given by

$$U(x) = \begin{bmatrix}0 & 1 \\ 1 & 0\end{bmatrix}$$

has Lipschitz constant $\|U\|_{\text{op}} = 1$, but it is easy to see that $U(x)$ cannot be the gradient of any function $\phi : \mathbb{R}^d \to \mathbb{R}$. We hope that by tailoring $I_t(\cdot, \cdot)$ to μ_0, μ_1 and incorporating *rotational alignment* of the two distributions, it will be possible to automatically estimate maps like $U(x)$ from data, which would in some cases have *stronger* contractivity properties than that of optimal transport.

References

Michael S. Albergo, Nicholas M. Boffi, and Eric Vanden-Eijnden. Stochastic interpolants: A unifying framework for flows and diffusions, 2023. URL https://arxiv.org/abs/2303.08797.

Michael S. Albergo, Mark Goldstein, Nicholas M. Boffi, Rajesh Ranganath, and Eric Vanden-Eijnden. Stochastic interpolants with data-dependent couplings. In *Proceedings of the 41st International Conference on Machine Learning*, 2024. URL https://arxiv.org/abs/2310.03725.

Michael Samuel Albergo and Eric Vanden-Eijnden. Building normalizing flows with stochastic interpolants. In *The Eleventh International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=li7qeBbCR1t.

Dominique Bakry, Ivan Gentil, and Michel Ledoux. *Analysis and Geometry of Markov Diffusion Operators*. Springer, Cham, 2014. ISBN 978-3-319-00227-9. doi: 10.1007/978-3-319-00227-9. URL https://link.springer.com/book/10.1007/978-3-319-00227-9.

Jean-David Benamou and Yann Brenier. A computational fluid mechanics solution to the Monge–Kantorovich mass transfer problem. *Numerische Mathematik*, 84(3):375–393, January 2000. doi: 10.1007/s002110050002.

Joe Benton, Valentin De Bortoli, Arnaud Doucet, and George Deligiannidis. Nearly d-linear convergence bounds for diffusion models via stochastic localization. In *International Conference on Learning Representations*, 2024. URL https://openreview.net/forum?id=r5njV3BsuD.

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. doi: 10.1007/PL00001645.

Luis A. Caffarelli. Monotonicity Properties of Optimal Transportation and the FKG and Related Inequalities. *Communications in Mathematical Physics*, 214(3):547–563, January 2000. doi: 10.1007/s002200000257.

Sitan Chen, Sinho Chewi, Holden Lee, Yuanzhi Li, Jianfeng Lu, and Adil Salim. The probability flow ode is provably fast. In *Proceedings of the 37th International Conference on Neural Information Processing Systems*, NIPS '23, Red Hook, NY, USA, 2023a. Curran Associates Inc.

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 l2 accuracy. In *International Conference on Learning Representations*, 2023b. URL https://openreview.net/forum?id=zyLVMgsZ0U_.

Sinho Chewi and Aram-Alexandre Pooladian. An entropic generalization of caffarelli's contraction theorem via covariance inequalities. *Comptes Rendus Mathématique*, 361(9):1471–1482, 2023. doi: 10.5802/crmath.452. URL https://doi.org/10.5802/crmath.452.

Yin Dai, Yuan Gao, Jian Huang, Yuling Jiao, Lican Kang, and Jin Liu. Lipschitz transport maps via the follmer flow, 2023. URL https://arxiv.org/abs/2309.03490.

Zhao Ding, Yuling Jiao, Xiliang Lu, Zhijian Yang, and Cheng Yuan. Sampling via föllmer flow, 2023. URL https://arxiv.org/abs/2311.03660.

Wenhua Jiang and Cun-Hui Zhang. General maximum likelihood empirical bayes estimation of normal means. *The Annals of Statistics*, 37(4):1647–1684, 2009. ISSN 00905364, 21688966. URL http://www.jstor.org/stable/30243683.

Kacper Kapusniak, Peter Potaptchik, Teodora Reu, Leo Zhang, Alexander Tong, Michael M. Bronstein, Avishek Joey Bose, and Francesco Di Giovanni. Metric flow matching for smooth interpolations on the data manifold. In *Advances in Neural Information Processing Systems*, 2024. URL https://arxiv.org/abs/2405.14780.

Parnian Kassraie, Aram-Alexandre Pooladian, Michal Klein, James Thornton, Jonathan Niles-Weed, and marco cuturi. Progressive entropic optimal transport solvers. In *The Thirty-eighth Annual Conference on Neural Information Processing Systems*, 2024. URL https://openreview.net/forum?id=7WvwzuYkUq.

Leon Klein, Andreas Krämer, and Frank Noe. Equivariant flow matching. In *Thirty-seventh Conference on Neural Information Processing Systems*, 2023. URL https://openreview.net/forum?id=eLH2NFOO1B.

Frederic Koehler, Alexander Heckett, and Andrej Risteski. Statistical efficiency of score matching: The view from isoperimetry. In *International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=TD7AnQjNzR6.

Gen Li and Yuling Yan. Adapting to unknown low-dimensional structures in score-based diffusion models. In *Advances in Neural Information Processing Systems*, 2024. URL https://arxiv.org/abs/2405.14861.

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. URL https://openreview.net/forum?id=PqvMRDCJT9t.

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. URL https://openreview.net/forum?id=XVjTT1nw5z.

Nanye Ma, Mark Goldstein, Michael S. Albergo, Nicholas M. Boffi, Eric Vanden-Eijnden, and Saining Xie. SiT: Exploring flow and diffusion-based generative models with scalable interpolant transformers. In *Proceedings of the European Conference on Computer Vision (ECCV)*, 2024. URL https://arxiv.org/abs/2401.08740.

Dan Mikulincer and Yair Shenfeld. *On the Lipschitz Properties of Transportation Along Heat Flows*, pages 269–290. Springer International Publishing, Cham, 2023. ISBN 978-3-031-26300-2. doi: 10.1007/978-3-031-26300-2_9. URL https://doi.org/10.1007/978-3-031-26300-2_9.

Dan Mikulincer and Yair Shenfeld. The brownian transport map. *Probability Theory and Related Fields*, 190(1):379–444, Oct 2024. ISSN 1432-2064. doi: 10.1007/s00440-024-01286-0. URL https://doi.org/10.1007/s00440-024-01286-0.

Kazusato Oko, Shunta Akiyama, and Taiji Suzuki. Diffusion models are minimax optimal distribution estimators. In *Proceedings of the 40th International Conference on Machine Learning*, ICML'23. JMLR.org, 2023.

Chirag Pabbaraju, Dhruv Rohatgi, Anish Prasad Sevekari, Holden Lee, Ankur Moitra, and Andrej Risteski. Provable benefits of score matching. In A. Oh, T. Naumann, A. Globerson, K. Saenko, M. Hardt, and S. Levine, editors, *Advances in Neural Information Processing Systems*, volume 36, pages 61306–61326. Curran Associates, Inc., 2023. URL https://proceedings.neurips.cc/paper_files/paper/2023/file/c11f8d40c119867e30e3421f696f931d-Paper-Conference.pdf.

George Papamakarios, Eric Nalisnick, Danilo Jimenez Rezende, Shakir Mohamed, and Balaji Lakshminarayanan. Normalizing flows for probabilistic modeling and inference. *J. Mach. Learn. Res.*, 22(1), January 2021. ISSN 1532-4435.

Aram-Alexandre Pooladian and Jonathan Niles-Weed. Plug-in estimation of schrödinger bridges, 2024. URL https://arxiv.org/abs/2408.11686.

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*, 2023. URL https://arxiv.org/abs/2304.14772.

Sujayam Saha and Adityanand Guntuboyina. On the nonparametric maximum likelihood estimator for Gaussian location mixture densities with application to Gaussian denoising. *The Annals of Statistics*, 48(2):738 – 762, 2020. doi: 10.1214/19-AOS1817. URL https://doi.org/10.1214/19-AOS1817.

Neta Shaul, Ricky T. Q. Chen, Maximilian Nickel, Matt Le, and Yaron Lipman. On kinetic optimal probability paths for generative models. In *Proceedings of the 40th International Conference on Machine Learning*, 2023. URL https://arxiv.org/abs/2306.06626.

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*, 2021. URL https://openreview.net/forum?id=PxTIG12RRHS.

Hannes Stark, Bowen Jing, Chenyu Wang, Gabriele Corso, Bonnie Berger, Regina Barzilay, and Tommi Jaakkola. Dirichlet flow matching with applications to dna sequence design. In *Proceedings of the 41st International Conference on Machine Learning*, ICML'24. JMLR.org, 2024.

Andre Wibisono, Yihong Wu, and Kaylee Yingxi Yang. Optimal score estimation via empirical bayes smoothing. In Shipra Agrawal and Aaron Roth, editors, *Proceedings of Thirty Seventh Conference on Learning Theory*, volume 247 of *Proceedings of Machine Learning Research*, pages 4958–4991. PMLR, 30 Jun–03 Jul 2024. URL https://proceedings.mlr.press/v247/wibisono24a.html.

Yilun Xu, Ziming Liu, Max Tegmark, and Tommi Jaakkola. Poisson flow generative models. In S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh, editors, *Advances in Neural Information Processing Systems*, volume 35, pages 16782–16795. Curran Associates, Inc., 2022. URL https://proceedings.neurips.cc/paper_files/paper/2022/file/6ad68a54eaa8f9bf6ac698b02ec05048-Paper-Conference.pdf.

Appendix A. Grönwall Lemma

Lemma 14 *Let $\lambda_t : [0, 1] \to \mathbb{R}$ be such that*

$$\|Dv_t(x)\|_{op} \leq \lambda_t$$

for any $x \in \mathbb{R}^d$. The flow map Jacobian satisfies the bound

$$\|Df_t^v\|_{op} \leq e^{\int_0^t \lambda_s \, ds}.$$

Proof Define $\alpha_w(t) = Df_t^v(x)w$. We have the differential inequality

$$\begin{aligned}
\partial_t \|\alpha_w(t)\|_2 &= \frac{1}{\|\alpha_w(t)\|_2} \alpha_w(t)^T \partial_t \alpha_w(t) \\
&= \frac{1}{\|\alpha_w(t)\|_2} w^T (Df_t^v(x))^T Dv_t(f_t^v(x))(Df_t^v(x))w \\
&\leq \lambda_t \frac{1}{\|\alpha_w(t)\|_2} w^T (Df_t^v(x))^T (Df_t^v(x))w \\
&= \lambda_t \|\alpha_w(t)\|_2.
\end{aligned}$$

Since $\|\alpha_w(0)\| = \|w\|_2$ it follows that $\|\alpha_w(t)\|_2 \leq \|w\|_2 \exp\left(\int_0^t \lambda_s \, ds\right)$. ■

Appendix B. Stochastic interpolation and optimal transport for Gaussian endpoints

We will show that, if μ_0 and μ_1 are Gaussian measures with commuting covariance matrices, then the isotropic stochastic interpolant yields a flow map that coincides with the optimal transport map between the two Gaussians. Recall that the optimal transport map is given by

$$T(x) = m_1 + \Sigma_1^{1/2} \Sigma_0^{-1/2} (x - m_0).$$

The stochastic interpolant flow map can also be calculated in closed form. Interestingly, it has the property that $f_1^v(x) = T(x)$, regardless of the choice of coefficients α_t, β_t, which mirrors the behavior of Theorem 6.

Proposition 15 *Suppose that $\mu_0 = \mathcal{N}(m_0, \Sigma_0)$ and $\mu_1 = \mathcal{N}(m_1, \Sigma_1)$ where Σ_0, Σ_1 are commuting positive definite matrices. Let $(\mu_t)_{t \in [0,1]}$ be the isotropic stochastic interpolant with coefficients $(\alpha_t)_{t \in [0,1]}$, $(\beta_t)_{t \in [0,1]}$, and let v_t, f_t^v be the drift and flow respectively. Then,*

$$\begin{aligned}
v_t(x) &= \dot{m}_t + \frac{1}{2} \dot{\Sigma}_t \Sigma_t^{-1} (x - m_t) \\
f_t^v(x) &= m_t + \Sigma_t^{1/2} \Sigma_0^{-1/2} (x - m_t)
\end{aligned}$$

where $m_t = \alpha_t m_0 + \beta_t m_1$ and $\Sigma_t = \alpha_t^2 \Sigma_0 + \beta_t^2 \Sigma_1$.

Proof In this setting, the stochastic interpolant drift $v_t(x)$ is computed explicitly in (Albergo et al., 2023, Equation (A.7)). Hence,

$$\partial_t(f_t^v(x) - m_t) = v_t(f_t^v(x)) - \dot{m}_t = \frac{1}{2}\dot{\Sigma}_t\Sigma_t^{-1}(f_t^v(x) - m_t).$$

Since Σ_0, Σ_1 commute, there exists an orthonormal matrix $U \in \mathbb{R}^{d\times d}$ so that $\Sigma_0 = U\Lambda_0 U^T$, $\Sigma_1 = U\Lambda_1 U^T$ are simultaneously diagonalized. Then $\Sigma_t = U\Lambda_t U^T$ for $\Lambda_t = \alpha_t^2\Lambda_0 + \beta_t^2\Lambda_1$ and, by linearity, we have

$$\partial_t U^T(f_t^v(x) - m_t) = \dot{\Lambda}_t\Lambda_t^{-1}U^T(f_t^v - m_t).$$

Let $w_{i,t} = [U^T(f_t^v(x) - m_t)]_i$ be the i-th coordinate and let $\lambda_{i,t}$ be the i-th diagonal entry of Λ_t. By the previous display we have,

$$\partial_t w_{i,t} = \frac{1}{2}\frac{\dot{\lambda}_{i,t}}{\lambda_{i,t}}w_{i,t} \implies w_{i,t} = \exp\left(\frac{1}{2}(\log(\lambda_{i,t}) - \log(\lambda_{i,0}))\right)w_0 = \sqrt{\frac{\lambda_{i,t}}{\lambda_{i,0}}}w_{i,0}.$$

It follows that,

$$U^T(f_t^v(x) - m_t) = \Lambda_t^{1/2}\Lambda_0^{-1/2}U^T(f_t^v(x) - m_t)$$

from which the desired representation follows by rearranging. ∎

Appendix C. Non-Gaussian Endpoints

We assume in this section that $\mu_0(x) = e^{-V_0(x)}$ and $\mu_1(x) = e^{-V_1(x)}$ satisfy the log-concavity bounds

$$0 \preccurlyeq \nabla^2 V_0(x) \preccurlyeq A \qquad 0 \prec B \preccurlyeq \nabla^2 V_1(x) \qquad B \preccurlyeq A$$

for $x \in \mathbb{R}^d$, $\kappa_0^- > 0$, and where $A, B \in \mathbb{R}^{d\times d}$ are commuting, symmetric positive semidefinite matrices. Define $\kappa_0^+ = \lambda_{\max}(A)$ and $\kappa_1^- = \lambda_{\min}(B)$ be the upper and lower eigenvalues of A, B, respectively, and define also $\kappa_0^- = \inf_{x\in\mathbb{R}^d}\lambda_{\min}(\nabla^2 V_0(x))$ and $\kappa_1^+ = \sup_{x\in\mathbb{R}^d}\lambda_{\max}(\nabla^2 V_1(x))$. We begin by rewriting slightly Definition 1 in terms of the variables

$$\begin{bmatrix} X_t \\ R_t \end{bmatrix} := \underbrace{\begin{bmatrix} \alpha_t I_d & \beta_t I_d \\ \dot{\alpha}_t I_d & \dot{\beta}_t I_d \end{bmatrix}}_{:=J_t} \begin{bmatrix} X_0 \\ X_1 \end{bmatrix}$$

and we assume further that J_t is invertible at each $t \in [0, 1]$. We will prove the following.

Theorem 16 *Let (μ_t) be the stochastic interpolant between μ_0 and μ_1 with velocity field (v_t), for isotropic coefficients (α_t, β_t) such that $t \mapsto \alpha_t^2 + \beta_t^2$ is constant and $\alpha_t\dot{\beta}_t - \beta_t\dot{\alpha}_t \neq 0$ for $t \in (0, 1)$. Then,*

$$\|Df_t^v(x)\|_{op} \leq \exp\left(\int_0^t \left(\frac{\dot{\alpha}_t\alpha_t\kappa_1^- + \dot{\beta}_t\beta_t\kappa_0^+}{(\alpha_t^2\kappa_1^- + \beta_t^2\kappa_0^+)^{1/2}(\alpha_t^2\kappa_1^- + \beta_t^2\kappa_0^-)^{1/2}}\right)dt\right)$$

To interpret this bound, suppose first that $\kappa_0^- > 0$ and α_t decreases monotonically. Then since $\alpha_t^2 + \beta_t^2$ is constant, we have

$$\sup_{t\in[0,1]} \frac{\alpha_t^2\kappa_1^- + \beta_t^2\kappa_0^+}{\alpha_t^2\kappa_1^- + \beta_t^2\kappa_0^-} = \frac{\kappa_0^+}{\kappa_0^-}$$

and so the integrand is bounded by

$$\begin{aligned}\frac{\dot\alpha_t\alpha_t\kappa_1^- + \dot\beta_t\beta_t\kappa_0^+}{(\alpha_t^2\kappa_1^- + \beta_t^2\kappa_0^+)^{1/2}(\alpha_t^2\kappa_1^- + \beta_t^2\kappa_0^-)^{1/2}} &\le \frac{\dot\alpha_t\alpha_t\kappa_1^- + \dot\beta_t\beta_t\kappa_0^+}{\alpha_t^2\kappa_1^- + \beta_t^2\kappa_0^+}\cdot\left(\frac{\kappa_0^+}{\kappa_0^-}\right)^{1/2}\\ &= \left(\frac{d}{dt}\log(\kappa_1^-\alpha_t^2 + \kappa_0^+\beta_t^2)\right)\cdot\left(\frac{\kappa_0^+}{\kappa_0^-}\right)^{1/2}.\end{aligned}$$

and the bound simplifies as in the following corollary.

Corollary 17 *In the setting of Theorem 11, if α_t is decreasing monotonically and $\kappa_0^- > 0$, then*

$$\|Df_t^v(x)\|_{op} \le \left(\frac{\kappa_0^+}{\kappa_1^-}\right)^{\sqrt{\kappa_0^+/\kappa_0^-}}.$$

It is also possible to use Theorem 11 to bound the Lipschitz constant of f_t^v when $\kappa_0^- = 0$.

Corollary 18 *In the setting of Theorem 11, if $\kappa_0^- = 0$ and if $\alpha_t = O(t^{1/2+\varepsilon})$, then $\|Df_1^v\| < \infty$.*

Proof Apply Cauchy-Schwartz to the integral,

$$\begin{aligned}&\int_0^1 \left(\frac{d}{dt}\log(\kappa_1^-\alpha_t^2 + \kappa_0^+\beta_t^2)\right)\left(1 + \frac{\beta_t^2\kappa_0^+}{\alpha_t^2\kappa_1^-}\right)dt\\ &= \left(\int_0^1\left(\frac{d}{dt}\log(\kappa_1^-\alpha_t^2 + \kappa_0^+\beta_t^2)\right)^2 dt\right)^{1/2}\left(1 + \int_0^1 \frac{\beta_t^2\kappa_0^+}{\alpha_t^2\kappa_1^-}dt\right)^{1/2} < \infty.\end{aligned}$$

■

Now we will prove Theorem 11.

Lemma 19 *The stochastic interpolant has at time $t \in (0,1)$, $R_t \mid X_t = x \sim \tilde\mu_{1,t}$, with density*

$$\tilde\mu_{1,t}(r_t \mid X_t = x_t) = Z^{-1}(x_t)\exp\left(-V_t(x_t, r_t)\right)|\det J_t^{-1}| \tag{7}$$

$$V_t(x_t, r_t) \coloneqq V_0\left(\frac{\dot\beta_t x_t - \beta_t r_t}{\alpha_t\dot\beta_t - \dot\alpha_t\beta_t}\right) + V_1\left(\frac{\alpha_t r_t - \dot\alpha_t x_t}{\alpha_t\dot\beta_t - \dot\alpha_t\beta_t}\right) \tag{8}$$

where $Z^{-1}(x_t)$ is a normalization constant. The velocity $v_t(x) = \mathbb{E}[R_t \mid X_t = x]$ has Jacobian matrix,

$$Dv_t(x) = \mathrm{cov}_{\tilde\mu_{1,t}}[\mathsf{id}, -\nabla_x V_t(\cdot, x)]$$

where $\mathrm{cov}_{\tilde\mu_{1,t}}[f,g] = \mathbb{E}[f(R_t)g(R_t)^T] - \mathbb{E}[f(R_t)]\mathbb{E}[g(R_t)]^T$ over $R_t \sim \tilde\mu_{1,t}(\cdot \mid X_t = x)$.

Proof The density (7) follows by the change of variables formula with $(X_t, R_t) = J_t(X_0, X_1)$ and the fact that

$$J_t^{-1} = \frac{1}{\alpha_t \dot{\beta}_t - \dot{\alpha}_t \beta_t} \begin{bmatrix} \dot{\beta}_t I_d & -\beta_t I_d \\ -\dot{\alpha}_t I_d & \alpha_t I_d \end{bmatrix}.$$

Differentiating $v_t(x)$ yields

$$\begin{aligned} Dv_t(x) &= D_x \left\{ \frac{\int R_t e^{-V_t(x,R_t)}\, dR_t}{\int e^{-V_t(x,R_t)}\, dR_t} \right\} \\ &= \frac{\int R_t(-\nabla_x V_t(x, R_t))^T e^{-V_t(x,R_t)}\, dR_t}{\int e^{-V_t(x,R_t)}\, dR_t} \\ &\quad - \left(\frac{\int R_t e^{-V_t(x,R_t)}\, dR_t}{\int e^{-V_t(x,R_t)}\, dR_t} \right) \left(\frac{\int (-\nabla_x V_t(x, R_t)) e^{-V_t(x,R_t)}\, dR_t}{\int e^{-V_t(x,R_t)}\, dR_t} \right) \\ &= \mathrm{cov}_{\tilde{\mu}_{1,t}}[\mathsf{id}, -\nabla_x V_t(\cdot, x)]. \end{aligned}$$

■

Here is the main lemma needed to prove Theorem 11.

Lemma 20 *Suppose that $\alpha_t^2 + \beta_t^2$ is constant. The velocity jacobian satisfies*

$$\sup_{\substack{u,v \in \mathbb{R}^d \\ \|u\| \le 1}} u^T Dv_t(x) v \le \frac{\dot{\alpha}_t \alpha_t \kappa_1^- + \dot{\beta}_t \beta_t \kappa_0^+}{(\alpha_t^2 \kappa_1^- + \beta_t^2 \kappa_0^+)^{1/2} (\alpha_t^2 \kappa_1^- + \beta_t^2 \kappa_0^-)^{1/2}}$$

Proof It will be convienent to use the notation

$$X_0' = \frac{\dot{\beta} x - \beta_t R_t}{\alpha_t \dot{\beta}_t - \dot{\alpha}_t \beta_t} \qquad X_1' = \frac{\alpha_t R_t - \dot{\alpha}_t x}{\alpha_t \dot{\beta}_t - \dot{\alpha}_t \beta_t}$$

as shorthand for some affine transformations of the random variable $R_t \sim \tilde{\mu}_{1,t}(\cdot \mid X_t = x)$. Call $f_u(R_t) = \langle u, R_t \rangle$ and $g_v(R_t) = -\langle v, \nabla_x V_t(x, R_t) \rangle$. Using Lemma 19,

$$\begin{aligned} u^T Dv_t(x) v &= u^T \mathrm{cov}_{\tilde{\mu}_{1,t}}[\mathsf{Id}, -\nabla_x V_t(x, \cdot)] v \\ &\le \inf_{\varepsilon \ge 0} \frac{1}{2\varepsilon} \mathrm{cov}_{\tilde{\mu}_{1,t}}[f_u, f_u] + \frac{\varepsilon}{2} \mathrm{cov}_{\tilde{\mu}_{1,t}}[g_v, g_v]. \end{aligned} \tag{9}$$

Applying the Brascamp-Lieb inequality,

$$\begin{aligned} \mathrm{cov}_{\tilde{\mu}_{1,t}}[f_u, f_u] &\le \mathbb{E}_{R_t}[\langle u, (\nabla^2_{R_t} V_t(x, R_t))^{-1} u \rangle] \\ &= \left(\alpha_t \dot{\beta}_t - \dot{\alpha}_t \beta_t \right)^2 \mathbb{E}_{R_t}[\langle u, \left((\beta_t^2 \nabla^2 V_0(X_0') + \alpha_t^2 \nabla^2 V_1(X_1') \right)^{-1} u \rangle] \\ &\le \frac{\left(\alpha_t \dot{\beta}_t - \dot{\alpha}_t \beta_t \right)^2}{\kappa_0^- \beta_t^2 + \kappa_1^- \alpha_t^2} \end{aligned}$$

For the second term, observe that

$$\begin{aligned}\nabla_{R_t} g_v(R_t) &= -\left[D_{R_t} \nabla_x V_t(x, R_t)\right] v \\ &= \frac{1}{\left(\alpha_t \dot{\beta}_t - \dot{\alpha}_t \beta_t\right)^2} \left(\alpha_t \dot{\alpha}_t \nabla^2 V_1(X_1') + \beta_t \dot{\beta}_t \nabla^2 V_0(X_0')\right) \\ &= \frac{\alpha_t \dot{\alpha}_t}{\left(\alpha_t \dot{\beta}_t - \dot{\alpha}_t \beta_t\right)^2} \left(\nabla^2 V_1(X_1') - \nabla^2 V_0(X_0')\right)\end{aligned}$$

where we used that $\alpha_t^2 + \beta_t^2 = 0$ which implies $\alpha_t \dot{\alpha}_t = -\beta_t \dot{\beta}_t$. Plugging in,

$$\begin{aligned}\mathrm{cov}_{\tilde{\mu}_{1,t}}[g_v, g_v] &\leq \mathbb{E}_{R_t}[\langle \nabla_{R_t} g_v(R_t), (\nabla^2_{R_t} V_t(x, R_t))^{-1} \nabla_{R_t} g_v(R_t)\rangle] \\ &\leq \frac{\alpha_t \dot{\alpha}_t}{\left(\alpha_t \dot{\beta}_t - \dot{\alpha}_t \beta_t\right)^2} \mathbb{E}_{R_t}\left[\langle v, \Gamma(X_0', X_1') v\rangle\right] \\ \Gamma(X_0', X_1') &\coloneqq \left(\nabla^2_{R_t} V_0(X_0') - \nabla^2_{R_t} V_1(X_1')\right) \left(\alpha_t^2 \nabla^2 V_1(X_1') + \beta_t^2 \nabla^2 V_0(X_0')\right)^{-1} \left(\nabla^2_{R_t} V_0(X_0') - \nabla^2_{R_t} V_1(X_1')\right)\end{aligned}$$

Apply Lemma 21 to the random matrix $\Gamma(X_0', X_1')$ with $P = A$, $S = B$, to see that

$$\Gamma(X_0', X_1') \preccurlyeq (A - B)(\alpha_t^2 B + \beta_t^2 A)^{-1}(A - B).$$

and altogether that

$$\mathrm{cov}_{\tilde{\mu}_{1,t}}[g_v, g_v] \leq \frac{1}{\left(\alpha_t \dot{\beta}_t - \dot{\alpha}_t \beta_t\right)^2} \left(\alpha_t^2 B + \beta_t^2 A\right)^{-1} \left(\alpha_t \dot{\alpha}_t B + \beta_t \dot{\beta}_t A\right)^2$$

where we used that A, B commute to bound $\Gamma(X_0', X_1')$ and reorganize terms. Applying Lemma 21 again with $A \preccurlyeq \kappa_0^+ I_d$ and $B \succcurlyeq \kappa_1^- I_d$, we have

$$\mathrm{cov}_{\tilde{\mu}_{1,t}}[g_v, g_v] \leq \frac{1}{\left(\alpha_t \dot{\beta}_t - \dot{\alpha}_t \beta_t\right)^2} \frac{(\alpha_t \dot{\alpha}_t \kappa_1^+ + \beta_t \dot{\beta}_t \kappa_0^-)^2}{\alpha_t^2 \kappa_1^- + \beta_t^2 \kappa_0^+}$$

and now taking the minimizer in (9) we get the desired bound

$$u^T D_x v_t(x) v \leq \frac{\dot{\alpha}_t \alpha_t \kappa_1^- + \dot{\beta}_t \beta_t \kappa_0^+}{(\alpha_t^2 \kappa_1^- + \beta_t^2 \kappa_0^+)^{1/2} (\alpha_t^2 \kappa_1^- + \beta_t^2 \kappa_0^-)^{1/2}}.$$

■

Lemma 21 *For positive definite $P \succcurlyeq Q$, $R \succcurlyeq S$, and $\gamma \geq 0$,*

$$(P - S)(P + \gamma S)^{-1}(P - S) \succcurlyeq (Q - R)(Q + \gamma R)^{-1}(Q - R)$$

Proof First we will prove the lemma with $R = S$. Without loss of generality we may further assume $R = I$ by diagonalizing, since if $R = U\Sigma U^T$ then for $\tilde{P} = \Sigma^{-1/2}U^T PU\Sigma^{-1/2}$ and $\tilde{Q} = \Sigma^{-1/2}U^T QU\Sigma^{-1/2}$ the claim is equivalent to

$$(\tilde{P} - I)(\tilde{P} + \gamma I)^{-1}(\tilde{P} - I) \succcurlyeq (\tilde{Q} - I)(\tilde{Q} + \gamma I)^{-1}(\tilde{Q} - I).$$

which follows from the fact that $\tilde{P} \succcurlyeq \tilde{Q} \iff P \succcurlyeq Q$ and that $\lambda \mapsto \frac{(\lambda-1)^2}{\lambda+\gamma}$ is increasing for $\lambda \geq 1$ and $\gamma \geq 0$.

In the case $R \neq S$, we have that

$$(P - R)(P + \gamma R)^{-1}(P - R) \succcurlyeq (Q - R)(Q + \gamma R)^{-1}(Q - R)$$

and since $-S \succcurlyeq -R$ we can reverse the inequality to prove

$$(P - S)(P + \gamma S)^{-1}(P - S) \succcurlyeq (P - R)(P + \gamma R)^{-1}(P - R).$$

$\blacksquare$