
Theoretical guarantees in KL for Diffusion Flow Matching

Marta Gentiloni Silveri

École polytechnique
Route de Saclay, 91120 Palaiseau, France
marta.gentiloni-silveri@polytechnique.edu

Giovanni Conforti

Università degli Studi di Padova
Via Trieste, 63, 35131 Padova, Italia
giovanni.conforti@math.unipd.it

Alain Durmus

École polytechnique
Route de Saclay, 91120 Palaiseau, France
alain.durmus@polytechnique.edu

Abstract

Flow Matching (FM) (also referred to as stochastic interpolants or rectified flows) stands out as a class of generative models that aims to bridge in finite time the target distribution ν^* with an auxiliary distribution μ , leveraging a fixed coupling π and a bridge which can either be deterministic or stochastic. These two ingredients define a path measure which can then be approximated by learning the drift of its Markovian projection. The main contribution of this paper is to provide relatively mild assumptions on ν^* , μ and π to obtain non-asymptotic guarantees for Diffusion Flow Matching (DFM) models using as bridge the conditional distribution associated with the Brownian motion. More precisely, we establish bounds on the Kullback-Leibler divergence between the target distribution and the one generated by such DFM models under moment conditions on the score of ν^* , μ and π , and a standard L^2 -drift-approximation error assumption.

1 Introduction

A significant task in statistics and machine learning currently revolves around generating samples from a target distribution that is only accessible via a dataset. To tackle this challenge, generative models have become prominent as effective computational tools for learning to simulate new data. Essentially, these models involve learning a *generator* capable of mapping a source distribution into new approximate samples from the target distribution.

One of the most productive approaches to generative modeling is based on deterministic and stochastic transport dynamics, that connect a target distribution with a base distribution. Typically, the target distribution represents the data set from which we want to generate new samples, while the base distribution is one that can be easily simulated or sampled. Regarding the dynamics, they correspond to SDEs Stochastic Differential Equations (SDEs) or Ordinary Differential Equations (ODEs), where the drift (for SDEs) or velocity field (for ODEs) is determined by solving a regression problem. This regression problem is usually addressed with appropriate neural networks and related training techniques [SDWVG15a, OFLR21, FJNO20, DBMP19, CLT22, DBTHD21, SE19, LYB⁺23, GCB19].

Among these methods, score-based generative models (SGMs) and in particular diffusion models based on score matching [SDWVG15b, HJA20, SE20, SE19] was an important milestone. In a nutshell, these models involve transforming an arbitrary density into a standard Gaussian model and consists in learning the drift of the corresponding reversal process. More precisely, the idea is to first consider

an Ornstein-Uhlenbeck (OU) process $(X_t^{\text{OU}})_{t \in [0, T]}$, over a time interval $[0, T]$,

$$dX_t^{\text{OU}} = -(1/2)X_t^{\text{OU}}dt + dB_t, \quad X_0^{\text{OU}} \sim \nu^*, \quad (1)$$

where $(B_t)_{t \geq 0}$ is a d -dimensional Brownian motion. Then, the reversal process $(\overleftarrow{X}_t^{\text{OU}})_{t \in [0, T]}$, which is defined from the non-homogeneous SDE [And82]:

$$d\overleftarrow{X}_t^{\text{OU}} = \{(1/2)\overleftarrow{X}_t^{\text{OU}} + \nabla \log p_{T-t}^{\text{OU}}(\overleftarrow{X}_t^{\text{OU}})\}dt + dB_t, \quad t \in [0, T] \quad , \quad \text{with } \overleftarrow{X}_0^{\text{OU}} \sim \nu^* P_T^{\text{OU}}, \quad (2)$$

allows the law of X_T^{OU} to be transported to ν^* : [And82, Equations 3.11, 3.12] show that $\overleftarrow{X}_T^{\text{OU}}$ has distribution ν^* . The initialization $\nu^* P_T^{\text{OU}}$ in (2) is the distribution of X_T^{OU} defined in (1). The drift in (2) can be decomposed as the sum of a linear function and the score associated with the density of X_{T-t}^{OU} with respect to the Lebesgue measure, denoted by p_{T-t}^{OU} . From the Tweedie identity, this score is the solution to a regression problem that can be solved efficiently by score matching [HD05, Hyv05, Vin11]. Learning this score at different times can also be formulated as a sequence of denoising problems. Once the drift of the reversal process is learned or equivalently the scores $(\nabla \log p_t^{\text{OU}})_{t \in [0, T]}$, score-based generative models consist in following the reversal dynamics over $[0, T]$ or, more commonly, a discretization of it, starting with a sample from $N(0, \text{Id})$. The final sample at time T is then approximatively distributed according to ν^* . Note that an approximation is made even if the reversal dynamics were simulated exactly, because for full accuracy, the model would need to start from a sample of $\nu^* P_T^{\text{OU}}$. However, it is well known that for sufficiently large T , $\nu^* P_T^{\text{OU}}$ is (exponentially) close to $N(0, \text{Id})$.

When exploring diffusion models, it has been realized that the generation of approximate data samples could also be achieved using an ODE instead of the reversal diffusion:

$$d\overleftarrow{\mathbf{x}}_t^{\text{OU}}/dt = (1/2)(\overleftarrow{\mathbf{x}}_t^{\text{OU}} + \nabla \log p_{T-t}^{\text{OU}}(\overleftarrow{\mathbf{x}}_t^{\text{OU}})).$$

Similarly to the drift of the reversal diffusion, the velocity fields at time $t \in [0, T]$ associated with this ODE is the sum of a linear function and the score of the density of X_{T-t}^{OU} . This observation has prompted the introduction of the Probability Flow ODE implementation of diffusion models [CCL+23a]. SGMs in their standard and probability flow ODE implementations have achieved notable success in a range of applications; see e.g., [RBL+22, RDN+22, PVG+21].

While diffusion-based methods have now become popular generative models, they can suffer from two limitations. First, there is a trade-off in selecting the time horizon T and second, they rely solely on Gaussian distributions as base distributions, in general. Therefore, there remains considerable interest in developing methods that consider a more general base distribution μ and that accomplish the transport between ν^* and μ relying on dynamics defined on a fixed finite time interval. Defining a generative process in finite time by means of a coupling and an interpolating process, [Pel22], laid the foundation for Flow Matching (FM) models [AVE22, ABVE23, LCBH+23, Liu22, LGL23], finally addressing this problem.

In its simplest form, the main strategy employed by FMs to bridge two distributions, involves a fixed coupling π between ν^* and μ and the use of a bridge, *i.e.*, a conditional distribution on the path space $\mathcal{C}([0, 1], \mathbb{R}^d)$ of a reference process $(R_t)_{t \in [0, 1]}$ given its starting point R_0 and end point R_1 . In case R_t is a deterministic function of R_0 and R_1 , we say that the bridge is deterministic and stochastic otherwise. As $(R_t)_{t \in [0, 1]}$ corresponds to the solution of a stochastic differential equation, we coin the term Diffusion Flow Matching (DFM) to distinguish the latter case from the former one and focus on it. Then, this bridge and the coupling π between ν^* and μ define an interpolated process $(X_t^{\text{I}})_{t \in [0, 1]}$, referred to as an interpolant, defined as $(X_0^{\text{I}}, X_1^{\text{I}}) \sim \pi$ and $(X_t^{\text{I}})_{t \in [0, 1]}$ given $(X_0^{\text{I}}, X_1^{\text{I}})$ has the same conditional distribution as $(R_t)_{t \in [0, 1]}$ given (R_0, R_1) . However, $(X_t^{\text{I}})_{t \in [0, 1]}$ does not correspond in general to the distribution of a diffusion or even to the one of a Markov process. This characteristic poses a challenge when dealing with potential stochastic sampling procedures: indeed, similarly to SGMs, FMs and DFMs aim to design a Markov process that approximatively transport μ to ν^* . To address this issue, most works proposing FM and DFM models [SBCD23, LWYql23] rely on mimicking the marginal flow of the interpolated process $(X_t^{\text{I}})_{t \in [0, 1]}$ through a diffusion process known as the Markovian projection. A remarkable feature of this diffusion lies in the fact that its drift is also a solution of a regression problem that can be approximatively solved using only samples from the interpolant $(X_t^{\text{I}})_{t \in [0, 1]}$. Then, an approximate samples from ν^* is obtained by following a discretization of the dynamics associated with the considered Markovian projection, starting with a sample of μ .

While there exists now an important literature on theoretical guarantees for SGMs [CDS23, CLL23, CCL⁺23b, PMM23, LLT23, Bor22a], only a few works have been considering FMs. In addition, up to our knowledge, these works on FMs only consider deterministic interpolants [AVE22, BDD23, GHJZ24]. The main objective of this paper is to fill this gap and to analyze DFMs using the d -dimensional Brownian motion as reference process, in which case the bridge is simply the d -dimensional standard Brownian bridge. We provide theoretical guarantees, upper bounding the Kullback Leibler divergence between the target distribution and the one resulting from the DFM. Our results consider the two sources of error coming from the DFM model, namely drift-estimation and time-discretization. This pursuit underscores the significance of comprehending and quantifying the factors influencing the performance of DFMs, paving the way for further advancements in generative modeling techniques.

Our contribution. In this work, we analyze a DFM model using as bridge the d -dimensional Brownian bridge and examine how it performs in two distinct scenarios: one without early-stopping and another with early-stopping. In our first main contribution Theorem 2, we establish an explicit and simple bound on the KL divergence between the data distribution and the distribution at the final time of the DFM model. We achieve our bound without early stopping, by assuming only (1) moment conditions on the target ν^* and the base μ (H1); (2) integrability conditions on the scores associated with the data distribution ν^* , the base distribution μ and the coupling π (H2); (3) a L^2 -drift-approximation error (H3) (an assumption commonly considered in previous works). Note that condition (2) implies that ν^* necessarily admits a density. We relax this condition in our second contribution. In Theorem 3, we establish an explicit bound on the KL divergence between a smoothed version of the target distribution and an early stopped version of the DFM model, assuming (1) and (3), but replacing the condition (2) by assuming (4) $\pi = \mu \otimes \nu^*$ and integrability conditions only on the score associated with μ .

To the best of our knowledge, our paper provides the first convergence analysis for diffusion-type FMs, that tackles all the sources of error, *i.e.*, the drift-approximation-error and the time-discretization error. In addition, previous studies concerning FMs and Probability Flow ODEs, with deterministic or mixed sampling procedure, either rely on at least some Lipschitz regularity of the flow velocity field or its estimator and/or do not take the time-discretization error into consideration. Also, in the context of SGMs with constant step-size, most of existing works without early-stopping are obtained assuming either the score (or its estimator) to be Lipschitz or the data distribution to satisfy some additional conditions (e.g., manifold hypothesis, bounded support, etc.); the unique exception being [CDS23]. We refer to Section 3.2 for a more in depth literature comparison.

Notation. Given a measurable space (E, \mathcal{E}) , we denote by $\mathcal{P}(E)$ the set of probability measures of E . Also, given a topological space (E, τ) , we use $\mathcal{B}(E)$ to denote the Borel σ -algebra on E . Given two random variables Y, \tilde{Y} , we write $Y \perp \tilde{Y}$ to say that Y and \tilde{Y} are independent. Denote by $(B_t)_{t \in [0,1]}$ a d -dimensional Brownian motion. We denote by Leb^d the Lebesgue measure on \mathbb{R}^d . Given two real numbers $u, v \in \mathbb{R}$, we write $u \lesssim v$ (resp. $u \gtrsim v$) to mean $u \leq Cv$ (resp. $u \geq Cv$) for a universal constant $C > 0$. Also, we denote by $\|x\|$ the Euclidean norm of $x \in \mathbb{R}^d$, by $\langle x, y \rangle$ the scalar product between $x, y \in \mathbb{R}^d$, and by x^T the transpose of x . Given a matrix $\mathbf{A} \in \mathbb{R}^{d \times s}$, we denote by $\|\mathbf{A}\|_{\text{op}}$ the operator norm of \mathbf{A} . For $f : [0, 1] \times \mathbb{R}^d \rightarrow \mathbb{R}$ regular enough, we denote by $\nabla_x f(t, x), \nabla_x^2 f(t, x)$ and $\Delta_x f(t, x)$ respectively the gradient, hessian and laplacian of f , defined for $t, x \in [0, 1] \times \mathbb{R}^d$ by $\nabla_x f(t, x) := (\partial_{x_i} f(t, x))_i, \nabla_x^2 f(t, x) := (\partial_{x_i} \partial_{x_j} f(t, x))_{i,j}, \Delta f(t, x) := \sum_{i=1}^d \partial_{x_i}^2 f(t, x)$, where ∂_{x_j} denotes the partial derivative with respect to the j -th variable. For $F : [0, 1] \times \mathbb{R}^d \rightarrow \mathbb{R}^d$ regular enough, we denote by $D_x F, \text{div}_x F$ and $\Delta_x F$ respectively, the Jacobian matrix, the divergence and the vectorial laplacian of F , defined for $t, x \in [0, 1] \times \mathbb{R}^d$ by $D_x F(t, x) = (\partial_{x_j} F_i(t, x))_{i,j}, \text{div}_x F(t, x) := \sum_{j=1}^d \partial_{x_j} F_j(t, x), \Delta_x F(t, x) = (\Delta_x F_1(t, x), \dots, \Delta_x F_d(t, x))$.

2 Diffusion Flow Matching.

Given a target distribution $\nu^* \in \mathcal{P}(\mathbb{R}^d)$ and a base distribution $\mu \in \mathcal{P}(\mathbb{R}^d)$, the idea at the core of FM models is intuitively to construct a path between these two by considering two ingredients (1) a coupling π and (2) a bridge (or an interpolant following [AVE22]) between μ and ν^* (or more precisely a bridge with foundations π). More formally, here we say that π is a coupling

between μ and ν^* if for any $A \in \mathcal{B}(\mathbb{R}^d)$, $\pi(A \times \mathbb{R}^d) = \mu(A)$ and $\pi(\mathbb{R}^d \times A) = \nu^*(A)$, and denote by $\Pi(\mu, \nu^*)$ the set of coupling between μ and ν^* . Then, based on a probability measure on $\mathbb{W} = C([0, 1], \mathbb{R}^d)$ the set of continuous functions from $[0, 1]$ to \mathbb{R}^d , we define the bridge bQ associated with \mathbb{Q} as the Markov kernel bQ on $\mathbb{R}^{2d} \times \mathbb{W}$, such that, for any $A \in \mathcal{B}(\mathbb{W})$, $\mathbb{Q}(A) = \int \mathbb{Q}_{0,1} d(x_0, x_1) \text{bQ}((x_0, x_1), A)$ (see e.g., [Kle13, Theorem 8.37] for the existence of this kernel), where for any $l = \{t_1, \dots, t_n\} \subset [0, 1]$, $t_1 < \dots < t_n$, \mathbb{Q}_l is the l -marginal distribution of \mathbb{Q} , i.e., the pushforward measure of \mathbb{Q} by $(x_t)_{t \in [0,1]} \mapsto (x_{t_1}, \dots, x_{t_n})$. From a probabilistic perspective, it implies that if $(W_t)_{t \in [0,1]} \sim \mathbb{Q}$, then bQ is a conditional distribution of $(W_t)_{t \in [0,1]} \sim \mathbb{Q}$ given (W_0, W_1) : for any bounded and measurable function on \mathbb{W} , $\mathbb{E}[f((W_t)_{t \in [0,1]})|X_0, X_1] = \int f((w_t)_{t \in [0,1]}) \text{bQ}((W_0, W_1), d(w_t)_{t \in [0,1]})$.

2.1 Definition of the interpolated process

Consider now a coupling π and a bridge bQ^β associated to $\mathbb{Q}^\beta \in \mathcal{P}(\mathbb{W})$. We suppose here that \mathbb{Q}^β is the distribution of $(Y_t)_{t \in [0,1]}$ solution of the stochastic differential equation

$$dY_t = \beta(Y_t)dt + \sqrt{2}dB_t, \quad t \in [0, 1], \quad Y_0 \sim \mu \in \mathcal{P}(\mathbb{R}^d), \quad (3)$$

where $(B_t)_{t \in \mathbb{R}_+}$ is a standard d -dimensional Brownian motion. In addition, we suppose $\beta \in C^\infty(\mathbb{R}^d, \mathbb{R}^d)$ for simplicity and that (3) admits a unique strong solution. Consider now, the interpolated measure $\mathbb{I}(\pi, \mathbb{Q}^\beta)$ corresponding to the distribution of the process defined by $(X_0^I, X_1^I) \sim \pi$ and $(X_t^I)_{t \in [0,1]} | (X_0^I, X_1^I) \sim \text{bQ}^\beta((X_0^I, X_1^I), \cdot)$. In [ABVE23], $(X_t^I)_{t \in [0,1]}$ is referred to as a stochastic interpolant. Denote by $(s, t, x, y) \mapsto p_{t|s}^Y(y|x)$ the conditional density of Y_t given Y_s with respect to the Lebesgue measure and by $(p_t^I)_{t \in [0,1]}$ the time marginal densities of $(X_t^I)_{t \in [0,1]}$ with respect to the Lebesgue measure, that is $\mathbb{P}(Y_t \in A | Y_s) = \int_A p_{t|s}(x|Y_s) dx$ and $\mathbb{P}(X_t^I \in A) = \int_A p_t^I(x) dx$, for $A \in \mathcal{B}(\mathbb{R}^d)$ and $s, t \in [0, 1]$, $s \neq t$. Then, note that, as a straightforward consequence of the definition of $(X_t^I)_{t \in [0,1]}$, it holds

$$p_t^I(x) = \int_{\mathbb{R}^{2d}} p_{t|0}^Y(x|x_0) p_{1|t}^Y(x_1|x) \tilde{\pi}(dx_0, dx_1), \quad (4)$$

where

$$\tilde{\pi}(dx_0, dx_1) = \frac{\pi(dx_0, dx_1)}{p_{1|0}^Y(x_1|x_0)}.$$

An example that we will focus on in this paper is $\mathbb{Q}^\beta = \mathbb{B}$ the distribution of the Brownian motion $(\sqrt{2}B_t)_{t \in \mathbb{R}}$ solution of (3) with $\beta \equiv 0$. Then, it is well known that bB is then the Markov kernel associated with the Brownian bridge and the resulting stochastic interpolant satisfies for any $t \in [0, 1]$

$$X_t^I \stackrel{\text{dist}}{=} (1-t)X_0^I + tX_1^I + \sqrt{2t(1-t)}Z, \quad Z \sim \mathcal{N}(0, \text{Id}), \quad (5)$$

where $\stackrel{\text{dist}}{=}$ denotes the equality in distribution.

It is well-known that for any $x_0, x_1 \in \mathbb{R}^d$, the distribution $\text{bQ}^\beta((x_0, x_1), \cdot)$ is diffusion-like under appropriate conditions. More precisely, $\text{bQ}^\beta((x_0, x_1), \cdot)$ is the distribution of $(\mathbf{Y}_t)_{t \in [0,1]}$ solution to

$$d\mathbf{Y}_t = \{\beta(\mathbf{Y}_t) + 2\nabla\phi_t^{x_1}(\mathbf{Y}_t)\}dt + \sqrt{2}d\mathbf{B}_t, \quad t \in [0, 1], \quad \mathbf{Y}_0 = x_0, \quad (6)$$

where $\phi_t^{x_1}(y) = \log p_{t|1}^Y(x_1|y)$. For instance, for $x_0, x_1 \in \mathbb{R}^d$, the bridge $\text{bB}((x_0, x_1), \cdot)$ associated to \mathbb{B} is the distribution of a Brownian bridge $(\bar{B}_t^{x_0, x_1})_{t \in [0,1]}$ solution to the SDE

$$d\bar{B}_t^{x_0, x_1} = \frac{x_1 - \bar{B}_t^{x_0, x_1}}{1-t} dt + \sqrt{2}d\mathbf{B}_t, \quad t \in [0, 1], \quad \bar{B}_0^{x_0, x_1} = x_0.$$

From (6), it turns out that $(X_t^I)_{t \in [0,1]}$ given (X_0^I, X_1^I) is therefore solution to

$$dX_t^I = \{\beta(X_t^I) + 2\nabla\phi_t^{X_1^I}(X_t^I)\}dt + \sqrt{2}d\mathbf{B}_t, \quad t \in [0, 1], \quad X_0^I \sim \mu. \quad (7)$$

¹It corresponds to $\pi \otimes \text{bQ}^\beta$ the tensor product between π and bQ^β : $\mathbb{I}(\pi, \mathbb{Q}^\beta)(A) = \int \text{bQ}^\beta((x_0, x_1), A) d\pi(x_0, x_1)$, $A \in \mathcal{B}(\mathbb{R}^d)$. In other words, π -mixture of \mathbb{Q}^β -bridges.

Note that the drift coefficient in (7) depends on X_1^I , and therefore, $(X_t^I)_{t \in [0,1]}$ is not Markov, which is a natural property if we are interested in constructing a generative process. To circumvent this issue, based on $\mathbb{I}(\pi, \mathbb{Q}^\beta)$, we aim to define a distribution $\mathbb{M}(\pi, \mathbb{Q}^\beta)$ such that it has the same one-dimensional time marginals as $\mathbb{I}(\pi, \mathbb{Q}^\beta)$ and corresponds to a diffusion, i.e., if $(X_t^I)_{t \in [0,1]} \sim \mathbb{I}(\pi, \mathbb{Q}^\beta)$ and $(X_t^M)_{t \in [0,1]} \sim \mathbb{M}(\pi, \mathbb{Q}^\beta)$, for any $t \in [0, 1]$, $X_t^I \stackrel{\text{dist}}{=} X_t^M$ and $(X_t^M)_{t \in \mathbb{N}}$ is solution of a diffusion with Markov coefficient. This can be done through the Markovian projection.

2.2 Markovian projection and Diffusion Flow Matching

Markovian projection. The idea of Markovian projection originally dates back to [Gyö86] and [Kry]. Its main idea is in essence to define a diffusion Markov process which “mimics” the time-marginal of an Itô process:

$$d\mathbf{X}_t = \mathbf{b}_t dt + \sqrt{2}dB_t, \quad t \in [0, 1], \quad \mathbf{X}_0 \sim \mu,$$

for some adapted process $(\mathbf{b}_t)_{t \in [0,1]}$ and initial distribution μ . Under appropriate conditions (see [BS13, Corollary 3.7]), this diffusion process, denoted by $(\mathbf{X}_t^M)_{t \in [0,1]}$, exists and is solution to an SDE with a (relatively) simple modification of the drift \mathbf{b}_t , namely

$$d\mathbf{X}_t^M = \tilde{\mathbf{b}}_t(\mathbf{X}_t^M)dt + \sqrt{2}dB_t, \quad t \in [0, 1], \quad \mathbf{X}_0^M \sim \mu,$$

where $\tilde{\mathbf{b}}_t(\mathbf{X}_t^M) = \mathbb{E}[\mathbf{b}_t | \mathbf{X}_t^M]$. This result can be applied to the non-Markov process $(X_t^I)_{t \in [0,1]}$ solution of (7): its Markovian projection is solution of

$$dX_t^M = \tilde{\beta}_t(X_t^M)dt + \sqrt{2}dB_t, \quad t \in [0, 1],$$

for some function $\tilde{\beta} : [0, 1] \times \mathbb{R}^d \rightarrow \mathbb{R}^d$. It turns out that we can identify $\tilde{\beta}$, relying on the family of conditional densities $(p_{t|s}^Y)_{0 \leq s \leq t \leq 1}$ and marginal densities $(p_t^I)_{0 \leq t \leq 1}$. This is the content of the following result.

Theorem 1. Consider a $\pi \in \Pi(\mu, \nu^*)$ and \mathbb{Q}^β associated with (3). Consider the drift field

$$\tilde{\beta}_t^Y(x) = \beta(x) + 2 \frac{\int \nabla_x \log p_{1|t}^Y(x_1|x) p_{t|0}^Y(x|x_0) p_{1|t}^Y(x_1|x) \tilde{\pi}(dx_0, dx_1)}{p_t^I(x)}. \quad (8)$$

Under appropriate conditions (see Appendix A.1), the Markov process $(X_t^M)_{t \in [0,1]}$ solution of

$$dX_t^M = \tilde{\beta}_t^Y(X_t^M)dt + \sqrt{2}dB_t, \quad t \in [0, 1], \quad X_0^M \sim \mu, \quad (9)$$

mimics the one-dimensional time marginals of $\mathbb{I}(\pi, \mathbb{Q}^\beta)$, i.e., for any $t \in [0, 1]$, $X_t^I \stackrel{\text{dist}}{=} X_t^M$.

This result is well-known, but, for sake of completeness, we provide the proof in Appendix A.1. The process (9) is known as the *Markovian projection* of $\mathbb{I}(\pi, \mathbb{Q}^\beta)$, and its drift (8) as *mimicking drift*. In what follows, we denote by $\mathbb{M}(\pi, \mathbb{Q}^\beta)$ the distribution of $(X_t^M)_{t \in [0,1]}$ on \mathbb{W} .

Remark 1. It can be easily shown by continuity that $X_t^M = X_t^I \Rightarrow X_1^I$ for $t \rightarrow 1$, where \Rightarrow denotes the convergence in distribution. As $\text{Law}(X_1^I) = \nu^*$, the Markovian projection therefore gives an *ideal* generative model which would consist in following the SDE (9) with initial distribution μ .

Remark 2. Note that, because of (4), the mimicking drift (8) rewrites as

$$\tilde{\beta}_t^Y(X_t^I) = \mathbb{E} \left[2 \nabla_x \log p_{1|t}^Y(X_1^I | X_t^I) + \beta(X_t^I) | X_t^I \right].$$

Diffusion Flow Matching. Eventually, as pointed out in Remark 1, the Markovian projection gives an *ideal* generative model. However, a) the mimicking drift (8) is intractable and b) the continuous-time SDE (9) can not be numerically simulated. Thus, in order to implement the proposed theoretical idea, we first need to address and overcome the aforementioned computational challenges. To circumvent a), observe that, because of Remark 2 and [Kle13, Corollary 8.17], we can approximate the mimicking drift via solving

$$\min_{\theta \in \Theta} \mathbb{E} \left[\left\| s_\theta^Y(t, X_t^I) - \tilde{\beta}_t^Y(X_t^I) \right\|^2 \right], \quad (10)$$

for a properly chosen class of neural networks $\{(t, x) \mapsto s_\theta^Y(t, x)\}_{\theta \in \Theta}$, and replace $\tilde{\beta}^Y$ in (9) with $s_{\theta^*}^Y$, where $\theta^* \in \Theta$ denotes a minimizer of (10). To deal with b), we simply make use of the Euler-Maruyama scheme, *i.e.*, for a choice of sequence of step sizes $\{h_k\}_{k=1}^N$, $N \geq 1$, and the corresponding time discretization $t_k = \sum_{i=1}^k h_i$, such that $t_0 = 0$ and $t_N = 1$, we define the continuous process $(X_t^{\theta^*})_{t \in [0,1]}$ recursively on the intervals $[t_k, t_{k+1}]$ by

$$dX_t^{\theta^*} = s_{\theta^*}^Y(t_k, X_{t_k}^{\theta^*})dt + \sqrt{2}dB_t, \quad t \in [t_k, t_{k+1}], \quad \text{with } X_0^{\theta^*} \sim \mu. \quad (11)$$

(11) is the DFM generative model we are going to analyze.

3 Main results

In this section, we provide convergence guarantees in Kullback-Leibler divergence for the Diffusion Flow Matching model (11), under mild assumptions on the μ, ν^*, π and s_{θ^*} , either within a non-early-stopping regime or within an early-stopping regime.

From now on, we consider the case $\beta \equiv 0$, *i.e.*, $\mathbb{Q}^\beta = \mathbb{B}$. We show in Appendix A.2, Remark 9, that, under our set of assumptions, the conditions of Theorem 1 hold for this setup. Moreover, in this case, for any $s, t \in [0, 1]$, $s < t$ and $x, y \in \mathbb{R}^d$, the conditional density $p_{t|s}^Y(y|x) \equiv p_{t-s}(y|x)$ where $(t, x, y) \mapsto p_t(y|x)$ is the heat kernel:

$$p_t(y|x) = \frac{1}{(4\pi t)^{d/2}} \exp\left(-\frac{\|y-x\|^2}{4t}\right), \quad t \in (0, 1]. \quad (12)$$

In the following, we set $\tilde{\beta}^Y \equiv \tilde{\beta}$ and $s_{\theta^*}^Y \equiv s_{\theta^*}$.

3.1 Convergence Bounds.

We assume moment conditions on the probability measures μ and ν^* , and mild integrability conditions on the probability distributions μ, ν^* and on the coupling π .

For $p \geq 1$, we denote for $\zeta \in \mathcal{P}(\mathbb{R}^d)$,

$$\mathbf{m}_p[\zeta] = \int \|x\|^p d\zeta(x).$$

H1. The probability distributions μ, ν^* satisfy $\mathbf{m}_s[\mu] + \mathbf{m}_s[\nu^*] < +\infty$.

H2. The probability distributions ν^*, μ and π are absolutely continuous with respect to the Lebesgue measure on \mathbb{R}^d and \mathbb{R}^{2d} respectively, and satisfy

(i) The functions $\log d\mu/d\text{Leb}^d$ and $\log d\nu^*/d\text{Leb}^d$ are continuously differentiable and satisfy $\|\nabla \log \nu^*\|_{L^s(\nu^*)}^s + \|\nabla \log \mu\|_{L^s(\mu)}^s < +\infty$ where for $\zeta \in \{\nu^*, \mu\}$,

$$\|\nabla \log \zeta\|_{L^s(\zeta)}^s = \int \left\| \nabla \log \left(\frac{d\zeta}{d\text{Leb}^d} \right) (x_0) \right\|^s d\zeta(x_0).$$

(ii) The function $\log d\pi/d\text{Leb}^{2d}$ is continuously differentiable and satisfies $\|\nabla \log \tilde{\pi}\|_{L^s(\pi)}^s < +\infty$ where

$$\|\nabla \log \tilde{\pi}\|_{L^s(\pi)}^s = \int \|\nabla \log(\tilde{\pi})(x_0, x_1)\|^s d\pi(x_0, x_1), \quad \tilde{\pi}(x_0, x_1) = \frac{1}{p_1(x_1|x_0)} \frac{d\pi}{d\text{Leb}^{2d}}(x_0, x_1), \quad (13)$$

and p_1 is defined in (12).

Remark 3. Under **H1**, note that $\|\nabla \log \tilde{\pi}\|_{L^s(\pi)}^s < +\infty$ is equivalent by (12) and $\pi \in \Pi(\mu, \nu^*)$ to

$$\int \|\nabla \log(d\pi/d\text{Leb}^{2d})(x_0, x_1)\|^s d\pi(x_0, x_1) < +\infty.$$

Remark 4. We can relax the condition that $\log d\mu/d\text{Leb}^d$, $\log d\nu^*/d\text{Leb}^d$ and $\log d\pi/d\text{Leb}^{2d}$ are continuously differentiable assuming that $\sqrt[8]{d\mu/d\text{Leb}^d}$, $\sqrt[8]{d\nu^*/d\text{Leb}^d}$ and $\sqrt[8]{d\pi/d\text{Leb}^{2d}}$ belongs to some Sobolev space, but, for ease of presentation, we prefer not to delve into these technical details.

Moreover, we assume to have estimated the mimicking drift with an ε^2 -precision, for some $\varepsilon^2 > 0$ sufficiently small.

H3. There exist $\theta^* \in \Theta$ and $\varepsilon^2 > 0$ such that

$$\sum_{k=0}^{N-1} h_{k+1} \mathbb{E} \left[\left\| s_{\theta^*}(t_k, X_{t_k}^M) - \tilde{\beta}_{t_k}(X_{t_k}^M) \right\|^2 \right] \leq \varepsilon^2.$$

Remark 5. To be coherent with the previous section, observe that, as a consequence of Theorem 1, for any $k = 0, \dots, N$, it holds

$$\mathbb{E} \left[\left\| s_{\theta^*}(t_k, X_{t_k}^M) - \tilde{\beta}_{t_k}(X_{t_k}^M) \right\|^2 \right] = \mathbb{E} \left[\left\| s_{\theta^*}(t_k, X_{t_k}^I) - \tilde{\beta}_{t_k}(X_{t_k}^I) \right\|^2 \right].$$

Under such assumptions, we derive an upper bound on the KL divergence between the data distribution ν^* and the output of the DFM (11):

Theorem 2. Consider a uniform partition of $[0, 1]$ with a constant stepsize $h_k \equiv h$, $h = 1/N_h > 0$, for $N_h \in \mathbb{N}^*$ and consider the corresponding process $(X_t^{\theta^*})_{t \in [0,1]}$ defined in (11). Assume **H1** to **3**. Denoting by $\nu_1^{\theta^*}$ the distribution of $X_1^{\theta^*}$, we have that

$$\text{KL}(\nu^* | \nu_1^{\theta^*}) \lesssim \varepsilon^2 + h(h^{1/8} + 1) \left(d^4 + \mathbf{m}_8[\mu] + \mathbf{m}_8[\nu^*] + \|\nabla \log \tilde{\pi}\|_{L^8(\pi)}^8 + \|\nabla \log \mu\|_{L^8(\mu)}^8 + \|\nabla \log \nu^*\|_{L^8(\nu^*)}^8 \right). \quad (14)$$

Remark 6. Under almost the same conditions as Theorem 2, i.e., **H1** **H2**, **H3**, replacing $\tilde{\pi}$ in (13) by

$$\tilde{\pi}_T(x_0, x_1) = \frac{1}{p_T(x_1|x_0)} \frac{d\pi}{d\text{Leb}^{2d}}(x_0, x_1), \quad (15)$$

our proofs apply also to DFM using a time horizon $T > 0$ and the Brownian bridge on $[0, T]$. In particular, we would have obtained similar bounds as the ones derived in Theorem 2 but with a factor $\max(1, T^8)$ in front of the second addend.

Remark 7. Choosing, in Theorem 2,

$$N_h = \frac{d^4 + \mathbf{m}_8[\mu] + \mathbf{m}_8[\nu^*] + \|\nabla \log \tilde{\pi}\|_{L^8(\pi)}^8 + \|\nabla \log \mu\|_{L^8(\mu)}^8 + \|\nabla \log \nu^*\|_{L^8(\nu^*)}^8}{\varepsilon^2}$$

makes the approximation error of order $\mathcal{O}(\varepsilon^2)$ and the complexity of order $\mathcal{O}(\varepsilon^{-2})$.

Using an early-stopping procedure, we obtain in the case π is the independent coupling:

Theorem 3. Fix $0 < \delta < 1/2$. Consider a uniform partition of $[0, 1]$ with a constant stepsize $h_k \equiv h$, $h = 1/N_h > 0$, for $N_h \in \mathbb{N}^*$ and consider the corresponding process $(X_t^{\theta^*})_{t \in [0,1]}$ defined in (11). Assume **H1**, **H3** and $\pi = \mu \otimes \nu^*$ to be the independent coupling. Suppose in addition that μ is absolutely continuously with respect to the Lebesgue measure, $\log d\mu/d\text{Leb}^d$ is continuously differentiable and satisfies $\|\nabla \log \mu\|_{L^8(\mu)} < +\infty$.

Then, denoting by $\nu_{1-\delta}^*$ and $\nu_{1-\delta}^{\theta^*}$ the distribution of $X_{1-\delta}^M$ and $X_{1-\delta}^{\theta^*}$ respectively, we have that

$$\text{KL}(\nu_{1-\delta}^* | \nu_{1-\delta}^{\theta^*}) \lesssim \varepsilon^2 + h(h^{1/8} + 1) \left(\frac{d^4}{\delta^4} + \mathbf{m}_8[\mu] + \mathbf{m}_8[\nu^*] \frac{1}{\delta^8} + \|\nabla \log \mu\|_{L^8(\mu)}^8 \right).$$

Corollary 1. Fix $\delta = \mathcal{O}(\sqrt{\varepsilon})$. Consider a uniform partition of $[0, 1]$ with a constant stepsize $h = \mathcal{O}(\min(\varepsilon^4/d^4, \varepsilon^6))$, and consider the corresponding process $(X_t^{\theta^*})_{t \in [0,1]}$ defined in (11). Assume **H1**, **H3** and $\pi = \mu \otimes \nu^*$ to be the independent coupling. Suppose in addition that μ

is absolutely continuously with respect to the Lebesgue measure, $\log d\mu/d\text{Leb}^d$ is continuously differentiable and satisfies $\|\nabla \log \mu\|_{L^s(\mu)}^8 < +\infty$.

Then, denoting by $\nu_{1-\delta}^*$ and $\nu_{1-\delta}^{\theta^*}$ the distribution of $X_{1-\delta}^M$ and $X_{1-\delta}^{\theta^*}$ respectively, we have that

$$\mathcal{W}_{2,FM}^2(\nu_{1-\delta}^*|\nu_{1-\delta}^{\theta^*}) \lesssim \varepsilon^2,$$

where $\mathcal{W}_{2,FM}$ denotes the Fortet- Mourier distance of order 2, i.e.

$$\mathcal{W}_{2,FM}^2(\mu, \nu) = \inf_{\pi \in \Pi(\mu, \nu)} \int \min\{\|x - y\|^2, 1\} d\pi(x, y).$$

3.2 Related works and comparison with existing literature.

FMs stand at the forefront of innovation in generative modeling, offering a practical solution to the longstanding challenge of bridging two arbitrary distributions within a finite time interval. Their consequent immense potential has prompted substantial research efforts aimed at providing a theoretical explanation for their effectiveness and has put SGMs and Probability Flow ODEs all in perspective. In this section we report and discuss previous researches on FMs, SGMs and Probability Flow ODEs with the purpose of highlighting the links and differences with our work and contextualizing our contribution.

Non-early-stopping setting. In the context of FMs, [AVE22] and [BDD23] seek convergence guarantees in 2-Wasserstein distance. Both consider more general designs for the stochastic interpolant and a deterministic sampling procedure, rather than a stochastic one. However, both works rely on some regularity condition on the approximated flow velocity field, i.e., that it is Lipschitz. Namely, [AVE22] works under a K -Lipschitz (uniform in time and space) assumption on the estimator of the exact flow velocity field. How such assumption pertains to the flow matching framework for generative modeling is not articulated and remains unclear. On the other hand, [BDD23] assumes the estimator of the exact flow velocity field to be L_t -Lipschitz for any $t \in [0, 1]$ and discuss in [BDD23, Theorem 2] how such assumption relate to the setting : under the additional assumption [BDD23, Assumption 4], the true flow velocity field is proven to be L_t -Lipschitz in space for any $t \in [0, 1]$. Therefore, [BDD23] enhances the findings of [AVE22]. However, [BDD23, Assumption 4] is not an usual conditions considered in papers about convergence guarantees for SGM and it is unclear which type of distributions satisfy [BDD23, Assumption 4]. Moreover, both works [BDD23, AVE22] do not take into account the discretization error in their analysis.

In the context of Probability Flow ODEs, [LWCC24] and [GZ24] investigate the performance of such models in Total Variation distance and 2-Wasserstein distance. In contrast to [BDD23] and [AVE22], [LWCC24] and [GZ24] examine the error coming from the (prerequisite when implementing an algorithm) introduction of a time-discretization scheme. However, once again, the provided bounds work under smoothness assumptions either on the score or on its estimator. More precisely, the result reported in [LWCC24] depends on a small L^2 -Jacobian-estimation error assumption, besides a classical small L^2 -score-estimation error assumption. As for [GZ24], they assume the score to be Lipschitz in time and the data to be smooth and log-concave. In contrast, our result do not make such assumptions. Finally, to the best of our knowledge, the recent work [CDS23] represents the state of art in the context of SGMs without early-stopping procedure: [CDS23, Theorem 2.1] provides a sharp bound in KL divergence between the data distribution and the law of the SGM both in the overdamped and kinetic setting under the sole assumptions of an L^2 -score-approximation error and that the data distribution has finite Fisher information with respect to the standard Gaussian distribution. All previous results are obtained assuming either some Lipschitz condition on the score and/or its estimator ([CCL⁺23b], [CLL23]) or a manifold hypothesis on the data distribution ([Bor22b]). However, we underline that the FM framework enables to consider a significantly wider range of interpolating paths compared to SGMs and to avoid the trade-off concerning the time horizon T which is inevitable when dealing with SGMs.

Early-stopping procedure. The recent work [GHJZ24] establishes convergence guarantees in 2-Wasserstein distance for FMs based on a deterministic sampling procedure. However, the results of [GHJZ24] requires to interpolate with a Gaussian distribution and applies only to data distributions which either have a bounded support, are strongly log-concave, or are the convolution between a Gaussian and an other probability distribution supported on an Euclidean Ball. In contrast, for our bound to hold true we only need the data distribution ν^* and its score to have finite eight-order

moment. Furthermore, even if [GHJZ24] goes into depth when dealing with the statistical analysis of the estimator and the L^2 -estimation error, the entire investigation therein pursued depends on the choice of ReLUNetworks with Lipschitz regularity control to approximate the velocity field. They motivate such choice by proving (see [GHJZ24, Theorem 5.1]) Lipschitz properties in time and space of the true velocity field under the aforementioned assumptions on the data. On the contrary, we do not assume any regularity on the estimator of the mimicking drift. In the context of Probability Flow ODEs, [CCL+23a] provides bound in Total Variation distance, but assuming both the score and its estimator to be Lipschitz in space. So, also in the early-stopping regime, our bound improves previously obtained one.

To conclude, in the context of SGMs, [CLL23, CDS23, BDBDD23] are able to cover any data distribution with bounded second moment at the cost of using exponentially decreasing step-sizes. However, [CDS23, Corollary 2.4] and [BDBDD23] improves upon [CLL23, Theorem 2.2]: the term that takes track of the time-discretization error is linearly dependent on the dimension d in the former works, whereas quadratically dependent on d in the latter.

3.3 The proposed methodology.

In what follows, we provide a sketch of the proofs of Theorem 2 and Theorem 3 in order to outline and delineate our methodology.

The starting point of our proof of Theorem 2 is the following (by now) standard [CCL+23b, CLL23, CDS23] decomposition of the KL divergence which is derived from Girsanov theorem:

$$\begin{aligned} \text{KL}(\nu^* | \nu_1^{\theta^*}) &\leq \text{KL}(\mathbb{M}(\pi, \mathbb{B}) | \text{Law}(X_{[0,1]}^{\theta^*})) \lesssim \sum_{k=0}^{N-1} \int_{t_k}^{t_{k+1}} \mathbb{E} \left[\left\| s_{\theta^*}(t_k, X_{t_k}^M) - \tilde{\beta}_t(X_t^M) \right\|^2 \right] dt \\ &\lesssim \varepsilon^2 + \sum_{k=0}^{N-1} \int_{t_k}^{t_{k+1}} \mathbb{E} \left[\left\| \tilde{\beta}_{t_k}(X_{t_k}^M) - \tilde{\beta}_t(X_t^M) \right\|^2 \right] dt, \end{aligned}$$

where, for the first inequality, we used the data processing inequality [Nut21, Lemma 1.6] and the last inequality follows from the triangle inequality and the assumption **H3**. In order to conclude, we should bound the L^2 norm of the adjoint process in the Pontryagin system associated with the Markovian projection of the interpolant. We do so by introducing a novel quantity in the generative model literature (see [Kre97]), namely the so-called *reciprocal characteristic* of the mimicking drift, *i.e.*,

$$(\partial_t + \mathcal{L}_t^M) \tilde{\beta}_t,$$

where \mathcal{L}^M denotes the generator of $(X_t^M)_{t \in [0,1]}$. This quantity may be viewed as some sort of mean acceleration field and guides the time evolution of the mimicking drift, as

$$d\tilde{\beta}_t(X_t^M) = (\partial_t + \mathcal{L}_t^M) \tilde{\beta}_t(X_t^M) dt + \sqrt{2} D_x \tilde{\beta}_t(X_t^M) dB_t, \quad t \in [0, 1].$$

The main efforts in our proof are directed towards bounding the L^2 norm of the reciprocal characteristic whose representation in terms of either conditional moments or higher-order logarithmic derivatives of conditional densities is quite intricate, see (39). Trying to bound each of these terms separately requires strong assumptions on the initial distributions and couplings leading to sub-optimal results. However, using integration by parts both in time and space and a double change of measure argument, and profiting from symmetry properties of the heat kernel, we managed to bound these terms under assumptions comparable to the minimal ones required in the analysis of SGMs. Note that the analysis of the reciprocal characteristic is not required for SGMs (it is always 0) and that controlling it also requires new tricks and ideas, since its representation contains up to three logarithmic derivatives of conditional distributions, whereas the analysis of SGMs requires at most two such derivatives to be analyzed.

Regarding the proof of Theorem 3, we consider the interpolated process $(X_t^I)_{t \in [0, \delta]}$ restricted to $[0, 1 - \delta]$. Denoting by $\pi_{1-\delta}$, the coupling between μ and $\nu_{1-\delta}^*$ corresponding to the distribution of the couple $(X_0^I, X_{1-\delta}^I)$. By the property of the Brownian bridge, $(X_t^I)_{t \in [0, 1-\delta]}$ is a stochastic interpolant resulting from $\pi_{1-\delta}$ and the Brownian bridge on $[0, 1 - \delta]$. Therefore based on Remark 6,

we only have to show $\nu_{1-\delta}^*$ and $\pi_{1-\delta}$ satisfy **H1** and **H2**, replacing $\tilde{\pi}$ in **H2** by $\tilde{\pi}_{1-\delta}$ defined in (15). More precisely, we show that they hold and that

$$\begin{aligned} \|\nabla \log \tilde{\pi}_{1-\delta}\|_{L^s(\pi_{1-\delta})}^s &\leq \|\nabla \log \mu\|_{L^s(\mu)}^s + \mathbf{m}_s[\mu] \frac{1}{(1-\delta)^8} + \mathbf{m}_s[\nu^*] \frac{1}{\delta^8} + d^4 \frac{1}{\delta^4(1-\delta)^4} \\ \|\nabla \log \nu_{1-\delta}^*\|_{L^s(\nu_{1-\delta}^*)}^s &\leq \mathbf{m}_s[\mu] \frac{1}{(1-\delta)^8} + \mathbf{m}_s[\nu^*] \frac{1}{\delta^8} + d^4 \frac{1}{\delta^4(1-\delta)^4}. \end{aligned}$$

4 Conclusion

In this work, we have investigated a Diffusion Flow Matching model built around the Markovian projection of the d -dimensional Brownian bridge between the data distribution ν^* and the base distribution μ . In particular, we have derived convergence guarantees in Kullback-Leibler divergence, which take account of all the sources of error - time-discretization error and drift-estimation error - that arise when implementing the model, and which hold under mild moments conditions on μ, ν^* , the scores of μ, ν^* and the score of the coupling π between μ and ν^* . However, there are several questions remaining open. First, it would be worthy to understand if we could lower the order of integrability of the score associated with μ, ν and π . Second, it would be interesting to complement our analysis by a statistical analysis of DFM (11), similarly to what have been achieved in [GHJZ24] for a particular deterministic FM model. Finally, it would be valuable to obtain better dimension dependence with respect to the space dimension d when applying early-stopping procedure.

Acknowledgments and Disclosure of Funding

The work of Marta Gentiloni-Silveri has been supported by the Paris Ile-de-France Région in the framework of DIM AI4IDF. The work by Alain Durmus is partially funded by the European Union (ERC-2022-SYG-OCEAN-101071601). Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Council Executive Agency. Neither the European Union nor the granting authority can be held responsible for them.

References

- [ABVE23] Michael S Albergo, Nicholas M Boffi, and Eric Vanden-Eijnden. Stochastic interpolants: A unifying framework for flows and diffusions. *arXiv preprint arXiv:2303.08797*, 2023.
- [And82] Brian DO Anderson. Reverse-time diffusion equation models. *Stochastic Processes and their Applications*, 12(3):313–326, 1982.
- [AVE22] Michael S Albergo and Eric Vanden-Eijnden. Building normalizing flows with stochastic interpolants. *arXiv preprint arXiv:2209.15571*, 2022.
- [BB17] Paolo Baldi and Paolo Baldi. *Stochastic calculus*. Springer, 2017.
- [BDBDD23] Joe Benton, Valentin De Bortoli, Arnaud Doucet, and George Deligiannidis. Linear convergence bounds for diffusion models via stochastic localization. *arXiv preprint arXiv:2308.03686*, 2023.
- [BDD23] Joe Benton, George Deligiannidis, and Arnaud Doucet. Error bounds for flow matching methods. *arXiv preprint arXiv:2305.16860*, 2023.
- [BKRS15] Vladimir I. Bogachev, Nicolai V. Krylov, Michael Röckner, and Stanislav V. Shaposhnikov. Fokker-planck-kolmogorov equations. 2015.
- [Bor22a] Valentin De Bortoli. Convergence of denoising diffusion models under the manifold hypothesis. *Transactions on Machine Learning Research*, 2022. Expert Certification.
- [Bor22b] Valentin De Bortoli. Convergence of denoising diffusion models under the manifold hypothesis. *Transactions on Machine Learning Research*, 2022. Expert Certification.

- [BS13] Gerard Brunick and Steven Shreve. Mimicking an Itô process by a solution of a stochastic differential equation. *The Annals of Applied Probability*, 23(4):1584 – 1628, 2013.
- [CCL⁺23a] Sitan Chen, Sinho Chewi, Holden Lee, Yuanzhi Li, Jianfeng Lu, and Adil Salim. The probability flow ODE is provably fast. In *Thirty-seventh Conference on Neural Information Processing Systems*, 2023.
- [CCL⁺23b] Sitan Chen, Sinho Chewi, Jerry Li, Yuanzhi Li, Adil Salim, and Anru R. Zhang. Sampling is as easy as learning the score: theory for diffusion models with minimal data assumptions. *International Conference on Learning Representations*, 2023.
- [CDS23] Giovanni Conforti, Alain Durmus, and Marta Gentiloni Silveri. Score diffusion models without early stopping: finite fisher information is all you need. *arXiv preprint arXiv:2308.12240*, 2023.
- [CLL23] Hongrui Chen, Holden Lee, and Jianfeng Lu. Improved analysis of score-based generative modeling: User-friendly bounds under minimal smoothness assumptions. In *International Conference on Machine Learning*, pages 4735–4763. PMLR, 2023.
- [CLT22] Tianrong Chen, Guan-Hong Liu, and Evangelos Theodorou. Likelihood training of schrödinger bridge using forward-backward SDEs theory. In *International Conference on Learning Representations*, 2022.
- [DBMP19] Conor Durkan, Artur Bekasov, Iain Murray, and George Papamakarios. Neural spline flows. *Advances in neural information processing systems*, 32, 2019.
- [DBTHD21] Valentin De Bortoli, James Thornton, Jeremy Heng, and Arnaud Doucet. Diffusion schrödinger bridge with applications to score-based generative modeling. *Advances in Neural Information Processing Systems*, 34:17695–17709, 2021.
- [FJNO20] Chris Finlay, Jörn-Henrik Jacobsen, Levon Nurbekyan, and Adam Oberman. How to train your neural ode: the world of jacobian and kinetic regularization. In *International conference on machine learning*, pages 3154–3164. PMLR, 2020.
- [G CBD19] Will Grathwohl, Ricky TQ Chen, Jesse Bettencourt, and David Duvenaud. Scalable reversible generative models with free-form continuous dynamics. In *International Conference on Learning Representations*, page 7, 2019.
- [GHJZ24] Yuan Gao, Jian Huang, Yuling Jiao, and Shurong Zheng. Convergence of continuous normalizing flows for learning probability distributions. *arXiv preprint arXiv:2404.00551*, 2024.
- [Gyö86] István Gyöngy. Mimicking the one-dimensional marginal distributions of processes having an itô differential. *Probability theory and related fields*, 71(4):501–516, 1986.
- [GZ24] Xuefeng Gao and Lingjiong Zhu. Convergence analysis for general probability flow odes of diffusion models in wasserstein distances. *arXiv preprint arXiv:2401.17958*, 2024.
- [HD05] Aapo Hyvärinen and Peter Dayan. Estimation of non-normalized statistical models by score matching. *Journal of Machine Learning Research*, 6(4), 2005.
- [HJA20] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models, 2020.
- [Hyv05] Aapo Hyvärinen. Estimation of non-normalized statistical models by score matching. *Journal of Machine Learning Research*, 6(24):695–709, 2005.
- [Kle13] Achim Klenke. *Probability theory: a comprehensive course*. Springer Science & Business Media, 2013.
- [Kre97] Arthur J Krener. Reciprocal diffusions in flat space. *Probability theory and related fields*, 107(2):243–281, 1997.

- [Kry] N. V. Krylov. On the relation between differential operators of second order and the solutions of stochastic differential equations. *Steklov Seminar*.
- [LCBH⁺23] 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*.
- [LGL23] Xingchao Liu, Chengyue Gong, and Qiang Liu. Flow straight and fast: Learning to generate and transfer data with rectified flow. *International Conference on Learning Representations (ICLR), 2023*.
- [Liu22] Qiang Liu. Rectified flow: A marginal preserving approach to optimal transport. *arXiv preprint arXiv:2209.14577, 2022*.
- [LLT23] Holden Lee, Jianfeng Lu, and Yixin Tan. Convergence of score-based generative modeling for general data distributions. In *International Conference on Algorithmic Learning Theory*, pages 946–985. PMLR, 2023.
- [LWCC24] Gen Li, Yuting Wei, Yuxin Chen, and Yuejie Chi. Towards non-asymptotic convergence for diffusion-based generative models. In *The Twelfth International Conference on Learning Representations, 2024*.
- [LWYql23] Xingchao Liu, Lemeng Wu, Mao Ye, and qiang liu. Learning diffusion bridges on constrained domains. In *The Eleventh International Conference on Learning Representations, 2023*.
- [LYB⁺23] Sungbin Lim, EUN BI YOON, Taehyun Byun, Taewon Kang, Seungwoo Kim, Kyungjae Lee, and Sungjoon Choi. Score-based generative modeling through stochastic evolution equations in hilbert spaces. In A. Oh, T. Naumann, A. Globerson, K. Saenko, M. Hardt, and S. Levine, editors, *Advances in Neural Information Processing Systems*, volume 36, pages 37799–37812. Curran Associates, Inc., 2023.
- [Nut21] Marcel Nutz. Introduction to entropic optimal transport. *Lecture notes, Columbia University, 2021*.
- [OFLR21] 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.
- [Pel22] Stefano Peluchetti. Non-denoising forward-time diffusions, 2022.
- [PMM23] Francesco Pedrotti, Jan Maas, and Marco Mondelli. Improved convergence of score-based diffusion models via prediction-correction. *arXiv preprint arXiv:2305.14164, 2023*.
- [PVG⁺21] Vadim Popov, Ivan Vovk, Vladimir Gogoryan, Tasnima Sadekova, and Mikhail Kudinov. Grad-tts: A diffusion probabilistic model for text-to-speech. In *International Conference on Machine Learning*, pages 8599–8608. PMLR, 2021.
- [RBL⁺22] 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.
- [RDN⁺22] Aditya Ramesh, Prafulla Dhariwal, Alex Nichol, Casey Chu, and Mark Chen. Hierarchical text-conditional image generation with clip latents. *arXiv preprint arXiv:2204.06125, 1(2):3, 2022*.
- [RY13] Daniel Revuz and Marc Yor. *Continuous martingales and Brownian motion*, volume 293. Springer Science & Business Media, 2013.
- [SBCD23] Yuyang Shi, Valentin De Bortoli, Andrew Campbell, and Arnaud Doucet. Diffusion schrödinger bridge matching. In *Thirty-seventh Conference on Neural Information Processing Systems, 2023*.

- [SDWVG15a] Jascha Sohl-Dickstein, Eric Weiss, Niru Maheswaranathan, and Surya Ganguli. Deep unsupervised learning using nonequilibrium thermodynamics. In Francis Bach and David Blei, editors, *Proceedings of the 32nd International Conference on Machine Learning*, volume 37 of *Proceedings of Machine Learning Research*, pages 2256–2265, Lille, France, 07–09 Jul 2015. PMLR.
- [SDWVG15b] Jascha Sohl-Dickstein, Eric A. Weiss, Niru Maheswaranathan, and Surya Ganguli. Deep unsupervised learning using nonequilibrium thermodynamics, 2015.
- [SE19] Yang Song and Stefano Ermon. Generative modeling by estimating gradients of the data distribution. *Advances in neural information processing systems*, 32, 2019.
- [SE20] Yang Song and Stefano Ermon. Generative modeling by estimating gradients of the data distribution, 2020.
- [VEH14] Tim Van Erven and Peter Harremo. Rényi divergence and kullback-leibler divergence. *IEEE Transactions on Information Theory*, 60(7):3797–3820, 2014.
- [Vin11] Pascal Vincent. A connection between score matching and denoising autoencoders. *Neural Computation*, 23:1661–1674, 2011.

A Postponed proofs

A.1 The Markovian Projection

First, we introduce the set of assumptions under which Theorem 1 holds true.

H4. For $t > s$, $(s, t, x, y) \mapsto p_{t|s}^Y(y|x)$ is continuously differentiable in the t and s variables and twice continuously differentiable in the x and y variables. Furthermore, $\partial_t p_{t|s}^Y(y|x)$, $\partial_s p_{t|s}^Y(y|x)$, $\nabla_x p_{t|s}^Y(y|x)$, $\nabla_y p_{t|s}^Y(y|x)$, $\nabla_x^2 p_{t|s}^Y(y|x)$, $\nabla_y^2 p_{t|s}^Y(y|x)$ are bounded.

H4 ensures that $(s, t, x, y) \mapsto p_{t|s}^Y(y|x)$ is enough regular to allow a series of algebraic manipulations.

H5. $\tilde{\beta}^Y$ is a locally bounded Borel vector field on $\mathbb{R}^d \times (0, 1)$ such that, for at least one probability solution μ_t to the Fokker-Planck equation

$$\partial_t \mu_t + \operatorname{div}(\tilde{\beta}_t^Y \mu_t) - \Delta \mu_t = 0, \quad t \in (0, 1), \quad \mu_0 = \mu, \quad (16)$$

it holds $\int \|\tilde{\beta}_t^Y(x)\| \mu_t(dx) dt < +\infty$.

H5 provides uniqueness of the solution to the Fokker Planck equation with drift field $\tilde{\beta}^Y$, see [BKRS15, Theorem 9.4.3].

We are now ready to rigorously state and prove Theorem 1:

Theorem 4. Consider a $\pi \in \Pi(\mu, \nu^*)$, \mathbb{Q}^β associated with (3) and the drift field defined in (8). Under **H4** and **5**, the Markov process $(X_t^M)_{t \in [0,1]}$ solution of (9) is such that, for any $t \in [0, 1)$, $X_t^I \stackrel{\text{dist}}{=} X_t^M$.

Proof of Theorem 4: We start by reminding that $(s, t, x, y) \mapsto p_{t|s}^Y(y|x)$ satisfies for any $x, y \in \mathbb{R}^d$ and $s, t \in [0, 1]$ with $s < t$ both the Fokker-Planck equation

$$\partial_t p_{t|s}^Y(y|x) + \operatorname{div}_y(p_{t|s}^Y(y|x)\beta(x)) - \Delta_y p_{t|s}^Y(y|x) = 0,$$

and the Kolmogorov backward equation

$$\partial_s p_{t|s}^Y(y|x) + \langle \beta(x), \nabla_x p_{t|s}^Y(y|x) \rangle + \Delta_x p_{t|s}^Y(y|x) = 0.$$

If we exploit these well-known results, (4) and **H4**, we get that for $t \in (0, 1)$

$$\begin{aligned} \partial_t p_t^I(x) &= \partial_t \left(\int_{\mathbb{R}^{2d}} p_{t|0}^Y(x|x_0) p_{1|t}^Y(x_1|x) \tilde{\pi}(dx_0, dx_1) \right) \\ &= \int_{\mathbb{R}^{2d}} \left(\Delta_x p_{t|0}^Y(x|x_0) p_{1|t}^Y(x_1|x) - p_{t|0}^Y(x|x_0) \Delta_x p_{1|t}^Y(x_1|x) \right) \tilde{\pi}(dx_0, dx_1) \\ &\quad - \int_{\mathbb{R}^{2d}} \operatorname{div}_x(p_{t|0}^Y(x|x_0)\beta(x)) p_{1|t}^Y(x_1|x) \tilde{\pi}(dx_0, dx_1) \\ &\quad - \int_{\mathbb{R}^{2d}} \langle \beta(x), \nabla_x p_{1|t}^Y(x_1|x) \rangle p_{t|0}^Y(x|x_0) \tilde{\pi}(dx_0, dx_1) \\ &= \int_{\mathbb{R}^{2d}} \left(\Delta_x p_{t|0}^Y(x|x_0) p_{1|t}^Y(x_1|x) + \nabla_x p_{t|0}^Y(x|x_0) \nabla_x p_{1|t}^Y(x_1|x) \right) \tilde{\pi}(dx_0, dx_1) \\ &\quad - \int_{\mathbb{R}^{2d}} \left(\nabla_x p_{t|0}^Y(x|x_0) \nabla_x p_{1|t}^Y(x_1|x) + p_{t|0}^Y(x|x_0) \Delta_x p_{1|t}^Y(x_1|x) \right) \tilde{\pi}(dx_0, dx_1) \\ &\quad - \langle \beta(x), \nabla_x p_t^I(x) \rangle - \operatorname{div}_x \beta(x) p_t^I(x) \\ &= \operatorname{div}_x \left(\int_{\mathbb{R}^{2d}} \{ \nabla_x p_{t|0}^Y(x|x_0) p_{1|t}^Y(x_1|x) - p_{t|0}^Y(x|x_0) \nabla_x p_{1|t}^Y(x_1|x) \} \tilde{\pi}(dx_0, dx_1) \right) \\ &\quad - \operatorname{div}_x(\beta(x) p_t^I(x)) \\ &= \operatorname{div}_x \left(-\beta(x) p_t^I(x) \right. \\ &\quad \left. + \int_{\mathbb{R}^{2d}} \{ \nabla_x \log p_{t|0}^Y(x|x_0) - \nabla_x \log p_{1|t}^Y(x_1|x) \} p_{t|0}^Y(x|x_0) p_{1|t}^Y(x_1|x) \tilde{\pi}(dx_0, dx_1) \right). \end{aligned}$$

Therefore, if we set

$$v_t^Y(x) = \frac{\int_{\mathbb{R}^{2d}} \{ \nabla_x \log p_{1|t}^Y(x_1|x) - \nabla_x \log p_{t|0}^Y(x|x_0) \} p_{t|0}^Y(x|x_0) p_{1|t}^Y(x_1|x) \tilde{\pi}(dx_0, dx_1)}{p_t^I(x)} + \beta(x),$$

we have proven that $p_t^I(x)$ satisfies the following continuity equation

$$\partial_t p_t^I(x) + \operatorname{div}_x(v_t^Y(x) p_t^I(x)) = 0, \quad t \in (0, 1), \quad x \in \mathbb{R}^d.$$

Consequently, if we define

$$b_t^Y(x) = v_t^Y(x) + \nabla_x \log p_t^I(x),$$

we have that, under **H4**, $p_t^I(x)$ satisfies the following Fokker-Planck equation

$$\partial_t p_t^I(x) + \operatorname{div}_x(b_t^Y(x) p_t^I(x)) - \Delta_x p_t^I(x) = 0, \quad t \in (0, 1), \quad x \in \mathbb{R}^d. \quad (17)$$

So, $b_t(x)$ is a mimicking drift. It remains to show that $b^Y \equiv \tilde{\beta}^Y$: the thesis will then follow from the uniqueness of the solution to the Fokker Planck equation (17) under **H5**, see [BKRS15, Theorem 9.4.3]. To this aim, note that

$$\begin{aligned} \nabla_x \log p_t^I(x) &= \frac{\int_{\mathbb{R}^{2d}} \{ \nabla_x p_{t|0}^Y(x|x_0) p_{1|t}^Y(x_1|x) + p_{t|0}^Y(x|x_0) \nabla_x p_{1|t}^Y(x_1|x) \} \tilde{\pi}(dx_0, dx_1)}{p_t^I(x)} \\ &= \frac{\int_{\mathbb{R}^{2d}} \{ \nabla_x \log p_{t|0}^Y(x|x_0) + \nabla_x \log p_{1|t}^Y(x_1|x) \} p_{t|0}^Y(x|x_0) p_{1|t}^Y(x_1|x) \tilde{\pi}(dx_0, dx_1)}{p_t^I(x)}. \end{aligned}$$

Therefore, for any $t \in [0, 1)$ and $x \in \mathbb{R}^d$, we have that

$$b_t^Y(x) = 2 \frac{\int_{\mathbb{R}^{2d}} \nabla_x \log p_{t|0}^Y(x|x_0) p_{t|0}^Y(x|x_0) p_{1|t}^Y(x_1|x) \tilde{\pi}(dx_0, dx_1)}{p_t^I(x)} + \beta(x) = \tilde{\beta}_t^Y(x).$$

□

A.2 Preliminary results

We begin this section with two remarks aiming at highlighting some of the properties of the heat kernels (12) and some of their important consequences.

Remark 8. It is well-known that $(s, x, y) \mapsto p_s(y|x)$ defined in (12) is twice continuously differentiable in the space variables x and y and satisfies for $x, y \in \mathbb{R}^d, s \in (0, 1]$,

$$\nabla_x p_s(y|x) = -\frac{x-y}{2s} p_s(y|x) = -\nabla_y p_s(y|x), \quad (18)$$

$$\nabla_x^2 p_s(y|x) = -\frac{1}{2s} p_s(y|x) \operatorname{Id} + \frac{(x-y)(x-y)^T}{4s^2} p_s(y|x) = \nabla_y^2 p_s(y|x), \quad (19)$$

$$\Delta_x p_s(y|x) = -\frac{d}{2s} p_s(y|x) + \left\| \frac{x-y}{2s} \right\|^2 p_s(y|x) = \Delta_y p_s(y|x). \quad (20)$$

Moreover (12) satisfies the heat equation, *i.e.*,

$$\partial_s p_s(y|x) = \Delta_x p_s(y|x), \quad s \in (0, 1], \quad x, y \in \mathbb{R}^d. \quad (21)$$

Thus, in particular, $(s, x, y) \mapsto p_s(y|x)$ is continuously differentiable in the time variable s .

Remark 9. In the case $\beta \equiv 0$, *i.e.*, $\mathbb{Q}^\beta = \mathbb{B}$, under **H1**, the conditions of Theorem 1 hold. Indeed, as highlighted in Remark 8, $(s, x, y) \mapsto p_s(y|x)$ defined in (12) is continuously differentiable in the time variable s and twice continuously differentiable in the space variables x and y . Also, using Equation (18), Equation (19) and (12), it's straightforward to verify that $\nabla_x p_{t|s}^Y(y|x), \nabla_y p_{t|s}^Y(y|x), \nabla_x^2 p_{t|s}^Y(y|x), \nabla_y^2 p_{t|s}^Y(y|x)$ are bounded. Additionally, using (20) and (21), it's easy to argue that also $\partial_t p_{t|s}^Y(y|x)$ and $\partial_s p_{t|s}^Y(y|x)$ are bounded. So **H4** is verified and p_t^I solves the Fokker-Planck equation

(16). Moreover $\tilde{\beta}$ is clearly locally bounded on $\mathbb{R}^d \times (0, 1)$ and, as a consequence of (18), (5), and Jensen inequality, it satisfies uniformly in time

$$\begin{aligned} \int_{\mathbb{R}^d} \|\tilde{\beta}_t(x)\|^2 p_t^I(x) dx &= \int_{\mathbb{R}^d} \left\| \frac{1}{p_t^I(x)} \int_{\mathbb{R}^{2d}} \frac{x_1 - x}{1-t} p_t(x|x_0) p_{1-t}(x_1|x) \tilde{\pi}(dx_0, dx_1) \right\|^2 p_t^I(x) dx \\ &= \mathbb{E} \left[\left\| \mathbb{E} \left[\frac{X_1^I - X_t^I}{1-t} \middle| X_t^I \right] \right\|^2 \right] \\ &= \mathbb{E} \left[\left\| \mathbb{E} \left[X_1^I - X_0^I - \frac{\sqrt{2t}}{\sqrt{1-t}} Z \middle| X_t^I \right] \right\|^2 \right] \\ &\lesssim \mathbb{E} \left[\left\| X_1^I - X_0^I - \sqrt{2t} Z \right\|^2 \right] \\ &\lesssim \mathbf{m}_2[\mu] + \mathbf{m}_2[\nu^*] + d, \end{aligned}$$

which is finite under **H1**. It follows that $\int \|\tilde{\beta}_t^Y(x)\| p_t^I(x) dx dt < +\infty$. So, **H5** is verified.

Additionally, hereunder, we state three lemmas that will be crucial in the derivation of the bounds provided in Theorems 2 and 3.

Lemma 1. For any $p \geq 1$, they hold

$$\mathbb{E}[\|X_s^I - X_0^I\|^{2p}] \lesssim s^{2p} \mathbf{m}_{2p}[\mu] + s^{2p} \mathbf{m}_{2p}[\nu^*] + d^p s^p (1-s)^p,$$

and

$$\mathbb{E}[\|X_1^I - X_s^I\|^{2p}] ds \lesssim (1-s)^{2p} \mathbf{m}_{2p}[\mu] + (1-s)^{2p} \mathbf{m}_{2p}[\nu^*] + d^p s^p (1-s)^p.$$

Proof of Lemma 1: As a direct consequence of (5) and Young inequality, it holds

$$\begin{aligned} \mathbb{E}[\|X_s^I - X_0^I\|^{2p}] &= \mathbb{E} \left[\left\| s(X_1^I - X_0^I) + \sqrt{2s(1-s)} Z \right\|^{2p} \right] \\ &\lesssim s^{2p} \mathbb{E}[\|X_0^I - X_1^I\|^{2p}] + (2s(1-s))^p \mathbb{E}[\|Z\|^{2p}] \\ &\lesssim s^{2p} \mathbf{m}_{2p}[\mu] + s^{2p} \mathbf{m}_{2p}[\nu^*] + d^p s^p (1-s)^p. \end{aligned}$$

A similar argument holds for $\mathbb{E}[\|X_1^I - X_s^I\|^{2p}]$. □

We preface the next lemma with a definition.

Definition 1. Consider $\mathbb{Q} \in \mathcal{P}(\mathbb{W})$. The reverse time measure $\mathbb{Q}^R \in \mathcal{P}(\mathbb{W})$ of \mathbb{Q} is defined as follows: for any $A \in \mathcal{B}(\mathbb{W})$

$$\mathbb{Q}^R(A) = \mathbb{Q}(A^R),$$

where $A^R := \{t \mapsto \omega(1-t) : \omega \in A\}$.

Lemma 2. Assume $\pi \ll \text{Leb}^{2d}$ and $\mu, \nu^* \ll \text{Leb}^d$. Then $(X_t^I)_{t \in [0,1]}$ solves weakly

$$d\vec{X}_t = 2\vec{b}_t(\vec{X}_0, \vec{X}_t) dt + \sqrt{2} d\vec{B}_t, \quad t \in [0,1], \quad \vec{X}_0 \sim \mu. \quad (22)$$

with $(\vec{B}_t)_{t \in [0,1]}$ d -dimensional Brownian motion independent of \vec{X}_0 and

$$\vec{b}_t(x_0, x) = \nabla_x \psi_t^{x_0}(x),$$

where $\psi_t^{x_0}(x)$ solves

$$\partial_t \psi_t^{x_0} + \Delta_x \psi_t^{x_0} + \|\nabla_x \psi_t^{x_0}\|^2 = 0, \quad t \in [0,1], \quad \psi_1^{x_0} = \log \tilde{\pi}_0^{x_0}, \quad (23)$$

with

$$\tilde{\pi}_0^{x_0}(x) := \frac{d\pi(x_0, x)}{d\text{Leb}^{2d}} \frac{1}{p_1(x|x_0)} \bigg/ \frac{d\mu(x_0)}{d\text{Leb}^d}, \quad (24)$$

and $p_1(x|x_0)$ defined in (12).
 Similarly, $((X_t^1)_{t \in [0,1]})^{\mathbb{R}}$ solves weakly

$$d\overleftarrow{X}_t = 2\overleftarrow{b}_t(\overleftarrow{X}_0, \overleftarrow{X}_t)dt + \sqrt{2}d\overleftarrow{B}_t, \quad t \in [0, 1], \quad \overleftarrow{X}_0 \sim \nu^*.$$

with $(\overleftarrow{B}_t)_{t \in [0,1]}$ d -dimensional Brownian motion independent of \overleftarrow{X}_0 and

$$\overleftarrow{b}_t(x_1, x) = \nabla_x \varphi_{1-t}^{x_1}(x),$$

where $\varphi_t^{x_1}(x)$ solves

$$-\partial_t \varphi_t^{x_1} + \Delta_x \varphi_t^{x_1} + \|\nabla_x \varphi_t^{x_1}\|^2 = 0, \quad t \in [0, 1], \quad \varphi_0^{x_1} = \log \tilde{\pi}_1^{x_1},$$

with

$$\tilde{\pi}_1^{x_1}(x) := \frac{d\pi(x, x_1)}{d\text{Leb}^{2d}} \frac{1}{p_1(x_1|x)} \Big/ \frac{d\nu^*(x_1)}{d\text{Leb}^d}, \quad (25)$$

and $p_1(x_1|x)$ defined in (12).

Proof of Lemma 2: We just show that $(X_t^1)_{t \in [0,1]}$ solves weakly (22). The argument for $((X_t^1)_{t \in [0,1]})^{\mathbb{R}}$ is similar and therefore omitted.

For a fixed $x_0 \in \mathbb{R}^d$, we denote by $(x_0, \mathbb{A}) \mapsto \mathbb{I}_{\pi, \mathbb{B}}(x_0, \mathbb{A})$ the conditional distribution of $(X_t^1)_{t \in [0,1]}$ given X_0^1 (see e.g. [Kle13, Theorem 8.37] for the existence of this conditional distribution) and by $(B_t^{x_0})_{t \in [0,1]}$ the solution to

$$dB_t^{x_0} = \sqrt{2}dB_t, \quad t \in [0, 1], \quad B_0^{x_0} = x_0.$$

Also, we denote by $(W_t)_{t \in [0,1]}$ the canonical process on the Wiener space \mathbb{W} and by $(\mathcal{F}_t)_{t \in [0,1]}$ the corresponding natural filtration. Note that, as a consequence of the very definition of $(X_t^1)_{t \in [0,1]}$, (23) and Ito's formula applied to $(\psi_t^{x_0}(B_t^{x_0}))_{t \in [0,1]}$, it holds

$$\begin{aligned} & \frac{d\mathbb{I}_{\pi, \mathbb{B}}(x_0, \cdot)}{d\text{Law}((B_t^{x_0})_{t \in [0,1]})}((B_t^{x_0})_{t \in [0,1]}) \\ &= \tilde{\pi}_0^{x_0}(B_1^{x_0}) \\ &= \exp\left(\psi_1^{x_0}(B_1^{x_0})\right) \\ &= \exp\left(\psi_1^{x_0}(B_1^{x_0}) - \psi_0^{x_0}(x_0)\right) \\ &= \exp\left(\psi_1^{x_0}(B_1^{x_0}) - \psi_0^{x_0}(B_0^{x_0}) - \int_0^1 \left\{ \partial_t \psi_t^{x_0} + \Delta_x \psi_t^{x_0} + \|\nabla_x \psi_t^{x_0}\|^2 \right\} (B_t^{x_0}) dt\right) \\ &= \exp\left(\int_0^1 \nabla_x \psi_t^{x_0}(B_t^{x_0}) dB_t^{x_0} - \int_0^1 \|\nabla_x \psi_t^{x_0}(B_t^{x_0})\|^2 dt\right). \end{aligned}$$

Hence, for any $t \in [0, 1]$ we have that

$$\left. \frac{d\mathbb{I}_{\pi, \mathbb{B}}(x_0, \cdot)}{d\text{Law}((B_t^{x_0})_{t \in [0,1]})} \right|_{\mathcal{F}_t} = D_t,$$

where

$$D_t = \exp\left(\int_0^t \nabla_x \psi_s^{x_0}(W_s) dW_s - \int_0^t \|\nabla_x \psi_s^{x_0}(W_s)\|^2 ds\right)$$

is continuous $\text{Law}((B_t^{x_0})_{t \in [0,1]})$ -almost surely and is such that

$$dD_t = D_t \nabla_x \psi_s^{x_0}(W_s) dW_t.$$

Therefore, being $(W_t)_{t \in [0,1]}$ a martingale under $\text{Law}((B_t^{x_0})_{t \in [0,1]})$, as a consequence of Girsanov theorem (see [RY13, Theorem 1.4]), we have that

$$(W_t - (D^{-1}\langle W, D \rangle)_t)_{t \in [0,1]} = \left(W_t - 2 \int_0^t \nabla_x \psi_s^{x_0}(W_s) ds \right)_{t \in [0,1]}$$

is a $((\mathcal{F}_t)_t, \mathbb{I}_{\pi, \mathbb{B}}(x_0, \cdot))$ -martingale with bracket $\sqrt{2}t$. It follows that, under $\mathbb{I}_{\pi, \mathbb{B}}(x_0, \cdot)$, $(W_t)_{t \in [0, 1]}$ solves

$$dW_t = 2\nabla_x \psi_t^{x_0}(W_s)dt + \sqrt{2}dB_t, \quad t \in [0, 1], \quad W_0 = x_0.$$

The thesis is now a direct consequence of the fact that

$$\mathbb{I}(\pi, \mathbb{B})(A) = \int_{\mathbb{R}^d} \mathbb{P}(X^I \in A | X_0^I = x_0) \mu(dx_0) = \int_{\mathbb{R}^d} \mathbb{I}_{\pi, \mathbb{B}}(x_0, A) \mu(dx_0),$$

for any measurable $A \in \mathbb{W}$. □

Lemma 3. Assume **H2**. Then, almost surely, it holds

$$\vec{b}_t(X_0^I, X_t^I) = \mathbb{E} \left[\frac{\nabla \tilde{\pi}_0^{X_0^I}(X_1^I)}{\tilde{\pi}_0^{X_0^I}(X_1^I)} \middle| (X_0^I, X_t^I) \right], \quad (26)$$

where $\tilde{\pi}_0^{x_0}$ is defined as in (24). Moreover, for any $u \in [0, 1 - s]$, $s \in [0, 1]$ and $p \in \{2, 4, 8\}$ it holds

$$\mathbb{E} \left[\left\| \int_u^{u+s} \vec{b}_r(\vec{X}_0, \vec{X}_r) dr \right\|^p \right] \lesssim s^p \left(\|\nabla \log \tilde{\pi}\|_{L^p(\pi)}^p + \|\nabla \log \mu\|_{L^p(\mu)}^p \right). \quad (27)$$

Similarly, almost surely it holds

$$\overleftarrow{b}_t(X_0^I, X_t^I) = \mathbb{E} \left[\frac{\nabla \tilde{\pi}_1^{X_1^I}(X_0^I)}{\tilde{\pi}_1^{X_1^I}(X_0^I)} \middle| (X_0^I, X_t^I) \right], \quad (28)$$

where $\tilde{\pi}_1^{x_1}$ is defined as in (25). Moreover, for any $u \in [0, 1 - s]$, $s \in [0, 1]$ and $p \in \{2, 4, 8\}$ it holds

$$\mathbb{E} \left[\left\| \int_u^{u+s} \overleftarrow{b}_r(\overleftarrow{X}_0, \overleftarrow{X}_r) dr \right\|^p \right] \lesssim s^p \left(\|\nabla \log \tilde{\pi}\|_{L^p(\pi)}^p + \|\nabla \log \nu^*\|_{L^p(\nu^*)}^p \right). \quad (29)$$

Proof of Lemma 3: We just show (26) and (27). The proof of (28) and (29) is similar and therefore omitted. First of all, note that (26) is trivial for $t = 1$. We therefore focus on $t \in [0, 1)$. It's well known that the solution $\psi_t^{x_0}(x)$ to the Hamilton-Jacobi-Bellman equation (23) is given by

$$\psi_t^{x_0}(x) = \log \left(\int \tilde{\pi}_0^{x_0}(x_1) p_{1-t}(x_1|x) dx_1 \right), \quad t \in [0, 1).$$

Therefore, we have that for $t \in [0, 1)$

$$\vec{b}_t(x_0, x) = \nabla_x \psi_t^{x_0}(x) = \frac{\int \tilde{\pi}_0^{x_0}(x_1) \nabla_x p_{1-t}(x_1|x) dx_1}{\int \tilde{\pi}_0^{x_0}(\tilde{x}_1) p_{1-t}(\tilde{x}_1|x) d\tilde{x}_1}.$$

Using (18) and integrating by parts, we get for $t \in [0, 1)$ that

$$\vec{b}_t(x_0, x) = - \frac{\int \tilde{\pi}_0^{x_0}(x_1) \nabla_{x_1} p_{1-t}(x_1|x) dx_1}{\int \tilde{\pi}_0^{x_0}(\tilde{x}_1) p_{1-t}(\tilde{x}_1|x) d\tilde{x}_1} = \frac{\int (\nabla \tilde{\pi}_0^{x_0}(x_1) / \tilde{\pi}_0^{x_0}(x_1)) \tilde{\pi}_0^{x_0}(x_1) p_{1-t}(x_1|x) dx_1}{\int \tilde{\pi}_0^{x_0}(\tilde{x}_1) p_{1-t}(\tilde{x}_1|x) d\tilde{x}_1}.$$

But then, it suffices to prove that for $t \in [0, 1)$ it holds

$$\frac{\tilde{\pi}_0^{x_0}(x_1) p_{1-t}(x_1|x)}{\int \tilde{\pi}_0^{x_0}(\tilde{x}_1) p_{1-t}(\tilde{x}_1|x) d\tilde{x}_1} = p_{1|0,t}^I(x_1|x_0, x),$$

to conclude. To do so, we simply use (4).

$$\begin{aligned} p_{1|0,t}^I(x_1|x_0, x) &= \frac{p_{0,1,t}^I(x_0, x_1, x)}{\int p_{0,1,t}^I(x_0, \tilde{x}_1, x) d\tilde{x}_1} \\ &= \frac{\pi(x_0, x_1) p_t(x|x_0) p_{1-t}(x_1|x)}{p_1(x_1|x_0)} \bigg/ \int \frac{\pi(x_0, \tilde{x}_1) p_t(x|x_0) p_{1-t}(\tilde{x}_1|x)}{p_1(\tilde{x}_1|x_0)} d\tilde{x}_1 \\ &= \frac{\pi(x_0, x_1) p_{1-t}(x_1|x)}{p_1(x_1|x_0)} \bigg/ \int \frac{\pi(x_0, \tilde{x}_1) p_{1-t}(\tilde{x}_1|x)}{p_1(\tilde{x}_1|x_0)} d\tilde{x}_1 \\ &= \frac{\pi(x_0, x_1) p_{1-t}(x_1|x)}{\mu(x_0) p_1(x_1|x_0)} \bigg/ \int \frac{\pi(x_0, \tilde{x}_1) p_{1-t}(\tilde{x}_1|x)}{\mu(x_0) p_1(\tilde{x}_1|x_0)} d\tilde{x}_1 \\ &= \frac{\tilde{\pi}_0^{x_0}(x_1) p_{1-t}(x_1|x)}{\int \tilde{\pi}_0^{x_0}(\tilde{x}_1) p_{1-t}(\tilde{x}_1|x) d\tilde{x}_1}. \end{aligned}$$

We are left with the bounds (27).

We prove (27) only for $p = 1$. The other two cases are analogous. To this aim, we simply make use of (22), Hölder and Jensen inequalities and the properties of the conditional expectation.

$$\begin{aligned} \mathbb{E} \left[\left\| \int_u^{u+s} \vec{b}_r(\vec{X}_0, \vec{X}_r) dr \right\|^2 \right] &= \mathbb{E} \left[\left\| \int_u^{u+s} \mathbb{E} \left[\frac{\nabla \tilde{\pi}_0^{X_0^1}(X_1^1)}{\tilde{\pi}_0^{X_0^1}(X_1^1)} \middle| (X_0^1, X_r^1) \right] dr \right\|^2 \right] \\ &\lesssim s \mathbb{E} \left[\int_u^{u+s} \left\| \mathbb{E} \left[\frac{\nabla \tilde{\pi}_0^{X_0^1}(X_1^1)}{\tilde{\pi}_0^{X_0^1}(X_1^1)} \middle| (X_0^1, X_r^1) \right] \right\|^2 dr \right] \\ &\lesssim s \mathbb{E} \left[\int_u^{u+s} \mathbb{E} \left[\left\| \frac{\nabla \tilde{\pi}_0^{X_0^1}(X_1^1)}{\tilde{\pi}_0^{X_0^1}(X_1^1)} \right\|^2 \middle| (X_0^1, X_r^1) \right] dr \right] \\ &= s \mathbb{E} \left[\int_u^{u+s} \left\| \frac{\nabla \tilde{\pi}_0^{X_0^1}(X_1^1)}{\tilde{\pi}_0^{X_0^1}(X_1^1)} \right\|^2 dr \right] \\ &\lesssim s^2 \int \|\nabla \log \tilde{\pi}_0^{x_0}(x_1)\|^2 d\pi(x_0, x_1) \\ &\lesssim s^2 \left(\|\nabla \log \tilde{\pi}\|_{L^2(\pi)}^2 + \|\nabla \log \mu\|_{L^2(\mu)}^2 \right), \end{aligned}$$

where, in the last inequality, we have used the very definition of $\tilde{\pi}_0^{x_0}$ given in (24). \square

A.3 Main results

We preface the proofs of our main results with some extra notation. For sake of brevity, we denote by

$$\vec{f}_u^t = 2 \int_u^{t+u} \vec{b}_r(\vec{X}_0, \vec{X}_r) dr, \quad \vec{g}_u^t = \vec{B}_{t+u} - \vec{B}_u,$$

and similarly

$$\overleftarrow{f}_u^t = 2 \int_u^{t+u} \overleftarrow{b}_r(\overleftarrow{X}_0, \overleftarrow{X}_r) dr, \quad \overleftarrow{g}_u^t = \overleftarrow{B}_{t+u} - \overleftarrow{B}_u.$$

Remark 10. With this new notation, according to Lemma 2, we have that

$$(X_t^I)_{t \in [0,1]} \stackrel{\text{dist}}{=} (\vec{X}_t)_{t \in [0,1]}, \quad (X_t^I)_{t \in [0,1]} \stackrel{\text{dist}}{=} ((\overleftarrow{X}_t)_{t \in [0,1]})^R,$$

and that for any $u \in [0, 1]$ and $t \in [0, 1 - u]$,

$$X_{t+u}^I - X_u^I \stackrel{\text{dist}}{=} \vec{f}_u^t + \vec{g}_u^t, \quad X_{1-(t+u)}^I - X_{1-u}^I \stackrel{\text{dist}}{=} \overleftarrow{f}_u^t + \overleftarrow{g}_u^t.$$

Furthermore, according to Lemma 3, under **H2**, for any $u \in [0, 1]$, $t \in [0, 1 - u]$, $p \in \{2, 4, 8\}$, we have that

$$\mathbb{E} \left[\left\| \vec{f}_u^t \right\|^p \right] \lesssim t^p (\|\nabla \log \mu\|_{L^p(\mu)}^p + \|\nabla \log \tilde{\pi}\|_{L^p(\pi)}^p),$$

and

$$\mathbb{E} \left[\left\| \overleftarrow{f}_u^t \right\|^p \right] \lesssim t^p (\|\nabla \log \nu^*\|_{L^p(\nu^*)}^p + \|\nabla \log \tilde{\pi}\|_{L^p(\pi)}^p).$$

Remark 11. Note that for any $u \in [0, 1]$ and $t \in [0, 1 - u]$

$$\vec{g}_u^t \perp (\vec{X}_r)_{r \leq u}, \quad \overleftarrow{g}_u^t \perp (\overleftarrow{X}_r)_{r \leq u}.$$

This fact is an almost immediate consequence of the Markov property of $(\vec{B}_t)_{t \in [0,1]}$, see [BB17, Theorem 3.3]: consider the filtration $(\mathcal{F}_t)_{t \in [0,1]}$ defined by $\mathcal{F}_t = \sigma(\vec{X}_0, (\vec{B}_u)_{u \leq t})$. Then, being $(\vec{B}_t)_{t \in [0,1]} \perp \vec{X}_0$, $(\vec{B}_t)_{t \in [0,1]}$ is $(\mathcal{F}_t)_{t \in [0,1]}$ -adapted and, as a consequence of the Markov property of $(\vec{B}_t)_{t \in [0,1]}$, $(\vec{g}_u^t)_{t \in [0,1-u]} = (\vec{B}_{t+u} - \vec{B}_u)_{t \in [0,1-u]} \perp \mathcal{F}_u$. On the other hand, it's well known that $(\vec{X}_t)_{t \in [0,1]}$ is $(\mathcal{F}_t)_{t \in [0,1]}$ -adapted. It therefore follows that $\vec{g}_u^t \perp (\vec{X}_r)_{r \leq u}$. A similar arguments holds for $(\overleftarrow{X}_t)_{t \in [0,1]}$.

A.4 Proof of Theorem 2

We fix $0 < \epsilon < \min\{h, 1/2\}$ and, for any $t \in [0, 1 - \epsilon]$, we denote by $\nu_t^* = \text{Law}(X_t^M)$ and $\nu_t^{\theta^*} = \text{Law}(X_t^{\theta^*})$.

First, using the data processing inequality [Nut21, Lemma 1.6], the standard decomposition of the KL divergence [CCL+23b, CLL23, CDS23] based on Girsanov theorem, triangle inequality and H3, we bound the KL divergence between $\nu_{1-\epsilon}^*$ and $\nu_{1-\epsilon}^{\theta^*}$ as follows

$$\begin{aligned} & \text{KL}(\nu_{1-\epsilon}^* | \nu_{1-\epsilon}^{\theta^*}) \\ & \leq \text{KL}(\text{Law}((X_t^M)_{t \in [0, 1-\epsilon]}) | \text{Law}((X_t^{\theta^*})_{t \in [0, 1-\epsilon]})) \\ & \lesssim \sum_{k=0}^{N-2} \int_{t_k}^{t_{k+1}} \mathbb{E} \left[\left\| s_{\theta^*}(t_k, X_{t_k}^M) - \tilde{\beta}_t(X_{t_k}^M) \right\|^2 \right] dt + \int_{1-h}^{1-\epsilon} \mathbb{E} \left[\left\| s_{\theta^*}(1-h, X_{1-h}^M) - \tilde{\beta}_t(X_{1-h}^M) \right\|^2 \right] dt \\ & \lesssim \epsilon^2 + \sum_{k=0}^{N-2} \int_{t_k}^{t_{k+1}} \mathbb{E} \left[\left\| \tilde{\beta}_{t_k}(X_{t_k}^M) - \tilde{\beta}_t(X_{t_k}^M) \right\|^2 \right] dt + \int_{1-h}^{1-\epsilon} \mathbb{E} \left[\left\| \tilde{\beta}_{1-h}(X_{1-h}^M) - \tilde{\beta}_t(X_{1-h}^M) \right\|^2 \right] dt . \end{aligned} \tag{30}$$

Second, we aim at bounding the RHS of (30) uniformly in ϵ . Indeed, if we assume to be able to bound it with a constant A independent of ϵ , then, using the weak convergence of $X_{1-\epsilon}^M$ to X_1^I (whose law is given by ν^*) as $\epsilon \rightarrow 0$, the continuity of $(X_t^{\theta^*})_{t \in [0, 1]}$, hence the weak convergence of $X_{1-\epsilon}^{\theta^*}$ to $X_1^{\theta^*}$ (whose law is given by $\nu_1^{\theta^*}$) as $\epsilon \rightarrow 0$, and the lower semi-continuity of the KL-divergence with respect to the weak convergence [VEH14, Theorem 19], we will get

$$\text{KL}(\nu^* | \nu_1^{\theta^*}) \leq \liminf_{\epsilon \rightarrow 0} \text{KL}(\nu_{1-\epsilon}^* | \nu_{1-\epsilon}^{\theta^*}) \lesssim \liminf_{\epsilon \rightarrow 0} A = A . \tag{31}$$

Let us therefore bound the RHS of (30). We will do so by using stochastic calculus tools, and, more precisely, Ito's formula. To this aim let us introduce the generator of $(X_t^M)_{t \in [0, 1-\epsilon]}$, which is defined for any $t \in [0, 1 - \epsilon]$ and $\rho \in C^2(\mathbb{R}^d)$ as

$$\mathcal{L}_t^M \rho := \langle \nabla_x \rho, \tilde{\beta}_t \rangle + \Delta_x \rho .$$

Using Ito's formula, we get that

$$d\tilde{\beta}_t(X_t^M) = (\partial_t + \mathcal{L}_t^M)\tilde{\beta}_t(X_t^M)dt + \sqrt{2}D_x\tilde{\beta}_t(X_t^M)dB_t , \quad t \in [0, 1 - \epsilon] .$$

So, applying Young inequality and Ito's isometry, we have that, for any $k = 0, \dots, N - 1$

$$\begin{aligned} & \mathbb{E} \left[\left\| \tilde{\beta}_{t_k}(X_{t_k}^M) - \tilde{\beta}_t(X_t^M) \right\|^2 \right] \\ & = \mathbb{E} \left[\left\| \int_{t_k}^t (\partial_s + \mathcal{L}_s^M)\tilde{\beta}_s(X_s^M)ds + \sqrt{2} \int_{t_k}^t D_x\tilde{\beta}_s(X_s^M)dB_s \right\|^2 \right] \\ & \lesssim \mathbb{E} \left[\left\| \int_{t_k}^t (\partial_s + \mathcal{L}_s^M)\tilde{\beta}_s(X_s^M)ds \right\|^2 \right] + 2 \int_{t_k}^t \mathbb{E} \left[\left\| D_x\tilde{\beta}_s(X_s^M) \right\|^2 \right] ds . \end{aligned}$$

We now bound separately the two upper addends. To do so, we introduce the auxiliary measures $\lambda_k^h(ds) \in \mathcal{P}([t_k, t_{k+1}])$ for $k = 0, \dots, N - 2$ and $\lambda_{N-1}^h(ds) \in \mathcal{P}([1-h, 1-\epsilon])$ which will help us, via a double change of measure argument, to mitigate the bad behaviour at $t = 0$ and $t = 1$ of the reciprocal characteristic of the mimicking drift (i.e., $\partial_s + \mathcal{L}_s^M$), which is the trickiest addend. Namely, for $k = 0, \dots, N - 2$ we consider the measures $\lambda_k^h(ds) \in \mathcal{P}([t_k, t_{k+1}])$ defined as

$$\lambda_k^h(ds) = \frac{\rho(s)^{-1}}{Z_k} \mathbb{1}_{[t_k, t_{k+1}]} ds ,$$

with

$$\rho(s)^{-1} = s^{-7/8} \mathbb{1}_{\{s \leq 1/2\}} + (1-s)^{-7/8} \mathbb{1}_{\{s > 1/2\}} ,$$

and

$$Z_k = \int_{\min\{t_k, 1/2\}}^{\min\{t_{k+1}, 1/2\}} r^{-7/8} dr + \int_{\max\{t_k, 1/2\}}^{\max\{t_{k+1}, 1/2\}} (1-r)^{-7/8} dr .$$

Whereas, for $k = N - 1$, we consider the measures $\lambda_{N-1}^h(ds) \in \mathcal{P}([1 - h, 1 - \epsilon])$ defined as

$$\lambda_{N-1}^h(ds) = \frac{\rho(s)^{-1}}{Z_{N-1}} \mathbb{1}_{[1-h, 1-\epsilon]} ds,$$

with $\rho(s)^{-1}$ as above and

$$Z_{N-1} = \int_{\min\{1-h, 1/2\}}^{\min\{1-\epsilon, 1/2\}} r^{-7/8} dr + \int_{\max\{1-h, 1/2\}}^{\max\{1-\epsilon, 1/2\}} (1-r)^{-7/8} dr.$$

Note that, for any $s \in [0, 1 - \epsilon]$ and for any $k = 0, \dots, N - 1$, they hold

$$\rho(s) \lesssim 1, \quad Z_k \lesssim h^{1/8}, \quad \int_0^{1-\epsilon} \rho^{-1}(s) ds = 16 \left(\frac{1}{2}\right)^{1/8} - 8\epsilon^{1/8} \lesssim 1. \quad (32)$$

We start by bounding the first addend, that is the one that involves the reciprocal characteristic of the mimicking drift. With a first change of measure argument, we get for any $k = 0, \dots, N - 1$,

$$\mathbb{E} \left[\left\| \int_{t_k}^t (\partial_s + \mathcal{L}_s^M) \tilde{\beta}_s(X_s^M) ds \right\|^2 \right] = Z_k^2 \mathbb{E} \left[\left\| \int_{t_k}^t (\partial_s + \mathcal{L}_s^M) \tilde{\beta}_s(X_s^M) \rho(s) \lambda_k^h(ds) \right\|^2 \right],$$

where, in the last inequality, we used (32). But then, if we apply Jensen inequality and use an other change of measure argument, we get

$$\begin{aligned} \mathbb{E} \left[\left\| \int_{t_k}^t (\partial_s + \mathcal{L}_s^M) \tilde{\beta}_s(X_s^M) ds \right\|^2 \right] &\leq Z_k^2 \mathbb{E} \left[\int_{t_k}^t \left\| (\partial_s + \mathcal{L}_s^M) \tilde{\beta}_s(X_s^M) \right\|^2 \rho(s)^2 \lambda_k^h(ds) \right] \\ &\leq Z_k \int_{t_k}^t \mathbb{E} \left[\left\| (\partial_s + \mathcal{L}_s^M) \tilde{\beta}_s(X_s^M) \right\|^2 \right] \rho(s) ds \\ &\lesssim h^{1/8} \int_{t_k}^t \mathbb{E} \left[\left\| (\partial_s + \mathcal{L}_s^M) \tilde{\beta}_s(X_s^M) \right\|^2 \right] \rho(s) ds. \end{aligned} \quad (33)$$

Let us now focus on the second addend. Remarkably, this addend can be bounded via the reciprocal characteristic of $\tilde{\beta}$: because of Ito's formula, for $t \in [0, 1 - \epsilon]$, it holds true

$$d \left\| \tilde{\beta}_t(X_t^M) \right\|^2 = \left\{ 2 \langle \tilde{\beta}_t, (\partial_t + \mathcal{L}_t^M) \tilde{\beta}_t \rangle + 2 \left\| D_x \tilde{\beta}_t \right\|^2 \right\} (X_t^M) dt + 2\sqrt{2} \langle \tilde{\beta}_t, D_x \tilde{\beta}_t \rangle (X_t^M) dB_t.$$

Consequently, if we assume that the process $(\int_0^s \langle \tilde{\beta}_t, D_x \tilde{\beta}_t \rangle (X_t^M) dB_t)_{s \in [0, 1-\epsilon]}$ is a true martingale (see Lemma 4 below), we have that

$$\begin{aligned} &2 \int_{t_k}^t \mathbb{E} \left[\left\| D_x \tilde{\beta}_s(X_s^M) \right\|^2 \right] ds \\ &\leq \mathbb{E} \left[\left\| \tilde{\beta}_t(X_t^M) \right\|^2 \right] - \mathbb{E} \left[\left\| \tilde{\beta}_{t_k}(X_{t_k}^M) \right\|^2 \right] + 2 \left| \int_{t_k}^t \mathbb{E}[\langle \tilde{\beta}_s(X_s^M), (\partial_s + \mathcal{L}_s^M) \tilde{\beta}_s(X_s^M) \rangle] ds \right|. \end{aligned}$$

But then, using, as before, a double change of measure argument and applying Cauchy-Schwartz inequality, we can bound the above expression as follows

$$\begin{aligned} &2 \int_{t_k}^t \mathbb{E} \left[\left\| D_x \tilde{\beta}_s(X_s^M) \right\|^2 \right] ds \\ &\leq \mathbb{E} \left[\left\| \tilde{\beta}_t(X_t^M) \right\|^2 \right] - \mathbb{E} \left[\left\| \tilde{\beta}_{t_k}(X_{t_k}^M) \right\|^2 \right] \\ &\quad + 2Z_k \left| \int_{t_k}^t \mathbb{E}[\langle \tilde{\beta}_s(X_s^M), (\partial_s + \mathcal{L}_s^M) \tilde{\beta}_s(X_s^M) \rangle] \rho(s) \lambda_k^h(ds) \right| \\ &\leq \mathbb{E} \left[\left\| \tilde{\beta}_t(X_t^M) \right\|^2 \right] - \mathbb{E} \left[\left\| \tilde{\beta}_{t_k}(X_{t_k}^M) \right\|^2 \right] + Z_k \int_{t_k}^t \mathbb{E} \left[\left\| \tilde{\beta}_s(X_s^M) \right\|^2 \right] \lambda_k^h(ds) \\ &\quad + Z_k \int_{t_k}^t \mathbb{E} \left[\left\| (\partial_s + \mathcal{L}_s^M) \tilde{\beta}_s(X_s^M) \right\|^2 \right] \rho(s)^2 \lambda_k^h(ds) \\ &= \mathbb{E} \left[\left\| \tilde{\beta}_t(X_t^M) \right\|^2 \right] - \mathbb{E} \left[\left\| \tilde{\beta}_{t_k}(X_{t_k}^M) \right\|^2 \right] + \int_{t_k}^t \mathbb{E} \left[\left\| \tilde{\beta}_s(X_s^M) \right\|^2 \right] \rho(s)^{-1} ds \\ &\quad + \int_{t_k}^t \mathbb{E} \left[\left\| (\partial_s + \mathcal{L}_s^M) \tilde{\beta}_s(X_s^M) \right\|^2 \right] \rho(s) ds. \end{aligned}$$

Plugging this bound and (33) in (30), we get

$$\begin{aligned} \text{KL}(\nu_{1-\epsilon}^* | \nu_{1-\epsilon}^{\theta^*}) &\lesssim \varepsilon^2 + h\mathbb{E} \left[\left\| \tilde{\beta}_{1-\epsilon}(X_{1-\epsilon}^M) \right\|^2 \right] + h \int_0^{1-\epsilon} \mathbb{E} \left[\left\| \tilde{\beta}_s(X_s^M) \right\|^2 \right] \rho(s)^{-1} ds \\ &\quad + h(h^{1/8} + 1) \int_0^{1-\epsilon} \mathbb{E} \left[\left\| (\partial_s + \mathcal{L}_s^M) \tilde{\beta}_s(X_s^M) \right\|^2 \right] \rho(s) ds . \end{aligned} \quad (34)$$

We now compute explicitly and upper bound each term appearing in the RHS of (34), recalling that, because of Theorem 1, for any $s \in [0, 1 - \epsilon]$, $\nu_s^* = \text{Law}(X_s^I)$. We start with

$$\mathbb{E} \left[\left\| \tilde{\beta}_{1-\epsilon}(X_{1-\epsilon}^M) \right\|^2 \right] \lesssim \int_{\mathbb{R}^d} \left\| \frac{\int_{\mathbb{R}^{2d}} p_{1-\epsilon}(x|x_0) \nabla_x p_\epsilon(x_1|x) \tilde{\pi}(x_0, x_1) dx_1 dx_0}{p_{1-\epsilon}^I(x)} \right\|^2 p_{1-\epsilon}^I(x) dx .$$

First we use (18), second we integrate by part, third we apply Jensen inequality and last we rely on **H 2**.

$$\begin{aligned} &\mathbb{E} \left[\left\| \tilde{\beta}_{1-\epsilon}(X_{1-\epsilon}^M) \right\|^2 \right] \\ &\lesssim \int_{\mathbb{R}^d} \left\| \frac{\int_{\mathbb{R}^{2d}} p_{1-\epsilon}(x|x_0) p_\epsilon(x_1|x) (\nabla_{x_1} \tilde{\pi}(x_0, x_1) / \tilde{\pi}(x_0, x_1)) \tilde{\pi}(x_0, x_1) dx_0 dx_1}{p_{1-\epsilon}^I(x)} \right\|^2 p_{1-\epsilon}^I(x) dx \\ &= \mathbb{E} \left[\left\| \mathbb{E} \left[\frac{\nabla_{x_1} \tilde{\pi}}{\tilde{\pi}}(X_0^I, X_1^I) \middle| X_{1-\epsilon}^I \right] \right\|^2 \right] \leq \mathbb{E} \left[\left\| \frac{\nabla_{x_1} \tilde{\pi}}{\tilde{\pi}}(X_0^I, X_1^I) \right\|^2 \middle| X_{1-\epsilon}^I \right] \\ &= \mathbb{E} \left[\left\| \frac{\nabla_{x_1} \tilde{\pi}}{\tilde{\pi}}(X_0^I, X_1^I) \right\|^2 \right] = \left\| \frac{\nabla_{x_1} \tilde{\pi}}{\tilde{\pi}} \right\|_{L^2(\pi)}^2 \leq \|\nabla \log \tilde{\pi}\|_{L^2(\pi)}^2 . \end{aligned} \quad (35)$$

In the very same way we deal with the third term of the RHS of (34).

$$\begin{aligned} &\int_0^{1-\epsilon} \mathbb{E} \left[\left\| \tilde{\beta}_s(X_s^M) \right\|^2 \right] \rho(s)^{-1} ds \\ &\lesssim \int_0^{1-\epsilon} \rho(s)^{-1} \int_{\mathbb{R}^d} \left\| \frac{\int_{\mathbb{R}^{2d}} p_s(x|x_0) \nabla_x p_{1-s}(x_1|x) \tilde{\pi}(x_0, x_1) dx_0 dx_1}{p_s^I(x)} \right\|^2 p_s^I(x) dx ds \\ &= \int_0^{1-\epsilon} \rho(s)^{-1} \int_{\mathbb{R}^d} \left\| \frac{1}{p_s^I(x)} \int_{\mathbb{R}^{2d}} p_s(x|x_0) p_{1-s}(x_1|x) \frac{\nabla_{x_1} \tilde{\pi}}{\tilde{\pi}}(x_0, x_1) \tilde{\pi}(x_0, x_1) dx_0, dx_1 \right\|^2 p_s^I(x) dx \\ &= \int_0^{1-\epsilon} \rho(s)^{-1} \mathbb{E} \left[\left\| \mathbb{E} \left[\frac{\nabla_{x_1} \tilde{\pi}}{\tilde{\pi}}(X_0^I, X_1^I) \middle| X_s^I \right] \right\|^2 \right] ds \\ &\leq \int_0^{1-\epsilon} \rho(s)^{-1} \mathbb{E} \left[\left\| \frac{\nabla_{x_1} \tilde{\pi}}{\tilde{\pi}}(X_0^I, X_1^I) \right\|^2 \middle| X_s^I \right] ds \\ &= \int_0^{1-\epsilon} \rho(s)^{-1} \mathbb{E} \left[\left\| \frac{\nabla_{x_1} \tilde{\pi}}{\tilde{\pi}}(X_0^I, X_1^I) \right\|^2 \right] ds \\ &= \int_0^{1-\epsilon} \rho(s)^{-1} ds \|\nabla \log \tilde{\pi}\|_{L^2(\pi)}^2 \lesssim \|\nabla \log \tilde{\pi}\|_{L^2(\pi)}^2 , \end{aligned} \quad (36)$$

where, in the last inequality, we used (32). We now turn to the last term, *i.e.*, to

$$\int_0^{1-\epsilon} \mathbb{E} \left[\left\| (\partial_s + \mathcal{L}_s^M) \tilde{\beta}_s(X_s^M) \right\|^2 \right] \rho(s) ds = \int_0^{1-\epsilon} \int_{\mathbb{R}^d} \left\| (\partial_s + \mathcal{L}_s^M) \tilde{\beta}_s(x) \right\|^2 \rho(s) p_s^I(x) dx ds .$$

Some computations and (21) lead to

$$\begin{aligned}
 & \partial_s \tilde{\beta}_s(x) \\
 &= 2 \frac{\int_{\mathbb{R}^{2d}} \Delta_x p_s(x|x_0) \nabla_x p_{1-s}(x_1|x) \tilde{\pi}(x_0, x_1) dx_0 dx_1}{p_s^I(x)} \\
 & - 2 \frac{\int_{\mathbb{R}^{2d}} p_s(x|x_0) \nabla_x \Delta_x p_{1-s}(x_1|x) \tilde{\pi}(x_0, x_1) dx_0 dx_1}{p_s^I(x)} \\
 & - 2 \frac{\int_{\mathbb{R}^{2d}} p_s(x|x_0) \nabla_x p_{1-s}(x_1|x) \tilde{\pi}(x_0, x_1) dx_0 dx_1 \int_{\mathbb{R}^{2d}} \Delta_x p_s(x|x_0) p_{1-s}(x_1|x) \tilde{\pi}(x_0, x_1) dx_0 dx_1}{(p_s^I(x))^2} \\
 & + 2 \frac{\int_{\mathbb{R}^{2d}} p_s(x|x_0) \nabla_x p_{1-s}(x_1|x) \tilde{\pi}(x_0, x_1) dx_0 dx_1 \int_{\mathbb{R}^{2d}} p_s(x|x_0) \Delta_x p_{1-s}(x_1|x) \tilde{\pi}(x_0, x_1) dx_0 dx_1}{(p_s^I(x))^2};
 \end{aligned}$$

$$\begin{aligned}
 & D_x \tilde{\beta}_s(x) \\
 &= 2 \frac{\int_{\mathbb{R}^{2d}} \nabla_x p_{1-s}(x_1|x) (\nabla_x p_s(x|x_0))^T \tilde{\pi}(x_0, x_1) dx_0 dx_1}{p_s^I(x)} \\
 & + 2 \frac{\int_{\mathbb{R}^{2d}} p_s(x|x_0) \nabla_x^2 p_{1-s}(x_1|x) \tilde{\pi}(x_0, x_1) dx_0 dx_1}{p_s^I(x)} \\
 & - 2 \left(\frac{\int_{\mathbb{R}^{2d}} p_s(x|x_0) \nabla_x p_{1-s}(x_1|x) \tilde{\pi}(x_0, x_1) dx_0 dx_1}{p_s^I(x)} \right) \left(\frac{\int_{\mathbb{R}^{2d}} \nabla_x p_s(x|x_0) p_{1-s}(x_1|x) \tilde{\pi}(x_0, x_1) dx_0 dx_1}{p_s^I(x)} \right)^T \\
 & - 2 \left(\frac{\int_{\mathbb{R}^{2d}} p_s(x|x_0) \nabla_x p_{1-s}(x_1|x) \tilde{\pi}(x_0, x_1) dx_0 dx_1}{p_s^I(x)} \right) \left(\frac{\int_{\mathbb{R}^{2d}} p_s(x|x_0) \nabla_x p_{1-s}(x_1|x) \tilde{\pi}(x_0, x_1) dx_0 dx_1}{p_s^I(x)} \right)^T,
 \end{aligned} \tag{37}$$

hence

$$\begin{aligned}
 & D_x \tilde{\beta}_s(x) \tilde{\beta}_s(x) \\
 &= 4 \frac{\int_{\mathbb{R}^{2d}} \nabla_x p_{1-s}(x_1|x) (\nabla_x p_s(x|x_0))^T \tilde{\pi}(x_0, x_1) dx_0 dx_1 \int_{\mathbb{R}^{2d}} p_s(x|x_0) \nabla_x p_{1-s}(x_1|x) \tilde{\pi}(x_0, x_1) dx_0 dx_1}{p_s^I(x) p_s^I(x)} \\
 & + 4 \frac{\int_{\mathbb{R}^{2d}} p_s(x|x_0) \nabla_x^2 p_{1-s}(x_1|x) \tilde{\pi}(x_0, x_1) dx_0 dx_1 \int_{\mathbb{R}^{2d}} p_s(x|x_0) \nabla_x p_{1-s}(x_1|x) \tilde{\pi}(x_0, x_1) dx_0 dx_1}{p_s^I(x) p_s^I(x)} \\
 & - 4 \left(\frac{\int_{\mathbb{R}^{2d}} p_s(x|x_0) \nabla_x p_{1-s}(x_1|x) \tilde{\pi}(x_0, x_1) dx_0 dx_1}{p_s^I(x)} \right) \left(\frac{\int_{\mathbb{R}^{2d}} \nabla_x p_s(x|x_0) p_{1-s}(x_1|x) \tilde{\pi}(x_0, x_1) dx_0 dx_1}{p_s^I(x)} \right)^T \\
 & \quad \cdot \frac{\int_{\mathbb{R}^{2d}} p_s(x|x_0) \nabla_x p_{1-s}(x_1|x) \tilde{\pi}(x_0, x_1) dx_0 dx_1}{p_s^I(x)} \\
 & - 4 \frac{\int_{\mathbb{R}^{2d}} p_s(x|x_0) \nabla_x p_{1-s}(x_1|x) \tilde{\pi}(x_0, x_1) dx_0 dx_1}{p_s^I(x)} \left\| \frac{\int_{\mathbb{R}^{2d}} p_s(x|x_0) \nabla_x p_{1-s}(x_1|x) \tilde{\pi}(x_0, x_1) dx_0 dx_1}{p_s^I(x)} \right\|^2;
 \end{aligned} \tag{38}$$

$$\begin{aligned}
& \Delta_x \tilde{\beta}_s(x) \\
&= 2 \frac{\int_{\mathbb{R}^{2d}} \Delta_x p_s(x|x_0) \nabla_x p_{1-s}(x_1|x) \tilde{\pi}(x_0, x_1) dx_0 dx_1}{p_s^I(x)} \\
&+ 4 \frac{\int_{\mathbb{R}^{2d}} \nabla_x^2 p_{1-s}(x|x_0) \nabla_x p_s(x_1|x) \tilde{\pi}(x_0, x_1) dx_0 dx_1}{p_s^I(x)} \\
&- 4 \frac{\int_{\mathbb{R}^{2d}} \nabla_x p_{1-s}(x_1|x) (\nabla_x p_s(x|x_0))^T \tilde{\pi}(x_0, x_1) dx_0 dx_1}{p_s^I(x)} \\
&\quad \cdot \frac{\int_{\mathbb{R}^{2d}} \nabla_x p_s(x|x_0) p_{1-s}(x_1|x) \tilde{\pi}(x_0, x_1) dx_0 dx_1}{p_s^I(x)} \\
&- 4 \frac{\int_{\mathbb{R}^{2d}} \nabla_x p_{1-s}(x_1|x) (\nabla_x p_s(x|x_0))^T \tilde{\pi}(x_0, x_1) dx_0 dx_1}{p_s^I(x)} \\
&\quad \cdot \frac{\int_{\mathbb{R}^{2d}} p_s(x|x_0) \nabla_x p_{1-s}(x_1|x) \tilde{\pi}(x_0, x_1) dx_0 dx_1}{p_s^I(x)} \\
&- 4 \frac{\int_{\mathbb{R}^{2d}} \langle \nabla_x p_{1-s}(x_1|x), \nabla_x p_s(x|x_0) \rangle \tilde{\pi}(x_0, x_1) dx_0 dx_1}{p_s^I(x)} \\
&\quad \cdot \frac{\int_{\mathbb{R}^{2d}} p_s(x|x_0) \nabla_x p_{1-s}(x_1|x) \tilde{\pi}(x_0, x_1) dx_0 dx_1}{p_s^I(x)} \\
&+ 2 \frac{\int_{\mathbb{R}^{2d}} p_s(x|x_0) \nabla_x \Delta_x p_{1-s}(x_1|x) \tilde{\pi}(x_0, x_1) dx_0 dx_1}{p_s^I(x)} \\
&- 4 \frac{\int_{\mathbb{R}^{2d}} p_s(x|x_0) \nabla_x^2 p_{1-s}(x_1|x) \tilde{\pi}(x_0, x_1) dx_0 dx_1 \int_{\mathbb{R}^{2d}} \nabla_x p_s(x|x_0) p_{1-s}(x_1|x) \tilde{\pi}(x_0, x_1) dx_0 dx_1}{(p_s^I(x))^2} \\
&- 4 \frac{\int_{\mathbb{R}^{2d}} p_s(x|x_0) \nabla_x^2 p_{1-s}(x_1|x) \tilde{\pi}(x_0, x_1) dx_0 dx_1 \int_{\mathbb{R}^{2d}} p_s(x|x_0) \nabla_x p_{1-s}(x_1|x) \tilde{\pi}(x_0, x_1) dx_0 dx_1}{(p_s^I(x))^2} \\
&- 2 \frac{\int_{\mathbb{R}^{2d}} \Delta_x p_s(x|x_0) p_{1-s}(x_1|x) \tilde{\pi}(x_0, x_1) dx_0 dx_1 \int_{\mathbb{R}^{2d}} p_s(x|x_0) \nabla_x p_{1-s}(x_1|x) \tilde{\pi}(x_0, x_1) dx_0 dx_1}{(p_s^I(x))^2} \\
&- 2 \frac{\int_{\mathbb{R}^{2d}} p_s(x|x_0) \Delta_x p_{1-s}(x_1|x) \tilde{\pi}(x_0, x_1) dx_0 dx_1 \int_{\mathbb{R}^{2d}} p_s(x|x_0) \nabla_x p_{1-s}(x_1|x) \tilde{\pi}(x_0, x_1) dx_0 dx_1}{(p_s^I(x))^2} \\
&+ 8 \frac{\int_{\mathbb{R}^{2d}} p_s(x|x_0) \nabla_x p_{1-s}(x_1|x) \tilde{\pi}(x_0, x_1) dx_0 dx_1}{p_s^I(x)} \left(\frac{\int_{\mathbb{R}^{2d}} \nabla_x p_s(x|x_0) p_{1-s}(x_1|x) \tilde{\pi}(x_0, x_1) dx_0 dx_1}{p_s^I(x)} \right)^T \\
&\quad \cdot \frac{\int_{\mathbb{R}^{2d}} p_s(x|x_0) \nabla_x p_{1-s}(x_1|x) \tilde{\pi}(x_0, x_1) dx_0 dx_1}{p_s^I(x)} \\
&+ 4 \left\| \frac{\int_{\mathbb{R}^{2d}} p_s(x|x_0) \nabla_x p_{1-s}(x_1|x) \tilde{\pi}(x_0, x_1) dx_0 dx_1}{p_s^I(x)} \right\|^2 \cdot \frac{\int_{\mathbb{R}^{2d}} p_s(x|x_0) \nabla_x p_{1-s}(x_1|x) \tilde{\pi}(x_0, x_1) dx_0 dx_1}{p_s^I(x)} \\
&+ 4 \left\| \frac{\int_{\mathbb{R}^{2d}} \nabla_x p_s(x|x_0) p_{1-s}(x_1|x) \tilde{\pi}(x_0, x_1) dx_0 dx_1}{p_s^I(x)} \right\|^2 \cdot \frac{\int_{\mathbb{R}^{2d}} p_s(x|x_0) \nabla_x p_{1-s}(x_1|x) \tilde{\pi}(x_0, x_1) dx_0 dx_1}{p_s^I(x)}.
\end{aligned}$$

So, we get

$$(\partial_s + \mathcal{L}_s^M) \tilde{\beta}_s(x) = \sum_{k=1}^6 A_s^k(x), \tag{39}$$

where we have defined

$$\begin{aligned}
A_s^1(x) = & -4 \frac{\int_{\mathbb{R}^{2d}} \nabla_x p_{1-s}(x_1|x) (\nabla_x p_s(x|x_0))^T \tilde{\pi}(x_0, x_1) dx_0 dx_1}{p_s^I(x)} \\
& \cdot \frac{\int_{\mathbb{R}^{2d}} \nabla_x p_s(x|x_0) p_{1-s}(x_1|x) \tilde{\pi}(x_0, x_1) dx_0 dx_1}{p_s^I(x)},
\end{aligned}$$

$$\begin{aligned}
A_s^2(x) &= -4 \frac{\int_{\mathbb{R}^{2d}} \langle \nabla_x p_{1-s}(x_1|x), \nabla_x p_s(x|x_0) \rangle \tilde{\pi}(x_0, x_1) dx_0 dx_1}{p_s^I(x)} \cdot \frac{\int_{\mathbb{R}^{2d}} p_s(x|x_0) \nabla_x p_{1-s}(x_1|x) \tilde{\pi}(x_0, x_1) dx_0 dx_1}{p_s^I(x)}, \\
A_s^3(x) &= 4 \frac{\left\| \int_{\mathbb{R}^{2d}} \nabla_x p_s(x|x_0) p_{1-s}(x_1|x) \tilde{\pi}(x_0, x_1) dx_0 dx_1 \right\|^2}{(p_s^I(x))^2} \cdot \frac{\int_{\mathbb{R}^{2d}} p_s(x|x_0) \nabla_x p_{1-s}(x_1|x) \tilde{\pi}(x_0, x_1) dx_0 dx_1}{p_s^I(x)}, \\
A_s^4(x) &= 4 \frac{\int_{\mathbb{R}^{2d}} p_s(x|x_0) \nabla_x p_{1-s}(x_1|x) \tilde{\pi}(x_0, x_1) dx_0 dx_1}{p_s^I(x)} \\
&\quad \cdot \left(\frac{\int_{\mathbb{R}^{2d}} \nabla_x p_s(x|x_0) p_{1-s}(x_1|x) \tilde{\pi}(x_0, x_1) dx_0 dx_1}{p_s^I(x)} \right)^T \frac{\int_{\mathbb{R}^{2d}} p_s(x|x_0) \nabla_x p_{1-s}(x_1|x) \tilde{\pi}(x_0, x_1) dx_0 dx_1}{p_s^I(x)}, \\
A_s^5(x) &= 4 \frac{\int_{\mathbb{R}^{2d}} \Delta_x p_s(x|x_0) \nabla_x p_{1-s}(x_1|x) \tilde{\pi}(x_0, x_1) dx_0 dx_1}{p_s^I(x)} \\
&\quad - 4 \frac{\int_{\mathbb{R}^{2d}} \Delta_x p_s(x|x_0) p_{1-s}(x_1|x) \tilde{\pi}(x_0, x_1) dx_0 dx_1 \int_{\mathbb{R}^{2d}} p_s(x|x_0) \nabla_x p_{1-s}(x_1|x) \tilde{\pi}(x_0, x_1) dx_0 dx_1}{(p_s^I(x))^2}, \\
A_s^6(x) &= 4 \frac{\int_{\mathbb{R}^{2d}} \nabla_x^2 p_{1-s}(x_1|x) \nabla_x p_s(x|x_0) \tilde{\pi}(x_0, x_1) dx_0 dx_1}{p_s^I(x)} \\
&\quad - 4 \frac{\int_{\mathbb{R}^{2d}} p_s(x|x_0) \nabla_x^2 p_{1-s}(x_1|x) \tilde{\pi}(x_0, x_1) dx_0 dx_1 \int_{\mathbb{R}^{2d}} \nabla_x p_s(x|x_0) p_{1-s}(x_1|x) \tilde{\pi}(x_0, x_1) dx_0 dx_1}{(p_s^I(x))^2}.
\end{aligned}$$

Therefore

$$\int_0^{1-\epsilon} \mathbb{E} \left[\left\| (\partial_s + \mathcal{L}_s^M) \tilde{\beta}_s(X_s^M) \right\|^2 \right] \rho(s) ds \lesssim \sum_{k=1}^6 \int_0^{1-\epsilon} \int_{\mathbb{R}^d} \|A_s^k(x)\|^2 \rho(s) p_s^I(x) dx ds.$$

We now bound each term A_s^k in the sum. Using (32) and Young inequality, we have

$$\begin{aligned}
&\int_0^{1-\epsilon} \int_{\mathbb{R}^d} \|A_s^1(x)\|^2 \rho(s) p_s^I(x) dx ds \\
&\lesssim \int_0^{1-\epsilon} \int_{\mathbb{R}^d} \left\| \frac{\int_{\mathbb{R}^{2d}} \nabla_x p_s(x|x_0) p_{1-s}(x_1|x) \tilde{\pi}(x_0, x_1) dx_0 dx_1}{p_s^I(x)} \right\|^4 p_s^I(x) dx ds \\
&\quad + \int_0^{1-\epsilon} \int_{\mathbb{R}^d} \left\| \frac{\int_{\mathbb{R}^{2d}} \nabla_x p_s(x|x_0) (\nabla_x p_{1-s}(x_1|x))^T \tilde{\pi}(x_0, x_1) dx_0 dx_1}{p_s^I(x)} \right\|_{\text{op}}^4 p_s^I(x) dx ds.
\end{aligned}$$

To bound the first term we proceed as in (36), that is, we first exploit (18) and the integration by part formula and we then use Jensen inequality and the properties of the conditional expectation.

$$\int_0^{1-\epsilon} \int_{\mathbb{R}^d} \left\| \frac{\int_{\mathbb{R}^{2d}} \nabla_x p_s(x|x_0) p_{1-s}(x_1|x) \tilde{\pi}(x_0, x_1) dx_0 dx_1}{p_s^I(x)} \right\|^4 p_s^I(x) dx ds \leq \|\nabla \log \tilde{\pi}\|_{L^4(\pi)}^4. \quad (40)$$

To bound the second term we first split the time interval $[0, 1 - \epsilon]$ in two, $[0, 1/2]$ and $[1/2, 1 - \epsilon]$, we second make either $\nabla_x p_s(x|x_0)$ or $\nabla_x p_{1-s}(x_1|x)$ explicit and we last proceed as before, *i.e.*, we exploit (18), we integrate by parts and we use Young and Jensen inequalities.

$$\begin{aligned}
& \int_0^{1-\epsilon} \int_{\mathbb{R}^d} \left\| \frac{\int_{\mathbb{R}^{2d}} \nabla_x p_{1-s}(x_1|x) (\nabla_x p_s(x|x_0))^T \tilde{\pi}(x_0, x_1) dx_0 dx_1}{p_s^I(x)} \right\|_{\text{op}}^4 p_s^I(x) dx ds \\
&= \int_0^{1/2} \int_{\mathbb{R}^d} \left\| \frac{\int_{\mathbb{R}^{2d}} \nabla_x p_s(x|x_0) (\nabla_x p_{1-s}(x_1|x))^T \tilde{\pi}(x_0, x_1) dx_0 dx_1}{p_s^I(x)} \right\|_{\text{op}}^4 p_s^I(x) dx ds \\
&\quad + \int_{1/2}^{1-\epsilon} \int_{\mathbb{R}^d} \left\| \frac{\int_{\mathbb{R}^{2d}} \nabla_x p_{1-s}(x_1|x) (\nabla_x p_s(x|x_0))^T \tilde{\pi}(x_0, x_1) dx_0 dx_1}{p_s^I(x)} \right\|_{\text{op}}^4 p_s^I(x) dx ds \\
&\lesssim \int_0^{1/2} \int_{\mathbb{R}^d} \left\| \frac{1}{p_s^I(x)} \int_{\mathbb{R}^{2d}} \frac{\nabla_{x_0} \tilde{\pi}}{\tilde{\pi}}(x_0, x_1) \frac{(x_1 - x)^T}{1-s} p_s(x|x_0) p_{1-s}(x_1|x) \tilde{\pi}(x_0, x_1) dx_0 dx_1 \right\|_{\text{op}}^4 \\
&\quad \cdot p_s^I(x) dx ds \\
&\quad + \int_{1/2}^{1-\epsilon} \int_{\mathbb{R}^d} \left\| \frac{1}{p_s^I(x)} \int_{\mathbb{R}^{2d}} \frac{\nabla_{x_1} \tilde{\pi}}{\tilde{\pi}}(x_0, x_1) \frac{(x - x_0)^T}{s} p_s(x|x_0) p_{1-s}(x_1|x) \tilde{\pi}(x_0, x_1) dx_0 dx_1 \right\|_{\text{op}}^4 \\
&\quad \cdot p_s^I(x) dx ds \\
&\lesssim \int_0^{1/2} \mathbb{E} \left[\left\| \left(\frac{\nabla_{x_0} \tilde{\pi}}{\tilde{\pi}}(X_0^I, X_1^I) \right) (X_1^I - X_s^I)^T \right\|_{\text{op}}^4 \right] ds \\
&\quad + \int_{1/2}^1 \mathbb{E} \left[\left\| \left(\frac{\nabla_{x_1} \tilde{\pi}}{\tilde{\pi}}(X_0^I, X_1^I) \right) (X_s^I - X_0^I)^T \right\|_{\text{op}}^4 \right] ds \\
&\lesssim \int_0^{1/2} \mathbb{E} \left[\|X_1^I - X_s^I\|^8 \right] ds + \int_{1/2}^1 \mathbb{E} \left[\|X_s^I - X_0^I\|^8 \right] ds + \|\nabla \log \tilde{\pi}\|_{L^8(\pi)}^8 \\
&\lesssim d^4 + \mathbf{m}_s[\mu] + \mathbf{m}_s[\nu^*] + \|\nabla \log \tilde{\pi}\|_{L^8(\pi)}^8,
\end{aligned}$$

where, in the last inequality, we have used Lemma 1.

In a similar way we get

$$\int_0^{1-\epsilon} \int_{\mathbb{R}^d} \|A_s^2(x)\|^2 \rho(s) p_s^I(x) dx ds \lesssim d^4 + \mathbf{m}_s[\mu] + \mathbf{m}_s[\nu^*] + \|\nabla \log \tilde{\pi}\|_{L^8(\pi)}^8.$$

The argument to bound A_s^2 resembles the one used to bound A_s^1 and we therefore omit it.

We now focus on A_s^3 . To bound such term, we first use (32) and Young inequality and we second proceed as in (40) and (36).

$$\begin{aligned}
& \int_0^{1-\epsilon} \int_{\mathbb{R}^d} \|A_s^3(x)\|^2 \rho(s) p_s^I(x) dx ds \\
&\lesssim \int_0^{1-\epsilon} \int_{\mathbb{R}^d} \left\| \frac{\int_{\mathbb{R}^{2d}} \nabla_x p_s(x|x_0) p_{1-s}(x_1|x) \tilde{\pi}(x_0, x_1) dx_0 dx_1}{p_s^I(x)} \right\|_{\text{op}}^8 p_s^I(x) dx ds \\
&\quad + \int_0^{1-\epsilon} \int_{\mathbb{R}^d} \left\| \frac{\int_{\mathbb{R}^{2d}} p_s(x|x_0) \nabla_x p_{1-s}(x_1|x) \tilde{\pi}(x_0, x_1) dx_0 dx_1}{p_s^I(x)} \right\|_{\text{op}}^4 p_s^I(x) dx ds \\
&\lesssim \mathbb{E} \left[\left\| \frac{\nabla_{x_0} \tilde{\pi}}{\tilde{\pi}}(X_0^I, X_1^I) \right\|^8 \right] + \mathbb{E} \left[\left\| \frac{\nabla_{x_1} \tilde{\pi}}{\tilde{\pi}}(X_0^I, X_1^I) \right\|^4 \right] \\
&\lesssim \|\nabla \log \tilde{\pi}\|_{L^8(\pi)}^8 + \|\nabla \log \tilde{\pi}\|_{L^4(\pi)}^4.
\end{aligned}$$

Proceeding in a similar way (we omit the argument, as it is almost a duplication of the previous one), we get

$$\int_0^{1-\epsilon} \int_{\mathbb{R}^d} \|A_s^4(x)\|^2 \rho(s) p_s^I(x) dx ds \lesssim \|\nabla \log \tilde{\pi}\|_{L^8(\pi)}^8 + \|\nabla \log \tilde{\pi}\|_{L^4(\pi)}^4 .$$

We now turn to A_s^5 . Because of (20), A_s^5 rewrites as

$$\begin{aligned} A_s^5(x) &= \frac{1}{p_s^I(x)} \int_{\mathbb{R}^{2d}} \left\{ \frac{-d}{2s} + \frac{\|x - x_0\|^2}{4s^2} \right\} \frac{x_1 - x}{2(1-s)} p_s(x|x_0) p_{1-s}(x_1|x) \tilde{\pi}(x_0, x_1) dx_0 dx_1 \\ &\quad - \left(\frac{1}{p_s^I(x)} \int_{\mathbb{R}^{2d}} \left\{ \frac{-d}{2s} + \frac{\|x - x_0\|^2}{4s^2} \right\} p_s(x|x_0) p_{1-s}(x_1|x) \tilde{\pi}(x_0, x_1) dx_0 dx_1 \right) \\ &\quad \cdot \left(\frac{1}{p_s^I(x)} \int_{\mathbb{R}^{2d}} \frac{x_1 - x}{2(1-s)} p_s(x|x_0) p_{1-s}(x_1|x) \tilde{\pi}(x_0, x_1) dx_0 dx_1 \right) \\ &\lesssim \frac{1}{p_s^I(x)} \int_{\mathbb{R}^{2d}} \frac{\|x - x_0\|^2}{s^2} \frac{x_1 - x}{1-s} p_s(x|x_0) p_{1-s}(x_1|x) \tilde{\pi}(x_0, x_1) dx_0 dx_1 \\ &\quad - \left(\frac{1}{p_s^I(x)} \int_{\mathbb{R}^{2d}} \frac{\|x - x_0\|^2}{s^2} p_s(x|x_0) p_{1-s}(x_1|x) \tilde{\pi}(x_0, x_1) dx_0 dx_1 \right) \\ &\quad \cdot \left(\frac{1}{p_s^I(x)} \int_{\mathbb{R}^{2d}} \frac{x_1 - x}{1-s} p_s(x|x_0) p_{1-s}(x_1|x) \tilde{\pi}(x_0, x_1) dx_0 dx_1 \right) . \end{aligned}$$

Therefore, we have that

$$\begin{aligned} \int_0^{1-\epsilon} \int_{\mathbb{R}^d} \|A_s^5(x)\|^2 \rho(s) p_s^I(x) dx ds &\lesssim \int_0^{1-\epsilon} \mathbb{E} \left[\left\| \mathbb{E} \left[\frac{\|X_s^I - X_0^I\|^2}{s^2} \frac{X_1^I - X_s^I}{1-s} \middle| X_s^I \right] \right\|^2 \right. \\ &\quad \left. - \mathbb{E} \left[\frac{\|X_s^I - X_0^I\|^2}{s^2} \middle| X_s^I \right] \mathbb{E} \left[\frac{X_1^I - X_s^I}{1-s} \middle| X_s^I \right] \right\|^2 \rho(s) ds . \end{aligned}$$

We now split the time interval $[0, 1 - \epsilon]$ in two, $[0, 1/2]$, $[1/2, 1 - \epsilon]$ and we focus on the first one. So we look at $s \in [0, 1/2]$ and we try to bound the integrand. Using Lemma 1, Lemma 2, Lemma 3 and standard and well-known inequalities (Cauchy-Schwarz, Young and Jensen inequalities), we get

that for $s \in [0, 1/2]$

$$\begin{aligned}
& \mathbb{E} \left[\left\| \mathbb{E} \left[\frac{\|X_s^I - X_0^I\|^2}{s^2} \frac{X_1^I - X_s^I}{1-s} \middle| X_s^I \right] - \mathbb{E} \left[\frac{\|X_s^I - X_0^I\|^2}{s^2} \middle| X_s^I \right] \mathbb{E} \left[\frac{X_1^I - X_s^I}{1-s} \middle| X_s^I \right] \right\|^2 \right] s^{7/8} \\
&= \mathbb{E} \left[\left\| \mathbb{E} \left[\frac{\|\widehat{X}_{1-s} - \widehat{X}_1\|^2}{s^2} \frac{\widehat{X}_0 - \widehat{X}_{1-s}}{1-s} \middle| \widehat{X}_{1-s} \right] \right. \right. \\
&\quad \left. \left. - \mathbb{E} \left[\frac{\|\widehat{X}_{1-s} - \widehat{X}_1\|^2}{s^2} \middle| \widehat{X}_{1-s} \right] \mathbb{E} \left[\frac{\widehat{X}_0 - \widehat{X}_{1-s}}{1-s} \middle| \widehat{X}_{1-s} \right] \right\|^2 \right] s^{7/8} \\
&= \mathbb{E} \left[\left\| \mathbb{E} \left[\frac{\|\overleftarrow{f}_{1-s}^s\|^2 + 2\langle \overleftarrow{f}_{1-s}^s, \overleftarrow{g}_{1-s}^s \rangle + \|\overleftarrow{g}_{1-s}^s\|^2}{s^2} (\widehat{X}_0 - \widehat{X}_{1-s}) \middle| \widehat{X}_{1-s} \right] \right. \right. \\
&\quad \left. \left. - \mathbb{E} \left[\frac{\|\overleftarrow{f}_{1-s}^s\|^2 + 2\langle \overleftarrow{f}_{1-s}^s, \overleftarrow{g}_{1-s}^s \rangle + \|\overleftarrow{g}_{1-s}^s\|^2}{s^2} \middle| \widehat{X}_{1-s} \right] \mathbb{E} [\widehat{X}_0 - \widehat{X}_{1-s} \middle| \widehat{X}_{1-s}] \right\|^2 \right] s^{7/8} \\
&= \mathbb{E} \left[\left\| \mathbb{E} \left[\frac{\|\overleftarrow{f}_{1-s}^s\|^2 + 2\langle \overleftarrow{f}_{1-s}^s, \overleftarrow{g}_{1-s}^s \rangle}{s^2} (\widehat{X}_0 - \widehat{X}_{1-s}) \middle| \widehat{X}_{1-s} \right] \right. \right. \\
&\quad \left. \left. - \mathbb{E} \left[\frac{\|\overleftarrow{f}_{1-s}^s\|^2 + 2\langle \overleftarrow{f}_{1-s}^s, \overleftarrow{g}_{1-s}^s \rangle}{s^2} \middle| \widehat{X}_{1-s} \right] \mathbb{E} [\widehat{X}_0 - \widehat{X}_{1-s} \middle| \widehat{X}_{1-s}] \right\|^2 \right] s^{7/8} \\
&\lesssim \mathbb{E} \left[\left\| \mathbb{E} \left[\frac{\|\overleftarrow{f}_{1-s}^s\|^2 + 2\langle \overleftarrow{f}_{1-s}^s, \overleftarrow{g}_{1-s}^s \rangle}{s^2} (\widehat{X}_0 - \widehat{X}_{1-s}) \middle| \widehat{X}_{1-s} \right] \right\|^2 \right] s^{7/8} \\
&\quad + \mathbb{E} \left[\left\| \mathbb{E} \left[\frac{\|\overleftarrow{f}_{1-s}^s\|^2 + 2\langle \overleftarrow{f}_{1-s}^s, \overleftarrow{g}_{1-s}^s \rangle}{s^2} \middle| \widehat{X}_{1-s} \right] \mathbb{E} [\widehat{X}_0 - \widehat{X}_{1-s} \middle| \widehat{X}_{1-s}] \right\|^2 \right] s^{7/8} \\
&\lesssim \mathbb{E} \left[\left\| \frac{\|\overleftarrow{f}_{1-s}^s\|^2 + 2\langle \overleftarrow{f}_{1-s}^s, \overleftarrow{g}_{1-s}^s \rangle}{s^2} (\widehat{X}_0 - \widehat{X}_{1-s}) \right\|^2 \right] s^{7/8} \\
&\quad + \mathbb{E} \left[\left\| \mathbb{E} \left[\frac{\|\overleftarrow{f}_{1-s}^s\|^2 + 2\langle \overleftarrow{f}_{1-s}^s, \overleftarrow{g}_{1-s}^s \rangle}{s^2} \middle| \widehat{X}_{1-s} \right] \right\|^4 \right] s^{7/4} + \mathbb{E} \left[\left\| \mathbb{E} [\widehat{X}_0 - \widehat{X}_{1-s} \middle| \widehat{X}_{1-s}] \right\|^4 \right] \\
&\lesssim \mathbb{E} \left[\left\| \frac{\|\overleftarrow{f}_{1-s}^s\|^2 + 2\langle \overleftarrow{f}_{1-s}^s, \overleftarrow{g}_{1-s}^s \rangle}{s^2} \right\|^4 \right] s^{7/4} + \mathbb{E} \left[\|\widehat{X}_0 - \widehat{X}_{1-s}\|^4 \right] \\
&\lesssim \mathbb{E} \left[\frac{\|\overleftarrow{f}_{1-s}^s\|^8}{s^8} \right] + \mathbb{E} \left[\frac{\langle \overleftarrow{f}_{1-s}^s, \overleftarrow{g}_{1-s}^s \rangle^4}{s^8} \right] s^{7/4} + \mathbb{E} \left[\|\widehat{X}_0 - \widehat{X}_{1-s}\|^4 \right]
\end{aligned}$$

$$\begin{aligned}
&\lesssim \mathbb{E} \left[\frac{\|\overleftarrow{f}_{1-s}^s\|^8}{s^8} \right] + \mathbb{E} \left[\frac{\|\overleftarrow{g}_{1-s}^s\|^8}{s^8} \right] s^{7/2} + \mathbb{E} \left[\|\overleftarrow{X}_0 - \overleftarrow{X}_{1-s}\|^4 \right] \\
&\lesssim \mathbb{E} \left[\frac{\|\overleftarrow{f}_{1-s}^s\|^8}{s^8} \right] + d^4 s^{-1/2} + \mathbb{E} \left[\|X_1^I - X_s^I\|^4 \right] \\
&\lesssim \|\nabla \log \tilde{\pi}\|_{L^s(\pi)}^8 + \|\nabla \log \nu^*\|_{L^s(\nu^*)}^8 + d^4 s^{-1/2} + d^2 + \mathbf{m}_4[\mu] + \mathbf{m}_4[\nu^*].
\end{aligned}$$

But then, we obtain that

$$\begin{aligned}
\int_0^{1/2} \int_{\mathbb{R}^d} \|A_s^5(x)\|^2 \rho(s) p_s^I(x) dx ds &\lesssim d^4 + \mathbf{m}_4[\mu] + \mathbf{m}_4[\nu^*] + \|\nabla \log \tilde{\pi}\|_{L^s(\pi)}^8 \\
&\quad + \|\nabla \log \nu^*\|_{L^s(\nu^*)}^8.
\end{aligned}$$

We now focus on the second time interval, that is we look at $s \in [1/2, 1 - \epsilon]$ and try to bound the integrand. Using (32) and proceeding in a similar way, we get that

$$\begin{aligned}
&\mathbb{E} \left[\left\| \mathbb{E} \left[\frac{\|X_s^I - X_0^I\|^2}{s^2} \frac{X_1^I - X_s^I}{1-s} \middle| X_s^I \right] - \mathbb{E} \left[\frac{\|X_s^I - X_0^I\|^2}{s^2} \middle| X_s^I \right] \mathbb{E} \left[\frac{X_1^I - X_s^I}{1-s} \middle| X_s^I \right] \right\|^2 \right] \rho(s) \\
&\lesssim \mathbb{E} \left[\left\| \mathbb{E} \left[\|\overrightarrow{X}_s - \overrightarrow{X}_0\|^2 \frac{\overrightarrow{f}_s^{1-s} + \overrightarrow{g}_s^{1-s}}{1-s} \middle| \overrightarrow{X}_s \right] \right. \right. \\
&\quad \left. \left. - \mathbb{E} \left[\|\overrightarrow{X}_s - \overrightarrow{X}_0\|^2 \middle| \overrightarrow{X}_s \right] \mathbb{E} \left[\frac{\overrightarrow{f}_s^{1-s} + \overrightarrow{g}_s^{1-s}}{1-s} \middle| \overrightarrow{X}_s \right] \right\|^2 \right] \\
&= \mathbb{E} \left[\left\| \mathbb{E} \left[\|\overrightarrow{X}_s - \overrightarrow{X}_0\|^2 \frac{\overrightarrow{f}_s^{1-s}}{1-s} \middle| \overrightarrow{X}_s \right] - \mathbb{E} \left[\|\overrightarrow{X}_s - \overrightarrow{X}_0\|^2 \middle| \overrightarrow{X}_s \right] \mathbb{E} \left[\frac{\overrightarrow{f}_s^{1-s}}{1-s} \middle| \overrightarrow{X}_s \right] \right\|^2 \right] \\
&\lesssim \mathbb{E} \left[\|\overrightarrow{X}_s - \overrightarrow{X}_0\|^8 \right] + \mathbb{E} \left[\frac{\|\overrightarrow{f}_s^{1-s}\|^4}{(1-s)^4} \right] = \mathbb{E} \left[\|X_s^I - X_0^I\|^8 \right] + \mathbb{E} \left[\frac{\|\overrightarrow{f}_s^{1-s}\|^4}{(1-s)^4} \right] \\
&\lesssim d^4 + \mathbf{m}_8[\mu] + \mathbf{m}_8[\nu^*] + \|\nabla \log \tilde{\pi}\|_{L^4(\pi)}^4 + \|\nabla \log \mu\|_{L^4(\mu)}^4.
\end{aligned}$$

But then, we have that

$$\begin{aligned}
\int_{1/2}^{1-\epsilon} \int_{\mathbb{R}^d} \|A_s^5(x)\|^2 \rho(s) p_s^I(x) dx ds &\lesssim d^4 + \mathbf{m}_8[\mu] + \mathbf{m}_8[\nu^*] + \|\nabla \log \tilde{\pi}\|_{L^s(\pi)}^8 \\
&\quad + \|\nabla \log \mu\|_{L^s(\mu)}^8.
\end{aligned}$$

To conclude, we can bound A_s^5 as follows

$$\begin{aligned}
\int_0^{1-\epsilon} \int_{\mathbb{R}^d} \|A_s^5(x)\|^2 \rho(s) p_s^I(x) dx ds &\lesssim d^4 + \mathbf{m}_8[\mu] + \mathbf{m}_8[\nu^*] + \|\nabla \log \tilde{\pi}\|_{L^s(\pi)}^8 \\
&\quad + \|\nabla \log \mu\|_{L^s(\mu)}^8 + \|\nabla \log \nu^*\|_{L^s(\nu^*)}^8.
\end{aligned}$$

We are left with A_s^6 . Using (19), we can rewrite A_s^6 as follows

$$\begin{aligned}
 & A_s^6(x) \\
 &= \frac{1}{p_s^I(x)} \int_{\mathbb{R}^{2d}} \left\{ \frac{-1}{2(1-s)} \text{Id} + \frac{(x_1-x)(x_1-x)^T}{4(1-s)^2} \right\} \frac{x-x_0}{2s} p_s(x|x_0) p_{1-s}(x_1|x) \tilde{\pi}(x_0, x_1) dx_0 dx_1 \\
 &\quad - \left(\frac{1}{p_s^I(x)} \int_{\mathbb{R}^{2d}} \left\{ \frac{-1}{2(1-s)} \text{Id} + \frac{(x_1-x)(x_1-x)^T}{4(1-s)^2} \right\} p_s(x|x_0) p_{1-s}(x_1|x) \tilde{\pi}(x_0, x_1) dx_0 dx_1 \right) \\
 &\quad \cdot \left(\frac{1}{p_s^I(x)} \int_{\mathbb{R}^{2d}} \frac{x-x_0}{2s} p_s(x|x_0) p_{1-s}(x_1|x) \tilde{\pi}(x_0, x_1) dx_0 dx_1 \right) \\
 &\lesssim \left(\frac{1}{p_s^I(x)} \int_{\mathbb{R}^{2d}} \frac{(x_1-x)(x_1-x)^T}{(1-s)^2} \frac{x-x_0}{s} p_s(x|x_0) p_{1-s}(x_1|x) \tilde{\pi}(x_0, x_1) dx_0 dx_1 \right. \\
 &\quad \left. - \left(\frac{1}{p_s^I(x)} \int_{\mathbb{R}^{2d}} \frac{(x_1-x)(x_1-x)^T}{(1-s)^2} p_s(x|x_0) p_{1-s}(x_1|x) \tilde{\pi}(x_0, x_1) dx_0 dx_1 \right) \right. \\
 &\quad \left. \cdot \left(\frac{1}{p_s^I(x)} \int_{\mathbb{R}^{2d}} \frac{x-x_0}{s} p_s(x|x_0) p_{1-s}(x_1|x) \tilde{\pi}(x_0, x_1) dx_0 dx_1 \right) \right).
 \end{aligned}$$

It follows that

$$\begin{aligned}
 & \int_0^{1-\epsilon} \int_{\mathbb{R}^d} \|A_s^6(x)\|^2 \rho(s) p_s^I(x) dx ds \\
 &\lesssim \int_0^{1-\epsilon} \mathbb{E} \left[\left\| \mathbb{E} \left[\frac{(X_1^I - X_s^I)(X_1^I - X_s^I)^T}{(1-s)^2} \frac{X_s^I - X_0^I}{s} \middle| X_s^I \right] \right. \right. \\
 &\quad \left. \left. - \mathbb{E} \left[\frac{(X_1^I - X_s^I)(X_1^I - X_s^I)^T}{(1-s)^2} \middle| X_s^I \right] \mathbb{E} \left[\frac{X_s^I - X_0^I}{s} \middle| X_s^I \right] \right\|^2 \right] \rho(s) ds
 \end{aligned}$$

At this point, we proceed as for A_s^5 , that is we split the time interval in two and we use Lemma 1, Lemma 2 and Lemma 3. By doing so and by using (32), we get that for $s \in [0, 1/2]$

$$\begin{aligned}
 & \mathbb{E} \left[\left\| \mathbb{E} \left[\frac{(X_1^I - X_s^I)(X_1^I - X_s^I)^T}{(1-s)^2} \frac{X_s^I - X_0^I}{s} \middle| X_s^I \right] \right. \right. \\
 &\quad \left. \left. - \mathbb{E} \left[\frac{(X_1^I - X_s^I)(X_1^I - X_s^I)^T}{(1-s)^2} \middle| X_s^I \right] \mathbb{E} \left[\frac{X_s^I - X_0^I}{s} \middle| X_s^I \right] \right\|^2 \right] \rho(s) \\
 &\lesssim \mathbb{E} \left[\left\| \mathbb{E} \left[(\hat{X}_0 - \hat{X}_{1-s})(\hat{X}_0 - \hat{X}_{1-s})^T \frac{\hat{X}_1 - \hat{X}_{1-s}}{s} \middle| \hat{X}_{1-s} \right] \right. \right. \\
 &\quad \left. \left. - \mathbb{E} \left[(\hat{X}_0 - \hat{X}_{1-s})(\hat{X}_0 - \hat{X}_{1-s})^T \middle| \hat{X}_{1-s} \right] \mathbb{E} \left[\frac{\hat{X}_1 - \hat{X}_{1-s}}{s} \middle| \hat{X}_{1-s} \right] \right\|^2 \right] \\
 &= \mathbb{E} \left[\left\| \mathbb{E} \left[(\hat{X}_0 - \hat{X}_{1-s})(\hat{X}_0 - \hat{X}_{1-s})^T \frac{\overleftarrow{f}_{1-s}^s}{s} \middle| \hat{X}_{1-s} \right] \right. \right. \\
 &\quad \left. \left. - \mathbb{E} \left[(\hat{X}_0 - \hat{X}_{1-s})(\hat{X}_0 - \hat{X}_{1-s})^T \middle| \hat{X}_{1-s} \right] \mathbb{E} \left[\frac{\overleftarrow{f}_{1-s}^s}{s} \middle| \hat{X}_{1-s} \right] \right\|^2 \right] \\
 &\lesssim \mathbb{E} \left[\left\| (\hat{X}_0 - \hat{X}_{1-s})(\hat{X}_0 - \hat{X}_{1-s})^T \right\|_{\text{op}}^4 \right] + \mathbb{E} \left[\left\| \frac{\overleftarrow{f}_{1-s}^s}{s} \right\|^4 \right] \\
 &= \mathbb{E} [\|X_1^I - X_s^I\|^8] + \mathbb{E} \left[\left\| \frac{\overleftarrow{f}_{1-s}^s}{s} \right\|^4 \right] \\
 &\lesssim d^4 + \mathbf{m}_s[\mu] + \mathbf{m}_s[\nu^*] + \|\nabla \log \tilde{\pi}\|_{L^s(\pi)}^8 + \|\nabla \log \nu^*\|_{L^s(\nu^*)}^8.
 \end{aligned}$$

Whereas, for $s \in [1/2, 1 - \epsilon]$, we get

$$\begin{aligned}
& \mathbb{E} \left[\left\| \mathbb{E} \left[\frac{(X_1^I - X_s^I)(X_1^I - X_s^I)^T}{(1-s)^2} \frac{X_s^I - X_0^I}{s} \middle| X_s^I \right] \right\|^2 \right] \\
& - \mathbb{E} \left[\frac{(X_1^I - X_s^I)(X_1^I - X_s^I)^T}{(1-s)^2} \middle| X_s^I \right] \mathbb{E} \left[\frac{X_s^I - X_0^I}{s} \middle| X_s^I \right] \left\| \right\|^2 \rho(s) \\
& \lesssim \mathbb{E} \left[\left\| \mathbb{E} \left[\frac{(\vec{X}_1 - \vec{X}_s)(\vec{X}_1 - \vec{X}_s)^T}{(1-s)^2} (\vec{X}_s - \vec{X}_0) \middle| \vec{X}_s \right] \right\|^2 \right] \\
& - \mathbb{E} \left[\frac{(\vec{X}_1 - \vec{X}_s)(\vec{X}_1 - \vec{X}_s)^T}{(1-s)^2} \middle| \vec{X}_s \right] \mathbb{E} \left[\vec{X}_s - \vec{X}_0 \middle| \vec{X}_s \right] \left\| \right\|^2 (1-s)^{7/8} \\
& = (1-s)^{7/8} \mathbb{E} \left[\left\| \mathbb{E} \left[\frac{\vec{f}_s^{1-s}(\vec{f}_s^{1-s})^T + \vec{f}_s^{1-s}(\vec{g}_s^{1-s})^T + \vec{g}_s^{1-s}(\vec{f}_s^{1-s})^T}{(1-s)^2} (\vec{X}_s - \vec{X}_0) \middle| \vec{X}_s \right] \right\|^2 \right] \\
& - \mathbb{E} \left[\frac{\vec{f}_s^{1-s}(\vec{f}_s^{1-s})^T + \vec{f}_s^{1-s}(\vec{g}_s^{1-s})^T + \vec{g}_s^{1-s}(\vec{f}_s^{1-s})^T}{(1-s)^2} \middle| \vec{X}_s \right] \mathbb{E} \left[\vec{X}_s - \vec{X}_0 \middle| \vec{X}_s \right] \left\| \right\|^2 \\
& \lesssim \mathbb{E} \left[\frac{\left\| \vec{f}_s^{1-s}(\vec{f}_s^{1-s})^T + \vec{f}_s^{1-s}(\vec{g}_s^{1-s})^T + \vec{g}_s^{1-s}(\vec{f}_s^{1-s})^T \right\|_{\text{op}}^4}{(1-s)^8} \right] (1-s)^{7/4} + \mathbb{E} \left[\left\| \vec{X}_s - \vec{X}_0 \right\|^4 \right] \\
& \lesssim \mathbb{E} \left[\frac{\left\| \vec{f}_s^{1-s} \right\|^8}{(1-s)^8} \right] + \mathbb{E} \left[\frac{\left\| \vec{g}_s^{1-s} \right\|^8}{(1-s)^8} \right] (1-s)^{7/2} + \mathbb{E} \left[\left\| X_s^I - X_0^I \right\|^4 \right] \\
& \lesssim \|\nabla \log \tilde{\pi}\|_{L^8(\pi)}^8 + \|\nabla \log \mu\|_{L^8(\mu)}^8 + d^4(1-s)^{-1/2} + d^2 + \mathbf{m}_4[\mu] + \mathbf{m}_4[\nu^*].
\end{aligned}$$

Consequently, we have that

$$\begin{aligned}
\int_0^{1-\epsilon} \int_{\mathbb{R}^d} \|A_s^6(x)\|^2 \rho(s) p_s^I(x) dx ds & \lesssim d^4 + \mathbf{m}_8[\mu] + \mathbf{m}_8[\nu^*] + \|\nabla \log \tilde{\pi}\|_{L^8(\pi)}^8 \\
& + \|\nabla \log \mu\|_{L^8(\mu)}^8 + \|\nabla \log \nu^*\|_{L^8(\nu^*)}^8.
\end{aligned}$$

Putting together the bounds on the $\{A_s^k\}_{k=1}^6$ derived so far, we eventually obtain

$$\begin{aligned}
& \int_0^{1-\epsilon} \mathbb{E} \left[\left\| (\partial_s + \mathcal{L}_s^M) \tilde{\beta}_s(\vec{X}_s) \right\|^2 \right] \rho(s) ds \\
& \lesssim d^4 + \mathbf{m}_8[\mu] + \mathbf{m}_8[\nu^*] + \|\nabla \log \tilde{\pi}\|_{L^8(\pi)}^8 + \|\nabla \log \mu\|_{L^8(\mu)}^8 + \|\nabla \log \nu^*\|_{L^8(\nu^*)}^8.
\end{aligned} \tag{41}$$

Plugging (41), (35) and (36) into (34), we get

$$\begin{aligned}
\text{KL}(\nu_{1-\epsilon}^* \nu_{1-\epsilon}^{\theta^*}) & \lesssim \varepsilon^2 + h(h^{1/8} + 1) \left(d^4 + \mathbf{m}_8[\mu] + \mathbf{m}_8[\nu^*] + \|\nabla \log \tilde{\pi}\|_{L^8(\pi)}^8 \right. \\
& \left. + \|\nabla \log \mu\|_{L^8(\mu)}^8 + \|\nabla \log \nu^*\|_{L^8(\nu^*)}^8 \right).
\end{aligned}$$

The estimate (14) then follows from the above estimate and (31).

However, for (14) to hold true, we still need to prove that

Lemma 4. $(\int_0^s \langle \tilde{\beta}_t, D_x \tilde{\beta}_t \rangle (X_t^M) dB_t)_{s \in [0, 1-\epsilon]}$ is a martingale.

Proof of Lemma 4. If we show that for any $s \in [0, 1 - \epsilon]$, it holds $\mathbb{E}[\|\langle \tilde{\beta}_s, D_x \tilde{\beta}_s \rangle (X_s^M)\|^2] < C$, for some $C > 0$ independent of time, then by Fubini's theorem and [BB17, Theorem 7.3] we are done. To do so, by the Cauchy-Schwarz inequality, we just need to show that $\mathbb{E}[\|\tilde{\beta}_s(X_s^M)\|^4]$ and

$\mathbb{E}[\|D_x \tilde{\beta}_s(X_s^M)\|_{\text{op}}^4]$ are bounded from above by constants which are independent of time. To this aim, note that, as a direct consequence of (8), Theorem 1, (18), Jensen inequality and **H2**, it holds

$$\begin{aligned}
& \mathbb{E} \left[\left\| \tilde{\beta}_s(X_s^M) \right\|^4 \right] \\
& \lesssim \int_{\mathbb{R}^d} \left\| \frac{\int_{\mathbb{R}^{2d}} p_s(x|x_0) \nabla_x p_{1-s}(x_1|x) \tilde{\pi}(x_0, x_1) dx_0 dx_1}{p_s^I(x)} \right\|^4 p_s^I(x) dx \\
& = \int_{\mathbb{R}^d} \left\| \frac{1}{p_s^I(x)} \int_{\mathbb{R}^{2d}} p_s(x|x_0) p_{1-s}(x_1|x) \frac{\nabla_{x_1} \tilde{\pi}}{\tilde{\pi}}(x_0, x_1) \tilde{\pi}(x_0, x_1) dx_0, dx_1 \right\|^4 p_s^I(x) dx \quad (42) \\
& = \mathbb{E} \left[\left\| \mathbb{E} \left[\frac{\nabla_{x_1} \tilde{\pi}}{\tilde{\pi}}(X_0^I, X_1^I) \middle| X_s^I \right] \right\|^4 \right] \leq \mathbb{E} \left[\left\| \frac{\nabla_{x_1} \tilde{\pi}}{\tilde{\pi}}(X_0^I, X_1^I) \right\|^4 \middle| X_s^I \right] \\
& = \mathbb{E} \left[\left\| \frac{\nabla_{x_1} \tilde{\pi}}{\tilde{\pi}}(X_0^I, X_1^I) \right\|^4 \right] = \|\nabla \log \tilde{\pi}\|_{L^4(\pi)}^4.
\end{aligned}$$

Similarly, recalling (38), we have that

$$\begin{aligned}
& \mathbb{E} \left[\left\| D_x \tilde{\beta}_s(X_s^M) \right\|_{\text{op}}^4 \right] \\
& \lesssim \int_{\mathbb{R}^d} \left\| \frac{\int_{\mathbb{R}^{2d}} \nabla_x p_{1-s}(x_1|x) (\nabla_x p_s(x|x_0))^T \tilde{\pi}(x_0, x_1) dx_0 dx_1}{p_s^I(x)} \right\|_{\text{op}}^4 p_s^I(x) dx \\
& + \int_{\mathbb{R}^d} \left\| \frac{\int_{\mathbb{R}^{2d}} p_s(x|x_0) \nabla_x^2 p_{1-s}(x_1|x) \tilde{\pi}(x_0, x_1) dx_0 dx_1}{p_s^I(x)} \right\|_{\text{op}}^4 p_s^I(x) dx \\
& + \int_{\mathbb{R}^d} \left\| \frac{\int_{\mathbb{R}^{2d}} p_s(x|x_0) \nabla_x p_{1-s}(x_1|x) \tilde{\pi}(x_0, x_1) dx_0 dx_1}{p_s^I(x)} \right\|^8 p_s^I(x) dx \\
& + \int_{\mathbb{R}^d} \left\| \frac{\int_{\mathbb{R}^{2d}} \nabla_x p_s(x|x_0) p_{1-s}(x_1|x) \tilde{\pi}(x_0, x_1) dx_0 dx_1}{p_s^I(x)} \right\|^8 p_s^I(x) dx.
\end{aligned}$$

To bound the first term of the RHS of the above expression, we integrate by parts and use Lemma 1 and **H2**.

$$\begin{aligned}
& \int_{\mathbb{R}^d} \left\| \frac{\int_{\mathbb{R}^{2d}} \nabla_x p_{1-s}(x_1|x) (\nabla_x p_s(x|x_0))^T \tilde{\pi}(x_0, x_1) dx_0 dx_1}{p_s^I(x)} \right\|_{\text{op}}^4 p_s^I(x) dx \\
& = \int_{\mathbb{R}^d} \left\| \frac{\int_{\mathbb{R}^{2d}} \nabla_x p_s(x|x_0) (\nabla_x p_{1-s}(x_1|x))^T \tilde{\pi}(x_0, x_1) dx_0 dx_1}{p_s^I(x)} \right\|_{\text{op}}^4 p_s^I(x) dx \\
& = \int_{\mathbb{R}^d} \left\| \frac{1}{p_s^I(x)} \int_{\mathbb{R}^{2d}} \frac{\nabla_{x_0} \tilde{\pi}}{\tilde{\pi}}(x_0, x_1) \frac{(x_1 - x)^T}{1-s} p_s(x|x_0) p_{1-s}(x_1|x) \tilde{\pi}(x_0, x_1) dx_0 dx_1 \right\|_{\text{op}}^4 p_s^I(x) dx \\
& \lesssim \mathbb{E} \left[\left\| \frac{\nabla_{x_0} \tilde{\pi}}{\tilde{\pi}}(X_0^I, X_1^I) \frac{(X_1^I - X_s^I)^T}{\epsilon} \right\|_{\text{op}}^4 \right] \\
& \lesssim \frac{1}{\epsilon} (d^4 + \mathbf{m}_8[\mu] + \mathbf{m}_8[\nu^*]) + \|\nabla \log \tilde{\pi}\|_{L^8(\pi)}^8.
\end{aligned}$$

To bound the second term, we integrate by parts and proceed as before.

$$\begin{aligned}
 & \int_{\mathbb{R}^d} \left\| \frac{\int_{\mathbb{R}^{2d}} p_s(x|x_0) \nabla_x^2 p_{1-s}(x_1|x) \tilde{\pi}(x_0, x_1) dx_0 dx_1}{p_s^I(x)} \right\|_{\text{op}}^4 p_s^I(x) dx \\
 &= \int_{\mathbb{R}^d} \left\| \frac{\int_{\mathbb{R}^{2d}} p_s(x|x_0) (\nabla_{x_1} \tilde{\pi}(x_0, x_1) / \tilde{\pi}(x_0, x_1)) (\nabla_x p_{1-s}(x_1|x))^T \tilde{\pi}(x_0, x_1) dx_0 dx_1}{p_s^I(x)} \right\|_{\text{op}}^4 p_s^I(x) dx \\
 &= \int_{\mathbb{R}^d} \left\| \frac{1}{p_s^I(x)} \int_{\mathbb{R}^{2d}} p_s(x|x_0) p_{1-s}(x_1|x) \frac{\nabla_{x_1} \tilde{\pi}}{\tilde{\pi}}(x_0, x_1) \frac{(x-x_1)^T}{1-s} \tilde{\pi}(x_0, x_1) dx_0 dx_1 \right\|_{\text{op}}^4 p_s^I(x) dx \\
 &\lesssim \frac{1}{\epsilon} (d^4 + \mathbf{m}_s[\mu] + \mathbf{m}_s[\nu^*]) + \|\nabla \log \tilde{\pi}\|_{L^s(\pi)}^8.
 \end{aligned}$$

To bound the third and last term, we proceed as in (42), getting

$$\int_{\mathbb{R}^d} \left\| \frac{\int_{\mathbb{R}^{2d}} p_s(x|x_0) \nabla_x p_{1-s}(x_1|x) \tilde{\pi}(x_0, x_1) dx_0 dx_1}{p_s^I(x)} \right\|_{\text{op}}^8 p_s^I(x) dx \lesssim \|\nabla \log \tilde{\pi}\|_{L^s(\pi)}^8,$$

and

$$\int_{\mathbb{R}^d} \left\| \frac{\int_{\mathbb{R}^{2d}} \nabla_x p_s(x|x_0) p_{1-s}(x_1|x) \tilde{\pi}(x_0, x_1) dx_0 dx_1}{p_s^I(x)} \right\|_{\text{op}}^8 p_s^I(x) dx \lesssim \|\nabla \log \tilde{\pi}\|_{L^s(\pi)}^8.$$

□

A.5 Proof of Theorem 3

Fix $\delta > 0$. Then, because of Theorem 1, (5) and H1, it holds

$$\mathbf{m}_s[\nu_{1-\delta}^*] = \mathbb{E}[\|X_{1-\delta}^I\|^8] \lesssim \delta^8 \mathbf{m}_s[\mu] + (1-\delta)^8 \mathbf{m}_s[\nu^*] + d^4 \delta^4 (1-\delta)^4 < +\infty. \quad (43)$$

Moreover $\nu_{1-\delta}^* \ll \text{Leb}^d$ with density $p_{1-\delta}^I$ and

$$\nabla \log \left(\frac{d\nu_{1-\delta}^*}{d\text{Leb}^d} \right) \in L^s(\nu_{1-\delta}^*). \quad (44)$$

Indeed because of (18), it holds

$$\begin{aligned}
 & \nabla \log \frac{d\nu_{1-\delta}^*}{d\text{Leb}^d}(x_{1-\delta}) \\
 &= \frac{\nabla \left(\int_{\mathbb{R}^{2d}} p_{1-\delta}(x_{1-\delta}|x_0) p_\delta(x_1|x_{1-\delta}) \tilde{\pi}(dx_0, dx_1) \right)}{p_{1-\delta}^I(x_{1-\delta})} \\
 &= \frac{1}{p_{1-\delta}^I(x_{1-\delta})} \int_{\mathbb{R}^{2d}} \left\{ \frac{x_{1-\delta} - x_0}{1-\delta} + \frac{x_1 - x_{1-\delta}}{\delta} \right\} p_{1-\delta}(x_{1-\delta}|x_0) p_\delta(x_1|x_{1-\delta}) \tilde{\pi}(dx_0, dx_1) \\
 &= \frac{1}{p_{1-\delta}^I(x_{1-\delta})} \int_{\mathbb{R}^{2d}} \frac{(2\delta-1)x_{1-\delta} - \delta x_0 + (1-\delta)x_1}{\delta(1-\delta)} p_{1-\delta}(x_{1-\delta}|x_0) p_\delta(x_1|x_{1-\delta}) \tilde{\pi}(dx_0, dx_1) \\
 &\lesssim \mathbb{E} \left[\frac{X_{1-\delta}^I - \delta X_0^I + (1-\delta)X_1^I}{\delta(1-\delta)} \middle| X_{1-\delta}^I = x_{1-\delta} \right].
 \end{aligned}$$

But then, using Jensen inequality we obtain that

$$\begin{aligned} \int_{\mathbb{R}^d} \left\| \nabla_{x_{1-\delta}} \log \frac{d\nu_{1-\delta}^*}{d\text{Leb}^d} \right\|^8 d\nu_{1-\delta}^* &= \mathbb{E} \left[\left\| \mathbb{E} \left[\frac{X_{1-\delta}^I - \delta X_0^I + (1-\delta)X_1^I}{\delta(1-\delta)} \middle| X_{1-\delta}^I \right] \right\|^8 \right] \\ &\lesssim \mathbb{E} \left[\left\| \frac{X_{1-\delta}^I - \delta X_0^I + (1-\delta)X_1^I}{\delta(1-\delta)} \right\|^8 \right] \\ &\lesssim \mathbf{m}_8[\nu_{1-\delta}^*] \frac{1}{\delta^8(1-\delta)^8} + \mathbf{m}_8[\mu] \frac{1}{(1-\delta)^8} + \mathbf{m}_8[\nu^*] \frac{1}{\delta^8} \\ &\lesssim \mathbf{m}_8[\mu] \frac{1}{(1-\delta)^8} + \mathbf{m}_8[\nu^*] \frac{1}{\delta^8} + d^4 \frac{1}{\delta^4(1-\delta)^4}. \end{aligned}$$

Also, consider

$$\pi_{1-\delta}(x_0, x_{1-\delta}) = \mu(x_0) \int_{\mathbb{R}^d} p_{1-\delta|0,1}^I(x_{1-\delta}|x_0, x_1) \nu^*(dx_1),$$

where $(x_0, x_1, x_{1-\delta}) \mapsto p_{1-\delta|0,1}^I(x_{1-\delta}|x_0, x_1)$ denotes the density of $X_{1-\delta}^I$ given (X_0^I, X_1^I) with respect to the Lebesgue measure. Then $\pi_{1-\delta} \in \Pi(\mu, \nu_{1-\delta}^*)$ and $\pi_{1-\delta} \ll \text{Leb}^{2d}$. Moreover

$$\nabla \log \left(\frac{1}{p_{1-\delta}} \frac{d\pi_{1-\delta}}{d\text{Leb}^{2d}} \right) \in L^8(\pi_{1-\delta}). \quad (45)$$

Indeed, because of (5),

$$p_{1-\delta|0,1}^I(x_{1-\delta}|x_0, x_1) = \frac{1}{(4\pi\delta(1-\delta))^{d/2}} \exp \left(- \frac{\|x_{1-\delta} - \delta x_0 - (1-\delta)x_1\|^2}{4\delta(1-\delta)} \right).$$

Therefore

$$\nabla_{x_0} p_{1-\delta|0,1}^I(x_{1-\delta}|x_0, x_1) = \frac{x_{1-\delta} - \delta x_0 - (1-\delta)x_1}{2(1-\delta)} p_{1-\delta|0,1}^I(x_{1-\delta}|x_0, x_1),$$

and

$$\nabla_{x_{1-\delta}} p_{1-\delta|0,1}^I(x_{1-\delta}|x_0, x_1) = - \frac{x_{1-\delta} - \delta x_0 - (1-\delta)x_1}{2\delta(1-\delta)} p_{1-\delta|0,1}^I(x_{1-\delta}|x_0, x_1).$$

Furthermore

$$\begin{aligned} \frac{p_{1-\delta|0,1}^I(x_{1-\delta}|x_0, x_1) \nu^*(dx_1)}{\int_{\mathbb{R}^d} p_{1-\delta|0,1}^I(x_{1-\delta}|x_0, \tilde{x}_1) \nu^*(d\tilde{x}_1)} &= \frac{p_{1-\delta|0,1}^I(x_{1-\delta}|x_0, x_1) \mu(x_0) \nu^*(dx_1)}{\int_{\mathbb{R}^d} p_{1-\delta|0,1}^I(x_{1-\delta}|x_0, \tilde{x}_1) \mu(x_0) \nu^*(d\tilde{x}_1)} \\ &= p_{1|0,1-\delta}^I(x_1|x_0, x_{1-\delta}) dx_1. \end{aligned}$$

Consequently, we have that

$$\begin{aligned} &\frac{\nabla_{x_0} \pi_{1-\delta}}{\pi_{1-\delta}}(x_0, x_{1-\delta}) \\ &= \frac{\nabla \mu(x_0) \int_{\mathbb{R}^d} p_{1-\delta|0,1}^I(x_{1-\delta}|x_0, x_1) \nu^*(dx_1) + \mu(x_0) \int_{\mathbb{R}^d} \nabla_{x_0} p_{1-\delta|0,1}^I(x_{1-\delta}|x_0, x_1) \nu^*(dx_1)}{\mu(x_0) \int_{\mathbb{R}^d} p_{1-\delta|0,1}^I(x_{1-\delta}|x_0, \tilde{x}_1) \nu^*(d\tilde{x}_1)} \\ &= \frac{\nabla \mu(x_0)}{\mu(x_0)} + \int_{\mathbb{R}^d} \frac{x_{1-\delta} - \delta x_0 - (1-\delta)x_1}{2(1-\delta)} p_{1|0,1-\delta}^I(x_1|x_0, x_{1-\delta}) dx_1, \end{aligned}$$

and that

$$\begin{aligned} \frac{\nabla_{x_{1-\delta}} \pi_{1-\delta}}{\pi_{1-\delta}}(x_0, x_{1-\delta}) &= \frac{\mu(x_0) \int_{\mathbb{R}^d} \nabla_{x_{1-\delta}} p_{1-\delta|0,1}^I(x_{1-\delta}|x_0, x_1) \nu^*(dx_1)}{\mu(x_0) \int_{\mathbb{R}^d} p_{1-\delta|0,1}^I(x_{1-\delta}|x_0, \tilde{x}_1) \nu^*(d\tilde{x}_1)} \\ &= - \int_{\mathbb{R}^d} \frac{x_{1-\delta} - \delta x_0 - (1-\delta)x_1}{2\delta(1-\delta)} p_{1|0,1-\delta}^I(x_1|x_0, x_{1-\delta}) dx_1. \end{aligned}$$

But then, if we use Jensen inequality and (43), we get

$$\begin{aligned}
 \int_{\mathbb{R}^{2d}} \left\| \nabla_{x_0} \log \frac{d\pi_{1-\delta}}{d\text{Leb}^{2d}} \right\|^8 d\pi_{1-\delta} &\lesssim \int_{\mathbb{R}^d} \left\| \nabla_{x_0} \log \frac{d\mu}{d\text{Leb}^d} \right\|^8 d\mu \\
 &\quad + \mathbb{E} \left[\left\| \mathbb{E} \left[\frac{X_{1-\delta}^I - \delta X_0^I - (1-\delta)X_1^I}{1-\delta} \middle| (X_0^I, X_{1-\delta}^I) \right] \right\|^8 \right] \\
 &\lesssim \|\nabla \log \mu\|_{L^s(\mu)}^8 + \mathbb{E} \left[\left\| \frac{X_{1-\delta}^I - \delta X_0^I - (1-\delta)X_1^I}{1-\delta} \right\|^8 \right] \\
 &\lesssim \|\nabla \log \mu\|_{L^s(\mu)}^8 + \mathbf{m}_8[\nu_{1-\delta}^*] \frac{1}{(1-\delta)^8} + \mathbf{m}_8[\mu] \frac{\delta^8}{(1-\delta)^8} + \mathbf{m}_8[\nu^*] \\
 &\lesssim \|\nabla \log \mu\|_{L^s(\mu)}^8 + \mathbf{m}_8[\mu] \frac{\delta^8}{(1-\delta)^8} + \mathbf{m}_8[\nu^*] + d^4 \frac{\delta^4}{(1-\delta)^4} \\
 &\lesssim \|\nabla \log \mu\|_{L^s(\mu)}^8 + \mathbf{m}_8[\mu] \frac{1}{(1-\delta)^8} + \mathbf{m}_8[\nu^*] \frac{1}{\delta^8} + d^4 \frac{1}{\delta^4(1-\delta)^4},
 \end{aligned}$$

and (similarly)

$$\begin{aligned}
 \int_{\mathbb{R}^{2d}} \left\| \nabla_{x_{1-\delta}} \log \frac{d\pi_{1-\delta}}{d\text{Leb}^{2d}} \right\|^8 d\pi_{1-\delta} &\lesssim \mathbb{E} \left[\left\| \mathbb{E} \left[\frac{X_{1-\delta}^I - \delta X_0^I - (1-\delta)X_1^I}{\delta(1-\delta)} \middle| (X_0^I, X_{1-\delta}^I) \right] \right\|^8 \right] \\
 &\lesssim \mathbf{m}_8[\mu] \frac{1}{(1-\delta)^8} + \mathbf{m}_8[\nu^*] \frac{1}{\delta^8} + d^4 \frac{1}{\delta^4(1-\delta)^4}.
 \end{aligned}$$

Additionally, because of Remark 8 and (43), they hold

$$\begin{aligned}
 &\int_{\mathbb{R}^{2d}} \|\nabla_{x_0} \log p_{1-\delta}(x_{1-\delta}|x_0)\|^8 d\pi_{1-\delta}(x_0, x_{1-\delta}) \\
 &= \int_{\mathbb{R}^{2d}} \|\nabla_{x_{1-\delta}} \log p_{1-\delta}(x_{1-\delta}|x_0)\|^8 d\pi_{1-\delta}(x_0, x_{1-\delta}) \\
 &= \mathbb{E} \left[\left\| \frac{X_{1-\delta}^I - X_0^I}{1-\delta} \right\|^8 \right] \lesssim \mathbf{m}_8[\nu_{1-\delta}^*] \frac{1}{(1-\delta)^8} + \mathbf{m}_8[\mu] \frac{1}{(1-\delta)^8} \\
 &\lesssim \mathbf{m}_8[\mu] \frac{\delta^8}{(1-\delta)^8} + \mathbf{m}_8[\nu^*] + d^4 \frac{\delta^4}{(1-\delta)^4} + \mathbf{m}_8[\mu] \frac{1}{(1-\delta)^8} \\
 &\lesssim \mathbf{m}_8[\mu] \frac{1}{(1-\delta)^8} + \mathbf{m}_8[\nu^*] \frac{1}{\delta^8} + d^4 \frac{1}{\delta^4(1-\delta)^4}.
 \end{aligned}$$

It follows from **H1**, **H2(i)**, (43), (44) and (45) that the probability distributions $\mu, \nu_{1-\delta}^*$ and the coupling $\pi_{1-\delta} \in \Pi(\mu, \nu_{1-\delta}^*)$ satisfy **H1**. The bound in Theorem 3 is now a straightforward consequence of Theorem 2 and the bounds on the scores derived so far.

NeurIPS Paper Checklist

1. Claims

Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

Answer: [Yes]

Justification: The claims made in the abstract and introduction make clear the goals attained by the paper and match the results therein provided. Additionally, a discussion on the contributions made by the paper and the comparison with the literature is held.

Guidelines:

- The answer NA means that the abstract and introduction do not include the claims made in the paper.
- The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A No or NA answer to this question will not be perceived well by the reviewers.
- The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
- It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. Limitations

Question: Does the paper discuss the limitations of the work performed by the authors?

Answer: [Yes]

Justification: The paper discusses the limitations of the work performed in Section 4.

Guidelines:

- The answer NA means that the paper has no limitation while the answer No means that the paper has limitations, but those are not discussed in the paper.
- The authors are encouraged to create a separate "Limitations" section in their paper.
- The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
- The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
- The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
- The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
- If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
- While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. Theory Assumptions and Proofs

Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

Answer: [Yes]

Justification: The paper provides the full set of assumptions for either Theorem 2 and Theorem 3, see Section 3.1. Moreover, it includes the detailed proofs of all the stated theorems in Appendix A and a sketch of the proof of our main contributions Theorem 2, Theorem 3 in Section 3.3.

Guidelines:

- The answer NA means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. Experimental Result Reproducibility

Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

Answer: [NA]

Justification: Since this work is of theoretical nature, we do not provide experimental data, hence our answer.

Guidelines:

- The answer NA means that the paper does not include experiments.
- If the paper includes experiments, a No answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
- If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
- Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general, releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
- While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
 - (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
 - (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
 - (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
 - (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. Open access to data and code

Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

Answer: [NA]

Justification: Since this work is of theoretical nature, we do not provide experimental data, hence our answer.

Guidelines:

- The answer NA means that paper does not include experiments requiring code.
- Please see the NeurIPS code and data submission guidelines (<https://nips.cc/public/guides/CodeSubmissionPolicy>) for more details.
- While we encourage the release of code and data, we understand that this might not be possible, so “No” is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
- The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (<https://nips.cc/public/guides/CodeSubmissionPolicy>) for more details.
- The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
- The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
- At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. Experimental Setting/Details

Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

Answer: [NA]

Justification: Since this work is of theoretical nature, we do not provide experimental data, hence our answer.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
- The full details can be provided either with the code, in appendix, or as supplemental material.

7. Experiment Statistical Significance

Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

Answer: [NA]

Justification: Since this work is of theoretical nature, we do not provide experimental data, hence our answer.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.

- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g. negative error rates).
- If error bars are reported in tables or plots, The authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. Experiments Compute Resources

Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

Answer: [NA]

Justification: Since this work is of theoretical nature, we do not provide experimental data, hence our answer.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
- The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
- The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. Code Of Ethics

Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics <https://neurips.cc/public/EthicsGuidelines?>

Answer: [Yes]

Justification: We confirm that the research conducted our paper conform, in every respect, with the NeurIPS Code of Ethics.

Guidelines:

- The answer NA means that the authors have not reviewed the NeurIPS Code of Ethics.
- If the authors answer No, they should explain the special circumstances that require a deviation from the Code of Ethics.
- The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. Broader Impacts

Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

Answer: [NA]

Justification: Since this work is of theoretical nature, it does not present societal impacts up to our knowledge.

Guidelines:

- The answer NA means that there is no societal impact of the work performed.
- If the authors answer NA or No, they should explain why their work has no societal impact or why the paper does not address societal impact.
- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. Safeguards

Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pretrained language models, image generators, or scraped datasets)?

Answer: [NA]

Justification: Since this work is of theoretical nature, it does not present such risks.

Guidelines:

- The answer NA means that the paper poses no such risks.
- Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
- Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
- We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. Licenses for existing assets

Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

Answer: [NA]

Justification: In the present work, we do not use any assets to be credited.

Guidelines:

- The answer NA means that the paper does not use existing assets.
- The authors should cite the original paper that produced the code package or dataset.
- The authors should state which version of the asset is used and, if possible, include a URL.
- The name of the license (e.g., CC-BY 4.0) should be included for each asset.
- For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.

- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, paperswithcode.com/datasets has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New Assets**

Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

Answer: [NA]

Justification: In the present work, we do not introduce new assets.

Guidelines:

- The answer NA means that the paper does not release new assets.
- Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
- The paper should discuss whether and how consent was obtained from people whose asset is used.
- At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and Research with Human Subjects**

Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

Answer: [NA]

Justification: The present work does not include any experiment involving human subjects, hence our answer.

Guidelines:

- The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
- Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
- According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional Review Board (IRB) Approvals or Equivalent for Research with Human Subjects**

Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

Answer: [NA]

Justification: The present work does not include any experiment involving human subjects, hence our answer.

Guidelines:

- The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.

- Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.