Appendix A. Theoretical Result and Its Proof

A-1. Preliminaries

Let p^{src} and p^{tgt} be the source distribution and target distribution, respectively on $\mathbb{R}^d$ of dimension d. We consider the Ornstein-Uhlenbeck (OU) process[2] on an interval $[0, T]$

$$dx_t = -x_t dt + \sqrt{2} dw_t, \quad x_0 \sim p^{\diamond}, \tag{6}$$

where $\diamond \in \{\text{src}, \text{tgt}\}$ and $\{w_t\}_{t\in[0,T]}$ is a Wiener process. We denote p_t as the marginal density of the stochastic process $\{x_t\}_{t\in[0,T]}$ given by Eq. (6). It associates with the following PF-ODE [28]

$$\frac{d}{dt} x_t = x_t + \nabla \log p_t^{\diamond}(x_t), \quad \text{where } \diamond \in \{\text{src}, \text{tgt}\}. \tag{7}$$

When $t = 0$, it represents the clean data space (where p^{src} and p^{tgt} are supported on), and when $t = T$, it represents the latent noisy space.

For $\diamond \in \{\text{src}, \text{tgt}\}$, diffusion model $D_\theta^{\diamond}(x, t)$ is trained to approximate $\nabla \log p_t^{\diamond}(x)$ and leads to the following *empirical PF-ODE*

$$\frac{d}{dt} \hat{x}_t = \hat{x}_t + D_\theta^{\diamond}(\hat{x}_t, t). \tag{8}$$

Let p and q be two densities defined on $\mathbb{R}^d$. We define the *total variation distance* between p and q as

$$\text{TV}(p, q) := \frac{1}{2} \int |p(x) - q(x)| \, dx.$$

Starting from $x^{\text{src}} \sim p^{\text{src}}$, the following ODEs solving defines a *cycle manner procedure*

$$\begin{aligned} \hat{x}^{\text{latent}} &= \text{ODESolve}(x^{\text{src}}; D_\theta^{\text{src}}, 0, T), \\ \hat{x}^{\text{tgt}} &= \text{ODESolve}(\hat{x}^{\text{latent}}; D_\theta^{\text{tgt}}, T, 0), \end{aligned} \tag{9}$$

and then

$$\begin{aligned} \hat{\hat{x}}^{\text{latent}} &= \text{ODESolve}(\hat{x}^{\text{tgt}}; D_\theta^{\text{tgt}}, 0, T), \\ \hat{\hat{x}}^{\text{src}} &= \text{ODESolve}(\hat{\hat{x}}^{\text{latent}}; D_\theta^{\text{src}}, T, 0), \end{aligned} \tag{10}$$

DDIB proves the *cycle consistency property* that $\hat{\hat{x}}^{\text{src}} = x^{\text{src}}$, but assumes perfect diffusion model training and no ODE discretization errors, which are unrealistic. In Theorem 1', we establish distributional cycle consistency by accounting for diffusion model training errors and ODESolve discretization errors.

A-2. Assumptions

We list up the assumptions which are mostly similar to those in [36].

Assumption A (Compactly supported densities)**.** *Both p^{src} and p^{tgt} are compactly supported on a compact set in $\mathbb{R}^d$.*

Assumption B (Training accuracy of diffusion model)**.** *Let $\epsilon_{DM} > 0$. For $\diamond \in \{src, tgt\}$,*

$$\int_0^T \mathbb{E}_{x_t \sim p_t(x)} \left[\|D_\theta^{\diamond}(x_t, t) - \nabla \log p_t^{\diamond}(x_t)\|_2^2 \right] dt \le \epsilon_{DM}^2$$

Assumption C (Smoothness of diffusion model)**.** *For $\diamond \in \{src, tgt\}$, assume that $D_\theta^{\diamond}(\cdot, t)$ is $\mathscr{C}^2(\mathbb{R}^d)$ for all $t \in [0, T]$. That is, it is twice continuously differentiable. Additionally, we assume that there is a constant $L_t > 0$ so that*

$$\|D_\theta^{\diamond}(\cdot, t)\|_{\mathscr{C}^2(\mathbb{R}^d)} \le L_t.$$

We denote $L := \int_0^T L_t dt$ and assume that $L < \infty$.

A-3. Full Statement of Theorem 1 and Its Proof

To ensure precision, we slightly modify the notations used in the main manuscript. We present the theorem with time discretization, corresponding one-to-one with variance discretization [29]. Let $t_{N-1} = T > \cdots > t_{i+1} > t_i > \cdots > t_0 = 0$ be the discretization timestep on $[0, T]$, and define $h := \max_{i\in\{0,\cdots,N-1\}} |t_{i+1} - t_i|$.

Starting from $x^{(s)} \sim p^{\text{src}}$, let p^{latent} be the oracle density obtained by the forward-in-time PF-ODE (Eq. (7) with $\diamond = \text{src}$), and $\hat{p}^{\text{latent}}$ be the pushforward density obtained by solving the ODE (Eq. (8) with $\diamond = \text{src}$) numerically:

$$\hat{x}^{(l)} = \text{ODESolve}(x^{(s)}; D_\theta^{\text{src}}, 0, T), \quad x^{(s)} \sim p^{\text{src}}.$$

Now starting from the noisy latent space, let $\hat{p}^{\text{tgt}}$ be the density obtained by solving the ODE (Eq. (8) with $\diamond = \text{tgt}$), starting from $\hat{x}^{(l)} \sim \hat{p}^{\text{latent}}$:

$$\hat{x}^{(t)} = \text{ODESolve}(\hat{x}^{(l)}; D_\theta^{\text{tgt}}, T, 0), \quad \hat{x}^{(l)} \sim \hat{p}^{\text{latent}}.$$

We now present the full statement of Theorem 1 along with its proof.

[2] The statement and argument may be extended to a more general diffusion process. However, we leave it as a future work.

Theorem 1' (Distributional Cycle Consistency)**.** *Consider the ODE solvers are κ^{th}-order RK method. Under Assumptions A, B, and C, the total variation distance TV between $\hat{p}^{tgt}$ and p^{tgt} is bounded by:*

$$TV\big(\hat{p}^{tgt}, p^{tgt}\big) \lesssim \mathcal{O}(\epsilon_{DM}) + \mathcal{O}(h^{\kappa}).$$

Here, $\lesssim$ and $\mathcal{O}(\cdot)$ conceals a multiplication constant depending only on dimensionality d, $p^{\diamond}$ with $\diamond \in \{src, tgt\}$, and the pre-defined Runge–Kutta matrix [37].

Proof. Applying [36]'s Theorem 3.10 and its Remark C.2 backward in time (from T to 0) to Eqs. (8) and (7) with $\diamond = \text{tgt}$, we obtain

$$\text{TV}\big(\hat{p}^{\text{tgt}}, p^{\text{tgt}}\big) \lesssim \text{TV}\big(\hat{p}^{\text{latent}}, p^{\text{latent}}\big) + \mathcal{O}(\epsilon_{\text{DM}}) + \mathcal{O}(h^{\kappa}).$$

Now applying the same theorem but forward in time (from 0 to T) to Eqs. (8) and (7) with $\diamond = \text{src}$, we obtain

$$\text{TV}\big(\hat{p}^{\text{latent}}, p^{\text{latent}}\big) \lesssim \mathcal{O}(\epsilon_{\text{DM}}) + \mathcal{O}(h^{\kappa}),$$

as we start from the same initial distribution p^{src}. Combining these two inequalities, we derive the desired bound:

$$\text{TV}\big(\hat{p}^{\text{tgt}}, p^{\text{tgt}}\big) \lesssim \mathcal{O}(\epsilon_{\text{DM}}) + \mathcal{O}(h^{\kappa}).$$

□

We note that a sample-wise bound (instead of a distributional bound) can also be derived by analyzing the RK-solver in detail. Additionally, the bounds in Theorem 1' can be further refined using advanced techniques, but we do not pursue this overly complex mathematical analysis in this work.