Incorporating Stability Into Flow Matching

Christopher Iliffe Sprague [1 2] **Arne Elofsson** [1] **Hossein Azizpour** [2]

Abstract

In contexts where data samples represent a physically stable state, it is often assumed that the data points represent the local minima of an energy landscape. In control theory, it is well-known that energy can serve as an effective Lyapunov function. Despite this, connections between control theory and generative models in the literature are sparse, even though there are several machine learning applications with physically stable data points. In this paper, we focus on such data and a recent class of deep generative models called flow matching. We apply tools of stochastic stability to flow matching models. In doing so, we formally characterize a space of flow matching models that are amenable to this treatment, as well as draw connections to other control theory principles. We demonstrate our theoretical results on a toy example.

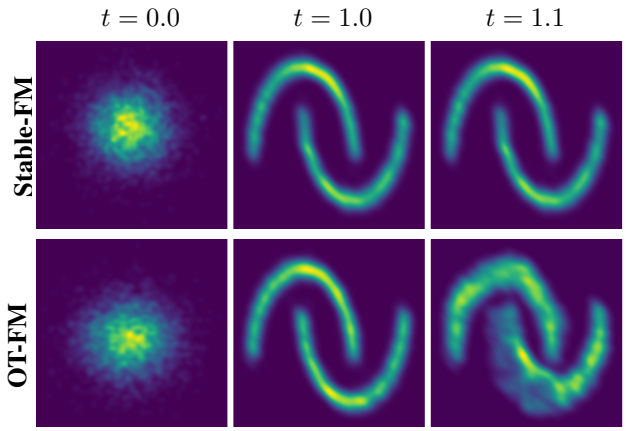

Figure 1. Flows of the Stable-FM model (top) and OT-FM model (bottom) from the standard normal distribution to the moons distribution. Note that the flow of OT-FM model does not stabilize to the distribution at $t = 1$, while the Stable-FM model remains stable to the distribution as $t \to \infty$. See Figure 2 for the corresponding vector field.

1. Introduction

Generative modeling is a fundamental problem in machine learning, where the goal is usually to learn a model that can generate samples from a target distribution. In recent years, deep generative models based on continuous-time dynamics (Song et al., 2020; Lipman et al., 2022) have shown exceptional capabilities in a variety of tasks ranging from image generation (Rombach et al., 2021) to complex structural biology applications (Corso et al., 2023; Ketata et al., 2023).

In contexts where samples from the target distribution represent a physically stable configuration, e.g. molecular conformations (Parr et al., 1979) or robotic formations (Sun et al., 2017), it is natural to posit that incorporating their physical characteristics into the generative model will bolster performance. Indeed, a variety of works have attempted to do so through force-field analogies (Shi et al., 2021; Feng et al., 2023; Luo et al., 2021; Zaidi et al., 2022), which rely on the idea that such samples represent local minima on a free-energy landscape (Dill, 1985).

In control theory, it is well-known that an energy function can serve as an effective Lyapunov function and help to endow a dynamical system with stability (Khalil, 2002). Despite this, connections in the literature between stability theory and generative models for physically stable data are sparse, even though there are existing works on applying Lyapunov stability to machine learning (Kang et al., 2021; Rodriguez et al., 2022; Zhang et al., 2022a; Kolter & Manek, 2019; Zhang et al., 2022b; Lawrence et al., 2020).

To bridge this gap, we apply a stochastic version of La Salle's invariance principle (La Salle, 1966; Mao, 1999) to flow matching (FM) models (Lipman et al., 2022), a recent alternative to diffusion models based on continuous normalizing flows (CNFs) (Chen et al., 2018). The main contribution of our work is equiping the dynamics of the FM model with stability to the support of the data distribution, with the goal to reflect the physical stability of its samples.

[1]Science for Life Laboratory, Solna, Sweden [2]KTH Royal Institute of Technology, Stockholm, Sweden. Correspondence to: Christopher Iliffe Sprague <sprague@kth.se>.

Accepted by the Structured Probabilistic Inference & Generative Modeling workshop of ICML 2024, Vienna, Austria.

2. Related work

In recent years, continuous-time dynamics-based deep generative models (Chen et al., 2018; Song et al., 2020; Lipman et al., 2022) have come to the forefront of the field of deep generative modeling.

Notably, diffusion models (Song et al., 2020) were shown to be state-of-the-art in image generation tasks (Dhariwal & Nichol, 2021), and have since garnered multiple applications ranging from structural biology tasks (Corso et al., 2023; Yim et al., 2023b; Ketata et al., 2023) to video generation (Ho et al., 2022; Blattmann et al., 2023; Esser et al., 2023), with several extensions such as latent representations (Vahdat et al., 2021; Blattmann et al., 2023) and geometric priors (Bortoli et al., 2022; Dockhorn et al., 2021). Diffusion models are based on the SDEs (e.g. (Särkkä & Solin, 2019)) and score matching (Hyvärinen, 2005; Song et al., 2019), where one learns a time-dependent vector-valued score function $\nabla_{\boldsymbol{x}} \log(p(\boldsymbol{x}, t))$ over diffused data samples and then plugs it into the well-known time-reversed SDE (Anderson, 1982; Haussmann & Pardoux, 1986; Lindquist & Picci, 1979).

FM models (Lipman et al., 2022) were proposed as an alternative to diffusion models (Song et al., 2020) that enjoy fast training and sampling, while maintaining competitive performance. They rely on CNFs (Chen et al., 2018) and can be seen as a generalization of diffusion models, as demonstrated by the existence of the probability flow ordinary differential equation (ODE) that induces the same marginal PDF as the SDE of diffusion models (Maoutsa et al., 2020; Song et al., 2020). Several applications of FM models have been made, ranging from structural biology (Yim et al., 2023a; Bose et al., 2023) to media (Le et al., 2023; Liu et al., 2023), as well as fundamental extensions (Tong et al., 2023; Pooladian et al., 2023; Shaul et al., 2023; Chen & Lipman, 2023b; Klein et al., 2023).

In the context of data that describes physically stable states, e.g. molecular conformations, several works (Zaidi et al., 2022; Feng et al., 2023; Shi et al., 2021; Luo et al., 2021) have learned force fields by leveraging the connection between the score function and Boltzmann distributions, i.e. $\nabla_{\boldsymbol{x}} \log(p(\boldsymbol{x}, t)) = -\nabla_{\boldsymbol{x}} H(\boldsymbol{x}, t)$ where H is a scalar energy function. In (Zaidi et al., 2022), an equivalence between denoising score matching (Vincent, 2011) and force-field learning is pointed out. This idea was extended in (Feng et al., 2023) to incorporate off-equilibrium data and NN gradient fields. In (Shi et al., 2021; Luo et al., 2021), a NN gradient field is learned to model a psuedo-force field, which is then used to sample energy-minimizing molecular conformations via annealed Langevin dynamics (Song & Ermon, 2019).

Along the same lines, Poisson Flow Generative Models (Xu et al., 2023; 2022), use the solution of Poisson's equation to define a force field as the gradient of the Poisson potential. Similar to our work, they augment the spatial state with an auxiliary state that acts as an interpolant and replaces time. However, the gradient field of their spatial state is linked to that of their auxiliary variable. In contrast, our work deals with quadratic potentials that yield independent dynamics for the spatial and auxiliary states.

In control theory, physical stability is often analyzed through Lyapunov stability (Khalil, 2002). As a result, there are numerous works on applying Lyapunov stability to ODEs/SDEs in the context of dynamics learning (Kang et al., 2021; Rodriguez et al., 2022; Zhang et al., 2022a; Kolter & Manek, 2019; Zhang et al., 2022b; Lawrence et al., 2020). Despite this, virtually none of the aforementioned works consider stochastic stability to support of the target distribution. In our work, we apply a stochastic invariance principle (Theorem 3.5) to construct stable CNFs in the context of FM.

3. Preliminaries

In this paper, we assume that we have a dataset $\mathcal{D} \subset \mathcal{X}$, existing in an ambient space $\mathcal{X} \subseteq \mathbb{R}^n$ that we assume to be described by a latent PDF $q \in \mathcal{P}(\mathcal{X})$ s.t. $\mathcal{D} \subset \text{supp}(q)$. We want to generate new samples from the latent PDF q.

Notation Denote the space of continuous functions from $\mathcal{X}$ to $\mathcal{Y}$ as $C(\mathcal{X}, \mathcal{Y})$. Denote the space of continuous function-valued functions s.t. if $f \in C(\mathcal{X}, C(\mathcal{Y}, \mathcal{Z}))$ then $f(\cdot \mid \boldsymbol{x}) \in C(\mathcal{Y}, \mathcal{Z})$ and $f(\boldsymbol{y} \mid \boldsymbol{x}) \in \mathcal{Z}$. Denote the space of non-negative reals as $\mathbb{R}_{\geq 0}$ and positive reals as $\mathbb{R}_{>0}$. Denote the space of positive definite matrices as $\mathbb{R}^{n \times n}_{>0}$ and positive semi-definite matrices as $\mathbb{R}^{n \times n}_{\geq 0}$. Denote the space of probability density functions (PDFs) as $\mathcal{P}(\mathcal{X}) = \{f \in C(\mathcal{X}, \mathbb{R}_{\geq 0}) \mid \int_{\mathcal{X}} f(\boldsymbol{x}) \mathrm{d}\boldsymbol{x} = 1\}$ and PDF paths as $\mathcal{P}(\mathcal{X} \times \mathbb{R}_{\geq 0}) = \{f \in C(\mathcal{X} \times \mathbb{R}_{\geq 0}, \mathbb{R}_{\geq 0}) \mid \int_{\mathcal{X}} f(\boldsymbol{x}, t) \mathrm{d}\boldsymbol{x} = 1\}$. Denote the support of a PDF $f \in \mathcal{P}(\mathcal{X})$ as $\text{supp}(f) = \{\boldsymbol{x} \in \mathcal{X} \mid f(\boldsymbol{x}) \neq 0\}$. Denote the Dirac delta PDF centered at $\boldsymbol{x}_1 \in \mathcal{X}$ as $\delta(\boldsymbol{x} - \boldsymbol{x}_1)$.

3.1. Flow Matching

Explicitly modeling the latent PDF q comes with the burden of constraining the model to be a proper PDF with a suitable normalizing constant, which is often intractable. Continuity-based generative models (Song et al., 2020; Lipman et al., 2022), on the other hand, circumvent this burden by modeling vector fields (VFs) that transport samples from a tractable PDF (e.g. Gaussian) to the latent PDF via the well-known continuity equation. These models are often called continuous normalizing flows (CNFs).

In flow matching (FM) (Lipman et al., 2022), we assume

that there exists a latent marginal PDF path $p \in \mathcal{P}(\mathcal{X} \times \mathbb{R}^n)$ that transports $\mathcal{N}(\mathbf{0}, \boldsymbol{I}) = p(\boldsymbol{x}, 0)$ to $q(\boldsymbol{x}) = p(\boldsymbol{x}, T)$ at $T = 1$. The continuity equation then tells us that there exists a conservative VF $\boldsymbol{v} \in C(\mathcal{X} \times \mathbb{R}_{\geq 0}, \mathbb{R}^n)$ that generates this PDF path (Villani et al., 2009; Ambrosio et al., 2005). Therefore, we define the space of CNFs as ordered pairs of a PDF path, VF, and flow map in Definition 3.1.

Definition 3.1 (CNF Space)**.** A subset

$$\mathfrak{F} \subset \mathcal{P}(\mathcal{X} \times \mathbb{R}_{\geq 0}) \times C(\mathcal{X} \times \mathbb{R}_{\geq 0}, \mathbb{R}^n) \times C(\mathcal{X} \times \mathbb{R}_{\geq 0}, \mathcal{X}) \tag{1}$$

s.t. for all CNFs $(p, \boldsymbol{v}, \boldsymbol{\phi}) \in \mathfrak{F}$:

$$\frac{\partial p(\boldsymbol{x}, t)}{\partial t} = -\nabla_{\boldsymbol{x}} \cdot (p(\boldsymbol{x}, t)\boldsymbol{v}(\boldsymbol{x}, t)) \tag{2}$$

$$\frac{\mathrm{d}\boldsymbol{\phi}(\boldsymbol{x}, t)}{\mathrm{d}t} = \boldsymbol{v}(\boldsymbol{\phi}(\boldsymbol{x}, t), t)\,. \tag{3}$$

Since we do not have access to the latent CNF, we assume that we can approximate it via a mixture of conditional CNFs, conditioned on data samples $\boldsymbol{x}_1 \in \operatorname{supp}(q)$, as shown in Definition 3.2.

Definition 3.2 (CNF Pair Space)**.** A subset

$$\mathfrak{F}_{\text{FM}} \subset \mathfrak{F} \times C(\mathcal{X}, \mathfrak{F}) \quad \text{s.t.} \quad \exists q \in \mathcal{P}(\mathcal{X}) \tag{4}$$

and for all CNF pairs $((p, \boldsymbol{v}, \cdot), (p_1, \boldsymbol{v}_1, \cdot)) \in \mathfrak{F}_{\text{FM}}$:

$$p(\boldsymbol{x}, t) = \int_{\mathcal{X}} p_1(\boldsymbol{x}, t \mid \boldsymbol{x}_1) q(\boldsymbol{x}_1) \mathrm{d}\boldsymbol{x}_1, \tag{5}$$

$$\boldsymbol{v}(\boldsymbol{x}, t) = \frac{1}{p(\boldsymbol{x}, t)} \int_{\mathcal{X}} \boldsymbol{v}_1(\boldsymbol{x}, t \mid \boldsymbol{x}_1) p_1(\boldsymbol{x}, t \mid \boldsymbol{x}_1) q(\boldsymbol{x}_1) \mathrm{d}\boldsymbol{x}_1$$

s.t. there exists $T \in \mathbb{R}_{\geq 0} \cup \{\infty\}$ s.t. for all $\boldsymbol{x}_1 \in \operatorname{supp}(q)$ there exists $\boldsymbol{\Sigma}_1 \in \mathbb{R}^{n \times n}_{\geq 0}$ s.t. $p_1(\boldsymbol{x}, T \mid \boldsymbol{x}_1) = \mathcal{N}(\boldsymbol{x}_1, \boldsymbol{\Sigma}_1)$.

The importance of Definition 3.2 is that it allows us to regress the conditional CNF's VF $\boldsymbol{v}_1$ instead of the unknown latent CNF's VF $\boldsymbol{v}$, as shown in Theorem 3.3.

Theorem 3.3 (FM Loss)**.** *For all $\boldsymbol{v}_\theta \in C(\mathcal{X} \times \mathbb{R}_{\geq 0}, \mathcal{X})$ and $((p, \boldsymbol{v}, \cdot), (p_1, \boldsymbol{v}_1, \cdot)) \in \mathfrak{F}_{FM}$:*

$$L_{FM}(\theta) = \mathop{\mathbb{E}}_{\substack{t \sim \mathcal{U}[0,T] \\ \boldsymbol{x}_1 \sim q(\boldsymbol{x}_1) \\ \boldsymbol{x} \sim p_1(\boldsymbol{x}, t \mid \boldsymbol{x}_1)}} \|\boldsymbol{v}_\theta(\boldsymbol{x}, t) - \boldsymbol{v}_1(\boldsymbol{x}, t \mid \boldsymbol{x}_1)\|_2^2 \tag{6}$$

$$= \mathop{\mathbb{E}}_{\substack{t \sim \mathcal{U}[0,T] \\ \boldsymbol{x} \sim p(\boldsymbol{x}, t)}} \|\boldsymbol{v}_\theta(\boldsymbol{x}, t) - \boldsymbol{v}(\boldsymbol{x}, t)\|_2^2 + \mathit{Const}. \tag{7}$$

Proof. See (Lipman et al., 2022). □

3.2. Stochastic Stability

In this paper, we will analyze FM from the perspective of stochastic differential equations (SDEs). A key link between FM (Lipman et al., 2022) and SDE-based generative models (Song et al., 2020) is the probability flow ODE, whose solutions induce the same PDF path as its corresponding SDE, see Theorem 3.4.

Theorem 3.4 (Probability Flow)**.** *An SDE*

$$\mathrm{d}\boldsymbol{x} = \boldsymbol{f}(\boldsymbol{x}, t)\mathrm{d}t + \boldsymbol{g}(\boldsymbol{x}, t)\mathrm{d}\boldsymbol{w}. \tag{8}$$

with drift $\boldsymbol{f} \in C(\mathcal{X}, \mathbb{R}^n)$ and diffusion $\boldsymbol{g} \in C(\mathcal{X}, \mathbb{R}^{n \times n})$ induces the same PDF path as the ODE

$$\frac{\mathrm{d}\boldsymbol{x}}{\mathrm{d}t} = \boldsymbol{f}(\boldsymbol{x}, t) - \frac{1}{2}\nabla_{\boldsymbol{x}} \cdot \boldsymbol{g}(\boldsymbol{x}, t)^2 - \frac{1}{2}\boldsymbol{g}(\boldsymbol{x}, t)^2 \nabla_{\boldsymbol{x}} \ln(p(\boldsymbol{x}, t)) \tag{9}$$

where $\boldsymbol{g}(\boldsymbol{x}, t)^2 = \boldsymbol{g}(\boldsymbol{x}, t)\boldsymbol{g}(\boldsymbol{x}, t)^\top$, and $\nabla_{\boldsymbol{x}} \cdot$ is the divergence operator.

Proof. See (Song et al., 2020; Maoutsa et al., 2020). □

We are interested in modeling a CNF such that its PDF path is stable to the latent PDF q.

In ODEs, stability is a well-studied topic (Khalil, 2002; Meiss, 2007; Chossat & Lauterbach, 2000), where one typically seeks to characterize stability to a point $\boldsymbol{x}_1 \in \mathcal{X}$ s.t. $\boldsymbol{\phi}(\boldsymbol{x}_0, \infty) = \boldsymbol{x}_1$ for all starting states $\boldsymbol{x}_0 \in \mathcal{B}_0 \supset \{\boldsymbol{x}_1\}$. To do this, one typically applies Lyapunov's direct method, where if one can find a scalar function $V \in C(\mathcal{X}, \mathbb{R}_{\geq 0})$ whose gradient along the VF resembles a funnel centered at $\boldsymbol{x}_1$, then $\boldsymbol{x}_1$ is an asymptotically stable point. Similarly, one can also characterize stability to a set $\mathcal{B}_1 \subset \mathcal{X}$, s.t. $\boldsymbol{\phi}(\boldsymbol{x}_0, \infty) \in \mathcal{B}_1$ for all starting states $\boldsymbol{x}_0 \in \mathcal{B}_0 \supset \mathcal{B}_1$. In this case, one can apply La Salle's invariance principle (La Salle, 1966), which employs a Lyapunov-like function to locate invariant such stable sets.

In SDEs, notions of the aforementioned methods have been developed in (Kushner, 1965; Khasminskii, 1980; Mao, 1999). As the latent PDF q may characterize a set of stable states, we are interested in applying the stochastic version (Mao, 1999) of La Salle's invariance principle. We adapt the principle to this setting in Theorem 3.5.

Theorem 3.5 (Stochastic Invariance)**.** *For an SDE, an operator $\mathcal{L}$ acting on a scalar function $V \in C(\mathcal{X}, \mathbb{R}_{>0})$*

$$\mathcal{L}V(\boldsymbol{x}) := \nabla_{\boldsymbol{x}} V(\boldsymbol{x})^\top \boldsymbol{f}(\boldsymbol{x}) + \frac{1}{2}\operatorname{tr}\left(\nabla^2_{\boldsymbol{x}} V(\boldsymbol{x})\boldsymbol{g}(\boldsymbol{x})^2\right), \tag{10}$$

where $\nabla^2_{\boldsymbol{x}}$ is the Hessian operator and tr *is the trace operator, and two sets $\mathcal{B}_0 \supset \mathcal{B}_1$*

$$\mathcal{B}_0 := \{\boldsymbol{x} \in \mathcal{X} \mid \mathcal{L}V(\boldsymbol{x}) \leq 0\}, \tag{11}$$

$$\mathcal{B}_1 := \{\boldsymbol{x} \in \mathcal{X} \mid \mathcal{L}V(\boldsymbol{x}) = 0\}, \tag{12}$$

if $\boldsymbol{x}_0 \in \mathcal{B}_0$ then $\boldsymbol{\phi}(\boldsymbol{x}_0, \infty) \in \mathcal{B}_1$ almost surely.

Proof. See Corollary 4.1 in (Mao, 1999). □

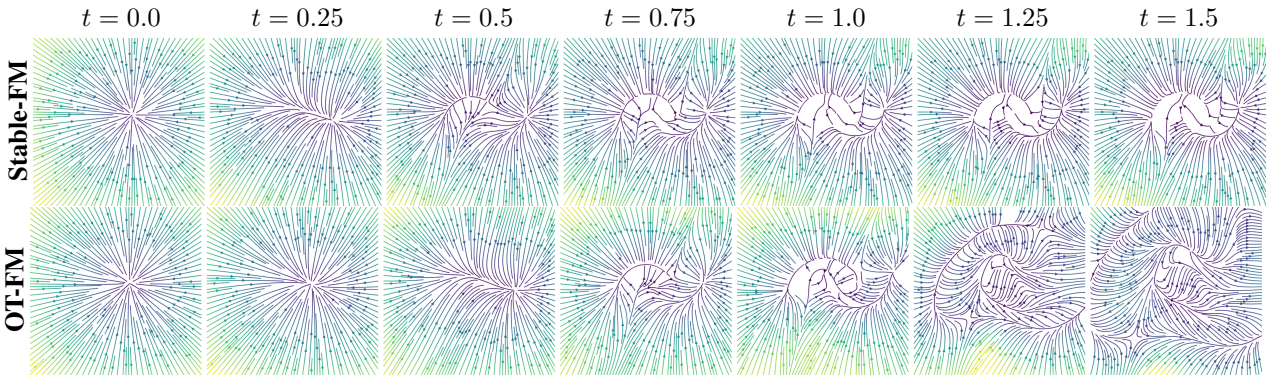

Figure 2. Stream plots of the vector fields of Stable-FM (top) and OT-FM (bottom) corresponding to Figure 1. Note that beyond $t = 1$, the OT-FM vector field diverges, while the Stable-FM VF stabilizes. Lighter colors indicate larger vector magnitudes and vice versa.

4. Main Result

4.1. Autonomous CNF Pair Space

We aim to construct CNF pairs that satisfy the invariance principle of Theorem 3.5; however, the principle only admits autonomous systems. The marginal VF $\boldsymbol{v}(\boldsymbol{x}, t)$ in Definition 3.2 does not admit such a form, even if the conditional VF $\boldsymbol{v}_1(\boldsymbol{x} \mid \boldsymbol{x}_1)$ does, due to its dependence on the constructed PDF path $p(\boldsymbol{x}, t)$. To alleviate this, we consider that t in the context of CNFs serves merely as an interpolation parameter between the initial and target PDFs. Thus, we ask if it would be possible to use other variables in the state space to perform the same job while also rendering an autonomous form for the marginal VF $\boldsymbol{v}(\boldsymbol{x}, t)$.

To explore this line of reasoning, we decompose the state space into two subspaces $\mathcal{X} := \mathcal{Y} \times \mathcal{Z}$: the data space $\mathcal{Y} \subset \mathbb{R}^{n-1}$ and the interpolant space $\mathcal{Z} \subset \mathbb{R}$. We then suppose that the interpolant state $z \in \mathcal{Z}$ evolves deterministically over time via a dirac-delta PDF path, but is still involved as arguments to the VFs. We define this new CNF pair space in Definition 4.1.

Definition 4.1 (Auto CNF Pair Space)**.** A subset

$$\mathfrak{F}_{\text{Auto}} \subset \mathfrak{F}_{\text{FM}} \quad \text{s.t.} \quad \begin{array}{c} \mathcal{X} = \mathcal{Y} \times \mathcal{Z} \subset \mathbb{R}^{n-1} \times \mathbb{R} \\ q(\boldsymbol{x}) = q_y(\boldsymbol{y}) q_z(z) \end{array} \tag{13}$$

and for all $(\cdot, (p, \boldsymbol{v}, \boldsymbol{\phi})) \in \mathfrak{F}_{\text{Auto}}$:

$$p(\boldsymbol{x}, t \mid \boldsymbol{x}_1) = p_y(\boldsymbol{y}, t \mid \boldsymbol{y}_1) p_z(z, t \mid z_1) \tag{14}$$

$$\boldsymbol{v}(\boldsymbol{x}, t \mid \boldsymbol{x}_1) = \begin{bmatrix} \boldsymbol{v}_y(\boldsymbol{x}, t \mid \boldsymbol{x}_1), \\ v_z(\boldsymbol{x}, t \mid \boldsymbol{x}_1) \end{bmatrix}, \tag{15}$$

and furthermore the flow map ϕ_z of z is bijective with t s.t. $t = \phi_z^{-1}(z_0, \phi_z(z_0, t \mid z_1) \mid z_1)$, and with $z_1, z_0 \in \mathcal{Z}$, s.t. $z_0 \neq z_1$ we have

$$p_z(z, t \mid z_1) = \delta(z - \phi_z(z_0, t \mid z_1)), \quad q_z(z) = \delta(z - z_1). \tag{16}$$

The key part of Definition 4.1 is that the PDF q_z is a Dirac delta PDF centered at z_1, and the conditional PDF path p_z is a Dirac delta PDF centered at $\phi_z(z_0, t \mid z_1)$. Thus, z will evolve deterministically over time from z_0 to z_1, and hence its flow map ϕ_z forms a bijection with time. We can then use the inverse of μ_z to map z to t and put the constructed CNF in autonomous form, as shown in Theorem 4.2.

Theorem 4.2. *For all* $((p, \boldsymbol{v}, \cdot), (p_1, \boldsymbol{v}_1, \boldsymbol{\phi}_1)) \in \mathfrak{F}_{\textit{Auto}}$ *there exists* $\bar{p} \in \mathcal{P}(\mathcal{X})$ *and* $\bar{\boldsymbol{v}} \in C(\mathcal{X}, \mathbb{R}^n)$ *corresponding to* $p(\boldsymbol{x}, t)$ *and* $\boldsymbol{v}(\boldsymbol{x}, t)$*:*

$$\begin{aligned} \bar{p}(\boldsymbol{x}) &= \int_{\mathcal{Y}} \bar{p}_y(\boldsymbol{x} \mid \boldsymbol{y}_1) q_y(\boldsymbol{y}_1) \mathrm{d}\boldsymbol{y}_1 \\ \bar{\boldsymbol{v}}(\boldsymbol{x}) &= \frac{1}{\bar{p}(\boldsymbol{x})} \int_{\mathcal{Y}} \bar{\boldsymbol{v}}_1(\boldsymbol{x} \mid \boldsymbol{x}_1) \bar{p}_y(\boldsymbol{x} \mid \boldsymbol{y}_1) q_y(\boldsymbol{y}_1) \mathrm{d}\boldsymbol{y}_1 \\ \textit{s.t.} \quad & \begin{aligned} \bar{p}_y(\boldsymbol{x} \mid \boldsymbol{y}_1) &= p_y(\boldsymbol{y}, \phi_z^{-1}(z_0, z \mid z_1) \mid \boldsymbol{y}_1) \\ \bar{\boldsymbol{v}}_1(\boldsymbol{x} \mid \boldsymbol{x}_1) &= \boldsymbol{v}_1(\boldsymbol{x}, \phi_z^{-1}(z_0, z \mid z_1) \mid \boldsymbol{x}_1). \end{aligned} \end{aligned} \tag{17}$$

Proof. See Appendix A.1. □

The key takeaway of Theorem 4.2 is that we can now consider modeling a VF in autonomous form.

Since we are interested in stability, we are interested in conditional CNFs s.t. $p(\boldsymbol{x}, \infty \mid \boldsymbol{x}_1) = \mathcal{N}(\boldsymbol{u}_1, \boldsymbol{\Sigma}_1)$ for some $\boldsymbol{\Sigma} \in \mathbb{R}^{n \times n}_{\succeq 0}$. This implies that $\bar{v}_z(z_1 \mid z_1) = 0$, which poses a problem in the denominator of the loss function in Lemma A.1. Furthermore, we are interested in minimizing the loss over $t \in [0, \infty]$ which poses a problem due to the presence of the $1/T$ term in the original loss. To circumvent these problems, we elect to minimize a loss that directly integrates over z, as shown in Definition 4.3.

Definition 4.3 (Autonomous CNF Loss)**.** For all $\boldsymbol{v}_\theta \in C(\mathcal{X}, \mathbb{R}^n)$ and $(\cdot, (p_1, \boldsymbol{v}_1, \boldsymbol{\phi}_1)) \in \mathfrak{F}_{\text{Auto}}$:

$$L_{\text{Auto}}(\theta) = \mathop{\mathbb{E}}_{\substack{z \sim \mathcal{U}[z_0, z_1] \\ \boldsymbol{y}_1 \sim q_y(\boldsymbol{y}) \\ \boldsymbol{y} \sim \bar{p}_y(\boldsymbol{y}, z \mid \boldsymbol{y}_1)}} \| \boldsymbol{v}_\theta(\boldsymbol{x}) - \bar{\boldsymbol{v}}_1(\boldsymbol{x} \mid \boldsymbol{x}_1) \|^2. \tag{18}$$

Importantly, the loss function in Definition 4.3 does not suffer from the ailments of the loss function in Lemma A.1. The training procedure then becomes: sample the uniform PDF $z \sim \mathcal{U}[z_0, z_1]$, sample a data point $\boldsymbol{y}_1 \sim q_y(\boldsymbol{y})$ from the dataset, sample the conditional PDF path $\boldsymbol{y} \sim \bar{p}_y(\boldsymbol{y}, z \mid \boldsymbol{y}_1)$, then minimize $\|\boldsymbol{v}_\theta(\boldsymbol{y}, z) - \boldsymbol{v}_1(\boldsymbol{y}, z \mid \boldsymbol{y}_1, z_1)\|^2$, where z_0 and z_1 are held constant. Assuming that the initial PDF is $\bar{p}_y(\boldsymbol{y}, z_0 \mid \boldsymbol{y}_1) = \mathcal{N}(\boldsymbol{y} \mid \boldsymbol{0}_{n-1}, \boldsymbol{I}_{n-1})$, the inference procedure is then: sample $\boldsymbol{x}_0 = (\boldsymbol{y}_0, z_0) \sim \mathcal{N}(\boldsymbol{0}_{n-1}, \boldsymbol{I}_{n-1}) \times \delta(z - z_0)$, then integrate $\boldsymbol{v}_\theta(\boldsymbol{z}, \tau)$ from $\boldsymbol{x}_0$ until $\boldsymbol{x}_1 = (\boldsymbol{y}_1, z_1) \sim q_y \times \delta(z - z_1)$.

4.2. Stable and Finite CNF Pair Space

With an autonomous form established in Definition 4.1, we are now ready to apply the invariance principle of Theorem 3.5. The first thing we would like to do is ensure that the conditional CNFs we use for training inherently satisfy the invariance principle. To do this, we borrow the idea of quadratic stability (Boyd et al., 1994) and model the conditional CNF's VFs as quadratic gradient flows, where there is a quadratic energy function with a global minimum centered at the conditioned data point. Rather than the conditional PDF paths leading to a single data point, we would like them to lead to a distribution around the data point as in (Lipman et al., 2022). We also need the corresponding SDE to be linear so that it gives samples of its PDF path in closed-form. To do this, we first analyze a linear time-invariant SDE (see section 6.2 of (Särkkä & Solin, 2019)) of the form

$$\mathrm{d}\boldsymbol{y} = -\boldsymbol{A}(\boldsymbol{y} - \boldsymbol{y}_1)\mathrm{d}t + \sqrt{2\boldsymbol{A}\boldsymbol{\Sigma}_1}\mathrm{d}\boldsymbol{w}(t). \tag{19}$$

Assuming that $\boldsymbol{A} \in \mathbb{R}^{n\times n}_{>0}$ is positive definite and commutes (i.e. is simultaneously diagonalizable) with $\boldsymbol{\Sigma}_1 \in \mathbb{R}^{n\times n}_{\geq 0}$, then the SDE will have an exponentially decaying PDF path that stabilizes to a stationary PDF given by $p_y(\boldsymbol{y}, \infty) = \mathcal{N}(\boldsymbol{y}_1, \boldsymbol{\Sigma}_1)$. However, if $\boldsymbol{A} \in \mathbb{R}^{n\times n}_{\geq 0}$ is positive semidefinite, then it has some zero eigenvalues, which can be used to enforce some equivariant degrees of freedom. In this case, the flow will be stationary along the eigenvectors corresponding to zero eigenvalues. An equivalent statement is that the flow will be stationary in the nullspace $\mathrm{null}(\boldsymbol{A})$ of $\boldsymbol{A}$, which is the space spanned by such eigenvectors, a.k.a. the center subspace (Carr, 1981; Kœnig, 1997). The flow along other eigenvectors, in the orthogonal complement of the nullspace, $\mathrm{null}(\boldsymbol{A})^\perp$ (a.k.a. stable subspace), will decay as expected. Thus, in order to ensure that the flow converges to the desired stationary PDF in the general case, one can enforce that the nullspace projection of the initial mean and covariance equals that of the target mean and covariance (see Definition 4.4). The projection of the initial mean and covariance in the orthogonal complement of the nullspace can be left as is.

As we will later show, it is desirable to have the dynamics of the interpolant state z track the PDF of the data state $\boldsymbol{y}$. To do this, we take inspiration from the conditional flow construction in Sec. 3.2 of (Chen & Lipman, 2023a), where a premetric is defined. For our interpolant flow, we define a probabilistic metric space $(\mathcal{P}(\mathcal{Y}), d)$, where $d : \mathcal{P}(\mathcal{Y}) \times \mathcal{P}(\mathcal{Y}) \to \mathbb{R}_{\geq 0}$ is a chosen metric (e.g. Wasserstein distance), satisfying the usual axioms. Since we are working with linear SDEs with Gaussian PDFs, such a metric can be expressed in a tractable form in terms of the mean and covariance. As previously explained, the flows of our data state $\boldsymbol{y}$ will converge to a stationary PDF, thus it is convenient to track the distance of the data PDF from its stationary PDF $\mathcal{N}(\boldsymbol{y}_1, \boldsymbol{\Sigma}_1)$. Intuitively, this relative distance will decay to zero, and hence can be used to define a flow for the interpolant state.

Since both the data PDF and interpolant PDF will converge to a stationary PDF, they satisfy the invariance principle Theorem 3.5. We define the space of CNFs that produce these flows in Definition 4.4.

Definition 4.4 (Stable CNF Pair Space)**.** A subset

$$\mathfrak{F}_{\text{Stable}} \subset \mathfrak{F}_{\text{Auto}} \tag{20}$$

s.t. for all $(\cdot, (p, \cdot, \cdot)) \in \mathfrak{F}_{\text{Stable}}$ the following holds.

$$\begin{aligned} p_y(\boldsymbol{y}, t \mid \boldsymbol{y}_1) &= \mathcal{N}(\boldsymbol{y} \mid \boldsymbol{\mu}_y(t), \boldsymbol{\Sigma}_y(t)) \\ \text{s.t.} \quad \boldsymbol{\mu}_y(t) &= \boldsymbol{y}_1 + e^{-\boldsymbol{A}t}(\boldsymbol{y}_0 - \boldsymbol{y}_1) \\ \boldsymbol{\Sigma}_y(t) &= \boldsymbol{\Sigma}_1 + e^{-2\boldsymbol{A}t}(\boldsymbol{\Sigma}_0 - \boldsymbol{\Sigma}_1) \end{aligned} \tag{21}$$

where $\boldsymbol{A}, \boldsymbol{\Sigma}_0, \boldsymbol{\Sigma}_1 \in \mathbb{R}^{n\times n}_{\geq 0}$ are positive semidefinite and simultaneously diagonalizable by orthogonal eigenvectors $\boldsymbol{P} = [\boldsymbol{P}_C, \boldsymbol{P}_H] \in \mathbb{R}^{n\times n}$ s.t. $\mathrm{span}(\boldsymbol{P}_C) = \mathrm{null}(\boldsymbol{A})$ is the nullspace of $\boldsymbol{A}$ and $\mathrm{span}(\boldsymbol{P}_H) = \mathrm{null}(\boldsymbol{A})^\perp$ is its orthogonal complement, and furthermore $\boldsymbol{\mu}_0 = \boldsymbol{P}_C^2\boldsymbol{x}_1 + \boldsymbol{P}_H^2\boldsymbol{\mu}_0'$ for some $\boldsymbol{\mu}_0' \in \mathcal{X}$ and $\boldsymbol{\Sigma}_0 = \boldsymbol{P}_C^2\boldsymbol{\Sigma}_1\boldsymbol{P}_C^2 + \boldsymbol{P}_H^2\boldsymbol{\Sigma}_0'\boldsymbol{P}_H^2$ for some $\boldsymbol{\Sigma}_0' \in \mathbb{R}^{n\times n}_{\geq 0}$, where $\boldsymbol{P}^2 = \boldsymbol{P}\boldsymbol{P}^\top$.

$$\begin{aligned} p_z(z, t \mid z_1) &= \delta(z - \mu_z(t)) \\ \text{s.t.} \quad \mu_z(t) &= z_1 + \left(\frac{d(\boldsymbol{\mu}_y(t), \boldsymbol{\Sigma}_y(t))}{d(\boldsymbol{\mu}_y(0), \boldsymbol{\Sigma}_y(0))}\right)^\kappa (z_0 - z_1) \end{aligned} \tag{22}$$

where $\kappa \in \mathbb{R}_{>0}$ is a scalar and $d : \mathcal{P}(\mathcal{Y}) \times \mathcal{P}(\mathcal{Y}) \to \mathbb{R}_{\geq 0}$ is a metric abbreviated as the distance between the current and stationary Guassian PDFs

$$d(\boldsymbol{\mu}_y(t), \boldsymbol{\Sigma}_y(t)) = d(p_y(\boldsymbol{y}, t \mid \boldsymbol{y}_1), p_y(\boldsymbol{y}, \infty \mid \boldsymbol{y}_1)). \tag{23}$$

The equations for the mean and covariance of the data state in Definition 4.4 correspond to the aforementioned LTI SDE, thus its corresponding VF is straightforwardly obtained via the PF-ODE in Theorem 3.4. Since the interpolant state has deterministic dynamics, its time-derivative gives its VF directly. Without a declared distance metric, the resulting VF can be obtained in terms of the mean and covariance

dynamics (see Eq. 6.18 in (Särkkä & Solin, 2019)) and the partial derivatives of the distance metric w.r.t. the mean and covariance of the data state. We establish this result in Lemma 4.5.

Lemma 4.5. *For all* $(\cdot, (\cdot, \boldsymbol{v}, \cdot)) \in \mathfrak{F}_{\textit{Stable}}$*:*

$$\begin{aligned} \boldsymbol{v}_y(\boldsymbol{x}, t \mid \boldsymbol{x}_1) &= -\boldsymbol{A}\left(\boldsymbol{y} - \boldsymbol{y}_1 - \boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_y^{-1}(t)\left(\boldsymbol{y} - \boldsymbol{\mu}_y(t)\right)\right) \\ v_z(\boldsymbol{x}, t \mid \boldsymbol{x}_1) &= -\frac{\kappa\left(h_\mu(t) + h_\Sigma(t)\right)}{h(t)}(z - z_1) \end{aligned} \tag{24}$$

where $h(t) = d(\boldsymbol{\mu}_y(t), \boldsymbol{\Sigma}_y(t))$ *and*

$$\begin{aligned} h_\mu(t) &= \left\langle \frac{\partial d(\boldsymbol{\mu}_y(t), \boldsymbol{\Sigma}_y(t))}{\partial \boldsymbol{\mu}_y(t)}, \boldsymbol{A}\left(\boldsymbol{\mu}_y(t) - \boldsymbol{y}_1\right) \right\rangle \\ h_\Sigma(t) &= \left\langle \frac{\partial d(\boldsymbol{\mu}_y(t), \boldsymbol{\Sigma}_y(t))}{\partial \boldsymbol{\Sigma}_y(t)}, 2\boldsymbol{A}\left(\boldsymbol{\Sigma}_y(t) - \boldsymbol{\Sigma}_1\right) \right\rangle_F, \end{aligned} \tag{25}$$

where $\langle \cdot, \cdot \rangle$ *and* $\langle \cdot, \cdot \rangle_F$ *are the Euclidean and Frobenius inner products respectively. Additionally,* $\boldsymbol{v}_y(\boldsymbol{x}, \infty \mid \boldsymbol{x}_1) = 0$ *and* $v_z(\boldsymbol{x}, \infty \mid \boldsymbol{x}_1) = 0$ *for all* $\kappa \in [1, \infty)$*.*

Proof. See Appendix A.3. □

Remark 4.6 (Choosing $\boldsymbol{A}$)**.** The form of $\mathfrak{F}_{\text{Stable}}$ in Definition 4.4 grants one freedom to design conditional CNFs by manipulating $\boldsymbol{A}$. In contexts where data describes physically stable states, $\boldsymbol{A}$ could be used to incorporate local dynamics information. It is well-known that linearizing a VF about a hyperbolic equilibrium point yields a locally topologically conjugate approximation to the nonlinear dynamics (see e.g. Section 5 of (Ozaki, 1993; Meiss, 2007)). Thus, we could choose model $\boldsymbol{A}$ as the Jacobian of a VF or the Hessian of a potential function. However, it is well-known that these quantities will have zero eigenvalues when the VF is equivariant or the potential is invariant to some group action, see (Carr, 1981; Kœnig, 1997; Chossat & Lauterbach, 2000). Using a non-scalar form of $\boldsymbol{A}$ to incorporate local information in generative models has been suggested in (Yu et al., 2024; Singhal et al., 2024).

The conditional CNFs defined so far satisfy the invariance principle; however, since they stabilize to the stationary PDF at $t = \infty$, it is unclear what the time window should be for computing the likelihood of data samples under the model (see App. C of (Lipman et al., 2022)). One approach is to enforce the decay rate of the flow (via the eigenvalues) to guarantee convergence to $\boldsymbol{y}_1$ within some epsilon distance and some consistent time window. A more exact way is to define a transformation of the VF that describes the dynamics of the data state w.r.t. the interpolant state. We can do this by inverting the distance metric, i.e. instead of thinking of the data mean being y_1 when $t = \infty$, we can think of it when the interpolant mean is z_1. We identify this transformation in Theorem 4.7.

Theorem 4.7. *There exists a subset*

$$\mathfrak{F}_{\textit{Finite}} \subset \mathfrak{F}_{\textit{Auto}} \tag{26}$$

s.t. for all $(\cdot, (p, \boldsymbol{v}, \cdot)) \in \mathfrak{F}_{\textit{Stable}}$ *there exists a transformation* $(\cdot, (r, \boldsymbol{w}, \cdot)) \in \mathfrak{F}_{\textit{Finite}}$ *s.t.*

$$r(\boldsymbol{x}, t \mid \boldsymbol{x}_1) = p(\boldsymbol{x}, \tau(t) \mid \boldsymbol{x}_1) \tag{27}$$

$$\textit{s.t.} \quad \tau(t) = h^{-1}\left(h(0)\left(\frac{t - z_1}{z_0 - z_1}\right)\right) \tag{28}$$

where $r(\boldsymbol{x}, z_1 \mid \boldsymbol{x}_1) = p(\boldsymbol{x}, \infty \mid \boldsymbol{x}_1)$ *and furthermore*

$$\boldsymbol{w}_y(\boldsymbol{x}, t \mid \boldsymbol{x}_1) = \frac{\boldsymbol{v}_y(\boldsymbol{x}, \tau(t) \mid \boldsymbol{x}_1)}{v_z(\boldsymbol{x}, \tau(t) \mid \boldsymbol{x}_1)}, \quad w_z(\boldsymbol{x}, t \mid \boldsymbol{x}_1) = 1.$$

Proof. See Appendix A.4. □

We will now consider the simple case of the above CNFs when the chosen metric is the the Euclidean distance between the mean and the stationary mean. Note that the stable VF in Lemma 4.5 requires the distance metric to be invertible. To satisfy this, we use the bound (Kågström, 1977; Van Loan, 1977; Moler & Van Loan, 1978)

$$\begin{aligned} h(t) =& \|\boldsymbol{\mu}_y(t) - \boldsymbol{\mu}_y(\infty)\|_2 = \|e^{-\boldsymbol{A}t}(\boldsymbol{y}_0 - \boldsymbol{y}_1)\|_2 \\ \leq& \|e^{-\boldsymbol{A}t}\|_2 \|\boldsymbol{y}_0 - \boldsymbol{y}_1\|_2 \leq e^{-\alpha_{\min} t}\|\boldsymbol{y}_0 - \boldsymbol{y}_1\|_2, \end{aligned} \tag{29}$$

to characterize the worst-case distance, and define such a CNF in Definition 4.8.

Definition 4.8 (μ-Stable CNF Pair Space)**.** A subset

$$\mathfrak{F}_{\mu\text{-Stable}} \subset \mathfrak{F}_{\text{Stable}} \quad \text{s.t.} \quad h(t) = e^{-\alpha_{\min} t}\|\boldsymbol{y}_0 - \boldsymbol{y}_1\|_2 \tag{30}$$

s.t. $\alpha_{\min} \in \mathbb{R}_{>0}$ is the minimum non-zero eigenvalue of $\boldsymbol{A}$.

With a stable CNF based on this mean distance bound, we can apply the transformation in Theorem 4.7, and then get the autonomous forms from Theorem 4.2 for both the stable and finite versions, see Lemma 4.9.

Lemma 4.9. *For all* $(\cdot, (p, \boldsymbol{v}, \cdot)) \in \mathfrak{F}_{\mu\textit{-Stable}}$*:*

$$\begin{aligned} \bar{\boldsymbol{v}}_y(\boldsymbol{x} \mid \boldsymbol{x}_1) &= -\boldsymbol{A}\left(\boldsymbol{y} - \boldsymbol{y}_1 - \boldsymbol{\Sigma}_1\bar{\boldsymbol{\Sigma}}_y^{-1}(z)\left(\boldsymbol{y} - \bar{\boldsymbol{\mu}}_y(z)\right)\right) \\ \bar{v}_z(\boldsymbol{x} \mid \boldsymbol{x}_1) &= -\kappa\alpha_{\min}(z - z_1), \end{aligned} \tag{31}$$

and there exists a subset $\mathfrak{F}_{\mu\textit{-Finite}} \subset \mathfrak{F}_{\textit{Finite}}$ *s.t. for all corresponding transformations* $(\cdot, (r, \boldsymbol{w}, \cdot)) \in \mathfrak{F}_{\mu\textit{-Finite}}$*:*

$$\begin{aligned} \bar{\boldsymbol{w}}_y(\boldsymbol{x} \mid \boldsymbol{x}_1) &= \frac{\boldsymbol{A}\left(\boldsymbol{y} - \boldsymbol{y}_1 - \boldsymbol{\Sigma}_1\bar{\boldsymbol{\Sigma}}_y^{-1}(z)\left(\boldsymbol{y} - \bar{\boldsymbol{\mu}}_y(z)\right)\right)}{\kappa\alpha_{\min}(z - z_1)} \\ \bar{w}_z(\boldsymbol{x} \mid \boldsymbol{x}_1) &= 1 \end{aligned} \tag{32}$$

where $\bar{p}_y(\boldsymbol{x} \mid \boldsymbol{y}_1) = \bar{r}_y(\boldsymbol{x} \mid \boldsymbol{y}_1)$ *are equivalently given by*

$$\begin{aligned} \bar{\boldsymbol{\mu}}_y(z) =& \boldsymbol{y}_1 + \left(\frac{z - z_1}{z_0 - z_1}\right)^{\frac{\boldsymbol{A}}{\kappa\alpha_{\min}}}(\boldsymbol{y}_0 - \boldsymbol{y}_1) \\ \bar{\boldsymbol{\Sigma}}_y(z) =& \boldsymbol{\Sigma}_1 + \left(\frac{z - z_1}{z_0 - z_1}\right)^{\frac{2\boldsymbol{A}}{\kappa\alpha_{\min}}}(\boldsymbol{\Sigma}_0 - \boldsymbol{\Sigma}_1) \end{aligned} \tag{33}$$

Lipschitz constants Note that the autonomous form of $\mathfrak{F}_{\mu\text{-Finite}}$ can be simply obtained by dividing the VF of $\mathfrak{F}_{\mu\text{-Stable}}$ by its own interpolant state's VF. Thus, one could learn with $\mathfrak{F}_{\mu\text{-Stable}}$ and straightforwardly attain the attributes of $\mathfrak{F}_{\mu\text{-Finite}}$; however the converse is not true. Aside from stability, learning with $\mathfrak{F}_{\mu\text{-Stable}}$ may present some advantages in terms of its VF's Lipschitz constant (LC). The LC of the drift of the SDE for $\mathfrak{F}_{\mu\text{-Stable}}$ yields $\max\{\alpha_{\max}, \kappa\alpha_{\min}\}$, while for $\mathfrak{F}_{\mu\text{-Stable}}$ the LC depends on $\boldsymbol{y}_1$. Since the marginal VF is a convex combination of conditional VFs (see Lemma 4.10), the LC for the marginal VF will be the greatest LC of all the conditional VFs. Thus, for $\mathfrak{F}_{\mu\text{-Finite}}$, its marginal VF's LC depends on the scale of the dataset, while for $\mathfrak{F}_{\mu\text{-Stable}}$ it does not, unless $\boldsymbol{A}$ is chosen based on $\boldsymbol{y}_1$ (see Remark 4.6).

Lemma 4.10 (Differential Inclusion)**.** *For all* $((p, \boldsymbol{v}, \boldsymbol{\phi}), (p_1, \boldsymbol{v}_1, \boldsymbol{\phi}_1)) \in \mathfrak{F}_{FM}$*:*

$$\boldsymbol{v}(\boldsymbol{x}, t) \in \operatorname{co}\left\{\boldsymbol{v}_1(\boldsymbol{x}, t \mid \boldsymbol{x}_1) \mid \boldsymbol{x}_1 \in \operatorname{supp}(q)\right\}, \tag{34}$$

where co *is the convex hull operator.*

Proof. The statement is evident from the fact that

$$\int_{\mathcal{X}} \frac{p_1(\boldsymbol{x}, t \mid \boldsymbol{x}_1) q(\boldsymbol{x}_1) \mathrm{d}\boldsymbol{x}_1}{p(\boldsymbol{x}, t)} = 1. \tag{35}$$

□

Remark 4.11. Differential inclusions (Filippov, 1960) occur frequently in the hybrid dynamical systems literature (Goebel et al., 2009; Sanfelice, 2021). E.g. in switched systems (Liberzon & Morse, 1999), it represents arbitrary switching between different VFs. In discontinuous dynamical systems (Cortes, 2008; Sprague & Ögren, 2021), it represents sliding dynamics between VFs. Stability results for differential inclusions have been explored in (Boyd et al., 1994; Molchanov & Pyatnitskiy, 1989), and in (Hafez & Broucke, 2022; Veer & Poulakakis, 2019; Alpcan & Basar, 2010) when multiple equilibria are present. These results are highly relevant as the marignal VF is a convex combination of linear VFs (see Lemma 4.10). At present, these ideas have only appeared in (Zhu & Lin, 2024).

We will now make a direct comparison to OT FM of (Lipman et al., 2022). We start by defining an adaptation of OT FM in Definition 4.12.

Definition 4.12 (OT CNF Pair Space)**.** A subset

$$\mathfrak{F}_{\text{OT}} \subset \mathfrak{F}_{\text{Auto}} \tag{36}$$

s.t. for all $(\cdot, (p, \boldsymbol{v}, \cdot)) \in \mathfrak{F}_{\text{OT}}$ we have $\mu_z(t) = t$ and

$$\boldsymbol{\mu}_y(t) = t\boldsymbol{y}_1, \quad \boldsymbol{\Sigma}_y(t) = \left(1 - (1 - \sigma_{\min})t\right)^2 \boldsymbol{I} \tag{37}$$

and furthermore:

$$\boldsymbol{v}_y(\boldsymbol{x}, t \mid \boldsymbol{x}_1) = \frac{\boldsymbol{y}_1 - (1 - \sigma_{\min})\boldsymbol{y}}{1 - (1 - \sigma_{\min})t}, \quad v_z(\boldsymbol{x}, t \mid \boldsymbol{x}_1) = 1 \tag{38}$$

In Lemma 4.13 we show that the mean of the autonomous conditional PDF path (see Theorem 4.2) of $\mathfrak{F}_{\mu\text{-Finite}}$ is the same as for $\mathfrak{F}_{\text{OT}}$, while the covariances differ slightly but are the same at the endpoints z_0, z_1.

Lemma 4.13 (Optimal Transport FM Comparison)**.** *For all* $(\cdot, (p, \cdot, \cdot)) \in \mathfrak{F}_{\mu\text{-Finite}}$ *s.t.*

$$\boldsymbol{A} = \alpha \boldsymbol{I}, \quad \kappa = 1, \quad \boldsymbol{y}_0 = \boldsymbol{0}, \quad \boldsymbol{\Sigma}_0 = \boldsymbol{I}, \tag{39}$$

$$\boldsymbol{\Sigma}_1 = \sigma_{\min}^2 \boldsymbol{I}, \quad z_0 = 0, \quad z_1 = 1 \tag{40}$$

with $\bar{p}_y$ *given by* $\bar{\boldsymbol{\mu}}_y(z)$ *and* $\bar{\boldsymbol{\Sigma}}_y(z)$ *and for all* $(\cdot, (r, \cdot, \cdot)) \in \mathfrak{F}_{OT}$ *with* $\bar{r}_y$ *given by* $\bar{\boldsymbol{\mu}}'_y(z)$ *and* $\bar{\boldsymbol{\Sigma}}'_y(z)$*, we have that*

$$\bar{\boldsymbol{\mu}}_y(z) = \bar{\boldsymbol{\mu}}'_y(z), \quad \bar{\boldsymbol{\Sigma}}_y(0) = \bar{\boldsymbol{\Sigma}}'_y(0), \quad \bar{\boldsymbol{\Sigma}}_y(1) = \bar{\boldsymbol{\Sigma}}'_y(1), \tag{41}$$

and furthermore

$$\max_{z \in [0,1]} \left\| \bar{\boldsymbol{\Sigma}}_y(z) - \bar{\boldsymbol{\Sigma}}'_y(z) \right\|_2 = \frac{\sigma_{\min} |\sigma_{\min} - 1|}{2} \tag{42}$$

at $z = 1/2$.

Note that a Guassian PDF path can be described by a linear time-varying SDE (see Sec. 6.1 of (Särkkä & Solin, 2019)). Thus, it may be possible to identify a different diffusion coefficient in the SDE in Equation (19) by matching Eq. 6.2 in (Särkkä & Solin, 2019) with the time-derivative of the mean and covariance of $\mathfrak{F}_{\text{OT}}$.

4.3. Experiments and Discussion

Conditional Flows In this work, we first consider stable CNF flow with the mean distance metric (referred to as *Stable-FM*), $\mathfrak{F}_{\mu\text{-Stable}}$ from Equation (31) in Lemma 4.9 with $\boldsymbol{A} = \alpha \boldsymbol{I}$, $\alpha = 4$, $\boldsymbol{y}_0 = \boldsymbol{0}$, $\boldsymbol{\Sigma}_0 = \boldsymbol{I}$, $\boldsymbol{\Sigma}_1 = (10^{-5})^2 \boldsymbol{I}$, $z_0 = 0$, $z_1 = 1$. Note that $\alpha = 4$ is chosen so that $\|\mu_y(1) - \boldsymbol{y}_1\|_2 \leq 10^{-4}$ to converge at roughly the same time as the OT flow. We consider OT FM $\mathfrak{F}_{\text{OT}}$ flow (referred to as *OT-FM*) from Definition 4.12 with $\sigma_{\min} = 10^{-5}$, meaning $\boldsymbol{\Sigma}_1 = (10^{-5})^2 \boldsymbol{I}$, as in (Lipman et al., 2022).

Marginal Flows With the autonomous form in Theorem 4.2, we model the marginal VF $\boldsymbol{v}_\theta : \mathcal{X} \to \mathbb{R}^n$ as both an arbitrary function with a feed-forward architecture (referred to as *simple*) and a *gradient field* $\boldsymbol{v}_\theta(\boldsymbol{x}) = -\nabla_{\boldsymbol{x}} H_\theta(\boldsymbol{x})$ where $H_\theta : \mathcal{X} \to \mathbb{R}_{\geq 0}$ is an energy function (using H in spirit of the Hamiltonian). Note that the gradient field model is curl-free, and thus is unique in the context of the

continuity equation (does not add any redundant divergence-free components). Also note that the gradient field model straightforwardly satisfies the invariance principle in Theorem 3.5 with

$$\mathcal{B}_0 = \mathcal{X}, \quad \mathcal{B}_1 = \{\boldsymbol{x} \in \mathcal{X} | \nabla_{\boldsymbol{x}} H(\boldsymbol{x}) = 0\}, \tag{43}$$

where $\boldsymbol{\phi}(\boldsymbol{x}_0, \infty) \in \mathcal{B}_1$ for all $\boldsymbol{x}_0 \in \mathcal{B}_0$. The simple model does not possess these guarantees, although it is easier to train.

Experiments We test the combinations of the aforementioned conditional and marginal flow models on generating data from the moons distribution. As can be seen in Figure 1 and Figure 2 the invariance principle of Theorem 3.5 is strongly reflected: after $t = 1$ the OT-FM flow diverges, while the Stable-FM remains stable. We also plot the maximum mean discrepancy (MMD) between inference flows and the dataset over time in Figure 3. As can be seen, the MMD of both Stable-FM and OT-FM approach zero; however, OT-FM only does so momentarily, while Stable-FM approaches zero MMD asymptotically. One can also see that the MMD curves are similar for both marginal flow models (simple and gradient field), however the invariance guarantees of the gradient field still stand.

Note on invariance If we modeled the marginal VF as a non-autonomous gradient field $v_\theta(\boldsymbol{x}, t) = -\nabla_{\boldsymbol{x}} H_\theta(\boldsymbol{x}, t)$ we would need to establish conditions on $\partial H_\theta(\boldsymbol{x}, t)/\partial t$ to satisfy the invariance principles (see Sec. 2 of (Mao, 1999)). To ensure non-positivity of $\partial H_\theta(\boldsymbol{x}, t)/\partial t$ we initially explored using partially input-convex neural networks (ICNNs) (Amos et al., 2017), but did not have success.

Note on energy Recall from Remark 4.6 that $\boldsymbol{A}$ could be defined based on each data sample $\boldsymbol{y}_1$ However, if $\boldsymbol{A}$ is held constant for every $\boldsymbol{y}_1$, then the VF for the interpolant state $\bar{v}_z$ will be constant for all data samples. In this case, one may consider simply setting $-\nabla_z H(\boldsymbol{y}, z) = -\alpha_{\min}(z - z_1)$; however, this would imply separability of the form

$$H_\theta(\boldsymbol{y}, z) = H_y(\boldsymbol{y}) + \frac{\alpha_{\min}}{2}(z - z_1)^2 \tag{44}$$

which would imply the inability to correlate $\boldsymbol{y}$ and $\boldsymbol{z}$.

Note on idempotence Due to the invariance principle, the flow map of the conditional flows queried at $t = \infty$ always yields the same result, i.e. $\boldsymbol{\phi}(\boldsymbol{x}_0, \infty \mid \boldsymbol{x}_1) \in \mathcal{B}_1$ for all $\boldsymbol{x}_0 \in \mathcal{B}_0$ almost surely (see Theorem 3.5), where $\mathcal{B}_1 \supseteq \{\boldsymbol{x}_1\}$ is an ellipsoid centered at $\boldsymbol{x}_1$ described by the stationary distribution $\mathcal{N}(\boldsymbol{x}_1, \boldsymbol{\Sigma}_1)$ and the eigenvectors of $\boldsymbol{A}$. Thus the flow map may be seen as stochastically indempotent w.r.t. the starting set $\mathcal{B}_0$: $\boldsymbol{\phi}_\infty(\boldsymbol{\phi}_\infty(\boldsymbol{x}_0)) \in \mathcal{B}_1$ for all $\boldsymbol{x}_0 \in \mathcal{B}_0$, where $\boldsymbol{\phi}_\infty(\boldsymbol{x}) = \boldsymbol{\phi}(\boldsymbol{x}, \infty \mid \boldsymbol{x}_1)$. This result is relevant to Idempotent Generative Networks (Shocher et al., 2023) and Consistency Models (Song et al., 2023).

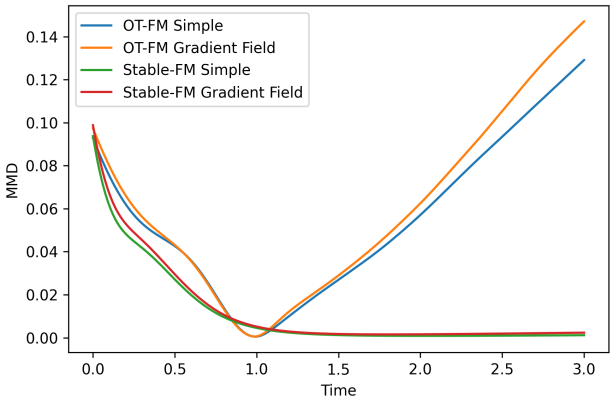

Figure 3. Maximum mean discrepancy (MMD) between the generated samples and the moons dataset over time. The models were trained with 500 data points. The MMD was computed for 200 time steps between $t = 0$ and $t = 3$ with 1000 generated samples and 10000 dataset samples. "Simple" refers to an arbitrary architecture as described above. "Gradient field" refers to when the architecture is the gradient of a scalar function. Here, the time input to OT-FM is clipped at $t = 1$.

5. Conclusion

In this paper we applied a stochastic version of La Salle's invariance principle (La Salle, 1966; Mao, 1999) to FM (Lipman et al., 2022) to enforce stability of the model's flows to the support of the target distribution, a desirable property when the data represents a physically stable state. In doing so, we showed how to render CNFs autonomous by introducing an extra state variable z and performing interpolation over it instead of time. Additionally, we showed how to construct stable CNFs and we showed how they and their PDF trajectories compare to the OT CNF from (Lipman et al., 2022). Lastly, we have made several connections from control theory to generative modeling. Overall, we demonstrated that our approach is effective with theoretical and experimental results.

Acknowledgments

We thank Ricky Chen, Ruibo Tu, Henrik Sandberg, and Dimos Dimarogonas for insightful discussions. This work was partially supported by the Wallenberg AI, Autonomous Systems and Software Program (WASP) funded by the Knut and Alice Wallenberg Foundation.

References

Alpcan, T. and Basar, T. A stability result for switched systems with multiple equilibria. *Dynamics of Continuous, Discrete and Impulsive Systems Series A: Mathematical Analysis*, 17(4):949–958, 2010.

Ambrosio, L., Gigli, N., and Savaré, G. *Gradient flows: in metric spaces and in the space of probability measures*. Springer Science & Business Media, 2005.

Amos, B., Xu, L., and Kolter, J. Z. Input convex neural networks. In *International Conference on Machine Learning*, pp. 146–155. PMLR, 2017.

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

Blattmann, A., Rombach, R., Ling, H., Dockhorn, T., Kim, S. W., Fidler, S., and Kreis, K. Align your latents: High-resolution video synthesis with latent diffusion models. *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 22563–22575, 2023.

Bortoli, V. D., Mathieu, E., Hutchinson, M. J., Thornton, J., Teh, Y. W., and Doucet, A. Riemannian score-based generative modelling. In *Neural Information Processing Systems*, 2022.

Bose, A. J., Akhound-Sadegh, T., Fatras, K., Huguet, G., Rector-Brooks, J., Liu, C.-H., Nica, A. C., Korablyov, M., Bronstein, M., and Tong, A. Se(3)-stochastic flow matching for protein backbone generation. *ArXiv*, abs/2310.02391, 2023.

Boyd, S., El Ghaoui, L., Feron, E., and Balakrishnan, V. *Linear matrix inequalities in system and control theory*. SIAM, 1994.

Carr, J. Applications of centre manifold theory, volume 35 of. *Applied Mathematical Sciences*, 1981.

Chen, R. T. and Lipman, Y. Riemannian flow matching on general geometries. *arXiv preprint arXiv:2302.03660*, 2023a.

Chen, R. T. Q. and Lipman, Y. Riemannian flow matching on general geometries. *ArXiv*, abs/2302.03660, 2023b.

Chen, T. Q., Rubanova, Y., Bettencourt, J., and Duvenaud, D. K. Neural ordinary differential equations. In *Neural Information Processing Systems*, 2018.

Chossat, P. and Lauterbach, R. *Methods in equivariant bifurcations and dynamical systems*, volume 15. World Scientific Publishing Company, 2000.

Corso, G., Jing, B., Barzilay, R., Jaakkola, T., et al. Diffdock: Diffusion steps, twists, and turns for molecular docking. In *International Conference on Learning Representations (ICLR 2023)*, 2023.

Cortes, J. Discontinuous dynamical systems. *IEEE Control systems magazine*, 28(3):36–73, 2008.

Dhariwal, P. and Nichol, A. Diffusion models beat gans on image synthesis. *Advances in neural information processing systems*, 34:8780–8794, 2021.

Dill, K. A. Theory for the folding and stability of globular proteins. *Biochemistry*, 24(6):1501–1509, 1985.

Dockhorn, T., Vahdat, A., and Kreis, K. Score-based generative modeling with critically-damped langevin diffusion. In *International Conference on Learning Representations*, 2021.

Esser, P., Chiu, J., Atighehchian, P., Granskog, J., and Germanidis, A. Structure and content-guided video synthesis with diffusion models. *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 7312–7322, 2023.

Feng, R., Zhu, Q., Tran, H., Chen, B., Toland, A., Ramprasad, R., and Zhang, C. May the force be with you: Unified force-centric pre-training for 3d molecular conformations. In *Thirty-seventh Conference on Neural Information Processing Systems*, 2023.

Filippov, A. F. Differential equations with discontinuous right-hand side. *Matematicheskii sbornik*, 93(1):99–128, 1960.

Goebel, R., Sanfelice, R. G., and Teel, A. R. Hybrid dynamical systems. *IEEE control systems magazine*, 29(2): 28–93, 2009.

Hafez, M. A. and Broucke, M. E. Stability of discrete-time switched systems with multiple equilibria using a common quadratic lyapunov function. *IEEE Control Systems Letters*, 6:2497–2502, 2022.

Haussmann, U. G. and Pardoux, É. Time reversal of diffusions. *Annals of Probability*, 14:1188–1205, 1986.

Ho, J., Salimans, T., Gritsenko, A., Chan, W., Norouzi, M., and Fleet, D. J. Video diffusion models. *ArXiv*, abs/2204.03458, 2022.

Hyvärinen, A. Estimation of non-normalized statistical models by score matching. *J. Mach. Learn. Res.*, 6:695–709, 2005.

Kågström, B. Bounds and perturbation bounds for the matrix exponential. *BIT Numerical Mathematics*, 17:39–57, 1977.

Kang, Q., Song, Y., Ding, Q., and Tay, W. P. Stable neural ode with lyapunov-stable equilibrium points for defending against adversarial attacks. *Advances in Neural Information Processing Systems*, 34:14925–14937, 2021.

Ketata, M. A., Laue, C., Mammadov, R. A., Stärk, H., Wu, M., Corso, G., Marquet, C., Barzilay, R., and Jaakkola, T. Diffdock-pp: Rigid protein-protein docking with diffusion models. *ArXiv*, abs/2304.03889, 2023.

Khalil, H. K. *Nonlinear Systems*. 2002.

Khasminskii, R. Z. Stochastic stability of differential equations. 1980.

Klein, L., Krämer, A., and Noe, F. Equivariant flow matching. In *Thirty-seventh Conference on Neural Information Processing Systems*, 2023.

Kœnig, M. Linearization of vector fields on the orbit space of the action of a compact lie group. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 121, pp. 401–424. Cambridge University Press, 1997.

Kolter, J. Z. and Manek, G. Learning stable deep dynamics models. *Advances in neural information processing systems*, 32, 2019.

Kushner, H. J. On the stability of stochastic dynamical systems. *Proceedings of the National Academy of Sciences of the United States of America*, 53 1:8–12, 1965.

La Salle, J. P. An invariance principle in the theory of stability. Technical report, 1966.

Lawrence, N., Loewen, P., Forbes, M., Backstrom, J., and Gopaluni, B. Almost surely stable deep dynamics. *Advances in Neural Information Processing Systems*, 33: 18942–18953, 2020.

Le, M., Vyas, A., Shi, B., Karrer, B., Sari, L., Moritz, R., Williamson, M., Manohar, V., Adi, Y., Mahadeokar, J., et al. Voicebox: Text-guided multilingual universal speech generation at scale. In *Thirty-seventh Conference on Neural Information Processing Systems*, 2023.

Liberzon, D. and Morse, A. S. Basic problems in stability and design of switched systems. *IEEE control systems magazine*, 19(5):59–70, 1999.

Lindquist, A. and Picci, G. On the stochastic realization problem. *Siam Journal on Control and Optimization*, 17: 365–389, 1979.

Lipman, Y., Chen, R. T. Q., Ben-Hamu, H., Nickel, M., and Le, M. Flow matching for generative modeling. *ArXiv*, abs/2210.02747, 2022.

Liu, G.-H., Vahdat, A., Huang, D.-A., Theodorou, E. A., Nie, W., and Anandkumar, A. I2sb: Image-to-image schrödinger bridge. In *International Conference on Machine Learning*, 2023.

Luo, S., Shi, C., Xu, M., and Tang, J. Predicting molecular conformation via dynamic graph score matching. In *Neural Information Processing Systems*, 2021.

Mao, X. Stochastic versions of the lasalle theorem. *Journal of differential equations*, 153(1):175–195, 1999.

Maoutsa, D., Reich, S., and Opper, M. Interacting particle solutions of fokker–planck equations through gradient–log–density estimation. *Entropy*, 22, 2020.

Meiss, J. D. *Differential dynamical systems*. SIAM, 2007.

Molchanov, A. P. and Pyatnitskiy, Y. S. Criteria of asymptotic stability of differential and difference inclusions encountered in control theory. *Systems & Control Letters*, 13(1):59–64, 1989.

Moler, C. and Van Loan, C. Nineteen dubious ways to compute the exponential of a matrix. *SIAM review*, 20(4): 801–836, 1978.

Ozaki, T. A local linearization approach to nonlinear filtering. *International Journal of Control*, 57(1):75–96, 1993.

Parr, R. G., Gadre, S. R., and Bartolotti, L. J. Local density functional theory of atoms and molecules. *Proceedings of the National Academy of Sciences*, 76(6):2522–2526, 1979.

Pooladian, A.-A., Ben-Hamu, H., Domingo-Enrich, C., Amos, B., Lipman, Y., and Chen, R. T. Q. Multisample flow matching: Straightening flows with minibatch couplings. In *International Conference on Machine Learning*, 2023.

Rodriguez, I. D. J., Ames, A., and Yue, Y. Lyanet: A lyapunov framework for training neural odes. In *International Conference on Machine Learning*, 2022.

Rombach, R., Blattmann, A., Lorenz, D., Esser, P., and Ommer, B. High-resolution image synthesis with latent diffusion models. *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 10674–10685, 2021.

Sanfelice, R. G. *Hybrid feedback control*. Princeton University Press, 2021.

Särkkä, S. and Solin, A. Applied stochastic differential equations. 2019.

Shaul, N., Chen, R. T., Nickel, M., Le, M., and Lipman, Y. On kinetic optimal probability paths for generative models. In *International Conference on Machine Learning*, pp. 30883–30907. PMLR, 2023.

Shi, C., Luo, S., Xu, M., and Tang, J. Learning gradient fields for molecular conformation generation. In *International Conference on Machine Learning*, 2021.

Shocher, A., Dravid, A., Gandelsman, Y., Mosseri, I., Rubinstein, M., and Efros, A. A. Idempotent generative network. *arXiv preprint arXiv:2311.01462*, 2023.

Singhal, R., Goldstein, M., and Ranganath, R. What's the score? automated denoising score matching for nonlinear diffusions. *arXiv preprint arXiv:2407.07998*, 2024.

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

Song, Y., Garg, S., Shi, J., and Ermon, S. Sliced score matching: A scalable approach to density and score estimation. In *Conference on Uncertainty in Artificial Intelligence*, 2019.

Song, Y., Sohl-Dickstein, J., Kingma, D. P., Kumar, A., Ermon, S., and Poole, B. Score-based generative modeling through stochastic differential equations. In *International Conference on Learning Representations*, 2020.

Song, Y., Dhariwal, P., Chen, M., and Sutskever, I. Consistency models. *arXiv preprint arXiv:2303.01469*, 2023.

Sprague, C. I. and Ögren, P. Continuous-time behavior trees as discontinuous dynamical systems. *IEEE Control Systems Letters*, 6:1891–1896, 2021.

Sun, Z., Anderson, B. D., Deghat, M., and Ahn, H.-S. Rigid formation control of double-integrator systems. *International Journal of Control*, 90(7):1403–1419, 2017.

Tong, A., Malkin, N., Fatras, K., Atanackovic, L., Zhang, Y., Huguet, G., Wolf, G., and Bengio, Y. Simulation-free schrödinger bridges via score and flow matching. In *ICML Workshop on New Frontiers in Learning, Control, and Dynamical Systems*, 2023.

Vahdat, A., Kreis, K., and Kautz, J. Score-based generative modeling in latent space. In *Neural Information Processing Systems*, 2021.

Van Loan, C. The sensitivity of the matrix exponential. *SIAM Journal on Numerical Analysis*, 14(6):971–981, 1977.

Veer, S. and Poulakakis, I. Switched systems with multiple equilibria under disturbances: Boundedness and practical stability. *IEEE Transactions on Automatic Control*, 65 (6):2371–2386, 2019.

Villani, C. et al. *Optimal transport: old and new*, volume 338. Springer, 2009.

Vincent, P. A connection between score matching and denoising autoencoders. *Neural Computation*, 23:1661–1674, 2011.

Xu, Y., Liu, Z., Tegmark, M., and Jaakkola, T. Poisson flow generative models. *Advances in Neural Information Processing Systems*, 35:16782–16795, 2022.

Xu, Y., Liu, Z., Tian, Y., Tong, S., Tegmark, M., and Jaakkola, T. Pfgm++: Unlocking the potential of physics-inspired generative models. In *International Conference on Machine Learning*, 2023.

Yim, J., Campbell, A., Foong, A. Y. K., Gastegger, M., Jiménez-Luna, J., Lewis, S., Satorras, V. G., Veeling, B. S., Barzilay, R., Jaakkola, T., and Noé, F. Fast protein backbone generation with se(3) flow matching. 2023a.

Yim, J., Trippe, B. L., Bortoli, V. D., Mathieu, E., Doucet, A., Barzilay, R., and Jaakkola, T. Se(3) diffusion model with application to protein backbone generation. *ArXiv*, abs/2302.02277, 2023b.

Yu, X., Gu, X., Liu, H., and Sun, J. Constructing non-isotropic gaussian diffusion model using isotropic gaussian diffusion model for image editing. *Advances in Neural Information Processing Systems*, 36, 2024.

Zaidi, S., Schaarschmidt, M., Martens, J., Kim, H., Teh, Y. W., Sanchez-Gonzalez, A., Battaglia, P., Pascanu, R., and Godwin, J. Pre-training via denoising for molecular property prediction. In *NeurIPS 2022 AI for Science: Progress and Promises*, 2022.

Zhang, J., Mohammadi, H. B., and Rozo, L. D. Learning riemannian stable dynamical systems via diffeomorphisms. In *Conference on Robot Learning*, 2022a.

Zhang, J., Zhu, Q., and Lin, W. Neural stochastic control. *Advances in Neural Information Processing Systems*, 35: 9098–9110, 2022b.

Zhu, Q. and Lin, W. Switched flow matching: Eliminating singularities via switching odes. *arXiv preprint arXiv:2405.11605*, 2024.

A. Proofs

A.1. Proof of Theorem 4.2

We start with the definition of the marginal PDF trajectory from Definition 3.2:

$$p(\boldsymbol{x}, t) = \int_{\mathcal{X}} p(\boldsymbol{x}, t \mid \boldsymbol{x}_1) q(\boldsymbol{x}_1) \mathrm{d}\boldsymbol{x}_1. \tag{45}$$

With Definition 4.1 and recalling $\boldsymbol{x} = [\boldsymbol{y}^\top, z]$ and $\boldsymbol{x}_1 = [\boldsymbol{y}_1^\top, z_1]$ we can write

$$\begin{aligned} p(\boldsymbol{x}, t) &= \int_{\mathcal{Y}} \int_{\mathcal{Z}} p_y(\boldsymbol{y}, t \mid \boldsymbol{y}_1) p_z(z, t \mid z_1) q(\boldsymbol{y}_1) q(z_1) \mathrm{d}\boldsymbol{y}_1 \mathrm{d}z_1 && (46) \\ &= \int_{\mathcal{Y}} p_y(\boldsymbol{y}, t \mid \boldsymbol{y}_1) q(\boldsymbol{y}_1) \mathrm{d}\boldsymbol{y}_1 \int_{\mathcal{Z}} p_z(z, t \mid z_1) q(z_1) \mathrm{d}z_1 && (47) \\ &= \int_{\mathcal{Y}} p_y(\boldsymbol{y}, t \mid \boldsymbol{y}_1) q(\boldsymbol{y}_1) \mathrm{d}\boldsymbol{y}_1 \delta(z - \phi_z(z_0, t \mid z_1)) && (48) \end{aligned}$$

where the data state's marginal PDF path and the interpolant state's marginal PDF path are independent. Since the mean of the interpolant state's marginal PDF path is bijective in time $t = \phi_z^{-1}(z_0, \phi_z(z_0, t \mid z_1) \mid z_1)$, we can write

$$\begin{aligned} \bar{p}(\boldsymbol{x}) &= p(\boldsymbol{x}, \phi_z^{-1}(z_0, z \mid z_1)) && (49) \\ &= \int_{\mathcal{Y}} p_y(\boldsymbol{y}, \phi_z^{-1}(z_0, z \mid z_1) \mid \boldsymbol{y}_1) q(\boldsymbol{y}_1) \mathrm{d}\boldsymbol{y}_1 && (50) \\ &= \int_{\mathcal{Y}} \bar{p}_y(\boldsymbol{x} \mid \boldsymbol{y}_1) q(\boldsymbol{y}_1) \mathrm{d}\boldsymbol{y}_1. && (51) \end{aligned}$$

We continue with the definition of the marginal VF from Definition 3.2:

$$\boldsymbol{v}(\boldsymbol{x}, t) = \frac{1}{p(\boldsymbol{x}, t)} \int_{\mathcal{X}} \boldsymbol{v}_1(\boldsymbol{x}, t \mid \boldsymbol{x}_1) p_1(\boldsymbol{x}, t \mid \boldsymbol{x}_1) q(\boldsymbol{x}_1) \mathrm{d}\boldsymbol{x}_1. \tag{52}$$

With Definition 4.1 we can write

$$\begin{aligned} \boldsymbol{v}(\boldsymbol{x}, t) &= \frac{1}{p(\boldsymbol{x}, t)} \int_{\mathcal{Y}} \int_{\mathcal{Z}} \boldsymbol{v}_1(\boldsymbol{x}, t \mid \boldsymbol{x}_1) p_y(\boldsymbol{y}, t \mid \boldsymbol{y}_1) p_z(z, t \mid z_1) q(\boldsymbol{y}_1) q(z_1) \mathrm{d}\boldsymbol{y}_1 \mathrm{d}z_1 && (53) \\ &= \frac{1}{p(\boldsymbol{x}, t)} \int_{\mathcal{Y}} \boldsymbol{v}_1(\boldsymbol{x}, t \mid \boldsymbol{x}_1) p_y(\boldsymbol{y}, t \mid \boldsymbol{y}_1) q(\boldsymbol{y}_1) \mathrm{d}\boldsymbol{y}_1 \delta(z - \phi_z(z_0, t \mid z_1)) && (54) \end{aligned}$$

where the data state's marginal VF and the interpolant state's marginal VF are decoupled. Using the bijectivity of the mean of the interpolant state's marginal PDF path, we can write

$$\begin{aligned} \bar{\boldsymbol{v}}(\boldsymbol{x}) &= \boldsymbol{v}(\boldsymbol{x}, \phi_z^{-1}(z_0, z \mid z_1)) && (55) \\ &= \frac{1}{\bar{p}(\boldsymbol{x})} \int_{\mathcal{Y}} \boldsymbol{v}_1(\boldsymbol{x}, \phi_z^{-1}(z_0, z \mid z_1) \mid \boldsymbol{x}_1) p_y(\boldsymbol{y}, \phi_z^{-1}(z_0, z \mid z_1) \mid \boldsymbol{y}_1) q(\boldsymbol{y}_1) \mathrm{d}\boldsymbol{y}_1 && (56) \\ &= \frac{1}{\bar{p}(\boldsymbol{x})} \int_{\mathcal{Y}} \bar{\boldsymbol{v}}_1(\boldsymbol{x} \mid \boldsymbol{x}_1) \bar{p}_y(\boldsymbol{x} \mid \boldsymbol{y}_1) q(\boldsymbol{y}_1) \mathrm{d}\boldsymbol{y}_1, && (57) \end{aligned}$$

where $\bar{p}(\boldsymbol{x})$ is given above.

A.2. Lemma A.1

Lemma A.1 (Transformed CNF Loss)**.** *For all $\boldsymbol{v}_\theta \in C(\mathcal{X}, \mathbb{R}^n)$ and $(\cdot, (p_1, \boldsymbol{v}_1, \boldsymbol{\phi}_1)) \in \mathfrak{F}_{\text{Auto}}$:*

$$L_{FM}(\theta) = \mathop{\mathbb{E}}_{\substack{z \sim \mathcal{U}[z_0, z_1] \\ \boldsymbol{y}_1 \sim q(\boldsymbol{x}_1) \\ \boldsymbol{y} \sim \bar{p}_y(\boldsymbol{y} \mid \boldsymbol{y}_1)}} \frac{\|\bar{\boldsymbol{v}}_\theta(\boldsymbol{x}) - \bar{\boldsymbol{v}}_1(\boldsymbol{x} \mid \boldsymbol{x}_1)\|^2}{\bar{v}_z(z \mid z_1)} \tag{58}$$

$$\text{s.t} \quad \begin{aligned} &\bar{\boldsymbol{v}}_\theta(\boldsymbol{x}) = \boldsymbol{v}_\theta(\boldsymbol{x}, \phi_z^{-1}(z_0, z \mid z_1)) \\ &z_0 = \phi_z(z_0, 0 \mid z_1) \quad z_1 = \phi_z(z_0, T \mid z_1). \end{aligned} \tag{59}$$

Proof. We start with the loss function in Theorem 3.3:

$$L_{\text{FM}}(\theta) = \mathop{\mathbb{E}}_{\substack{t\sim\mathcal{U}[0,T]\\ \boldsymbol{x}_1\sim q(\boldsymbol{x}_1)\\ \boldsymbol{x}\sim p_1(\boldsymbol{x},t|\boldsymbol{x}_1)}} \|\boldsymbol{v}_\theta(\boldsymbol{x},t) - \boldsymbol{v}_1(\boldsymbol{x},t \mid \boldsymbol{x}_1)\|^2 \tag{60}$$

$$= \frac{1}{T}\int_0^T \int_{\mathcal{X}}\int_{\mathcal{X}} \|\boldsymbol{v}_\theta(\boldsymbol{x},t) - \boldsymbol{v}_1(\boldsymbol{x},t \mid \boldsymbol{x}_1)\|^2 p_1(\boldsymbol{x},t \mid \boldsymbol{x}_1) q(\boldsymbol{x}_1) \mathrm{d}\boldsymbol{x}\mathrm{d}\boldsymbol{x}_1\mathrm{d}t. \tag{61}$$

With Definition 4.1 we can write

$$L_{\text{FM}}(\theta) = \frac{1}{T}\int_0^T \int_{\mathcal{Y}}\int_{\mathcal{Z}}\int_{\mathcal{Y}}\int_{\mathcal{Z}} \|\boldsymbol{v}_\theta(\boldsymbol{x},t) - \boldsymbol{v}_1(\boldsymbol{x},t \mid \boldsymbol{x}_1)\|^2 p_y(\boldsymbol{y},t \mid \boldsymbol{y}_1) p_z(z,t \mid z_1) q(\boldsymbol{y}_1) q(z_1) \mathrm{d}\boldsymbol{y}\mathrm{d}z\mathrm{d}\boldsymbol{y}_1\mathrm{d}z_1\mathrm{d}t \tag{62}$$

$$= \frac{1}{T}\int_0^T \int_{\mathcal{Y}}\int_{\mathcal{Z}}\int_{\mathcal{Y}} \|\boldsymbol{v}_\theta(\boldsymbol{x},t) - \boldsymbol{v}_1(\boldsymbol{x},t \mid \boldsymbol{x}_1)\|^2 p_y(\boldsymbol{y},t \mid \boldsymbol{y}_1) q(\boldsymbol{y}_1) \delta(z - \phi_z(z_0,t \mid z_1)) \mathrm{d}\boldsymbol{y}\mathrm{d}z\mathrm{d}\boldsymbol{y}_1\mathrm{d}t \tag{63}$$

$$= \frac{1}{z_1 - z_0}\int_{z_0}^{z_1} \int_{\mathcal{Y}}\int_{\mathcal{Y}} \frac{\|\bar{\boldsymbol{v}}_\theta(\boldsymbol{x}) - \bar{\boldsymbol{v}}_1(\boldsymbol{x} \mid \boldsymbol{x}_1)\|^2 \bar{p}_y(\boldsymbol{y} \mid \boldsymbol{y}_1) q(\boldsymbol{y}_1)}{\bar{v}_z(z \mid z_1)} \mathrm{d}\boldsymbol{y}\mathrm{d}\boldsymbol{y}_1\mathrm{d}z. \tag{64}$$

In expectation form, we can write

$$L_{\text{FM}}(\theta) = \mathop{\mathbb{E}}_{\substack{z\sim\mathcal{U}[z_0,z_1]\\ \boldsymbol{y}_1\sim q(\boldsymbol{x}_1)\\ \boldsymbol{y}\sim\bar{p}_y(\boldsymbol{y}|\boldsymbol{y}_1)}} \frac{\|\bar{\boldsymbol{v}}_\theta(\boldsymbol{x}) - \bar{\boldsymbol{v}}_1(\boldsymbol{x} \mid \boldsymbol{x}_1)\|^2}{\bar{v}_z(z \mid z_1)}. \tag{65}$$

□

A.3. Proof of Lemma 4.5

Proof. We start by verifying

$$\boldsymbol{v}_y(\boldsymbol{x},t \mid \boldsymbol{x}_1) = -\boldsymbol{A}\left(\boldsymbol{y} - \boldsymbol{y}_1 - \boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_y^{-1}(t)\left(\boldsymbol{y} - \boldsymbol{\mu}_y(t)\right)\right). \tag{66}$$

Consider a linear time-invariant SDE given by

$$\mathrm{d}\boldsymbol{y} = -\boldsymbol{A}(\boldsymbol{y} - \boldsymbol{y}_1)\mathrm{d}t + \sqrt{2\boldsymbol{A}\boldsymbol{\Sigma}_1}\mathrm{d}\boldsymbol{w}(t). \tag{67}$$

Section 6.2 of (Särkkä & Solin, 2019) (Eq. 6.19-6.20) gives us the closed-form solution to the PDF path of this SDE as $p_y(\boldsymbol{y},t \mid \boldsymbol{y}_1) = \mathcal{N}(\boldsymbol{y} \mid \boldsymbol{\mu}_y(t), \boldsymbol{\Sigma}_y(t))$, where the mean is given as

$$\boldsymbol{\mu}_y(t) = \boldsymbol{y}_1 + e^{-\boldsymbol{A}t}(\boldsymbol{y}_0 - \boldsymbol{y}_1). \tag{68}$$

and, assuming $\boldsymbol{A}$, $\boldsymbol{\Sigma}_0$, $\boldsymbol{\Sigma}_1$ commute with each other (i.e. simultaneously diagonalizable) and $\boldsymbol{B} = \sqrt{2\boldsymbol{A}\boldsymbol{\Sigma}_1}$, the covariance is given as

$$\boldsymbol{\Sigma}_y(t) = e^{-\boldsymbol{A}t}\boldsymbol{\Sigma}_0 e^{-\boldsymbol{A}t^\top} + \int_0^t e^{-\boldsymbol{A}(t-s)}\boldsymbol{B}^2 e^{-\boldsymbol{A}^\top(t-s)}\mathrm{d}s \tag{69}$$

$$= \boldsymbol{\Sigma}_0 e^{-2\boldsymbol{A}t} + \boldsymbol{B}^2\int_0^t e^{-2\boldsymbol{A}(t-s)}\mathrm{d}s \tag{70}$$

$$= \boldsymbol{\Sigma}_0 e^{-2\boldsymbol{A}t} - \frac{\boldsymbol{B}^2\boldsymbol{A}^{-1}}{2}\left(e^{-2\boldsymbol{A}t} - \boldsymbol{I}\right) \tag{71}$$

$$= \boldsymbol{\Sigma}_0 e^{-2\boldsymbol{A}t} - \boldsymbol{\Sigma}_1\left(e^{-2\boldsymbol{A}t} - \boldsymbol{I}\right) \tag{72}$$

$$= \boldsymbol{\Sigma}_1 + e^{-2\boldsymbol{A}t}\left(\boldsymbol{\Sigma}_0 - \boldsymbol{\Sigma}_1\right) \tag{73}$$

The score of a Gaussian PDF path is (see e.g. Section 3.1 of (Hyvärinen, 2005))

$$\frac{\partial \ln(p_y(\boldsymbol{y},t \mid \boldsymbol{y}_1))}{\partial \boldsymbol{y}} = -\boldsymbol{\Sigma}_y^{-1}(t)\left(\boldsymbol{y} - \boldsymbol{\mu}_y(t)\right), \tag{74}$$

and hence the probability flow ODE (Maoutsa et al., 2020; Song et al., 2020) verifies the result.

We proceed to verify

$$v_z(\boldsymbol{x}, t \mid \boldsymbol{x}_1) = -\frac{\kappa\left(h_\mu(t) + h_\Sigma(t)\right)}{h(t)}(z - z_1) \tag{75}$$

where $h(t) = d(\boldsymbol{\mu}_y(t), \boldsymbol{\Sigma}_y(t))$ and

$$\begin{aligned} h_\mu(t) &= \left\langle \frac{\partial d(\boldsymbol{\mu}_y(t), \boldsymbol{\Sigma}_y(t))}{\partial \boldsymbol{\mu}(t)}, \boldsymbol{A}\left(\boldsymbol{\mu}_y(t) - \boldsymbol{y}_1\right) \right\rangle \\ h_\Sigma(t) &= \left\langle \frac{\partial d(\boldsymbol{\mu}_y(t), \boldsymbol{\Sigma}_y(t))}{\partial \boldsymbol{\Sigma}(t)}, 2\boldsymbol{A}\left(\boldsymbol{\Sigma}_y(t) - \boldsymbol{\Sigma}_1\right) \right\rangle_F, \end{aligned} \tag{76}$$

where $\langle \cdot, \cdot \rangle$ and $\langle \cdot, \cdot \rangle_F$ are the Euclidean and Frobenius inner products respectively. Start with

$$\mu_z(t) = z_1 + \left(\frac{d(\boldsymbol{\mu}_y(t), \boldsymbol{\Sigma}_y(t))}{d(\boldsymbol{\mu}_y(0), \boldsymbol{\Sigma}_y(0))}\right)^\kappa (z_0 - z_1) \tag{77}$$

$$= z_1 + \left(\frac{h(t)}{h(0)}\right)^\kappa (z_0 - z_1) \tag{78}$$

from Definition 4.4. The time derivative of this is

$$\frac{\mathrm{d}\mu_z(t)}{\mathrm{d}t} = \frac{\kappa(z_0 - z_1)}{h(0)}\left(\frac{h(t)}{h(0)}\right)^{\kappa-1}\frac{\mathrm{d}h(t)}{\mathrm{d}t} \tag{79}$$

$$= \frac{\kappa(z_0 - z_1)}{h(0)}\left(\frac{h(t)}{h(0)}\right)^{\kappa-1}\left(\left\langle \frac{\partial d(\boldsymbol{\mu}_y(t), \boldsymbol{\Sigma}_y(t))}{\partial \boldsymbol{\mu}_y(t)}, \frac{\mathrm{d}\boldsymbol{\mu}_y(t)}{\mathrm{d}t} \right\rangle + \left\langle \frac{\partial d(\boldsymbol{\mu}_y(t), \boldsymbol{\Sigma}_y(t))}{\partial \boldsymbol{\Sigma}_y(t)}, \frac{\mathrm{d}\boldsymbol{\Sigma}_y(t)}{\mathrm{d}t} \right\rangle_F\right) \tag{80}$$

where the chain rule is applied. Eq. 6.2 in (Särkkä & Solin, 2019) gives the dynamics of the mean and covariance for the above SDE, and hence we can write

$$\frac{\mathrm{d}\boldsymbol{\mu}_y(t)}{\mathrm{d}t} = -\boldsymbol{A}(\boldsymbol{\mu}_y(t) - \boldsymbol{y}_1) \tag{81}$$

$$\frac{\mathrm{d}\boldsymbol{\Sigma}_y(t)}{\mathrm{d}t} = -2\boldsymbol{A}(\boldsymbol{\Sigma}_y(t) - \boldsymbol{\Sigma}_1). \tag{82}$$

Thus, we can write

$$\frac{\mathrm{d}\mu_z(t)}{\mathrm{d}t} = -\frac{\kappa(z_0 - z_1)}{h(0)}\left(\frac{h(t)}{h(0)}\right)^{\kappa-1}\left(h_\mu(t) + h_\Sigma(t)\right) \tag{83}$$

To simplify further, and remove dependence on z_0, we solve $\mu_z(t)$ for $z_0 - z_1 = (\mu_z(t) - z_1)(h(t)/h(0))^{-\kappa}$ and substitute it in to get

$$\frac{\mathrm{d}\mu_z(t)}{\mathrm{d}t} = -\frac{\kappa(\mu_z(t) - z_1)}{h(0)}\left(\frac{h(t)}{h(0)}\right)^{-1}\left(h_\mu(t) + h_\Sigma(t)\right) \tag{84}$$

$$= -\frac{\kappa\left(h_\mu(t) + h_\Sigma(t)\right)}{h(t)}(\mu_z(t) - z_1) \tag{85}$$

Note that since $h(\infty) = 0$ (the distance to the stationary distribution goes to zero), having $\kappa \in [1, 0)$ in the expression before substitution prevents $\frac{\mathrm{d}\mu_z(t)}{\mathrm{d}t}$ from becoming undefined. Eq. 6.2 in (Särkkä & Solin, 2019) shows that $\frac{\mathrm{d}\mu_z(t)}{\mathrm{d}t}$ describes the dynamics of the mean of a linear time-varying SDE without a diffusion coefficient

$$\mathrm{d}z = -\frac{\kappa\left(h_\mu(t) + h_\Sigma(t)\right)}{h(t)}(z - z_1)\mathrm{d}t + 0\mathrm{d}w(t) \tag{86}$$

Hence the probability flow ODE gives the result. □

A.4. Proof of Theorem 4.7

Proof. Consider the expression

$$\mu_z(t) = z_1 + \left(\frac{h(t)}{h(0)}\right)^{\kappa} (z_0 - z_1)\,. \tag{87}$$

Assuming the h is bijective, we can solve for the inverse mapping of $\mu_z(t)$ as

$$\tau(z) = h^{-1}\left(h(0)\left(\frac{z - z_1}{z_0 - z_1}\right)^{1/\kappa}\right). \tag{88}$$

Plugging this into $\boldsymbol{\mu}_y(t)$ gives its solution in terms of $\mu_z(t)$:

$$\boldsymbol{\mu}_y(t) = \boldsymbol{y}_1 + \exp\left(-\boldsymbol{A}\tau(\mu_z(t))\right)(\boldsymbol{y}_0 - \boldsymbol{y}_1) \tag{89}$$

which is a solution to

$$\frac{\mathrm{d}\boldsymbol{\mu}_y(t)}{\mathrm{d}\mu_z(t)} = \frac{\mathrm{d}\boldsymbol{\mu}_y(t)}{\mathrm{d}t}\frac{\mathrm{d}t}{\mathrm{d}\mu_z(t)} = \frac{\boldsymbol{A}(\boldsymbol{\mu}_y(t) - \boldsymbol{y}_1)}{f\left(\tau(\mu_z(t))\right)(\mu_z(t) - z_1)} \tag{90}$$

where (see Lemma 4.5)

$$f(t) = \frac{\kappa\left(h_\mu(t) + h_\Sigma(t)\right)}{h(t)}. \tag{91}$$

Similarly, for the covariance $\boldsymbol{\Sigma}_y(t)$ we have

$$\boldsymbol{\Sigma}_y(\mu_z(t)) = \boldsymbol{\Sigma}_1 + \exp\left(-2\boldsymbol{A}\tau(\mu_z(t))\right)(\boldsymbol{\Sigma}_0 - \boldsymbol{\Sigma}_1)\,, \tag{92}$$

which is a solution to

$$\frac{\mathrm{d}\boldsymbol{\Sigma}_y(t)}{\mathrm{d}\mu_z(t)} = \frac{\mathrm{d}\boldsymbol{\Sigma}_y(t)}{\mathrm{d}t}\frac{\mathrm{d}t}{\mathrm{d}\mu_z(t)} = \frac{2\boldsymbol{A}(\boldsymbol{\Sigma}_y(t) - \boldsymbol{\Sigma}_1)}{f\left(\tau(\mu_z(t))\right)(\mu_z(t) - z_1)}. \tag{93}$$

Inserting the expression for t into $\mu_z(t)$ yields

$$\mu_z(\tau(\mu_z(t))) = \mu_z(t) \tag{94}$$

by definition.

Thus, by taking $\mu_z(t)$ as a new time, we obtain a new system

$$\boldsymbol{\mu}'_y(t) = \boldsymbol{y}_1 + \exp\left(-\boldsymbol{A}\tau(t)\right)(\boldsymbol{y}_0 - \boldsymbol{y}_1) = \boldsymbol{\mu}_y(\tau(t)) \tag{95}$$

$$\boldsymbol{\Sigma}'_y(t) = \boldsymbol{\Sigma}_1 + \exp\left(-2\boldsymbol{A}\tau(t)\right)(\boldsymbol{\Sigma}_0 - \boldsymbol{\Sigma}_1) = \boldsymbol{\Sigma}_y(\tau(t)) \tag{96}$$

$$\mu'_z(t) = z_1 + \left(\frac{h(\tau(t))}{h(0)}\right)(z_0 - z_1) = t = \mu_z(\tau(t)) \tag{97}$$

which are solutions to

$$\frac{\mathrm{d}\boldsymbol{\mu}'_y(t)}{\mathrm{d}t} = \frac{\boldsymbol{A}(\boldsymbol{\mu}'_y(t) - \boldsymbol{y}_1)}{f\left(\tau(t)\right)(t - z_1)} \tag{98}$$

$$\frac{\mathrm{d}\boldsymbol{\Sigma}'_y(t)}{\mathrm{d}t} = \frac{2\boldsymbol{A}(\boldsymbol{\Sigma}'_y(t) - \boldsymbol{\Sigma}_1)}{f\left(\tau(t)\right)(t - z_1)} \tag{99}$$

$$\frac{\mathrm{d}\mu'_z(t)}{\mathrm{d}t} = 1. \tag{100}$$

Now let

$$\boldsymbol{G}(t) = \frac{\boldsymbol{A}}{f(\tau(t))(t - z_1)}. \tag{101}$$

Then the above dynamics describe a linear time-varying SDE

$$\mathrm{d}y = \boldsymbol{G}(t)(\boldsymbol{y} - \boldsymbol{y}_1)\mathrm{d}t + \sqrt{2\boldsymbol{G}(t)\boldsymbol{\Sigma}_1}\mathrm{d}\boldsymbol{w}(t) \tag{102}$$

$$\mathrm{d}z = 1\mathrm{d}t + 0\mathrm{d}w(t). \tag{103}$$

Applying the formula for the probability flow ODE then gives us

$$\frac{\mathrm{d}\boldsymbol{y}}{\mathrm{d}t} = \boldsymbol{G}(t)(\boldsymbol{y} - \boldsymbol{y}_1) - \boldsymbol{G}(t)\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_y'(t)^{-1}\left(\boldsymbol{y} - \boldsymbol{\mu}_y'(t)\right) = \frac{\boldsymbol{v}_y(\boldsymbol{x}, \tau(t) \mid \boldsymbol{x}_1)}{\boldsymbol{v}_z(\boldsymbol{x}, \tau(t) \mid \boldsymbol{x}_1)} = \boldsymbol{w}_y(\boldsymbol{x}, t \mid \boldsymbol{x}_1). \tag{104}$$

□