GRADUAL FINE-TUNING FOR FLOW MATCHING MODELS

Anonymous authors
Paper under double-blind review

ABSTRACT

Fine-tuning flow matching models is a central challenge in settings with limited data, evolving distributions, or computational constraints. While recent work has produced significant advances, particularly in the area of reward-based fine-tuning, current methods fail to demonstrate both theoretical correctness and strong empirical results in terms of stability, efficiency, and accuracy. In this work, we propose Gradual Fine-Tuning (GFT), a simple yet principled framework for fine-tuning flow generative models using samples from the target distribution. For stochastic flows, GFT defines a temperature-controlled sequence of intermediate objectives that smoothly interpolate between the pretrained and target drifts, provably converging to the true target as the temperature approaches zero. Empirically, GFT significantly improves convergence stability, while maintaining or improving generation quality, inference efficiency, and generation diversity compared to other fine-tuning methods. When extended to reward-based fine-tuning, GFT outperforms Adjoint Matching on several metrics while achieving over an order of magnitude reduction in training time. Our results position GFT as a simple yet theoretically grounded and practically effective alternative for scalable adaptation of flow matching models under distribution shift.

1 INTRODUCTION

Recent advances in flow-based generative modeling (Lipman et al., 2023; Liu et al., 2023) have achieved remarkably high generation quality, with widespread use in image generation (Esser et al., 2024), as well as scientific applications such as molecule generation (Jing et al., 2024; Klein et al., 2023; Song et al., 2023), biological sequence modeling (Stark et al., 2024; Davis et al., 2024), and atmospheric forecasting (Fotiadis et al., 2025; Bao et al., 2024). In settings with limited computational resources, data-scarcity, or evolving distributions, fine-tuning a pretrained model can be the most effective approach for achieving accurate generation. An ideal method should therefore be *theoretically correct*, with provable convergence to the target; *stable* and *efficient* during training; and compatible with methods for *inference efficiency*, such as few-step generation (Geng et al., 2026; Boffi et al., 2025) and Optimal Transport (OT) couplings (Tong et al., 2024; Pooladian et al., 2023).

Recent work in fine-tuning for flow generative models has made significant progress, particularly in reward-based methods, but existing methods do not fully satisfy these key criteria. Adjoint Matching (Domingo-Enrich et al., 2025) and other methods based on Stochastic Optimal Control (SOC) provide convergence guarantees but require multiple trajectory simulations for each fine-tuning step, and optimize separate control vector fields which are added linearly to the pretrained drift, making them incompatible with nonlinear fine-tuning techniques such as full fine-tuning and LoRA (Hu et al., 2022). The broader class of reward-based methods (Clark et al., 2024; Xu et al., 2023; Prabhudesai et al., 2023; Black et al., 2024; Fan et al., 2023; 2025b) similarly require trajectory simulation, and largely suffer from high gradient instability. An important and practically relevant setting remains underexplored: fine-tuning when samples from the target distribution are directly available, removing the need to define or learn a reward function. A common method in this category is standard fine-tuning, or continued training with the same algorithm on a new dataset, treating the pretrained model merely as an initialization. Without a principled mechanism for controlled adaptation, this often leads to highly unstable training.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

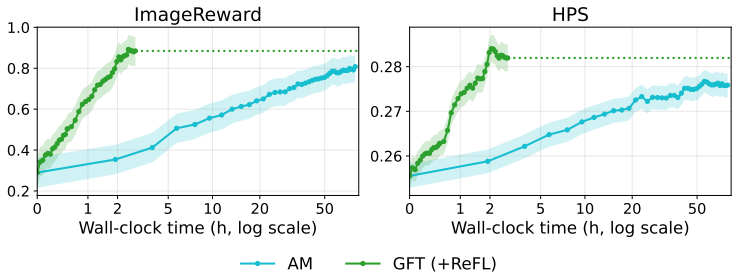


Figure 1: Comparison of GFT (+ReFL) (Algorithm 5) and Adjoint Matching performance against wall-clock training time on a logarithmic scale. After training on prompts from the ImageReward dataset, the GFT-based method outperforms Adjoint Matching on ImageReward and Human Preference Score (HPS), two independently trained human preference reward models, while achieving over an order of magnitude reduction in training time. Further details given in Section 5.3.

We introduce **Gradual Fine-Tuning** (GFT), a principled framework for adapting pretrained flow models using target distribution samples. GFT uses annealed regularization to create a sequence of intermediate objectives that smoothly interpolate between pretrained and target dynamics. The result is a gradual, highly stable, and reliably controllable transition to the fine-tuning target. GFT addresses the limitations of existing approaches while satisfying the key criteria for flow model fine-tuning outlined above. The main contributions of our work are as follows:

1. We prove that GFT converges to the target distribution under standard assumptions, including when conditioned on arbitrary source-target couplings, such as OT.
2. We show the natural extension of GFT to few step inference methods and reward fine-tuning.
3. We empirically demonstrate that GFT matches or exceeds the generation quality of existing methods while providing improved training stability and shorter trajectory lengths. In reward fine-tuning, GFT outperforms Adjoint Matching with an order of magnitude reduction in training time (Figure 1).

This work presents the first systematic investigation of annealed gradual fine-tuning for flow models using a temperature parameter, and establishes a foundation for scalable and efficient adaptation of flow-based generative models.

2 Related Work

Available Supervision for Fine-Tuning. A substantial body of work focuses on fine-tuning generative models to adapt to a tilted distribution implicitly defined by a reward function. When the reward function is differentiable, these methods can directly backpropagate the reward gradients (Clark et al., 2024; Domingo-Enrich et al., 2025; Xu et al., 2023; Prabhudesai et al., 2023). In contrast, policy gradient and reinforcement learning methods are commonly used when the reward function is not differentiable (Black et al., 2024; Fan et al., 2023; 2025a). Although reward-based methods can be sample efficient and highly generalizable, they often face high training instability and reward hacking when the true reward is approximated by a trained model. Another class of methods directly optimize over pairwise preference data when a reward function is unavailable or difficult to define (Wallace et al., 2023; Liang et al., 2025). When samples from the fine-tuning target distribution are available, the most straightforward approach is standard fine-tuning. GFT is also sample-based, but incorporates annealed regularization to yield strong theoretical guarantees and improve upon the instability associated with standard fine-tuning.

Fine-Tuning Strategy. Several recent works formulate fine-tuning as stochastic optimal control (SOC) (Domingo-Enrich et al., 2025; Uehara et al., 2024; Potaptchik et al., 2025). Although SOC methods provide strong theoretical guarantees, they are incompatible with one-step inference methods and non-linear fine-tuning techniques, and trajectory simulation during training makes them time and memory intensive. Beyond SOC formulations, many methods perform direct optimization using reward gradients (Clark et al., 2024; Prabhudesai et al., 2023) or samples, such as standard fine-tuning and GFT. A separate class of methods apply inference-time guidance or conditioning to

alter the generation of samples instead of fine-tuning the pretrained model (Ho & Salimans, 2022; Ben-Hamu et al., 2024). While computationally appealing, these methods typically increase inference cost, and may not be capable of handling large distribution shifts as the parameters of the pretrained model remain unchanged.

3 Background and Preliminaries

3.1 Flow Matching

Flow Matching (Lipman et al., 2023; Liu et al., 2023) is a family of generative modeling approaches which iteratively simulate a continuous process between a source distribution p_0 and a data-defined target distribution p_1. Flow Matching defines a generative model by learning a time-varying vector field $v_\theta(X_t, t) : \mathbb{R}^d \times \mathbb{R} \to \mathbb{R}^d$ which transports a family of probability densities $\{p_t\}_{t\in[0,1]}$ along a deterministic path. The evolution of a single sample through this vector field is described by the ordinary differential equation (ODE) $dX_t = v_\theta(X_t, t)\ dt,\ X_0 \sim p_0$. Integrating this ODE over time therefore enables simulated generation from the target distribution. To achieve a well-defined continuous transport process, each intermediate density p_t must also satisfy the continuity equation (Villani, 2008),

$$\partial_t p_t + \nabla_x \cdot (p_t v_t) = 0. \tag{1}$$

Aside from impressive generation quality, the major achievement of flow matching is its highly efficient simulation-free training. However, simulation is still required during inference, making flow-based generation inefficient compared to earlier single-step generative models (Kingma & Welling, 2014; Goodfellow et al., 2014). Several recent advances have sought to mitigate this inference cost, such as minimizing vector field curvature (Liu et al., 2023), sampling according to OT plans (Tong et al., 2024), and few-step inference methods (Geng et al., 2026; Boffi et al., 2025).

3.2 Stochastic View of Flow Matching

Despite their deterministic formulation, flow matching models can be embedded into a class of stochastic differential equations (SDEs) with time marginals that match p_t. We consider the SDE $dX_t = \tilde{v}_t(X_t)dt + \sigma_t(X_t)dB_t,\ X_0 \sim p_0$ with drift term $\tilde{v} : \mathbb{R}^d \times \mathbb{R} \to \mathbb{R}^d$ and diffusion term $\sigma : \mathbb{R}^d \times \mathbb{R} \to \mathbb{R}^{d\times d}$. Assuming isotropic diffusion, the Fokker-Planck equation describing the evolution of the density p_t under the dynamics of this SDE is

$$\partial_t p_t = -\nabla_x \cdot (\tilde{v}_t p_t) + \frac{1}{2}\nabla_x \cdot \nabla_x \cdot (\sigma_t^2 I p_t)\ . \tag{2}$$

We can find a drift term to exactly match the ODE density evolution by substituting the continuity equation 1 into the Fokker-Planck equation 2 (details shown in Appendix E).

$$\tilde{v}_t = v_t + \frac{1}{2}\sigma_t^2 \nabla_x \log p_t\ . \tag{3}$$

This formulation explicitly introduces the score function of the density p_t, and has therefore motivated the exploration of the relationship between deterministic Neural ODEs and diffusion models (Song et al., 2021b), as well as the related stochastic interpolant model (Albergo et al., 2025). We recover the exact ODE drift term by setting $\sigma_t = 0$.

4 Gradual Fine-Tuning

4.1 Problem Setup

Assume that we have access to a pretrained flow matching model with parameters θ_0 which generates samples from a distribution p_1. The model has been trained to simulate the base continuous process

$$\mathbb{P}_{\theta_0} : dX_t = v_{\theta_0}(X_t, t)dt + \sigma_t dB_t, \qquad X_0 \sim p_0\ , \tag{4}$$

where v_{θ_0} is the vector field transforming samples from the source distribution p_0 to p_1, and B_t is a standard d-dimensional Brownian motion. We now want to fine-tune this model to generate samples

from a new target distribution q, from which we have access to samples. The target fine-tuned continuous process is defined as

$$\mathbb{P}_q : dX_t = v_q(X_t, t)dt + \sigma_t dB_t, \qquad X_0 \sim p_0 \ , \tag{5}$$

where the target vector field v_q, which generates samples from q, is unknown during fine-tuning. The goal of fine-tuning is to learn a new process $\mathbb{P}_\theta$ which closely approximates $\mathbb{P}_q$:

$$\mathbb{P}_\theta : dX_t = v_\theta(X_t, t)dt + \sigma_t dB_t, \qquad X_0 \sim p_0 \ . \tag{6}$$

The fine-tuned vector field v_θ is initialized with the parameters of the base vector field v_{θ_0}, and then optimized to match the target vector field v_q through fine-tuning. Here, $\mathbb{P}_{\theta_0}, \mathbb{P}_q,$ and $\mathbb{P}_\theta$ are probability path measures on the path space $C([0,1], \mathbb{R}^d)$ induced by their corresponding SDEs (Karatzas & Shreve, 1991; Song et al., 2021a; Berner et al., 2024). As such, they assign likelihoods to trajectories $\{X_t\}_{t\in[0,1]}$, rather than individual points X_t, and therefore allow us to reason about the full generative process. For each process, the distribution of the terminal state X_1 generated by the SDE is defined by its corresponding terminal marginal. In particular, under $\mathbb{P}_{\theta_0}$ we have $X_1 \sim p_1$, and under $\mathbb{P}_q$ we have $X_1 \sim q$. Therefore, when fine-tuning successfully adapts $\mathbb{P}_\theta$ to closely match $\mathbb{P}_q$, its terminal marginal will match the target distribution q, under suitable assumptions.

4.2 Gradual Fine-Tuning (GFT) Objective

Consider the following optimization over SDE-induced path measures:

$$\begin{aligned} \min_\theta \quad & \mathrm{KL}(\mathbb{P}_\theta \| \mathbb{P}_q) + \beta \mathrm{KL}(\mathbb{P}_\theta \| \mathbb{P}_{\theta_0}) \\ \text{s.t.} \quad & dX_t = v_\theta(X_t, t)dt + \sigma_t dB_t \ . \end{aligned} \tag{7}$$

The first term of the cost expression promotes alignment with the target process, while the second term acts as a divergence regularizer that penalizes large deviations from the pretrained dynamics. The coefficient $\beta \in \mathbb{R}_+$ is a temperature parameter that controls the strength of this regularization. The GFT objective can therefore be viewed as an interpolation between two transport processes, one which guides samples to the terminal distribution q by matching the target process, and another which maintains the pretrained dynamics by penalizing deviation from the base process. Balancing these costs suggests that the optimal fine-tuned model should be a weighted average of the base and target dynamics, controlled by the temperature β. We now formalize this intuition in the following theorem.

Theorem 4.1 *Let $\mathbb{P}_{\theta_0}, \mathbb{P}_q,$ and $\mathbb{P}_\theta$ be path measures induced by SDEs with drift terms $v_{\theta_0}, v_q,$ and v_θ, respectively. Suppose these processes share a diffusion coefficient σ_t. Then, for a given temperature β, the vector field minimizing the GFT objective (7) is given by the convex combination*

$$v^*_\theta(X_t, t) = \left(\frac{1}{1+\beta}\right) v_q(X_t, t) + \left(\frac{\beta}{1+\beta}\right) v_{\theta_0}(X_t, t) \ . \tag{8}$$

Proof given in Appendix H.

By Theorem 4.1, minimizing the GFT objective yields a closed-form and continuous optimal drift which is a weighted arithmetic mean of the pretrained and target vector fields. The weighting is directly controlled by the temperature β, resulting in limiting behavior which recovers the pretrained and target dynamics:

$$\lim_{\beta\to\infty} v^*_\theta = v_{\theta_0} \ , \qquad \lim_{\beta\to 0^+} v^*_\theta = v_q \ . \tag{9}$$

Annealed Regularization. The limiting behavior of the optimal vector field v^*_θ provides a clear motivation for the gradual nature of our method. To control the tradeoff between preserving pretrained knowledge and adapting to the new target distribution, GFT uses a time-dependent cooling schedule. At the start of fine-tuning, a high temperature strongly penalizes deviation from the base model, keeping updates close to the pretrained drift and preventing premature collapse towards the target. As training progresses, β is gradually annealed toward 0^+, smoothly relaxing this constraint and allowing the fine-tuned drift v_θ to asymptotically recover the target vector field v_q (9). This progression creates a gradual transition between the pretrained and target dynamics, improving training stability and final generation diversity. Similar gradual adaptation methods have previously been

shown to stabilize and improve classical domain adaptation (He et al., 2024), and related trust region methods have recently been considered for diffusion model fine-tuning (Blessing et al., 2025).

Tractable Reformulation. Although the GFT objective provides a simple and elegant formulation for gradual fine-tuning of flow-based models, it cannot be directly optimized for two reasons. First, it implicitly depends on the vector field v_q through the path measure $\mathbb{P}_q$, which is unknown during fine-tuning. Second, the objective is defined over the space of path measures, which are not differentiable with respect to the model parameters θ. Fundamentally, the neural network being optimized during fine-tuning represents the drift term of the SDE which induces the path measure $\mathbb{P}_\theta$, not the measure itself. In order to perform optimization, the objective must be rewritten in terms of the corresponding drift terms of each path measure. In the following proposition, we demonstrate that the application of Girsanov's Theorem achieves this result.

Proposition 4.2 *Let $\mathbb{P}_f$ and $\mathbb{P}_g$ be mutually absolutely continuous path measures induced by SDEs with shared diffusion coefficients and source distributions, with drift terms f and g respectively. Then, by Girsanov's theorem, $KL(\mathbb{P}_f \| \mathbb{P}_g) = \mathbb{E}_{t\sim\mathcal{U}[0,1], X_t\sim\mathbb{P}_f}\left[\frac{1}{2}\|\sigma_t^{-1}(f(X_t,t) - g(X_t,t))\|^2\right]$. Proof given in Appendix F.*

We apply Proposition 4.2 to (7) to write the tractable form of the GFT objective, with $\ell_\theta(X_t, t) := \|\sigma_t^{-1}(v_\theta(X_t,t) - v_q(X_t,t))\|^2 + \beta\|\sigma_t^{-1}(v_\theta(X_t,t) - v_{\theta_0}(X_t,t))\|^2$ (details given in Appendix J).

$$\mathcal{L}(\theta) = \frac{1}{2}\mathbb{E}_{t\sim\mathcal{U}[0,1],\ X_t\sim\mathbb{P}_\theta}\left[\ell_\theta(X_t,t)\right] \tag{10}$$

While this reformulation successfully shifts optimization from the path space to the space of vector fields, enabling optimization of v_θ, the dependence on the unknown true v_q still exists.

4.3 Conditional GFT with Source-Target Couplings

To remove the dependence on the unknown v_q, we follow CFM (Lipman et al., 2023) and condition the target process on samples. Let $\pi \in \Pi(p_0, q)$ be any coupling between the source and target distributions with marginals $\int \pi(X_0, X_1)dX_1 = p_0(X_0)$ and $\int \pi(X_0, X_1)dX_0 = q(X_1)$, and define $Z := (X_0, X_1) \sim \pi$. The conditional GFT objective is

$$\mathcal{L}_\pi(\theta) = \mathbb{E}_{Z\sim\pi}\left[KL(\mathbb{P}_\theta \| \mathbb{P}_q(\cdot|Z)) + \beta KL(\mathbb{P}_\theta \| \mathbb{P}_{\theta_0})\right] \tag{11}$$

where $\mathbb{P}_q(\cdot|Z)$ is induced by the SDE with the conditional drift term $v_q(X_t, t|Z)$ which marginalizes to the true target, such that $\mathbb{E}_{Z\sim\pi}[v_q(X_t,t|Z)] = v_q(X_t,t)$ (Lipman et al., 2023). Recall that the pretrained base process $\mathbb{P}_{\theta_0}$ is fixed, and therefore remains unconditional. Crucially, the fine-tuned process $\mathbb{P}_\theta$ must also be unconditional to perform generation at inference time without access to target distribution samples. Its drift term v_θ is therefore trained to approximate the optimal conditional vector field marginalized over $Z \sim \pi$.

Simulation-Free GFT. Applying Proposition 4.2 to both KL terms yields an expectation over trajectories from the current model $\mathbb{P}_\theta$, making the objective simulation-based. However, it can be exactly rewritten in a simulation-free version through change of measure. Girsanov's theorem provides a closed-form expression for the Radon-Nikodym derivative $w_\theta(X|Z) := \frac{d\mathbb{P}_\theta(\cdot|Z)}{d\mathbb{P}_q(\cdot|Z)}(X)$, with $X := (X_t)_{t\in[0,1]}$ denoting the full trajectory. Reweighting by this density moves the expectation onto the conditional target process, which is known analytically (proof given in Appendix K). With $\ell_\theta(X_t, t|Z) := \|\sigma_t^{-1}(v_\theta(X_t,t) - v_q(X_t,t|Z))\|^2 + \beta\|\sigma_t^{-1}(v_\theta(X_t,t) - v_{\theta_0}(X_t,t))\|^2$, from (11) we obtain:

$$\mathcal{L}_\pi(\theta) = \frac{1}{2}\mathbb{E}_{\substack{Z\sim\pi,\ t\sim\mathcal{U}[0,1],\\ X_t\sim\mathbb{P}_\theta(\cdot|Z)}}\left[\ell_\theta(X_t,t|Z)\right] = \frac{1}{2}\mathbb{E}_{\substack{Z\sim\pi,\ t\sim\mathcal{U}[0,1],\\ X_t\sim\mathbb{P}_q(\cdot|Z)}}\left[w_\theta(X|Z)\ell_\theta(X_t,t|Z)\right]. \tag{12}$$

Under standard assumptions, this loss is exactly gradient-equivalent to the unconditional objective (10) (proof given in Appendix I), and therefore preserves the theoretical guarantees of the unconditional objective while enabling the use of arbitrary couplings such as OT. Although the change of measure in (12) is exact and importance sampling with density ratios such as $w_\theta(X|Z)$ is unbiased, it is also well known to suffer from high variance (Chatterjee & Diaconis, 2018). In our setting, this is exacerbated by the fact that $w_\theta(X|Z)$ depends on the mismatch between the fine-tuned and conditional target drifts accumulated over the full sampling trajectory (Appendix K). Despite this

high variance, we empirically demonstrate strong convergence results through proper tuning of the diffusion coefficient σ_t, along with a proposed algorithm for the practical implementation of the simulation-free objective (12) (Appendix K.1). Interestingly, our ablations revealed that simulation-free GFT, which is valid under the assumption that $w_\theta(X|Z) \approx 1$, achieves almost identical convergence results while lowering variance and increasing training efficiency (Appendix P.4). For this reason, the main experimental results presented in this work use simulation-free GFT with the approximation that $w_\theta \approx 1$. We note that this is not a new challenge, and that trading unbiasedness for variance reduction has been used as a training-time strategy in related fields (Schulman et al., 2017; Black et al., 2024).

4.4 GFT Convergence

After initialization, standard CFM fine-tuning regresses directly onto the target vector field, discarding the pretrained dynamics and providing no control over the transition from v_{θ_0}. In Section 5 we empirically demonstrate that this abrupt shift often leads to highly unstable fine-tuning. In contrast, the sequence of intermediary targets created by GFT results in a smooth and controlled transition to the new target distribution, creating stability while maintaining accurate convergence. In Theorem 4.3, we demonstrate that the temperature parameter β directly controls the rate of convergence, and can be used to guarantee convergence to the true target.

Theorem 4.3 *Suppose v_{θ_0} and v_q are L-Lipschitz continuous and have a bounded mismatch given by C', and that the conditional variance of v_q is finite. For a fixed β, let v^*_β be the GFT minimizer given in* (8)*, with terminal distribution p^*_β. Then,*

$$\|v^*_\beta - v_q\|^2_{L^2(\mathbb{P}_q)} \leq \left(\tfrac{\beta}{1+\beta}\right)^2 C' \,, \qquad W_2^2(p^*_\beta, q) \leq e^{2L+1}\left(\tfrac{\beta}{1+\beta}\right)^2 C' \,.$$

Then, as $\beta \to 0^+$, $v^*_\beta \to v_q$ in $L^2(\mathbb{P}_q)$ and $p^*_\beta \to q$ in W_2. Proof given in Appendix L.

Theorem 4.3 bounds the deviation from the GFT minimizer to the target, in both its vector field and terminal distribution, by a function of β. As a result, any cooling schedule with $\beta \to 0$ is guaranteed to recover the target distribution.

4.5 Extension to Few-Step Inference

Recent few-step inference methods for flow generative models (Boffi et al., 2025; Geng et al., 2026) have significantly mitigated the inefficiency of iterative generation. However, Adjoint Matching and other SOC-based methods inherently optimize over full sampling trajectories (Domingo-Enrich et al., 2025; Havens et al., 2025), and would therefore require a non-trivial reformulation to extend to such methods. In contrast, the vector field regression objective of GFT extends naturally to few-step inference, which we demonstrate by adapting GFT to fine-tuning for MeanFlow (MF) models (Geng et al., 2026) in the MF-GFT algorithm (2).

In this framework, we define u_{θ_0}, u_q, and u_θ to be the average velocity fields related to v_{θ_0}, v_q, and v_θ by the MF identity (Geng et al., 2026, Eq. 6). When v_θ takes the form of the GFT optimum given in (8), we show that the corresponding optimal average velocity field is provably the same convex combination: $u_\theta = \frac{1}{1+\beta} u_q + \frac{\beta}{1+\beta} u_{\theta_0}$ (Appendix M). Intuitively, this relationship simply lifts the gradual sequence of instantaneous target vector fields created by GFT to the space of average vector fields, allowing MF-GFT to retain many of the benefits of gradual adaptation. Empirically, our results demonstrate that MF-GFT leads to a significant improvement in inference efficiency without sacrificing generation quality or training stability.

4.6 Extension to Reward-Based Fine-Tuning

In all preceding analysis, we develop GFT and its theoretical guarantees under sample-based fine-tuning, a commonly-used yet under-explored setting in the fine-tuning literature for flow generative models. A strong thread of recent work has demonstrated significant advancement in reward-based fine-tuning (Domingo-Enrich et al., 2025; Clark et al., 2024; Black et al., 2024), which may be preferable when a reward function can be naturally defined and samples from the target distribution are unavailable or sparse. Notably, sample-based GFT extends naturally to reward fine-tuning

Table 1: Convergence and stability analysis on the Camelyon17 dataset. GFT yields greater stability and a higher convergence rate compared to CFM. Extended results in Appendix R.5.

	Objective	Inst. Variance ↓	Convergence Rate ↑	Spearman ρ ↓
Cross-Domain Adaptation	Full (CFM)	99.242	3.133×10^{-2}	-0.663
	LoRA (CFM)	70.604	2.504×10^{-2}	-0.597
	Full (GFT)	**63.132**	$\mathbf{3.647 \times 10^{-2}}$	**-0.932**
In-Domain Adaptation	Full (CFM)	19.387	4.693×10^{-3}	-0.253
	LoRA (CFM)	13.134	1.182×10^{-2}	-0.499
	LoRA (GFT)	**8.429**	$\mathbf{2.213 \times 10^{-2}}$	**-0.966**

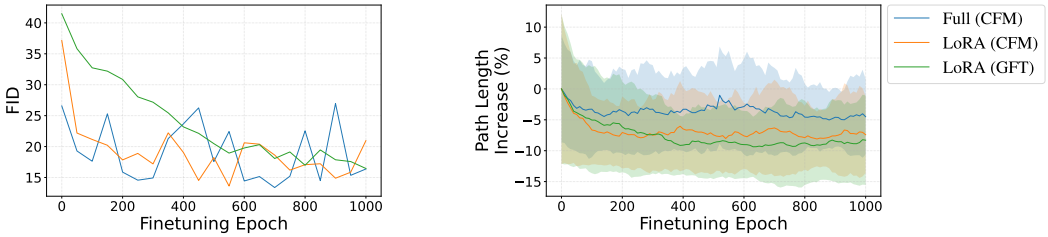


Figure 2: In-domain adaptation on Camelyon17. GFT achieves the same accuracy as CFM at the end of fine-tuning, while maintaining a lower average path length. Further results in Appendix R.3.

through a simple change of measure. Given a reward function $R(X_1)$, a target defined as the tilted distribution $q(x) \propto p_{\theta_0}(x) \exp(R(x)/\alpha)$ for $\alpha > 0$, and an independent coupling $\pi_0 = p_0 \otimes p_{\theta_0}$, the GFT objective can be rewritten as a reward-based objective (proof given in Appendix N):

$$\mathcal{L}_R(\theta) = \frac{1}{2}\mathbb{E}_{Z\sim\pi_0,\, t\sim\mathcal{U}[0,1],\, X_t\sim\mathbb{P}_q(\cdot|Z)}\left[\exp\left(\frac{R(X_1)}{\alpha}\right) w_\theta(X|Z)\ell_\theta(X_t,t|Z)\right]. \tag{13}$$

The resulting objective retains the two-term structure of target matching and regularization, but does not depend on the target distribution q. Crucially, since the samples $X_1 \sim p_{\theta_0}$ are drawn from the fixed pretrained model, they are independent of θ and can be generated once prior to the start of training. This suggests a simple implementation of (13), where rather than computing importance weights within each batch, the reward weighting can be applied once at the dataset level. This can be done simply by generating samples from the pretrained model and filtering based on reward (Algorithm 4). This practical implementation of the derived objective creates a truncated approximation of the tilted distribution, avoiding the high variance of single-batch importance weights and allowing for the use of a variation of the original sample-based GFT algorithm (Algorithm 5). More details on the reward-based GFT implementation can be found in Appendix N.

5 Experiments

We evaluate GFT by fine-tuning pretrained flow generative models and comparing against baseline fine-tuning methods in two settings: sample-based and reward-based fine-tuning.

Sample Fine-Tuning Experiments. To evaluate GFT under natural distribution shifts, we compare it to standard fine-tuning on three WILDS datasets (Koh et al., 2021). Each dataset partitions images into groups that share natural factors such as experimental batch, geographic region, or time of acquisition, and its training and validation splits contain disjoint groups, creating a distribution shift between them. We define two scenarios to evaluate GFT across varying degrees of distribution shift. In **Cross-domain adaptation** we fine-tune a model pretrained on Cifar-10 to the training split of a WILDS dataset, representing a large shift from a general-purpose prior to a specific scientific domain. For **In-domain adaptation** we further fine-tune the resulting model on the corresponding validation split, simulating a subtler shift within the same scientific context. We report FID and average path length, along with three metrics that quantify the speed and stability of convergence. Instantaneous variance measures local volatility, convergence rate measures adaptation speed, and Spearman correlation between training epoch and FID measures global consistency.

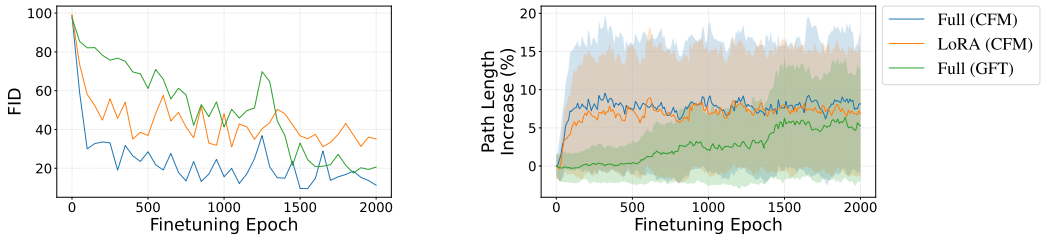
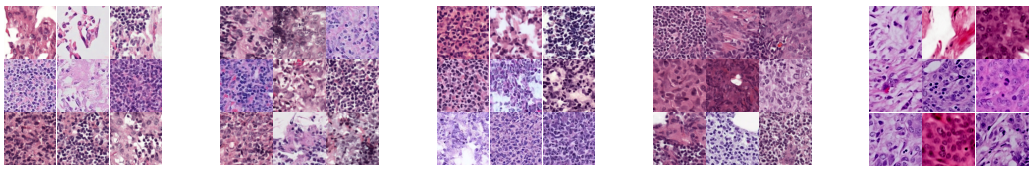


Figure 3: Cross-domain adaptation on Camelyon17. By the end of fine-tuning, GFT has achieved the same accuracy as CFM with a lower average path length. Further results in Appendix R.1.



(a) Full (CFM). (b) Full (GFT). (c) LoRA (CFM). (d) LoRA (GFT). (e) Full (GFT-MF).

Figure 4: Generated images of the Camelyon17 dataset with varying fine-tuning methods. Images shown in (e) are generated using two-step inference. Extended results in Appendix R.4.

Reward Fine-Tuning Experiments. To compare GFT against Adjoint Matching (AM), we fine-tune a Stable Diffusion model (Rombach et al., 2022) to optimize ImageReward (Xu et al., 2023), a predefined reward model trained on human preference data. These reward-based experiments represent an important test of GFT in a complex and practical setting, including a conditional text-to-image pipeline, and stochastic training and generation. We report performance throughout training in terms of ImageReward, as well as CLIPScore (Hessel et al., 2021) to evaluate adherence to prompt, and HPS (Wu et al., 2023) to test generalization to a separately-trained human preference reward model. All evaluation metrics are reported as averages over 10 samples generated per prompt using 100 unseen prompts. Additional implementation details are given in Appendix N.2.

5.1 Sample Fine-tuning for In-Domain Adaptation

For the in-domain setting, where the pretraining and fine-tuning target distribution are similar, low rank updates demonstrate superior performance. LoRA with the CFM objective matches the accuracy of full fine-tuning throughout training, and LoRA with the GFT objective reaches a similar accuracy after about 600 epochs (Figure 2)). However, GFT converges with significantly higher stability than CFM (Table 1). GFT achieves the lowest instantaneous variance, indicating low oscillation throughout fine-tuning, and a Spearman correlation coefficient is close to -1, pointing to strong monotonic convergence towards optimal model parameters. In addition to high accuracy and stable convergence, GFT achieves a lower average path length than both CFM methods, supporting efficient inference-time generation.

5.2 Sample Fine-tuning for Cross-Domain Adaptation

The evolution of FID in the cross-domain adaptation experiments demonstrates the benefits of GFT and the key differences between LoRA and full fine-tuning under large distribution shifts (Figure 3). LoRA fails to match the accuracy of full fine-tuning, indicating that low-rank updates are too restrictive to capture the substantial differences between the pretrained and fine-tuned target distributions. Although full fine-tuning with GFT requires more epochs to reach high accuracy generation than CFM, it attains comparable FID at the end of fine-tuning with notably improved stability (Table 1) and higher diversity (Table 4). This behavior is consistent with the use of the GFT gradual cooling schedule, which moderates model updates to achieve more stable optimization. In addition, GFT yields shorter average probability path lengths with lower variance (Figure 3), indicating better preservation of the pretrained vector field and pointing to more efficient generation at inference time.

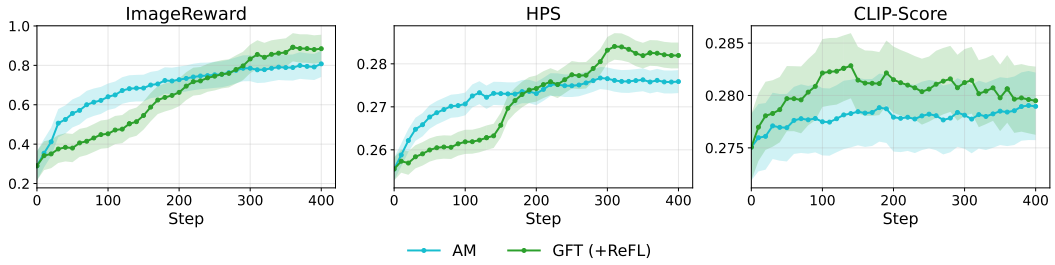


Figure 5: Reward Fine-Tuning comparison between GFT (+ReFL) and Adjoint Matching. The GFT-based method outperforms Adjoint Matching on both ImageReward and HPS, and maintains a higher CLIP score through 400 steps of fine-tuning.

Mean Flow GFT. We perform further Cross-Domain experiments using the MF-GFT algorithm (2) (Appendix M). Compared to standard MF fine-tuning, MF-GFT increases stability (Table 3) while reaching the same FID at the end of fine-tuning (Figure 16). Notably, the MF-GFT model almost matches the FID of standard CFM, with a 50-100x reduction in number of model evaluations.

5.3 Reward Fine-Tuning

Following Section 4.6, to evaluate reward fine-tuning with GFT we construct a training dataset using only the reward function and samples generated from the pretrained model (Algorithm 4), and perform sample-based GFT on this dataset. In this setting, GFT is a zeroth order method, as it relies only on the reward values rather than their gradients. This is a beneficial property in setting where the reward function is not differentiable, but creates an unfair comparison to Adjoint Matching, which is a first order method and therefore utilizes strictly more information throughout fine-tuning. To address this, we augment the GFT objective with a ReFL term (Xu et al., 2023) to incorporate reward gradients, resulting in the GFT (+ReFL) algorithm (Algorithm 5). We note that generating the high-reward dataset required 12 GPU hours, along with approximately 5 hours of GPU time for training GFT (+ReFL), giving a total of 17 GPU hours compared to a total of 163 GPU hours for Adjoint Matching (AM) (Domingo-Enrich et al., 2025) training which directly optimizes ImageReward (recall Figure 1). As shown in Figure 5, GFT-based training outperforms Adjoint Matching on both ImageReward and HPS, showing both strong reward optimization and high generalization. GFT also maintains a higher CLIPScore throughout fine-tuning. Combined with an order of magnitude reduction in total training cost (Figure 1), these results demonstrate that GFT provides an efficient and effective alternative to Adjoint Matching based reward fine-tuning. Additional results in the reward fine-tuning setting are shown in Appendix N.4.

6 Discussion

We have introduced **GFT**, a principled framework for fine-tuning flow generative models. Derived from an optimization over path space, GFT yields a closed-form optimal drift that interpolates between pretrained and target dynamics through an annealing cooling schedule, resulting in a method which is simple to implement yet demonstrates strong theoretical and empirical performance.

GFT addresses several critical gaps in the existing literature. It provably converges to the target distribution at a rate directly controllable through the cooling schedule, and exhibits highly stable and monotonic convergence compared to standard fine-tuning, while consistently maintaining lower trajectory length. GFT is also highly efficient, achieving an order of magnitude reduction in training cost compared to Adjoint Matching, and supports efficient inference through compatibility with OT couplings and few-step inference methods such as MeanFlow. Across all experimental settings, GFT matches or outperforms alternative fine-tuning methods in generation accuracy. Our results demonstrate that GFT offers a simple yet principled approach to adapting flow generative models under distribution shift.

REFERENCES

Michael S. Albergo, Nicholas M. Boffi, and Eric Vanden-Eijnden. Stochastic interpolants: A unifying framework for flows and diffusions, 2025. URL https://arxiv.org/abs/2303.08797.

Feng Bao, Zezhong Zhang, and Guannan Zhang. An ensemble score filter for tracking high-dimensional nonlinear dynamical systems. *Computer Methods in Applied Mechanics and Engineering*, 432:117447, 2024.

Heli Ben-Hamu, Omri Puny, Itai Gat, Brian Karrer, Uriel Singer, and Yaron Lipman. D-flow: Differentiating through flows for controlled generation. In *Forty-first International Conference on Machine Learning*, 2024. URL https://openreview.net/forum?id=SE20BFqj6J.

Julius Berner, Lorenz Richter, and Karen Ullrich. An optimal control perspective on diffusion-based generative modeling. *Transactions on Machine Learning Research*, 2024. ISSN 2835-8856. URL https://openreview.net/forum?id=oYIjw37pTP.

Kevin Black, Michael Janner, Yilun Du, Ilya Kostrikov, and Sergey Levine. Training diffusion models with reinforcement learning. In *The Twelfth International Conference on Learning Representations*, 2024. URL https://openreview.net/forum?id=YCWjhGrJFD.

Denis Blessing, Julius Berner, Lorenz Richter, Carles Domingo-Enrich, Yuanqi Du, Arash Vahdat, and Gerhard Neumann. Trust region constrained measure transport in path space for stochastic optimal control and inference. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2025. URL https://openreview.net/forum?id=6RlbOEcOS4.

Nicholas M. Boffi, Michael S. Albergo, and Eric Vanden-Eijnden. How to build a consistency model: Learning flow maps via self-distillation, 2025. URL https://arxiv.org/abs/2505.18825.

Péter Bándi, Oscar Geessink, Quirine Manson, Marcory Van Dijk, Maschenka Balkenhol, Meyke Hermsen, Babak Ehteshami Bejnordi, Byungjae Lee, Kyunghyun Paeng, Aoxiao Zhong, Quanzheng Li, Farhad Ghazvinian Zanjani, Svitlana Zinger, Keisuke Fukuta, Daisuke Komura, Vlado Ovtcharov, Shenghua Cheng, Shaoqun Zeng, Jeppe Thagaard, Anders B. Dahl, Huangjing Lin, Hao Chen, Ludwig Jacobsson, Martin Hedlund, Melih Çetin, Eren Halıcı, Hunter Jackson, Richard Chen, Fabian Both, Jörg Franke, Heidi Küsters-Vandevelde, Willem Vreuls, Peter Bult, Bram van Ginneken, Jeroen van der Laak, and Geert Litjens. From detection of individual metastases to classification of lymph node status at the patient level: The camelyon17 challenge. *IEEE Transactions on Medical Imaging*, 38(2):550–560, 2019. doi: 10.1109/TMI.2018.2867350.

Sourav Chatterjee and Persi Diaconis. The sample size required in importance sampling. *The Annals of Applied Probability*, 28(2):1099–1135, 2018.

Gordon Christie, Neil Fendley, James Wilson, and Ryan Mukherjee. Functional map of the world. In *CVPR*, 2018.

Kevin Clark, Paul Vicol, Kevin Swersky, and David J. Fleet. Directly fine-tuning diffusion models on differentiable rewards. In *The Twelfth International Conference on Learning Representations*, 2024. URL https://openreview.net/forum?id=1vmSEVL19f.

Oscar Davis, Samuel Kessler, Mircea Petrache, İsmail İlkan Ceylan, Michael Bronstein, and Avishek Joey Bose. Fisher flow matching for generative modeling over discrete data. In A. Globerson, L. Mackey, D. Belgrave, A. Fan, U. Paquet, J. Tomczak, and C. Zhang (eds.), *Advances in Neural Information Processing Systems*, volume 37, pp. 139054–139084. Curran Associates, Inc., 2024. doi: 10.52202/079017-4413. URL https://proceedings.neurips.cc/paper_files/paper/2024/file/fadec8f2e65f181d777507d1df69b92f\-Paper-Conference.pdf.

Carles Domingo-Enrich, Michal Drozdzal, Brian Karrer, and Ricky T. Q. Chen. Adjoint matching: Fine-tuning flow and diffusion generative models with memoryless stochastic optimal control. In *The Thirteenth International Conference on Learning Representations*, 2025. URL https://openreview.net/forum?id=xQBRrtQM8u.

Patrick Esser, Sumith Kulal, Andreas Blattmann, Rahim Entezari, Jonas Müller, Harry Saini, Yam Levi, Dominik Lorenz, Axel Sauer, Frederic Boesel, Dustin Podell, Tim Dockhorn, Zion English, and Robin Rombach. Scaling rectified flow transformers for high-resolution image synthesis. In Ruslan Salakhutdinov, Zico Kolter, Katherine Heller, Adrian Weller, Nuria Oliver, Jonathan Scarlett, and Felix Berkenkamp (eds.), *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 12606–12633. PMLR, 21–27 Jul 2024. URL https://proceedings.mlr.press/v235/esser24a.html.

Jiajun Fan, Chaoran Cheng, Shuaike Shen, Xiangxin Zhou, and Ge Liu. Fine-tuning flow matching generative models with intermediate feedback, 2025a. URL https://arxiv.org/abs/2510.18072.

Jiajun Fan, Shuaike Shen, Chaoran Cheng, Yuxin Chen, Chumeng Liang, and Ge Liu. Online reward-weighted fine-tuning of flow matching with wasserstein regularization. In *The Thirteenth International Conference on Learning Representations*, 2025b. URL https://openreview.net/forum?id=2IoFFexvuw.

Ying Fan, Olivia Watkins, Yuqing Du, Hao Liu, Moonkyung Ryu, Craig Boutilier, Pieter Abbeel, Mohammad Ghavamzadeh, Kangwook Lee, and Kimin Lee. Reinforcement learning for fine-tuning text-to-image diffusion models. In *Thirty-seventh Conference on Neural Information Processing Systems*, 2023. URL https://openreview.net/forum?id=8OTPepXzeh.

Stathi Fotiadis, Noah D Brenowitz, Tomas Geffner, Yair Cohen, Michael Pritchard, Arash Vahdat, and Morteza Mardani. Adaptive flow matching for resolving small-scale physics. In *Forty-second International Conference on Machine Learning*, 2025. URL https://openreview.net/forum?id=YJ1My9ttEN.

Zhengyang Geng, Mingyang Deng, Xingjian Bai, J Zico Kolter, and Kaiming He. Mean flows for one-step generative modeling. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2026. URL https://openreview.net/forum?id=uWj4s7rMnR.

Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In Z. Ghahramani, M. Welling, C. Cortes, N. Lawrence, and K.Q. Weinberger (eds.), *Advances in Neural Information Processing Systems*, volume 27. Curran Associates, Inc., 2014. URL https://proceedings.neurips.cc/paper_files/paper/2014/file/f033ed80deb0234979a61f95710dbe25-Paper.pdf.

Aaron J Havens, Benjamin Kurt Miller, Bing Yan, Carles Domingo-Enrich, Anuroop Sriram, Daniel S. Levine, Brandon M Wood, Bin Hu, Brandon Amos, Brian Karrer, Xiang Fu, Guan-Horng Liu, and Ricky T. Q. Chen. Adjoint sampling: Highly scalable diffusion samplers via adjoint matching. In *Forty-second International Conference on Machine Learning*, 2025. URL https://openreview.net/forum?id=6Eg1OrHmg2.

Yifei He, Haoxiang Wang, Bo Li, and Han Zhao. Gradual domain adaptation: Theory and algorithms. *Journal of Machine Learning Research*, 25(361):1–40, 2024. URL http://jmlr.org/papers/v25/23-1180.html.

Jack Hessel, Ari Holtzman, Maxwell Forbes, Ronan Le Bras, and Yejin Choi. CLIPScore: A reference-free evaluation metric for image captioning. In Marie-Francine Moens, Xuanjing Huang, Lucia Specia, and Scott Wen-tau Yih (eds.), *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pp. 7514–7528, Online and Punta Cana, Dominican Republic, November 2021. Association for Computational Linguistics. doi: 10.18653/v1/2021.emnlp-main.595. URL https://aclanthology.org/2021.emnlp-main.595/.

Jonathan Ho and Tim Salimans. Classifier-free diffusion guidance. In *NeurIPS 2021 Workshop on Deep Generative Models and Downstream Applications*, 2021. URL https://openreview.net/forum?id=qw8AKxfYbI.

Jonathan Ho and Tim Salimans. Classifier-free diffusion guidance, 2022. URL https://arxiv.org/abs/2207.12598.

Edward J Hu, yelong shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. LoRA: Low-rank adaptation of large language models. In *International Conference on Learning Representations*, 2022. URL https://openreview.net/forum?id=nZeVKeeFYf9.

Bowen Jing, Bonnie Berger, and Tommi Jaakkola. Alphafold meets flow matching for generating protein ensembles. In *Forty-first International Conference on Machine Learning*, 2024.

Ioannis Karatzas and Steven E. Shreve. *Brownian motion and stochastic calculus*. Springer, 2 edition, 1991.

Diederik P. Kingma and Max Welling. Auto-Encoding Variational Bayes. In *2nd International Conference on Learning Representations, ICLR 2014, Banff, AB, Canada, April 14-16, 2014, Conference Track Proceedings*, 2014.

S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi. Optimization by simulated annealing. *Science*, 220 (4598):671–680, 1983. doi: 10.1126/science.220.4598.671. URL https://www.science.org/doi/abs/10.1126/science.220.4598.671.

Leon Klein, Andreas Krämer, and Frank Noe. Equivariant flow matching. In A. Oh, T. Naumann, A. Globerson, K. Saenko, M. Hardt, and S. Levine (eds.), *Advances in Neural Information Processing Systems*, volume 36, pp. 59886–59910. Curran Associates, Inc., 2023. URL https://proceedings.neurips.cc/paper_files/paper/2023/file/bc827452450356f9f558f4e4568d553b\-Paper-Conference.pdf.

Pang Wei Koh, Shiori Sagawa, Henrik Marklund, Sang Michael Xie, Marvin Zhang, Akshay Balsubramani, Weihua Hu, Michihiro Yasunaga, Richard Lanas Phillips, Irena Gao, Tony Lee, Etienne David, Ian Stavness, Wei Guo, Berton A. Earnshaw, Imran S. Haque, Sara Beery, Jure Leskovec, Anshul Kundaje, Emma Pierson, Sergey Levine, Chelsea Finn, and Percy Liang. WILDS: A benchmark of in-the-wild distribution shifts. In *International Conference on Machine Learning (ICML)*, 2021.

Alex Krizhevsky. Learning multiple layers of features from tiny images. Technical report, 2009.

Zhanhao Liang, Yuhui Yuan, Shuyang Gu, Bohan Chen, Tiankai Hang, Mingxi Cheng, Ji Li, and Liang Zheng. Aesthetic post-training diffusion models from generic preferences with step-by-step preference optimization, 2025. URL https://arxiv.org/abs/2406.04314.

Yaron Lipman, Ricky T. Q. Chen, Heli Ben-Hamu, Maximilian Nickel, and Matthew Le. Flow matching for generative modeling. In *The Eleventh International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=PqvMRDCJT9t.

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

Aram-Alexandre Pooladian, Heli Ben-Hamu, Carles Domingo-Enrich, Brandon Amos, Yaron Lipman, and Ricky T. Q. Chen. Multisample flow matching: straightening flows with minibatch couplings. In *Proceedings of the 40th International Conference on Machine Learning*, ICML'23. JMLR.org, 2023.

Peter Potaptchik, Cheuk-Kit Lee, and Michael S. Albergo. Tilt matching for scalable sampling and fine-tuning, 2025. URL https://arxiv.org/abs/2512.21829.

Mihir Prabhudesai, Anirudh Goyal, Deepak Pathak, and Katerina Fragkiadaki. Aligning text-to-image diffusion models with reward backpropagation, 2023.

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 (CVPR)*, pp. 10684–10695, June 2022.

John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. Proximal policy optimization algorithms. *CoRR*, abs/1707.06347, 2017. URL http://arxiv.org/abs/1707.06347.

Yang Song, Conor Durkan, Iain Murray, and Stefano Ermon. Maximum likelihood training of score-based diffusion models. In *Thirty-Fifth Conference on Neural Information Processing Systems*, 2021a.

Yang Song, Jascha Sohl-Dickstein, Diederik P Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. In *International Conference on Learning Representations*, 2021b. URL https://openreview.net/forum?id=PxTIG12RRHS.

Yuxuan Song, Jingjing Gong, Minkai Xu, Ziyao Cao, Yanyan Lan, Stefano Ermon, Hao Zhou, and Wei-Ying Ma. Equivariant flow matching with hybrid probability transport for 3d molecule generation. In A. Oh, T. Naumann, A. Globerson, K. Saenko, M. Hardt, and S. Levine (eds.), *Advances in Neural Information Processing Systems*, volume 36, pp. 549–568. Curran Associates, Inc., 2023. URL https://proceedings.neurips.cc/paper_files/paper/2023/file/01d64478381c33e29ed611f1719f5a37-Paper\-Conference.pdf.

Hannes Stark, Bowen Jing, Chenyu Wang, Gabriele Corso, Bonnie Berger, Regina Barzilay, and Tommi Jaakkola. Dirichlet flow matching with applications to DNA sequence design. In *ICLR 2024 Workshop on Machine Learning for Genomics Explorations*, 2024. URL https://openreview.net/forum?id=ehYe5bz8H3.

Maciej Sypetkowski, Morteza Rezanejad, Saber Saberian, Oren Kraus, John Urbanik, James Taylor, Ben Mabey, Mason Victors, Jason Yosinski, Alborz Rezazadeh Sereshkeh, Imran Haque, and Berton Earnshaw. RxRx1: A Dataset for Evaluating Experimental Batch Correction Methods . In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 4285–4294, Los Alamitos, CA, USA, June 2023. IEEE Computer Society. doi: 10.1109/CVPRW59228.2023.00451. URL https://doi.ieeecomputersociety.org/10.1109/CVPRW59228.2023.00451.

Alexander Tong, Nikolay Malkin, Kilian FATRAS, Lazar Atanackovic, Yanlei Zhang, Guillaume Huguet, Guy Wolf, and Yoshua Bengio. Simulation-free schrödinger bridges via score and flow matching. In *ICML Workshop on New Frontiers in Learning, Control, and Dynamical Systems*, 2023. URL https://openreview.net/forum?id=adkj23mvB0.

Alexander Tong, Kilian Fatras, Nikolay Malkin, Guillaume Huguet, Yanlei Zhang, Jarrid Rector-Brooks, Guy Wolf, and Yoshua Bengio. Improving and generalizing flow-based generative models with minibatch optimal transport. *Transactions on Machine Learning Research*, 2024. ISSN 2835-8856. URL https://openreview.net/forum?id=CD9Snc73AW. Expert Certification.

Masatoshi Uehara, Yulai Zhao, Kevin Black, Ehsan Hajiramezanali, Gabriele Scalia, Nathaniel Lee Diamant, Alex M Tseng, Tommaso Biancalani, and Sergey Levine. Fine-tuning of continuous-time diffusion models as entropy-regularized control, 2024. URL https://arxiv.org/abs/2402.15194.

C. Villani. *Optimal Transport: Old and New*. Grundlehren der mathematischen Wissenschaften. Springer Berlin Heidelberg, 2008. ISBN 9783540710493. URL https://books.google.com/books?id=NZXiNAEACAAJ.

Bram Wallace, Meihua Dang, Rafael Rafailov, Linqi Zhou, Aaron Lou, Senthil Purushwalkam, Stefano Ermon, Caiming Xiong, Shafiq Joty, and Nikhil Naik. Diffusion model alignment using direct preference optimization, 2023.

Xiaoshi Wu, Keqiang Sun, Feng Zhu, Rui Zhao, and Hongsheng Li. Human preference score: Better aligning text-to-image models with human preference. In *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 2096–2105, 2023. doi: 10.1109/ICCV51070.2023.00200.

Jiazheng Xu, Xiao Liu, Yuchen Wu, Yuxuan Tong, Qinkai Li, Ming Ding, Jie Tang, and Yuxiao Dong. Imagereward: learning and evaluating human preferences for text-to-image generation. In *Proceedings of the 37th International Conference on Neural Information Processing Systems*, pp. 15903–15935, 2023.

A AI USE STATEMENT

AI-based tools were used to assist with this work in the following ways:

1. writing, debugging, and modifying code for experiments,
2. organizing experimental results and developing plotting code,
3. generating LaTeX tables and figures based on the experimental results.

The authors take full responsibility for the accuracy and correctness of all information presented in the paper.

B ETHICS STATEMENT

This paper presents theoretical and methodological work with the primary goal of advancing the field of machine learning by improving the understanding and fine-tuning of flow-based generative models. The proposed method and accompanying analysis are general-purpose, and do not target a specific application domain. As such, the potential societal impacts of this work are indirect and depend on possible downstream use-cases, which may include both beneficial and harmful applications. We do not foresee any immediate ethical concerns directly related to this work, beyond those already broadly understood for generative models.

C REPRODUCIBILITY STATEMENT

All theorems presented in this paper are proven in the appendix, with all relevant assumptions stated explicitly. Complete implementation details for all experiments are provided either in the main paper or in Appendices O and N.2, and the code used for major experiments has been included with this submission.

D SOC-BASED FINE-TUNING AND NON-LINEAR METHODS

Adjoint Matching and many other SOC-based methods are incompatible specifically with nonlinear fine-tuning techniques. The main reason for this is the fundamental difference between output space linearity and parameter space linearity. Methods such as Adjoint Matching parameterize the fine-tuned vector field as an additive correction in the output space of the models,

$$v_{ft}(x,t) = v_{\theta_0}(x,t) + u_\phi(x,t) \tag{14}$$

where v_{θ_0} is the fixed pretrained model and u_ϕ is the control vector field that is optimized during fine-tuning. Concretely, these methods train u_ϕ to match $v_{ft} - v_{\theta_0}$, which is only a meaningful training target when the fine-tuned vector field decomposes additively in the output space. This output space linearity is therefore an inherent structural assumption used during training.

In contrast, many popular fine-tuning methods, including full fine-tuning and LoRA, make updates in the parameter space of a model,

$$\theta = \theta_0 + \Delta\theta\ , \tag{15}$$

where $\Delta\theta$ differs based on the chosen fine-tuning method. Crucially, this parameter space linearity does not translate to output space linearity. In general,

$$v_{\theta_0+\Delta\theta} \neq v_{\theta_0} + v_{\Delta\theta} \tag{16}$$

due to the nonlinear relationship between a neural network's parameters and its output. Since Adjoint Matching and similar methods require output space additivity, they are not compatible with nonlinear fine-tuning methods in the above sense, as they make updates directly in the parameter space.

E Recovering the Score in the Stochastic View of Flow Matching

Substituting the continuity equation (1) into the Fokker-Planck equation (2) yields

$$\nabla_x \cdot (p_t v_t) = \nabla_x \cdot (\tilde{v}_t p_t) - \frac{1}{2}\nabla_x \cdot \nabla_x \cdot (\sigma_t^2 I p_t). \tag{17}$$

Since the diffusion coefficient σ_t^2 depends only on the timestep t, we can move it outside the divergence operators.

$$\nabla_x \cdot (p_t v_t) = \nabla_x \cdot (\tilde{v}_t p_t) - \frac{1}{2}\sigma_t^2(\nabla_x \cdot \nabla_x \cdot p_t) \tag{18}$$

We rearrange terms to isolate the drift term divergence.

$$\nabla_x \cdot (\tilde{v}_t p_t) = \nabla_x \cdot (p_t v_t) + \frac{1}{2}\sigma_t^2(\nabla_x \cdot \nabla_x \cdot p_t) \tag{19}$$

As the divergence operator commutes with scalar multiplication, the terms on the right hand side can be combined.

$$\nabla_x \cdot (\tilde{v}_t p_t) = \nabla_x \cdot \left[p_t v_t + \frac{1}{2}\sigma_t^2 \nabla_x p_t \right] \tag{20}$$

Two vector fields have equivalent divergence when they differ only by a divergence-free component Z_t, such that $\nabla_x \cdot (p_t Z_t) = 0$. Therefore, this equivalence can be rewritten without the leading divergence operator.

$$\tilde{v}_t p_t = p_t v_t + \frac{1}{2}\sigma_t^2 \nabla_x p_t + p_t Z_t \tag{21}$$

Finally, we divide by p_t, using the identity that $\frac{\nabla_x p_t}{p_t} = \nabla \log p_t$.

$$\tilde{v}_t = v_t + \frac{1}{2}\sigma_t^2 \nabla_x \log p_t + Z_t \tag{22}$$

Z_t is a divergence-free velocity that can simply be set to 0 to yields the final expression.

F Derivation of Path Length by KL Divergence

We derive a distance measure between the probability distribution of the paths generated by two SDEs using Girsanov's Theorem. Let $\mathbb{P}_f$ and $\mathbb{P}_g$ be the path measures of two SDEs on the time interval $[0, 1]$, with initial samples from the same source distribution.

$$\mathbb{P}_f : \quad dX_t = f_t(X_t, t)dt + \sigma_t dB_t, \qquad X_0 \sim p_0 \tag{23}$$

$$\mathbb{P}_g : \quad dX_t = g_t(X_t, t)dt + \sigma_t dB_t, \qquad X_0 \sim p_0 \tag{24}$$

We assume that $\mathbb{P}_f$ and $\mathbb{P}_g$ are mutually absolutely continuous, and that the diffusion coefficient σ_t is non-zero. The KL-divergence between the probability measures of these processes is

$$KL(\mathbb{P}_f \| \mathbb{P}_g) = \mathbb{E}_{\mathbb{P}_f}\left[\log \frac{d\mathbb{P}_f}{d\mathbb{P}_g} \right] = -\mathbb{E}_{\mathbb{P}_f}\left[\log \frac{d\mathbb{P}_g}{d\mathbb{P}_f} \right]. \tag{25}$$

Girsanov's Theorem provides the formulation for the Radon-Nikodym derivative $\frac{d\mathbb{P}_g}{d\mathbb{P}_f}$, which specifies how the probability measure $\mathbb{P}_g$ changes with respect to $\mathbb{P}_f$.

$$\frac{d\mathbb{P}_g}{d\mathbb{P}_f} = \exp\left(\int_0^1 \sigma_t^{-1}(g_t - f_t)^\intercal dB_t - \frac{1}{2}\int_0^1 \|\sigma_t^{-1}(g_t - f_t)\|^2 dt\right) \tag{26}$$

$$\log\left(\frac{d\mathbb{P}_g}{d\mathbb{P}_f}\right) = \int_0^1 \sigma_t^{-1}(g_t - f_t)^\intercal dB_t - \frac{1}{2}\int_0^1 \|\sigma_t^{-1}(g_t - f_t)\|^2 dt \tag{27}$$

We now take the expectation with respect to the target process.

$$KL(\mathbb{P}_f\|\mathbb{P}_g) = -\mathbb{E}_{\mathbb{P}_f}\left[\int_0^1 \sigma_t^{-1}(g_t - f_t)^\intercal dB_t\right] + \mathbb{E}_{\mathbb{P}_f}\left[\frac{1}{2}\int_0^1 \|\sigma_t^{-1}(g_t - f_t)\|^2 dt\right] \tag{28}$$

Note that since the integral in the first term is with respect to a Brownian motion, this term is a martingale under the measure $\mathbb{P}_f$, and has an expected value of 0. Therefore, our final expression for Proposition 4.2 is

$$KL(\mathbb{P}_f\|\mathbb{P}_g) = \mathbb{E}_{\mathbb{P}_f}\left[\frac{1}{2}\int_0^1 \|\sigma_t^{-1}(f_t - g_t)\|^2 dt\right]. \tag{29}$$

G Relationship Between Path Length and Fisher Information

Consider the practical pretrained and fine-tuned processes

$$\mathbb{P}_{\theta_0}: \quad dX_t = v_{\theta_0}(X_t, t)dt + \sigma_t dB_t, \qquad X_0 \sim p_0 \tag{30}$$

$$\mathbb{P}_{\theta}: \quad dX_t = v_{\theta}(X_t, t)dt + \sigma_t dB_t, \qquad X_0 \sim p_0 \tag{31}$$

Applying Proposition 4.2 to these processes gives us a measure of the divergence between the probability distribution of their generated paths. In this section we set $\sigma_t = 1$ for simplicity.

$$KL(\mathbb{P}_\theta\|\mathbb{P}_{\theta_0}) = \mathbb{E}_{\mathbb{P}_\theta}\left[\frac{1}{2}\int_0^1 \|(v_\theta - v_{\theta_0})\|^2 dt\right] \tag{32}$$

With fine-tuning methods such as LoRA and head adaptation, we can assume that $\mathbb{P}_\theta \approx \mathbb{P}_{\theta_0}$ throughout the fine-tuning process, and we can further simplify this expression using first-order Taylor approximation.

$$v_\theta - v_{\theta_0} \approx \mathbf{J}_t(X_t, t) \cdot (\theta - \theta_0) \tag{33}$$

Here, $\mathbf{J}_t(X_t, t) = \nabla_\theta v_\theta(X_t, t)\big|_{\theta_0} \in \mathbb{R}^{D\times P}$ is the Jacobian matrix of the vector field v with respect to parameters θ, evaluated at θ_0. Note that when fine-tuning with LoRA or head adaptation, $\Delta\theta = \theta - \theta_0$, which is the guidance of the fine-tuning process. Substituting this back into the KL expression yields

$$KL(\mathbb{P}_\theta\|\mathbb{P}_{\theta_0}) \approx \mathbb{E}_{\mathbb{P}_\theta}\left[\frac{1}{2}\int_0^1 \|\mathbf{J}_t\Delta\theta\|^2 dt\right] \tag{34}$$

$$= \mathbb{E}_{\mathbb{P}_\theta}\left[\frac{1}{2}\int_0^1 (\Delta\theta^\intercal \mathbf{J}_t^\intercal \mathbf{J}_t \Delta\theta) dt\right]. \tag{35}$$

Note that the guidance $\Delta\theta$ is not dependent on x_t or t.

$$= \frac{1}{2}\Delta\theta^\intercal \left(\int_0^1 \mathbb{E}_{\mathbb{P}_\theta} \left[\mathbf{J}_t^\intercal \mathbf{J}_t \right] dt \right) \Delta\theta \tag{36}$$

Notice that the integral term is related to the Fisher Information matrix centered at θ_0, $\mathbf{F}_{\theta_0} \in \mathbb{R}^{P \times P}$.

$$\mathbf{F}_{\theta_0} = \int_0^1 \mathbb{E}_{\mathbb{P}_\theta} [\mathbf{J}_t^\intercal \mathbf{J}_t] dt \tag{37}$$

This is the time-integrated expected Fisher Information matrix over the time range $[0, 1]$ for the path measure $\mathbb{P}_\theta$. For practical implementation, we can rewrite the integral as an expectation over a uniformly sampled time $\tau \sim \mathcal{U}[0, 1]$.

$$\mathbf{F}_{\theta_0} = \mathbb{E}_{\tau \sim \mathcal{U}[0,1], X_\tau \sim \mathbb{P}_\theta} [\mathbf{J}_\tau^\intercal \mathbf{J}_\tau] \tag{38}$$

Therefore, the final simplified approximated divergence between the probability distribution of the paths generated by the base and fine-tuned models is

$$KL(\mathbb{P}_\theta \| \mathbb{P}_{\theta_0}) \approx \frac{1}{2}\Delta\theta^\intercal \mathbf{F}_{\theta_0} \Delta\theta. \tag{39}$$

H Minimizing the Gradual Fine-Tuning Objective

The GFT objective function is defined as

$$\mathcal{L}(\theta) = \text{KL}(\mathbb{P}_\theta \| \mathbb{P}_q) + \beta \text{KL}(\mathbb{P}_\theta \| \mathbb{P}_{\theta_0}) \tag{40}$$

Applying Proposition 4.2 to both terms, with $t \sim \mathcal{U}[0, 1]$ throughout, yields

$$\mathcal{L}(\theta) = \mathbb{E}_{X_t \sim \mathbb{P}_\theta} \left[\frac{1}{2} \| \sigma_t^{-1} (v_\theta(X_t, t) - v_q(X_t, t)) \|^2 \right] + \beta \cdot \mathbb{E}_{X_t \sim \mathbb{P}_\theta} \left[\frac{1}{2} \| \sigma_t^{-1} (v_\theta(X_t, t) - v_{\theta_0}(X_t, t)) \|^2 \right] \tag{41}$$

$$= \mathbb{E}_{X_t \sim \mathbb{P}_\theta} \left[\frac{1}{2} \| \sigma_t^{-1} (v_\theta(X_t, t) - v_q(X_t, t)) \|^2 + \frac{\beta}{2} \| \sigma_t^{-1} (v_\theta(X_t, t) - v_{\theta_0}(X_t, t)) \|^2 \right] \tag{42}$$

Minimizing this function over $\mathbb{P}_\theta$ is equivalent to finding the optimal drift v_θ^*. We do this by minimizing the expression inside the expectation.

$$C = \frac{1}{2} \| \sigma_t^{-1} (v_\theta(X_t, t) - v_q(X_t, t)) \|^2 + \frac{\beta}{2} \| \sigma_t^{-1} (v_\theta(X_t, t) - v_{\theta_0}(X_t, t)) \|^2 \tag{43}$$

We define $\Sigma_t = \sigma_t \sigma_t^\intercal$.

$$C = \frac{1}{2} (v_\theta(X_t, t) - v_q(X_t, t))^\intercal \Sigma_t^{-1} (v_\theta(X_t, t) - v_q(X_t, t)) + \frac{\beta}{2} (v_\theta(X_t, t) - v_{\theta_0}(X_t, t))^\intercal \Sigma_t^{-1} (v_\theta(X_t, t) - v_{\theta_0}(X_t, t)) \tag{44}$$

$$\frac{dC}{dv_\theta} = \Sigma_t^{-1} (v_\theta(X_t, t) - v_q(X_t, t)) + \beta \Sigma_t^{-1} (v_\theta(X_t, t) - v_{\theta_0}(X_t, t)) = 0 \tag{45}$$

$$\Sigma_t^{-1} \left((v_\theta(X_t, t) - v_q(X_t, t)) + \beta (v_\theta(X_t, t) - v_{\theta_0}(X_t, t)) \right) = 0 \tag{46}$$

$$(1+\beta)v_\theta(X_t,t) - v_q(X_t,t) - \beta v_{\theta_0}(X_t,t) = 0 \tag{47}$$

$$v_\theta^*(X_t,t) = \left(\frac{1}{1+\beta}\right) v_q(X_t,t) + \left(\frac{\beta}{1+\beta}\right) v_{\theta_0}(X_t,t) \tag{48}$$

While the expectation in the objective function depends on v_θ through the path measure $\mathbb{P}_\theta$, the pointwise minimizer of the integrand remains consistent. Since the optimizer v_θ^* is a convex combination of continuous vector fields, the resulting optimal path measure $\mathbb{P}_\theta^*$ is well-defined and achieves the lower bound of the objective.

I Gradient Equivalence between Total and OT-Conditional KL Divergence

Let $\pi \in \Pi(p_0, q)$ be any coupling between the source and target distributions that admits the marginals $\int \pi(X_0, X_1)dX_1 = p_0(X_0)$ and $\int \pi(X_0, X_1)dX_0 = q(X_1)$. For each drawn pair $Z := (X_0, X_1) \sim \pi$, we define the conditional probability path $p_t(X_t|Z)$ with the corresponding conditional vector field $v_t(X_t|Z)$, which must satisfy the continuity equation.

$$\partial_t p_t(X_t|Z) + \nabla_{X_t} \cdot (p_t(X_t|Z)v_t(X_t|Z)) = 0 \tag{49}$$

The unconditional path measure $\mathbb{P}_\theta$ is then obtained by marginalizing over π.

$$p_t(X_t) = \int p_t(X_t|Z)d\pi(Z) \tag{50}$$

$$v_t(X_t) = \frac{\int v_t(X_t|Z)p_t(X_t|Z)d\pi(Z)}{p_t(X_t)} \tag{51}$$

We use this to define the conditional GFT objective for an arbitrary coupling π.

$$\mathcal{L}_\pi(\theta) = \mathbb{E}_{Z\sim\pi}\left[KL(\mathbb{P}_\theta(\cdot|Z)\|\mathbb{P}_q(\cdot|Z)) + \beta KL(\mathbb{P}_\theta(\cdot|Z)\|\mathbb{P}_{\theta_0}(\cdot|Z))\right] \tag{52}$$

Here, $\mathbb{P}_\theta$ denotes the path measure induced by the conditional SDE:

$$dX_t = v_\theta(X_t, t|Z)dt + \sigma_t dB_t. \tag{53}$$

The conditional path measure $\mathbb{P}_q$ is defined similarly using the unknown conditional target vector field $v_q(X_t, t|Z)$. Note that despite this new conditioning, the base process $\mathbb{P}_{\theta_0}$ remains fixed. We apply Proposition 4.2 to the conditional GFT objective to write it in terms of these conditional vector fields.

$$\mathcal{L}_\pi(\theta) = \mathbb{E}_{Z\sim\pi}\mathbb{E}_{X_{[0,1]}\sim\mathbb{P}_\theta(\cdot|Z)}\left[\frac{1}{2}\int_0^1 \|\sigma_t^{-1}(v_\theta(X_t,t|Z) - v_q(X_t,t|Z))\|^2 dt + \frac{\beta}{2}\int_0^1 \|\sigma_t^{-1}(v_\theta(X_t,t|Z) - v_{\theta_0}(X_t,t|Z))\|^2 dt\right] \tag{54}$$

$$= \mathbb{E}_{Z\sim\pi}\mathbb{E}_{X_{[0,1]}\sim\mathbb{P}_\theta(\cdot|Z)}\left[\int_0^1 \frac{1}{2}\left[\|\sigma_t^{-1}(v_\theta(X_t,t|Z) - v_q(X_t,t|Z))\|^2 dt + \beta\|\sigma_t^{-1}(v_\theta(X_t,t|Z) - v_{\theta_0}(X_t,t|Z))\|^2\right]dt\right] \tag{55}$$

Following the same logic used to minimize the unconditional objective (Appendix H) the integrand is minimized by the following conditional drift pointwise for each (X_0, X_1), yielding the conditional minimizer.

$$v_\theta^*(X_t, t|Z) = \left(\frac{1}{1+\beta}\right) v_q(X_t, t|Z) + \left(\frac{\beta}{1+\beta}\right) v_{\theta_0}(X_t, t|Z) \tag{56}$$

Note that when the unconditional vector fields v_q and v_{θ_0} are defined as the marginalization of their conditional counterparts over the same coupling π, the unconditional optimum is precisely the marginalization of the conditional optimum. However, this optimum is still not usable for training, as it is written in terms of the unknown target vector field v_q. We can solve this intractable dependency by following CFM Lipman et al. (2023), and replacing v_q with a chosen conditional vector field which marginalizes to the true target:

$$\mathbb{E}_{Z\sim\pi}[v_q(X_t, t|Z)] = v_q(X_t, t). \tag{57}$$

Crucially, the fine-tuned model v_θ is unconditional, and is trained to approximate the expected marginal optimal vector field

$$v_\theta(X_t, t) \approx \mathbb{E}_{Z\sim\pi}[v_\theta^*(X_t, t|Z)]. \tag{58}$$

Applying Proposition 4.2 and substituting the chosen tractable conditional target vector field, we get the final GFT loss:

$$\mathcal{L}_\pi(\theta) = \mathbb{E}_{Z\sim\pi}\mathbb{E}_{X_t\sim\mathbb{P}(\cdot|Z)}\Bigg[\frac{1}{2}\int_0^1 (\|\sigma_t^{-1}(v_\theta(X_t, t) - v_q(X_t, t|Z))\|^2 + \beta\|\sigma_t^{-1}(v_\theta(X_t, t) - v_{\theta_0}(X_t, t|Z))\|^2)dt\Bigg]. \tag{59}$$

As a result, the gradient of the conditional objective $\mathcal{L}_\pi$ with respect to the model parameters θ will coincide with the gradient of the original unconditional objective $\mathcal{L}$ when the model is parameterized to satisfy the marginalization property. This follows from the same arguments used to justify the original CFM objective Lipman et al. (2023). Differentiating with respect to the model parameters θ gives

$$\nabla_\theta\mathcal{L}_\pi(\theta) = \mathbb{E}_{Z\sim\pi}\mathbb{E}_{X_{[0,1]}\sim\mathbb{P}(\cdot|Z)}\Bigg[\int_0^1 \langle\Sigma_t^{-1}\left[(v_\theta(X_t, t) - v_q(X_t, t|Z)) + \beta(v_\theta(X_t, t) - v_{\theta_0}(X_t, t|Z))\right], \nabla_\theta v_\theta(X_t, t)\rangle dt\Bigg]. \tag{60}$$

By the construction of the conditional vector fields, for $\phi = \theta, \theta_0, q$, we have

$$v_\phi(X_t, t) = \mathbb{E}_{Z\sim\pi}\left[v_\phi(X_t, t|Z)|X_t\right]. \tag{61}$$

Using these identities and applying the tower property of conditional expectations,

$$\nabla_\theta\mathcal{L}_\pi(\theta) = \mathbb{E}_{X_{[0,1]}\sim\mathbb{P}}\Bigg[\int_0^1 \langle\Sigma_t^{-1}\left[(v_\theta(X_t, t) - \mathbb{E}_\pi[v_q(X_t, t|X_t)]) + \beta(v_\theta(X_t, t) - \mathbb{E}_\pi[v_{\theta_0}(X_t, t|X_t)])\right], \nabla_\theta v_\theta(X_t, t)\rangle dt\Bigg]. \tag{62}$$

This is exactly the gradient of $\mathcal{L}(\theta)$ when the unconditional objective uses the marginal vector fields, so we can conclude that

$$\nabla_\theta \mathcal{L}_\pi(\theta) = \nabla_\theta \mathcal{L}(\theta). \tag{63}$$

J Practical Implementation of Gradual Fine-Tuning

Our proposed objective function consists of two KL divergence terms.

$$\mathcal{L}(\theta) = KL(\mathbb{P}_\theta \| \mathbb{P}_q) + \beta KL(\mathbb{P}_\theta \| \mathbb{P}_{\theta_0}) \tag{64}$$

We can therefore apply Proposition 4.2 to both terms to rewrite the expression.

$$KL(\mathbb{P}_\theta \| \mathbb{P}_q) = \mathbb{E}_{\mathbb{P}_\theta}\left[\frac{1}{2}\int_0^1 \|\sigma_t^{-1}(v_\theta - v_q)\|^2 dt\right] \tag{65}$$

$$KL(\mathbb{P}_\theta \| \mathbb{P}_{\theta_0}) = \mathbb{E}_{\mathbb{P}_\theta}\left[\frac{1}{2}\int_0^1 \|\sigma_t^{-1}(v_\theta - v_{\theta_0})\|^2 dt\right] \tag{66}$$

Here, v_{θ_0} is the pretrained base vector field, and v_q is the optimal vector field that transports samples from the source distribution p_0 to the target fine-tuning distribution q. Since this vector field is unknown, it can be defined in the same way that the CFM Lipman et al. (2023) defines a target vector field to be matched. We show empirically that the diffusion coefficient σ_t can simply be set to 1 while maintaining high performance. Note that both terms of the objective are now written using the drift terms of the base and target processes, and that the parameters θ directly determine the path measure $\mathbb{P}_\theta$. The otherwise intractable objective therefore now optimizes directly over the parameters θ. We can now write a tractable version of the proposed objective.

$$\mathcal{L}(\theta) = \mathbb{E}_{\mathbb{P}_\theta}\left[\frac{1}{2}\int_0^1 \|\sigma_t^{-1}(v_\theta - v_q)\|^2 dt\right] + \beta \cdot \mathbb{E}_{\mathbb{P}_\theta}\left[\frac{1}{2}\int_0^1 \|\sigma_t^{-1}(v_\theta - v_{\theta_0})\|^2 dt\right] \tag{67}$$

K Simulation-Free GFT

We begin by restating Girsanov's theorem.

Lemma K.1 *Suppose that X solves $dX_t = b_0(X_t, t)\,dt + \sigma_t\,dW_t^0$ under some reference measure $\mathbb{P}_0$, where σ_t is invertible, and let*

$$h_t := \sigma_t^{-1}\big(b_1(X_t, t) - b_0(X_t, t)\big) \tag{68}$$

for some other drift b_1. We define the process

$$Z_t := \exp\left(\int_0^t h_s \cdot dW_s^0 \;-\; \frac{1}{2}\int_0^t \|h_s\|^2\,ds\right), \qquad Z_0 = 1. \tag{69}$$

If h satisfies Novikov's condition,

$$\mathbb{E}_{P_0}\left[\exp\left(\frac{1}{2}\int_0^1 \|h_s\|^2\,ds\right)\right] < \infty, \tag{70}$$

then Z_1 has $\mathbb{E}_{\mathbb{P}_0}[Z_1] = 1$, and defining a new measure $\mathbb{P}_1$ by $d\mathbb{P}_1 := Z_1\,d\mathbb{P}_0$ makes X solve $dX_t = b_1(X_t, t)\,dt + \sigma_t\,dB_t^1$ under $\mathbb{P}_1$, for some $\mathbb{P}_1$-Brownian motion B^1. In particular, Z_1 is the Radon-Nikodym derivative $d\mathbb{P}_1/d\mathbb{P}_0$.

We now apply this with $\mathbb{P}_0 = \mathbb{P}_q(\cdot|Z)$, $b_0 = v_q(\cdot,\cdot|Z)$, $B^0 = B^q$, $b_1 = v_\theta(\cdot,\cdot|Z)$, $h_t = \sigma_t^{-1}\Delta_\theta(X_t, t|Z)$, and $\mathbb{P}_1 = \mathbb{P}_\theta(\cdot|Z)$. For simplicity, we have defined a shorthand notation for the drift mismatch between the current fine-tuned drift and the target drift:

$$\Delta_\theta(X_t, t|Z) := v_\theta(X_t, t|Z) - v_q(X_t, t|Z)\,. \tag{71}$$

This gives the Radon-Nikodym derivative

$$w_\theta(X|Z) \;:=\; \frac{d\mathbb{P}_\theta(\cdot|Z)}{d\mathbb{P}_q(\cdot|Z)}(X) \;=\; \exp\!\left(\int_0^1 \sigma_t^{-1}\Delta_\theta(X_t,t|Z)\cdot dB_t^q \;-\; \frac{1}{2}\int_0^1 \left\|\sigma_t^{-1}\Delta_\theta(X_t,t|Z)\right\|^2 dt\right). \tag{72}$$

With the definition of the Girsanov density w_θ established, we can apply an exact change of measure operation to the simulation-based GFT objective:

$$\mathcal{L}_\pi(\theta) = \frac{1}{2}\mathbb{E}_{Z\sim\pi,\; t\sim\mathcal{U}[0,1],\; X_t\sim\mathbb{P}_\theta(\cdot|Z)}\left[\|\sigma_t^{-1}(v_\theta(X_t,t)-v_q(X_t,t|Z))\|^2+\beta\|\sigma_t^{-1}(v_\theta(X_t,t)-v_{\theta_0}(X_t,t))\|^2\right] \tag{73}$$

$$= \frac{1}{2}\mathbb{E}_{Z\sim\pi,\; t\sim\mathcal{U}[0,1],\; X_t\sim\mathbb{P}_q(\cdot|Z)}\left[w_\theta(X|Z)(\|\sigma_t^{-1}(v_\theta(X_t,t)-v_q(X_t,t|Z))\|^2+\beta\|\sigma_t^{-1}(v_\theta(X_t,t)-v_{\theta_0}(X_t,t))\|^2)\right]. \tag{74}$$

Note that this change of measure is an exact equality, and not a bound. The simulation-free version of the GFT objective is therefore exactly equivalent to the simulation-based version.

K.1 Simulation-Free GFT Algorithms and Results

To implement the simulation-free GFT objective (12) without requiring evaluation of the Girsanov density at every training step, we propose a Moving Proposal GFT Algorithm (1). Rather than using the conditional target process $\mathbb{P}_q(\cdot|Z)$ as the reference measure for the Girsanov weight throughout the entire training process, this algorithm periodically freezes a copy of the fine-tuned model every k steps, and uses its terminal samples to construct an OT coupling with the target distribution. At each training step, both the conditional target process and the frozen model are considered as candidate proposals, and the candidate with the larger effective sample size is selected for computing the Girsanov weight and the corresponding simulation-free GFT loss. Since the frozen base model is often much closer to the current fine-tuned model than the conditional target process, the purpose of this algorithm is to significantly lower variance in the Girsanov weight, allowing for the use of smaller values of σ_t at the cost of a moderate increase in computation.

We evaluate the efficacy of this method by comparing to training using simulation-free GFT with $w_\theta(X|Z) \approx 1$ and standard CFM in the 2D experimental setting described in Appendix O.2 (Figure 6). The GFT and CFM fine-tuning methods successfully adapt the pretrained model to both target modes, with tightly concentrated generated samples around the target modes. Moving Proposal GFT also recovers both target modes, but produces samples with a larger spread, particularly towards the region between the modes, reflecting the higher variance introduced by computing the Girsanov weight during training.

Finally, we perform another comparison between the performance of weighted and unweighted simulation-free GFT at the image scale, as shown in the ablation results in Appendix P.4 (Figure 14). This ablation demonstrates the similarity in convergence behavior between these two versions of the objective. In fact, the unweighted version of the objective (with $w_\theta(X|Z) \approx 1$ achieves the lowest final FID, along with an apparently lower variance during training than the weighted objective, reflecting the high variance introduced by using the Girsanov density as an importance weight during training.

L GFT Convergence Analaysis

L.1 Assumptions

Assumption 0 (Function space minimization) *At each $\beta > 0$, the vector field v_θ exactly minimizes $\mathcal{L}_\pi(\theta)$ in function space, yielding a pointwise minimizer v_θ^*.*

Algorithm 1 GFT with Moving Proposals

Require: Pretrained drift v_{θ_0}, target samples from q, cooling schedule $\{\beta_s\}$, refresh interval k, pool size m, batch size B, ESS margin γ, diffusion coefficient σ

$\theta \leftarrow \theta_0$
for each training step $s = 0, 1, 2, \ldots$ **do**
  if $s \bmod k = 0$ **then** ▷ Refresh coupling and proposal
    Sample $\{X_0^{(i)}\}_{i=1}^m \sim p_0$ and $\{X_1^{(j)}\}_{j=1}^m \sim q$
    $\hat{X}_1^{(i)} \leftarrow \mathrm{ODESolve}(v_\theta, X_0^{(i)})$ ▷ Terminal samples of current model
    Compute OT plan between $\{\hat{X}_1^{(i)}\}$ and $\{X_1^{(j)}\}$, and pair each $X_0^{(i)}$ with $X_1^{(j)}$ accordingly to form coupling π
    $\tilde{\theta} \leftarrow \theta$ ▷ Freeze current model as proposal
    Optionally decrease σ if the proposal's ESS remained high over the previous window
  end if
  Sample $\{Z_b\}_{b=1}^B \sim \pi$, with $Z_b = (X_0^{(b)}, X_1^{(b)})$
  for each candidate proposal drift $u \in \{v_q(\cdot,\cdot|Z),\ v_{\tilde{\theta}}\}$ **do**
    Simulate $dX_t = u(X_t, t)\,dt + \sigma\,dB_t$ from $X_0^{(b)}$ for each b
    $\log w_b \leftarrow \int_0^1 \sigma^{-1}\big(v_\theta - u\big)(X_t, t) \cdot dB_t - \frac{1}{2}\int_0^1 \big\|\sigma^{-1}\big(v_\theta - u\big)(X_t, t)\big\|^2 dt$
    $\mathrm{ESS}(u) \leftarrow \big(\sum_b w_b\big)^2 / \sum_b w_b^2$
  end for
  $u^* \leftarrow v_{\tilde{\theta}}$ if $\mathrm{ESS}(v_{\tilde{\theta}}) > \gamma \cdot \mathrm{ESS}(v_q)$, else $u^* \leftarrow v_q(\cdot,\cdot|Z)$ ▷ Select proposal
  Simulate paths $X^{(b)}$ under u^* and compute self-normalized weights $\bar{w}_b \leftarrow \mathrm{sg}(w_b)/\sum_{b'} \mathrm{sg}(w_{b'})$
  $\mathcal{L} \leftarrow \sum_b \bar{w}_b \int_0^1 \frac{1}{2}\,\ell_\theta(X_t^{(b)}, t|Z_b)\,dt$ ▷ With $\beta = \beta_s$
  Update θ with $\nabla_\theta \mathcal{L}$
end for
return v_θ

Assumption 1 (Bounded mismatch under $\mathbb{P}_q$) *The pretrained vector field v_{θ_0} and the marginal CFM target v_q satisfy:*

$$C' := \int_0^1 \mathbb{E}_{\mathbb{P}_q} \|v_{\theta_0}(X_t^q, t) - v_q(X_t^q, t)\|^2\,dt < \infty \tag{75}$$

where the expectation is over X_t^q drawn from the process induced by v_q with initial state $X_0 \sim p_0$. The constant C' therefore measures the L^2 mismatch between the pretrained and target vector fields under the target path measure, and is large when the distribution shift from p_1 to q is large.

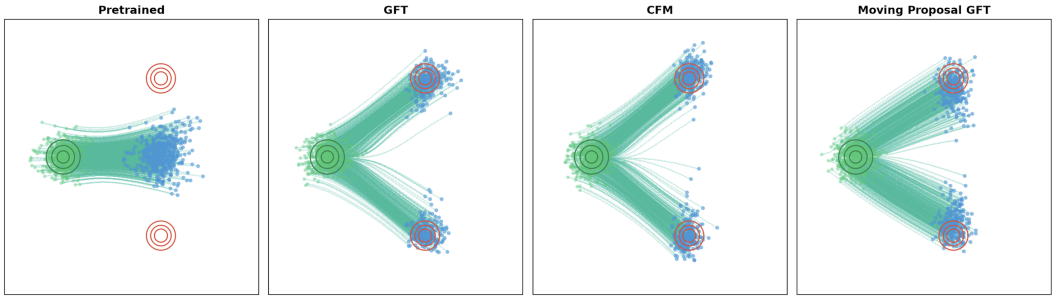


Figure 6: A comparison of fine-tuning with simulation-free and unweighted GFT, CFM, and Moving Proposal GFT. All three algorithms acheive similar performance and mode coverage at the end of fine-tuning, with Moving Proposal GFT ending with a slightly larger spread of generated samples.

Assumption 2 (Bounded conditional variance) *The conditional variance of $v_q(\cdot|Z)$ around its mean $v_q(\cdot)$ is finite:*

$$V := \mathbb{E}_{Z\sim\pi} \int_0^1 \|v_q(X_t, t) - v_q(X_t, t \mid Z)\|^2 \, dt < \infty \tag{76}$$

This quantity is the irreducible variance introduced by the coupling π, and is equal to the standard CFM loss evaluated at $v_q(\cdot)$.

Assumption 3 (Lipschitz vector fields) *Both v_q and v_{θ_0} are Lipschitz continuous in x, uniformly in t, such that there exist constants $L_q, L_0 < \infty$ such that for all $x, y \in \mathbb{R}^d$ and $t \in [0, 1]$:*

$$\|v_q(x,t) - v_q(y,t)\| \le L_q\|x-y\|, \qquad \|v_{\theta_0}(x,t) - v_{\theta_0}(y,t)\| \le L_0\|x-y\| \tag{77}$$

We write $L = \max(L_q, L_0)$.

L.2 Main Result

Theorem L.1 (GFT Convergence) *Under Assumptions 0–3, the pointwise minimizer of $\mathcal{L}_\pi$ at temperature $\beta > 0$ is:*

$$v^*_\theta(X_t, t) = \frac{1}{1+\beta}\, v_q(X_t, t) + \frac{\beta}{1+\beta}\, v_{\theta_0}(X_t, t) \tag{78}$$

and the following hold:

*(i) L^2 **convergence of vector fields.** $v^*_\beta \to v_q$ in $L^2(\mathbb{P}_q)$ as $\beta \to 0^+$, with the explicit rate:*

$$\int_0^1 \mathbb{E}_{\mathbb{P}_q}\|v^*_\beta - v_q\|^2 \, dt \le \frac{\beta^2}{(1+\beta)^2}\, C' \tag{79}$$

*(ii) **Terminal distribution convergence.** The terminal distribution p^*_β of the GFT process converges to q in Wasserstein-2 distance as $\beta \to 0^+$:*

$$W_2^2(p^*_\beta, q) \le \frac{\beta^2\, e^{2L+1}}{(1+\beta)^2}\, C' \tag{80}$$

*(iii) **Loss decomposition.** The GFT loss at the minimizer v^*_β satisfies:*

$$\mathcal{L}_\pi(v^*_\beta) = \frac{\beta}{1+\beta}\, C + V \tag{81}$$

where $C = \mathbb{E}_{Z\sim\pi} \int_0^1 \|v_{\theta_0} - v_q\|^2 \, dt$, and $\mathcal{L}_\pi(v^*_\beta) \to V$ as $\beta \to 0^+$.

proof.
We begin by deriving the pointwise minimizer for a fixed $\beta > 0$. Since $v_\theta(X_t, t)$ is unconditional and does not depend on Z, we minimize $\mathcal{L}_\pi(\theta)$ pointwise in (X_t, t) by differentiating inside the expectation over $Z \sim \pi$ and setting the result to zero:

$$\mathbb{E}_{Z\sim\pi}\Big[2\big(v_\theta - v_q(X_t, t \mid Z)\big) + 2\beta\big(v_\theta - v_{\theta_0}(X_t, t)\big)\Big] = 0 \tag{82}$$

Expanding and collecting all terms in v_θ:

$$(1+\beta)v_\theta = \mathbb{E}_{Z\sim\pi}[v_q(X_t, t \mid Z)] + \beta\, \mathbb{E}_{Z\sim\pi}[v_{\theta_0}(X_t, t)] \tag{83}$$

Since v_{θ_0} is unconditional, $\mathbb{E}_{Z\sim\pi}[v_{\theta_0}(X_t, t)] = v_{\theta_0}(X_t, t)$. Applying the marginalization property of the target vector field, $\mathbb{E}_{Z\sim\pi}[v_q(X_t, t \mid Z) \mid X_t] = v_q(X_t, t)$, and dividing by $(1+\beta)$ gives the minimizer (78). The second derivative of the objective with respect to v_θ is $2(1+\beta) > 0$, confirming this is a minimum.

Result (i): L^2 convergence of $v^*_\beta \to v_q$.

Subtracting v_q from both sides of (78) gives the pointwise identity:

$$v^*_\beta(X_t, t) - v_q(X_t, t) = \frac{\beta}{1+\beta}\big(v_{\theta_0}(X_t, t) - v_q(X_t, t)\big) \tag{84}$$

This holds for every $(X_t, t) \in \mathbb{R}^d \times [0,1]$. Taking the squared norm, integrating over t, and taking the expectation under $\mathbb{P}_q$ gives

$$\int_0^1 \mathbb{E}_{\mathbb{P}-q}\|v_\beta^* - v_q\|^2\, dt = \frac{\beta^2}{(1+\beta)^2}\int_0^1 \mathbb{E}_{\mathbb{P}_q}\|v_{\theta_0} - v_q\|^2\, dt \le \frac{\beta^2}{(1+\beta)^2}\, C' \tag{85}$$

using Assumption 1. As $\beta \to 0^+$, the prefactor $\frac{\beta^2}{(1+\beta)^2} \to 0$, establishing $v_\beta^* \to v_q$ in $L^2(\mathbb{P}_q)$ at a rate of $O(\beta^2)$.

Result (ii): Terminal distribution convergence via Gronwall's Inequality.

We couple the GFT process X_t^* (driven by v_β^*) and the target process X_t^q (driven by v_q) through the *same* initial condition $X_0 \sim p_0$ and the *same* Brownian motion B_t:

$$dX_t^* = v_\beta^*(X_t^*, t)\, dt + \sigma_t\, dB_t, \qquad X_0^* = X_0 \tag{86}$$

$$dX_t^q = v_q(X_t^q, t)\, dt + \sigma_t\, dB_t, \qquad X_0^q = X_0 \tag{87}$$

We now define a difference process $\delta_t = X_t^* - X_t^q$. Since both processes share the same X_0 and B_t, the diffusion terms cancel exactly:

$$d\delta_t = \big(v_\beta^*(X_t^*, t) - v_q(X_t^q, t)\big)dt \tag{88}$$

with $\delta_0 = X_0^* - X_0^q = 0$. Thus δ_t evolves as a deterministic ODE given X_0, with randomness entering only through the shared initial condition.

We compute the time derivative of $\|\delta_t\|^2$:

$$\begin{aligned}\frac{d}{dt}\|\delta_t\|^2 &= 2\delta_t^\top \frac{d\delta_t}{dt} \\ &= 2\delta_t^\top \big(v_\beta^*(X_t^*, t) - v_q(X_t^q, t)\big)\end{aligned} \tag{89}$$

Adding and subtracting $v_\beta^*(X_t^q, t)$:

$$\frac{d}{dt}\|\delta_t\|^2 = 2\delta_t^\top \big(v_\beta^*(X_t^*, t) - v_\beta^*(X_t^q, t)\big) + 2\delta_t^\top \big(v_\beta^*(X_t^q, t) - v_q(X_t^q, t)\big) \tag{90}$$

For the first term, by Cauchy-Schwarz and the Lipschitz continuity of v_β^*, since v_β^* is a convex combination of v_q and v_{θ_0} (78), it inherits Lipschitz continuity with constant at most $\frac{L_q + \beta L_0}{1+\beta} \le \max(L_q, L_0) = L$ from Assumption 3. Using $\|X_t^* - X_t^q\| = \|\delta_t\|$:

$$2\delta_t^\top \big(v_\beta^*(X_t^*, t) - v_\beta^*(X_t^q, t)\big) \le 2\|\delta_t\| \cdot L\|\delta_t\| = 2L\|\delta_t\|^2 \tag{91}$$

For the second term, by Young's inequality $2a^\top b \le \|a\|^2 + \|b\|^2$:

$$2\delta_t^\top \big(v_\beta^*(X_t^q, t) - v_q(X_t^q, t)\big) \le \|\delta_t\|^2 + \|v_\beta^*(X_t^q, t) - v_q(X_t^q, t)\|^2 \tag{92}$$

Combining (91) and (92):

$$\frac{d}{dt}\|\delta_t\|^2 \le (2L+1)\|\delta_t\|^2 + \|v_\beta^*(X_t^q, t) - v_q(X_t^q, t)\|^2 \tag{93}$$

Taking expectation over $X_0 \sim p_0$:

$$\frac{d}{dt}\mathbb{E}\|\delta_t\|^2 \le (2L+1)\mathbb{E}\|\delta_t\|^2 + \mathbb{E}_{\mathbb{P}_q}\|v_\beta^*(X_t^q, t) - v_q(X_t^q, t)\|^2 \tag{94}$$

Applying Gronwall's inequality with initial condition $\mathbb{E}\|\delta_0\|^2 = 0$:

$$\mathbb{E}\|\delta_1\|^2 \le \int_0^1 e^{(2L+1)(1-s)}\mathbb{E}_{\mathbb{P}_q}\|v_\beta^*(X_s^q, s) - v_q(X_s^q, s)\|^2\, ds \le e^{2L+1}\int_0^1 \mathbb{E}_{\mathbb{P}_q}\|v_\beta^*(X_s^q, s) - v_q(X_s^q, s)\|^2\, ds \tag{95}$$

Substituting the pointwise identity (84) and applying (85):

$$\mathbb{E}\|\delta_1\|^2 \le \frac{\beta^2\, e^{2L+1}}{(1+\beta)^2}\, C' \tag{96}$$

To obtain the W_2 bound, we note that our coupled construction (X_1^*, X_1^q) is one particular coupling of p_β^* and q with marginals p_β^* and q respectively. Since $W_2^2(p_\beta^*, q)$ is defined as the infimum of $\mathbb{E}\|X - Y\|^2$ over *all* couplings of p_β^* and q, it is bounded above by the expectation under any specific coupling, including ours:

$$W_2^2(p_\beta^*, q) \le \mathbb{E}\|X_1^* - X_1^q\|^2 = \mathbb{E}\|\delta_1\|^2 \le \frac{\beta^2\, e^{2L+1}}{(1+\beta)^2}\, C' \tag{97}$$

This gives $W_2(p_\beta^*, q) \to 0$ at rate $O(\beta)$ as $\beta \to 0^+$.

Results (iii): Loss decomposition.

We substitute v_β^* into $\mathcal{L}_\pi$ and handle each term. For the first term, we decompose:

$$v_\beta^* - v_q(X_t, t \mid Z) = \underbrace{(v_\beta^* - v_q)}_{=\frac{\beta}{1+\beta}(v_{\theta_0} - v_q)} + (v_q - v_q(X_t, t \mid Z)) \tag{98}$$

using (84). Expanding the squared norm and taking expectation over $Z \sim \pi$:

$$\begin{aligned}\mathbb{E}_Z\|v_\beta^* - v_q(\cdot \mid Z)\|^2 = \frac{\beta^2}{(1+\beta)^2}\|v_{\theta_0} - v_q\|^2 \\ + \frac{2\beta}{1+\beta}(v_{\theta_0} - v_q)^\top \underbrace{\mathbb{E}_Z[v_q - v_q(\cdot \mid Z)]}_{=0} \\ + \mathbb{E}_Z\|v_q - v_q(\cdot \mid Z)\|^2\end{aligned} \tag{99}$$

The cross term vanishes since $\mathbb{E}_{Z\sim\pi}[v_q(X_t, t) - v_q(X_t, t \mid Z)] = v_q - v_q = 0$ by the marginalization property of v_q. For the second term of the loss:

$$\|v_\beta^* - v_{\theta_0}\|^2 = \frac{1}{(1+\beta)^2}\|v_q - v_{\theta_0}\|^2 \tag{100}$$

using (84) applied to $v_\beta^* - v_{\theta_0} = \frac{1}{1+\beta}(v_q - v_{\theta_0})$. Combining (99) and (100), integrating over t, and collecting:

$$\begin{aligned}\mathcal{L}_\pi(v_\beta^*) &= \left(\frac{\beta^2}{(1+\beta)^2} + \frac{\beta}{(1+\beta)^2}\right)\mathbb{E}_{Z\sim\pi}\int_0^1 \|v_{\theta_0} - v_q\|^2\, dt + V \\ &= \frac{\beta(\beta+1)}{(1+\beta)^2}\, C + V = \frac{\beta}{1+\beta}\, C + V\end{aligned} \tag{101}$$

As $\beta \to 0^+$, $\mathcal{L}_\pi(v_\beta^*) \to V$, which is the standard CFM loss. The β-dependent component $\frac{\beta}{1+\beta}C$ vanishes at rate $O(\beta)$.

M GFT FOR MEAN FLOW MODELS

M.1 RELATIONSHIP BETWEEN AVERAGE AND INSTANTANEOUS OPTIMUMS

We assume u_{θ_0}, u_q, and u_θ be average velocity fields related to their corresponding instantaneous fields v_{θ_0}, v_q, and v_θ via the Mean Flow Identity:

$$u(X_t, r, t) = v(X_t, t) - (t - r)\frac{d}{dt}u(X_t, r, t)\,. \tag{102}$$

The Mean Flow Identity can then be rewritten to express the average velocity field directly as an integral over the instantaneous velocity field.

$$u(X_t, r, t) + (t - r)\frac{d}{dt}u(X_t, r, t) = v(X_t, t) \tag{103}$$

$$\frac{d}{dt}[(t - r)u(X_t, r, t)] = v(X_t, t) \tag{104}$$

$$\int_r^t \frac{d}{d\tau}[(\tau - r)u(X_\tau, r, \tau)]d\tau = \int_r^t v(X_\tau, \tau)d\tau \tag{105}$$

$$(t - r)u(X_t, r, t) = \int_r^t v(X_\tau, \tau)d\tau \tag{106}$$

$$u(X_t, r, t) = \frac{1}{(t - r)} \int_r^t v(X_\tau, \tau)d\tau \tag{107}$$

Now suppose that the instantaneous vector field v_θ takes the following form:

$$v_\theta(X_t, t) = \left(\frac{1}{1+\beta}\right) v_q(X_t, t) + \left(\frac{\beta}{1+\beta}\right) u_{\theta_0}(X_t, t) \, . \tag{108}$$

Let X_τ be the trajectory induced by v_θ, such that $\frac{dX_\tau}{d\tau} = v_\theta(X_\tau, \tau)$. Despite this trajectory being induced by the fine-tuned model, all models will now be evaluated along the points of this given trajectory. We substitute (108) into (107).

$$u_\theta(X_t, r, t) = \frac{1}{(t - r)} \int_r^t v_\theta(X_\tau, \tau)d\tau \tag{109}$$

$$= \frac{1}{t - r} \int_r^t \left(\frac{1}{1+\beta}\right) v_q(X_\tau, \tau) + \left(\frac{\beta}{1+\beta}\right) v_{\theta_0}(X_\tau, \tau) \, d\tau \tag{110}$$

Since the trajectory X_τ is fixed throughout, the linearity of the integral operator gives:

$$u_\theta(X_t, r, t) = \frac{1}{t - r} \int_r^t \left(\frac{1}{1+\beta}\right) v_q(X_\tau, \tau) \, d\tau + \frac{1}{t - r} \int_r^t \left(\frac{\beta}{1+\beta}\right) v_{\theta_0}(X_\tau, \tau) \, d\tau \, . \tag{111}$$

Recognizing each integral as the average velocity field of v_q and v_{θ_0} respectively, both evaluated along the same trajectory X_τ, we obtain the exact result:

$$u_\theta(X_t, r, t) = \left(\frac{1}{1+\beta}\right) u_q(X_t, r, t) + \left(\frac{\beta}{1+\beta}\right) u_{\theta_0}(X_t, r, t) \, , \tag{112}$$

where $u_q(X_t, r, t)$ and $u_{\theta_0}(X_t, r, t)$ are evaluated at the point X_t induced by the fine-tuned process $u_\theta(X_t, r, t)$.

Note that this relationship between the optimal instantaneous and average velocity fields also works in reverse. Assume now that we start from the average velocity field, written as a convex combination as given in (112). We apply the Mean Flow Identity to all average velocity terms.

$$v_\theta^* - (t - r)\frac{d}{dt}u_\theta^* = \frac{1}{1+\beta}\left(v_q - (t - r)\frac{d}{dt}u_q\right) + \frac{\beta}{1+\beta}\left(v_{\theta_0} - (t - r)\frac{d}{dt}u_{\theta_0}\right) \tag{113}$$

Applying the linearity of $\frac{d}{dt}$ to (112) gives

$$\frac{d}{dt}u_\theta^* = \frac{1}{1+\beta}\frac{d}{dt}u_q + \frac{\beta}{1+\beta}\frac{d}{dt}u_{\theta_0} \, . \tag{114}$$

Substituting this into the left-hand side of (113):

$$v_\theta^* - (t - r)\left(\frac{1}{1+\beta}\frac{d}{dt}u_q + \frac{\beta}{1+\beta}\frac{d}{dt}u_{\theta_0}\right) = \frac{v_q + \beta v_{\theta_0}}{1+\beta} - (t - r)\left(\frac{1}{1+\beta}\frac{d}{dt}u_q + \frac{\beta}{1+\beta}\frac{d}{dt}u_{\theta_0}\right). \tag{115}$$

The second terms on both sides cancel, leaving the original instantaneous convex combination.

$$v_\theta^* = \frac{v_q + \beta v_{\theta_0}}{1+\beta} \,, \tag{116}$$

M.2 Mean Flow GFT Algorithm

Algorithm 2 MeanFlow GFT fine-tuning step

Require: Current model u_θ, frozen base model u_{θ_0}, data sample x_1, noise x_0, reference time r, time t, temperature β

1. Compute interpolated sample and conditional target velocity
$x_t \leftarrow (1-t)x_0 + tx_1$
$v_q \leftarrow x_1 - x_0$

2. Recover instantaneous velocities from Mean Flow models
$v_{\theta_0} \leftarrow u_{\theta_0}(x_t,\, t,\, t)$ (forward pass of frozen model at $r = t$)
$v_\theta \leftarrow u_\theta(x_t,\, t,\, t)$ (forward pass of current model at $r = t$)

3. Compute the conditional GFT instantaneous target
$v^\star \leftarrow \dfrac{v_q + \beta\, v_{\theta_0}}{1+\beta}$

4. Compute JVP of current model using $v^\star$ as tangent
$(u_\theta^{\text{out}},\ \frac{d}{dt}u_\theta) \leftarrow \texttt{jvp}(u_\theta(x_t, r, t),\ (v^\star,\ 0,\ 1))$

5. Compute average velocity target via MeanFlow Identity
$u^\star \leftarrow \mathrm{sg}\left(v^\star - (t-r)\dfrac{d}{dt}u_\theta\right)$

6. Compute loss
$\mathcal{L} \leftarrow \|u_\theta^{\text{out}} - u^\star\|^2$
return $\mathcal{L}$

N Reward-based GFT

N.1 Derivation of the Reward-based GFT Objective

We begin from the sample-based GFT objective derived in Section 4.3, restated here for simplicity:

$$\mathcal{L}_\pi(\theta) = \mathbb{E}_{\substack{Z\sim\pi,\ t\sim\mathcal{U}[0,1],\\ X_t\sim\mathbb{P}_q(\cdot|Z)}}\left[\frac{w_\theta}{2}\left(\|\sigma_t^{-1}(v_\theta(X_t,t) - v_q(X_t,t|Z))\|^2 + \beta\|\sigma_t^{-1}(v_\theta(X_t,t) - v_{\theta_0}(X_t,t))\|^2\right)\right]. \tag{117}$$

First consider the standard reward maximization problem with weighted KL divergence regularization:

$$\max_q\ \mathbb{E}_{X_1\sim q}[R(X_1)] - \alpha\,\mathrm{KL}(q\|p_{\theta_0})\,. \tag{118}$$

This problem has a known and unique solution of the following form:

$$q(X_1) = \frac{1}{\mathcal{Z}}\,p_{\theta_0}(X_1)\,\exp\left(\frac{R(X_1)}{\alpha}\right)\,, \tag{119}$$

$$\mathcal{Z} = \mathbb{E}_{X_1\sim p_{\theta_0}}\left[\exp(R(X_1)/\alpha)\right]\,. \tag{120}$$

The current sample-based GFT objective samples $Z = (X_0, X_1) \sim \pi$ where $\pi \in \Pi(p_0, q)$ is any arbitrary coupling between the source and target distributions. With this definition, the objective is inherently sample-based, as it requires the ability to draw samples from q. In order to achieve a reward-based objective, we must remove this dependency on q. To this end, we consider the following couplings:

$$\pi(X_0, X_1) = p_0(X_0)\, q(X_1)\,, \tag{121}$$

$$\pi_0(X_0, X_1) = p_0(X_0)\, p_{\theta_0}(X_1)\,, \tag{122}$$

where p_0 is the shared source distribution for the pretrained, fine-tuned, and target processes, and p_{θ_0} is the generative distribution of the pretrained model. We assume mutual absolute continuity between these couplings, and solve for their Radon-Nikodym derivative. The function $\frac{d\pi}{d\pi_0}(X_0, X_1)$ is the Radon-Nikodym derivative if for every $\mathcal{A} \subseteq \mathbb{R}^d \times \mathbb{R}^d$:

$$\pi(\mathcal{A}) = \int_{\mathcal{A}} \frac{d\pi}{d\pi_0}(X_0, X_1) d\pi_0(X_0, X_1)\,. \tag{123}$$

Starting from the LHS:

$$\pi(\mathcal{A}) = \int_{\mathcal{A}} p_0(X_0) q(X_1) dX_0 dX_1\,. \tag{124}$$

Substituting the definition of q:

$$= \int_{\mathcal{A}} p_0(X_0) \frac{1}{\mathcal{Z}}\, p_{\theta_0}(X_1)\, \exp\!\left(\frac{R(X_1)}{\alpha}\right) dX_0 dX_1 \tag{125}$$

$$= \int_{\mathcal{A}} \frac{1}{\mathcal{Z}}\, \exp\!\left(\frac{R(X_1)}{\alpha}\right) p_0(X_0) p_{\theta_0}(X_1) dX_0 dX_1\,. \tag{126}$$

We have now recovered the exact definition of $d\pi_0$:

$$= \int_{\mathcal{A}} \frac{1}{\mathcal{Z}}\, \exp\!\left(\frac{R(X_1)}{\alpha}\right) d\pi_0(X_0, X_1)\,. \tag{127}$$

This gives us the final definition of the Radon-Nikodym derivative:

$$\frac{d\pi}{d\pi_0}(X_0, X_1) = \frac{1}{\mathcal{Z}}\, \exp\!\left(\frac{R(X_1)}{\alpha}\right) = \frac{q(X_1)}{p_{\theta_0}(X_1)}\,. \tag{128}$$

Importantly, the Radon-Nikodym derivative depends only on the terminal sample X_1. Given this, for any arbitrary function $f(X_0, X_1)$, we can apply the Radon-Nikodym derivative as an importance weight to perform change of measure between π and π_0:

$$\mathbb{E}_{(X_0,X_1)\sim\pi}\big[f(X_0, X_1)\big] = \mathbb{E}_{(X_0,X_1)\sim\pi_0}\left[\frac{d\pi}{d\pi_0}(X_0, X_1) \cdot f(X_0, X_1)\right] \tag{129}$$

$$= \frac{1}{\mathcal{Z}}\, \mathbb{E}_{(X_0,X_1)\sim\pi_0}\left[\exp\!\left(\frac{R(X_1)}{\alpha}\right) f(X_0, X_1)\right]\,. \tag{130}$$

Applying this to the original sample-based GFT objective gives

$$\mathcal{L}_R(\theta) = \frac{1}{\mathcal{Z}}\, \mathbb{E}_{\substack{X_0\sim p_0,\, X_1\sim p_{\theta_0} \\ X_t\sim\mathbb{P}_q(\cdot|X_0,X_1) \\ t\sim\mathcal{U}[0,1]}}\Bigg[\exp\!\left(\frac{R(X_1)}{\alpha}\right) \frac{w_\theta}{2}(\|\sigma_t^{-1}(v_\theta(X_t, t) - v_q(X_t, t|X_0, X_1))\|^2 + \beta\, \|\sigma_t^{-1}(v_\theta(X_t, t) - v_{\theta_0}(X_t, t))\|^2)\Bigg] \tag{131}$$

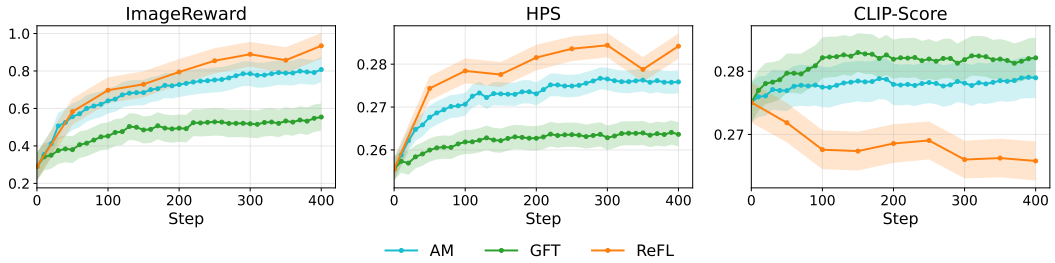


Figure 7: Extended reward fine-tuning results for ReFL and sample-based GFT.

The terminal point X_1 is now sampled through rollouts from the frozen pretrained model, and the conditional v_q remains well defined using X_0 and X_1. X_t is sampled based on the target process $\mathbb{P}_q$, which is possible in closed-form given the endpoints X_0 and X_1. The objective no longer depends on samples from the target distribution q, and uses the reward function as an importance weight. We note that this derivation depends on the use of independent couplings for π and π_0, and that the GFT property of allowing for OT source-target couplings does not carry to the reward fine-tuning setting.

N.2 Reward Fine-Tuning Implementation Details

For the pretrained base model, we use the Stable Diffusion 1.5 model (stable-diffusion-v1-5). All methods are trained on 10,000 prompts, or prompt-image pairs generated according to Algorithm 4. All experiments use an Adam optimizer, and the GFT experiments use an inverse sigmoid cooling schedule. All reward-based experiments are run in parallel over two A100 GPUs.

The reward-based experiments use Classifier-Free Guidance (CFG) Ho & Salimans (2021) with a guidance scale of 7.5 for both GFT and AM. Both methods keep the unconditional branch trainable by dropping prompts with a 20% likelihood. The Adjoint Matching experiments use the code from the paper's official Github repository Domingo-Enrich et al. (2025). We note that the training time required for AM depends on several hyperparameter settings, including dataset size, batch size, and number of epochs. The training time reported in this paper is a result of using the default hyperparameter setting given in this Github repository.

Optimal Transport coupling are not used in the GFT reward experiments, in accordance with the derivation of the reward-based GFT objective given in Appendix N.1.

For the GFT experiments with an augmented ReFL term, as described in Algorithm 5, we find that 130 warmup steps of pure sample-based GFT yields the highest performance across the reported metrics. Using 50 discretization steps for full sample generation, we set the snapshot indices to $\mathcal{K} = \{38, 42, 45, 48\}$, and average the reward loss over these indices. The rollout for the ReFL term is on policy, while the GFT term continues to use the precomputed dataset. We set the ReFL weight term $\lambda_r = 25$ for the reported training runs. In these experiments, we perform EMA parameter updates to the fine-tuned model, as this counteracts the increased training instability that is observed after the addition of the ReFL term to the GFT objective. We use an EMA decay parameter of 0.95.

N.3 Reward-GFT Algorithms

All algorithms used to implement reward-based GFT are shown in 3, 4, and 5.

N.4 Extended Results and Baselines

We now present baselines for reward-based fine-tuning in addition to Adjoint Matching. ReFL Xu et al. (2023) is used as an augmentation to the GFT objective in the GFT (+ReFL) algorithm, but is shown here individually for comparison. ReFL is able to outperform adjoint matching on both ImageReward and HPS, but reduces CLIP in comparison. We also report sample-based GFT without any ReFL augmentation. This method is zeroth order, meaning that it does not use the reward gradients, and therefore has less information at training time than Adjoint Matching or ReFL, both

Algorithm 3 GFT + ReFL Fine-Tuning for Diffusion Models

Require: Pretrained checkpoint θ_{pre}, reward model R, dataset $\mathcal{D}$, snapshot indices $\mathcal{K}$, fixed weights λ_r, β_r, cooling schedule β_s, EMA decay μ

$\theta \leftarrow \theta_{\text{pre}}, \quad \theta_0 \leftarrow \theta_{\text{pre}}, \quad \theta_{\text{EMA}} \leftarrow \theta_{\text{pre}}$ ▷ freeze θ_0
for each optimizer step s **do**
  — Part A: ReFL (on-policy, from noise) —
  Sample prompts P
  $x_T \sim \mathcal{N}(0, I)$
  Roll out DDIM to step $\max(\mathcal{K})$, saving snapshots $\{x_k\}_{k \in \mathcal{K}}$ ▷ No grad
  for $k \in \mathcal{K}$ **do**
    $\epsilon_\theta \leftarrow f_\theta(x_k, t_k)$ ▷ With grad
    $\hat{x}_0 \leftarrow \text{denoise}(x_k, \epsilon_\theta, t_k)$
    $\mathcal{L}_r \leftarrow -\mathbb{E}[R(\hat{x}_0, P)]$
    $\epsilon_0 \leftarrow f_{\theta_0}(x_k, t_k)$ ▷ No grad
    $\mathcal{L}_{\text{anchor}}^{\text{ReFL}} \leftarrow \beta_r \|\epsilon_\theta - \epsilon_0\|^2$
    Accumulate $\nabla_\theta \frac{1}{|\mathcal{K}|}(\lambda_r \mathcal{L}_r + \mathcal{L}_{\text{anchor}}^{\text{ReFL}})$
  end for
  — Part B: GFT (off-policy, on dataset) —
  $(x_1, c_d) \sim \mathcal{D}$
  $t \sim \mathcal{U}\{0, \ldots, T\}, \quad n \sim \mathcal{N}(0, I)$
  $x_t \leftarrow \sqrt{\bar{\alpha}_t}\, x_1 + \sqrt{1 - \bar{\alpha}_t}\, n$
  $\mathcal{L}_{\text{gft}} \leftarrow \|f_\theta(x_t, t, c_d) - n\|^2 + \beta_s \|f_\theta(x_t, t, c_d) - f_{\theta_0}(x_t, t, c_d)\|^2$
  Accumulate $\nabla_\theta(\mathcal{L}_{\text{gft}})$
  Update θ
  Update EMA parameters θ_{EMA}
end for

Algorithm 4 Reward-Filtered Dataset Construction. The gradient ascent refinement (lines 6–10) is optional and is skipped when $M = 0$.

Require: Pretrained model θ_0, reward model R, prompt set $\mathcal{P}$, samples per prompt K, samples retained per prompt n, ascent steps $M \geq 0$, step size η

1: $\mathcal{D} \leftarrow \emptyset$
2: **for** each prompt $p \in \mathcal{P}$ **do**
3: Generate $\{x^{(i)}\}_{i=1}^{K}$ using θ_0 with prompt p
4: $\mathcal{S}_p \leftarrow$ top-n samples by reward $R(x^{(i)}, p)$
5: **for** each $x \in \mathcal{S}_p$ **do** ▷ Optional refinement
6: **for** $m = 1, \ldots, M$ **do**
7: $x \leftarrow x + \eta \nabla_x R(x, p)$
8: **end for**
9: **end for**
10: $\mathcal{D} \leftarrow \mathcal{D} \cup \{(x, p) : x \in \mathcal{S}_p\}$
11: **end for**
12: **return** $\mathcal{D}$

Algorithm 5 Reward-Based GFT with ReFL

Require: Pretrained model v_{θ_0}, reward R, prompt set $\mathcal{C}$, cooling schedule $\{\beta_e\}$, warmup epochs E_w, total epochs E

$\mathcal{D} \leftarrow$ GENERATEDATASET$(v_{\theta_0}, R, \mathcal{C})$ ▷ Algorithm 4
Initialize $\theta \leftarrow \theta_0$
for $e = 1, \ldots, E$ **do**
  if $e \leq E_w$ **then**
    for each minibatch in $\mathcal{D}$ **do** ▷ Warmup: GFT only
      Update θ by a gradient step on $\mathcal{L}_\pi(\theta)$ (12) with $\beta = \beta_e$
    end for
  else
    Update θ for one epoch by Algorithm 3 with $\beta = \beta_e$ ▷ GFT + ReFL
  end if
end for
return v_θ

of which are first order. As a result, this version of GFT is not able to match Adjoint Matching or ReFL on ImageReward or HPS, but does outperform both methods in terms of CLIP.

O Sample Fine-Tuning Implementation Details

We use the exact U-Net architecture given by Tong et al. (2024) and Tong et al. (2023), which has been pretrained on the Cifar-10 dataset Krizhevsky (2009) for 400,000 epochs. Minibatch OT couplings are used during pretraining and fine-tuning of the model, in accordance with the convergence results given in Section 4.3. All GFT experiments use an inverse sigmoid β_s cooling schedule and an Adam Optimizer. To ensure rigorous evaluation, we hold out a random 20% of each training dataset for exclusive use in calculating Fréchet Inception Distance (FID).

All experiments were run on a single A100 GPU. The large scale image generation experiments each ran for approximately 12 hours for Cross-Domain and 8 hours for In-Domain, for a total of about 60 GPU hours. The 2D experiments ran for approximately 1 hour in total. This does not include the time or computational resources used for preliminary experiments that are not included in this paper.

O.1 Datasets

We fine-tune the pretrained image generation model for three datasets within the WILDS benchmarks datasets Koh et al. (2021) (MIT Liscense).

Camelyon17 Bándi et al. (2019) is a dataset of H&E-stained histopathology images of lymph node sections of breast cancer patients. The dataset includes both tumor and non-tumor images. The distribution shift between each split of this dataset is caused experimental batch effects and population variations. The Camelyon17 train set has a total of 302,436 samples, and 60,490 of these samples are held out from training for use in the FID calculations. The Camelyon17 validation set is substantially smaller than the train dataset, with a total of 34,904 samples, 6,981 of which are held out for use in the FID calculations.

RxRx1 Sypetkowski et al. (2023) is a dataset of fluorescent microscopy images of human tissue sections. The distribution shift between each split of this dataset is caused experimental batch effects, such as variations in temperature, humidity, and reagent concentration. The RxRx1 experiments are conducted on a compressed version of the dataset to enable running on a single GPU. The RxRx1 train set has a total of 40,612 samples, 8,123 of which are held out for FID calculations. The RxRx1 validation set is smaller than the train dataset, with a total of 9,854 samples, and 1,971 of these samples are held out from fine-tuning for use in the FID calculations.

FMoW Christie et al. (2018) is a dataset of global satellite images of regions with varying land use. The experiments shown below are conducted on the RGB image version of the dataset, rather than the multispectral data. The distribution shift between each split of this dataset is caused temporal

evolution. The FMoW experiments are conducted on a compressed version of the dataset to enable running on a single GPU. The FMoW train set has a total of 76,863 samples, and 15,373 are held out for use in the FID calculations. The FMoW validation set has a total of 19,915 samples, 3,983 of which are held out from fine-tuning for FID calculations.

Stability and Convergence Analysis We report three metrics to quantitatively analyze the convergence speed and stability of GFT. *Instantaneous Variance* measures the local variance in accuracy convergence. At 50 epoch intervals, the instantaneous mean is found using a Radial Basis Function (RBF) kernel over a window of 10 FID measurements. This mean is then used to calculate variance with the same weighting and window size. The reported value is the average of all instantaneous variances over the scope of fine-tuning. *Average convergence rate* is calculated by performing linear regression on windows of 10 consecutive FID measurements. The reported value is the average of the absolute value of the resulting slopes. Finally, *Spearman correlation* is calculated between fine-tuning epochs and FID. For a stable fine-tuning process, we expect a strong negative correlation. A Spearman correlation coefficient close to -1 would indicate that as fine-tuning progresses, the model performance consistently improves with minimal regressive spikes or oscillation.

Path Length Calculation As a proxy to estimate the generation efficiency of each method we test, we report average probability path length. We can define the path length of a single sample through a vector field as its energy cost, or its accumulated kinetic energy over time. By the Benamou-Brenier formulation of optimal transport, this can be calculated by integrating the square of the velocity magnitude over the path of the given sample.

$$L(x_0) = \int_0^1 \|\frac{dx}{dt}\|^2 dt = \int_0^1 \|\mathbf{v}(x_t, t)\|^2 dt \tag{132}$$

Given a prior distribution over initial samples $p(x_0)$, we can get the expected length of the vector field.

$$\mathbb{E}_{p(x_0)}[L(x_0)] = \mathbb{E}_{x_0 \sim p(x_0)} \left[\int_0^1 \|\mathbf{v}(x_t, t)\|^2 dt \right] \tag{133}$$

The expectation over the prior distribution p_0 can be approximated by averaging over a minibatch of M samples. We can apply Euler discretization with some chosen step size Δt to approximate the integral.

$$L(\theta) = \frac{1}{M} \sum_{j=1}^{M} \sum_{i=1}^{N} \|\mathbf{v}_\theta(x_i^{(j)}, t_i)\|^2 \cdot \Delta t \tag{134}$$

We estimate average path length for a model with parameters θ with 100 discretization steps and 64 initial samples.

O.2 Diversity and Mode Collapse Experiments

We additionally compare GFT to reward-based fine-tuning methods on simulated 2D data, which enables a fair comparison of unconditional generation, without the need for the supplemental stability or guidance mechanisms often required for reward methods in higher dimensional settings. In these experiments, a flow matching model is pretrained to generate samples from a unimodal Gaussian distribution, and then fine-tuned to a bimodal Gaussian target using GFT, as well as standard CFM fine-tuning Lipman et al. (2023), Adjoint Matching Domingo-Enrich et al. (2025), and ORW-CFM-W2 Fan et al. (2025b). In these experiments, we report mode coverage as the ratio of generated samples that fall within a fixed radius of each target mean, and covariance error as the L^2 norm of the difference between the empirical and true covariances for each mode.

All small scale 2D experiments are run on a single A100 GPU using simulated datasets. All fine-tuning methods are run until convergence, with GFT and CFM taking 500 epochs, Adjoint Matching

Table 2: Fine-tuning performance comparison for 2D experiments. GFT and CFM achieve better mode coverage and covariance accuracy than both reward-based methods. They also have significantly higher training efficiency, as reported in units of average seconds per epoch during pretraining.

	Upper Mode		Lower Mode		
Method	**Cov. Error ↓**	**Coverage ↑**	**Cov. Error ↓**	**Coverage ↑**	**Epoch Time**
CFM	0.112	**0.446**	**0.059**	0.394	0.994
GFT	**0.039**	0.362	0.071	**0.472**	1.556
Adjoint Matching	0.198	0.390	0.196	0.228	26.111
ORW-CFM-W2	0.211	0.042	inf	0.000	3.722

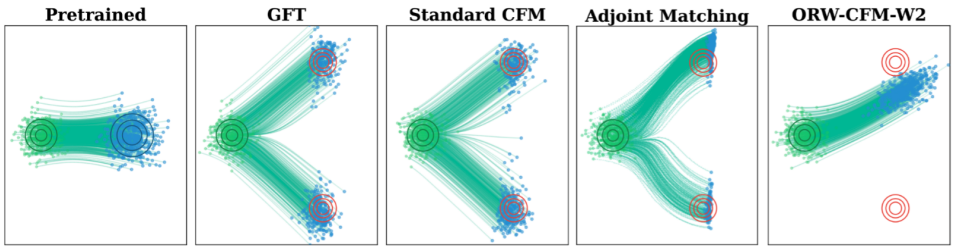


Figure 8: Simulated trajectories for 2D experiments. GFT and CFM recover both target modes with high accuracy and diversity. The reward-based methods fall into varying degrees of mode collapse.

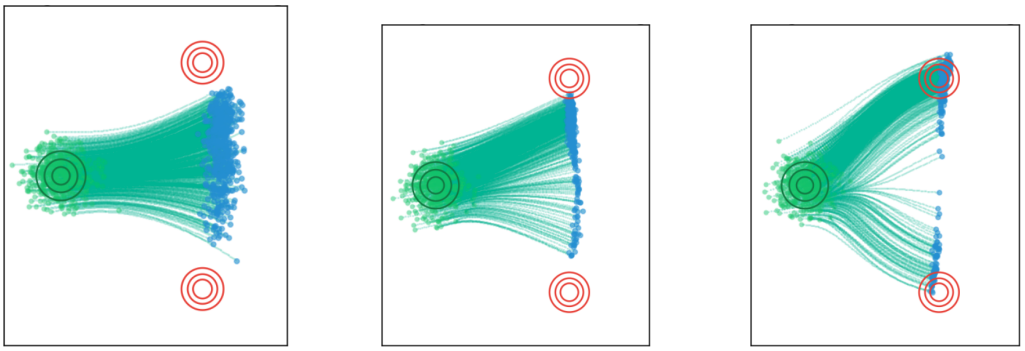
taking 1500 epochs, and ORW-CFM-W2 taking 1000 epochs. The ORW-CFM-W2 experiment uses the hyperparameters $\tau = 4.5$ and $\alpha = 0.001$.

The resulting trajectory plots (Figure 8) demonstrate the stronger performance of the sample-based fine-tuning methods over the reward methods. GFT and CFM are able to cover both fine-tuning target modes with high accuracy and diversity. Adjoint Matching suffers from significantly reduced diversity compared to the pretrained model, with all generated samples tightly clustered within the distribution. The reduced coverage and mode collapse demonstrated by Adjoint Matching in these experiments is further quantified in Table 2, where Adjoint Matching is shown to have significantly higher covariance error than both GFT and CFM. While ORW-CFM-W2 maintains a high generative diversity, it is only able to acquire one of the target distribution modes with low accuracy.

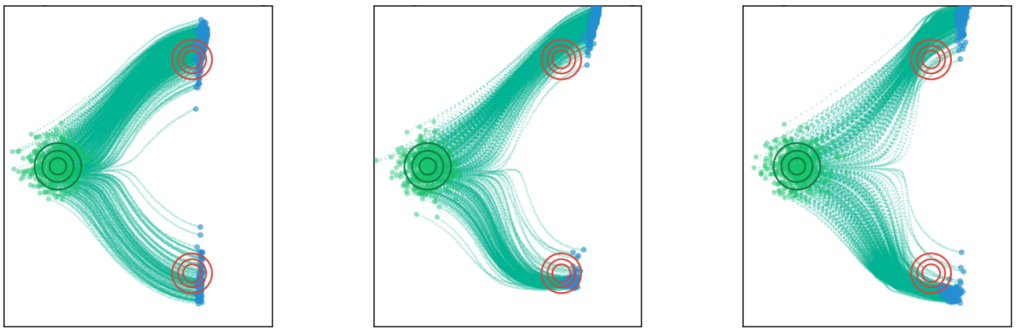
We further comment on the significantly increased training time required for reward-based compared to sample-based methods. Table 2 reports average epoch time for each method as a unit of the average seconds per epoch during pretraining with the CFM objective. Standard CFM fine-tuning retains the same training speed, with GFT taking slightly longer due to the need for two model evaluations for each epoch. However, the epoch time of Adjoint Matching is over an order of magnitude higher than that of GFT, as it involves two expensive ODE simulations for each fine-tuning step.

O.3 Extended Adjoint Matching Results

Fine-tuning with Adjoint Matching does not recover both modes of the target distribution until after 1000 epochs, compared to around 500 epochs for both GFT and CFM (Figure 9). Additionally, the target modes are recovered with a significantly decreased diversity than exhibited by the pretrained model, and the generation diversity decreases as fine-tuning proceeds. Finally, when trained further after recovering the target modes, Adjoint Matching shifts the solution away from the mode means, decreasing accuracy as fine-tuning continues (Figure 9f). This is in contrast to GFT, which maintains the diversity of the pretrained model while closely matching the covariance of both modes in the new target distribution.



(a) 200 Epochs. (b) 500 Epochs. (c) 1000 Epochs.



(d) 1500 Epochs. (e) 2000 Epochs. (f) 3000 Epochs.

Figure 9: Generated trajectories after fine-tuning with Adjoint Matching. Extended fine-tuning steadily reduces diversity and shifts the generated samples away from the mode centers.

P Ablation Experiments

P.1 Effect of Regularization Strength

To demonstrate the effect of regularization strength, we conduct a set of GFT experiments which vary only in the minimum temperature β reached at the end of fine-tuning. This minimum value of the cooling schedule dictates the strength of regularization at the final stages of fine-tuning, and therefore determines the distribution from which the model generates images. The results of these experiments reveal an important inverse relationship between final FID and average path length, which can be reliably controlled by altering the cooling schedule (Figure 10). Setting the minimum temperature to zero for GFT results in a similar final path length and FID as the CFM objective. However, every uniform increase in the minimum β value results in a highly consistent and predictable decrease in final average path length and corresponding increase in final FID. According to these results, the tradeoff between FID and path length is likely directly caused by the strength of regularization, and can therefore be explicitly controlled by varying the cooling schedule.

P.2 Cooling Schedule

All experiments reported in this work use a sigmoid cooling schedule, as experimental ablations found this schedule to result in the lowest final FID and average path length. However, we note that GFT is robust to several choices of cooling schedule. Figure 11 shows the results from cross-domain adaptation to the Camelon17 train dataset with three types of cooling schedules: Linear, Exponential, and Sigmoid. All three cooling schedules show stable convergence to very similar final FID values, although at slightly varying rates. The linear cooling schedule maintains a lower Path Length increase than sigmoid or exponential throughout the majority of training, but faces a slight increase at the end of 2000 epochs which means that the sigmoid schedule achieves the lowest final value. Following this, all experiments reported in this work use a sigmoid cooling schedule.

P.3 LoRA Rank Ablation

To study the effect of the chosen LoRA rank on fine-tuning performance, we report an ablation across four values of LoRA rank for both CFM and GFT across Cross- and In-Domain adaptation.

In the Cross-Domain setting, a higher LoRA rank tends to lead to a lower final FID for both GFT and CFM. CFM has a faster initial convergence than GFT for all chosen LoRA ranks, but CFM with ranks 4 and 8 results in a higher final FID than any of the GFT training runs. In Path Length, a higher LoRA rank also tends to result in a higher final Path Length. CFM with any LoRA rank has a faster initial increase in Path Length than all GFT experiments.

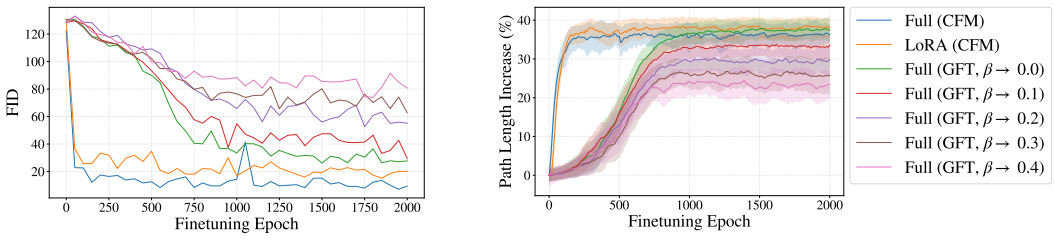
The In-Domain adaptation experiments show a similar pattern. The lowest final FID values are achieves by GFT and CFM fine-tuning with ranks 16 and 32. Lower ranks lead to slightly higher final FID values for both fine-tuning methods. However, in the In-Domain setting, GFT with rank 32 achieves the lowest final Path Length, with the lower-rank experiments retaining higher Path Lengths throughout training.

P.4 Weighted vs. Unweighted Simulation-Free GFT

In this experimental setting, we test the efficacy of the simulation-free GFT objective with the Girsanov density set to $w_\theta \approx 1$. Figure 14 compares unweighted GFT with the weighted objective given in equation 12. This ablation reveals that both the weighted and unweighted versions of the GFT objective have similar convergence behavior, but that unweighted GFT actually results in the lowest final FID. Weighted GFT with $\sigma = 3.0$ achieves the lowest final Path Length, but has a significantly higher variance in both metrics than unweighted GFT. This instability in both FID and Path Length for weighted GFT across all tested σ values can likely be attributed to the high variance of the Girsanov density w_θ across batches and in the high dimensional space of image data. Following the implication of this result, all reported GFT experiments in this work use the unweighted objective where the Girsanov density is approximated to be 1.

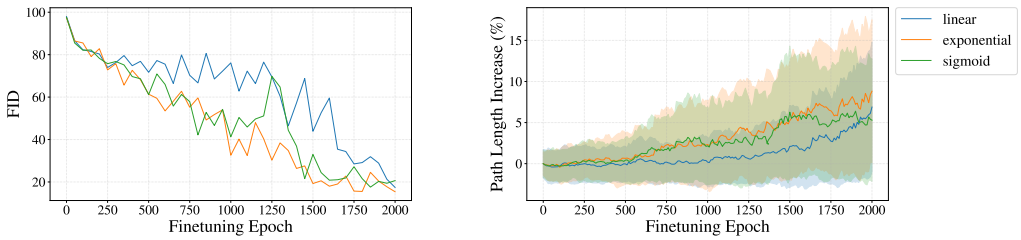
Q Cross-Domain Adaptation vs Random Initialization

We perform an additional evaluation of fine-tuning performance for cross-domain adaptation by comparing the fine-tuning of a base Cifar-10 model on the Camelyon17 train dataset to training on the same dataset from random initialization (Figure 15). FID analysis highlights the clear advantage of fine-tuning over training from random initialization (Figure 15a). Before fine-tuning has commenced, the base Cifar-10 model achieves a substantially lower FID than random initialization, despite the large distribution shift between it and Camelyon17. Fine-tuning with either of the tested objectives maintains this advantage over 2000 epochs. Despite its higher convergence rate, training from random initialization is highly unstable compared to fine-tuning, as measured by both instantaneous velocity and Spearman correlation coefficient (Table 1). The variation in path length also strongly points to the benefits of fine-tuning (Figures 15b, 15c). Training from random initialization results in a sharp rise in average path length, which does not decrease over the scope of the experiment. Although fine-tuning from the pretrained model also results in a slight increase in average path length, the relative scale is significantly smaller than training from scratch. This suggests that



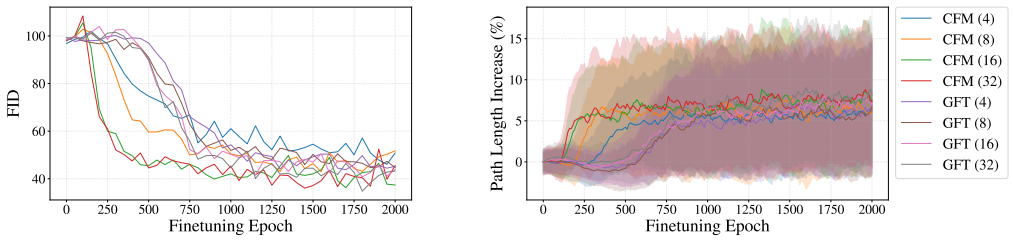
(a) FID. (b) Path Length.

Figure 10: Cross-domain adaptation with varying terminal temperatures. A lower final path length corresponds to higher final FID, and this tradeoff can be predictably controlled by changing β.



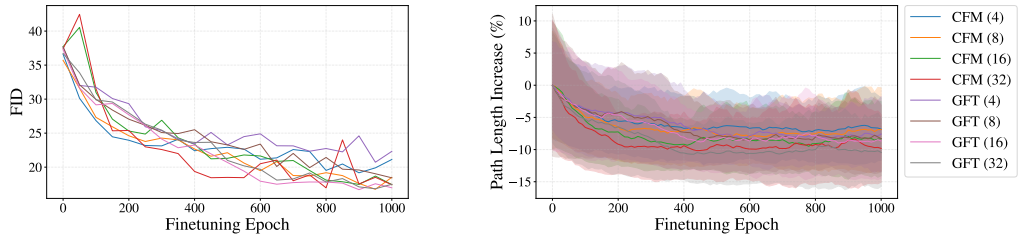
(a) FID. (b) Path Length.

Figure 11: Cross-domain adaptation with varying types of cooling schedules. All three schedule achieve very similar final values for FID and Path Length.



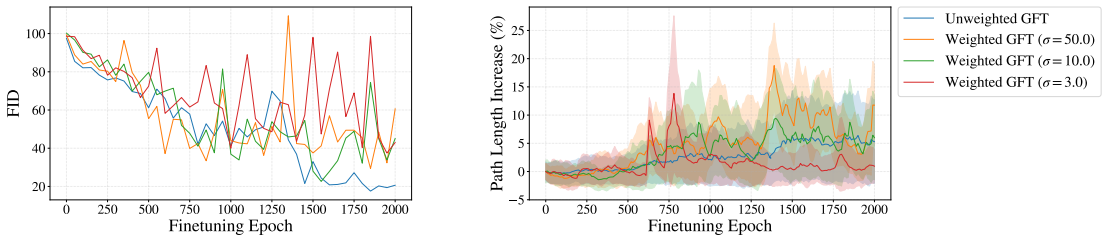
(a) FID. (b) Path Length.

Figure 12: Cross-domain adaptation with varying values for LoRA Rank for both CFM and GFT. Generally, an increase rank leads to a lower final FID but higher final Path Length.



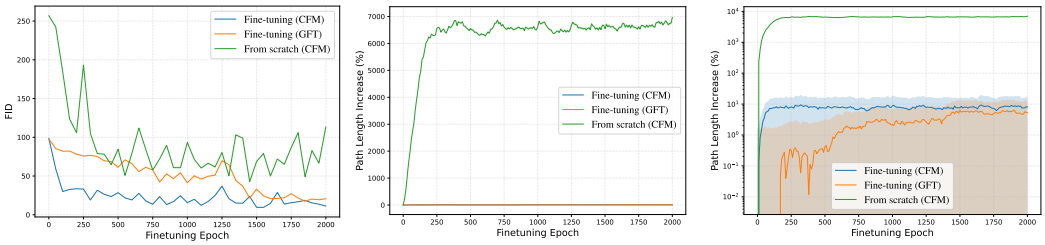
(a) FID. (b) Path Length.

Figure 13: In-domain adaptation with varying values for LoRA Rank for both CFM and GFT. Generally, an increase rank leads to a lower final FID but higher final Path Length.



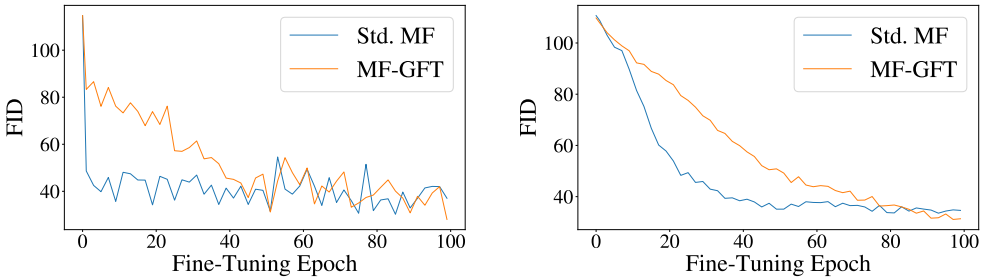
(a) FID. (b) Path Length.

Figure 14: Cross-domain adaptation with and without Girsanov weighting, with varying values for the diffusion coefficient. Unweighted GFT acheives the lowest final FID, and although Weighted GFT with $\sigma = 3.0$
acheives the lowest final peth length, it has a significantly higher variance than Unweighted GFT.



(a) FID. (b) Path length. (c) Path length (logarithmic scale).

Figure 15: Cross-domain adaptation results, as opposed to training from random initialization. Fine-tuning and pretraining are done on the Camelyon17 train dataset. The shaded regions of the path length graphs (b, c) represent one standard deviation from the mean.



(a) No EMA. (b) EMA.

Figure 16: Mean Flow Cross-domain adaptation on Camelyon17, with one-step generation. GFT achieves the same accuracy as standard MF fine-tuning, but with significantly higher stability.

the pretrained Cifar-10 vector field provides a structural prior that remains partially valid even under large shifts.

Q.1 Mean Flow Empirical Results

We begin with a Mean Flow model pretrained on the Cifar10 dataset, and fine-tuning it to the Camelyon17 train dataset using the standard Mean Flow training algorithm, and the GFT Mean Flow algorithm (2). The GFT Mean Flow algorithm reaches the same accuracy as the standard fine-tuning (Figure 16), but with significantly improved stability in both the normal update and EMA update settings (Table 3). The increased stability provided by GFT is particularly important when working with one-step generative models, which are known to be highly unstable during training. Compared to an iterative generative model fine-tuned from Cifar10 to the Camelyon17 dataset, the MF-GFT model achieves comparable FID with a 50-100x reduction in model evaluations during inference, with only a minor decrease in performance. In fact, the final FID of the MF-GFT model without EMA updates (26.665) was lower than that of the LoRA CFM model with 100 NFEs (35.120).

Table 3: Convergence and stability analysis for MeanFlow on the Camelyon17 dataset. GFT increases stability for both normal and EMA-updated fine-tuning.

	Objective	**Inst. Variance** $\downarrow$	**Convergence Rate** $\uparrow$	**Spearman** $\rho \downarrow$
No EMA	Standard MF	87.202	0.368	-0.42
	MF-GFT	**44.066**	**0.648**	**-0.88**
EMA	Standard MF	34.433	**0.805**	-0.93
	MF-GFT	**16.341**	0.794	**-1.00**

Table 4: Cross-Domain Adaptation Vendi scores across datasets and fine-tuning methods. GFT matches or exceeds the diversity of CFM across all datasets.

Dataset	Method	Vendi Score	% of Original
Camelyon17	Original Data	8.51	100.00%
	Full (GFT)	7.52	88.37%
	Full (CFM)	6.63	77.91%
RxRx1	Original Data	5.71	100.00%
	Full (GFT)	4.66	81.61%
	Full (CFM)	3.59	62.87%
FMoW	Original Data	21.57	100.00%
	Full (GFT)	11.45	53.08%
	Full (CFM)	11.75	54.47%

R Additional Results

R.1 Extended Results on Cross-Domain Adaptation

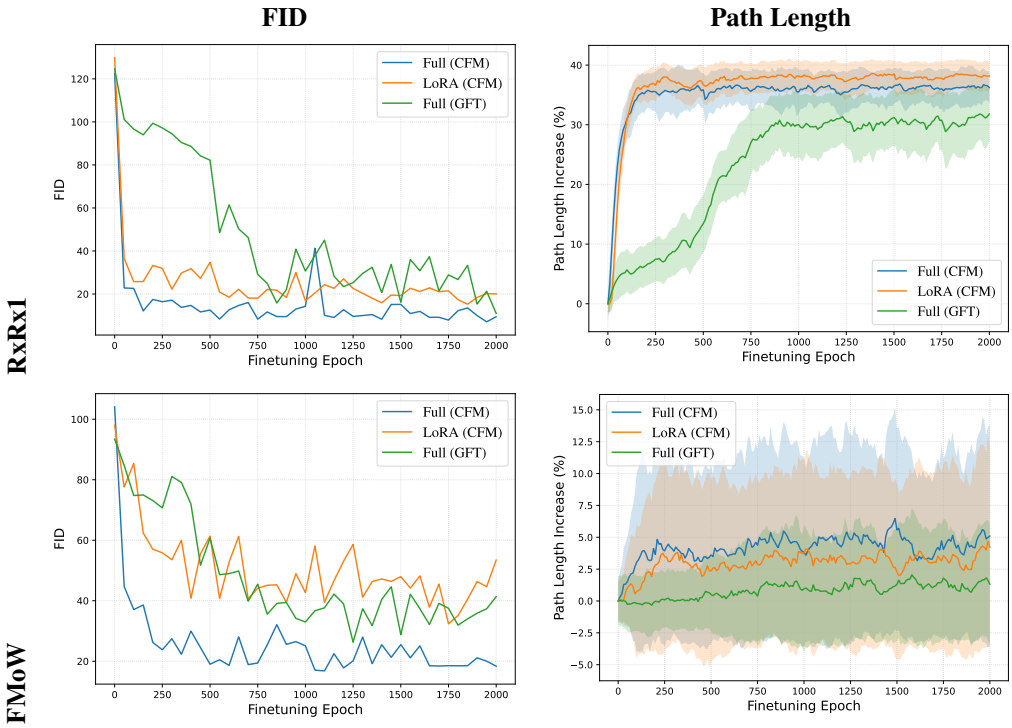


Figure 17: Fine-tuning performance from a pretrained Cifar-10 model to various WILDS train datasets. Rows corresponding to individual datasets, and the columns show FID and probability path lengths through 2000 epochs of fine-tuning. The shaded regions of the path length graphs correspond to one standard deviation from the mean.

R.2 Cross-Domain Adaptation Diversity

R.3 Extended Results on In-Domain Adaptation

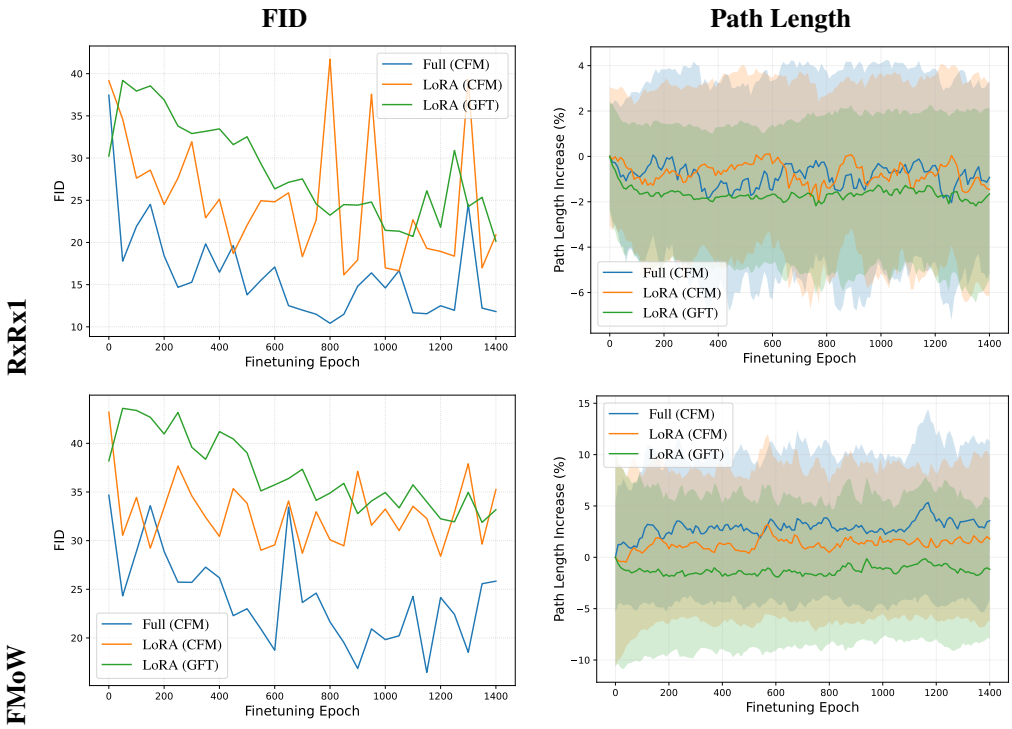


Figure 18: Fine-tuning performance from a fully fine-tuned model to two WILDS validation datasets. Rows corresponding to individual datasets, and the columns show FID and probability path lengths through fine-tuning. The shaded region of the path length graphs show one standard deviation from the mean.

R.4 GENERATED IMAGES

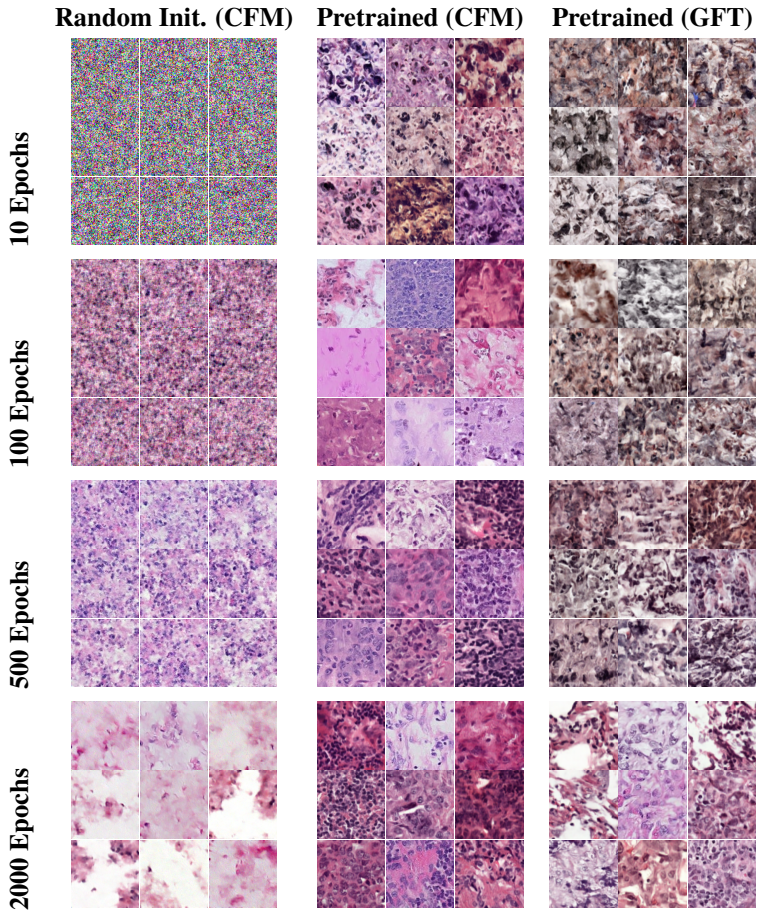


Figure 19: Generated images at various training stages from models trained or fine-tuned on the Camelyon17 train dataset. Rows correspond to the number of training/fine-tuning steps, and columns correspond to training initialization and objective.

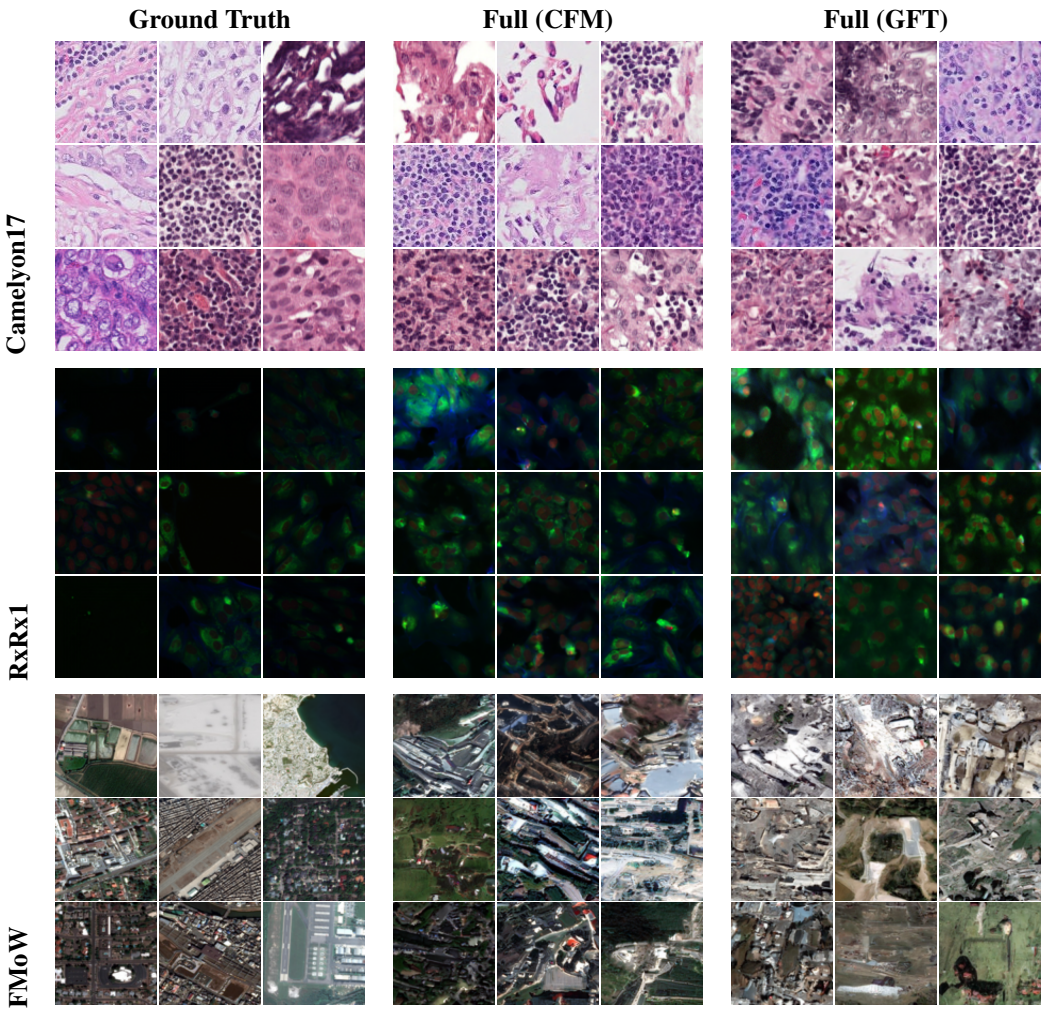


Figure 20: Generated images after 2000 epochs of fine-tuning on three WILDS datasets, starting from a pretrained Cifar-10 model.

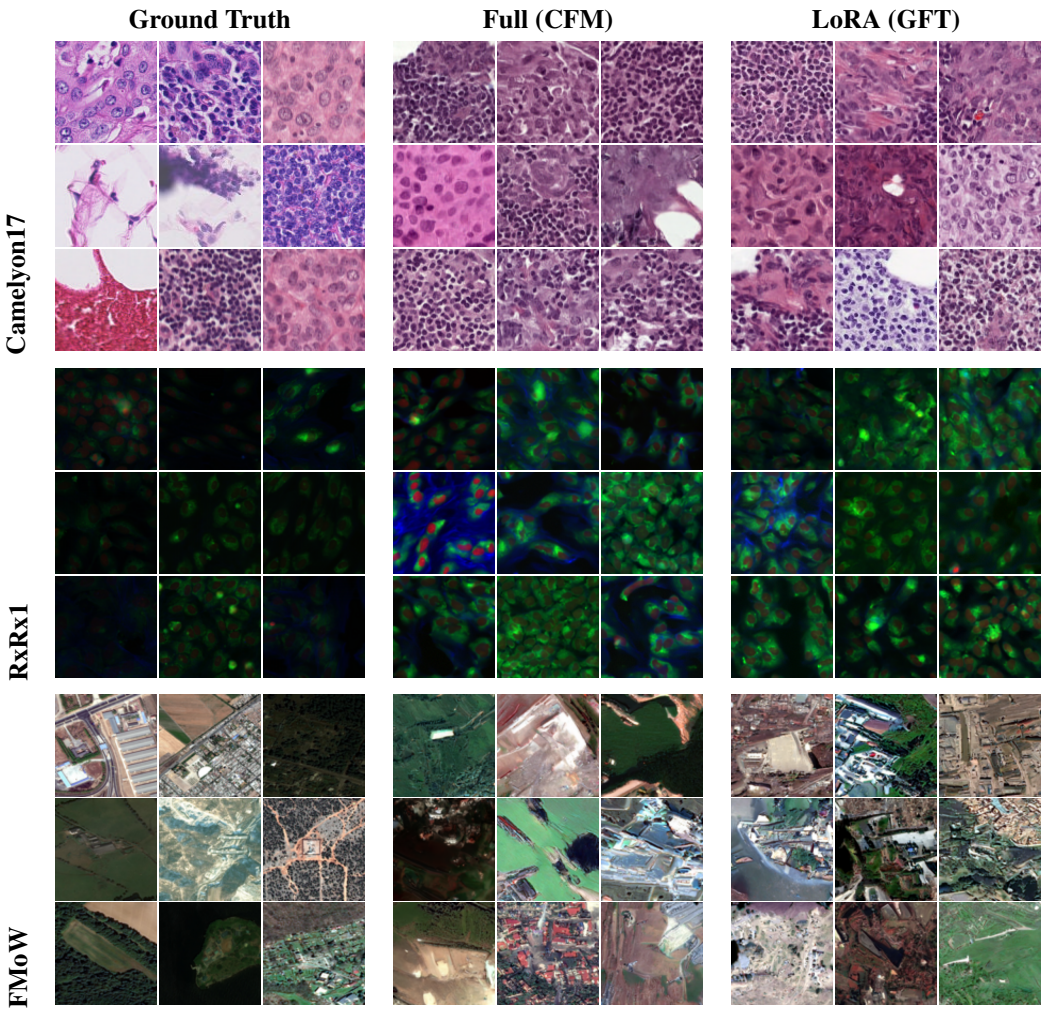


Figure 21: Generated images after fine-tuning on the validation splits of three WILDS datasets. Fine-tuning starts from a pretrained Cifar-10 model which has been fully fine-tuned on the train set of the corresponding WILDS dataset.

R.5 Extended Stability and Convergence Results

Table 5: Convergence stability comparison over large distribution shifts. Each method starts from a pretrained Cifar-10 model, and is fine-tuned on a WILDS train set.

	Objective	**Inst. Variance** $\downarrow$	**Avg. Conv. Rate** $\uparrow$	**Spearman** $\rho \downarrow$
RxRx1	Full (CFM)	191.685	3.110×10^{-2}	-0.560
	LoRA (CFM)	166.996	3.058×10^{-2}	-0.602
	Full (GFT)	**117.470**	$\mathbf{4.820 \times 10^{-2}}$	**-0.800**
FMoW	Full (CFM)	90.369	$\mathbf{2.827 \times 10^{-2}}$	-0.584
	LoRA (CFM)	66.178	2.154×10^{-2}	-0.552
	Full (GFT)	**43.226**	2.419×10^{-2}	**-0.759**

Table 6: Convergence stability comparison over small distribution shifts. Each method starts from the base Cifar-10 model after it has been fully fine-tuned on a WILDS train set, and is fine-tuned for the corresponding validation set.

	Objective	Inst. Variance ↓	Avg. Conv. Rate ↑	Spearman ρ ↓
RxRx1	Full (CFM)	10.654	9.779×10^{-3}	-0.165
	LoRA (CFM)	45.366	$\mathbf{1.014 \times 10^{-2}}$	-0.329
	LoRA (GFT)	**2.347**	3.743×10^{-3}	**-0.518**
FMoW	Full (CFM)	12.334	4.456×10^{-3}	-0.550
	LoRA (CFM)	9.236	1.406×10^{-3}	-0.125
	LoRA (GFT)	**2.881**	$\mathbf{6.157 \times 10^{-3}}$	**-0.884**

S Reward vs Sample-based Fine-Tuning Distribution

Reward-based fine-tuning methods assume access to a known reward model $r(x)$, and the goal of fine-tuning is to train a model to generate samples from the tilted distribution

$$p^*(x) \propto p^{\text{base}}(x) \exp(r(x))$$

In the case where we are directly given a fine-tuning target distribution q (in the form of a dataset of samples from q), our goal is now to train a model which generates samples from

$$q(x) = p(x)w(x), \qquad w(x) = \frac{q(x)}{p(x)} \tag{135}$$

In practice, we do not know the true target distribution q. However, we can prove that the density ratio w is exactly the Radon-Nikodym derivative.

proof.

Let $\mathbb{P}_p$ and $\mathbb{P}_q$ be the path measures associated with the continuous processes which generate samples from the base and fine-tuned target distributions p and q, respectively. In order to apply the Radon-Nikodym theorem, we assume that $\mathbb{P}_q$ is absolutely continuous with respect to $\mathbb{P}_p$. Then, there exists a unique function $w(x)$ such that

$$\mathbb{P}_q(A) = \int_A w(x) d\mathbb{P}_p$$

where A is any measurable set in the sample space. The function w is the Radon-Nikodym derivative, and is denoted as

$$w = \frac{d\mathbb{P}_q}{d\mathbb{P}_p}$$

Given that $\mathbb{P}_p, \mathbb{P}_q$ are defined over a space with an underlying Lebesgue base measure λ, we can write them in terms of their probability density functions.

$$\mathbb{P}_p = \int p(x) d\lambda \qquad \mathbb{P}_q = \int q(x) d\lambda$$
$$p(x) = \frac{d\mathbb{P}_p}{d\lambda} \qquad q(x) = \frac{d\mathbb{P}_q}{d\lambda}$$

We can then apply the chain rule for Radon-Nikodym derivatives to recover the density ratio.

$$w(x) = \frac{d\mathbb{P}_q}{d\mathbb{P}_p}(x) = \frac{d\mathbb{P}_q/d\lambda}{d\mathbb{P}_p/d\lambda}(x) = \frac{q(x)}{p(x)}$$

T Corresponding Reward Function

Reward-based methods fine-tune the base model to achieve generation from a tilted distribution defined by an external reward function $r(x)$.

$$p^*_\theta(x) \propto p_{\theta_0}(x) \exp(r(x)) \tag{136}$$

The reward function which results in an equivalent optimum to our proposed objective function is

$$r(x) = \frac{1}{1+\beta} \log\left(\frac{q(x)}{p_{\theta_0}(x)}\right) \tag{137}$$

We can easily verify this by substituting this expression into the tilted distribution.

$$p^*_\theta(x) \propto p_{\theta_0}(x) \exp\left(\frac{1}{1+\beta} \log\left(\frac{q(x)}{p_{\theta_0}(x)}\right)\right) \tag{138}$$

$$= p_{\theta_0}(x) \left(\frac{q(x)}{p_{\theta_0}(x)}\right)^{\frac{1}{1+\beta}} \tag{139}$$

$$= p_{\theta_0}(x)^{\frac{\beta}{1+\beta}} q(x)^{\frac{1}{1+\beta}} \tag{140}$$

Since the distribution $q(x)$ is unknown, this reward function is not tractable in it's current form. However, in Appendix S we prove the relationship between the density ratio and the Radon-Nikodym derivative, and in Appendix F we give the expression for the logarithm of the Radon-Nikodym derivative. Putting these together, we can rewrite the reward function as:

$$r(x) = \frac{1}{1+\beta} \left(\int_0^1 \sigma_t^{-1}(v_q - v_{\theta_0})^\intercal dB_t - \frac{1}{2}\int_0^1 \|\sigma_t^{-1}(v_q - v_{\theta_0})\|^2 dt\right) \tag{141}$$

The reward function is now written in terms of the drift terms of the base, target, and fine-tuned processes. As a result, this reward is only tractable when it is conditioned on given source and target samples. This is because the vector field v_q is unknown, and can only be defined conditionally given known samples. In the setting where a memoryless noise schedule is required to achieve unbiased convergence, this reward function is therefore intractable.

U Equivalence Between Total and Source-Conditional KL Divergence

Let $\mathbb{P}_f$ and $\mathbb{P}_g$ be the path measures of two SDEs on the time interval $[0, 1]$, with initial samples from the same source distribution and the same diffusion coefficient σ_t.

$$\mathbb{P}_f: \quad dX_t = f_t(X_t, t)dt + \sigma_t dB_t, \qquad X_0 \sim p_0 \tag{142}$$

$$\mathbb{P}_g: \quad dX_t = g_t(X_t, t)dt + \sigma_t dB_t, \qquad X_0 \sim p_0 \tag{143}$$

We assume that the standard conditions of Girsanov's Theorem hold, such that $\mathbb{P}_f$ and $\mathbb{P}_g$ are mutually absolutely continuous. The path measure $\mathbb{P}$ for any stochastic process starting at the state X_0 can be decomposed into the product of the initial distribution $p_0(X_0)$ and the conditional law of the paths given this starting point, $\mathbb{P}(\cdot|X_0)$.

$$\mathbb{P}(\cdot) = \int \mathbb{P}(\cdot|X_0)p_0(X_0)dX_0. \tag{144}$$

The path measure $\mathbb{P}$ then admits a disintegration with respect to X_0.

$$d\mathbb{P} = p_0(X_0)dX_0 \cdot d\mathbb{P}(A|X_0) \tag{145}$$

Since the path measures $\mathbb{P}_f$ and $\mathbb{P}_g$ share the same source distribution, applying this chain rule to both yields

$$d\mathbb{P}_f = p_0(X_0)dX_0 \cdot d\mathbb{P}_f(\cdot|X_0) \tag{146}$$

$$d\mathbb{P}_g = p_0(X_0)dX_0 \cdot d\mathbb{P}_g(\cdot|X_0) \tag{147}$$

We can then write the expression for the conditional Radon-Nikodym derivative of $\mathbb{P}_f$ and $\mathbb{P}_g$, pointwise over X_0.

$$\frac{d\mathbb{P}_f}{d\mathbb{P}_g} = \frac{p_0(X_0)}{p_0(X_0)} \times \frac{d\mathbb{P}_f(\cdot|X_0)}{d\mathbb{P}_g(\cdot|X_0)} \tag{148}$$

$$\frac{d\mathbb{P}_f}{d\mathbb{P}_g} = \frac{d\mathbb{P}_f(\cdot|X_0)}{d\mathbb{P}_g(\cdot|X_0)} \tag{149}$$

Therefore, the Radon-Nikodym derivative of the path measures conditioned on a given source sample X_0 is equivalent to the total derivative when the path measures share the same initial distribution. It follows from the proof given in appendix F that

$$KL(\mathbb{P}_f(\cdot|X_0)\|\mathbb{P}_g(\cdot|X_0)) = -\mathbb{E}_{\mathbb{P}_f(\cdot|X_0)}\left[\int_0^1 \sigma_t^{-1}(g_t - f_t)^\intercal dB_t\right] + \mathbb{E}_{\mathbb{P}_f(\cdot|X_0)}\left[\frac{1}{2}\int_0^1 \|\sigma_t^{-1}(g_t - f_t)\|^2 dt\right] \tag{150}$$

The stochastic integral is a Martingale under the path measure $\mathbb{P}_f$, and its expectation vanishes.

$$KL(\mathbb{P}_f(\cdot|X_0)\|\mathbb{P}_g(\cdot|X_0)) = \mathbb{E}_{\mathbb{P}_f(\cdot|X_0)}\left[\frac{1}{2}\int_0^1 \|\sigma_t^{-1}(g_t - f_t)\|^2 dt\right] \tag{151}$$

We can take the expectation over initial states X_0 to get

$$\mathbb{E}_{X_0\sim p_0}\left[KL(\mathbb{P}_f(\cdot|X_0)\|\mathbb{P}_g(\cdot|X_0))\right] = \mathbb{E}_{X_0\sim p_0}\left[\mathbb{E}_{\mathbb{P}_f(\cdot|X_0)}\left[\frac{1}{2}\int_0^1 \|\sigma_t^{-1}(g_t - f_t)\|^2 dt\right]\right] \tag{152}$$

By the Law of Total Expectation, nesting these expectations is equivalent to taking the expectation over the full path measure $\mathbb{P}_f$. We therefore find the final identity

$$KL(\mathbb{P}_f\|\mathbb{P}_g) = \mathbb{E}_{X_0\sim p_0}\left[KL(\mathbb{P}_f(\cdot|X_0)\|\mathbb{P}_g(\cdot|X_0))\right] = \mathbb{E}_{\mathbb{P}_f}\left[\frac{1}{2}\int_0^1 \|\sigma_t^{-1}(g_t - f_t)\|^2 dt\right] \tag{153}$$

V Equivalence Between Total and OT-Conditional KL Divergence

Let $\mathbb{P}_f$ and $\mathbb{P}_g$ be the path measures of two SDEs on the time interval $[0, 1]$, with initial samples from the same source distribution and the same diffusion coefficient σ_t.

$$\mathbb{P}_f: \quad dX_t = f_t(X_t, t)dt + \sigma_t dB_t, \qquad X_0 \sim p_0 \tag{154}$$

$$\mathbb{P}_g : \quad dX_t = g_t(X_t, t)dt + \sigma_t dB_t, \qquad X_0 \sim p_0 \tag{155}$$

We assume that the standard conditions of Girsanov's Theorem hold, such that $\mathbb{P}_f$ and $\mathbb{P}_g$ are mutually absolutely continuous. Let $\pi \in \Pi(p_0, q)$ be any coupling between the source and target distributions that admits the marginals $\int \pi(X_0, X_1)dX_1 = p_0(X_0)$ and $\int \pi(X_0, X_1)dX_0 = q(X_1)$. Consider the following expected KL divergence over jointly drawn samples:

$$\mathbb{E}_{(X_0,X_1)\sim\pi}\left[KL(\mathbb{P}_f(\cdot|X_0)\|\mathbb{P}_g(\cdot|X_0, X_1))\right] \tag{156}$$

Using the definition of π, we can write the integral form of this expectation.

$$\mathcal{L}_\pi = \int\int KL(\mathbb{P}_f(\cdot|X_0)\|\mathbb{P}_g(\cdot|X_0, X_1))\pi(X_0, X_1)dX_0dX_1 \tag{157}$$

Expanding the definition of the KL divergence for path measures:

$$\mathcal{L}_\pi = \int \pi(X_0, X_1)\left(\int \log \frac{d\mathbb{P}_f(\cdot|X_0)}{d\mathbb{P}_g(\cdot|X_0, X_1)} d\mathbb{P}_f(\cdot|X_0)\right) dX_0dX_1 \tag{158}$$

For the target process g, the relationship between the measure $\mathbb{P}_g(\cdot|X_0, X_1)$ and the source-conditioned measure $\mathbb{P}_g(\cdot|X_0)$ is given by Bayes Theorem,

$$d\mathbb{P}_g(\cdot|X_0, X_1) = \frac{d\mathbb{P}_g(\cdot|X_0)p_g(X_1|\cdot)}{p_g(X_1|X_0)} \tag{159}$$

where $p_g(X_1|\cdot)$ is the density of the terminal state of a path. Since we condition directly on X_1, this terminal-state likelihood evaluates to a Dirac measure, which is independent of f. We now substitute this into the Radon-Nikodym derivative.

$$\log \frac{d\mathbb{P}_f(\cdot|X_0)}{d\mathbb{P}_g(\cdot|X_0, X_1)} = \log \frac{d\mathbb{P}_f(\cdot|X_0)}{d\mathbb{P}_g(\cdot|X_0)} + \log \frac{p_g(X_1|X_0)}{p_g(X_1|\cdot)} \tag{160}$$

Finally, we plug this back into the coupled objective and distribute the expectation.

$$\mathcal{L}_\pi = \mathbb{E}_\pi\left[KL(\mathbb{P}_f(\cdot|X_0)\|\mathbb{P}_g(\cdot|X_0))\right] + \mathbb{E}_\pi\mathbb{E}_{\mathbb{P}_f(\cdot|X_0)}\left[\log \frac{p_g(X_1|X_0)}{p_g(X_1|\cdot)}\right] \tag{161}$$

$$= \int p_0(X_0)KL(\mathbb{P}_f(\cdot|X_0)\|\mathbb{P}_g(\cdot|X_0))dX_0 + \mathcal{C}(\pi, g) \tag{162}$$

The first term is exactly the global KL divergence $KL(\mathbb{P}_f\|\mathbb{P}_g)$. The second term depends only on the coupling and the target process, and is therefore constant with respect to the parameters of f. This is because the ratio is determined only by the process g and the terminal state X_1, so the expectation over $\mathbb{P}_f(\cdot|X_0)$ does not introduce any dependence on f. For gradient-based optimization, this term does not affect convergence.

W Terminal Distribution Objective

We examine a related objective function, which works on the terminal distributions of the generative process rather than the path measures. Consider the following constrained SDE,

$$\min_{p_\theta} \left(KL(p_\theta\|q) + \beta KL(p_\theta\|p_{\theta_0})\right) \tag{163}$$

$$\text{s.t. } dX_t = v_\theta(X_t, t)dt + \sigma_t dB_t \tag{164}$$

where p_θ and p_{θ_0} are the terminal distributions from which the fine-tuned and pretrained models generates samples, respectively, and β is a scalar hyperparameter. The objective function corresponding to this constraint is

$$\mathcal{L}(p_\theta) = KL(p_\theta \| q) + \beta KL(p_\theta \| p_{\theta_0}). \tag{165}$$

We want to show that minimizing this objective function results in sampling from the desired target distribution q. We start by expressing the constraint in integral form with respect to the Lebesgue base measure λ for the continuous Euclidean space $\mathbb{R}^d$.

$$\mathcal{L}(p_\theta) = \int p_\theta(X) \log \frac{p_\theta(X)}{q(X)} d\lambda(X) + \int \beta \cdot p_\theta(X) \log \frac{p_\theta(X)}{p_{\theta_0}(X)} d\lambda(X) \tag{166}$$

We minimize $\mathcal{L}(p_\theta)$ subject to the constraint that p_θ is a valid probability distribution: $\int_{\mathcal{X}} p_\theta(X) d\lambda(X) = 1$. To do so, we introduce a Lagrange multiplier μ for this constraint, and form the Lagrangian $L(p_\theta)$.

$$L(p_\theta) = \mathcal{L}(p_\theta) - \mu \left(\int_{\mathcal{X}} p_\theta(X) d\lambda(X) - 1 \right) \tag{167}$$

$$L(p_\theta) = \int_{\mathcal{X}} \left[p_\theta(X) \log \frac{p_\theta(X)}{q(X)} + \beta \cdot p_\theta(X) \log \frac{p_\theta(X)}{p_{\theta_0}(X)} - \mu p_\theta(X) \right] d\lambda(X) + \mu \tag{168}$$

We now minimize by setting the derivative with respect to p_θ to zero, and solving to find the optimal p^*_θ.

$$\frac{dL}{dp_\theta} = (\log p_\theta + 1 - \log q) + \beta(\log p_\theta + 1 - \log p_{\theta_0}) - \mu = 0 \tag{169}$$

Finally, we solve for the optimal $p^*_\theta(X)$.

$$(1+\beta) \log p_\theta + 1 + \beta - \log q - \beta \log p_{\theta_0} - \mu = 0 \tag{170}$$

$$(1+\beta) \log p_\theta = \log q + \beta \log p_{\theta_0} + \mu - 1 - \beta \tag{171}$$

$$\log p_\theta = \frac{1}{1+\beta}(\log q) + \frac{\beta}{1+\beta}(\log p_{\theta_0}) + \frac{\mu - 1 - \beta}{1+\beta} \tag{172}$$

$$p_\theta = \exp\left(\frac{1}{1+\beta}(\log q) + \frac{\beta}{1+\beta}(\log p_{\theta_0}) + \frac{\mu - 1 - \beta}{1+\beta} \right) \tag{173}$$

$$p^*_\theta = q^{\frac{1}{1+\beta}} \cdot p_{\theta_0}^{\frac{\beta}{1+\beta}} \exp\left(\frac{\mu - 1 - \beta}{1+\beta} \right) \tag{174}$$

We set Z to be a constant that absorbs all terms not related to X. Then, the final optimal distribution is

$$p^*_\theta(X) = Z \cdot q(X)^{\frac{1}{1+\beta}} \cdot p_{\theta_0}(X)^{\frac{\beta}{1+\beta}}. \tag{175}$$

Cooling Schedule Note that the exponents $\frac{1}{1+\beta}$ and $\frac{\beta}{1+\beta}$ sum to 1, meaning that we have defined a weighted average of the pretrained and fine-tuned target distributions. This means that hyperparameter β can be adjusted to achieve a fine-tuned model that generates from a user-defined target distribution.

$$p_\theta^*(X) \propto \begin{cases} q(X) & \text{if } \beta = 0 \\ \sqrt{p_{\theta_0}(X)q(X)} & \text{if } \beta = 1 \\ p_{\theta_0}(X) & \text{if } \beta = \infty \end{cases} \tag{176}$$

This means that the second KL term is acting as a regularization for the change from the pretrained model, and the hyperparameter β controls the strength of this regularization. β can therefore be used as a cooling schedule for simulated annealing Kirkpatrick et al. (1983) from the pretrained distribution $p_{\theta_0} \approx p$ to the fine-tuned target distribution $p_\theta \approx q$.