Generative Modeling and Stochastic Control as Dynamics on Probability Distributions

by

Carles Domingo i Enrich

A dissertation submitted in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

Department of Computer Science

New York University

May, 2024

Professor Joan Bruna

Acknowledgements

I would not have been able to complete this thesis without the help and support of many people. This is an attempt to acknowledge everyone's contributions, but it is not exhaustive by any means. First, I would like to thank my advisor Joan for his great guidance, for the many engaging research discussions that we have had, and for giving me the right amount of freedom to develop into an independent researcher. I also want to thank all of the members of my committee: Eric, Jon, Lester, and Ricky. It has been a pleasure to collaborate with and learn from you, and your advice has been invaluable.

During my PhD, I have been fortunate to do summer internships at IBM Research and at Microsoft Research, and to be a visiting researcher at Meta FAIR Labs on a longer term appointment which I have held until the end of my PhD. I want to thank Youssef from IBM for our collaboration and his advice in the first years of my PhD. I am also very grateful to Lester from Microsoft for his help and mentorship during my internship and afterwards, as well as to Raaz, who I also started working with while at MSR and who has given me very helpful advice. Within Meta, I want to send a sincere thank you to Ricky and Brandon for everything that I have learned during my stay. Spending time here has shown me a different perspective on research and has made me more well-rounded.

Beyond my advisor and mentors, I want to single out Samy and Aram as two collaborators and friends that I have enjoyed chatting and doing research with on multiple occasions. I would like to thank effusively other researchers and friends that I have worked with and learned from over

the course of my PhD: Alberto, Jiequn, Grant, Arthur, Damien, Fabian, Marylou, Yaron, Heli, Yair, and Vivien, among others. I also want to thank wholeheartedly other friends for interesting chats about research and other topics, and for your friendship throughout my PhD: Thomas, Matthew, Jaume, Irina, Maria, Alberto, Paola, Carles, Tomàs, Polina, Vlad, Richard, Rama, and Lei, among others (I apologize to those missing). I would also like to send special thanks to my housemates and friends Alejandro and Orestis, for their continued support and for all the chats that we have had in our living room.

Dono les gràcies als amics que mantinc a Vilanova, i als que conservo dels meus anys a la UPC. La llista és llarga, i no incloc els noms, però sabeu qui sou. Finalment, dono les gràcies a la meva família; en particular, als meus pares, Anna i Josep, pel vostre suport durant tots aquests anys, sense el qual res de tot això no hauria estat possible, i als meus germans, Roger i Emma, a qui desitjo el millor.

Abstract

Several modern machine learning algorithms can be studied from the perspective of evolution dynamics on the space of probability measures. Gradient descent-ascent algorithms that are used to solve minimax problems such as the ones arising in generative adversarial networks (GANs) can be interpreted as a joint evolution of two measures: one over the space of parameters of the generator, and one over the space of parameters of the discriminator. In chapter 2, we study systems of this form, and we provide convergence guarantees when possible. Diffusion models, which are another generative modeling technique, are also based on dynamics on probability measures, in this case over the space of samples. The dynamics are simulated at inference time; the starting distribution is a Gaussian, and the final distribution is meant to be the target data distribution. Diffusion models were generalized by the Flow Matching framework, which allows to construct different paths between the Gaussian noise distribution and the data distribution. In chapter 3, we introduce Multisample Flow Matching, which is a generalization of Flow Matching with intimate connections to optimal transport. Stochastic optimal control is a third problem where dynamics on measures play a critical role. The goal is to learn a vector field (the control) in order to drive the behavior of the solutions of a stochastic differential equation. In chapter 4, we present Stochastic Optimal Control Matching, a least-squares loss that is based on the same principles that are used to formulate diffusion model losses, and which achieves errors that are an order of magnitude lower than for existing methods.

Contents

List of Figures

List of Tables

List of Appendices

1 Introduction

A common thread among the three works covered in this thesis is that dynamics on probability distributions play a central role in all of them. These probability distributions can be defined over the space of parameters, or over the space of samples. A brief summary of each of the works can be found below.

Mean-field analysis of two-player zero-sum games [Domingo-Enrich et al. 2020] Finding Nash equilibria in two-player zero-sum continuous games is a central problem in machine learning, e.g. for training both GANs and robust models. The existence of pure Nash equilibria requires strong conditions which are not typically met in practice. Mixed Nash equilibria exist in greater generality and may be found using mirror descent. Yet this approach does not scale to high dimensions. To address this limitation, we parametrize mixed strategies as mixtures of particles, whose positions and weights are updated using gradient descent-ascent. We study this dynamics as an interacting gradient flow over measure spaces endowed with the Wasserstein-Fisher-Rao metric. We establish global convergence to an approximate equilibrium for the related Langevin gradient-ascent dynamic. We prove a law of large numbers that relates particle dynamics to mean-field dynamics. Our method identifies mixed equilibria in high dimensions and is demonstrably effective for training mixtures of GANs.

Multisample Flow Matching [Pooladian et al. 2023] Simulation-free methods for training continuous-time generative models construct probability paths that go between noise distribu-

tions and individual data samples. Recent works, such as Flow Matching, derived paths that are optimal for each data sample. However, these algorithms rely on independent data and noise samples, and do not exploit underlying structure in the data distribution for constructing probability paths. We propose Multisample Flow Matching, a more general framework that uses non-trivial couplings between data and noise samples while satisfying the correct marginal constraints. At very small overhead costs, this generalization allows us to (i) reduce gradient variance during training, (ii) obtain straighter flows for the learned vector field, which allows us to generate high-quality samples using fewer function evaluations, and (iii) obtain transport maps with lower cost in high dimensions, which has applications beyond generative modeling. Importantly, we do so in a completely simulation-free manner with a simple minimization objective. We show that our proposed methods improve sample consistency on downsampled ImageNet data sets, and lead to better low-cost sample generation.

Stochastic Optimal Control Matching [Domingo-Enrich et al. 2023] Stochastic optimal control, which has the goal of driving the behavior of noisy systems, is broadly applicable in science, engineering and artificial intelligence. Our work introduces Stochastic Optimal Control Matching (SOCM), a novel Iterative Diffusion Optimization (IDO) technique for stochastic optimal control that stems from the same philosophy as the conditional score matching loss for diffusion models. That is, the control is learned via a least squares problem by trying to fit a matching vector field. The training loss, which is closely connected to the cross-entropy loss, is optimized with respect to both the control function and a family of reparameterization matrices which appear in the matching vector field. The optimization with respect to the reparameterization matrices aims at minimizing the variance of the matching vector field. Experimentally, our algorithm achieves lower error than all the existing IDO techniques for stochastic optimal control for three out of four control problems, in some cases by an order of magnitude. The key idea underlying SOCM is the path-wise reparameterization trick, a novel tech-

nique that is of independent interest, e.g., for generative modeling. The code can be found at https://github.com/facebookresearch/SOC-matching.

2 Mean-field two-player zero-sum games

2.1 Introduction

Multi-objective optimization problems arise in many fields, from economics to civil engineering. Tasks that require optimizing multiple objectives have also become a routine part of many agent-based machine learning algorithms including generative adversarial networks [Goodfellow et al. 2014], imaginative agents [Racanière et al. 2017], hierarchical reinforcement learning [Wayne and Abbott 2014] and multi-agent reinforcement learning [Bu et al. 2008]. It not only remains difficult to carry out the necessary optimization, but also to assess the optimality of a given solution.

Multi-agent optimization is generally cast as finding equilibria in the space of strategies. The classic notion of equilibrium is due to Nash [Nash 1951]: a Nash equilibrium is a set of agent strategies for which no agent can unilaterally improve its loss value. Pure Nash equilibria, in which each agent adopts a single strategy, provide a limited notion of optimality because they exist only under restrictive conditions. On the other hand, mixed Nash equilibria (MNE), where agents adopt a strategy from a probability distribution over the set of all strategies, exist in much greater generality [Glicksberg 1952]. Importantly, MNE exist for games with infinite-dimensional compact strategy spaces, in which each player observes a loss function that is continuous in its strategy. We encounter this setting in different game formulations of machine learning problems,

like GANs [Goodfellow et al. 2014].

Although MNE are guaranteed to exist, it is difficult to identify them. Indeed, worst-case complexity analyses have shown that without additional assumptions on the losses there is no efficient algorithm for finding a MNE, even in the case of two-player finite games [Daskalakis et al. 2009]. Some recent progress has been made; [Hsieh et al. 2019] proposed a mirror-descent algorithm with convergence guarantees, which is approximately realizable in high-dimension.

Contributions. Following Hsieh et al. [2019], we formulate continuous two-player zero-sum games as a multi-agent optimization problem over the space of probability measures on strategies. We describe two gradient descent-ascent dynamics in this space, both involving a transport term.

- We show that the stationary points of a gradient ascent-descent flow with Langevin diffusion over the space of mixed strategies are approximate MNE.
- We analyse a gradient ascent-descent dynamics that jointly updates the positions and weights of two mixed strategies to converge to an *exact* MNE. This dynamics corresponds to a gradient descent-ascent flow over the space of measures endowed with a Wasserstein-Fisher-Rao (WFR) metric [Chizat et al. 2018].
- We discretize both dynamics in space and time to obtain implementable training algorithms. We provide mean-field type consistency results on the discretization. We demonstrate numerically how both dynamics overcome the curse of dimensionality for finding MNE on synthetic games. On real data, we use WFR flows to train mixtures of GANs, that explicitly discover data clusters while maintaining good performance.

2.2 Related work

Equilibria in continuous games. Most of the works that study convergence to equilibria in continuous games or GANs do not frame the problem in the infinite-dimensional space of mea-

sures, but on finite-dimensional spaces. That is because they either (i) restrict their attention to games with convexity-concavity assumptions in which pure equilibria exist [Mertikopoulos et al. 2019; Lin et al. 2018; Nouiehed et al. 2019], or (ii) provide algorithms with convergence guarantees to local notions of equilibrium such as stable fixed points, local Nash equilibria and local minimax points [Heusel et al. 2017; Adolphs et al. 2018; Mazumdar et al. 2019; Jin et al. 2019; Fiez et al. 2019; Balduzzi et al. 2018]. Both approaches differ from ours, which is to give global convergence guarantees without convexity assumptions. Some works have studied approximate MNE in infinite-dimensional measure spaces. Arora et al. [2017] proved the existence of approximate MNE and studied the generalization properties of this approximate solution; their analysis, however, does not provide a constructive method to identify such a solution. In a more explicit setting, Grnarova et al. [2017] designed an online-learning algorithm for finding a MNE in GANs under the assumption that the discriminator is a single hidden layer neural network. Balandat et al. [2016] apply the dual averaging algorithm to the minimax problem and show that it recovers a MNE, but they do not provide any convergence rate nor a practical algorithm for learning mixed NE. Our framework holds without making any assumption on the architectures of the discriminator and generator and provides explicit algorithms with some convergence guarantees.

Mean-field view of nonlinear gradient descent. Our approach is closely related to the mean-field perspective on wide neural networks [Mei et al. 2018; Rotskoff and Vanden-Eijnden 2018; Chizat and Bach 2018; Sirignano and Spiliopoulos 2019; Rotskoff et al. 2019]. These methods view training algorithms as approximations of Wasserstein gradient flows, which are dynamics on measures over the space of neurons. In our setting, a mixed strategy corresponds to a measure over the space of strategies.

Particle approaches for two-player games. Our theoretical work sheds a new light on the results of Hsieh et al. [2019], and rigorously justifies important algorithmic modifications the authors introduced. Specifically, they give rates of convergence for infinite-dimensional mirror

descent on measures (i.e. updating strategy weights but not their positions). The straightforward implementation of this algorithm performs poorly unless the dimension is low (Figure 2.1), which is why they proposed an 'implementable' two-timescale version, in which the inner loop is a transport-based sampling procedure closely related to our Algorithm 1. This implementable version is not studied theoretically, as the two-timescale structure hinders a thorough analysis. Our analysis includes transport on equal footing with mirror descent updates.

2.3 Problem setup and mean-field dynamics

Notation. For a topological space $\mathcal{X}$ we denote by $\mathcal{P}(\mathcal{X})$ the space of Borel probability measures on $\mathcal{X}$, and $\mathcal{M}_+(\mathcal{X})$ the space of Borel (positive) measures. For a given measure $\mu \in \mathcal{P}(\mathcal{X})$ that is absolutely continuous with respect to the canonical Borel measure dx of $\mathcal{X}$ and has Radon-Nikodym derivative $\frac{d\mu}{dx} \in C(\mathcal{X})$, we define its differential entropy $H(\mu) = -\int \log(\frac{d\mu}{dx}) d\mu$. For measures $\mu, \nu \in \mathcal{P}(\mathcal{X})$, $\mathcal{W}_2$ is the 2-Wasserstein distance.

2.3.1 Lifting differentiable games to spaces of strategy distributions

Differentiable two-player zero-sum games. We recall the definition of a differentiable zero-sum game, and show how finding a mixed Nash equilibrium to such a game is equivalent to solving a bi-linear game in the infinite dimensional space of distributions on strategies. We will use gradient flow approaches for solving the lifted problem.

Definition 2.1. A two-player zero-sum game consists of a set of two players with parameters $z = (x, y) \in \mathcal{Z} = \mathcal{X} \times \mathcal{Y}$, where players observe a loss functions $\ell_1 : \mathcal{Z} \to \mathbb{R}$ and $\ell_2 : \mathcal{Z} \to \mathbb{R}$ that satisfy for all $(x, y) \in \mathcal{Z}$, $\ell_1(x, y) + \ell_2(x, y) = 0$. $\ell \triangleq \ell_1 = -\ell_2$ is the loss of the game.

The compact finite-dimensional spaces of strategies $\mathcal{X}$ and $\mathcal{Y}$ are endowed with a certain distance function d (which we assume Euclidean in what follows—subsection A.7.5 derives our results

on arbitrary strategy manifolds). This allows to define differentiable games, amenable to first-order optimization. We make the following mild assumption over the regularity of losses and constraints [Glicksberg 1952].

Assumption 1. *The parameter spaces $\mathcal{X}$ and $\mathcal{Y}$ are compact Riemannian manifolds without boundary of dimensions d_x, d_y embedded in $\mathbb{R}^{D_x}, \mathbb{R}^{D_y}$ respectively. The loss ℓ is continuously differentiable and L-smooth with respect to each parameter. That is, for all $x, x' \in \mathcal{X}$ and $y, y' \in \mathcal{Y}$, $\|\nabla_x \ell(x,y) - \nabla_x \ell(x',y')\|_2 \leq L(d(x,x')+d(y,y'))$, $\|\nabla_y \ell(x,y) - \nabla_y \ell(x',y')\|_2 \leq L(d(x,x')+d(y,y'))$.*

From pure to mixed Nash equilibria. Assuming that both players play simultaneously, a pure Nash equilibrium point is a pair of strategies $(x^*, y^*) \in \mathcal{X} \times \mathcal{Y}$ such that, for all $(x,y) \in \mathcal{X} \times \mathcal{Y}$, $\ell(x^\star, y) \leq \ell(x^\star, y^\star) \leq \ell(x, y^\star)$. Such points do not always exist in continuous games. In contrast, mixed Nash equilibria (MNE) are guaranteed to exist [Glicksberg 1952] under Assumption 1. Those distributions $(\mu_x^\star, \mu_y^\star) \in \mathcal{P}(\mathcal{X}) \times \mathcal{P}(\mathcal{Y})$ are global saddle points of the expected loss $\mathcal{L}(\mu_x, \mu_y) \triangleq \iint \ell(x,y) d\mu_x(x) d\mu_y(y)$. Formally, for all $\mu_x, \mu_y \in \mathcal{P}(\mathcal{X}) \times \mathcal{P}(\mathcal{Y})$,

$$\mathcal{L}(\mu_x^*, \mu_y) \leq \mathcal{L}(\mu_x^*, \mu_y^*) \leq \mathcal{L}(\mu_x, \mu_y^*). \tag{2.1}$$

We quantify the accuracy of an estimation $(\hat{\mu}_x, \hat{\mu}_y)$ of a MNE using the Nikaidô and Isoda [1955] error

$$\mathrm{NI}(\hat{\mu}_x, \hat{\mu}_y) = \sup_{\mu_y \in \mathcal{P}(\mathcal{Y})} \mathcal{L}(\hat{\mu}_x, \mu_y) - \inf_{\hat{\mu}_x \in \mathcal{P}(\mathcal{X})} \mathcal{L}(\mu_x, \hat{\mu}_y). \tag{2.2}$$

We track the evolution of this metric in our theoretical results (subsection 2.4.2) and in our experiments. We obtain guarantees on finding ε-MNE $(\mu_x^\varepsilon, \mu_y^\varepsilon)$, i.e. distribution pairs such that $\mathrm{NI}(\mu_x^\varepsilon, \mu_y^\varepsilon) \leq \varepsilon$.

2.3.2 Training dynamics on discrete mixtures of strategies

We study three different dynamics for solving (2.1). Let us first assume that the two players play *finite* mixtures of n strategies $\mu_x = \sum_{i=1}^n w_x^i \delta_{x^i} \in \mathcal{P}(\mathcal{X})$, $\mu_y = \sum_{i=1}^n w_y^i \delta_{y^i} \in \mathcal{P}(\mathcal{Y})$, where $\{x^i, y^i\}_{i\in[1:n]}$ are the positions of the strategies and $w_x^i, w_y^i \geq 0$ are their weights. In the simplest setting, those mixtures are assumed uniform, i.e. $w_x^i = w_y^i = 1/n$. Finding the best $2n$ strategies involve finding a saddle point of $\mathcal{L}(\mu_x, \mu_y) = \frac{1}{n^2}\sum_i \sum_j \ell(x_i, y_j)$. Starting from random independent initial strategies $x_0^i = \xi_i \sim \mu_{x,0}$, $y_0^i = \bar{\xi}_i \sim \mu_{y,0}$, we may hope that the gradient descent-ascent dynamics

$$\frac{dx_t^i}{dt} = -\frac{1}{n}\sum_{j=1}^n \nabla_x \ell(x_t^i, y_t^j), \qquad \frac{dy_t^i}{dt} = \frac{1}{n}\sum_{j=1}^n \nabla_y \ell(x_t^j, y_t^i), \quad \forall i \in [1:n] \tag{2.3}$$

finds such a saddle point. Yet this may fail in simple nonconvex-nonconcave games, as illustrated in subsection A.7.2—the particle distributions collapse to a stationary point that is not a MNE.

To mitigate this convergence problem, we analyse a perturbed dynamics analogous to Langevin gradient descent. Using the same initialization as in (2.3), we add a small amount of noise in the gradient dynamics and obtain the stochastic differential equations

$$dX_t^i = -\frac{1}{n}\sum_{j=1}^n \nabla_x \ell(X_t^i, Y_t^j)dt + \sqrt{\frac{2}{\beta}}dW_t^i, \; dY_t^i = \frac{1}{n}\sum_{j=1}^n \nabla_y \ell(X_t^j, Y_t^i)dt + \sqrt{\frac{2}{\beta}}d\bar{W}_t^i, \tag{2.4}$$

where $W_t^i, \bar{W}_t^i$ are independent Brownian motions. The discretization of (2.4) results in algorithm 1; it is similar to Alg. 4 in Hsieh et al. [2019].

We propose a second alternative dynamics to (2.3), that updates both the positions and the weights of the particles, using relative updates for weights. We will show that it enjoys better convergence

Algorithm 1 Langevin Descent-Ascent (L-DA)

Input: IID samples $x_0^1, \dots, x_0^n$ from $\mu_{x,0} \in \mathcal{P}(\mathcal{X})$, IID samples $y_0^1, \dots, y_0^n \in \mathcal{Y}$ from $\mu_{y,0} \in \mathcal{P}(\mathcal{Y})$

for $t = 0, \dots, T$ **do**
  for $i = 1, \dots, n$ **do**
    Sample $\Delta W_t^i \sim \mathcal{N}(0, I)$ $\quad x_{t+1}^i = x_t^i - \frac{\eta}{n} \sum_{j=1}^n \nabla_x \ell(x_t^i, y_t^j) + \sqrt{2\eta\beta^{-1}} \Delta W_t^i$
    Sample $\Delta \bar{W}_t^i \sim \mathcal{N}(0, I)$ $\quad y_{t+1}^i = y_t^i + \frac{\eta}{n} \sum_{j=1}^n \nabla_y \ell(x_t^j, y_t^i) + \sqrt{2\eta\beta^{-1}} \Delta \bar{W}_t^i$
  end
end

Output: $\mu_{x,T}^n = \frac{1}{n} \sum_{i=1}^n \delta_{x_T^i}, \quad \mu_{y,T}^n = \frac{1}{n} \sum_{i=1}^n \delta_{y_T^i}$

Algorithm 2 Wasserstein-Fisher-Rao Descent-Ascent (WFR-DA)

Input: IID samples $x_0^{(1)}, \dots, x_0^{(n)}$ from $\nu_{x,0} \in \mathcal{P}(\mathcal{X})$, IID samples $y_0^{(1)}, \dots, y_0^{(n)}$ from $\nu_{y,0} \in \mathcal{P}(\mathcal{Y})$.
Initial weights: For all $i \in [1:n]$, $w_x^{(i)} = 1$, $w_y^{(i)} = 1$.

for $t = 0, \dots, T$ **do**
  $[x_{t+1}^{(i)}]_{i=1}^n = [x_t^{(i)} - \eta \sum_{j=1}^n w_{y,t}^{(i)} \nabla_x \ell(x_t^{(i)}, y_t^{(j)})]_{i=1}^n$
  $\left[\hat{w}_{x,t+1}^{(i)}\right]_{i=1}^n = \left[w_{x,t}^{(i)} \exp\left(-\eta' \sum_{j=1}^n w_{y,t}^{(j)} \ell(x_t^{(i)}, y_t^{(j)})\right)\right]_{i=1}^n \quad [w_{x,t+1}^{(i)}]_{i=1}^n = \left[\hat{w}_{x,t+1}^{(i)}\right]_{i=1}^n / \sum_{j=1}^n \hat{w}_{x,t+1}^{(j)}$
  $[y_{t+1}^{(i)}]_{i=1}^n = [y_t^{(i)} + \eta \sum_{j=1}^n w_{x,t}^{(j)} \nabla_y \ell(x_t^{(j)}, y_t^{(i)})]_{i=1}^n$
  $\left[\hat{w}_{y,t+1}^{(i)}\right]_{i=1}^n = \left[w_{y,t}^{(i)} \exp\left(\eta' \sum_{j=1}^n w_{x,t}^{(j)} \ell(x_t^{(j)}, y_t^{(i)})\right)\right]_{i=1}^n \quad [w_{y,t+1}^{(i)}]_{i=1}^n = \left[\hat{w}_{y,t+1}^{(i)}\right]_{i=1}^n / \sum_{j=1}^n \hat{w}_{y,t+1}^{(j)}$
end

Output: $\bar{\nu}_{x,T}^n = \frac{1}{T+1} \sum_{t=0}^T \sum_{i=1}^n w_{x,T}^{(i)} \delta_{x_T^{(i)}}, \quad \bar{\nu}_{y,T}^n = \frac{1}{T+1} \sum_{t=0}^T \sum_{i=1}^n w_{y,T}^{(i)} \delta_{y_T^{(i)}}$

properties in the mean-field limit.

$$\frac{dx_t^i}{dt} = -\gamma \sum_{j=1}^n w_{y,t}^j \nabla_x \ell(x_t^i, y_t^j), \quad \frac{dw_{x,t}^i}{dt} = \alpha \left(- \sum_{j=1}^n w_{y,t}^j \ell(x_t^i, y_t^j) + K(t) \right) w_{x,t}^i \tag{2.5}$$

and similarly for all y_t^i (flipping the sign of ℓ). $K(t) \triangleq \sum_{k=1}^n \sum_{j=1}^n w_{y,t}^j w_{x,t}^k \ell(x_t^i, y_t^j)$ keeps $w_{x,t}$ in the simplex. We use uniform weights for initialization. When $\gamma = 0$ and $\alpha = 1$, only the weights are updated: this results in the continuous-time version of the infinite-dimensional mirror descent studied by Hsieh et al. [2019]. The Euler discretization of (2.5) results in algorithm 2.

2.3.3 Training dynamics as gradient flows on measures

The three dynamics that we have introduced at the level of particles induces dynamics on the associated empirical probability measures. If $\{x_t^i, y_t^i\}_{i\in[1,n]}$ is a solution of (2.3), then $\mu_x(t) = \frac{1}{n}\sum_{i=1}^n \delta_{x_t^i}$ and $\mu_y(t) = \frac{1}{n}\sum_{i=1}^n \delta_{y_t^i}$ are solutions of the *Interacting Wasserstein Gradient Flow* (IWGF) of $\mathcal{L}$:

$$\begin{cases} \partial_t \mu_x = \nabla \cdot (\mu_x \nabla_x V_x(\mu_y, x)), \quad \mu_x(0) = \frac{1}{n}\sum_{i=1}^n \delta_{x_0^i}, \\ \partial_t \mu_y = -\nabla \cdot (\mu_y \nabla_y V_y(\mu_x, y)), \quad \mu_y(0) = \frac{1}{n}\sum_{i=1}^n \delta_{y_0^i}. \end{cases} \tag{2.6}$$

The derivation of (2.6) is provided in subsection A.7.3. We use the notation $V_x(\mu_y, x) \triangleq \frac{\delta \mathcal{L}}{\delta \mu_x}(\mu_x, \mu_y)(x) = \int \ell(x, y) d\mu_y(y)$ for the first variations of the functional $\mathcal{L}(\mu_x, \mu_y)$. Holding μ_y fixed, the evolution of μ_x is a Wasserstein gradient flow on $\mathcal{L}(\cdot, \mu_y)$.We interpret these PDEs in the weak sense, i.e. equality holds when integrating measures against bounded continuous functions.

The distributions $\mu_x(t) = \frac{1}{n}\sum_{i=1}^n \delta_{X_t^i}$ and $\mu_y(t) = \frac{1}{n}\sum_{i=1}^n \delta_{Y_t^i}$, where $\{X^i, Y^i\}_{i\in[1:n]}$ are solutions of (2.4) follows a *Entropy-Regularized Interacting Wasserstein Gradient Flow (ERIWGF)*:

$$\begin{cases} \partial_t \mu_x = \nabla_x \cdot (\mu_x \nabla_x V_x(\mu_y, x)) + \beta^{-1} \Delta_x \mu_x, \quad \mu_x(0) = \frac{1}{n}\sum_{i=1}^n \delta_{x_0^i} \\ \partial_t \mu_y = -\nabla_y \cdot (\mu_y \nabla_y V_y(\mu_x, y)) + \beta^{-1} \Delta_y \mu_y, \quad \mu_y(0) = \frac{1}{n}\sum_{i=1}^n \delta_{y_0^i} \end{cases} \tag{2.7}$$

The derivation of (2.7) is provided in Theorem A.19. It is a system of coupled nonlinear Fokker-Planck equations, that are the Kolmogorov forward equations of the SDE (2.4). They correspond to the IWGF of the entropy-regularized loss $\mathcal{L}_\beta(\mu_x, \mu_y) \triangleq \mathcal{L}(\mu_x, \mu_y) + \beta^{-1}(H(\mu_y) - H(\mu_x))$.

Finally, if $\{x^i, y^i, w_x^i, w_y^i\}_{i\in[1:n]}$ solve (2.5), then $\mu_x(t) = \sum_{i=1}^n w_{x,t}^i \delta_{x_t^i}$, $\mu_y(t) = \sum_{i=1}^n w_{y,t}^i \delta_{y_t^i}$ solve the

Interacting Wasserstein-Fisher-Rao Gradient Flow (IWFRGF) of $\mathcal{L}$:

$$\begin{cases} \partial_t \mu_x &= \gamma \nabla_x \cdot (\mu_x \nabla_x V_x(\mu_y, x)) - \alpha \mu_x (V_x(\mu_y, x) - \mathcal{L}(\mu_x, \mu_y)),\ \mu_x(0) = \sum_{i=1}^n w_{x,0}^i \delta_{x_0^i}, \\ \partial_t \mu_y &= -\gamma \nabla_y \cdot (\mu_y \nabla_y V_y(\mu_x, y)) + \alpha \mu_y (V_y(\mu_x, y) - \mathcal{L}(\mu_x, \mu_y)),\ \mu_y(0) = \sum_{i=1}^n w_{y,0}^i \delta_{y_0^i}. \end{cases} \tag{2.8}$$

The derivation of (2.8) is provided in section A.1 and Theorem A.22. The Wasserstein-Fisher-Rao or Hellinger-Kantorovich metric [Chizat et al. 2015; Kondratyev et al. 2016; Gallouët and Monsaingeon 2016] is a metric on the probability space $\mathcal{M}_+(\mathcal{X})$ induced by a lifting to the space $\mathcal{P}(\mathcal{X} \times \mathbb{R}^+)$ of the form $\nu \mapsto \mu = \int_{\mathbb{R}^+} w \, d\nu(\cdot, w)$. If we keep ν_y fixed, the first equation in (2.8) is a Wasserstein-Fisher-Rao gradient flow (slightly modified by the term $\alpha \mu_x \mathcal{L}(\mu_x, \mu_y)$ to constrain μ_x in $\mathcal{P}(\mathcal{X})$). The term $-\alpha \mu_x (V_x(\mu_y, x) - \mathcal{L}(\mu_x, \mu_y))$, which also arises in entropic mirror descent, allow mass to 'teleport' from bad strategies to better ones with finite cost by moving along the weight coordinate. Wasserstein-Fisher-Rao gradient flows have been used by Chizat [2019]; Rotskoff et al. [2019]; Liero et al. [2018] in the context of optimization.

Initialization of (2.6), (2.7) and (2.8) may be done with the measures $\mu_{x,0}$ and $\mu_{y,0}$ from which $\{x_0^i\}, \{y_0^i\}$ are sampled, in which case the measures $\mu_x(t)$ and $\mu_y(t)$ are not discrete and follow the *mean-field* dynamics. In subsection 2.4.3 we link the dynamics starting from discrete realizations to the mean-field dynamics.

2.4 Convergence analysis

We establish convergence results for the entropy-regularized dynamics and the WFR dynamics.

2.4.1 Convergence of the entropy-regularized Wasserstein dynamics

The following theorem characterizes the stationary points of the entropy-regularized dynamics.

Theorem 2.2. *Suppose that Assumption 1 holds, that $\ell \in C^2(\mathcal{X} \times \mathcal{Y})$ and that the initial measures $\mu_{x,0}, \mu_{y,0}$ have densities in $L^1(\mathcal{X}), L^1(\mathcal{Y})$. If a solution $(\mu_x(t), \mu_y(t))$ of the ERIWGF (2.7) converges in time, it must converge to the point $(\hat{\mu}_x, \hat{\mu}_y)$ which is the unique fixed point of the problem*

$$\rho_x(x) = \frac{1}{Z_x} e^{-\beta \int \ell(x,y)\, d\mu_y(y)}, \quad \rho_y(y) = \frac{1}{Z_y} e^{\beta \int \ell(x,y)\, d\mu_x(x)}. \tag{2.9}$$

$(\hat{\mu}_x, \hat{\mu}_y)$ is an ϵ-Nash equilibrium of the game given by $\mathcal{L}$ when $\beta \geq \frac{4}{\epsilon} \log\left(2\frac{1-V_\delta}{V_\delta}(2K_\ell/\epsilon - 1)\right)$, where $K_\ell := \max_{x,y} \ell(x,y) - \min_{x,y} \ell(x,y)$ is the length of the range of ℓ, $\delta := \epsilon/(2Lip(\ell))$ and V_δ is a lower bound on the volume of a ball of radius δ in $\mathcal{X}, \mathcal{Y}$.

The proof is in section A.3. Theorem 2.2 characterizes the stationary points of the ERIWGF but does not provide a guarantee of convergence in time. It implies that if the dynamics (2.7) converges in time, the limit will be an ϵ-Nash equilibrium of $\mathcal{L}$, with $\epsilon = \tilde{O}(1/\beta)$ (disregarding log factors). The dynamics (2.7) correspond to a McKean-Vlasov process on the joint probability measure $\mu_x \times \mu_y$. While convergence to stationary solutions of such processes have been studied in the Euclidean case [Eberle et al. 2019]l, their results would only guarantee convergence for temperatures $\beta^{-1} \gtrsim Lip(\ell)$ in our setup, which is not strong enough to certify convergence to arbitrary ϵ-NE.

There is a trade-off between setting a low temperature β^{-1}, which yields an ϵ-Nash equilibrium with small ϵ but possibly slow or no convergence, and setting a high temperature, which has the opposite effect. Linear potential Fokker-Planck equations (that we recover when both players are decoupled) indeed converge exponentially with rate $e^{-\lambda_\beta t}$ for all β, with λ_β decreasing exponentially with β for nonconvex potentials [Markowich and Villani 1999, sec. 5]. Entropic regularization also biases the dynamics towards measures with full support and hence precludes convergence to sparse equilibria even if they exist. This problem does not arise in the WFR dynamics.

2.4.2 Analysis of the Wasserstein-Fisher-Rao dynamics

Theorem 2.3 states that, at a certain time t_0, the time averaged measures of the solution (v_x, v_y) of (2.8) are an ϵ-MNE, where ϵ can be made arbitrarily small by adjusting the constants γ, α of the dynamics. We define $\bar{v}_x(t) = \frac{1}{t}\int_0^t v_x(s)\, ds$ and $\bar{v}_y(t) = \frac{1}{t}\int_0^t v_y(s)\, ds$, where v_x and v_y are solutions of (2.8).

Theorem 2.3. *Let $\epsilon > 0$ arbitrary. Suppose that $v_{x,0}, v_{y,0}$ are such that their Radon-Nikodym derivatives with respect to the Borel measures of $\mathcal{X}, \mathcal{Y}$ are lower-bounded by $e^{-K'_x}, e^{-K'_y}$ respectively. For any $\delta \in (0, 1/2)$, there exists a constant $C_{\delta,\mathcal{X},\mathcal{Y},K'_x,K'_y} > 0$ depending on the dimensions of $\mathcal{X}, \mathcal{Y}$, their curvatures and K'_x, K'_y, such that if $\gamma/\alpha < 1$, $\frac{\gamma}{\alpha} \leq \left(\epsilon / C_{\delta,\mathcal{X},\mathcal{Y},K'_x,K'_y}\right)^{\frac{2}{1-\delta}}$*

$$NI(\bar{v}_x(t_0), \bar{v}_y(t_0)) \leq \epsilon \quad \textit{where} \quad t_0 = (\alpha\gamma)^{-1/2}.$$

The proof (section A.4) builds on the convergence properties of continuous-time mirror descent and closely follows the proof of Theorem 3.8 from Chizat [2019]. We explicit the dependency of $C_{\delta,\mathcal{X},\mathcal{Y},K'_x,K'_y}$ on the dimensions of the manifolds and the properties of the loss ℓ. Notice that Theorem 2.3 ensures convergence towards an ϵ-Nash equilibrium of the non-regularized game. Following Chizat [2019], it is possible to replace the regularity assumption on the initial measures $v_{x,0}, v_{y,0}$ by a singular initialisation, at the expense of using $O(\exp(d))$ particles. This result is not a convergence result for the measures, but rather on the value of the NI error. Notice that it involves time-averaging and a finite horizon. Similar results are common for mirror descent in convex games [Juditsky et al. 2011], albeit in the discrete-time setting.

Theorem 2.3 does not capture the benefits of transport, as it regards it as a perturbation of mirror descent (which corresponds to $\gamma = 0$). When targetting a small error ϵ, we need to set $\gamma \ll \alpha$ because of the bound on γ/α. In this case, mirror descent is the main driver of the dynamics. However, it is seen empirically that taking much higher ratios γ/α (i.e. increasing the importance

of the transport term) results in better performance. A satisfying explanation of this phenomenon is still sought after in the simpler optimization setting [Chizat 2019].

2.4.3 Convergence to mean-field

The following theorem (proof in section A.6) links the empirical measures of the systems (2.4), (2.5) to the solutions of the mean field dynamics (2.7) and (2.8) respectively. It can be seen as a law of large numbers. It shows that by Theorem 2.4, algorithm 1 and algorithm 2 approximate the mean-field dynamics studied in subsection 2.4.1 and subsection 2.4.2.

Theorem 2.4. *(i) Let* $\mu_x^n = \frac{1}{n}\sum_{i=1}^n \delta_{X^{(i)}} \in C([0,T],\mathcal{P}(\mathcal{X}))$, $\mu_y^n = \frac{1}{n}\sum_{i=1}^n \delta_{Y^{(i)}} \in C([0,T],\mathcal{P}(\mathcal{Y}))$ *be the empirical measures of a solution of* (2.4) *up to an arbitrary time* T. *Let* $\mu_x \in C([0,T],\mathcal{P}(\mathcal{X}))$, $\mu_y \in C([0,T],\mathcal{P}(\mathcal{Y}))$ *be a solution of the ERIWGF* (2.7) *with mean-field initial conditions* $\mu_x(0) = \mu_{x,0}$, $\mu_y(0) = \mu_{y,0}$. *Then,*

$$\mathbb{E}[\mathcal{W}_2^2(\mu_{x,t}^n, \mu_{x,t}) + \mathcal{W}_2^2(\mu_{y,t}^n, \mu_{y,t})] \xrightarrow{n\to\infty} 0, \quad \mathbb{E}[|NI(\mu_{x,t}^n, \mu_{y,t}^n) - NI(\mu_{x,t}, \mu_{y,t})|] \xrightarrow{n\to\infty} 0,$$

uniformly over $t \in [0,T]$. *NI is the Nikaido-Isoda error defined in* (2.2).
(ii) Let $\nu_x^n = \sum_{i=1}^n w_{x,t}^i \delta_{X^{(i)}} \in C([0,T],\mathcal{P}(\mathcal{X}))$, $\mu_y^n = \sum_{i=1}^n w_{y,t}^i \delta_{Y^{(i)}} \in C([0,T],\mathcal{P}(\mathcal{Y}))$ *be the (projected) empirical measures of a solution of* (2.5) *up to an arbitrary time* T. *Let* $\nu_x \in C([0,T],\mathcal{P}(\mathcal{X}))$, $\nu_y \in C([0,T],\mathcal{P}(\mathcal{Y}))$ *be a solution of* (2.8) *with mean-field initial conditions* $\mu_x(0) = \mu_{x,0}$, $\mu_y(0) = \mu_{y,0}$. *Then,*

$$\mathbb{E}[\mathcal{W}_2^2(\nu_{x,t}^n, \nu_{x,t}) + \mathcal{W}_2^2(\nu_{y,t}^n, \nu_{y,t})] \xrightarrow{n\to\infty} 0, \quad \mathbb{E}[|NI(\bar{\nu}_{x,t}^n, \bar{\nu}_{y,t}^n) - NI(\bar{\nu}_{x,t}, \bar{\nu}_{y,t})|] \xrightarrow{n\to\infty} 0,$$

uniformly over $t \in [0,T]$. $\bar{\nu}_{x,t}, \bar{\nu}_{y,t}, \bar{\nu}_{x,t}^n, \bar{\nu}_{y,t}^n$ *are the time-averaged measures, as in Theorem 2.3.*

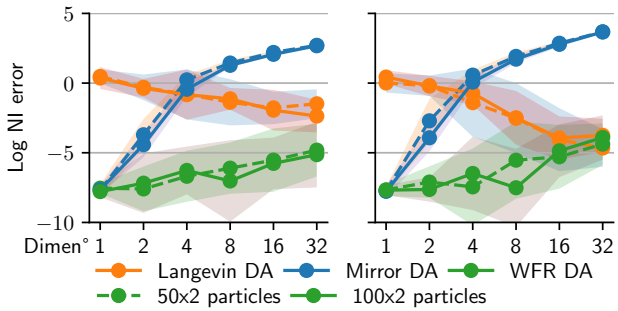


Figure 2.1: Nikaido-Isoida errors for L-DA, WFR-DA and mirror descent, as a function of the problem dimension, for a nonconvex loss ℓ_a (left) and convex loss ℓ_b (right). L-DA and WFR-DA outperforms mirror descent for large dimensions. Values averaged over 20 runs after 30000 iterations. Error bars show standard deviation across runs.

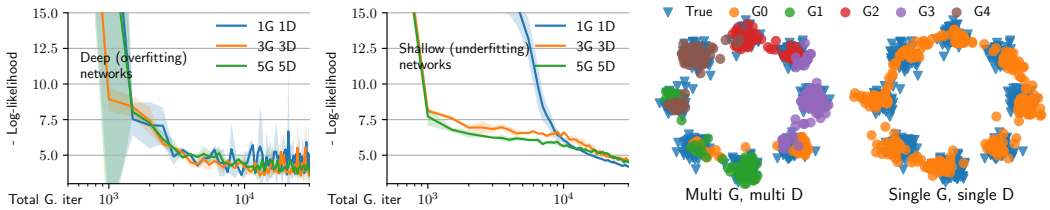


Figure 2.2: Training mixtures of GANs over a synthetic mixture of Gaussians in 2D. WFR-DA converges faster with models with low number of parameters, and similar performance with over-parametrized models. Mixtures naturally perform a form of clustering of the data. Errors bars show variance across 5 runs.

2.5 Numerical Experiments

We show that WFR and Langevin dynamics outperform mirror descent in high dimension, on synthetic games. We then show the interests of using WFR-DA for training GANs. Code has been made available for reproducibility.

Figure 2.3: Training mixtures of GANs over CIFAR10. We compare the algorithm that updates the mixture weights and parameters (WFR-DA flow) with the algorithm that only updates parameters (W-DA flow). Using several discriminators and a WFR-DA flow brings more stable convergence. Each generator tends to specialize in a type of images. Errors bars show variance across 5 runs.

2.5.1 Polynomial games on spheres

We study two different games with losses $\ell_a, \ell_b : \mathcal{S}^{d-1} \times \mathcal{S}^{d-1} \to \mathbb{R}$ of the form

$$\ell_a(x, y) = x^\top A_0 x + x^\top A_1 y + y^\top A_2 y + y^\top A_3 (x^2) + a_0^\top x + a_1^\top y$$
$$\ell_b(x, y) = x^\top A_0^\top A_0 x + x^\top A_1 y + y^\top A_2^\top A_2 y + a_0^\top x + a_1^\top y,$$

where $A_0, A_1, A_2, A_3, a_0, a_1$ are matrices and vectors with components sampled from a normal distribution $\mathcal{N}(0, 1)$, and x^2 is the vector given by component-wise multiplication of x. ℓ_b is a convex loss on the sphere, while ℓ_a is not. We run Langevin Descent-Ascent (updates of positions) and WFR Descent-Ascent (updates of weights and positions), and compare it with mirror descent (updates of weights).We note that the computation of the NI error (2.2) entails solving two optimization problems on measures, or equivalently in parameter space. We solve each of them by performing 2000 gradient acsent runs with random uniform initialization and selecting the highed minimum final value. This gives a lower bound on the NI error which is precise enough for our purposes. We perform time averaging on the weights of mirror descent and WFR-DA, but not on the positions of WFR-DA because that would incur an $O(t)$ overhead on memory.

Results. Wirror descent performs like WFR-DA in low dimensions, but suffers strongly from the curse of dimensionality (Figure 2.1). On the other hand, algorithms that incorporate a transport term keep performing well in high dimensions. In particular, WFR-DA is consistently the algorithm with lowest NI error. Notice that the errors in the $n = 50$ and $n = 100$ plots do not differ much, confirming that we reach a mean-field regime.

2.5.2 Training GAN mixtures

We now use WFR-DA to train mixtures of generator networks. We consider the Wasserstein-GAN [Arjovsky et al. 2017] setting. We seek to approximate a distribution $\mathcal{P}_{\text{data}}$ with a distribution

$\mathcal{G}_x$, defined as the push-forward of a noise distribution $\mathcal{N}(0, I)$ by a neural-network g_x. The discrepancy between $\mathcal{P}_{\text{data}}$ and $\mathcal{G}_x$ is estimated by a neural-network discriminator f_y, leading to the problem

$$\min_x \max_y \ell(x, y) \triangleq \mathbb{E}_{a\sim p_{\text{data}}}[f_y(a)] - \mathbb{E}_{\varepsilon\sim\mathcal{N}(0,I)}[f_y(g_x(\varepsilon))].$$

We lift this problem in the space of distributions over the parameters x and y (see subsection A.7.4), that we represent through weighted discrete distributions of $\sum_{i=1}^p w_x^{(i)}\delta_{x^{(i)}}$ and $\sum_{j=1}^q w_y^{(j)}\delta_{y^{(j)}}$. We solve

$$\min_{x^{(i)}, w_x\in\triangle_p} \max_{y^{(j)}, w_y\in\triangle^q} \sum_{i=1}^p \sum_{j=1}^q w_x^{(i)} w_y^{(j)} \ell(x^{(i)}, y^{(j)}) \,,$$

using algorithm 2, where $\triangle^q$ is the q-dimensional simplex. The optimal generation strategy corresponding to an equilibrium point $(x^{(i)})_i, w_x, (y^{(j)})_j, w_y$ is then to randomly select a generator g_{x_I} with I sampled among $[n]$ with probability $w_x^{(i)}$, and use it to generate $g_{x_I}(\varepsilon)$, with $\varepsilon \sim \mathcal{N}(0, I)$. Training mixtures of generators has been proposed by Ghosh et al. [2018], with a tweaked discriminator loss. Our formulation only involves a lifting in the space of measures, and uses a new training algorithm.

Results on 2D GMMs. We first set $\mathcal{P}_{\text{data}}$ to be an 8-mode mixture of Gaussians in two dimensions. We use the original W-GAN loss, with weight cropping for the discriminators $(f_{y^{(j)}})_j$. We measure the interest of using mixtures when a single generator $g_{x^{(i)}}$ cannot fit $\mathcal{P}_{\text{data}}$ (single-layer MLP), and when it can (4-layer MLP). We report results in Figure 2.2, measuring the log likelihood of $\mathcal{G}_x$ for the GMM during training. The WFR dynamic is stable even with few particles. When training under-parametrized generators, using mixtures permits faster convergence (in terms of generator updates). In the over-parametrized setting, training a single generator or a mixture of generators perform similarly. WFR-DA is thus useful to train mixtures of simple generators. In this setting, each simple generator identifies modes in the training data, doing data clustering at no cost (Figure 2.2 right).

Results on real data. We train a mixture of ResNet generators on CIFAR10 and MNIST. We replace the position updates in algorithm 2 by extrapolated Adam steps [Gidel et al. 2019] to achieve faster convergence, and perform grid search over generator and discriminators learning rates. Convergence curves for the best learning rates are displayed in Figure 2.3 right, measuring test FID [Heusel et al. 2017]. With a sufficient number of generators and discriminators ($G > 5, D > 2$), the model trains as fast as a normal GAN. WFR-DA is thus stable and efficient even with a reasonable number of particles. Using the discretized WFR versus the Wasserstein flow provides a slight improvement over updating parameters only. As with GMMs, each generator trained with WFR-DA becomes specialised in generating a fraction of the target data, thereby identifying clusters. Those could be used for unsupervised conditional generation of images.

2.6 Conclusions and future work

We have explored non-convex-non-concave, high-dimensional games from the perspective of optimal transport. As with non-convex optimization, framing the problem in terms of measures provides geometric benefits, at the expense of moving into non-Euclidean metric spaces over measures. Our theoretical results establish approximate mean-field convergence for two setups: Langevin Descent-Ascent and WFR D-A, and directly applies to GANs, for mixtures of generators and discriminators.

Despite the positive convergence guarantees our results are qualitative in nature, i.e. without rates. In the entropic case, the unfavorable tradeoff between temperature and convergence of the associated McKean-Vlasov scheme deserves further study, maybe through log-Sobolev-type inequalities [Markowich and Villani 1999]. In the WFR case, we lack a local convergence analysis explaining the benefits of transport observed empirically, perhaps leveraging sharpness Polyak-Łojasiewicz results such as those in [Chizat 2019] or [Sanjabi et al. 2018]. Finally, in our GAN formulation, each generator is associated to a single particle in a high-dimensional product space

of all network parameters, which is not scalable to large population sizes that would approximate their mean-field limit. A natural question is to understand to what extent our framework could be combined with specific choices of architecture, as recently studied in [Lei et al. 2019].

Broader impact

We study algorithms designed to find equilibria in games, provide theoretical guarantees of convergence and test their performance empirically. Among other applications, our results give insight into training algorithms for generative adversarial networks (GANs), which are useful for many relevant tasks such as image generation, image-to-image or text-to-image translation and video prediction. As always, we note that machine learning improvements like ours come in the form of "building machines to do X better". For a sufficiently malicious or ill-informed choice of X, such as surveillance or recidivism prediction, almost any progress in machine learning might indirectly lead to a negative outcome, and our work is not excluded from that.

3 Multisample Flow Matching

3.1 Introduction

Deep generative models offer an attractive family of paradigms that can approximate a data distribution and produce high quality samples, with impressive results in recent years [Ramesh et al. 2022; Saharia et al. 2022; Gafni et al. 2022]. In particular, these works have made use of simulation-free training methods for diffusion models [Ho et al. 2020; Song et al. 2021b]. A number of works have also adopted and generalized these simulation-free methods [Lipman et al. 2023; Albergo and Vanden-Eijnden 2023; Liu et al. 2022; Neklyudov et al. 2022] for continuous normalizing flows (CNF; Chen et al. [2018]), a family of continuous-time deep generative models that parameterizes a vector field which flows noise samples into data samples.

Recently, Lipman et al. [2023] proposed *Flow Matching* (FM), a method to train CNFs based on constructing explicit *conditional probability paths* between the noise distribution (at time $t = 0$) and each data sample (at time $t = 1$). Furthermore, they showed that these conditional probability paths can be taken to be the optimal transport path when the noise distribution is a standard Gaussian, a typical assumption in generative modeling. However, this does not imply that the *marginal probability path* (marginalized over the data distribution) is anywhere close to the optimal transport path between the noise and data distributions.

Most existing works, including diffusion models and Flow Matching, have only considered conditional sample paths where the endpoints (a noise sample and a data sample) are sampled indepen-

Figure 3.1: Multisample Flow Matching trained with batch optimal couplings produces more consistent samples across varying NFEs. Note that both flows on each row start from the same noise sample.

dently. However, this results in non-zero gradient variances even at convergence, slow training times, and in particular limits the design of probability paths. In turn, it becomes difficult to create paths that are fast to simulate, a desirable property for both likelihood evaluation and sampling.

Contributions: We present a tractable instance of Flow Matching with joint distributions, which we call *Multisample Flow Matching*. Our proposed method generalizes the construction of probability paths by considering non-independent couplings of k-sample empirical distributions. Among other theoretical results, we show that if an appropriate optimal transport (OT) inspired coupling is chosen, then sample paths become straight as the batch size $k \to \infty$, leading to more efficient simulation. In practice, we observe both improved sample quality on ImageNet using adaptive ODE solvers and using simple Euler discretizations with a low budget number of function evaluations. Empirically, we find that on ImageNet, we can *reduce the required sampling cost*

by 30% to 60% for achieving a low Fréchet Inception Distance (FID) compared to a baseline Flow Matching model, while introducing only 4% more training time. This improvement in sample efficiency comes at no degradation in performance, *e.g.* log-likelihood and sample quality.

Within the deep generative modeling paradigm, this allows us to regularize towards the optimal vector field in a *completely simulation-free manner* (unlike *e.g.* Finlay et al. [2020b]; Liu et al. [2022]), and avoids adversarial formulations (unlike *e.g.* Makkuva et al. [2020]; Albergo and Vanden-Eijnden [2023]). In particular, we are the first work to be able to make use of solutions from optimal solutions on minibatches while preserving the correct marginal distributions, whereas prior works would only fit to the barycentric average (see detailed discussion in subsection 3.5.1). Beyond generative modeling, we also show how our method can be seen as a new way to compute approximately optimal transport maps between arbitrary distributions in settings where the cost function is completely unknown and only minibatch optimal transport solutions are provided.

3.2 Preliminaries

3.2.1 Continuous Normalizing Flow

Let $\mathbb{R}^d$ denote the data space with data points $x = (x^1, \ldots, x^d) \in \mathbb{R}^d$. Two important objects we use in this paper are: the *probability path* $p_t : \mathbb{R}^d \to \mathbb{R}_{>0}$, which is a time dependent (for $t \in [0, 1]$) probability density function, *i.e.*, $\int p_t(x)dx = 1$, and a *time-dependent vector field*, $u_t : [0, 1] \times \mathbb{R}^d \to \mathbb{R}^d$. A vector field u_t constructs a time-dependent diffeomorphic map, called a *flow*, $\psi : [0, 1] \times \mathbb{R}^d \to \mathbb{R}^d$, defined via the ordinary differential equation (ODE):

$$\frac{d}{dt}\psi_t(x_0) = u_t(\psi_t(x_0)), \quad \psi_0(x_0) = x_0 . \tag{3.1}$$

To create a deep generative model, Chen et al. [2018] suggested modeling the vector field u_t with a neural network, leading to a deep parametric model of the flow ψ_t, referred to as a *Continuous Normalizing Flow* (CNF). A CNF is often used to transform a density p_0 to a different one, p_1, via the push-forward equation

$$p_t(x) = [\psi_t]_\sharp p_0(x) = p_0(\psi_t^{-1}(x)) \left| \det \left[\frac{\partial \psi_t^{-1}}{\partial x}(x) \right] \right|, \tag{3.2}$$

where the second equality defines the push-forward (or change of variables) operator $\sharp$. A vector field u_t is said to *generate* a probability path p_t if its flow ψ_t satisfies (3.2).

3.2.2 Flow Matching

A simple simulation-free method for training CNFs is the *Flow Matching* algorithm [Lipman et al. 2023], which regresses onto an (implicitly-defined) target vector field that generates the desired probability density path p_t. Given two marginal distributions $q_0(x_0)$ and $q_1(x_1)$ for which we would like to learn a CNF to transport between, Flow Matching seeks to optimize the simple regression objective,

$$\mathbb{E}_{t, p_t(x)} \left\| v_t(x; \theta) - u_t(x) \right\|^2, \tag{3.3}$$

where $v_t(x; \theta)$ is the parametric vector field for the CNF, and $u_t(x)$ is a vector field that generates a probability path p_t under the two marginal constraints that $p_{t=0} = q_0$ and $p_{t=1} = q_1$. While Equation 3.3 is the ideal objective function to optimize, not knowing (p_t, u_t) makes this computationally intractable.

Lipman et al. [2023] proposed a tractable method of optimizing (3.3), which first defines *conditional* probability paths and vector fields, such that when marginalized over $q_0(x_0)$ and $q_1(x_1)$, provide both $p_t(x)$ and $u_t(x)$. When targeted towards generative modeling, $q_0(x_0)$ is a simple

noise distribution and easy to directly enforce, leading to a one-sided construction:

$$p_t(x) = \int p_t(x|x_1)q_1(x_1)\ dx_1$$
$$u_t(x) = \int u_t(x|x_1)\frac{p_t(x|x_1)q_1(x_1)}{p_t(x)}\ dx_1, \tag{3.4}$$

where the conditional probability path is chosen such that

$$p_{t=0}(x|x_1) = q_0(x) \quad \text{and} \quad p_{t=1}(x|x_1) = \delta(x - x_1), \tag{3.5}$$

where $\delta(x - a)$ is a Dirac mass centered at $a \in \mathbb{R}^d$. By construction, $p_t(x|x_1)$ now satisfies both marginal constraints.

Lipman et al. [2023] shows that if $u_t(x|x_1)$ generates $p_t(x|x_1)$, then the marginalized $u_t(x)$ generates $p_t(x)$, and furthermore, one can train using the much simpler objective of *Conditional Flow Matching* (CFM):

$$\mathbb{E}_{t,q_1(x_1),p_t(x|x_1)} \left\| v_t(x;\theta) - u_t(x_t|x_1) \right\|^2, \tag{3.6}$$

with $x_t = \psi_t(x_0|x_1)$; see 3.2.2.1 for more details. Note that this objective has the same gradient with respect to the model parameters θ as Eq. (3.3) [Lipman et al. 2023, Theorem 2].

3.2.2.1 Conditional OT (CondOT) path

One particular choice of conditional path $p_t(x|x_1)$ is to use the flow that corresponds to the optimal transport displacement interpolant [McCann 1997] when $q_0(x_0)$ is the standard Gaussian, a common convention in generative modeling. The vector field that corresponds to this is

$$u_t(x_t|x_1) = \frac{x_1 - x}{1 - t}. \tag{3.7}$$

Using this conditional vector field in (3.1), this gives the conditional flow

$$x_t = \psi_t(x_0|x_1) = (1-t)x_0 + tx_1\,. \tag{3.8}$$

Substituting (3.8) into (3.7), one can also express the value of this vector field using a simpler expression,

$$u_t(x_t|x_1) = x_1 - x_0\,.$$

It is evident that this results in conditional flows that *(i)* tranports all points x_0 from $t = 0$ to x_1 at exactly $t = 1$ and *(ii)* are straight paths between the samples x_0 and x_1. This particular case of straight paths was also studied by Liu et al. [2022] and Albergo and Vanden-Eijnden [2023], where the conditional flow (3.8) is referred to as a stochastic interpolant. Lipman et al. [2023] additionally showed that the conditional construction can be applied to a large class of Gaussian conditional probability paths, namely when $p_t(x|x_1) = \mathcal{N}(x|\mu_t(x_1), \sigma_t(x_1)^2 I)$. This family of probability paths encompasses most prior diffusion models where probability paths are induced by simple diffusion processes with linear drift and constant diffusion (*e.g.* Ho et al. [2020]; Song et al. [2021b]). However, existing works mostly consider settings where $q_0(x_0)$ and $q_1(x_1)$ are sampled independently when computing training objectives such as (3.6).

3.2.3 Optimal Transport: Static & Dynamic

Optimal transport generally considers methodologies that define some notion of distance on the space of probability measures [Villani 2008, 2003; Santambrogio 2015]. Letting $\mathcal{P}(\mathbb{R}^d)$ be the space of probability measures over $\mathbb{R}^d$, we define the Wasserstein distance with respect to a cost function $c : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}_+$ between two measures $q_0, q_1 \in \mathcal{P}(\mathbb{R}^d)$ as [Kantorovitch 1942]

$$W_c(q_0, q_1) \coloneqq \min_{q \in \Gamma(q_0, q_1)} \mathbb{E}_{q(x_0, x_1)}[c(x_0, x_1)]\,, \tag{3.9}$$

where $\Gamma(q_0, q_1)$ is the set of joint measures with left marginal equal to q_0 and right marginal equal to q_1, called the set of *couplings*. The minimizer to Equation 3.9 is called the optimal coupling, which we denote by q_c^*. In the case where $c(x_0, x_1) := \|x_0 - x_1\|^2$, the squared-Euclidean distance, Equation 3.9 amounts to the (squared) 2-Wasserstein distance $W_2^2(q_0, q_1)$, and we simply write the optimal transport plan as q^*.

Considering again the squared-Euclidean cost, in the case where q_0 exhibits a density over $\mathbb{R}^d$ (e.g. if q_0 is the standard normal distribution), Benamou and Brenier [2000] states that $W_2^2(q_0, q_1)$ can be equivalently expressed as a *dynamic* formulation,

$$W_2^2(q_0, q_1) = \min_{p_t, u_t} \int_0^1 \int_{\mathbb{R}^d} \|u_t(x)\|^2 \, p_t(x) \mathrm{d}x_0 \mathrm{d}t. \tag{3.10}$$

where u_t generates p_t, and p_t satisfies boundary conditions $p_{t=0} = q_0$ and $p_{t=1} = q_1$. The optimality condition ensures that sample paths x_t are straight lines, i.e. minimize the length of the path, and leads to paths that are much easier to simulate. Some prior approaches have sought to regularize the model using this optimality objective (*e.g.* Tong et al. [2020]; Finlay et al. [2020b]). In contrast, instead of directly minimizing (3.10), we will discuss an approach based on using solutions of the optimal coupling q^* on minibatch problems, while leaving the marginal constraints intact.

3.3 Flow Matching with Joint Distributions

While Conditional Flow Matching in (3.6) leads to an unbiased gradient estimator for the Flow Matching objective, it was designed with independently sampled x_0 and x_1 in mind. We generalize the framework from Subsection 3.2.2 to a construction that uses arbitrary joint distributions of $q(x_0, x_1)$ which satisfy the correct marginal constraints, *i.e.*

$$\int q(x_0, x_1) \mathrm{d}x_1 = q_0(x_0) \,, \int q(x_0, x_1) \mathrm{d}x_0 = q_1(x_1). \tag{3.11}$$

We will show in Subsection 3.4 that this can potentially lead to lower gradient variance during training and allow us to design more optimal marginal vector fields $u_t(x)$ with desirable properties such as improved sample efficiency.

Building on top of Flow Matching, we propose modifying the conditional probability path construction (3.5) so that at $t = 0$, we define

$$p_{t=0}(x_0|x_1) = q(x_0|x_1).$$

where $q(x_0|x_1)$ is the conditional distribution $\frac{q(x_0,x_1)}{q_1(x_1)}$. Using this construction, we still satisfy the marginal constraint,

$$p_0(x) = \int p_0(x|x_1)q_1(x_1)dx_1 = \int q(x,x_1)dx_1 = q_0(x)$$

i.e. $p_{t=0}(x) = \int q(x,x_1)dx_1 = q_0(x)$ by the assumption made in (3.11). Then similar to Chen and Lipman [2023], we note that the conditional probability path $p_t(x|x_1)$ *need not be explicitly formulated* for training, and that only an appropriate conditional vector field $u_t(x|x_1)$ needs to be chosen such that all points arrive at x_1 at $t = 1$, which ensures $p_{t=1}(x|x_1) = \delta(x - x_1)$. As such, we can make use of the same conditional vector field as prior works, *e.g.* the choice in **??**.

We then propose the **Joint CFM** objective as

$$\mathcal{L}_{\text{JCFM}} = \mathbb{E}_{t,q(x_0,x_1)} \left\| v_t(x_t;\theta) - u_t(x_t|x_1) \right\|^2, \tag{3.12}$$

where $x_t = \psi_t(x_0|x_1)$ is the conditional flow. Training only involves sampling from $q(x_0,x_1)$ and does not require explicitly knowing the densities of $q(x_0,x_1)$ or $p_t(x|x_1)$. Note that Equation (3.12) reduces to the original CFM objective (3.6) when $q(x_0,x_1) = q_0(x_0)q_1(x_1)$.

A quick sanity check shows that this objective can be used with any choice of joint distribution $q(x_0,x_1)$.

Lemma 3.1. *The optimal vector field $v_t(\cdot;\theta)$ in (3.12), which is the marginal vector field u_t, maps between the marginal distributions $q_0(x_0)$ and $q_1(x_1)$.*

In the remainder of the section, we highlight some motivations for using joint distributions $q(x_0, x_1)$ that are different from the independent distribution $q_0(x_0)q_1(x_1)$.

VARIANCE REDUCTION Choosing a good joint distribution can be seen as a way to reduce the variance of the gradient estimate, which improves and speeds up training. We develop the gradient covariance at a fixed x and t, and bound its total variance:

Lemma 3.2. *The total variance (i.e. the trace of the covariance) of the gradient at a fixed x and t is bounded as:*

$$\begin{aligned}\sigma_{t,x}^2 &= \mathrm{Tr}\left[\mathrm{Cov}_{p_t(x_1|x)}\left(\nabla_\theta \left\|v_t(x;\theta) - u_t(x|x_1)\right\|^2\right)\right] \\ &\leq \left\|\nabla_\theta v_t(x;\theta)\right\|^2 \mathbb{E}_{p_t(x_1|x)}\left\|u_t(x) - u_t(x|x_1)\right\|^2\end{aligned} \tag{3.13}$$

Then $\mathbb{E}_{t,p_t(x)}[\sigma_{t,x}^2]$ is bounded above by:

$$\max_{t,x} \left\|\nabla_\theta v_t(x;\theta)\right\|^2 \times \mathcal{L}_{\mathrm{JCFM}} \tag{3.14}$$

This proves that $\mathbb{E}_{t,p_t(x)}[\sigma_{t,x}^2]$, which is the average gradient variance at fixed x and t, is upper bounded in terms of the Joint CFM objective. That means that minimizing the Joint CFM objective help in decreasing $\mathbb{E}_{t,p_t(x)}[\sigma_{t,x}^2]$. Note also that $\mathbb{E}_{t,p_t(x)}[\sigma_{t,x}^2]$ is not the gradient variance and is always smaller, as it does not account for variability over x and t, but it is a good proxy for it. The proof is in App. B.4.2.

Sampling x_0 and x_1 independently generally cannot achieve value zero for $\mathbb{E}_{t,p_t(x)}[\sigma_{t,x}^2]$ *even at the optimum*, since there are an infinite number of pairs (x_0, x_1) whose conditional path crosses

any particular x at a time t. As shown in (3.14), having a low optimal value for the Joint CFM objective is a good proxy for low gradient variance and hence a desirable property for choosing a joint distribution $q(x_0, x_1)$. In section 3.4, we show that certain joint distributions have optimal Joint CFM values close to zero.

Straight flows Ideally, the flow ψ_t of the marginal vector field u_t (and of the learned v_θ by extension) should be close to a straight line. The reason is that ODEs with straight trajectories can be solved with high accuracy using fewer steps (i.e. function evaluations), which speeds up sample generation. The quantity

$$S = \mathbb{E}_{t,q_0(x_0)}\left[\|u_t(\psi_t(x_0))\|^2 - \|\psi_1(x_0) - x_0\|^2\right], \tag{3.15}$$

which we call the *straightness* of the flow and was also studied by Liu [2022], measures how straight the trajectories are. Namely, we can rewrite it as

$$S = \mathbb{E}_{t,q_0(x_0)}\left[\|u_t(\psi_t(x_0)) - \mathbb{E}_{t'}\left[u_{t'}(\psi_{t'}(x_0))\right]\|^2\right], \tag{3.16}$$

which shows that $S \geq 0$ and only zero if $u_t(\psi_t(x_0))$ is constant along t, which is equivalent to $\psi_t(x_0)$ being a straight line.

When x_0 and x_1 are sampled independently, the straightness is in general far from zero. This can be seen in the CondOT plots in Figure 3.2 (right); if flows were close to straight lines, samples generated with one function evaluation (NFE=1) would be of high quality. In section 3.4, we show that for certain joint distributions, the straightness of the flow is close to zero.

Near-optimal transport cost By Lemma 3.1, the flow ψ_t corresponding to the optimal u_t satisfies that $\psi_0(x_0) = x_0 \sim q_0$ and $\psi_1(x_0) \sim q_1$. Hence, $x_0 \mapsto \psi_1(x_0)$ is a transport map between

q_0 and q_1 with an associated transport cost

$$\mathbb{E}_{q_0(x_0)}\|\psi_1(x_0) - x_0\|^2. \tag{3.17}$$

There is no reason to believe that when x_0 and x_1 are sampled independently, the transport cost $\mathbb{E}_{q_0(x_0)}\|\psi_1(x_0) - x_0\|^2$ will be anywhere near the optimal transport cost $W_2^2(p_0, p_1)$. Yet, in Section 3.4 we show that for well chosen q, the transport cost for ψ_1 does approach its optimal value. Computing optimal (or near-optimal) transport maps in high dimensions is a challenging task [Makkuva et al. 2020; Amos 2023] that extends beyond generative modeling and into the field of optimal transport, and it has applications in computer vision [Feydy et al. 2017; Solomon et al. 2015, 2016; Liu et al. 2023b] and computational biology [Lübeck et al. 2022; Bunne et al. 2021, 2022; Schiebinger et al. 2019], for instance. Hence, Joint CFM may also be viewed as a practical way to obtain approximately optimal transport maps in this context.

3.4 Multisample Flow Matching

Constructing a joint distribution satisfying the marginal constraints is difficult, especially since at least one of the marginal distributions is based on empirical data. We thus discuss a method to construct the joint distribution $q(x_0, x_1)$ implictly by designing a suitable sampling procedure that leaves the marginal distributions invariant. Note that training with (3.12) only requires sampling from $q(x_0, x_1)$.

We use a multisample construction for $q(x_0, x_1)$ in the following manner:

1. Sample $\{x_0^{(i)}\}_{i=1}^k \sim q_0(x_0)$ and $\{x_1^{(i)}\}_{i=1}^k \sim q_1(x_1)$.

2. Construct a doubly-stochastic matrix with probabilities $\pi(i, j)$ dependent on the samples $\{x_0^{(i)}\}_{i=1}^k$ and $\{x_1^{(i)}\}_{i=1}^k$.

3. Sample from the discrete distribution,
$q^k(x_0, x_1) = \frac{1}{k} \sum_{i,j=1}^k \delta(x_0 - x_0^i)\delta(x_1 - x_1^j)\pi(i, j)$.

Marginalizing $q^k(x_0, x_1)$ over samples from Step 1, we obtain the implicitly defined $q(x_0, x_1)$. By choosing different *couplings* $\pi(i, j)$, we induce different joint distributions. In this work, we focus on couplings that induce joint distributions which approximates, or at least partially satisfies, the optimal transport joint distribution. The following result, proven in App. B.4.3, guarantees that q has the right marginals.

Lemma 3.3. *The joint distribution $q(x_0, x_1)$ constructed in Steps [1-3] has marginals $q_0(x_0)$ and $q_1(x_1)$.*

That is, the marginal constraints (3.11) are satisfied and consequently we are allowed to use the framework of Section 3.3.

3.4.1 CondOT is Uniform Coupling

The aforementioned multisample construction subsumes the independent joint distribution used by prior works, when the joint coupling is taken to be uniformly distributed, *i.e.* $\pi(i, j) = \frac{1}{k}$. This is precisely the coupling used by [Lipman et al. 2023] under our introduced notion of Multisample Flow Matching, and acts as a natural reference point.

3.4.2 Batch Optimal Transport (BatchOT) Couplings

The natural connections between optimal transport theory and optimal sampling paths in terms of straight-line interpolations, lead us to the following pseudo-deterministic coupling, which we

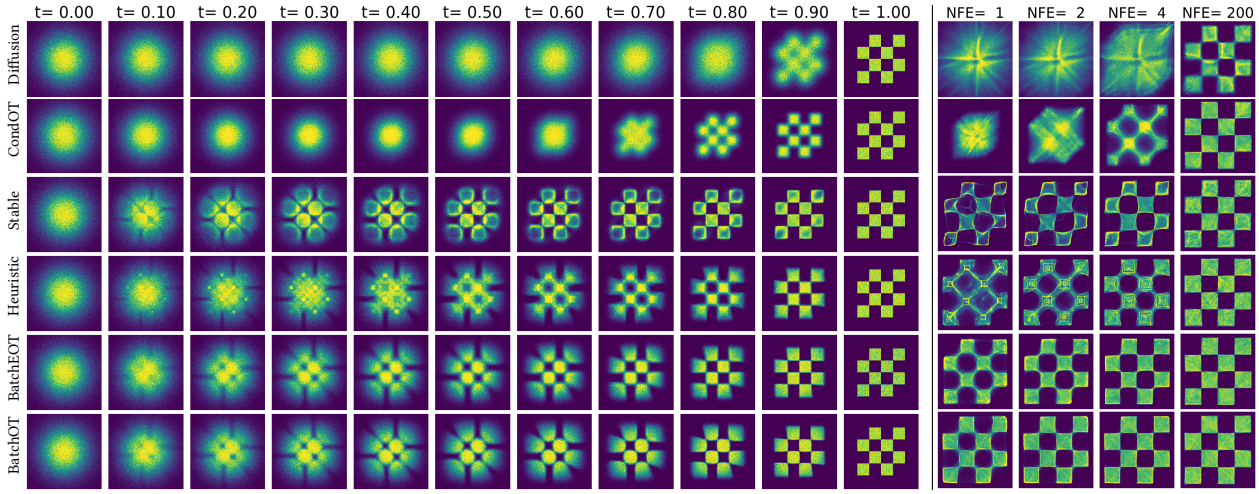


Figure 3.2: Multisample Flow Matching learn probability paths that are much closer to an optimal transport path than baselines such as Diffusion and CondOT paths. (*Left*) Exact marginal probability paths. (*Right*) Samples from trained models at $t = 1$ for different numbers of function evaluations (NFE), using Euler discretization. Furthermore, the final values of the Joint CFM objective (3.12)—upper bounds on the variance of u_t at convergence—are: CondOT: 10.72; Stable: 1.60, Heuristic: 1.56; BatchEOT: 0.57, BatchOT: 0.24.

call Batch Optimal Transport (BatchOT). While it is difficult to solve (3.9) at the population level, it can efficiently solved on the level of samples. Let $\{x_0^{(i)}\}_{i=1}^k \sim q_0(x_0)$ and $\{x_1^{(i)}\}_{i=1}^k \sim q_1(x_1)$. When defined on batches of samples, the OT problem (3.9) can be solved exactly and efficiently using standard solvers, as in `POT` [Flamary et al. 2021, Python Optimal Transport]. On a batch of k samples, the runtime complexity is well-understood via either the Hungarian algorithm or network simplex algorithm, with an overall complexity of $O(k^3)$ [Peyré and Cuturi 2019, Chapter 3]. The resulting coupling $\pi^{k,*}$ from the algorithm is a *permutation matrix*, which is a type of doubly-stochastic matrix that we can incorporate into Step 3 of our procedure.

We consider the effect that the sample size k has on the marginal vector field $u_t(x)$. The following theorem shows that in the limit of $k \to \infty$, BatchOT satisfies the three criteria that motivate Joint CFM: variance reduction, straight flows, and near-optimal transport cost.

Theorem 3.4 (Informal)**.** *Suppose that Multisample Flow Matching is run with BatchOT. Then, as* $k \to \infty$*,*

(i) *The value of the Joint CFM objective (Equation (3.12)) for the optimal u_t converges to 0.*

(ii) *The straightness S for the optimal marginal vector field u_t (Equation (3.15)) converges to zero.*

(iii) *The transport cost $\mathbb{E}_{q_0(x_0)}\|\psi_1(x_0) - x_0\|^2$ (Equation (3.17)) associated to u_t converges to the optimal transport cost $W_2^2(p_0, p_1)$.*

As $k \to \infty$, result *(i)* implies that the gradient variance both during training and at convergence is reduced due to Equation 3.14; result *(ii)* implies the optimal model will be easier to simulate between t=0 and t=1; result *(iii)* implies that Multisample Flow Matching can be used as a simulation-free algorithm for approximating optimal transport maps.

The full version of Thm. 3.4 can be found in App. B.4, and it makes use of standard, weak technical assumptions which are common in the optimal transport literature. While Thm. 3.4 only analyzes asymptotic properties, we provide theoretical evidence that the transport cost decreases with k, as summarized by a monotonicity result in Thm. B.8.

3.4.3 Batch Entropic OT (BatchEOT) Couplings

For k sufficiently large, the cubic complexity of the BatchOT approach is not always desirable, and instead one may consider approximate methods that produce couplings sufficiently close to BatchOT at a lower computational cost. A popular surrogate, pioneered in [Cuturi 2013], is to incorporate an entropic penalty parameter on the doubly stochastic matrix, pulling it closer to the independent coupling:

$$\min_{q \in \Gamma(q_0, q_1)} \mathbb{E}_{(x_0, x_1) \sim q} \|x_0 - x_1\|^2 + \varepsilon H(q)\,,$$

where $H(q) = -\sum_{i,j} q_{i,j}(\log(q_{i,j}) - 1)$ is the entropy of the doubly stochastic matrix q, and $\varepsilon > 0$ is some finite regularization parameter. The optimality conditions of this strictly convex program leads to Sinkhorn's algorithm, which has a runtime of $\tilde{O}(k^2/\varepsilon)$ [Altschuler et al. 2017].

The output of performing Sinkhorn's algorithm is a doubly-stochastic matrix. The two limit-

ing regimes of the regularization parameter are well understood (c.f. Peyré and Cuturi [2019], Proposition 4.1, for instance): as $\varepsilon \to 0$, BatchEOT recovers the BatchOT permutation matrix from subsection 3.4.2; as $\varepsilon \to \infty$, BatchEOT recovers the independent coupling on the indices from subsection 3.4.1.

3.4.4 Stable and Heuristic Couplings

An alternative approach is to consider faster algorithms that satisfy at least some desirable properties of an optimal coupling. In particular, an optimal coupling is *stable*. A permutation coupling is stable if *no pair of* $x_0^{(i)}$ *and* $x_1^{(j)}$ *favor each other over their assigned pairs based on the coupling.* Such a problem can be solved using the Gale-Shapeley algorithm [Gale and Shapley 1962] which has a compute cost of $O(k^2)$ given the cross set ranking of all samples. Starting from a random assignment, it is an iterative algorithm that reassigns pairs if they violate the stability property and can terminate very early in practice. Note that in a cost-based ranking, one has to sort the coupling costs of each sample with all samples in the opposing set, resulting in an overall $O(k^2 \log(k))$ compute cost.

The Gale-Shapeley algorithm is agnostic to any particular costs, however, as stability is only defined in terms of relative rankings of individual samples. We design a modified version of this algorithm based on a heuristic for satisfying the cyclical monotonicity property of optimal transport, namely that should pairs be reassigned, the reassignment should not increase the total cost of already matched pairs. We refer to the output of this modified algorithm as a *heuristic coupling* and discuss the details in Appendix B.1.2.

3.5 Related Work

Generative modeling and optimal transport are inherently intertwined topics, both often aiming to learn a transport between two distributions but with very different goals. Optimal transport

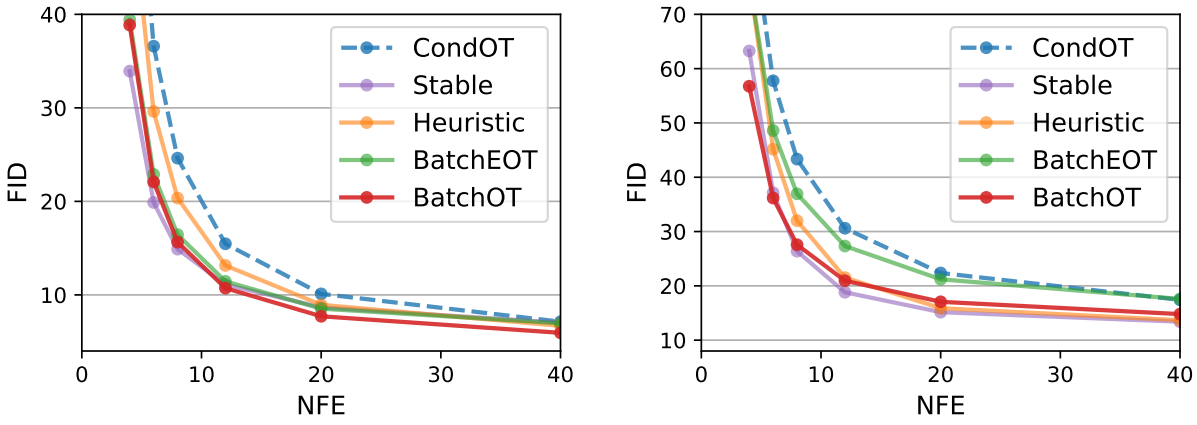


Figure 3.3: Sample quality (FID) vs compute cost (NFE) using Euler discretization. CondOT has significantly higher FID at lower NFE compared to proposed methods.

is widely recognized as a powerful tool for large-scale generative modeling as it can be used to stabilize training [Arjovsky et al. 2017]. In the context of continuous-time generative modeling, optimal transport has been used to regularize continuous normalizing flows for easier simulation [Finlay et al. 2020b; Onken et al. 2021], and increase interpretability [Tong et al. 2020]. However, the existing methods for encouraging optimality in a generative model generally require either solving a potentially unstable min-max optimization problem (*e.g.* [Arjovsky et al. 2017; Makkuva et al. 2020; Albergo and Vanden-Eijnden 2023]) or require simulation of the learned vector field as part of training (*e.g.* Finlay et al. [2020b]; Liu et al. [2022]). In contrast, the approach of using batch optimal couplings can be used to avoid the min-max optimization problem, but has not been successfully applied to generative modeling as they do not satisfy marginal constraints—we discuss this further in the following subsection 3.5.1. On the other hand, neural optimal transport approaches are mainly centered around the quadratic cost [Makkuva et al. 2020; Amos 2023; Finlay et al. 2020a] or rely heavily on knowing the exact cost function [Fan et al. 2021; Asadulaev et al. 2022]. Being capable of using batch optimal couplings allows us to build generative models to approximate optimal maps under any cost function, and even when the cost function is unknown.

3.5.1 Minibatch Couplings for Generative Modeling

Among works that use optimal transport for training generative models are those that make use of batch optimal solutions and their gradients such as Li et al. [2017]; Genevay et al. [2018]; Fatras et al. [2019]; Liu et al. [2019]. However, *naïvely using solutions to batches only produces, at best, the barycentric map, i.e.* the map that fits to average of the batch couplings [Ferradans et al. 2014; Seguy et al. 2017; Pooladian and Niles-Weed 2021], and does not correctly match the true marginal distribution. This is a well-known problem and while multiple works (*e.g.* Fatras et al. [2021]; Nguyen et al. [2022]) have attempted to circumvent the issue through alternative formulations of optimality, the lack of marginal preservation has been a major downside of using batch couplings for generative modeling as they do not have the ability to match the target distribution for finite batch sizes. This is due to the use of building models within the *static* setting, where the map is parameterized directly with a neural network. In contrast, we have shown in Theorem 3.3 that in our *dynamic* setting, where we parameterize the map as the solution of a neural ODE, it is possible to preserve the marginal distribution exactly. Furthermore, we have shown in Proposition B.7 (App. B.4.5) that our method produces a map that is no higher cost than the joint distribution induced from BatchOT couplings.

Concurrently, Tong et al. [2023] motivates the use of BatchOT solutions within a similar framework as our Joint CFM, but from the perspective of obtaining accurate solutions to dynamic optimal transport problems. Similarly, Lee et al. [2023] propose to explicitly learn a joint distribution, parameterized with a neural network, with the aim of minimizing trajectory curvature; this is done using through an auxiliary VAE-style objective function. In contrast, we propose a family of couplings that all satisfy the marginal constraints, all of which are easy to implement and have negligible cost during training. Our construction allow us to focus on (i) fixing consistency issues within simulation-free generative models, and (ii) using Joint CFM to obtain more optimal solutions than the original BatchOT solutions.

	ImageNet 32×32 NFE @ FID = 10	**ImageNet 64×64** NFE @ FID = 20
Diffusion	≥40	≥40
FM ^{w}/ CondOT	20	29
MultisampleFM ^{w}/ Heuristic	18	12
MultisampleFM ^{w}/ Stable	**14**	**11**
MultisampleFM ^{w}/ BatchOT	**14**	12

Table 3.1: Derived results shown in Figure 3.3, we can determine the approximate NFE required to achieve a certain FID across our proposed methods. The baseline diffusion-based methods (e.g. ScoreFlow and DDPM) require more than 40 NFE to achieve these FID values.

NFE	DDPM	ScoreSDE	BatchOT	Stable
Adaptive	5.72	6.84	**4.68**	5.79
40	19.56	16.96	**5.94**	7.02
20	63.08	58.02	**7.71**	8.66
8	232.97	218.66	15.64	**14.89**
6	275.28	266.76	22.08	**19.88**
4	362.37	340.17	38.86	**33.92**

Table 3.2: FID of model samples on ImageNet 32×32 using varying number of function evaluations (NFE) using Euler discretization.

3.6 Experiments

We empirically investigate Multisample Flow Matching on a suite of experiments. First, we show how different couplings affect the model on a 2D distribution. We then turn to benchmark, high-dimensional datasets, namely ImageNet [Deng et al. 2009]. We use the official *face-blurred* ImageNet data and then downsample to 32×32 and 64×64 using the open source preprocessing scripts from Chrabaszcz et al. [2017]. Finally, we explore the setting of unknown cost functions while only batch couplings are provided. Full details on the experimental setting can be found in Appendix B.5.2.

	ImageNet 32×32		**ImageNet 64×64**	
	CondOT	BatchOT	CondOT	BatchOT
Consistency(m=4)	0.141	**0.101**	0.174	**0.157**
Consistency(m=6)	0.105	**0.071**	0.151	**0.134**
Consistency(m=8)	0.079	**0.052**	0.132	**0.115**
Consistency(m=12)	0.046	**0.030**	0.106	**0.085**

Table 3.3: BatchOT produces samples with more similar content to its true samples at low NFEs (using midpoint discretization). Visual examples of this consistency are shown in Figure 3.1.

	2-D Cost			2-D KL		32-D Cost			32-D KL		64-D Cost			64-D KL	
Cost Fn. $c(x_0, x_1)$	B	B-ST	B-FM	B-ST	B-FM	B	B-ST	B-FM	B-ST	B-FM	B	B-ST	B-FM	B-ST	B-FM
$\|x_1 - x_0\|_2^2$	0.90	0.60	0.72	0.07	4E-3	41.08	31.58	38.73	151.47	0.06	92.90	65.57	87.97	335.38	0.14
$\|x_1 - x_0\|_1$	1.09	0.86	0.98	0.18	4E-3	27.92	24.51	27.26	254.59	0.08	60.27	50.49	58.38	361.16	0.16
$1 - \frac{\langle x_0, x_1 \rangle}{\|x_0\|\|x_1\|}$	0.03	2E-4	3E-3	5.91	4E-3	0.62	0.53	0.58	179.48	0.06	0.71	0.60	0.68	337.63	0.12
$\|A(x_1 - x_0)\|_2^2$	0.91	0.54	0.65	0.07	4E-3	32.66	24.61	30.13	256.90	0.06	78.70	58.11	78.50	529.09	0.19

Table 3.4: Matching couplings from an oracle BatchOT solver with unknown costs. Multisample Flow Matching is able to match the marginal distribution correctly while being at least a optimal as the oracle, but static maps fail to preserve the marginal distribution.

3.6.1 Insights from 2D experiments

Figure 3.2 shows the proposed Multisample Flow Matching algorithm on fitting to a checkboard pattern distribution in 2D. We show the marginal probability paths induced by different coupling algorithms, as well as low-NFE samples of trained models on these probability paths.

The diffusion and CondOT probability paths do not capture intricate details of the data distribution until it is almost at the end of the trajectory, whereas Multisample Flow Matching approaches provide a gradual transition to the target distribution along the flow. We also see that with a fixed step solver, the BatchOT method is able to produce an accurate target distribution in just one Euler step in this low-dimensional setting, while the other coupling approaches also get pretty close. Finally, it is interesting that both Stable and Heuristic exhibit very similar probability paths to optimal transport despite only satisfying weaker conditions.

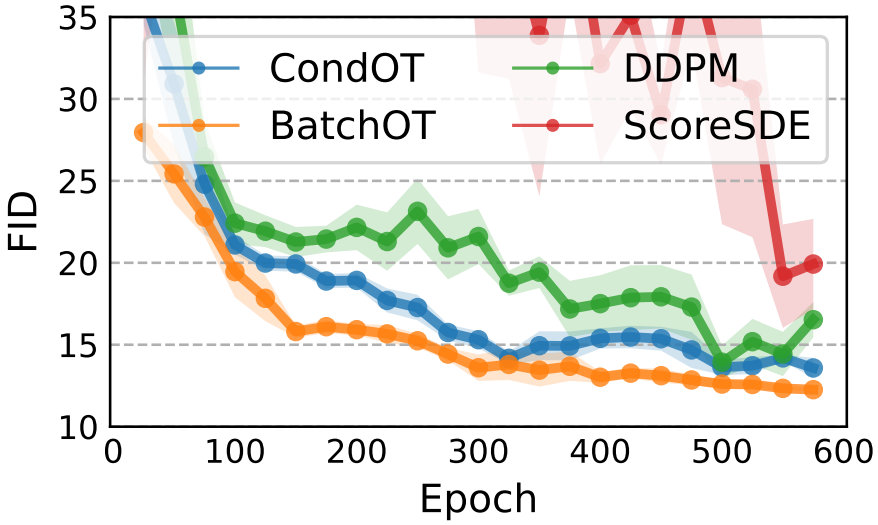


Figure 3.4: Multisample Flow Matching with BatchOT shows faster convergence due to reduced variance (ImageNet64).

3.6.2 Image Datasets

We find that Multisample Flow Matching retains the performance of Flow Matching while improving on sample quality, compute cost, and variance. In Table B.2 of subsection B.2.1, we report sample quality using the standard Fréchet Inception Distance (FID), negative log-likelihood values using bits per dimension (BPD), and compute cost using number of function evaluations (NFE); these are all standard metrics throughout the literature. Additionally, we report the variance of $u_t(x|x_0, x_1)$, estimated using the Joint CFM loss (3.12) which is an upper bound on the variance. We do not observe any performance degradations while simulation efficiency improves significantly, even with small batch sizes.

Additionally, in subsection B.2.5, we include runtime comparisons between Flow Matching and Multisample Flow Matching. On ImageNet32, we only observe a 0.8% relative increase in runtime compared to Flow Matching, and a 4% increase on ImageNet64.

Higher sample quality on a compute budget We observe that with a fixed NFE, models trained using Multisample Flow Matching generally achieve better sample quality. For these

experiments, we draw $x_0 \sim \mathcal{N}(0, I_d)$ and simulate $v_t(\cdot, \theta)$ up to time $t = 1$ using a fixed step solver with a fixed NFE. Figures 3.3 show that even on high dimensional data distributions, the sample quality of of multisample methods improves over the naïve CondOT approach as the number of function evaluations drops. We compare to the FID of diffusion baseline methods in Table 3.2, and provide additional results in subsection B.2.4.

Interestingly, we find that the Stable coupling actually performs on par, and some times better than the BatchOT coupling, despite having a smaller asymptotic compute cost and only satisfying a weaker condition within each batch.

As FID is computed over a full set of samples, it does not show how varying NFE affects individual sample paths. We discuss a notion of consistency next, where we analyze the similarity between low-NFE and high-NFE samples.

Consistency of individual samples In Figure 3.1 we show samples at different NFEs, where it can be qualitatively seen that BatchOT produces samples that are more consistent between high- and low-NFE solutions than CondOT, despite achieving similar FID values.

To evaluate this quantitatively, we define a metric for establishing the *consistency* of a model with respect to an integration scheme: let $x^{(m)}$ be the output of a numerical solver initialized at x using m function evalutions to reach $t = 1$, and let $x^{(*)}$ be a near-exact sample solved using a high-cost solver starting from x_0 as well. We define

$$\text{Consistency}(m) = \frac{1}{D}\mathbb{E}_{x\sim q_0}\|\mathcal{F}(x^{(m)}) - \mathcal{F}(x^{(*)})\|^2$$

where $\mathcal{F}(\cdot)$ outputs the hidden units from a pretrained InceptionNet[1], and D is the number of hidden units. These kinds of perceptual losses have been used before to check the content alignment between two image samples (*e.g.* Gatys et al. [2015]; Johnson et al. [2016]). We find that Multisample Flow Matching has better consistency at all values of NFE, shown in Table 3.3.

[1]We take the same layer as used in standard FID computation.

Training efficiency Figure 3.4 shows the convergence of Multisample Flow Matching with BatchOT coupling compared to Flow Matching with CondOT and diffusion-based methods. We see that by choosing better joint distributions, we obtain faster training. This is in line with our variance estimates reported in Table B.2 and supports our hypothesis that gradient variance is reduced by using non-trivial joint distributions.

3.6.3 Improved Batch Optimal Couplings

We further explore the usage of Multisample Flow Matching as an approach to improve upon batch optimal solutions. Here, we experiment with a different setting, where the cost is unknown and only samples from a batch optimal coupling are provided. In the real world, it is often the case that the preferences of each person are not known explicitly, but when given a finite number of choices, people can more easily find their best assignments. This motivates us to consider the case of unknown cost functions, and information regarding the optimal coupling is only given by a weak oracle that acts on finite samples, denoted $q^k_{OT,c}$. We consider two baselines: (i) the BatchOT cost (B) which corresponds to $\mathbb{E}_{q^k_{OT,c}(x_0,x_1)}[c(x_0,x_1)]$, and (ii) learning a static map that mimics the BatchOT couplings (B-ST) by minimizing the following objective:

$$\mathbb{E}_{q^k_{OT,c}(x_0,x_1)} \|x_1 - \psi_\theta(x_0)\|^2 .$$

This can be viewed as learning the barycentric projection [Ferradans et al. 2014; Seguy et al. 2017], *i.e.* $\psi^*(x_0) = E_{q^k_{OT,c}(x_1|x_0)}[x_1]$, a well-studied quantity but is known to not preserve the marginal distribution [Fatras et al. 2019].

We experiment with 4 different cost functions on three synthetic datasets in dimensions $\{2, 32, 64\}$ where both q_0 and q_1 are chosen to be Gaussian mixture models. In Table 3.4 we report both the transport cost and the KL divergence between q_1 and the distribution induced by the learned map, *i.e.* $[\psi_1]_\sharp q_0$. We observe that while B-ST always results in lower transport costs compared

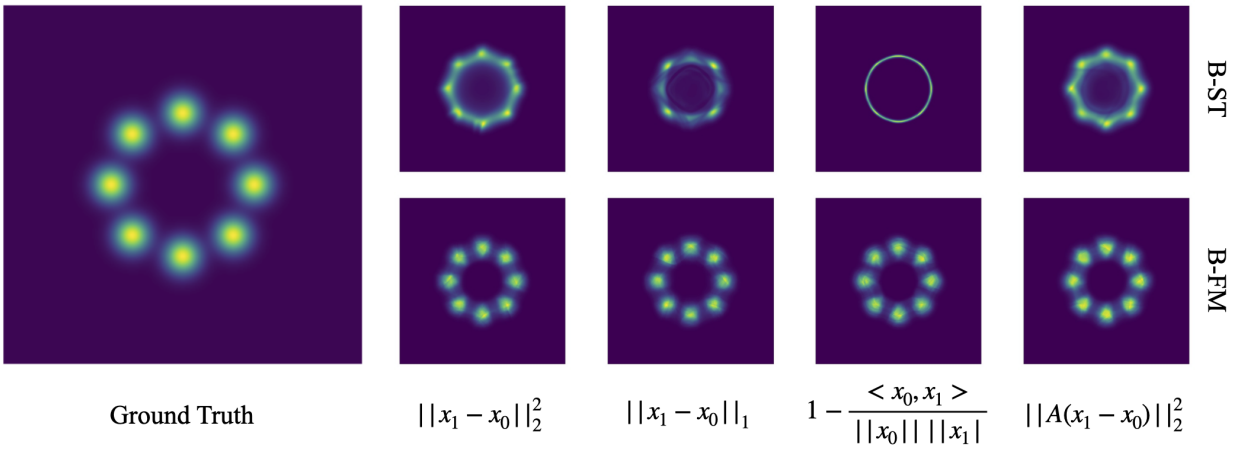


Figure 3.5: 2D densities on the 8-Gaussians target distribution. (Left) Ground truth density. (Right) Learned densities with static maps in the top row and Multisample Flow Matching dynamic maps in the bottom row. Models within each column were trained using batch optimal couplings with the corresponding cost function.

to B-FM, its KL divergence is always very high, meaning that the pushed-forward distribution by the learned static map poorly approximates q_1. Another interesting observation is that B-FM always reduces transport costs compared to B, providing experimental support to the theory (Theorem B.8).

Flow Matching improves optimality Figure 3.6 shows the cost of the learned model as we vary the batch size for computing couplings, where the models are trained sufficiently to achieve the same KL values as reported in Table 3.4. We see that our approach decreases the cost compared to the BatchOT oracle for any fixed batch size, and furthermore, converges to the OT solution faster than the batchOT oracle. Thus, since Multisample Flow Matching retains the correct marginal distributions, it can be used to better approximate optimal transport solutions than simply relying on a minibatch solution.

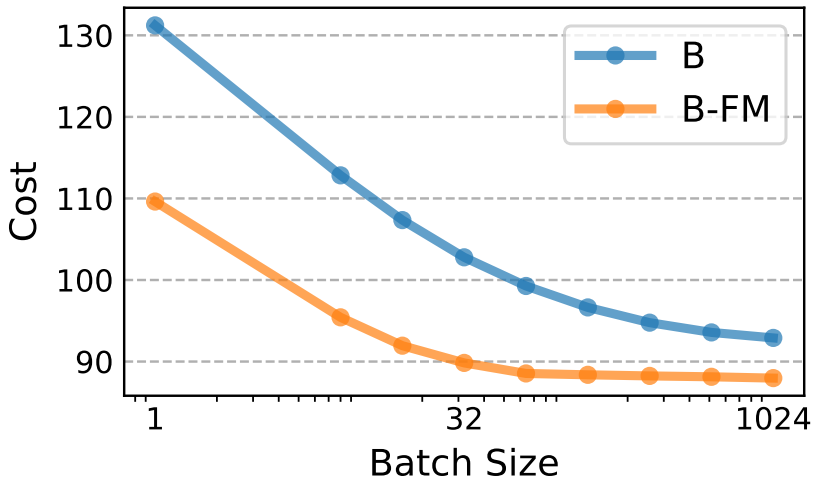


Figure 3.6: Transport cost vs. batch size (k) for computing couplings on the 64D synthetic dataset. The number of samples used for performing gradient steps during training and the resulting KL divergences were kept the same.

3.7 Conclusion

We propose Multisample Flow Matching, building on top of recent works on simulation-free training of continuous normalizing flows. While most prior works make use of training algorithms where data and noise samples are sampled independently, Multisample Flow Matching allows the use of more complex joint distribution. This introduces a new approach to designing probability paths. Our framework increases sample efficiency and sample quality when using low-cost solvers. Unlike prior works, our training method does not rely on simulation of the learned vector field during training, and does not introduce any min-max formulations. Finally, we note that our method of fitting to batch optimal couplings is the first to also preserve the marginal distributions, an important property in both generative modeling and solving transport problems.

4 Stochastic Optimal Control Matching

4.1 Introduction

Stochastic optimal control aims to drive the behavior of a noisy system in order to minimize a given cost. It has myriad applications in science and engineering: examples include the simulation of rare events in molecular dynamics [Hartmann et al. 2014; Hartmann and Schütte 2012; Zhang et al. 2014; Holdijk et al. 2023], finance and economics [Pham 2009; Fleming and Stein 2004], stochastic filtering and data assimilation [Mitter 1996; Reich 2019], nonconvex optimization [Chaudhari et al. 2018], power systems and energy markets [Belloni et al. 2016; Powell and Meisel 2016], and robotics [Theodorou et al. 2011; Gorodetsky et al. 2018]. Stochastic optimal has also been very impactful in neighboring fields such as mean-field games [Carmona et al. 2018], optimal transport [Villani 2003, 2008], backward stochastic differential equations (BSDEs) [Carmona 2016] and large deviations [Feng and Kurtz 2006].

For continuous-time problems with low-dimensional state spaces, the standard approach to learn the optimal control is to solve the Hamilton-Jacobi-Bellman (HJB) partial differential equation (PDE) by gridding the space and using classical numerical methods. For high-dimensional problems, a large number of works parameterize the control using a neural network and train it applying a stochastic optimization algorithm on a loss function. These methods are known as *Iterative*

Diffusion Optimization (IDO) techniques [Nüsken and Richter 2021] (see subsection 4.2.2).

It is convenient to draw an analogy between stochastic optimal control and *continuous normalizing flows* (CNFs), which are a generative modeling technique where samples are generated by solving an ordinary differential equation (ODE) for which the vector field has been learned, initialized at a Gaussian sample. CNFs were introduced by [Chen et al. 2018] (building on top of Rezende and Mohamed [2015]), and training them is similar to solving control problems because in both cases one needs to learn high-dimensional vector fields using neural networks, in continuous time.

The first algorithm developed to train normalizing flows was based on maximizing the likelihood of the generated samples [Chen et al. 2018, Sec. 4]. Obtaining the gradient of the maximum likelihood loss with respect to the vector field parameters requires backpropagating through the computation of the ODE trajectory, or equivalently, solving the *adjoint* ODE in parallel to the original ODE. Maximum likelihood CNFs (ML-CNFs) were superseded by diffusion models [Song and Ermon 2019; Ho et al. 2020; Song et al. 2021c] and flow-matching, a.k.a. stochastic interpolant, methods [Lipman et al. 2023; Albergo and Vanden-Eijnden 2023; Pooladian et al. 2023; Albergo et al. 2023], which are currently the preferred algorithms to train CNFs. Aside from architectural improvements such as the UNet [Ronneberger et al. 2015], a potential reason for the success of diffusion and flow matching models is that their *functional landscape* is convex, unlike for ML-CNFs. Namely, vector fields are learned by solving least squares regression problems where the goal is to fit a random matching vector field. Convex functional landscapes in combination with overparameterized models and moderate gradient variance can yield very stable training dynamics and help achieve low error.

Returning to stochastic optimal control, one of the best-performing IDO techniques amounts to choosing the control objective (equation 4.1) as the training loss (see (4.12)). As in ML-CNFs, computing the gradient of this loss requires backpropagating through the computation of the trajectories of the SDE (4.2), or equivalently, using an adjoint method. The functional landscape

of the loss is highly non-convex, and the method is prone to unstable training (see green curve in the bottom right plot of Figure 4.2). In light of this, a natural idea is to develop the analog of diffusion model losses for the stochastic optimal control problem, to obtain more stable training and lower error, and this is what we set out to do in our work. Our contributions are as follows:

- We introduce Stochastic Optimal Control Matching (SOCM), a novel IDO algorithm in which the control is learned by solving a least-squares regression problem where the goal is to fit a random *matching vector field* which depends on a family of *reparameterization matrices* that are also optimized.
- We derive a bias-variance decomposition of the SOCM loss (Theorem 4.5). The bias term is equal to an existing IDO loss: the *cross-entropy loss*, which shows that both algorithms have the same landscape in expectation. However, SOCM has an extra flexibility in the choice of reparameterization matrices, which affect only the variance. Hence, we propose optimizing the reparameterization matrices to reduce the variance of the SOCM objective.
- The key idea that underlies the SOCM algorithm is the *path-wise reparameterization trick* (Theorem 4.4), which is a novel technique for estimating gradients of an expectation of a functional of a random process with respect to its initial value. It is of independent interest and may be more generally applicable outside of the settings considered in this paper.
- We perform experiments on four different settings where we have access to the ground-truth control. For three of these, SOCM obtains a lower L^2 error with respect to the ground-truth control than all the existing IDO techniques, with around $10\times$ lower error than competing methods in some instances.

4.2 Framework

4.2.1 Setup and Preliminaries

Let $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathcal{P})$ be a fixed filtered probability space on which is defined a Brownian motion $B = (B_t)_{t\geq 0}$. We consider the control-affine problem

$$\min_{u\in\mathcal{U}} \mathbb{E}\Big[\int_0^T \big(\frac{1}{2}\|u(X_t^u, t)\|^2 + f(X_t^u, t)\big)\,\mathrm{d}t + g(X_T^u)\Big], \tag{4.1}$$

$$\text{where } \mathrm{d}X_t^u = (b(X_t^u, t) + \sigma(t)u(X_t^u, t))\,\mathrm{d}t + \sqrt{\lambda}\sigma(t)\mathrm{d}B_t, \qquad X_0^u \sim p_0. \tag{4.2}$$

and where $X_t^u \in \mathbb{R}^d$ is the state, $u : \mathbb{R}^d \times [0,T]$ is the feedback control and belongs to the set of admissible controls $\mathcal{U}$, $f : \mathbb{R}^d \times [0,T] \to \mathbb{R}$ is the state cost, $g : \mathbb{R}^d \to \mathbb{R}$ is the terminal cost, $b : \mathbb{R}^d \times [0,T] \to \mathbb{R}^d$ is the base drift, and $\sigma : [0,T] \to \mathbb{R}^{d\times d}$ is the invertible covariance matrix and $\lambda \in (0, +\infty)$ is the noise level. In section C.1 we formally define the set $\mathcal{U}$ of admissible controls and describe the regularity assumptions needed on the control functions. In the remainder of the section we introduce relevant concepts in stochastic optimal control; we provide the most relevant proofs in section C.2 and refer the reader to Oksendal [2013, Chap. 11] and Nüsken and Richter [2021, Sec. 2] for a similar, more extensive treatment.

Cost functional and value function The *cost functional* for the control u, point x and time t is defined as $J(u; x, t) := \mathbb{E}\big[\int_t^T \big(\frac{1}{2}\|u_s(X_s^u)\|^2 + f_s(X_s^u)\big)\,\mathrm{d}t + g(X_T^u)\big|X_t^u = x\big]$. That is, the cost functional is the expected value of the control objective restricted to the times $[t, T]$ with the initial value x at time t. The *value function* or *optimal cost-to-go* at a point x and time t is defined as the minimum value of the cost functional across all possible controls:

$$V(x, t) := \inf_{u\in\mathcal{U}} J(u; x, t). \tag{4.3}$$

HAMILTON-JACOBI-BELLMAN EQUATION AND OPTIMAL CONTROL If we define the infinitesimal generator $L := \frac{\lambda}{2}\sum_{i,j=1}^{d}(\sigma\sigma^\top)_{ij}(t)\partial_{x_i}\partial_{x_j} + \sum_{i=1}^{d} b_i(x,t)\partial_{x_i}$, the value function solves the following Hamilton-Jacobi-Bellman (HJB) partial differential equation:

$$(\partial_t + L)V(x,t) - \frac{1}{2}\|(\sigma^\top \nabla V)(x,t)\|^2 + f(x,t) = 0,$$
$$V(x,T) = g(x). \tag{4.4}$$

The *verification theorem* [Pavliotis 2014, Sec. 2.3] states that if a function V solves the HJB equation above and has certain regularity conditions, then V is the value function (4.3) of the problem (4.1)-(4.2). An implication of the verification theorem is that for every $u \in \mathcal{U}$,

$$V(x,t) + \mathbb{E}\Big[\frac{1}{2}\int_t^T \|\sigma^\top \nabla V + u\|^2(X_s^u, s)\,\mathrm{d}s \,\Big|\, X_t^u = x\Big] = J(u,x,t). \tag{4.5}$$

In particular, this implies that the unique optimal control is given in terms of the value function as $u^*(x,t) = -\sigma(t)^\top \nabla V(x,t)$. Equation (4.5) can be deduced by integrating the HJB equation (4.4) over $[t,T]$, and taking the conditional expectation with respect to $X_t^u = x$. We include the proof of (4.5) in section C.2 for completeness.

A PAIR OF FORWARD AND BACKWARD SDES (FBSDES) Consider the pair of SDEs

$$\mathrm{d}X_t = b(X_t,t)\,\mathrm{d}t + \sqrt{\lambda}\sigma(t)\mathrm{d}B_t, \qquad X_0 \sim p_0, \tag{4.6}$$
$$\mathrm{d}Y_t = (-f(X_t,t) + \frac{1}{2}\|Z_t\|^2)\,\mathrm{d}t + \sqrt{\lambda}\langle Z_t, \mathrm{d}B_t\rangle, \qquad Y_T = g(X_T). \tag{4.7}$$

where $Y : \Omega \times [0,T] \to \mathbb{R}$ and $Z : \Omega \times [0,T] \to \mathbb{R}^d$ are progressively measurable [1] random processes. It turns out that Y_t and Z_t defined as $Y_t := V(X_t,t)$ and $Z_t := \sigma(t)^\top \nabla V(X_t,t) = -u^*(X_t,t)$ satisfy (4.7). We include the proof in section C.2 for completeness.

[1] Being progressively measurable is a strictly stronger property than the notion of being a process adapted to the filtration $\mathcal{F}_t$ of B_t (see [Karatzas and Shreve 1991]).

AN ANALYTIC EXPRESSION FOR THE VALUE FUNCTION From the forward-backward equations (4.6)-(4.7), one can derive a closed-form expression for the value function V:

$$V(x,t) = -\lambda \log \mathbb{E}\Big[\exp\Big(-\lambda^{-1} \int_t^T f(X_s, s)\, \mathrm{d}s - \lambda^{-1} g(X_T)\Big) \Big| X_t = x \Big], \tag{4.8}$$

where X_t is the solution of the uncontrolled SDE (4.6). This is a classical result, but we still include its proof in section C.2. Given that $u^*(x,t) = -\sigma(t)^\top \nabla V(x,t)$, an immediate, yet important, consequence of (4.8) is the following representation of the optimal control:

Lemma 4.1 (Path-integral representation of the optimal control [Kappen 2005])**.**

$$u^*(x,t) = \lambda \sigma(t)^\top \nabla_x \log \mathbb{E}\Big[\exp\Big(-\lambda^{-1} \int_t^T f(X_s, s)\, \mathrm{d}s - \lambda^{-1} g(X_T)\Big) \Big| X_t = x \Big] \,. \tag{4.9}$$

Remark that the right-hand side of this equation involves the gradient of logarithm of a conditional expectation. This is reminiscent of the vector fields that are learned when training diffusion models or flow matching algorithms. For example, the target vector field for variance-exploding score-based diffusion loss [Song et al. 2021c] can be expressed as

$$\nabla_x \log p_t(x) = \nabla_x \log \mathbb{E}_{Y \sim p_{\text{data}}} [\frac{\exp(-\|x - Y\|^2/(2\sigma_t^2))}{(2\pi\sigma_t^2)^{d/2}}].$$

Note, however, that in (4.9) the gradient is taken with respect to the initial condition of the process, which requires the development of novel techniques.

CONDITIONED DIFFUSIONS Let $C = C([0,T]; \mathbb{R}^d)$ be the Wiener space of continuous functions from $[0,T]$ to $\mathbb{R}^d$ equipped with the supremum norm, and let $\mathcal{P}(C)$ be the space of Borel probability measures over C. For each control $u \in \mathcal{U}$, the controlled process in equation (4.2) induces a probability measure in $\mathcal{P}(C)$, as the law of the paths X_t^u, which we refer to as $\mathbb{P}^u$. We let $\mathbb{P}$ be

the probability measure induced by the uncontrolled process (4.6), and define the *work functional*

$$\mathcal{W}(X,t) := \int_t^T f(X_s,s)\,\mathrm{d}s + g(X_T). \tag{4.10}$$

It turns out (Theorem C.4 in section C.2) that the Radon-Nikodym derivative $\frac{d\mathbb{P}^{u^*}}{d\mathbb{P}}$ satisfies $\frac{d\mathbb{P}^{u^*}}{d\mathbb{P}}(X) = \exp\left(\lambda^{-1}\left(V(X_0,0) - \mathcal{W}(X,0)\right)\right)$. Also, a straight-forward application of the Girsanov theorem for SDEs (Theorem C.3) shows that

$$\frac{d\mathbb{P}^u}{d\mathbb{P}^{u^*}}(X^{u^*}) = \exp\left(-\lambda^{-1/2}\int_0^T \langle u^*(X_t^{u^*},t) - u(X_t^{u^*},t), \mathrm{d}B_t\rangle - \frac{\lambda^{-1}}{2}\int_0^T \|u^*(X_t^{u^*},t) - u(X_t^{u^*},t)\|^2\,\mathrm{d}t\right), \tag{4.11}$$

which means that the only control $u \in \mathcal{U}$ such that $\mathbb{P}^u = \mathbb{P}^{u^*}$ is the optimal control itself. Such changes of process are the basic tools to design IDO losses, and we leverage them as well.

4.2.2 Existing approaches and related work

Low-dimensional case: solving the HJB equation For low-dimensional control problems ($d \leq 3$), it is possible to grid the domain and use a numerical PDE solver to find a solution to the HJB equation (4.4). The main approaches include *finite difference methods* [Bonnans et al. 2004; Ma and Ma 2020; Baňas et al. 2022], which approximate the derivatives and gradients of the value function using finite differences, *finite element methods* [Jensen and Smears 2013], which involve restricting the solution to domain-dependent function spaces, and semi-Lagrangian schemes [Debrabant and Jakobsen 2013; Carlini et al. 2020; Calzola et al. 2022], which trace back characteristics and have better stability than finite difference methods. See Greif [2017] for an overview on these techniques, and Baňas et al. [2022] for a comparison between them. Hutzenthaler et al. [2016] introduced the multilevel Picard method, which leverages the Feynman-Kac and the Bismut-Elworthy-Li formulas to beat the curse of dimensionality in some settings [Beck et al. 2019; Hutzenthaler et al. 2019, 2018; Hutzenthaler and Kruse 2020].

High dimensional methods leveraging FBSDEs The FBSDE formulation in equations (4.6)-(4.7) has given rise to multiple methods to learn controls. One such approach is *least-squares Monte Carlo* (see Pham [2009, Chapter 3] and Gobet [2016] for an introduction, and Gobet et al. [2005]; Zhang et al. [2004] for an extensive analysis), where trajectories from the forward process (4.6) are sampled, and then regression problems are solved backwards in time to estimate the expected future cost in the spirit of dynamic programming. A second method that exploits FBSDEs was proposed by E et al. [2017]; Han et al. [2018]. They parameterize the control using a neural network u_θ, and use stochastic gradient algorithms to minimize the loss $\mathcal{L}(u_\theta, y_0) = \mathbb{E}[(Y_T(y_0, u_\theta) - g(X_T))^2]$, where $Y_T(y_0, u_\theta)$ is the process in (4.7) with initial condition y_0 and control u_θ. This algorithm can be seen as a shooting method, where the initial condition and the control are learned to match the terminal condition. Multiple recent works have combined neural networks with FBSDE Monte Carlo methods for parabolic and elliptic PDEs [Beck et al. 2018; Chan-Wai-Nam et al. 2019; Zhou et al. 2021], control [Becker et al. 2019; Hartmann et al. 2019], multi-agent games [Han and Hu 2020; Carmona and Laurière 2021, 2022]; see [E et al. 2021] for a more comprehensive review.

Many of the methods referenced above and some additional ones can be seen from a common perspective using controlled diffusions. As observed in equation (4.11), the key idea is that learning the optimal control is equivalent to finding a control u such that the induced probability measure $\mathbb{P}^u$ on paths is equal to the probability measure $\mathbb{P}^{u^*}$ for the optimal control. In the paragraphs below we cover several loss that fall into this framework. All the losses below can be optimized using a common algorithmic framework, which we describe in Algorithm 3. For more details, we refer the reader to Nüsken and Richter [2021], which introduced this perspective and named such methods *Iterative Diffusion Optimization* (IDO) techniques. For simplicity, we introduce the losses for the setting in which the initial distribution p_0 is concentrated at a single point x_{init}; we cover the general setting in section C.2.

Algorithm 3 Iterative Diffusion Optimization (IDO) algorithms for stochastic optimal control

Input: State cost $f(x,t)$, terminal cost $g(x)$, covariance matrix $\sigma(t)$, base drift $b(x,t)$, noise level λ, number of iterations N, batch size m, number of time steps K, initial control parameters θ_0, loss $\mathcal{L} \in \{\mathcal{L}_{\text{Adj}}(4.12), \mathcal{L}_{\text{CE}}(4.13), \mathcal{L}_{\text{Var}_v}(4.15), \mathcal{L}^{\log}_{\text{Var}_v}(4.16), \mathcal{L}_{\text{Mom}_v}(4.17)\}$

for $n \in \{0, \dots, N-1\}$ **do**
- Simulate m trajectories of the process X^v controlled by $v = u_{\theta_n}$, e.g., using Euler-Maruyama updates
- **if** $\mathcal{L} \neq \mathcal{L}_{\text{Adj}}$ **then** detach the m trajectories from the computational graph, so that gradients do not backpropagate;
- Using the m trajectories, compute an m-sample Monte Carlo approximation $\hat{\mathcal{L}}(u_{\theta_n})$ of the loss $\mathcal{L}(u_{\theta_n})$
- Compute the gradients $\nabla_\theta \hat{\mathcal{L}}(u_{\theta_n})$ of $\hat{\mathcal{L}}(u_{\theta_n})$ w.r.t. θ_n
- Obtain θ_{n+1} with via an Adam update on θ_n (or another stochastic algorithm)

end

Output: Learned control u_{θ_N}

THE RELATIVE ENTROPY LOSS AND THE ADJOINT METHOD The relative entropy loss is defined as the Kullback-Leibler divergence between $\mathbb{P}^u$ and $\mathbb{P}^{u^*}$: $\mathbb{E}_{\mathbb{P}^u}[\log \frac{d\mathbb{P}^u}{d\mathbb{P}^{u^*}}]$. Upon removing constant terms and factors, this loss is equivalent to (see Theorem C.5 in section C.2, or Hartmann and Schütte [2012]; Kappen et al. [2012]):

$$\mathcal{L}_{\text{Adj}}(u) := \mathbb{E}\Big[\int_0^T \Big(\frac{1}{2}\|u(X_t^u, t)\|^2 + f(X_t^u, t)\Big)\,\mathrm{d}t + g(X_T^u)\Big]. \tag{4.12}$$

This is exactly the control objective in (4.1). This connection has been studied extensively [Bierkens and Kappen 2014; Gómez et al. 2014; Hartmann and Schütte 2012; Kappen et al. 2012; Rawlik et al. 2013]. Hence, the relative entropy loss is a very natural one, and is widely used; see Onken et al. [2023]; Zhang and Chen [2022] for some examples on multiagent systems and sampling.

Solving optimization problems of the form (4.12) has a long history that dates back to Pontryagin [1962]. Note that $\mathcal{L}_{\text{Adj}}(u)$ depends on u both explicitly, and implicitly through the process X^u. To compute the gradient $\nabla_\theta \hat{\mathcal{L}}_{\text{Adj}}(u_{\theta_n})$ of a Monte Carlo approximation $\hat{\mathcal{L}}_{\text{Adj}}(u_{\theta_n})$ of $\mathcal{L}_{\text{Adj}}(u_{\theta_n})$ as required by Algorithm 3, we need to backpropagate through the simulation of the m trajectories, which is why we do *not* detach them from the computational graph. One can alternatively compute the gradient $\nabla_\theta \hat{\mathcal{L}}_{\text{Adj}}(u_{\theta_n})$ by explicitly solving an ODE, a technique which is known as the

adjoint method. The adjoint method was introduced by Pontryagin [1962], popularized in deep learning by Chen et al. [2018], and further developed for SDEs in Li et al. [2020].

THE CROSS-ENTROPY LOSS The cross-entropy loss is defined as the Kullback-Leibler divergence between $\mathbb{P}^{u^*}$ and $\mathbb{P}^u$, i.e., flipping the order of the two measures: $\mathbb{E}_{\mathbb{P}^{u^*}}[\log \frac{d\mathbb{P}^{u^*}}{d\mathbb{P}^u}]$. For an arbitrary $v \in \mathcal{U}$, this loss is equivalent to the following one (see Theorem C.6(i) in section C.2):

$$\mathcal{L}_{\mathrm{CE}}(u) := \mathbb{E}\Big[\big(-\lambda^{-1/2}\int_0^T \langle u(X_t^v,t), \mathrm{d}B_t\rangle - \lambda^{-1}\int_0^T \langle u(X_t^v,t), v(X_t^v,t)\rangle\,\mathrm{d}t + \frac{\lambda^{-1}}{2}\int_0^T \|u(X_t^v,t)\|^2\,\mathrm{d}t\big) \\ \times \exp\big(-\lambda^{-1}\mathcal{W}(X^v,0) - \lambda^{-1/2}\int_0^T \langle v(X_t^v,t), \mathrm{d}B_t\rangle - \frac{\lambda^{-1}}{2}\int_0^T \|v(X_t^v,t)\|^2\,\mathrm{d}t\big)\Big]. \quad (4.13)$$

The cross-entropy loss has a rich literature [Hartmann et al. 2017; Kappen and Ruiz 2016; Rubinstein and Kroese 2013; Zhang et al. 2014] and has been recently used in applications such as molecular dynamics [Holdijk et al. 2023].

Furthermore, we note that the cross-entropy loss can be significantly simplified and written in terms of the L^2 error of the control u with respect to the optimal control u^*:

Lemma 4.2 (Cross-entropy loss in terms of control L^2 error)**.**

$$\mathcal{L}_{\mathrm{CE}}(u) = \frac{\lambda^{-1}}{2}\mathbb{E}\Big[\int_0^T \|u^*(X_t^{u^*},t) - u(X_t^{u^*},t)\|^2\,\mathrm{d}t\,\exp\big(-\lambda^{-1}V(X_0^{u^*},0)\big)\Big].$$

This characterization, which is proven in Theorem C.6(ii) in section C.2, is relevant for us because a similar one can be written for the loss that we propose (see Theorem 4.5).

VARIANCE AND LOG-VARIANCE LOSSES For an arbitrary $v \in \mathcal{U}$, the *variance* and the *log-variance losses* are defined as $\tilde{\mathcal{L}}_{\mathrm{Var}_v}(u) = \mathrm{Var}_{\mathbb{P}^v}(\frac{\mathrm{d}\mathbb{P}^{u^*}}{\mathrm{d}\mathbb{P}^u})$ and $\tilde{\mathcal{L}}^{\log}_{\mathrm{Var}_v}(u) = \mathrm{Var}_{\mathbb{P}^v}(\log \frac{\mathrm{d}\mathbb{P}^{u^*}}{\mathrm{d}\mathbb{P}^u})$ whenever $\mathbb{E}_{\mathbb{P}^v}|\frac{\mathrm{d}\mathbb{P}^{u^*}}{\mathrm{d}\mathbb{P}^u}| <$

$+\infty$ and $\mathbb{E}_{\mathbb{P}^v}|\log \frac{d\mathbb{P}^{u^*}}{d\mathbb{P}^u}| < +\infty$, respectively. Define

$$\tilde{Y}_T^{u,v} = -\lambda^{-1} \int_0^T \langle u(X_t^v, t), v(X_t^v, t)\rangle \, \mathrm{d}t - \lambda^{-1} \int_0^T f(X_t^v, t) \, \mathrm{d}t - \lambda^{-1/2} \int_0^T \langle u(X_t^v, t), \mathrm{d}B_t\rangle + \frac{\lambda^{-1}}{2} \int_0^T \|u(X_t^v, t)\|^2 \, \mathrm{d}t. \tag{4.14}$$

Then, $\tilde{\mathcal{L}}_{\mathrm{Var}_v}$ and $\tilde{\mathcal{L}}^{\log}_{\mathrm{Var}_v}$ are equivalent, respectively, to the following losses (see Theorem C.7):

$$\mathcal{L}_{\mathrm{Var}_v}(u) := \mathrm{Var}\big(\exp\big(\tilde{Y}_T^{u,v} - \lambda^{-1} g(X_T^v)\big)\big), \tag{4.15}$$

$$\mathcal{L}^{\log}_{\mathrm{Var}_v}(u) := \mathrm{Var}\big(\tilde{Y}_T^{u,v} - \lambda^{-1} g(X_T^v)\big), \tag{4.16}$$

The variance and log-variance losses were introduced by Nüsken and Richter [2021]. Unlike for the cross-entropy loss, the choice of the control v does lead to different losses. When using $\mathcal{L}_{\mathrm{Var}_v}$ or $\mathcal{L}^{\log}_{\mathrm{Var}_v}$ in Algorithm 3, the variance is computed across the m trajectories in each batch.

Moment loss For an arbitrary $v \in \mathcal{U}$, the moment loss is defined as

$$\mathcal{L}_{\mathrm{Mom}_v}(u, y_0) = \mathbb{E}[(\tilde{Y}_T^{u,v} + y_0 - \lambda^{-1} g(X_T^v))^2], \tag{4.17}$$

where $\tilde{Y}_T^{u,v}$ is defined in (4.14). Note the similarity with the log-variance loss (4.16); the optimal value of y_0 for a fixed u is $y_0^* = \mathbb{E}[\lambda^{-1} g(X_T^v) - \tilde{Y}_T^{u,v}]$, and plugging this into (4.17) yields exactly the log-variance loss. The moment loss was introduced by Hartmann et al. [2019, Section III.B], and it is a generalization of the FBSDE method pioneered by E et al. [2017]; Han et al. [2018] and referenced earlier in this subsection. In fact, the original method corresponds to setting $v = 0$.

4.3 Stochastic Optimal Control Matching

In this section we present our loss, *Stochastic Optimal Control Matching* (SOCM). The corresponding method, which we describe in Algorithm 4, falls into the class of IDO techniques described in subsection 4.2.2. The general idea is to leverage the analytic expression of u^* in Theorem 4.1 to write a least squares loss for u, and the main challenge is to reexpress the gradient of a conditional expectation with respect to the initial condition of the process. We do that using a novel technique which introduces certain arbitrary matrix-valued functions M_t, that we also optimize.

Theorem 4.3 (SOCM loss)**.** *For each $t \in [0,T]$, let $M_t : [t,T] \to \mathbb{R}^{d\times d}$ be an arbitrary matrix-valued differentiable function such that $M_t(t) = \mathrm{Id}$. Let $v \in \mathcal{U}$ be an arbitrary control. Let $\mathcal{L}_{\mathrm{SOCM}} : L^2(\mathbb{R}^d \times [0,T]; \mathbb{R}^d) \times L^2([0,T]^2; \mathbb{R}^{d\times d}) \to \mathbb{R}$ be the loss function defined as*

$$\mathcal{L}_{\mathrm{SOCM}}(u, M) := \mathbb{E}\Big[\frac{1}{T}\int_0^T \big\|u(X_t^v, t) - w(t, v, X^v, B, M_t)\big\|^2 \,\mathrm{d}t \times \alpha(v, X^v, B)\Big], \tag{4.18}$$

where X^v is the process controlled by v (i.e., $dX_t^v = (b(X_t^v, t) + \sigma(t)v(X_t^v, t))\,\mathrm{d}t + \sqrt{\lambda}\sigma(t)\,\mathrm{d}B_t$ and $X_0^v \sim p_0$), and

$$\begin{aligned} w(t, v, X^v, B, M_t) = \sigma(t)^\top \Big(&- \int_t^T M_t(s)\nabla_x f(X_s^v, s)\,\mathrm{d}s - M_t(T)\nabla g(X_T^v) \\ &+ \int_t^T (M_t(s)\nabla_x b(X_s^v, s) - \partial_s M_t(s))(\sigma^{-1})^\top(s) v(X_s^v, s)\,\mathrm{d}s \\ &+ \lambda^{1/2}\int_t^T (M_t(s)\nabla_x b(X_s^v, s) - \partial_s M_t(s))(\sigma^{-1})^\top(s)\mathrm{d}B_s\Big), \\ \alpha(v, X^v, B) = \exp\Big(&- \lambda^{-1}\int_0^T f(X_t^v, t)\,\mathrm{d}s - \lambda^{-1} g(X_T^v) \\ &- \lambda^{-1/2}\int_0^T \langle v(X_t^v, t), \mathrm{d}B_t\rangle - \frac{\lambda^{-1}}{2}\int_0^T \|v(X_t^v, t)\|^2\,\mathrm{d}t\Big). \end{aligned} \tag{4.19}$$

$\mathcal{L}_{\mathrm{SOCM}}$ has a unique optimum (u^*, M^*), where u^* is the optimal control.

We refer to $M = (M_t)_{t\in[0,T]}$ as the family of *reparametrization matrices*, to the random vector field

w as the *matching vector field*, and to α as the *importance weight*. We present a proof sketch of Theorem 4.3; the full proofs for all the results in this section are in section C.3.

Proof sketch of Theorem 4.3 Let X be the uncontrolled process (4.6). Consider the loss

$$\begin{aligned}\tilde{\mathcal{L}}(u) &= \mathbb{E}\Big[\frac{1}{T}\int_0^T \big\|u(X_t,t) - u^*(X_t,t)\big\|^2 \,\mathrm{d}t \,\exp\big(-\lambda^{-1}\int_0^T f(X_t,t)\,\mathrm{d}t - \lambda^{-1}g(X_T)\big)\Big] \quad (4.20)\\ &= \mathbb{E}\Big[\frac{1}{T}\int_0^T \big(\big\|u(X_t,t)\big\|^2 - 2\langle u(X_t,t), u^*(X_t,t)\rangle + \|u^*(X_t,t)\big\|^2\big)\,\mathrm{d}t \\ &\qquad \times \exp\big(-\lambda^{-1}\int_0^T f(X_t,t)\,\mathrm{d}t - \lambda^{-1}g(X_T)\big)\Big].\end{aligned}$$

Clearly, the only optimum of this loss is the optimal control u^*. Using the analytic expression of u^* in Theorem 4.1, the cross-term can be rewritten as (see Theorem C.8 in section C.3):

$$\begin{aligned}&\mathbb{E}\Big[\frac{1}{T}\int_0^T \langle u(X_t,t), u^*(X_t,t)\rangle\,\mathrm{d}t \,\exp\big(-\lambda^{-1}\int_0^T f(X_t,t)\,\mathrm{d}t - \lambda^{-1}g(X_T)\big)\Big]\\ &= -\lambda\mathbb{E}\Big[\frac{1}{T}\int_0^T \Big\langle u(X_t,t), \sigma(t)^\top \nabla_x \mathbb{E}\Big[\exp\big(-\lambda^{-1}\int_t^T f(X_s,s)\,\mathrm{d}s - \lambda^{-1}g(X_T)\big)\Big|X_t = x\Big]\Big\rangle\,\mathrm{d}t\\ &\qquad \times \exp\big(-\lambda^{-1}\int_0^t f(X_s,s)\,\mathrm{d}s\big)\Big]. \qquad (4.21)\end{aligned}$$

It remains to evaluate the conditional expectation $\nabla_x \mathbb{E}\big[\exp\big(-\lambda^{-1}\int_t^T f(X_s,s)\,\mathrm{d}s - \lambda^{-1}g(X_T)\big)\big|X_t = x\big]$, which we do by a "reparameterization trick" that shifts the dependence on the initial value x into the stochastic processes—here we introduce a free variable M_t—and then applying Girsanov theorem. We coin this the *path-wise reparameterization trick*:

Proposition 4.4 (Path-wise reparameterization trick for stochastic optimal control)**.** *For each $t \in [0,T]$, let $M_t : [t,T] \to \mathbb{R}^{d\times d}$ be an arbitrary continuously differentiable function matrix-valued*

function such that $M_t(t) = \mathrm{Id}$. *We have that*

$$\begin{aligned}
&\nabla_x \mathbb{E}\big[\exp\big(-\lambda^{-1}\int_t^T f(X_s,s)\,\mathrm{d}s - \lambda^{-1} g(X_T)\big)\big|X_t = x\big] \\
&= \mathbb{E}\Big[\Big(-\lambda^{-1}\int_t^T M_t(s)\nabla_x f(X_s,s)\,\mathrm{d}s - \lambda^{-1} M_t(T)\nabla g(X_T) \\
&\qquad + \lambda^{-1/2}\int_t^T (M_t(s)\nabla_x b(X_s,s) - \partial_s M_t(s))(\sigma^{-1})^\top(s)\mathrm{d}B_s\Big) \\
&\qquad \times \exp\big(-\lambda^{-1}\int_t^T f(X_s,s)\,\mathrm{d}s - \lambda^{-1} g(X_T)\big)\big|X_t = x\Big]. \qquad (4.22)
\end{aligned}$$

We prove a more general form of this result (Theorem C.10) in subsection C.3.2 and also provide an intuitive derivation in subsection C.3.3. In the proof of Theorem C.10, the reparameterization matrices M_t arise as the gradients of a perturbation to the process X_t. Similar ideas can potentially be applied to derive losses for generative modeling. If we plug (4.22) into the right-hand side of (4.21), and then this back into (4.20), and we complete the square, we obtain that for some constant K independent of u,

$$\begin{aligned}
\tilde{\mathcal{L}}(u) = \mathbb{E}\Big[\frac{1}{T}\int_0^T \Big\|u(X_t,t) + \sigma(t)\Big(\int_t^T M_t(s)\nabla_x f(X_s,s)\,\mathrm{d}s + M_t(T)\nabla g(X_T) \\
- \lambda^{1/2}\int_t^T (M_t(s)\nabla_x b(X_s,s) - \partial_s M_t(s))(\sigma^{-1})^\top(s)\mathrm{d}B_s\Big)\Big\|^2 \mathrm{d}t \\
\times \exp\big(-\lambda^{-1}\int_0^T f(X_t,t)\,\mathrm{d}t - \lambda^{-1} g(X_T)\big)\Big] + K.
\end{aligned}$$

If we perform a change of process from X to X^v applying the Girsanov theorem (Theorem C.3 in section C.3), we obtain the loss $\mathcal{L}_{\mathrm{SOCM}}(u, M)$. □

The following proposition sheds some light onto the role of reparameterization matrices and connects the SOCM loss to the cross-entropy loss.

Proposition 4.5 (Bias-variance decomposition of the SOCM loss)**.** *The SOCM loss decomposes*

into a bias term that only depends on u and a variance term that only depends on M:

$$\mathcal{L}_{\mathrm{SOCM}}(u, M) = \underbrace{\mathbb{E}\Big[\frac{1}{T}\int_0^T \Big\|u(X_t^{u^*}, t) - u^*(X_t^{u^*}, t)\Big\|^2 \,\mathrm{d}t\, \exp(-\lambda^{-1} V(X_0^{u^*}, 0))\Big]}_{\textit{Bias of } u} + \underbrace{\mathrm{CondVar}(w; M)}_{\textit{Variance of } w}, \tag{4.23}$$

where

$$\begin{aligned}
&\mathrm{Var}(w; M) \\
&= \mathbb{E}\Big[\frac{1}{T}\int_0^T \Big\|\tilde{w}(t, X, B, M_t) - \frac{\mathbb{E}[\tilde{w}(t, X, B, M_t)\exp(-\lambda^{-1}\mathcal{W}(X, 0))|X_t]}{\mathbb{E}[\exp(-\lambda^{-1}\mathcal{W}(X, 0))|X_t]}\Big\|^2 \mathrm{d}t\, \exp(-\lambda^{-1}\mathcal{W}(X, 0))\Big] \\
&= \mathbb{E}\Big[\frac{1}{T}\int_0^T \Big\|w(t, v, X^v, B, M_t) - \frac{\mathbb{E}[w(t, v, X^v, B, M_t)\alpha(v, X^v, B)|X_t^v]}{\mathbb{E}[\alpha(v, X^v, B)|X_t^v]}\Big\|^2 \mathrm{d}t\, \alpha(v, X^v, B)\Big],
\end{aligned} \tag{4.24}$$

and

$$\begin{aligned}
\tilde{w}(t, X, B, M_t) = \sigma(t)^\top \Big(&- \int_t^T M_t(s)\nabla_x f(X_s, s)\,\mathrm{d}s - M_t(T)\nabla g(X_T) \\
&+ \lambda^{1/2}\int_t^T (M_t(s)\nabla_x b(X_s, s) - \partial_s M_t(s))(\sigma^{-1})^\top(s)\mathrm{d}B_s\Big).
\end{aligned} \tag{4.25}$$

Remark that the bias term in equation (4.23) is equal to the characterization of the cross-entropy loss in Theorem 4.2. In other words, the landscape of $\mathcal{L}_{\mathrm{SOCM}}(u, M)$ with respect to u is the landscape of the cross-entropy loss $\mathcal{L}_{\mathrm{CE}}(u)$. Thus, the SOCM loss can be seen as some form of variance reduction method for the cross-entropy loss, and performs substantially better experimentally (section 4.4). Yet, the expressions of the SOCM loss and the cross-entropy loss are very different; the former is a least squares loss and is expressed in terms of the gradients of the costs.

For good training performance, it is critical that the gradients have high signal-to-noise ratio. Looking at the SOCM loss, a good proxy for low gradient variance is to have low variance for $\frac{1}{T}\int_0^T \Big\|u(X_t^v, t) - w(t, v, X^v, B, M_t)\Big\|^2 \mathrm{d}t \times \alpha(v, X^v, B)$, and this holds when both $\alpha(v, X^v, B)$ and $w(t, v, X^v, B, M_t)$ have low variance. Next, we present strategies to lower the variance of these

Algorithm 4 Stochastic Optimal Control Matching (SOCM)

Input: State cost $f(x,t)$, terminal cost $g(x)$, covariance matrix $\sigma(t)$, base drift $b(x,t)$, noise level λ, number of iterations N, batch size m, number of time steps K, initial control parameters θ_0, initial matrix parameters ω_0, loss $\mathcal{L}_{\text{SOCM}}$ in (C.31)

for $n \in \{0, \dots, N-1\}$ **do**
- Simulate m trajectories of the process X^v controlled by $v = u_{\theta_n}$, e.g., using Euler-Maruyama updates
- Detach the m trajectories from the computational graph, so that gradients do not backpropagate
- Using the m trajectories, compute an m-sample Monte-Carlo approximation $\hat{\mathcal{L}}_{\text{SOCM}}(u_{\theta_n}, M_{\omega_n})$ of the loss $\mathcal{L}_{\text{SOCM}}(u_{\theta_n}, M_{\omega_n})$ in (C.31)
- Compute the gradients $\nabla_{(\theta,\omega)} \hat{\mathcal{L}}_{\text{SOCM}}(u_{\theta_n}, M_{\omega_n})$ of $\hat{\mathcal{L}}_{\text{SOCM}}(u_{\theta_n}, M_{\omega_n})$ at (θ_n, ω_n)
- Obtain θ_{n+1}, ω_{n+1} with via an Adam update on θ_n, ω_n, resp.

end

Output: Learned control u_{θ_N}

two objects.

Minimizing the variance of the importance weight α We want to use a vector field v such that $\text{Var}[\alpha(v, X^v, B)]$ is as low as possible. As shown by the following lemma, which is well-known in the literature, setting v to be the optimal control u^* actually achieves variance zero when we condition on the starting point of the controlled process X^v. The proof of this result can be found in Hartmann et al. [2017], but we include it in subsection C.3.5 for completeness.

Lemma 4.6. *When we set $v = u^*$, the conditional variance $\text{Var}[\alpha(v, X^v, B) | X_0^v = x_{\text{init}}]$ is zero for any $x_{\text{init}} \in \mathbb{R}^d$.*

Of course, we do not have access to the optimal control u^*, but it is still a good idea to set v as the closest vector field to u^* that we have access to, which is typically the currently learned control. In some instances, one may benefit from using a warm-started control parameterized as $u_{\text{WS}}(x,t) + u_\theta(x,t)$, where the warm-start u_{WS} is a reasonably good control obtained via a different strategy (see section C.4).

Minimizing the variance of the matching vector field w We are interested in finding the family $M = (M_t)_{t \in [0,T]}$ that minimizes the variance of $w(t, v, X^v, B, M_t)$ conditioned on t and X_t. Note that this is exactly the term $\text{CondVar}(w; M)$ in the right-hand side of equation (4.23). Since

$\mathrm{CondVar}(w; M)$ does not depend on the specific v, the optimal M does not depend on v either. And since the first term in the right-hand side of equation (4.23) does not depend on $M = (M_t)_{t\in[0,T]}$, minimizing $\mathrm{CondVar}(w; M)$ is equivalent to minimizing $\mathcal{L}(u)$ with respect to M. In practice, we parameterize M using a neural network with a two-dimensional input (t, s) and a d^2-dimensional output.

Furthermore, the following theorem shows that the optimal family $M^* = (M_t^*)_{t\in[0,T]}$ can be characterized as the solution of a linear equation in infinite dimensions. The proof is in subsection C.3.6.

Theorem 4.7 (Optimal reparameterization matrices)**.** *Let v be an arbitrary control in $\mathcal{U}$. Define the integral operator $\mathcal{T}_t : L^2([t,T];\mathbb{R}^{d\times d}) \to L^2([t,T];\mathbb{R}^{d\times d})$ as*

$$[\mathcal{T}_t(\dot{M}_t)](s) = \int_t^T \dot{M}_t(s')\mathbb{E}\big[\chi(s', X^v, B)\chi(s, X^v, B)^\top \times \alpha(v, X^v, B)\big]\,\mathrm{d}s',$$

where

$$\begin{aligned}\chi(t, X^v, B) := &\int_t^T \nabla_x f(X_s^v, s)\,\mathrm{d}s + \nabla g(X_T^v) + (\sigma_t^{-1})^\top(t)v(X_t^v, t)\\ &- \int_t^T \nabla_x b(X_s^v, s)(\sigma_s^{-1})^\top(s)v(X_t^v, t)\,\mathrm{d}s - \int_t^T \nabla_x b(X_s^v, s)(\sigma_s^{-1})^\top(s)\,\mathrm{d}B_s.\end{aligned}$$

If we define $N_t(s) = -\mathbb{E}\big[\big(\nabla g(X_T^v) + \int_t^T \nabla_x f(X_{s'}^v, s')\,\mathrm{d}s'\big)\chi(t, X^v, B)^\top \times \alpha(v, X^v, B)\big]$, the optimal $M^* = (M_t^*)_{t\in[0,T]}$ is of the form $M_t^*(s) = I + \int_t^s \dot{M}_t^*(s')\,\mathrm{d}s'$, where $\dot{M}_t^*$ is the unique solution of the following Fredholm equation of the first kind:

$$\mathcal{T}_t(\dot{M}_t) = N_t. \tag{4.26}$$

Solving the Fredholm equation (4.26) numerically is expensive, as the discretized linear system has d^2K equations and variables, K being the number of discretization time points. However,

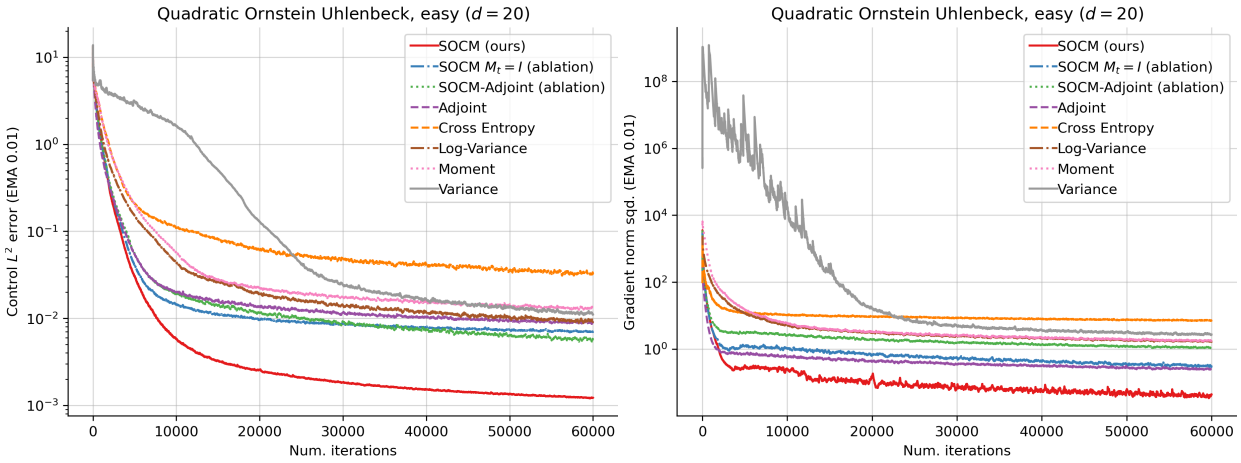


Figure 4.1: Plots of the L^2 error incurred by the learned control (*top*), and the norm squared of the gradient with respect to the parameters θ of the control (*bottom*), for the Quadratic Ornstein Uhlenbeck (easy) setting and for each IDO loss. Both plots show exponential moving averages computed from the trajectories used during training.

since the optimal M^* does not depend on v, this is a computation that must be done only once and that may be affordable in some settings.

Parameterizing the matrices M_t In practice, we parameterize the matrices $(M_t)_{t\in[0,T]}$ using a common function M_ω with two arguments (t, s). A simple way to enforce that $M_\omega(t, t) = \mathrm{Id}$ is to set $M_\omega(t, s) = e^{-\gamma(s-t)}\mathrm{Id} + (1 - e^{-\gamma(s-t)})\tilde{M}_{\tilde{\omega}}(t, s)$, where $\omega = (\gamma, \tilde{\omega})$, and $\tilde{M}_{\tilde{\omega}} : \mathbb{R} \times \mathbb{R} \to \mathbb{R}^{d\times d}$ is an unconstrained neural network.

4.4 Experiments

We consider four experimental settings that we adapt from Nüsken and Richter [2021]: Quadratic Ornstein Uhlenbeck (easy), Quadratic Ornstein Uhlenbeck (hard), Linear Ornstein Uhlenbeck and Double Well. We describe them in detail in section C.5. For all of them, we have access to the ground-truth optimal control, which means that we are able to estimate the L^2 error incurred by the learned control u.

In Figure 4.1 (*top*) we plot the control L^2 error for each IDO algorithm described in subsection 4.2.2,

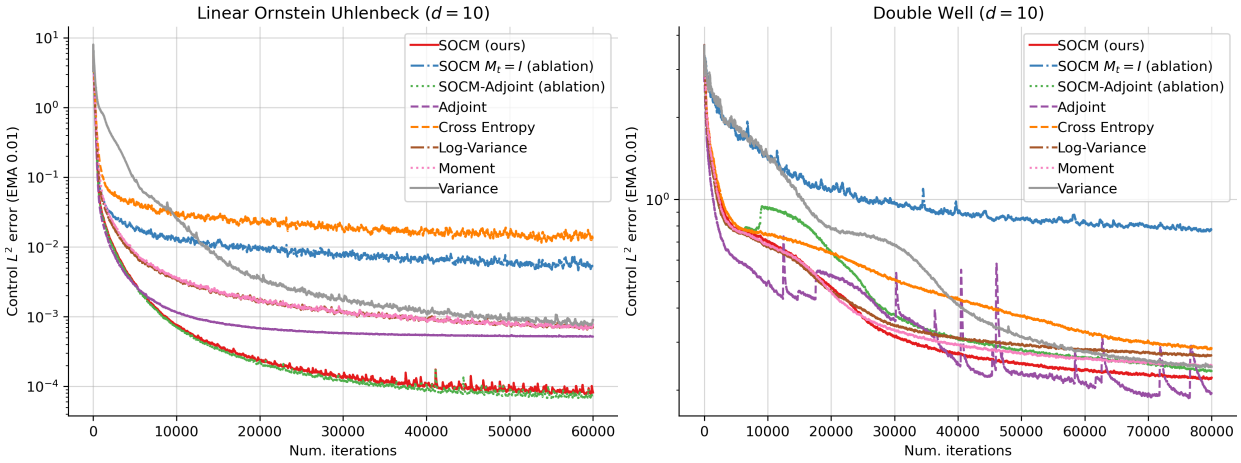


Figure 4.2: Plots of the L^2 error of the learned control for the Linear Ornstein Uhlenbeck and Double Well settings.

SOCM	SOCM $M_t = I$	SOCM adjoint	Adjoint
0.222	0.090	0.099	0.169
Cross entropy	Log-variance	Moment	Variance
0.086	0.117	0.087	0.086

Table 4.1: Time per iteration (exponential moving average) for various algorithms in seconds per iteration, for the Quadratic OU (easy) experiments (Figure 4.1).

and for the SOCM algorithm (Algorithm 4), for the Quadratic OU (easy) setting. We also include two ablations of SOCM: *(i)* a version of SOCM where the reparameterization matrices M_t are set fixed to the identity I, *(ii)* SOCM-Adjoint, where we estimate the conditional expectation in equation (4.22) using the adjoint method for SDEs instead of the path-wise reparameterization trick (see subsection C.3.4).

At the end of training, SOCM obtains the lowest L^2 error, improving over all existing methods by a factor of around ten. The two SOCM ablations come in second and third by a substantial difference, which underlines the importance of the path-wise reparameterization trick. The best among existing methods is the adjoint method (the relative entropy loss). In Figure 4.1 (*bottom*) we show the squared norm of the gradient of each loss with respect to the parameters θ of the control: algorithms with small noise variance tend to have low error values. Table 4.1 shows the average times per iteration for each algorithm.

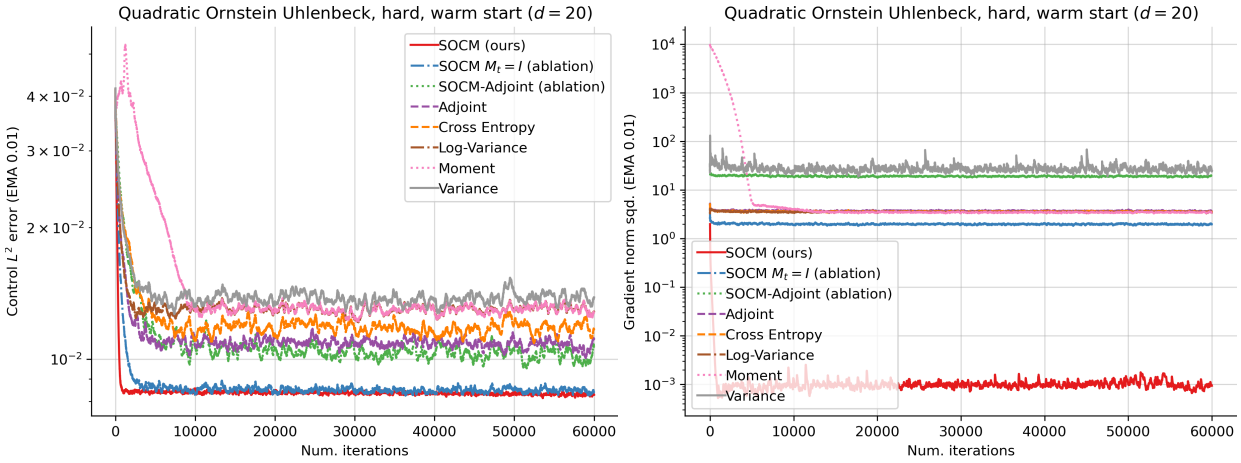


Figure 4.3: Plots of the L^2 error incurred by the learned control (*top*), and the norm squared of the gradient with respect to the parameters θ of the control (*bottom*), for the Quadratic Ornstein Uhlenbeck (hard) setting and for each IDO loss. All the algorithms use a warm-started control (see section C.4).

In Figure 4.2, we plot the control L^2 error for Linear Ornstein Uhlenbeck and Double Well. For Linear OU, the error is around five times smaller for SOCM than for any existing method. For Double Well, the SOCM algorithm achieves the second smallest error, slightly behind the adjoint method, but the latter shows instabilities. As we show in Figure C.4 in section C.5, these instabilities are inherent to the adjoint method and they do not disappear for small learning rates. Both in Figure 4.1 and Figure C.4, we observe that learning the reparameterization matrices is critical to obtain gradient estimates with high signal-to-noise ratio.

The costs f and g and the base drift b for Quadratic OU (hard) are five times those of Quadratic OU (easy). Consequently, the factor $\alpha(v, X^v, B)$ has a much larger variance, and initializing the control neural network without a warm-start yields poor results for the SOCM and cross-entropy losses (see Figure C.5 in section C.5). Yet, when we use the control warm-start strategy detailed in section C.4, Figure 4.3 shows that SOCM is once again the algorithm that achieves the lowest error and the smallest gradients. Remark that the warm-start control is a reasonable approximation of the optimal control, as the initial control L^2 error is much lower than in the other figures.

4.5 Conclusion

Our work introduces Stochastic Optimal Control Matching, a novel Iterative Diffusion Optimization technique for stochastic optimal control that stems from the same philosophy as the conditional score matching loss for diffusion models. That is, the control is learned via a least-squares problem by trying to fit a matching vector field. The training loss is optimized with respect to both the control function and a family of reparameterization matrices which appear in the matching vector field. The optimization with respect to the reparameterization matrices aims at minimizing the variance of the matching vector field. Experimentally, our algorithm achieves lower error than all the existing IDO techniques for stochastic optimal control for four different control settings.

One of the key ideas for deriving the SOCM algorithm is the path-wise reparameterization trick, a novel technique to obtain low-variance estimates of the gradient of the conditional expectation of a functional of a random process with respect to its initial value. An interesting future direction is to use the path-wise reparameterization trick to decrease the variance of the matching vector field for diffusion models.

The main roadblock when we try to apply SOCM to more challenging problems is that the variance of the factor $\alpha(v, X^v, B)$ explodes when f and/or g are large, or when the dimension d is high. We observe this in Figure C.5 in section C.5, which is for the Quadratic Ornstein Uhlenbeck (hard) setting but does not use warm-start. The control L^2 error for the SOCM and cross-entropy losses remains high and fluctuates heavily due to the large variance of α. The large variance of α is due to the mismatch between the probability measures induced by the learned control and the optimal control. Similar problems are encountered in out-of-distribution generalization for reinforcement learning, and some approaches may be carried over from that area [Munos et al. 2016].

5 Discussion

This thesis contains three works that study or propose machine learning algorithms from the lens of evolution dynamics on probability measures.

Looking ahead, several areas warrant further investigation. For the mean-field two-player dynamics, a reasonable goal is to get a better understanding of the convergence guarantees of the proposed algorithms. In the realm of Multisample Flow Matching, experimenting with different types of couplings and exploring the possibility of performing conditional generation are interesting directions. Lastly, for Stochastic Optimal Control Matching, the application of the reparameterization trick to other types of problems or its integration with reinforcement learning frameworks could be promising directions.

In conclusion, this thesis not only demonstrates the feasibility and effectiveness of novel approaches to managing dynamics on probability measures but also opens up numerous avenues of research. The methodologies developed here are poised to influence a range of applications, from deep learning and artificial intelligence to complex system optimization and beyond, underscoring the transformative potential of advanced mathematical frameworks in tackling real-world challenges.

A | Appendix: Mean-field two-player zero-sum games

A.1 Lifted dynamics for the Interacting Wasserstein-Fisher-Rao Gradient Flow

Recall the IWFRGF in (2.8), which we reproduce here for convenience.

$$\begin{cases} \partial_t \mu_x &= \gamma \nabla_x \cdot (\mu_x \nabla_x V_x(\mu_y, x)) - \alpha \mu_x (V_x(\mu_y, x) - \mathcal{L}(\mu_x, \mu_y)), \quad \mu_x(0) = \mu_{x,0} \\ \partial_t \mu_y &= -\gamma \nabla_y \cdot (\mu_y \nabla_y V_y(\mu_x, y)) + \alpha \mu_y (V_y(\mu_x, y) - \mathcal{L}(\mu_x, \mu_y)), \quad \mu_y(0) = \mu_{y,0} \end{cases}$$

Given $\nu_x \in \mathcal{P}(\mathcal{X} \times \mathbb{R}^+)$ define $\mu_x = \int_{\mathcal{X}} w_x \, d\nu_x(\cdot, w_x) \in \mathcal{P}(\mathcal{X})$, that is

$$\int_{\mathcal{X}} \psi(x) \, d\mu_x(x) = \int_{\mathcal{X} \times \mathbb{R}^+} w_x \psi(x) \, d\nu_x(x, w_x),$$

for all $\psi \in C(\mathcal{X})$. Given $\nu_y \in \mathcal{P}(\mathcal{Y} \times \mathbb{R}^+)$, define $\mu_y = \int_{\mathcal{X}} w_y \, d\nu_y(\cdot, w_y) \in \mathcal{P}(\mathcal{Y})$ analogously. We say that ν_x, ν_y are "lifted" measures of μ_x, μ_y, and reciprocally μ_x, μ_y are "projected" measures of ν_x, ν_y.

By Theorem A.1 below, we can view a solution of (2.8) as the projection of a solution of the

following dynamics on the lifted domains $\mathcal{X} \times \mathbb{R}^+$ and $\mathcal{Y} \times \mathbb{R}^+$:

$$\begin{cases} \partial_t \nu_x = \nabla_{w_x,x} \cdot (\nu_x g_{\mu_y}(x, w_x)), \quad \nu_x(0) = \mu_{x,0} \times \delta_{w_x=1} \\ \partial_t \nu_y = -\nabla_{w_y,y} \cdot (\nu_y g_{\mu_x}(y, w_y)), \quad \nu_y(0) = \mu_{y,0} \times \delta_{w_y=1} \end{cases} \tag{A.1}$$

where

$$g_{\mu_y}(x, w_x) = (\alpha w_x (V_x(\mu_y, x) - \mathcal{L}(\mu_x, \mu_y)), \gamma \nabla_x V_x(\mu_y, x))),$$
$$g_{\mu_x}(y, w_y) = (\alpha w_y (V_y(\mu_x, x) - \mathcal{L}(\mu_x, \mu_y)), \gamma \nabla_y V_y(\mu_x, y))).$$

Lemma A.1. *For a solution $\nu_x : [0,T] \to \mathcal{P}(\mathcal{X} \times \mathbb{R}^+), \nu_y : [0,T] \to \mathcal{P}(\mathcal{Y} \times \mathbb{R}^+)$ of (A.1), the projections μ_x, μ_y are solutions of (2.8).*

That is, given any $\psi_x \in C^1(\mathcal{X}), \psi_y \in C^1(\mathcal{Y})$, we have

$$\begin{aligned} \frac{d}{dt} \int_{\mathcal{X}} \psi_x(x)\, d\mu_x &= -\gamma \int_{\mathcal{X}} \nabla_x \psi_x(x) \cdot \nabla_x V_x(\mu_y, x)\, d\mu_x - \alpha \int_{\mathcal{X}} \psi_x(x)(V_x(\mu_y, x) - \mathcal{L}(\mu_x, \mu_y))\, d\mu_x, \\ \frac{d}{dt} \int_{\mathcal{Y}} \psi_y(y)\, d\mu_y &= \gamma \int_{\mathcal{Y}} \nabla_y \psi_y(y) \cdot \nabla_y V_y(\mu_x, y))\, d\mu_y + \alpha \int_{\mathcal{Y}} \psi_y(y)(V_y(\mu_x, y) - \mathcal{L}(\mu_x, \mu_y))\, d\mu_y, \\ \mu_x(0) &= \mu_{x,0}, \quad \mu_y(0) = \mu_{y,0} \end{aligned} \tag{A.2}$$

From (A.1) in the weak form, we obtain that given any $\psi_x \in C^1(\mathcal{X} \times \mathbb{R}^+), \psi_y \in C^1(\mathcal{Y} \times \mathbb{R}^+)$,

$$\begin{aligned} \frac{d}{dt} \int_{\mathcal{X}\times\mathbb{R}^+} \psi_x(x, w_x)\, d\nu_x(x, w_x) &= \int_{\mathcal{X}\times\mathbb{R}^+} -\gamma \nabla_x \psi_x(x, w_x) \cdot \nabla_x V_x(\mu_y, x) \\ &\quad - \alpha w_x \frac{d\psi_x}{dw_x}(x, w_x)(V_x(\mu_y, x) - \mathcal{L}(\mu_x, \mu_y))\, d\mu_x, \\ \frac{d}{dt} \int_{\mathcal{Y}\times\mathbb{R}^+} \psi_y(y, w_y)\, d\nu_y(y, w_y) &= \int_{\mathcal{Y}\times\mathbb{R}^+} \gamma \nabla_y \psi_y(y, w_y) \cdot \nabla_y V_y(\mu_x, y) \\ &\quad + \alpha w_y \frac{d\psi_y}{dw_y}(y, w_y)(V_y(\mu_x, y) - \mathcal{L}(\mu_x, \mu_y))\, d\mu_y, \\ \nu_x(0) = \mu_{x,0} \times \delta_{w_x=1}, &\quad \nu_y(0) = \mu_{y,0} \times \delta_{w_y=1}. \end{aligned} \tag{A.3}$$

Taking $\psi_x(x, w_x) = w_x\psi_x(x), \psi_y(y, w_y) = w_y\psi_y(y)$ yields

$$\begin{aligned}
\frac{d}{dt}\int_{\mathcal{X}\times\mathbb{R}^+} w_x\psi_x(x)\; d\nu_x(x, w_x) &= \int_{\mathcal{X}\times\mathbb{R}^+} -\gamma w_x \nabla_x\psi_x(x)\cdot\nabla_x V_x(\mu_y, x) \\
&\quad - \alpha w_x\psi_x(x)(V_x(\mu_y, x) - \mathcal{L}(\mu_x, \mu_y))\; d\mu_x, \\
\frac{d}{dt}\int_{\mathcal{Y}\times\mathbb{R}^+} w_y\psi_y(y, w_y)\; d\nu_y(y, w_y) &= \int_{\mathcal{Y}\times\mathbb{R}^+} \gamma w_y \nabla_y\psi_y(y)\cdot\nabla_y V_y(\mu_x, y) \\
&\quad + \alpha w_y\psi_y(y)(V_y(\mu_x, y) - \mathcal{L}(\mu_x, \mu_y))\; d\mu_y. \qquad \text{(A.4)}
\end{aligned}$$

Notice that (A.4) is indeed (A.2).

A.2 Continuity and convergence properties of the Nikaido-Isoda error

Lemma A.2. *The Nikaido-Isoda error $NI : \mathcal{P}(\mathcal{X})\times\mathcal{P}(\mathcal{Y}) \to \mathbb{R}$ defined in (2.2) is continuous when we endow $\mathcal{P}(\mathcal{X}), \mathcal{P}(\mathcal{Y})$ with the topology of weak convergence. Specifically, it is Lip(ℓ)-Lipschitz when we use the distance $\mathcal{W}_1(\mu_x, \mu'_x) + \mathcal{W}_1(\mu_y, \mu'_y)$ between (μ_x, μ_y) and (μ'_x, μ'_y) in $\mathcal{P}(\mathcal{X})\times\mathcal{P}(\mathcal{Y})$.*

Proof. For any μ_y, the function $V_x(\mu_y, \cdot) : \mathcal{X} \to \mathbb{R}$ defined as $x \mapsto \int \ell(x, y)\; d\mu_y$ is continuous and it has the same Lipschitz constant $\mathrm{Lip}(\ell)$ as ℓ. Hence, for any $\mu_x, \mu'_x \in \mathcal{P}(\mathcal{X})$,

$$\begin{aligned}
&\sup_{\mu_y\in\mathcal{P}(\mathcal{Y})} \mathcal{L}(\mu_x, \mu_y) - \sup_{\mu_y\in\mathcal{P}(\mathcal{Y})} \mathcal{L}(\mu'_x, \mu_y) = \sup_{\mu_y\in\mathcal{P}(\mathcal{Y})} \int V_x(\mu_y, x) d\mu_x - \sup_{\mu_y\in\mathcal{P}(\mathcal{Y})} \int V_x(\mu_y, x) d\mu'_x \\
&\leq \sup_{\mu_y\in\mathcal{P}(\mathcal{Y})} \int V_x(\mu_y, x) d\mu'_x + \sup_{\mu_y\in\mathcal{P}(\mathcal{Y})} \int V_x(\mu_y, x) d(\mu_x - \mu'_x) - \sup_{\mu_y\in\mathcal{P}(\mathcal{Y})} \int V_x(\mu_y, x) d\mu'_x \\
&= \sup_{\mu_y\in\mathcal{P}(\mathcal{Y})} \int V_x(\mu_y, x) d(\mu_x - \mu'_x) \leq \mathrm{Lip}(\ell)\mathcal{W}_1(\mu_x, \mu'_x)
\end{aligned}$$

The same inequality interchanging the roles of μ_x, μ_x' shows that

$$|\sup_{\mu_y \in \mathcal{P}(\mathcal{Y})} \mathcal{L}(\mu_x, \mu_y) - \sup_{\mu_y \in \mathcal{P}(\mathcal{Y})} \mathcal{L}(\mu_x', \mu_y)| \leq \mathrm{Lip}(\ell)\mathcal{W}_1(\mu_x, \mu_x')$$

holds. An analogous reasoning for $\ell(\mu_x, \cdot) : \mathcal{Y} \to \mathbb{R}$ and the triangle inequality complete the proof. □

Lemma A.3. *Suppose that $(\mu_x^n)_{n\in\mathbb{N}}$ is a sequence of random elements valued in $\mathcal{P}(X)$ such that*

$$\mathbb{E}[\mathcal{W}_2^2(\mu_x^n, \mu_x)] \xrightarrow{n\to\infty} 0,$$

where $\mu_x \in \mathcal{P}(X)$. Analogously, suppose that $(\mu_y^n)_{n\in\mathbb{N}}$ is a sequence of random elements valued in $\mathcal{P}(\mathcal{Y})$ such that

$$\mathbb{E}[\mathcal{W}_2^2(\mu_y^n, \mu_y)] \xrightarrow{n\to\infty} 0,$$

where $\mu_y \in \mathcal{P}(Y)$.

Then,

$$\mathbb{E}[|NI(\mu_x^n, \mu_y^n) - NI(\mu_x, \mu_y)|] \xrightarrow{n\to\infty} 0$$

Proof. First,

$$\mathbb{E}[\mathcal{W}_1(\mu_x^n, \mu_x)] \leq \mathbb{E}[\mathcal{W}_2(\mu_x^n, \mu_x)] \leq \left(\mathbb{E}[\mathcal{W}_2^2(\mu_x^n, \mu_x)]\right)^{1/2}, \tag{A.5}$$

which results from two applications of the Cauchy-Schwarz inequality on the appropriate scalar products. An analogous inequality holds for $\mathbb{E}[\mathcal{W}_1(\mu_y^n, \mu_y)]$. Hence, by Theorem A.2,

$$\begin{aligned}
\mathbb{E}[|\mathrm{NI}(\mu_x^n, \mu_y^n) - \mathrm{NI}(\mu_x, \mu_y)|] &\leq \mathrm{Lip}(\ell)\mathbb{E}[\mathcal{W}_1(\mu_x^n, \mu_x) + \mathcal{W}_1(\mu_y^n, \mu_y)] \\
&\leq \mathrm{Lip}(\ell)\left(\left(\mathbb{E}[\mathcal{W}_2^2(\mu_x^n, \mu_x)]\right)^{1/2} + \left(\mathbb{E}[\mathcal{W}_2^2(\mu_x^n, \mu_x)]\right)^{1/2}\right) \\
&\leq \mathrm{Lip}(\ell)\sqrt{2}\left(\mathbb{E}[\mathcal{W}_2^2(\mu_x^n, \mu_x)] + \mathbb{E}[\mathcal{W}_2^2(\mu_x^n, \mu_x)]\right)^{1/2},
\end{aligned}$$

where the second inequality uses (A.5) and the third inequality is another application of the Cauchy-Schwarz inequality. Since the right hand side converges to 0 by assumption, this concludes the proof. □

A.3 Proof of Theorem 2.2

We restate Theorem 2.2 for convenience.

Theorem 2.2. *Suppose that Assumption 1 holds, that $\ell \in C^{2,\alpha}(\mathcal{X} \times \mathcal{Y})$ for some $\alpha \in (0,1)$ and that the initial measures $\mu_{x,0}, \mu_{y,0}$ have densities in $L^1(\mathcal{X}), L^1(\mathcal{Y})$. If a solution $(\mu_x(t), \mu_y(t))$ of the ERIWGF (2.7) converges in time, it must converge to the point $(\hat{\mu}_x, \hat{\mu}_y)$ which is the unique fixed point of the problem*

$$\rho_x(x) = \frac{1}{Z_x} e^{-\beta \int \ell(x,y)\, d\mu_y(y)}, \quad \rho_y(y) = \frac{1}{Z_y} e^{\beta \int \ell(x,y)\, d\mu_x(x)}.$$

$(\hat{\mu}_x, \hat{\mu}_y)$ is an ϵ-Nash equilibrium of the game given by $\mathcal{L}$ when $\beta \geq \frac{4}{\epsilon} \log\left(2\frac{1-V_\delta}{V_\delta}(2K_\ell/\epsilon - 1)\right)$, where $K_\ell := \max_{x,y} \ell(x,y) - \min_{x,y} \ell(x,y)$ is the length of the range of ℓ, $\delta := \epsilon/(2Lip(\ell))$ and V_δ is a lower bound on the volume of a ball of radius δ in $\mathcal{X}, \mathcal{Y}$.

Theorem 2.2 is a consequence of the following three results, which we prove separately.

Theorem A.4. *Assume $\mathcal{X}, \mathcal{Y}$ are compact Polish metric spaces equipped with canonical Borel measures, and that ℓ is a continuous function on $\mathcal{X} \times \mathcal{Y}$. Let us consider the fixed point problem*

$$\begin{cases} \rho_x(x) &= \frac{1}{Z_x} e^{-\beta \int \ell(x,y)\, d\mu_y(y)}, \\ \rho_y(y) &= \frac{1}{Z_y} e^{\beta \int \ell(x,y)\, d\mu_x(x)}, \end{cases}$$

where Z_x and Z_y are normalization constants and ρ_x, ρ_y are the densities of μ_x, μ_y. This fixed point problem has a unique solution $(\hat{\mu}_x, \hat{\mu}_y)$ that is also the unique Nash equilibrium of the game given

by $\mathcal{L}_\beta(\mu_x, \mu_y) \triangleq \mathcal{L}(\mu_x, \mu_y) + \beta^{-1}(H(\mu_y) - H(\mu_x))$.

Theorem A.5. *Let* $K_\ell := \max_{x,y} \ell(x,y) - \min_{x,y} \ell(x,y)$ *be the length of the range of* ℓ*. Let* $\epsilon > 0$*,* $\delta := \epsilon/(2Lip(\ell))$ *and* V_δ *be a lower bound on the volume of a ball of radius* δ *in* $\mathcal{X}, \mathcal{Y}$*. Then the solution* $(\hat{\mu}_x, \hat{\mu}_y)$ *of* (2.9) *is an* ϵ*-Nash equilibrium of the game given by* $\mathcal{L}$ *when*

$$\beta \geq \frac{4}{\epsilon} \log\left(2\frac{1 - V_\delta}{V_\delta}(2K_\ell/\epsilon - 1)\right).$$

Theorem A.6. *Suppose that Assumption 1 holds and* $\ell \in C^{2,\alpha}(\mathcal{X} \times \mathcal{Y})$ *for some* $\alpha \in (0,1)$*, i.e. the second derivatives of* ℓ *are* α*-Hölder. Then, there exists only one stationary solution of the ERIWGF* (2.7) *and it is the solution of the fixed point problem* (2.9)*.*

A.3.1 Proof of Theorem A.4: Preliminaries

Definition A.7 (Upper hemicontinuity)**.** A set-valued function $\psi : X \to 2^Y$ is upper hemicontinuous if for every open set $W \subset Y$, the set $\{x | \psi(x) \subset W\}$ is open.

Alternatively, set-valued functions can be seen as correspondences $\Gamma : X \to Y$. The graph of Γ is $\mathrm{Gr}(\Gamma) = \{(a,b) \in X \times Y | b \in \Gamma(a)\}$. If Γ is upper hemicontinuous, then $\mathrm{Gr}(\Gamma)$ is closed. If Y is compact, the converse is also true.

Definition A.8 (Kakutani map)**.** Let X and Y be topological vector spaces and $\psi : X \to 2^Y$ be a set-valued function. If Y is convex, then ψ is termed a Kakutani map if it is upper hemicontinuous and $\psi(x)$ is non-empty, compact and convex for all $x \in X$.

Theorem A.9 (Kakutani-Glicksberg-Fan)**.** *Let* S *be a non-empty, compact and convex subset of a Hausdorff locally convex topological vector space. Let* $\psi : S \to 2^S$ *be a Kakutani map. Then* ψ *has a fixed point.*

Definition A.10 (Lower semi-continuity)**.** Suppose X is a topological space, x_0 is a point in X and $f : X \to \mathbb{R} \cup \{-\infty, \infty\}$ is an extended real-valued function. We say that f is lower semi-continuous

(l.s.c.) at x_0 if for every $\epsilon > 0$ there exists a neighborhood U of x_0 such that $f(x) \geq f(x_0) - \epsilon$ for all x in U when $f(x_0) < +\infty$, and $f(x)$ tends to $+\infty$ as x tends towards x_0 when $f(x_0) = +\infty$.

We can also characterize lower-semicontinuity in terms of level sets. A function is lower semicontinuous if and only if all of its lower level sets $\{x \in X : f(x) \leq \alpha\}$ are closed. This property will be useful.

Theorem A.11 (Weierstrass theorem for l.s.c. functions)**.** *Let $f : T \to (-\infty, +\infty]$ be a l.s.c. function on a compact Hausdorff topological space T. Then f attains its infimum over T, i.e. there exists a minimum of f in T.*

Proof. Proof. Let $\alpha_0 = \inf f(T)$. If $\alpha_0 = +\infty$, then f is infinite and the assertion trivially holds. Let $\alpha_0 < +\infty$. Then, for each real $\alpha > \alpha_0$, the set $\{f \leq \alpha\}$ is closed and nonempty. Any finite collection of such sets has a nonempty intersection. By compactness, also the set $\bigcap_{\alpha > \alpha_0} \{f \leq \alpha\} = \{f \leq \alpha_0\} = f^{-1}(\alpha_0)$ is nonempty. (In particular, this implies that α_0 is finite.) □

Remark 1. *By Prokhorov's theorem, since $\mathcal{X}$ and $\mathcal{Y}$ are compact separable metric spaces, $\mathcal{P}(\mathcal{X})$ and $\mathcal{P}(\mathcal{Y})$ are compact in the topology of weak convergence.*

A.3.2 Proof of Theorem A.4: Existence

Theorem A.12 and A.13 are intermediate results, and Theorem A.14 shows existence of the solution.

Lemma A.12. *For any $\mu_y \in \mathcal{P}(\mathcal{Y})$, $\mathcal{L}_\beta(\cdot, \mu_y) : \mathcal{P}(\mathcal{X}) \to \mathbb{R}$ is lower semicontinuous, and it achieves a unique minimum in $\mathcal{P}(\mathcal{X})$. Moreover, the minimum $m_x(\mu_y)$ is absolutely continuous with respect to the Borel measure, it has full support and its density takes the form*

$$\frac{dm_x(\mu_y)}{dx}(x) = \frac{1}{Z_{\mu_y}} e^{-\beta \int L(x,y) d\mu_y}, \tag{A.6}$$

where Z_{μ_y} is a normalization constant.

Analogously, for any $\mu_x \in \mathcal{P}(\mathcal{X})$, $-\mathcal{L}_\beta(\mu_x, \cdot) : \mathcal{P}(\mathcal{Y}) \to \mathbb{R}$ is lower semicontinuous, and it achieves a unique minimum in $\mathcal{P}(\mathcal{Y})$. The minimum $m_y(\mu_x)$ is absolutely continuous with respect to the Borel measure, it has full support and its density takes the form

$$\frac{dm_y(\mu_x)}{dy}(y) = \frac{1}{Z_{\mu_x}} e^{\beta \int L(x,y) d\mu_x},$$

where Z_{μ_x} is a normalization constant.

Proof. We will prove the result for $\mathcal{L}_\beta(\cdot, \mu_y)$, as the other one is analogous. Let dx denote the canonical Borel measure on $\mathcal{X}$, and let $\tilde{p}$ be the probability measure proportional to the canonical Borel measure, i.e. $\frac{d\tilde{p}}{dx} = \frac{1}{\text{vol}(\mathcal{X})}$. Notice that $\text{vol}(\mathcal{X})$ is by definition the value of the canonical Borel measure on the whole $\mathcal{X}$. We rewrite

$$\begin{aligned}
\mathcal{L}_\beta(\mu_x, \mu_y) &= \iint \ell(x,y) d\mu_y d\mu_x + \beta^{-1} \int \log\left(\frac{d\mu_x}{dx}\right) d\mu_x + \beta^{-1} H(\mu_y) \\
&= \iint \ell(x,y) d\mu_y d\mu_x + \beta^{-1} \int \log\left(\frac{d\mu_x}{d\tilde{p}} \frac{d\tilde{p}}{dx}\right) d\mu_x + \beta^{-1} H(\mu_y) \\
&= \iint \left(\ell(x,y) - \beta^{-1} \log\left(\text{vol}(\mathcal{X})\right)\right) d\mu_y d\mu_x + \beta^{-1} \int \log\left(\frac{d\mu_x}{d\tilde{p}}\right) d\mu_x + \beta^{-1} H(\mu_y)
\end{aligned}$$

Notice that the first term in the right hand side is a lower semi-continuous (in weak convergence topology) functional in μ_x when μ_y is fixed. That is because it is a linear functional in μ_x with a continuous integrand, which implies that it is continuous in the weak convergence topology. The second to last term can be seen as the relative entropy (or Kullback-Leibler divergence) between μ_x and $\tilde{p}$:

$$H_{\tilde{p}}(\mu_x) := \int \log\left(\frac{d\mu_x}{d\tilde{p}}\right) d\mu_x$$

The relative entropy $H_{\tilde{p}}(\mu_x)$ is a lower semi-continuous functional with respect to μ_x (see Theorem 1 of Posner [1975], which proves a stronger statement: joint semi-continuity with respect to

both measures).

Therefore, we conclude that $\mathcal{L}_\beta(\cdot, \mu_y)$ (with $\mu_y \in \mathcal{P}(\mathcal{Y})$ fixed) is a l.s.c. functional on $\mathcal{P}(\mathcal{X})$. By Theorem A.11 and using the compactness of $\mathcal{P}(\mathcal{X})$, there exists a minimum of $\mathcal{L}_\beta(\cdot, \mu_y)$ in $\mathcal{P}(\mathcal{X})$. Denote a minimum of $\mathcal{L}_\beta(\cdot, \mu_y)$ by $\hat{\mu}_x$. $\hat{\mu}_x$ must be absolutely continuous, because otherwise $-\beta^{-1}H(\hat{\mu}_x)$ would take an infinite value. By the Euler-Lagrange equations for functionals on probability measures, a necessary condition for $\hat{\mu}_x$ to be a minimum of $\mathcal{L}_\beta(\cdot, \mu_y)$ is that the first variation $\frac{\delta \mathcal{L}_\beta(\cdot, \mu_y)}{\delta \mu_x}(\hat{\mu}_x)(x)$ must take a constant value for all $x \in \text{supp}(\hat{\mu}_x)$ and values larger or equal outside of $\text{supp}(\hat{\mu}_x)$. The intuition behind this is that otherwise a zero-mean signed measure with positive mass on the minimizers of $\frac{\delta \mathcal{L}_\beta(\cdot, \mu_y)}{\delta \mu_x}(\hat{\mu}_x)$ and negative mass on the maximizers would provide a direction of decrease of the functional. We compute the first variation at $\hat{\mu}_x$:

$$\begin{aligned}\frac{\delta \mathcal{L}_\beta(\cdot, \mu_y)}{\delta \mu_x}(\hat{\mu}_x)(x) &= \frac{\delta}{\delta \mu_x}\left(\int L(x,y) d\mu_y d\mu_x - \beta^{-1}H(\hat{\mu}_x) + \beta^{-1}H(\mu_y)\right) \\ &= \int L(x,y) d\mu_y + \beta^{-1}\log\left(\frac{d\hat{\mu}_x}{dx}(x)\right),\end{aligned}$$

We equate $\int \ell(x,y) d\mu_y + \beta^{-1}\log(\frac{d\hat{\mu}_x}{dx}(x)) = K,\ \forall x \in \text{supp}(\hat{\mu}_x)$, where K is a constant. The first variation must take values larger or equal than K outside of $\text{supp}(\hat{\mu}_x)$, but since $\log(\frac{d\hat{\mu}_x}{dx}(x)) = -\infty$ outside of $\text{supp}(\hat{\mu}_x)$, we obtain that $\text{supp}(\hat{\mu}_x) = \mathcal{X}$. Then, for all $x \in \mathcal{X}$,

$$\frac{d\hat{\mu}_x}{dx}(x) = e^{-\beta \int L(x,y) d\mu_y + \beta K} = \frac{1}{Z_{\mu_y}} e^{-\beta \int L(x,y) d\mu_y}$$

where Z_{μ_y} is a normalization constant obtained from imposing $\int \frac{d\hat{\mu}_x}{dx}(x)\, dx = \int 1\, d\hat{\mu}_x = 1$. Since the necessary condition for optimality specifies a unique measure and the minimum exists, we obtain that $m_x(\mu_y) = \hat{\mu}_x$ is the unique minimum. An analogous argument holds for $m_y(\hat{\mu}_x)$ □

Lemma A.13. *Suppose that the measures $(\mu_{y,n})_{n\in\mathbb{N}}$ and μ_y are in $\mathcal{P}(\mathcal{Y})$. Recall the definition of $m_x : \mathcal{P}(\mathcal{Y}) \to \mathcal{P}(\mathcal{X})$ in equation (A.6). If $(\mu_{y,n})_{n\in\mathbb{N}}$ converges weakly to μ_y, then $(m_x(\mu_{y,n}))_{n\in\mathbb{N}}$ converges weakly to $m_x(\mu_y)$, i.e. m_x is a continuous mapping when we endow $\mathcal{P}(\mathcal{Y})$ and $\mathcal{P}(\mathcal{X})$ with*

their weak convergence topologies.

The same thing holds for m_y and measures $(\mu_{x,n})_{n\in\mathbb{N}}$ and μ_x on $\mathcal{X}$.

Proof. Given $x \in \mathcal{X}$, we have $\int \ell(x,y)d\mu_{y,n} \to \int \ell(x,y)d\mu_y$, because $\ell(x,\cdot)$ is a continuous bounded function on $\mathcal{Y}$. By continuity of the exponential function, we have that for all $x \in \mathcal{X}$, $e^{-\beta\int \ell(x,y)d\mu_{y,n}} \to e^{-\beta\int \ell(x,y)d\mu_y}$. Using the dominated convergence theorem,

$$\int_{\mathcal{X}} e^{-\beta\int \ell(x,y)d\mu_{y,n}}dx \to \int_{\mathcal{X}} e^{-\beta\int \ell(x,y)d\mu_y}dx$$

We need to find a dominating function. It is easy, because $\forall n \in \mathbb{N}$, $\forall x \in \mathcal{X}$, $e^{-\beta\int \ell(x,y)d\mu_{y,n}} \leq e^{-\beta\min_{(x,y)\in\mathcal{X}\times\mathcal{Y}}\ell(x,y)}$. And $\int_{\mathcal{X}} e^{-\beta\min_{(x,y)\in\mathcal{X}\times\mathcal{Y}}\ell(x,y)}dx = e^{-\beta\min_{(x,y)\in\mathcal{X}\times\mathcal{Y}}\ell(x,y)}\mathrm{vol}(\mathcal{X}) < \infty$. By the Portmanteau theorem, we just need to prove that for all continuity sets B of $m_x(\mu_y)$, we have $m_x(\mu_{y,n})(B) \to m_x(\mu_y)(B)$. This translates to

$$\frac{\int_B e^{-\beta\int \ell(x,y)d\mu_{y,n}}dx}{\int_{\mathcal{X}} e^{-\beta\int \ell(x,y)d\mu_{y,n}}dx} \to \frac{\int_B e^{-\beta\int \ell(x,y)d\mu_y}dx}{\int_{\mathcal{X}} e^{-\beta\int \ell(x,y)d\mu_y}dx}$$

We have proved that the denominators converge appropriately, and the numerator converges as well using the same reasoning with dominated convergence. And both the numerators and the denominators are positive and the numerator is always smaller denominator, the quotient must converge. □

Lemma A.14. *There exists a solution of (2.9), which is the Nash equilibrium of the game given by $\mathcal{L}_\beta$.*

Proof. We use Theorem A.9 on the set $\mathcal{P}(\mathcal{X}) \times \mathcal{P}(\mathcal{Y})$, with the map $m : \mathcal{P}(\mathcal{X}) \times \mathcal{P}(\mathcal{Y}) \to \mathcal{P}(\mathcal{X}) \times \mathcal{P}(\mathcal{Y})$ given by $m(\mu_x, \mu_y) = (m_x(\mu_y), m_y(\mu_x))$. The only condition to check is upper hemicontinuity of m. By Theorem A.13 we know that m_x, m_y are continuous, and since continuous functions are upper hemicontinuous as set valued functions, this concludes the argument.

Indeed, we could have used Tychonoff's theorem, which is similar to Theorem A.9 but for single-valued functions. □

A.3.3 Proof of Theorem A.4: Uniqueness

Lemma A.15. *The solution of (2.9) is unique.*

Proof. The argument is analogous to the proof of Theorem 2 of Rosen [1965]. Suppose $(\mu_{x,1}, \mu_{y,1})$ and $(\mu_{x,2}, \mu_{y,2})$ are two different solutions of (2.9). We use the notation $F_1(\mu_x, \mu_y) = \mathcal{L}_\beta(\mu_x, \mu_y)$, $F_2(\mu_x, \mu_y) = -\mathcal{L}_\beta(\mu_x, \mu_y)$. Hence, there exist constants $K_{x,1}, K_{y,1}, K_{x,2}, K_{y,2}$ such that

$$\frac{\delta F_1}{\delta \mu_x}(\mu_{x,1}, \mu_{y,1})(x) + K_{x,1} = 0, \frac{\delta F_2}{\delta \mu_y}(\mu_{x,1}, \mu_{y,1})(y) + K_{y,1} = 0,$$
$$\frac{\delta F_1}{\delta \mu_x}(\mu_{x,2}, \mu_{y,2})(x) + K_{x,2} = 0, \frac{\delta F_2}{\delta \mu_y}(\mu_{x,2}, \mu_{y,2})(y) + K_{y,2} = 0$$

On the one hand, we know that

$$\begin{aligned}
&\int \frac{\delta F_1}{\delta \mu_x}(\mu_{x,1}, \mu_{y,1})(x)\, d(\mu_{x,2} - \mu_{x,1}) + \int \frac{\delta F_2}{\delta \mu_y}(\mu_{x,1}, \mu_{y,1})(y)\, d(\mu_{y,2} - \mu_{y,1}) \\
&+ \int \frac{\delta F_1}{\delta \mu_x}(\mu_{x,2}, \mu_{y,2})(x)\, d(\mu_{x,1} - \mu_{x,2}) + \int \frac{\delta F_2}{\delta \mu_y}(\mu_{x,2}, \mu_{y,2})(y)\, d(\mu_{y,1} - \mu_{y,2}) \\
&= -\int K_{x,1}\, d(\mu_{x,2} - \mu_{x,1}) - \int K_{y,1}\, d(\mu_{y,2} - \mu_{y,1}) \\
&- \int K_{x,2}\, d(\mu_{x,1} - \mu_{x,2}) - \int K_{y,2}\, d(\mu_{y,1} - \mu_{y,2}) = 0
\end{aligned} \tag{A.7}$$

We will now prove that the left hand side of (A.7) must be strictly larger than 0, reaching a

contradiction. We can write

$$\begin{aligned}
\frac{\delta F_1}{\delta \mu_x}(\mu_{x,2}, \mu_{y,2})(x) - \frac{\delta F_1}{\delta \mu_x}(\mu_{x,1}, \mu_{y,1})(x) &= \int L(x,y)\, d(\mu_{y,2} - \mu_{y,1}) \\
&\quad + \beta^{-1}(\log(\mu_{x,2}(x)) - \log(\mu_{x,1}(x))), \\
\frac{\delta F_2}{\delta \mu_y}(\mu_{x,2}, \mu_{y,2})(x) - \frac{\delta F_2}{\delta \mu_y}(\mu_{x,1}, \mu_{y,1})(x) &= -\int L(x,y)\, d(\mu_{x,2} - \mu_{x,1}) \\
&\quad + \beta^{-1}(\log(\mu_{y,2}(x)) - \log(\mu_{y,1}(x)))
\end{aligned}$$

Hence, we rewrite the left hand side of (A.7) as

$$\begin{aligned}
&\iint L(x,y)\, d(\mu_{y,2} - \mu_{y,1}) d(\mu_{x,2} - \mu_{x,1}) + \beta^{-1} \int (\log(\mu_{x,2}(x)) - \log(\mu_{x,1}(x)))\, d(\mu_{x,2} - \mu_{x,1}) \\
&- \iint L(x,y)\, d(\mu_{x,2} - \mu_{x,1}) d(\mu_{y,2} - \mu_{y,1}) + \beta^{-1} \int (\log(\mu_{y,2}(x)) - \log(\mu_{y,1}(x)))\, d(\mu_{y,2} - \mu_{y,1}) \\
&= \beta^{-1}(H_{\mu_{x,1}}(\mu_{x,2}) + H_{\mu_{x,2}}(\mu_{x,1}) + H_{\mu_{y,1}}(\mu_{y,2}) + H_{\mu_{y,1}}(\mu_{y,2})).
\end{aligned}$$

Since the relative entropy is always non-negative and zero only if the two measures are equal, we have reached the desired contradiction. □

A.3.4 Proof of Theorem A.5

We will use the shorthand $V_x(x) = V_x(\hat{\mu}_y)(x) = \int \mathcal{L}(x,y) d\hat{\mu}_y$, $V_y(y) = V_y(\hat{\mu}_x)(y) = \int \mathcal{L}(x,y) d\hat{\mu}_x$. Since $\ell : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$ is a continuous function on a compact metric space, it is uniformly continuous. Hence,

$$\forall \epsilon > 0, \exists \delta > 0 \text{ st. } \sqrt{d(x,x')^2 + d(y,y')^2} < \delta \implies |\ell(x,y) - \ell(x',y')| < \epsilon$$

Which means that

$$d(x,x') < \delta \implies |V_x(x) - V_x(x')| = \left| \int (\ell(x,y) - \ell(x',y))dy \right| < \epsilon$$

This proves that V_x is uniformly continuous on $\mathcal{X}$ (and V_y is uniformly continuous on $\mathcal{Y}$ using the same argument).

We can write the Nikaido-Isoda function of the game with loss $\mathcal{L}$ (equation (2.2)) evaluated at $(\hat{\mu}_x, \hat{\mu}_y)$ as

$$\begin{aligned} \mathrm{NI}(\hat{\mu}_x, \hat{\mu}_y) &:= \mathcal{L}(\hat{\mu}_x, \hat{\mu}_y) - \min_{\mu'_x}\{\mathcal{L}(\mu'_x, \hat{\mu}_y)\} + (-\mathcal{L}(\hat{\mu}_x, \hat{\mu}_y) + \max_{\mu'_y}\{\mathcal{L}(\hat{\mu}_x, \mu'_y)\}) \\ &= \frac{\int V_x(x) e^{-\beta V_x(x)} dx}{\int e^{-\beta V_x(x)} dx} - \min_{x \in C_1} V_x(x) + \frac{-\int V_y(y) e^{\beta V_y(y)} dy}{\int e^{\beta V_y(y)} dy} + \max_{y \in C_2} V_y(y) \end{aligned} \tag{A.8}$$

The second equality follows from the definitions of $\mathcal{L}, V_x, V_y$. We observe that in the right-most expression the first two terms and the last two terms are analogous. Let us show the first two terms can be made smaller than an arbitrary $\epsilon > 0$ by taking β large enough; the last two will be dealt with in an analogous manner. Let us define $\tilde{V}_x(x) = V_x(x) - \min_{x' \in C_1} V_x(x')$.

$$\begin{aligned} &\frac{\int V_x(x) e^{-\beta V_x(x)} dx}{\int e^{-\beta V_x(x)} dx} - \min_{x \in C_1} V_x(x) = \frac{\int (V_x(x) - \min_{x' \in C_1} V_x(x')) e^{-\beta V_x(x)} dx}{\int e^{-\beta V_x(x)} dx} \\ &= \frac{\int \tilde{V}_x(x) e^{-\beta V_x(x)} \left(\mathbb{1}_{\{\tilde{V}_x(x) \le \epsilon/2\}} + \mathbb{1}_{\{\epsilon/2 < \tilde{V}_x(x) \le \epsilon\}} + \mathbb{1}_{\{\epsilon < \tilde{V}_x(x)\}} \right) dx}{\int e^{-\beta V_x(x)} \mathbb{1}_{\{\tilde{V}_x(x) \le \epsilon/2\}} dx + \int e^{-\beta V_x(x)} \mathbb{1}_{\{\epsilon/2 < \tilde{V}_x(x) \le \epsilon\}} dx + \int e^{-\beta V_x(x)} \mathbb{1}_{\{\epsilon < \tilde{V}_x(x)\}} dx} \end{aligned} \tag{A.9}$$

Let us define

$$q_{\{\tilde{V}_x(x) \le \epsilon/2\}} = \int e^{-\beta V_x(x)} \mathbb{1}_{\{\tilde{V}_x(x) \le \epsilon/2\}} dx,$$

and $q_{\{\epsilon/2 < \tilde{V}_x(x) \le \epsilon\}}$ and $q_{\{\epsilon < \tilde{V}_x(x)\}}$ analogously.

Similarly, let

$$r_{\{\tilde{V}_x(x)\le\epsilon/2\}} = \int \tilde{V}_x(x)e^{-\beta V_x(x)}\mathbb{1}_{\{\tilde{V}_x(x)\le\epsilon/2\}}dx,$$

and $r_{\{\epsilon/2<\tilde{V}_x(x)\le\epsilon\}}$ and $r_{\{\epsilon<\tilde{V}_x(x)\}}$ analogously.

Let

$$\tilde{p} = \frac{q_{\{\epsilon/2<\tilde{V}_x(x)\le\epsilon\}}}{q_{\{\tilde{V}_x(x)\le\epsilon/2\}} + q_{\{\epsilon/2<\tilde{V}_x(x)\le\epsilon\}} + q_{\{\epsilon<\tilde{V}_x(x)\}}}$$

Then, we can rewrite the right-most expression of (A.9) as

$$\begin{aligned}&\frac{r_{\{\tilde{V}_x(x)\le\epsilon/2\}} + r_{\{\epsilon/2<\tilde{V}_x(x)\le\epsilon\}} + r_{\{\epsilon<\tilde{V}_x(x)\}}}{q_{\{\tilde{V}_x(x)\le\epsilon/2\}} + q_{\{\epsilon/2<\tilde{V}_x(x)\le\epsilon\}} + q_{\{\epsilon<\tilde{V}_x(x)\}}}\\ &= \tilde{p}\frac{r_{\{\epsilon/2<\tilde{V}_x(x)\le\epsilon\}}}{q_{\{\epsilon/2<\tilde{V}_x(x)\le\epsilon\}}} + (1-\tilde{p})\frac{r_{\{\tilde{V}_x(x)\le\epsilon/2\}} + r_{\{\epsilon<\tilde{V}_x(x)\}}}{q_{\{\tilde{V}_x(x)\le\epsilon/2\}} + q_{\{\epsilon<\tilde{V}_x(x)\}}}\end{aligned} \tag{A.10}$$

Since $\tilde{V}(x) \le \epsilon$ in the set $\{x|\epsilon/2 < \tilde{V}_x(x) \le \epsilon\}$, $r_{\{\epsilon/2<\tilde{V}_x(x)\le\epsilon\}}/q_{\{\epsilon/2<\tilde{V}_x(x)\le\epsilon\}} \le \epsilon$.

Let $x_{\min}$ be such that $V(x_{\min}) = \min_{x\in C_1} V(x)$ (possibly not unique). By uniform continuity of V_x, we know there exists $\delta > 0$ (dependent only on ϵ) such that $B(x_{\min}, \delta) \subseteq \{x|\tilde{V}_x(x) \le \epsilon/2\}$. The following inequalities hold:

$$\begin{aligned}r_{\{\tilde{V}_x(x)\le\epsilon/2\}} &\le \frac{\epsilon}{2}q_{\{\tilde{V}_x(x)\le\epsilon/2\}},\\ r_{\{\epsilon<\tilde{V}_x(x)\}} &\le (\max_{x\in C_1} V_x(x) - \min_{x\in C_1} V_x(x))q_{\{\epsilon<\tilde{V}_x(x)\}} \le (\max_{x,y} L(x,y) - \min_{x,y} L(x,y))q_{\{\epsilon<\tilde{V}_x(x)\}}\\ &= K_L q_{\{\epsilon<\tilde{V}_x(x)\}}.\end{aligned} \tag{A.11}$$

where we define $K_\ell = \max_{x,y}\ell(x,y) - \min_{x,y}\ell(x,y)$. Using (A.11), we obtain

$$\frac{r_{\{\tilde{V}_x(x)\le\epsilon/2\}} + r_{\{\epsilon<\tilde{V}_x(x)\}}}{q_{\{\tilde{V}_x(x)\le\epsilon/2\}} + q_{\{\epsilon<\tilde{V}_x(x)\}}} \le \frac{\frac{\epsilon}{2}q_{\{\tilde{V}_x(x)\le\epsilon/2\}} + K_L q_{\{\epsilon<\tilde{V}_x(x)\}}}{q_{\{\tilde{V}_x(x)\le\epsilon/2\}} + q_{\{\epsilon<\tilde{V}_x(x)\}}}.$$

If the right-hand side is smaller or equal than ϵ, then equation (A.10) would be smaller than ϵ and the proof would be concluded. For that to happen, we need $(K_\ell - \epsilon) q_{\{\epsilon < \tilde{V}_x(x)\}} \leq \frac{\epsilon}{2} q_{\{\tilde{V}_x(x) \leq \epsilon/2\}} \Leftrightarrow q_{\{\tilde{V}_x(x) \leq \epsilon/2\}} / q_{\{\epsilon < \tilde{V}_x(x)\}} \geq 2(K_\ell/\epsilon - 1)$. The following bounds hold:

$$
\begin{aligned}
q_{\{\tilde{V}_x(x) \leq \epsilon/2\}} &\geq \mathrm{Vol}(B(x_{\min}, \delta)) e^{-\beta(\min_{x \in C_1} V_x(x) + \epsilon/2)}, \\
q_{\{\epsilon < \tilde{V}_x(x)\}} &\leq (1 - \mathrm{Vol}(B(x_{\min}, \delta))) e^{-\beta(\min_{x \in C_1} V_x(x) + \epsilon)}.
\end{aligned}
$$

Thus, the following condition is sufficient:

$$
\frac{\mathrm{Vol}(B(x_{\min}, \delta))}{1 - \mathrm{Vol}(B(x_{\min}, \delta))} e^{\beta \epsilon/2} \geq 2(K_L/\epsilon - 1).
$$

Hence, if we take

$$
\beta \geq \frac{2}{\epsilon} \log \left(2 \frac{1 - \mathrm{Vol}(B(x_{\min}, \delta))}{\mathrm{Vol}(B(x_{\min}, \delta))} (K_L/\epsilon - 1) \right) \tag{A.12}
$$

then $(\hat{\mu}_x, \hat{\mu}_y)$ is an ϵ-Nash equilibrium. Since we have only bound the first two terms in the right hand side of (A.8) and the other two are bounded in the same manner, the statement of the theorem results from setting $\epsilon = \epsilon/2$ in (A.12).

A.3.5 Proof of Theorem A.6

First, we show that any pair $\hat{\mu}_x, \hat{\mu}_y$ such that

$$
\frac{d\hat{\mu}_x}{dx}(x) = \frac{1}{Z_x} e^{-\beta \int \ell(x,y)\, d\hat{\mu}_y(y)}, \quad \frac{d\hat{\mu}_y}{dy}(y) = \frac{1}{Z_y} e^{\beta \int \ell(x,y)\, d\hat{\mu}_x(x)}
$$

is a stationary solution of (2.7). Denoting the Radon-Nikodym derivatives $\frac{d\hat{\mu}_x}{dx}, \frac{d\hat{\mu}_y}{dy}$ by $\hat{\rho}_x, \hat{\rho}_y$, it is sufficient to see that

$$\begin{cases} 0 = \nabla_x \cdot (\hat{\rho}_x \nabla_x V_x(\mu_y, x)) + \beta^{-1} \Delta_x \hat{\rho}_x \\ 0 = -\nabla_y \cdot (\hat{\rho}_y \nabla_y V_y(\mu_x, y)) + \beta^{-1} \Delta_y \hat{\rho}_y \end{cases} \tag{A.13}$$

holds weakly. And

$$\begin{aligned} \nabla_x \hat{\rho}_x &= \frac{1}{Z_x} e^{-\beta \int \ell(x,y)\, d\hat{\mu}_y(y)} \left(-\beta \nabla_x \int \ell(x,y)\, d\hat{\mu}_y(y) \right) = -\hat{\rho}_x \nabla_x V_x(\hat{\mu}_y, x), \\ \nabla_y \hat{\rho}_y &= \frac{1}{Z_y} e^{\beta \int \ell(x,y)\, d\hat{\mu}_x(x)} \left(\beta \nabla_y \int \ell(x,y)\, d\hat{\mu}_x(x) \right) = \hat{\rho}_y \nabla_y V_y(\hat{\mu}_x, y), \end{aligned}$$

implies that (A.13) holds.

Now we will prove the converse. Suppose that $\hat{\mu}_x, \hat{\mu}_y$ are (weak) stationary solutions of (2.7). That is, if $\psi_x \in C^2(\mathcal{X}), \psi_y \in C^2(\mathcal{Y})$ are arbitrary twice continuously differentiable functions, the following holds

$$\begin{aligned} 0 &= \int_{\mathcal{X}} \left(-\int_{\mathcal{Y}} \nabla_x \psi_x(x) \cdot \nabla_x \ell(x,y)\, d\hat{\mu}_y + \beta^{-1} \Delta_x \psi_x(x) \right) d\hat{\mu}_x \\ 0 &= \int_{\mathcal{Y}} \left(\int_{\mathcal{X}} -\nabla_y \psi_y(y) \cdot \nabla_y \ell(x,y)\, d\hat{\mu}_x - \beta^{-1} \Delta_y \psi_y(x,y) \right) d\hat{\mu}_y \end{aligned} \tag{A.14}$$

(A.14) can be seen as two measure-valued stationary Fokker-Planck equations. We want to see that they have densities and that the densities satisfy the corresponding classical stationary Fokker-Planck equations (A.13). Works in the theory of PDEs have studied sufficient conditions for measure-valued stationary Fokker-Planck equations to correspond to weak stationary Fokker-Planck equations, and further to classical stationary Fokker-Planck equations. See page 3 of Huang et al. [2015] for a more detailed explanation on the two steps. That measure-valued

stationary correspond to weak stationary solutions is shown in Theorem 2.2 of [Bogachev et al. 2001]. That weak stationary solutions are classical stationary solutions requires that the drift term is in $C^{1,\alpha}_{\text{loc}}$ (locally α-Hölder continuous with exponent 1), meaning that it is in C^1 and that its derivatives are α-Hölder in compact sets. The result follows from the theory of Schauder estimates. Differentiating under the integral sign, the drift terms $-\int_{\mathcal{Y}} \nabla_x \ell(x,y)\, d\hat{\mu}_y$, $\int_{\mathcal{X}} \nabla_y \ell(x,y)\, d\hat{\mu}_x$ fulfill the condition if $\ell \in C^{2,\alpha}$.

A.4 Proof of Theorem 2.3

Recall the expression of an *Interacting Wasserstein-Fisher-Rao Gradient Flow (IWFRGF)* in (2.8):

$$\begin{cases} \partial_t \mu_x &= \gamma \nabla \cdot (\mu_x \nabla_x V_x(\mu_y, x)) \\ &\quad -\alpha \mu_x (V_x(\mu_y, x) - \mathcal{L}(\mu_x, \mu_y)), \quad \mu_x(0) = \mu_{x,0} \\ \partial_t \mu_y &= -\gamma \nabla \cdot (\mu_y \nabla_y V_y(\mu_x, y)) \\ &\quad +\alpha \mu_y (V_y(\mu_x, y) - \mathcal{L}(\mu_x, \mu_y)), \quad \mu_y(0) = \mu_{y,0} \end{cases}$$

The aim is to obtain a global convergence result like the one in Theorem 3.8 of Chizat [2019]. First, we will rewrite Lemma 3.10 of Chizat [2019] in our case.

Lemma A.16. *Let μ_x, μ_y be the solution of the IWFRGF in (2.8). Let $\mu_x^\star, \mu_y^\star$ be arbitrary measures on $\mathcal{X}, \mathcal{Y}$. Let $\bar{\mu}_x(t) = \frac{1}{t}\int_0^t \mu_x(s)\, ds$ and $\bar{\mu}_y(t) = \frac{1}{t}\int_0^t \mu_y(s)\, ds$. Let $\|\cdot\|_{BL}$ be the bounded Lipschitz norm, i.e. $\|f\|_{BL} = \|f\|_\infty + Lip(f)$. Let*

$$Q_{\mu^\star, \mu_0}(\tau) = \inf_{\mu \in \mathcal{P}(\Theta)} \|\mu^\star - \mu\|_{BL}^* + \frac{1}{\tau}\mathcal{H}(\mu, \mu_0) \tag{A.15}$$

with $\Theta = \mathcal{X}$ *or* $\mathcal{Y}$. *Let*

$$B = \frac{1}{2}\left(\max_{x\in\mathcal{X},y\in\mathcal{Y}} \ell(x,y) - \min_{x\in\mathcal{X},y\in\mathcal{Y}} \ell(x,y)\right) + Lip(\ell) \tag{A.16}$$

Then,

$$\mathcal{L}(\bar{\mu}_x(t), \mu_y^\star) - \mathcal{L}(\mu_x^\star, \bar{\mu}_y(t)) \leq B\mathcal{Q}_{\mu_x^\star,\mu_{x,0}}(\alpha B t) + B\mathcal{Q}_{\mu_y^\star,\mu_{y,0}}(\alpha B t) + \gamma B^2 t \tag{A.17}$$

Proof. The proof is as in Lemma 3.10 of Chizat [2019], but in this case we have to do everything twice. Namely, we define the dynamics

$$\begin{aligned}\frac{d\mu_x^\epsilon}{dt} &= \gamma\nabla\cdot(\mu_x^\epsilon \nabla V_x(\mu_y, x))\\ \frac{d\mu_y^\epsilon}{dt} &= -\gamma\nabla\cdot(\mu_y^\epsilon \nabla V_y(\mu_x, y))\end{aligned}$$

initialized at $\mu_x^\epsilon(0) = \mu_{x,0}^\epsilon, \mu_y^\epsilon(0) = \mu_{y,0}^\epsilon$ arbitrary such that $\mu_{x,0}^\epsilon$ and $\mu_{y,0}^\epsilon$ are absolutely continuous with respect to $\mu_{x,0}$ and $\mu_{y,0}$ respectively.

Let us show that

$$\frac{1}{\alpha}\frac{d}{dt}\mathcal{H}(\mu_x^\epsilon, \mu_x) = \int \frac{\delta\mathcal{L}}{\delta\mu_x}(\mu_x,\mu_y)(x)\; d(\mu_x^\epsilon - \mu_x) \tag{A.18}$$

where $\mathcal{H}(\mu_x^\epsilon, \mu_x)$ is the relative entropy, i.e.

$$\frac{d}{dt}\mathcal{H}(\mu_x^\epsilon, \mu_x) = \frac{d}{dt}\int \log\left(\rho_x^\epsilon\right)\; d\mu_x^\epsilon,$$

ρ_x^ϵ being the Radon-Nikodym derivative $d\mu_x^\epsilon/d\mu_x$.

Assume to begin with that μ_x^ϵ remains absolutely continuous with respect to μ_x through time. We

can write

$$\frac{d}{dt}\int \psi_x(x)\rho_x^\epsilon(x)d\mu_x(x) = \frac{d}{dt}\int \psi(x)d\mu_x^\epsilon(x)$$

We can develop the left hand side into

$$\begin{aligned}\frac{d}{dt}\int \psi_x(x)\rho_x^\epsilon(x)d\mu_x(x) &= \int -\gamma\nabla(\psi_x(x)\rho_x^\epsilon(x))\cdot\nabla V_x(\mu_y,x)d\mu_x(x)\\ &+\int -\alpha\psi_x(x)\rho_x^\epsilon(x)(V_x(\mu_y,x)-\mathcal{L}(\mu_x,\mu_y))d\mu_x(x)\\ &+\int \psi_x(x)\frac{\partial\rho_x^\epsilon}{\partial t}(x)d\mu_x(x)\\ &=\int -\gamma(\nabla\psi_x(x)\rho_x^\epsilon(x)+\psi_x(x)\nabla\rho_x^\epsilon(x))\cdot\nabla V_x(\mu_y,x)\ d\mu_x(x)\\ &+\int -\alpha\psi_x(x)\rho_x^\epsilon(x)(V_x(\mu_y,x)-\mathcal{L}(\mu_x,\mu_y))d\mu_x(x)\\ &+\int \psi_x(x)\frac{\partial\rho_x^\epsilon}{\partial t}(x)d\mu_x(x)\end{aligned}$$

and the right hand side into

$$\frac{d}{dt}\int \psi(x)d\mu_x^\epsilon(x) = \int -\gamma\nabla\psi_x(x)\cdot\nabla V_x(\mu_y,x)d\mu_x^\epsilon(x)$$

Note that comparing terms, we obtain

$$\begin{aligned}&\int -\gamma\psi_x(x)\nabla\rho_x^\epsilon(x)\cdot\nabla V_x(\mu_y,x)\ d\mu_x(x)\\ &=\int \alpha\psi_x(x)\rho_x^\epsilon(x)(V_x(\mu_y,x)-\mathcal{L}(\mu_x,\mu_y))-\psi_x(x)\frac{\partial\rho_x^\epsilon}{\partial t}(x)\ d\mu_x(x)\end{aligned}$$

Since ψ_x is arbitrary, it must be that

$$-\gamma \nabla \rho_x^\epsilon(x) \cdot \nabla V_x(\mu_y, x) = \alpha \rho_x^\epsilon(x)(V_x(\mu_y, x) - \mathcal{L}(\mu_x, \mu_y)) - \frac{\partial}{\partial t}\rho_x^\epsilon(x) \tag{A.19}$$

holds μ_x-almost everywhere. Now,

$$\begin{aligned}
\frac{d}{dt}\int \log\left(\rho_x^\epsilon\right)\, d\mu_x^\epsilon &= -\gamma \int \nabla\left(\log\left(\rho_x^\epsilon(x)\right)\right) \cdot \nabla V_x(\mu_y, x)\, d\mu_x^\epsilon(x) \\
&= -\gamma \int \frac{1}{\rho_x^\epsilon(x)} \nabla\left(\rho_x^\epsilon(x)\right) \cdot \nabla V_x(\mu_y, x)\, d\mu_x^\epsilon(x) \\
&= \alpha \int (V_x(\mu_y, x) - \mathcal{L}(\mu_x, \mu_y))\, d\mu_x^\epsilon(x) - \int \frac{1}{\rho_x^\epsilon(x)} \frac{\partial}{\partial t}\rho_x^\epsilon(x) d\mu_x^\epsilon(x)
\end{aligned}$$

Here,

$$\int \frac{1}{\rho_x^\epsilon(x)} \frac{\partial}{\partial t}\rho_x^\epsilon(x) d\mu_x^\epsilon(x) = \int \frac{\partial}{\partial t}\rho_x^\epsilon(x) d\mu_x(x) = 0$$

And since

$$\mathcal{L}(\mu_x, \mu_y) = \int \frac{\delta \mathcal{L}}{\delta \mu_x}(\mu_x, \mu_y)(x)\, d\mu_x,$$

the first term yields (A.18). We assumed that ρ_x^ϵ existed and was regular enough. To make the argument precise, we can define the density of μ_x^ϵ with respect to μ_x to be a solution ρ_x^ϵ of (A.19), and thus specify μ_x^ϵ.

Now, recall that $\mu_x^\star$ is an arbitrary measure in $\mathcal{P}(\mathcal{X})$. By linearity of $\mathcal{L}$ with respect to μ_x,

$$\begin{aligned}
&\int \frac{\delta \mathcal{L}}{\delta \mu_x}(\mu_x, \mu_y)(x)\, d(\mu_x^\epsilon - \mu_x) = \int \frac{\delta \mathcal{L}}{\delta \mu_x}(\mu_x, \mu_y)(x)\, d(\mu_x^\star - \mu_x) + \int \frac{\delta \mathcal{L}}{\delta \mu_x}(\mu_x, \mu_y)(x)\, d(\mu_x^\epsilon - \mu_x^\star) \\
&\leq -(\mathcal{L}(\mu_x, \mu_y) - \mathcal{L}(\mu_x^\star, \mu_y)) + \|\frac{\delta \mathcal{L}}{\delta \mu_x}(\mu_x, \mu_y)\|_{\mathrm{BL}} \|\mu_x^\epsilon - \mu_x^\star\|_{\mathrm{BL}}^*
\end{aligned} \tag{A.20}$$

Notice that we can take $\|\frac{\delta \mathcal{L}}{\delta \mu_x}(\mu_x, \mu_y)\|_{\mathrm{BL}}$ to be smaller than B (defined in (A.16)). If we integrate

(A.18) and (A.20) from 0 to t and divide by t, we obtain

$$\begin{aligned}
&\frac{1}{t}\int_0^t \mathcal{L}(\mu_x(s),\mu_y(s))\,ds - \frac{1}{t}\int_0^t \mathcal{L}(\mu_x^\star,\mu_y(s))\,ds \\
&\leq \frac{1}{\alpha t}(\mathcal{H}(\mu_{x,0}^\epsilon,\mu_{x,0}) - \mathcal{H}(\mu_x^\epsilon(t),\mu_x(t))) + \frac{B}{t}\int_0^t \|\mu_x^\epsilon - \mu_x^\star\|_{\mathrm{BL}}^*\,ds
\end{aligned} \tag{A.21}$$

We bound the last term on the RHS:

$$\frac{B}{t}\int_0^t \|\mu_x^\epsilon - \mu_x^\star\|_{\mathrm{BL}}^*\,ds \leq B\|\mu_{x,0}^\epsilon - \mu_x^\star\|_{\mathrm{BL}}^* + \frac{B}{t}\int_0^t \|\mu_{x,0}^\epsilon - \mu_x^\epsilon\|_{\mathrm{BL}}^*\,ds \tag{A.22}$$

And

$$\begin{aligned}
\|\mu_x^\epsilon(t) - \mu_{x,0}^\epsilon\|_{\mathrm{BL}}^* &= \sup_{\|f\|_{\mathrm{BL}}\leq 1, f\in C^2(\mathcal{X})} \int f\,d(\mu_x^\epsilon(t) - \mu_{x,0}^\epsilon) = \sup_{\|f\|_{\mathrm{BL}}\leq 1, f\in C^2(\mathcal{X})} \int_0^t \frac{d}{ds}\int f\,d\mu_x^\epsilon(s)\,ds \\
&= \sup_{\|f\|_{\mathrm{BL}}\leq 1, f\in C^2(\mathcal{X})} -\int_0^t \int \gamma\nabla f(x)\cdot\nabla\frac{\delta\mathcal{L}}{\delta\mu_x}(\mu_x^\epsilon,\mu_y)(x)\,d\mu_x^\epsilon(s)\,ds \\
&\leq \int_0^t \int \gamma B\,d\mu_x^\epsilon(s)\,ds = \gamma B t
\end{aligned} \tag{A.23}$$

Also, by linearity of $\mathcal{L}$ with respect to μ_y,

$$-\frac{1}{t}\int_0^t \mathcal{L}(\mu_x^\star,\mu_y(s))\,ds = -\mathcal{L}(\mu_x^\star,\bar{\mu}_y(t)) \tag{A.24}$$

If we use (A.22), (A.23) and (A.24) and the non-negativeness of the relative entropy on (A.21), we obtain:

$$\frac{1}{t}\int_0^t \mathcal{L}(\mu_x(s),\mu_y(s))\,ds - \mathcal{L}(\mu_x^\star,\bar{\mu}_y(t)) \leq \frac{\mathcal{H}(\mu_{x,0}^\epsilon,\mu_{x,0})}{4\alpha t} + B\|\mu_{x,0}^\epsilon - \mu_x^\star\|_{\mathrm{BL}}^* + \frac{B^2\gamma}{2}t \tag{A.25}$$

$$-\frac{1}{t}\int_0^t \mathcal{L}(\mu_x(s),\mu_y(s))\,ds + \mathcal{L}(\bar{\mu}_x(t),\mu_y^\star) \leq \frac{\mathcal{H}(\mu_{y,0}^\epsilon,\mu_{y,0})}{4\alpha t} + B\|\mu_{y,0}^\epsilon - \mu_y^\star\|_{\mathrm{BL}}^* + \frac{B^2\gamma}{2}t \tag{A.26}$$

Equation (A.26) is obtained by performing the same argument switching the roles of x and y, and

$\mathcal{L}$ by $-\mathcal{L}$. By adding equations (A.25) and (A.26) and considering the definition of Q in (A.15), we obtain the inequality (A.17).

□

Notice that by taking the supremum wrt $\mu_x^\star, \mu_y^\star$ on (A.17) we obtain a bound on the Nikaido-Isoda error of $(\bar{\mu}_x(t), \bar{\mu}_y(t))$ (see (2.2)).

Next, we will obtain a result like Lemma E.1 from Chizat [2019] in which we bound Q. The proof is a variation of the argument in Lemma E.1 from Chizat [2019], as in our case no measures are necessarily sparse.

Lemma A.17. *Let Θ be a Riemannian manifold of dimension d. Assume that $Vol(B_{\theta,\epsilon}) \geq e^{-K}\epsilon^d$ for all $\theta \in \Theta$, where the volume is defined of course in terms of the Borel measure[1] of Θ. If $\rho := \frac{d\mu_0}{d\theta}$ is the Radon-Nikodym derivative of μ_0 with respect to the Borel measure of Θ, assume that $\rho(\theta) \geq e^{-K'}$ for all $\theta \in \Theta$. The function $Q_{\mu^\star,\mu_0}(\tau)$ defined in (A.15) can be bounded by*

$$Q_{\mu^\star,\mu_0}(\tau) \leq \frac{d}{\tau}(1 - \log d + \log \tau) + \frac{1}{\tau}(K + K')$$

Proof. We will choose μ^ϵ in order to bound the infimum. For $\theta \in \Theta, \epsilon > 0$, let $\xi_{\theta,\epsilon}$ be a probability measure on Θ with support on the ball $B_{\theta,\epsilon}$ of radius ϵ centered at θ and proportional to the Borel measure for all subsets of the ball. Let us define the measure

$$\mu^\epsilon(A) = \int_\Theta \xi_{\theta,\epsilon}(A)\, d\mu^\star(\theta)$$

for all Borel sets A of $\mathcal{X}$. Now, we can bound $\|\mu^\epsilon - \mu^\star\|_{\mathrm{BL}}^* \leq W_1(\mu^\epsilon, \mu^\star)$. Let us consider the

[1]The metric of the manifold gives a natural choice of a Borel (volume) measure, the one given by integrating the canonical volume form.

coupling γ between μ^ϵ and $\mu^\star$ defined as:

$$\gamma(A \times B) = \int_A \xi_{\theta,\epsilon}(B)\, d\mu^\star(\theta)$$

for A, B arbitrary Borel sets of Θ. Notice that γ is indeed a coupling between μ^ϵ and $\mu^\star$, because $\gamma(A \times \Theta) = \mu^\star(A)$ and $\gamma(\Theta \times B) = \mu^\epsilon(B)$. Hence,

$$W_1(\mu^\epsilon, \mu^\star) \leq \int_{\Theta\times\Theta} d_\Theta(\theta, \theta')\, d\gamma(\theta, \theta') = \int_\Theta \frac{1}{\mathrm{Vol}(B_{\theta',\epsilon})} \int_{B_{\theta',\epsilon}} d_\Theta(\theta, \theta')\, d\theta\, d\mu^\star(\theta') \tag{A.27}$$

where the inner integral is with respect to the Borel measure on Θ. Since $d_\Theta(\theta, \theta') \leq \epsilon$ for all $\theta \in B_{\theta',\epsilon}$, we conclude from that (A.27) that $W_1(\mu^\epsilon, \mu^\star) \leq \epsilon$.

Next, let us bound the relative entropy term. Define ρ_ϵ as the Radon-Nikodym derivative of μ^ϵ with respect to the Borel measure of Θ, i.e.

$$\rho_\epsilon(\theta) := \frac{d\mu^\epsilon}{d\theta}(\theta) = \int_\Theta \frac{1}{\mathrm{Vol}(B_{\theta',\epsilon})} \mathbb{1}_{B_{\theta',\epsilon}}(\theta)\, d\mu^\star(\theta').$$

Also, recall that $\rho := \frac{d\mu_0}{d\theta}$. Then, we write

$$\mathcal{H}(\mu^\epsilon, \mu_0) = \int_\Theta \log \frac{\rho_\epsilon}{\rho} d\mu^\epsilon = \int_\Theta \log(\rho_\epsilon)\rho_\epsilon d\theta - \int_\Theta \log(\rho)\rho_\epsilon d\theta. \tag{A.28}$$

On the one hand, we use the convexity of the function $x \to x \log x$:

$$\begin{aligned}\rho_\epsilon(\theta) \log \rho_\epsilon(\theta) &= \left(\int_\Theta \frac{1}{\mathrm{Vol}(B_{\theta',\epsilon})} \mathbb{1}_{B_{\theta',\epsilon}}(\theta)\, d\mu^\star(\theta')\right) \log \left(\int_\Theta \frac{1}{\mathrm{Vol}(B_{\theta',\epsilon})} \mathbb{1}_{B_{\theta',\epsilon}}(\theta)\, d\mu^\star(\theta')\right) \\ &\leq \int_\Theta \left(\frac{1}{\mathrm{Vol}(B_{\theta',\epsilon})} \mathbb{1}_{B_{\theta',\epsilon}}(\theta)\right) \log \left(\frac{1}{\mathrm{Vol}(B_{\theta',\epsilon})} \mathbb{1}_{B_{\theta',\epsilon}}(\theta)\right) d\mu^\star(\theta').\end{aligned}$$

We use Fubini's theorem:

$$\int_\Theta \rho_\epsilon(\theta)\log\rho_\epsilon(\theta)\,d\theta \le \int_\Theta\int_\Theta \left(\frac{1}{\mathrm{Vol}(B_{\theta',\epsilon})}\mathbb{1}_{B_{\theta',\epsilon}}(\theta)\right)\log\left(\frac{1}{\mathrm{Vol}(B_{\theta',\epsilon})}\mathbb{1}_{B_{\theta',\epsilon}}(\theta)\right)\,d\theta\,d\mu^\star(\theta')$$
$$= \int_\Theta \frac{1}{\mathrm{Vol}(B_{\theta',\epsilon})}\int_{B_{\theta',\epsilon}} -\log\left(\mathrm{Vol}(B_{\theta',\epsilon})\right)\,d\theta\,d\mu^\star(\theta') = -\int_\Theta \log\left(\mathrm{Vol}(B_{\theta',\epsilon})\right)d\mu^\star(\theta')$$
$$\le -d\log\epsilon + K \tag{A.29}$$

where d is the dimension of Θ and K is a constant such that $\mathrm{Vol}(B_{\theta',\epsilon}) \ge e^{-K}\epsilon^d$ for all $\theta' \in \Theta$.

On the other hand,

$$-\int_\Theta \log(\rho(\theta))\rho_\epsilon(\theta)\,d\theta = \int_\Theta \frac{1}{\mathrm{Vol}(B_{\theta',\epsilon})}\int_{\mathrm{Vol}(B_{\theta',\epsilon})} -\log(\rho(\theta))\,d\theta\,d\mu^\star(\theta')$$
$$\le \int_\Theta \frac{1}{\mathrm{Vol}(B_{\theta',\epsilon})}\int_{\mathrm{Vol}(B_{\theta',\epsilon})} K'\,d\theta\,d\mu^\star(\theta') = K' \tag{A.30}$$

where K' is defined such that $\rho(\theta) \ge e^{-K'}$ for all $\theta \in \Theta$.

By plugging (A.29) and (A.30) into (A.28) we obtain:

$$\|\mu^\star - \mu^\epsilon\|_{\mathrm{BL}}^* + \frac{1}{\tau}\mathcal{H}(\mu^\epsilon,\mu_0) \le \epsilon + \frac{1}{\tau}(-d\log\epsilon + K + K').$$

If we optimize the bound with respect to ϵ we obtain the final result. □

Theorem 2.3. *Let $\epsilon > 0$ arbitrary. Suppose that $\mu_{x,0}, \mu_{y,0}$ are such that their Radon-Nikodym derivatives with respect to the Borel measures of $\mathcal{X}, \mathcal{Y}$ are lower-bounded by $e^{-K'_x}, e^{-K'_y}$ respectively. For any $\delta \in (0, 1/2)$, there exists a constant $C_{\delta,\mathcal{X},\mathcal{Y},K'_x,K'_y} > 0$ depending on the dimensions of $\mathcal{X}, \mathcal{Y}$, their curvatures and K'_x, K'_y, such that if $\gamma/\alpha < 1$ and*

$$\frac{\gamma}{\alpha} \le \left(\frac{\epsilon}{C_{\delta,\mathcal{X},\mathcal{Y},K'_x,K'_y}}\right)^{\frac{2}{1-\delta}} \tag{A.31}$$

Then, at $t_0 = (\alpha\gamma)^{-1/2}$ *we have*

$$NI(\bar{\mu}_x(t_0), \bar{\mu}_y(t_0)) := \sup_{\mu_x^\star, \mu_y^\star} \mathcal{L}(\bar{\mu}_x(t_0), \mu_y^\star) - \mathcal{L}(\mu_x^\star, \bar{\mu}_y(t_0)) \leq \epsilon$$

Proof. We plug the bound of Theorem A.17 into the result of Theorem A.16, obtaining

$$\begin{aligned}\mathcal{L}(\bar{\mu}_x(t), \mu_y^\star) - \mathcal{L}(\mu_x^\star, \bar{\mu}_y(t)) \leq & \frac{d_x}{\alpha t}(1 - \log d_x + \log(\alpha B t)) \\ & + \frac{d_y}{\alpha t}(1 - \log d_y + \log(\alpha B t)) \\ & + \frac{1}{\alpha t}(K_x + K_x' + K_y + K_y') + \gamma B^2 t\end{aligned}$$

Now, we set $t = (\alpha\gamma)^{-1/2}$, and thus the right hand side becomes

$$\sqrt{\frac{\gamma}{\alpha}}\left(d_x\left(1 - \log\frac{d_x}{B} + \log\sqrt{\frac{\alpha}{\gamma}}\right) + d_y\left(1 - \log\frac{d_y}{B} + \log\sqrt{\frac{\alpha}{\gamma}}\right) + K_x + K_x' + K_y + K_y' + B^2\right) \tag{A.32}$$

Let $\epsilon > 0$ arbitrary. We want (A.32) to be lower or equal than ϵ. For any δ such that $0 < \delta < 1/2$, there exists C_δ such that $\log(x) \leq C_\delta x^\delta$. This yields

$$\begin{aligned}&\sqrt{\frac{\gamma}{\alpha}}\left(d_x\left(1 - \log\frac{d_x}{B} + C_\delta\left(\frac{\alpha}{\gamma}\right)^{-\delta/2}\right) + d_y\left(1 - \log\frac{d_y}{B} + C_\delta\left(\frac{\alpha}{\gamma}\right)^{-\delta/2}\right)\right) \\ &+ \sqrt{\frac{\gamma}{\alpha}}\left(K_x + K_x' + K_y + K_y' + B^2\right)\end{aligned} \tag{A.33}$$

If we set $\gamma < \alpha$, $(\gamma/\alpha)^{-\delta/2} > 1$ then (A.33) is upper-bounded by

$$\left(\frac{\gamma}{\alpha}\right)^{\frac{1-\delta}{2}}\left(d_x(1 - \log\frac{d_x}{B} + C_\delta) + d_y(1 - \log\frac{d_y}{B} + C_\delta) + K_x + K_x' + K_y + K_y' + B^2\right)$$

If we bound this by ϵ, we obtain the bound in (A.31). □

Corollary A.18. *Let $(\mathcal{X}_{d_x}, \mathcal{Y}_{d_y}, l_{d_x,d_y})_{d_x\in\mathbb{N},d_y\in\mathbb{N}}$ be a family indexed by $\mathbb{N}^2$. Assume that $\mu_{x,0}, \mu_{y,0}$ are set to be the Borel measures in $\mathcal{X}_{d_x}, \mathcal{Y}_{d_y}$, that $\mathcal{X}_{d_x}, \mathcal{Y}_{d_y}$ are locally isometric to the d_x, d_y-dimensional Euclidean spaces, and that the volumes of $\mathcal{X}_{d_x}, \mathcal{Y}_{d_y}$ grow no faster than exponentially on the dimensions d_x, d_y. Assume that l_{d_x,d_y} are such that B is constant. Then, we can rewrite (A.31) as*

$$\frac{\gamma}{\alpha} \leq O\left(\left(\frac{\epsilon}{(d_x+d_y)\log(B)+d_x\log(d_x)+d_y\log(d_y)+B^2}\right)^{\frac{2}{1-\delta}}\right)$$

Proof. The volume of n-dimensional ball of radius r in n-dimensional Euclidean space is

$$V_n(r) = \frac{\pi^{n/2}}{\Gamma(\frac{n}{2}+1)}R^n,$$

and hence, if $\mathcal{X}, \mathcal{Y}$ are locally isometric to the d_x and d_y-dimensional Euclidean spaces we can take

$$K_x = \log\Gamma\left(\frac{d_x}{2}+1\right) - \frac{d_x}{2}\log(\pi) \leq \left(\frac{d_x}{2}+1\right)\log\left(\frac{d_x}{2}+1\right) - \frac{d_x}{2}\log(\pi) \leq O(d_x\log d_x)$$
$$K_y = \log\Gamma(\frac{d_y}{2}+1) - \frac{n}{2}\log(\pi) \leq O(d_x\log d_x)$$

If the volumes of $\mathcal{X}, \mathcal{Y}$ grow no faster than an exponential of the dimensions d_x, d_y and we take $\mu_{x,0}, \mu_{y,0}$ to be the Borel measures, we can take $K'_x = \log(\mathrm{Vol}(\mathcal{X})), K'_y = \log(\mathrm{Vol}(\mathcal{Y}))$ to be constant with respect to the dimensions d_x, d_y. □

A.5 Proof of Theorem 2.4(i)

A.5.1 Preliminaries

Throughout the section we will use the techniques shown in subsection A.7.5 to deal with SDEs on manifolds. Effectively, this means that for SDEs we have additional drift terms $\hat{\mathbf{h}}_x$ or $\hat{\mathbf{h}}_x$ induced

by the geometry of the manifold, and that we must project the variations of the Brownian motion onto the tangent space.

Define the processes $\mathbf{X}^n = (X^1, \dots, X^n)$ and $\mathbf{Y}^n = (Y^1, \dots, Y^n)$ such that for all $i \in \{1, \dots, n\}$,

$$\begin{aligned} dX_t^i &= \left(-\frac{1}{n} \sum_{j=1}^{n} \nabla_x \ell(X_t^i, Y_t^j) + \hat{\mathbf{h}}_x(X_t^i) \right) dt + \sqrt{2\beta^{-1}} \operatorname{Proj}_{T_{X_t^i}\mathcal{X}}(dW_t^i), \quad X_0^{n,i} = \xi^i \sim \mu_{x,0} \\ dY_t^i &= \left(\frac{1}{n} \sum_{j=1}^{n} \nabla_y \ell(X_t^j, Y_t^i) + \hat{\mathbf{h}}_y(Y_t^i) \right) dt + \sqrt{2\beta^{-1}} \operatorname{Proj}_{T_{Y_t^i}\mathcal{Y}}(d\bar{W}_t^i), \quad Y_0^{n,i} = \bar{\xi}^i \sim \mu_{y,0} \end{aligned} \tag{A.34}$$

where $\mathbf{W}_t = (W_t^1, \dots, W_t^n)$, and $\bar{\mathbf{W}}_t = (\bar{W}_t^1, \dots, \bar{W}_t^n)$ are Brownian motions on $\mathbb{R}^{nD_x}$ and $\mathbb{R}^{nD_y}$ respectively. Notice that $\mathbf{X}_t$ is valued in $\mathcal{X}^n \subseteq \mathbb{R}^{nD_x}$ and $\mathbf{Y}_t$ is valued in $\mathcal{Y}^n \subseteq \mathbb{R}^{nD_y}$. (A.34) can be seen as a system of $2n$ interacting particles in which each particle of one player interacts with all the particles of the other one. It also corresponds to noisy continuous-time mirror descent on parameter spaces for an augmented game in which there are n replicas of each player, choosing $\frac{1}{2}\|\cdot\|_2^2$ for the mirror map.

Now, define $\tilde{\mathbf{X}} = (\tilde{X}^1, \dots, \tilde{X}^n)$ and $\tilde{\mathbf{Y}} = (\tilde{Y}^1, \dots, \tilde{Y}^n)$ for all $i \in \{1, \dots, n\}$ let

$$\begin{aligned} d\tilde{X}_t^i &= \left(-\int_{\mathcal{Y}} \nabla_x \ell(\tilde{X}_t^i, y)\, d\mu_{y,t} + \hat{\mathbf{h}}_x(\tilde{X}_t^i) \right) dt + \sqrt{2\beta^{-1}} \operatorname{Proj}_{T_{\tilde{X}_t^i}\mathcal{X}}(dW_t^i), \\ d\tilde{Y}_t^i &= \left(\int_{\mathcal{X}} \nabla_y \ell(x, \tilde{Y}_t^i)\, d\mu_{x,t} + \hat{\mathbf{h}}_y(\tilde{Y}_t^i) \right) dt + \sqrt{2\beta^{-1}} \operatorname{Proj}_{T_{\tilde{Y}_t^i}\mathcal{Y}}(d\bar{W}_t^i), \\ \tilde{X}_0^i &= \xi^i \sim \mu_{x,0}, \quad \mu_{y,t} = \operatorname{Law}(\tilde{Y}_t^i), \quad \tilde{Y}_0^i = \bar{\xi}^i \sim \mu_{y,0}, \quad \mu_{x,t} = \operatorname{Law}(\tilde{X}_t^i) \end{aligned} \tag{A.35}$$

Lemma A.19 (Forward Kolmogorov equation). *The laws $(\mu_x)_{t\in[0,T]}, (\mu_y)_{t\in[0,T]}$ of a solution $\tilde{X}, \tilde{Y}$*

of (A.35) *with* $n = 1$ *(seen as elements of* $C([0,T],\mathcal{P}(\mathcal{X})), C([0,T],\mathcal{P}(\mathcal{Y}))$*) are a solution of* (A.36)*.*

$$\begin{cases} \partial_t \mu_x = \nabla_x \cdot (\mu_x \nabla_x V_x(\mu_y, x)) + \beta^{-1} \Delta_x \mu_x, \quad \mu_x(0) = \mu_{x,0} \\ \partial_t \mu_y = -\nabla_y \cdot (\mu_y \nabla_y V_y(\mu_x, y)) + \beta^{-1} \Delta_y \mu_y, \quad \mu_y(0) = \mu_{y,0} \end{cases} \tag{A.36}$$

Proof. We sketch the derivation for the forward Kolmogorov equation on manifolds. First, we define the semigroups

$$P_t^x \psi_x(x) = \mathbb{E}[\psi_x(\tilde{X}_t) | \tilde{X}_0 = x], \quad P_t^y \psi_y(y) = \mathbb{E}[\psi_y(\tilde{Y}_t) | \tilde{Y}_0 = y],$$

where $\tilde{X}, \tilde{Y}$ are solutions of (A.35) with $n = 1$. We obtain that if $\mathcal{L}_t^x, \mathcal{L}_t^y$ are the infinitesimal generators (i.e., $\mathcal{L}_t^x \psi_x(x) = \lim_{t \to 0^+} \frac{1}{t}(P_t^x \psi_x(x) - \psi_x(x))$), the backward Kolmogorov equations $\frac{d}{dt} P_t^x \psi_x(x) = \mathcal{L}_t^x P_t^x \psi_x(x)$, $\frac{d}{dt} P_t^y \psi_y(y) = \mathcal{L}_t^y P_t^y \psi_y(y)$ hold for ψ_x, ψ_y in the domains of the generators. Since $\mathcal{L}_t^x$ and P_t^x commute for these choices of ψ_x, we have $\frac{d}{dt} P_t^x \psi_x(x) = P_t^x \mathcal{L}_t^x \psi_x(x)$, $\frac{d}{dt} P_t^y \psi_y(y) = P_t^y \mathcal{L}_t^y \psi_y(y)$. By integrating these two equations over the initial measures $\mu_{x,0}, \mu_{y,0}$, we get

$$\frac{d}{dt} \int \psi_x(x) \, d\mu_{x,t} = \int \mathcal{L}_t^x \psi_x(x) \, d\mu_{x,t}, \quad \frac{d}{dt} \int \psi_y(y) \, d\mu_{y,t} = \int \mathcal{L}_t^y \psi_y(y) \, d\mu_{y,t}.$$

We can write an explicit form for $\mathcal{L}_t^x P_t^x \psi_x(x)$ by using Itô's lemma on (A.35):

$$\mathcal{L}_t^x \psi_x(x) = \left(\int_{\mathcal{Y}} \nabla_x \ell(x,y) \, d\mu_{y,s} \, ds - \hat{\mathbf{h}}_x(x) \right) \nabla_x \psi_x(x) + \beta^{-1} \mathrm{Tr} \left(\left(\mathrm{Proj}_{T_x \mathcal{X}} \right)^\top H \psi_x(x) \, \mathrm{Proj}_{T_x \mathcal{X}} \right),$$

where we use $\mathrm{Proj}_{T_{\tilde{X}_t^i} \mathcal{X}}$ to denote its matrix in the canonical basis.

Let $\{\xi_k\}$ be a partition of unity for $\mathcal{X}$ (i.e. a set of functions such that $\sum_k \xi_k(x) = 1$) in which each

ξ_k is regular enough and supported on a patch of $\mathcal{X}$. We can write

$$\begin{aligned}\frac{d}{dt}\int_{\mathcal{X}} \psi_x(x)\, d\mu_{x,t}(x) = \frac{d}{dt}\int_{\mathcal{X}} \psi_x(x)\, d\mu_{x,t}(x) &= \sum_k \frac{d}{dt}\int_{\mathcal{X}} \xi_k(x)\psi_x(x)\, d\mu_{x,t}(x) \\ &= \sum_k \int \mathcal{L}_t^x(\xi_k(x)\psi_x(x))\, d\mu_{x,t}\end{aligned}$$

Now, let $\tilde{\psi}_x^k(x) = \xi_k(x)\psi_x(x)$.

$$\begin{aligned}&\int_{\mathcal{X}} \mathcal{L}_t^x \tilde{\psi}_x^k(x)\, d\mu_{x,t} \\ &= \int_{\mathcal{X}} \left(\nabla_x V_x(\mu_{y,s}, x) - \hat{\mathbf{h}}_x(x)\right) \nabla_x \tilde{\psi}_x^k(x) + \beta^{-1}\mathrm{Tr}\left(\left(\mathrm{Proj}_{T_x\mathcal{X}}\right)^\top H\tilde{\psi}_x^k(x)\, \mathrm{Proj}_{T_x\mathcal{X}}\right) d\mu_{x,t}\end{aligned}$$

Notice that this equation is analogous to (A.57). We reverse the argument made in subsection A.7.5. Using the fact that the support of $\tilde{\psi}_x^k(x)$ is contained on some patch of $\mathcal{X}$ given by the mapping $\psi_k : U_{\mathbb{R}^d} \subseteq \mathbb{R}^d \to U \subseteq \mathcal{X} \subseteq \mathbb{R}^D$, the corresponding Fokker-Planck on $U_{\mathbb{R}^d}$ is

$$\begin{aligned}&\frac{d}{dt}\int_{U_{\mathbb{R}^d}} \tilde{\psi}_x^k(\psi_k(q))\, d(\psi_k^{-1})_* \mu_{x,t}(q) \\ &= \int_{U_{\mathbb{R}^d}} \nabla V_x(\mu_{y,s}, \psi_k(q)) \cdot \nabla\tilde{\psi}_x^k(\psi_k(q)) + \beta^{-1}\Delta\tilde{\psi}_x^k(\psi_k(q))\, d(\psi_k^{-1})_*\mu_{x,t}(q),\end{aligned}$$

where the gradients and the Laplacian are in the metric inherited from the embedding (as in subsection A.7.5). The pushforward definition implies

$$\frac{d}{dt}\int_{\mathcal{X}} \tilde{\psi}_x^k(x)\, d\mu_{x,t}(x) = \int_{U_{\mathbb{R}^d}} \nabla V_x(\mu_{y,s}, x) \cdot \nabla\tilde{\psi}_x^k(x) + \beta^{-1}\Delta\tilde{\psi}_x^k(x)\, d\mu_{x,t}(x),$$

By substituting $\tilde{\psi}^k_x(x) = \xi_k(x)\psi_x(x)$, summing for all k and using $\sum_k \xi_k(x) = 1$, we obtain:

$$\frac{d}{dt}\int_{\mathcal{X}} \psi_x(x)\, d\mu_{x,t}(x) = \int_{\mathcal{X}} \nabla_x V_x(\mu_{y,s}, x)\cdot\nabla_x\psi_x(x) + \beta^{-1}\Delta_x\psi_x(x)\, d\mu_{x,t}(x)$$

which is the same as the first equation in (2.7). The second equation is obtained analogously. □

Let $\mu^n_x = \frac{1}{n}\sum_{i=1}^n \delta_{X^i}$ be a $\mathcal{P}(C([0,T],\mathcal{X}))$-valued random element that corresponds to the empirical measure of a solution $\mathbf{X}^n$ of (A.34). Analogously, let $\mu^n_y = \frac{1}{n}\sum_{i=1}^n \delta_{Y^i}$ be a $\mathcal{P}(C([0,T],\mathcal{Y}))$-valued random element corresponding to the empirical measure of $\mathbf{Y}^n$.

Define the 2-Wasserstein distance on $\mathcal{P}(C([0,T],\mathcal{X}))$ as

$$\mathcal{W}_2^2(\mu,\nu) := \inf_{\pi\in\Pi(\mu,\nu)}\int_{C([0,T],\mathcal{X})^2} d(x,y)^2\, d\pi(x,y) \tag{A.37}$$

where $d(x,y) = \sup_{t\in[0,T]} d_{\mathcal{X}}(x(t),y(t))$. Define it analogously on $\mathcal{P}(C([0,T],\mathcal{Y}))$.

We state a stronger version of the law of large numbers in the first statement of Theorem 2.4(i).

Theorem A.20. *There exists a solution of the coupled McKean-Vlasov SDEs (A.35). Pathwise uniqueness and uniqueness in law hold. Let $\mu_x \in \mathcal{P}(C([0,T],\mathcal{X}))$, $\mu_y \in \mathcal{P}(C([0,T],\mathcal{Y}))$ be the unique laws of the solutions for $n = 1$ (all pairs have the same solutions). Then,*

$$\mathbb{E}[\mathcal{W}_2^2(\mu^n_x,\mu_x) + \mathcal{W}_2^2(\mu^n_y,\mu_y)] \xrightarrow{n\to\infty} 0$$

Let us comment on why Theorem A.20 implies the first statement in Theorem 2.4(i). We make use of the mapping $\mathcal{P}(C([0,T],\mathcal{X})) \ni \mu \mapsto (\mu_t)_{t\in[0,T]} \in C([0,T],\mathcal{P}(\mathcal{X}))$ into the time marginals. By the definition (A.37), $\sup_{t\in[0,t]} \mathcal{W}_2^2(\mu^n_{x,t},\mu_{x,t}) \le \mathcal{W}_2^2(\mu^n_x,\mu_x)$ and the same holds for μ^n_y,μ_y. At this point, Theorem A.19 states that $(\mu_x)_{t\in[0,T]}, (\mu_y)_{t\in[0,T]}$ is a solution of the mean-field ERIWGF (A.36) and concludes the argument. The proof of Theorem A.20 uses a propagation of chaos

argument, originally due to Sznitman [1991] in the context of interacting particle systems. Our argument follows Theorem 3.3 of Lacker [2018].

A.5.2 Existence and uniqueness

We prove existence and uniqueness of the system given by

$$\begin{aligned}
\tilde{X}_t &= \int_0^t \left(-\int_{\mathcal{Y}} \nabla_x \ell(\tilde{X}_s, y)\, d\mu_{y,s}\, ds + \hat{\mathbf{h}}_x(\tilde{X}_s) \right) ds + \sqrt{2\beta^{-1}} \int_0^t \mathrm{Proj}_{T_{\tilde{X}_s}\mathcal{X}}(dW_s), \\
\tilde{Y}_t &= \int_0^t \left(\int_{\mathcal{X}} \nabla_y \ell(x, \tilde{Y}_s)\, d\mu_{x,s} + \hat{\mathbf{h}}_y(Y_s^{n,i}) \right) ds + \sqrt{2\beta^{-1}} \int_0^t \mathrm{Proj}_{T_{\tilde{Y}_s}\mathcal{Y}}(d\bar{W}_s), \\
\mu_{x,t} &= \mathrm{Law}(\tilde{X}_t^n), \quad \mu_{y,t} = \mathrm{Law}(\tilde{Y}_t^n), \quad \tilde{X}_0 = \xi \sim \mu_{x,0}, \quad \tilde{Y}_0 = \bar{\xi} \sim \mu_{y,0}.
\end{aligned} \tag{A.38}$$

Path-wise uniqueness means that given $W, \bar{W}, \xi, \bar{\xi}$, two solutions are equal almost surely. Uniqueness in law means that regardless of the Brownian motion and the initialization random variables chosen (as long as they are $\mu_{x,0}$ and $\mu_{y,0}$-distributed), the law of the solution is unique. We prove that both hold for (A.38).

We have that for all $x, x' \in \mathcal{X}, \mu, \nu \in \mathcal{P}(\mathcal{Y})$,

$$\left| \int \nabla_x \ell(x, y)\, d\mu - \int \nabla_x \ell(x', y)\, d\nu \right| \leq L(d(x, x') + \mathcal{W}_2(\mu, \nu)) \tag{A.39}$$

This is obtained by adding and subtracting the term $\int \nabla_x \ell(x'y)\, d\mu$, by using the triangle inequality and the inequality $\mathcal{W}_1(\mu, \nu)) \leq \mathcal{W}_2(\mu, \nu))$ (which is proven using the Cauchy-Schwarz inequality). Hence,

$$\left| \int \nabla_x \ell(x, y)\, d\mu - \int \nabla_x \ell(x', y)\, d\nu \right|^2 \leq 2L^2(d(x, x')^2 + \mathcal{W}_2^2(\mu, \nu)) \tag{A.40}$$

On the other hand, using the regularity of the manifold, there exists $\mathcal{L}_{\mathcal{X}}$ such that

$$|\hat{\mathbf{h}}_x(x) - \hat{\mathbf{h}}_x(x')| \leq L_{\mathcal{X}} d(x, x'),$$
$$|\mathrm{Proj}_{T_x\mathcal{X}} - \mathrm{Proj}_{T_{x'}\mathcal{X}}| \leq L_{\mathcal{X}} d(x, x')$$

where $\mathrm{Proj}_{T_x\mathcal{X}}$ denotes its matrix in the canonical basis and the norm in the second line is the Frobenius norm. Also, let $\|x - x'\|$ be the Euclidean norm of $\mathcal{X}$ in $\mathbb{R}^{D_x}$ (the Euclidean space where $\mathcal{X}$ is embedded) and let $K_{\mathcal{X}} > 1$ be such that $d(x, x') \leq K_{\mathcal{X}} \|x - x'\|$.

Let $\mu_y, \nu_y \in \mathcal{P}(C([0,T], \mathcal{X}))$ and let X^{μ_y}, X^{ν_y} be the solutions of the first equation of (A.38) when we plug μ_y (ν_y resp.) as the measure for the other player. X^{μ_y} and X^{ν_y} exist and are unique by the classical theory of SDEs (see Chapter 18 of Kallenberg [2002]). Following the procedure in Theorem 3.3 of Lacker [2018], we obtain

$$\begin{aligned}
\mathbb{E}[\|X^{\mu_y} - X^{\nu_y}\|_t^2] &\leq 3t\mathbb{E}\left[\int_0^t \left| \int \nabla_x \ell(X^{\mu_y}, y)\, d\mu_{y,r} - \int \nabla_x \ell(X^{\nu_y}, y)\, d\nu_{y,r} \right|^2 dr\right] \\
&\quad + 3t\mathbb{E}\left[\int_0^t |\hat{\mathbf{h}}_x(X^{\mu_y}) - \hat{\mathbf{h}}_x(X^{\nu_y})|^2\, dr\right] \\
&\quad + 12\mathbb{E}\left[\int_0^t |\mathrm{Proj}_{T_x\mathcal{X}} - \mathrm{Proj}_{T_{x'}\mathcal{X}}|^2\, dr\right] \\
&\leq 3(3t+4)\tilde{L}^2\mathbb{E}\left[\int_0^t (\|X^{\mu_y} - X^{\nu_y}\|_r^2 + \mathcal{W}_2^2(\mu_{y,r}, \nu_{y,r}))\, dr\right], \qquad \text{(A.41)}
\end{aligned}$$

where $\tilde{L}^2 = (L^2 + L_{\mathcal{X}}^2)K_{\mathcal{X}}^2$. Using Fubini's theorem and Gronwall's inequality, we obtain

$$\mathbb{E}[\|X^{\mu_y} - X^{\nu_y}\|_t^2] \leq 3(3T+4)\tilde{L}^2 \exp(3(3T+4)\tilde{L}^2) \int_0^t \mathcal{W}_2^2(\mu_{y,r}, \nu_{y,r}))\, dr \qquad \text{(A.42)}$$

Let $C_T := 3(3T+4)\tilde{L}^2 \exp(3(3T+4)\tilde{L}^2)$. For $\mu, \nu \in \mathcal{P}(C([0,T], \mathcal{X}))$, define

$$\mathcal{W}_{2,t}^2(\mu, \nu) := \inf_{\pi \in \Pi(\mu,\nu)} \int_{C([0,T],\mathcal{X})^2} \sup_{r \in [0,t]} d(x(r), y(r))\, \pi(dx, dy)$$

Hence, (A.42) and the bound $\mathcal{W}_2^2(\mu_{y,r}, \nu_{y,r}) \leq \mathcal{W}_{2,r}^2(\mu_y, \nu_y)$ yield

$$\mathbb{E}[\|X^{\mu_y} - X^{\nu_y}\|_t^2] \leq C_T \int_0^t \mathcal{W}_{2,r}^2(\mu_y, \nu_y)\, dr$$

Reasoning analogously for the other player, we obtain

$$\mathbb{E}[\|X^{\mu_y} - X^{\nu_y}\|_t^2 + \|Y^{\mu_x} - Y^{\nu_x}\|_t^2] \leq C_T \int_0^t \mathcal{W}_{2,r}^2(\mu_y, \nu_y)\, dr + C_T \int_0^t \mathcal{W}_{2,r}^2(\mu_x, \nu_x)\, dr$$

Given $\mu_y \in \mathcal{P}(C([0,T], \mathcal{Y}))$, define $\Phi_x(\mu_y) = \text{Law}(X^{\mu_y}) \in \mathcal{P}(C([0,T], \mathcal{X}))$, and define Φ_y analogously. Notice that $\mathcal{W}_{2,t}^2(\Phi_x(\mu_y), \Phi_x(\nu_y)) \leq \mathbb{E}[\|X^{\mu_y} - X^{\nu_y}\|_t^2], \mathcal{W}_{2,t}^2(\Phi_y(\mu_x), \Phi_y(\nu_x)) \leq \mathbb{E}[\|X^{\mu_x} - X^{\nu_x}\|_t^2]$. Hence, we obtain

$$\mathcal{W}_{2,t}^2(\Phi_x(\mu_y), \Phi_x(\nu_y)) + \mathcal{W}_{2,t}^2(\Phi_y(\mu_x), \Phi_y(\nu_x)) \leq C_T \int_0^t \mathcal{W}_{2,r}^2(\mu_y, \nu_y) + \mathcal{W}_{2,r}^2(\mu_x, \nu_x)\, dr$$

Observe that $\mathcal{W}_{2,t}^2(\mu_x, \nu_x) + \mathcal{W}_{2,t}^2(\mu_y, \nu_y)$ is the square of a distance between (μ_x, μ_y) and (ν_x, ν_y) on $\mathcal{P}(C([0,T], \mathcal{X})) \times \mathcal{P}(C([0,T], \mathcal{Y}))$. Hence, we can apply the Piccard iteration argument to obtain the existence result and another application of Gronwall's inequality yields pathwise uniqueness. Uniqueness in law (i.e., regardless of the specific Brownian motions and initialization random variables) follows from the typical uniqueness in law result for SDEs (see Chapter 18 of Kallenberg [2002] for example). The idea is that when we solve the SDEs with $W', \bar{W}', \xi', \bar{\xi}'$ plugging in the drift the laws of a solution for $W, \bar{W}, \xi, \bar{\xi}$, the solution has the same laws by uniqueness in law of SDEs. Hence, that new solution solves the coupled McKean-Vlasov for $W', \bar{W}', \xi', \bar{\xi}'$.

A.5.3 Propagation of chaos

Let $\mu_x^n = \frac{1}{n}\sum_{i=1}^n \delta_{X^i}, \mu_y^n = \frac{1}{n}\sum_{i=1}^n \delta_{Y^i}$. Using the argument from existence and uniqueness on the i-th components of $\mathbf{X}, \tilde{\mathbf{X}}$,

$$\mathbb{E}[\|X^i - \tilde{X}^i\|_t^2] \leq 3(3T+4)\tilde{L}^2\mathbb{E}\left[\int_0^t (\|X^i - \tilde{X}^i\|_r^2 + \mathcal{W}_2^2(\mu_{y,r}^n, \mu_{y,r}))\, dr\right]$$

Arguing as before, we obtain

$$\mathbb{E}[\|X^i - \tilde{X}^i\|_t^2] \leq C_T\mathbb{E}\left[\int_0^t \mathcal{W}_{2,r}^2(\mu_y^n, \mu_y)\, dr\right]$$

Let $\nu_x^n = \frac{1}{n}\sum_{i=1}^n \delta_{\tilde{X}^i}$ be the empirical measure of the mean field processes in (A.35). Notice that $\frac{1}{n}\sum_{i=1}^n \delta_{(X^i,\tilde{X}^i)}$ is a coupling between ν_x^n and μ_x^n, and so

$$\mathcal{W}_{2,t}^2(\mu_x^n, \nu_x^n) \leq \frac{1}{n}\sum_{i=1}^n \|X^i - \tilde{X}^i\|_t^2$$

Thus, we obtain

$$\mathbb{E}[\mathcal{W}_{2,t}^2(\mu_x^n, \nu_x^n)] \leq C_T\mathbb{E}\left[\int_0^t \mathcal{W}_{2,r}^2(\mu_y^n, \mu_y)\, dr\right]$$

We use the triangle inequality

$$\begin{aligned}\mathbb{E}[\mathcal{W}_{2,t}^2(\mu_x^n, \mu_x)] &\leq 2\mathbb{E}[\mathcal{W}_{2,t}^2(\mu_x^n, \nu_x^n)] + 2\mathbb{E}[\mathcal{W}_{2,t}^2(\nu_x^n, \mu_x)] \\ &\leq 2C_T\mathbb{E}\left[\int_0^t \mathcal{W}_{2,r}^2(\mu_y^n, \mu_y)\, dr\right] + 2\mathbb{E}[\mathcal{W}_{2,t}^2(\nu_x^n, \mu_x)]\end{aligned}$$

At this point we follow an analogous procedure for the other player and we end up with

$$\mathbb{E}[\mathcal{W}_{2,t}^2(\mu_x^n, \mu_x) + \mathcal{W}_{2,t}^2(\mu_y^n, \mu_y)] \leq 2C_T\mathbb{E}\left[\int_0^t \mathcal{W}_{2,r}^2(\mu_y^n, \mu_y) + \mathcal{W}_{2,r}^2(\mu_x^n, \mu_x)\, dr\right] + 2\mathbb{E}[\mathcal{W}_{2,t}^2(\nu_x^n, \mu_x) + \mathcal{W}_{2,t}^2(\nu_y^n, \mu_y)]$$

We use Fubini's theorem and Gronwall's inequality again.

$$\mathbb{E}[\mathcal{W}_{2,t}^2(\mu_x^n, \mu_x) + \mathcal{W}_{2,t}^2(\mu_y^n, \mu_y)] \leq 2\exp(2C_T T)\mathbb{E}[\mathcal{W}_{2,t}^2(\nu_x^n, \mu_x) + \mathcal{W}_{2,t}^2(\nu_y^n, \mu_y)]$$

If we set $t = T$ we get

$$\mathbb{E}[\mathcal{W}_2^2(\mu_x^n, \mu_x) + \mathcal{W}_2^2(\mu_y^n, \mu_y)] \leq 2\exp(2C_T T)\mathbb{E}[\mathcal{W}_2^2(\nu_x^n, \mu_x) + \mathcal{W}_2^2(\nu_y^n, \mu_y)]$$

and the factor $\mathbb{E}[\mathcal{W}_2^2(\nu_x^n, \mu_x) + \mathcal{W}_2^2(\nu_y^n, \mu_y)]$ goes to 0 as $n \to \infty$ by the law of large numbers (see Corollary 2.14 of [Lacker 2018]).

A.5.4 Convergence of the Nikaido-Isoda error

Corollary A.21. *For $t \in [0, T]$, if $\mu_{x,t}^n, \mu_{x,t}, \mu_{y,t}^n, \mu_{y,t}$ are the marginals of $\mu_x^n, \mu_x, \mu_y^n, \mu_y$ at time t, we have*

$$\mathbb{E}[|NI(\mu_{x,t}^n, \mu_{y,t}^n) - NI(\mu_{x,t}, \mu_{y,t})|] \xrightarrow{n\to\infty} 0$$

Proof. See Theorem A.3. □

A.6 Proof of Theorem 2.4(ii)

A.6.1 Preliminaries

Define the processes $\mathbf{X} = (X^1, \dots, X^n)$, $\mathbf{w}_x = (w_x^1, \dots, w_x^n)$ and $\mathbf{Y} = (Y^1, \dots, Y^n)$, $\mathbf{w}_y = (w_y^1, \dots, w_y^n)$ such that for all $i \in \{1, \dots, n\}$

$$\begin{aligned}
\frac{dX_t^i}{dt} &= -\gamma \frac{1}{n} \sum_{j=1}^n w_{y,t}^j \nabla_x \ell(X_t^i, Y_t^j), \quad X_0^i = \xi^i \sim \mu_{x,0} \\
\frac{dw_{x,t}^i}{dt} &= \alpha \left(-\frac{1}{n} \sum_{j=1}^n w_{y,t}^j \ell(X_t^i, Y_t^j) + \frac{1}{n^2} \sum_{k=1}^n \sum_{j=1}^n w_{y,t}^j w_{x,t}^k \ell(X_t^i, Y_t^j) \right) w_{x,t}^i, \quad w_{x,0}^i = 1 \\
\frac{dY_t^i}{dt} &= \gamma \frac{1}{n} \sum_{j=1}^n w_{x,t}^j \nabla_y \ell(X_t^j, Y_t^i), \quad Y_0^i = \bar{\xi}^i \sim \mu_{y,0} \\
\frac{dw_{y,t}^i}{dt} &= \alpha \left(\frac{1}{n} \sum_{j=1}^n w_{x,t}^j \ell(X_t^i, Y_t^j) - \frac{1}{n^2} \sum_{k=1}^n \sum_{j=1}^n w_{y,t}^j w_{x,t}^k \ell(X_t^i, Y_t^j) \right) w_{x,t}^i, \quad w_{y,0}^i = 1
\end{aligned} \tag{A.43}$$

Let $\nu_{x,t}^n = \frac{1}{n} \sum_{i=1}^n \delta_{(X_t^i, w_{x,t}^i)} \in \mathbb{P}(\mathcal{X} \times \mathbb{R}^+)$, $\nu_{y,t}^n = \frac{1}{n} \sum_{i=1}^n \delta_{(Y_t^i, r_{y,t}^{n,i})} \in \mathbb{P}(\mathcal{Y} \times \mathbb{R}^+)$. Let $\mu_{x,t}^n = \frac{1}{n} \sum_{i=1}^n w_{x,t}^i \delta_{X_t^i} \in \mathbb{P}(\mathcal{X})$, $\mu_{y,t}^n = \frac{1}{n} \sum_{i=1}^n w_{y,t}^i \delta_{Y_t^i} \in \mathbb{P}(\mathcal{Y})$ be the projections of $\nu_{x,t}^n$, $\nu_{y,t}^n$. Notice that we have changed the notation with respect to the main text, multiplying w_x^i by n: now $w_{x,0}^i = 1$ and $\sum_i w_{x,t}^i = n, \forall t \geq 0$ instead of $w_{x,0}^i = 1/n$ and $\sum_i w_{x,t}^i = 1, \forall t \geq 0$.

Let h_x, h_y be the projection operators, i.e. $h_x \nu_x = \int_{\mathcal{R}^+} w_x \nu_x(\cdot, w_x)$. We also define the mean field

processes $\tilde{\mathbf{X}}, \tilde{\mathbf{Y}}, \tilde{\mathbf{w}}_x, \tilde{\mathbf{w}}_y$ given component-wise by

$$\begin{aligned}
\frac{d\tilde{X}_t^i}{dt} &= -\gamma \nabla_x \int \ell(\tilde{X}_t^i, y) d\mu_{y,t}, \quad \tilde{X}_0^i = \xi^i \sim \mu_{x,0} \\
\frac{d\tilde{w}_{x,t}^i}{dt} &= \alpha \left(- \int \ell(\tilde{X}_t^i, y) d\mu_{y,t} + \mathcal{L}(\mu_{x,t}, \mu_{y,t}) \right) \tilde{w}_{x,t}^i, \quad \tilde{w}_{x,0}^i = 1 \\
\frac{d\tilde{Y}_t^i}{dt} &= \gamma \nabla_y \int \ell(x, \tilde{Y}_t^i) d\mu_{x,t}, \quad \tilde{Y}_0^i = \bar{\xi}^i \sim \mu_{y,0} \\
\frac{d\tilde{w}_{y,t}^i}{dt} &= \alpha \left(\int \ell(x, \tilde{Y}_t^i) d\mu_{x,t} - \mathcal{L}(\mu_{x,t}, \mu_{y,t}) \right) \tilde{w}_{x,t}^i, \quad \tilde{w}_{y,0}^i = 1 \\
\mu_{x,t} &= h_x \mathrm{Law}(\tilde{X}_t^i, \tilde{w}_{x,t}^i), \quad \mu_{y,t} = h_y \mathrm{Law}(\tilde{Y}_t^i, \tilde{w}_{y,t}^i)
\end{aligned} \tag{A.44}$$

for i between 1 and n.

Lemma A.22 (Forward Kolmogorov equation)**.** *If $\tilde{X}, \tilde{w}_x, \tilde{Y}, \tilde{w}_y$ is a solution of (A.44) with $n = 1$, then its laws ν_x, ν_y fulfill (A.1).*

Proof. Let $\psi_x : \mathcal{X} \times \mathbb{R}^+ \to \mathbb{R}$. Plug the laws ν_x, ν_y of the solution $(\tilde{X}, \tilde{w}_x), (\tilde{Y}, \tilde{w}_y)$ into the ODE (A.44). Let $\Phi_{x,t} = (X_{x,t}^\Phi, w_{x,t}^\Phi) : (\mathcal{X} \times \mathbb{R}^+) \to (\mathcal{X} \times \mathbb{R}^+)$ denote the flow that maps an initial condition of the ODE (A.44) to the corresponding solution at time t. Then, we can write $\nu_{x,t} = (\Phi_{x,t})_* \nu_{x,0}$, where $(\Phi_{x,t})_*$ is the pushforward. Hence,

$$\begin{aligned}
&\frac{d}{dt} \int_{\mathcal{X} \times \mathbb{R}^+} \psi_x(x, w_x) \, d\nu_{x,t}(x, w_x) \\
&= \frac{d}{dt} \int_{\mathcal{X} \times \mathbb{R}^+} \psi_x(\Phi_{x,t}(x, w_x)) \, d\nu_{x,0}(x, w_x) \\
&= \int_{\mathcal{X} \times \mathbb{R}^+} \left(\nabla_x \psi_x(\Phi_{x,t}(x, w_x)), \frac{d\psi_x}{dw_x}(\Phi_{x,t}(x, w_x)) \right) \cdot \frac{d}{dt} \Phi_{x,t}(x, w_x) \, d\nu_{x,0}(x, w_x) \\
&= \int_{\mathcal{X} \times \mathbb{R}^+} \nabla_x \psi_x(\Phi_{x,t}(x, w_x)) \cdot (-\gamma \nabla_x V_x(h_y \nu_{y,t}, X_{x,t}^\Phi)) \\
&+ \frac{d\psi_x}{dw_x}(\Phi_{x,t}(x, w_x)) \alpha(-V_x(h_y \nu_{y,t}, X_{x,t}^\Phi) + \mathcal{L}(h_x \nu_{x,t}, h_y \mu_{y,t})) \, d\nu_{x,0}(x, w_x)
\end{aligned}$$

And we can identify the right hand side as the weak form of (A.1), shown in (A.3). The argument

for ν_y is analogous. □

We state a stronger version of the law of large numbers in the first statement of Theorem 2.4(ii).

Theorem A.23. *There exists a solution of the coupled SDEs* (A.44)*. Pathwise uniqueness and uniqueness in law hold. Let* $\nu_x \in \mathcal{P}(C([0,T], \mathcal{X} \times \mathbb{R}^+))$, $\nu_y \in \mathcal{P}(C([0,T], \mathcal{Y} \times \mathbb{R}^+))$ *be the unique laws of the solutions for* $n = 1$ *(all pairs have the same solutions). Then,*

$$\mathbb{E}[\mathcal{W}_2^2(\nu_x^n, \nu_x) + \mathcal{W}_2^2(\nu_y^n, \nu_y)] \xrightarrow{n\to\infty} 0$$

Theorem A.23 is the law of large numbers for the WFR dynamics, and its proof follows the same argument of Theorem A.20. The reason Theorem A.23 implies Theorem 2.4(ii) is analogous to the reason for which Theorem A.20 implies Theorem 2.4(i), with the additional step that $\mathcal{W}_2^2(\mu_{x,t}^n, \mu_{x,t}) = \mathcal{W}_2^2(h_x \nu_{x,t}^n, h_x \nu_{x,t}) \le e^{4MT} \mathcal{W}_2^2(\nu_{x,t}^n, \nu_{x,t})$, and this inequality is shown in (A.46).

A.6.2 Existence and uniqueness

We choose to do an argument close to Sznitman [1991] (see Lacker [2018]), which yields convergence of the expectation of the square of the 2-Wasserstein distances between the empirical and the mean field measures.

First, to prove existence and uniqueness of the solution $(\mu_{x,t}, \mu_{y,t})$ in the time interval $[0,T]$ for

arbitrary T, we can use the same argument as in the section A.5. Now, instead of (A.38) we have

$$\begin{aligned}
\tilde{X}_t &= \xi - \gamma \int_0^t \int_{\mathcal{Y}} \nabla_x \ell(\tilde{X}_s, y)\, d\mu_{y,s}\, ds, \\
\tilde{w}_{x,t} &= 1 + \alpha \int_0^t \left(- \int \ell(\tilde{X}_t, y) d\mu_{y,t} + \mathcal{L}(\mu_{x,t}, \mu_{y,t}) \right) \tilde{w}_{x,s}\, ds, \\
\tilde{Y}_t &= \bar{\xi} + \gamma \int_0^t \int_{\mathcal{X}} \nabla_y \ell(x, \tilde{Y}_s)\, d\mu_{x,s}\, ds, \\
\tilde{w}_{y,t} &= 1 + \alpha \int_0^t \left(\int \ell(x, \tilde{Y}_t) d\mu_{x,t} - \mathcal{L}(\mu_{x,t}, \mu_{y,t}) \right) \tilde{w}_{y,s}\, ds, \\
\mu_{x,t} &= h_x \mathrm{Law}(\tilde{X}_t, \tilde{w}_{x,t}), \quad \mu_{y,t} = h_y \mathrm{Law}(\tilde{Y}_t, \tilde{w}_{y,t}),
\end{aligned}$$

where ξ and $\bar{\xi}$ are arbitrary random variables with laws $\mu_{x,0}, \mu_{y,0}$ respectively. For $x, x' \in \mathcal{X}$, $r, r' \in \mathbb{R}^+$, $\mu_x, \mu_x' \in \mathcal{P}(\mathcal{X})$, $\mu_y, \mu_y' \in \mathcal{P}(\mathcal{Y})$, notice that using an argument similar to (A.39) the following bound holds

$$\begin{aligned}
&\left| \left(- \int \ell(x, y) d\mu_y + \mathcal{L}(\mu_x, \mu_y) \right) w - \left(- \int \ell(x', y) d\mu_y' + \mathcal{L}(\mu_x', \mu_y') \right) w' \right| \\
&\leq 2M|w - w'| + |w'|\tilde{L}(|x - x'| + 3\mathcal{W}_1(\nu, \mu)) \leq 2M|w - w'| + |w'|\tilde{L}(|x - x'| + 3\mathcal{W}_2(\mu_y, \mu_y'))
\end{aligned}$$

$$\begin{aligned}
\implies &\left| \left(- \int \ell(x, y) d\mu_y + \mathcal{L}(\mu_x, \mu_y) \right) r - \left(- \int \ell(x', y) d\mu_y' + \mathcal{L}(\mu_x', \mu_y') \right) r' \right|^2 \\
&\leq 12M^2|w - w'|^2 + 3|w'|^2\tilde{L}^2(|x - x'|^2 + 9\mathcal{W}_2^2(\mu_y, \mu_y'))
\end{aligned}$$

Recall that M is a bound on the absolute value of ℓ and $\tilde{L}$ is the Lipschitz constant of the loss ℓ. A simple application of Gronwall's inequality shows $|\tilde{w}_{x,t}|$ is bounded by e^{2MT} for all $t \in [0, T]$.

Hence, we can write

$$\mathbb{E}[\|X^{\mu_y} - X^{\mu'_y}\|_t^2 + \|w_x^{\mu_y} - w_x^{\mu'_y}\|_t^2] \leq \gamma^2 t\mathbb{E}\left[\int_0^t \left|\nabla_x \int \ell(X_s^{\mu_y}, y)d\mu_{y,s} - \nabla_x \int \ell(X_s^{\mu'_y}, y)d\mu'_{y,s}\right|^2 ds\right]$$
$$+\alpha^2 t\mathbb{E}\left[\int_0^t \left|\left(-\int \ell(X_s^{\mu_y}, y)d\mu_y + \mathcal{L}(\mu_x, \mu_y)\right) w_x^{\mu_y} - \left(-\int \ell(X_s^{\mu'_y}, y)d\mu'_y + \mathcal{L}(\mu'_x, \mu'_y)\right) w_x^{\mu'_y}\right|^2 ds\right]$$
$$\leq Kt\mathbb{E}\left[\int_0^t \|X^{\mu_y} - X^{\mu'_y}\|_s^2 + \|w^{\mu_y} - w^{\mu'_y}\|_s^2 \, ds\right] + K't\mathbb{E}\left[\int_0^t \mathcal{W}_2^2(\mu_{y,s}, \mu'_{y,s}) \, ds\right],$$

where $K = \max\{12\alpha^2M^2, 2L^2\gamma^2 + 3\tilde{L}^2e^{4MT}\alpha^2\}, K' = 2L^2\gamma^2 + 27\tilde{L}^2e^{4MT}\alpha^2$. Notice that we have used (A.40) as well. This equation is analogous to equation (A.41), and upon application of Fubini's theorem and Gronwall's inequality it yields

$$\mathbb{E}[\|X^{\mu_y} - X^{\mu'_y}\|_t^2 + \|w_x^{\mu_y} - w_x^{\mu'_y}\|_t^2] \leq TK' \exp(TK)\mathbb{E}\left[\int_0^t \mathcal{W}_2^2(\mu_{y,s}, \mu'_{y,s}) \, ds\right] \tag{A.45}$$

Now we will prove that

$$\mathcal{W}_2^2(h_x\nu_x, h_x\nu'_x) \leq e^{4MT}\mathcal{W}_2^2(\nu_x, \nu'_x), \tag{A.46}$$

where $\nu_x, \nu'_x \in \mathcal{P}(\mathcal{X} \times [0, e^{2MT}])$. Define the homogeneous projection operator $\tilde{h} : \mathcal{P}((\mathcal{X} \times \mathbb{R}^+)^2) \to \mathcal{P}(\mathcal{X}^2)$ as $\forall f \in C(\mathcal{X}^2)$,

$$\int_{\mathcal{X}^2} f(x,y) \, d(\tilde{h}\pi)(x,y) = \int_{(\mathcal{X}\times[0,e^{2MT}])^2} w_x w_y f(x,y) \, d\pi(x, w_x, y, w_y), \ \forall \pi \in \mathcal{P}((\mathcal{X} \times \mathbb{R}^+)^2).$$

Let π be a coupling between $h_x\nu_x, h_x\nu'_x$. Then $\tilde{h}\pi$ is a coupling between $h_x\nu_x, h_x\nu'_x$ and

$$\begin{aligned}\int_{\mathcal{X}^2} \|x-y\|^2 \, d(\tilde{h}\pi)(x,y) &= \int_{(\mathcal{X}\times[0,e^{2MT}])^2} w_x w_y \|x-y\|^2 \, d\pi(x,w_x,y,w_y) \\ &\leq e^{4MT} \int_{(\mathcal{X}\times[0,e^{2MT}])^2} \|x-y\|^2 \, d\pi(x,w_x,y,w_y) \\ &\leq e^{4MT} \int_{(\mathcal{X}\times[0,e^{2MT}])^2} \|x-y\|^2 + |w_x-w_y|^2 \, d\pi'(x,w_x,y,w_y)\end{aligned}$$

Taking the infimum with respect to π on both sides we obtain the desired inequality.

Let $\nu_{x,t} = \text{Law}(X_t^{\mu_y}, w_{x,t}^{\mu_y}), \nu'_{x,t} = \text{Law}(X_t^{\mu'_y}, w_{x,t}^{\mu'_y})$ and recall that $\mu_{x,t} = h_x\nu_{x,t}, \mu'_{x,t} = h_x\nu'_{x,t}$. Given $\nu_y \in \mathcal{P}(C([0,T], \mathcal{Y}\times\mathbb{R}^+))$, define $\Phi_x(\nu_y) = \text{Law}(X^{\nu_y}, w_x^{\nu_y}) \in \mathcal{P}(C([0,T],\mathcal{X}))$ where we abuse the notation and use $(X^{\nu_y}, w_x^{\nu_y})$ to refer to $(X^{\mu_y}, w_x^{\mu_y})$. Notice also that

$$\begin{aligned}\mathcal{W}_{2,t}^2(\Phi_x(\nu_y), \Phi_x(\nu'_y)) &\leq \mathbb{E}\left[\sup_{s\in[0,t]} \|X_s^{\mu_y} - X_s^{\mu'_y}\|^2 + \|w_{x,s}^{\mu_y} - w_{x,s}^{\mu'_y}\|^2\right] \\ &\leq \mathbb{E}[\|X^{\mu_y} - X^{\mu'_y}\|_t^2 + \|w_x^{\mu_y} - w_x^{\mu'_y}\|_t^2] \qquad \text{(A.47)}\end{aligned}$$

We use (A.46) and (A.47) on (A.45) to conclude

$$\mathcal{W}_{2,t}^2(\Phi_x(\nu_y), \Phi_x(\nu'_y)) \leq TK' \exp(TK)\mathbb{E}\left[\int_0^t \mathcal{W}_{2,s}^2(\nu_y, \nu'_y)\, ds\right]$$

The rest of the argument is sketched in section A.5.

A.6.3 Propagation of chaos

Following the reasoning in the existence and uniqueness proof, we can write

$$\begin{aligned}&\mathbb{E}[\|X^i - \tilde{X}^i\|_t^2 + \|w_x^i - \tilde{w}_x^i\|_t^2] \\ &\leq Kt\mathbb{E}\left[\int_0^t \|X^i - \tilde{X}^i\|_s^2 + \|w_x^i - \tilde{w}_x^i\|_s^2 \, ds\right] + K't\mathbb{E}\left[\int_0^t \mathcal{W}_2^2(\mu_{y,s}^n, \mu_{y,s})\, ds\right],\end{aligned}$$

Hence, we obtain

$$\mathbb{E}[\|X^i - \tilde{X}^i\|_t^2 + \|w_x^i - \tilde{w}_x^i\|_t^2] \leq TK' \exp(TK)\mathbb{E}\left[\int_0^t \mathcal{W}_2^2(\mu_{y,s}^n, \mu_{y,s})\, ds\right]$$

Let $\tilde{\nu}_{x,t}^n = \frac{1}{n}\sum_{i=1}^n \delta_{(\tilde{X}_t^i, \tilde{w}_t^i)} \in \mathbb{P}(\mathcal{X}\times\mathbb{R}^+)$ be the marginal at time t of the empirical measure of (A.43). As in section A.5,

$$\mathcal{W}_{2,t}^2(\nu_x^n, \tilde{\nu}_x^n) \leq \frac{1}{n}\sum_{i=1}^n \sup_{s\in[0,t]} \|X_s^i - \tilde{X}_s^i\|^2 + |w_{x,s}^i - \tilde{w}_{x,s}^i|^2 \leq \frac{1}{n}\sum_{i=1}^n \|X^i - \tilde{X}^i\|_t^2 + \|w_x^i - \tilde{w}_x^i\|_t^2$$

which yields

$$\begin{aligned}\mathbb{E}[\mathcal{W}_{2,t}^2(\nu_x^n, \tilde{\nu}_x^n)] &\leq TK' \exp(TK)\mathbb{E}\left[\int_0^t \mathcal{W}_2^2(\mu_{y,s}^n, \mu_{y,s})\, ds\right] \\ &\leq TK' \exp((K+4M)T)\mathbb{E}\left[\int_0^t \mathcal{W}_{2,s}^2(\nu_y^n, \nu_y)\, ds\right]\end{aligned}$$

The second inequality above follows from inequality (A.46) $\mathcal{W}_2^2(\nu_{y,s}^n, \nu_{y,s}) \leq \mathcal{W}_{2,s}^2(\nu_y^n, \nu_y)$. Now we use the triangle inequality as in section A.5:

$$\begin{aligned}\mathbb{E}[\mathcal{W}_{2,t}^2(\nu_x^n, \nu_x)] &\leq 2\mathbb{E}[\mathcal{W}_{2,t}^2(\nu_x^n, \tilde{\nu}_x^n)] + 2\mathbb{E}[\mathcal{W}_{2,t}^2(\tilde{\nu}_x^n, \nu_x)] \\ &\leq 2TK' \exp((K+4M)T)\mathbb{E}\left[\int_0^t \mathcal{W}_{2,s}^2(\nu_y^n, \nu_y)\, ds\right] + 2\mathbb{E}[\mathcal{W}_{2,t}^2(\tilde{\nu}_x^n, \nu_x)]\end{aligned}$$

If we denote $C := 2TK' \exp((K+4M)T)$ and we make the same developments for the other player, we obtain

$$\begin{aligned}\mathbb{E}[\mathcal{W}_{2,t}^2(\nu_x^n, \nu_x) + \mathcal{W}_{2,t}^2(\nu_y^n, \nu_y)] \leq C\mathbb{E}&\left[\int_0^t \mathcal{W}_{2,s}^2(\nu_y^n, \nu_y) + \mathcal{W}_{2,s}^2(\nu_x^n, \nu_x)\, ds\right] \\ &+ 2\mathbb{E}[\mathcal{W}_{2,t}^2(\tilde{\nu}_x^n, \nu_x) + \mathcal{W}_{2,t}^2(\tilde{\nu}_y^n, \nu_y)]\end{aligned}$$

From this point on, the proof works as in section A.5.

A.6.4 Convergence of the Nikaido-Isoda error

Corollary A.24. *For $t \in [0,T]$, let $\bar{\mu}^n_{x,t} = \frac{1}{t}\int_0^t h_x v^n_{x,r}\, dr, \bar{\mu}_{x,t} = \frac{1}{t}\int_0^t h_x v_{x,r}\, dr$ and define $\bar{\mu}^n_{y,t}, \bar{\mu}_{y,t}$ analogously. Then,*

$$\mathbb{E}[|NI(\bar{\mu}^n_{x,t}, \bar{\mu}^n_{y,t}) - NI(\bar{\mu}_{x,t}, \bar{\mu}_{y,t})|] \xrightarrow{n\to\infty} 0$$

Proof. Notice that since the integral over time and the homogeneous projection commute, we have $\bar{\mu}^n_{x,t} = h_x(\frac{1}{t}\int_0^t v^n_{x,r}\, dr), \bar{\mu}_{x,t} = h_x(\frac{1}{t}\int_0^t v_{x,r}\, dr)$. Since $\frac{1}{t}\int_0^t v^n_{x,r}\, dr$ and $\frac{1}{t}\int_0^t v_{x,r}\, dr$ belong to $\mathcal{P}(\mathcal{X}\times[0,e^{2MT}])$, (A.46) implies

$$\mathcal{W}_2^2\left(h_x\left(\frac{1}{t}\int_0^t v^n_{x,r}\, dr\right), h_x\left(\frac{1}{t}\int_0^t v_{x,r}\, dr\right)\right) \le e^{4MT}\mathcal{W}_2^2\left(\frac{1}{t}\int_0^t v^n_{x,r}\, dr, \frac{1}{t}\int_0^t v_{x,r}\, dr\right)$$

Notice that $\mathcal{W}_2^2(\frac{1}{t}\int_0^t v^n_{x,r}\, dr, \frac{1}{t}\int_0^t v_{x,r}\, dr) \le \frac{1}{t}\int_0^t \mathcal{W}_2^2(v^n_{x,r}, v_{x,r})\, dr$. Indeed,

$$\begin{aligned}
\mathcal{W}_2^2\left(\frac{1}{t}\int_0^t v^n_{x,r}\, dr, \frac{1}{t}\int_0^t v_{x,r}\, dr\right) &= \max_{\psi\in\Psi_c(X)} \frac{1}{t}\int_0^t \int \psi\, dv^n_{x,r}\, dr + \frac{1}{t}\int_0^t \int \psi^c\, dv^n_{x,r}\, dr \\
&\le \frac{1}{t}\int_0^t \left(\max_{\psi\in\Psi_c(X)} \int \psi\, dv^n_{x,r} + \int \psi^c\, dv^n_{x,r}\right) dr \\
&= \frac{1}{t}\int_0^t \mathcal{W}_2^2(v^n_{x,r}, v_{x,r})\, dr
\end{aligned}$$

Hence, using the inequality $\mathcal{W}_2^2(v^n_{x,r}, v_{x,r}) \le \mathcal{W}_2^2(v^n_x, v_x)$:

$$\begin{aligned}
\mathbb{E}\left[\mathcal{W}_2^2\left(h_x\left(\frac{1}{t}\int_0^t v^n_{x,r}\, dr\right), h_x\left(\frac{1}{t}\int_0^t v_{x,r}\, dr\right)\right)\right] &\le e^{4MT}\mathbb{E}\left[\frac{1}{t}\int_0^t \mathcal{W}_2^2(v^n_{x,r}, v_{x,r})\, dr\right] \\
&\le e^{4MT}\mathbb{E}[\mathcal{W}_2^2(v^n_x, v_x)]
\end{aligned}$$

Since the right hand side goes to zero as $n \to \infty$ by Theorem A.23, we conclude by applying

Theorem A.3. □

A.6.5 Hint of the infinitesimal generator approach

Let $\psi_x : \mathcal{X} \to \mathbb{R}, \psi_y : \mathcal{Y} \to \mathbb{R}$ be arbitrary continuously differentiable functions, i.e. $\psi_x \in C^1(\mathcal{X}, \mathbb{R}), \psi_y \in C^1(\mathcal{Y}, \mathbb{R})$. Let us define the operators $\mathcal{L}^{(n)}_{x,t} : C^1(\mathcal{X}, \mathbb{R}) \to C^0(\mathcal{X}, \mathbb{R}), \mathcal{L}^{(n)}_{y,t} : C^1(\mathcal{Y}, \mathbb{R}) \to C^0(\mathcal{Y}, \mathbb{R})$ as

$$\begin{aligned}
\mathcal{L}^{(n)}_{x,t}\psi_x(x) &= -\gamma\nabla_x \int \ell(x,y)d\mu^n_{y,t} \cdot \nabla_x\psi_x(x) + \alpha\left(-\int \ell(x,y)d\mu^n_{y,t} + \mathcal{L}(\mu^n_{x,t}, \mu^n_{y,t})\right) \\
\mathcal{L}^{(n)}_{y,t}\psi_y(y) &= \gamma\nabla_y \int \ell(x,y)d\mu^n_{x,t} \cdot \nabla_y\psi_y(x) + \alpha\left(\int \ell(x,y)d\mu^n_{x,t} - \mathcal{L}(\mu^n_{x,t}, \mu^n_{y,t})\right)
\end{aligned} \tag{A.48}$$

Notice that from (A.43) and (A.48), we have

$$\begin{aligned}
\frac{d}{dt}\int_{\mathcal{X}} \psi_x(x)\, d\mu^n_{x,t}(x) &= \frac{d}{dt}\int_{\mathcal{X}\times\mathbb{R}^+} w_x\psi_x(x)\, dv^n_{x,t}(x, w_x) = \frac{d}{dt}\sum_{i=1}^n w^i_{x,t}\psi_x(X^i_t) \\
&= \sum_{i=1}^n \frac{dw^i_{x,t}}{dt}\psi_x(X^i_t) + \sum_{i=1}^n w^i_{x,t}\nabla_x\psi_x(X^i_t)\cdot\frac{dX^i_t}{dt} \\
&= \int_{\mathcal{X}\times\mathbb{R}^+} w_x\mathcal{L}^{(n)}_{x,t}\psi_x(x)\, dv^n_{x,t}(x, w_x) = \int_{\mathcal{X}} \mathcal{L}^{(n)}_{x,t}\psi_x(x)\, d\mu^n_{x,t}(x)
\end{aligned} \tag{A.49}$$

The analogous equation holds for $\mu^n_{x,t}$:

$$\frac{d}{dt}\int_{\mathcal{Y}} \psi_y(y)\, d\mu^n_{y,t}(y) = \int_{\mathcal{Y}} \mathcal{L}^{(n)}_{y,t}\psi_y(y)\, d\mu^n_{y,t}(y) \tag{A.50}$$

Formally taking the limit $n \to \infty$ on (A.49) and (A.50) yields

$$\begin{aligned}
\frac{d}{dt}\int_{\mathcal{X}} \psi_x(x)\, d\mu_{x,t}(x) &= \int_{\mathcal{X}} \mathcal{L}_{x,t}\psi_x(x)\, d\mu_{x,t}(x) \\
\frac{d}{dt}\int_{\mathcal{Y}} \psi_y(y)\, d\mu_{y,t}(y) &= \int_{\mathcal{Y}} \mathcal{L}_{y,t}\psi_y(y)\, d\mu_{y,t}(y),
\end{aligned}$$

where

$$\mathcal{L}_{x,t}\psi_x(x) = -\gamma\nabla_x \int \ell(x,y)d\mu_{y,t} \cdot \nabla_x\psi_x(x) + \alpha\left(-\int \ell(x,y)d\mu_{y,t} + \mathcal{L}(\mu_{x,t},\mu_{y,t})\right)$$
$$\mathcal{L}_{y,t}\psi_y(y) = \gamma\nabla_y \int \ell(x,y)d\mu_{x,t} \cdot \nabla_y\psi_y(x) + \alpha\left(\int \ell(x,y)d\mu_{x,t} - \mathcal{L}(\mu_{x,t},\mu_{y,t})\right)$$

and $\mu_{x,0}, \mu_{y,0}$ are set as in (A.43).

To make the limit $n \to \infty$ rigorous, an argument analogous to Theorem 2.6 of Chizat and Bach [2018] would result in almost sure convergence of the 2-Wasserstein distances between the empirical and the mean field measures. In our case almost sure convergence of the squared distance implies convergence of the expectation of the squared distance through dominated convergence, and hence the almost sure convergence result is stronger. Nonetheless, such an argument would require proving uniqueness of the mean field measure PDE through some notion of geodesic convexity, which is not clear in our case.

A.7 Auxiliary material

A.7.1 ϵ-Nash equilibria and the Nikaido-Isoda error

Recall that an ϵ-NE (μ_x, μ_y) satisfies $\forall \mu_x^* \in \mathcal{P}(\mathcal{X}),\ \mathcal{L}(\mu_x, \mu_y) \leq \mathcal{L}(\mu_x^*, \mu_y) + \epsilon$ and $\forall \mu_y^* \in \mathcal{P}(\mathcal{Y})$, $\mathcal{L}(\mu_x, \mu_y) \geq \mathcal{L}(\mu_x, \mu_y^*) - \epsilon$. That is, each player can improve its value by at most ϵ by deviating from the equilibrium strategy, supposing that the other player is kept fixed.

Recall the Nikaido-Isoda error defined in (2.2). This equation can be rewritten as:

$$\mathrm{NI}(\mu_x, \mu_y) = \sup_{\mu_y^* \in \mathcal{P}(\mathcal{Y})} \mathcal{L}(\mu_x, \mu_y^*) - \mathcal{L}(\mu_x, \mu_y) + \mathcal{L}(\mu_x, \mu_y) - \inf_{\mu_x^* \in \mathcal{P}(\mathcal{X})} \mathcal{L}(\mu_x^*, \mu_y)\ .$$

The terms $\sup_{\mu_y^* \in \mathcal{P}(\mathcal{Y})} \mathcal{L}(\mu_x, \mu_y^*) - \mathcal{L}(\mu_x, \mu_y) > 0$ measure how much player y can improve its value by deviating from μ_y while μ_x stays fixed. Analogously, the terms $\mathcal{L}(\mu_x, \mu_y) - \inf_{\mu_x^* \in \mathcal{P}(\mathcal{X})} \mathcal{L}(\mu_x^*, \mu_y) > 0$ measure how much player x can improve its value by deviating from μ_x while μ_y stays fixed.

Notice that

$$\forall \mu_x^* \in \mathcal{P}(\mathcal{X}),\ \mathcal{L}(\mu_x, \mu_y) \leq \mathcal{L}(\mu_x^*, \mu_y) + \epsilon \Leftrightarrow \mathcal{L}(\mu_x, \mu_y) - \inf_{\mu_x^* \in \mathcal{P}(\mathcal{X})} \mathcal{L}(\mu_x^*, \mu_y) \leq \epsilon$$

$$\forall \mu_y^* \in \mathcal{P}(\mathcal{Y}),\ \mathcal{L}(\mu_x, \mu_y) \geq \mathcal{L}(\mu_x, \mu_y^*) - \epsilon \Leftrightarrow \sup_{\mu_y^* \in \mathcal{P}(\mathcal{Y})} \mathcal{L}(\mu_x, \mu_y^*) - \mathcal{L}(\mu_x, \mu_y) \leq \epsilon$$

Thus, an ϵ-Nash equilibrium (μ_x, μ_y) fulfills $\mathrm{NI}(\mu_x, \mu_y) \leq 2\epsilon$, and any pair (μ_x, μ_y) such that $\mathrm{NI}(\mu_x, \mu_y) \leq \epsilon$ is an ϵ-Nash equilibrium.

A.7.2 Example: failure of the Interacting Wasserstein Gradient Flow

Let us consider the polynomial $f(x) = 5x^4 + 10x^2 - 2x$, which is an asymmetric double well as shown in Figure A.1.

Let us define the loss $\ell : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ as $\ell(x, y) = f(x) - f(y)$. That is, the two players are

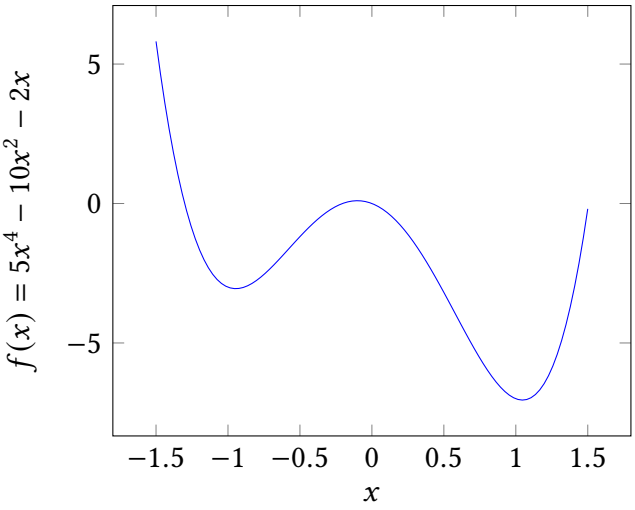


Figure A.1: Plot of the function $f(x) = 5x^4 + 10x^2 - 2x$.

non-interacting and hence we obtain $V_x(x, \mu_y) = f(x) + K$, $V_y(y, \mu_x) = -f(y) + K'$. This means that the IWGF in equation (2.6) becomes two independent Wasserstein Gradient Flows

$$\partial_t \mu_x = \nabla \cdot (\mu_x f'(x)), \quad \mu_x(0) = \mu_{x,0},$$
$$\partial_t \mu_y = -\nabla \cdot (\mu_y f'(y)), \quad \mu_y(0) = \mu_{y,0}.$$

The particle flows in (2.3) become

$$\frac{dx_i}{dt} = -f'(x_i), \quad \frac{dy_i}{dt} = f'(y_i).$$

That is, the particles of player x follow the gradient flow of f and the particles of player y follow the gradient flow of $-f$. It is clear from Figure A.1 that if the initializations $x_{0,i}, y_{0,i}$ are on the left of the barrier, they will not end up in the global minimum f (resp., the global maximum of $-f$). And in this case, the pair of measures supported on the global minimum of f is the only (pure) Nash equilibrium.

The game given by ℓ does not fall exactly in the framework that we describe in this work because ℓ is not defined on compact spaces. However, it is easy to construct very similar continuously

differentiable functions on compact spaces that display the same behavior.

A.7.3 Link between Interacting Wasserstein Gradient Flow and interacting particle gradient flows

Recall (2.3):

$$\frac{dx_i}{dt} = -\frac{1}{n}\sum_{j=1}^{n} \nabla_x \ell(x_i, y_j), \quad \frac{dy_i}{dt} = \frac{1}{n}\sum_{j=1}^{n} \nabla_x \ell(x_j, y_i).$$

Let $\Phi_t = (\Phi_{x,t}, \Phi_{y,t}) : \mathcal{X}^n \times \mathcal{Y}^n \to \mathcal{X}^n \times \mathcal{Y}^n$ be the flow mapping initial conditions $\mathbf{X}_0 = (x_{i,0})_{i\in[1:n]}$, $\mathbf{Y}_0 = (y_{i,0})_{i\in[1:n]}$ to the solution of (2.3). Let $\mu^n_{x,t} = \frac{1}{n}\sum_{i=1}^n \delta_{\Phi^{(i)}_{x,t}(\mathbf{X}_0,\mathbf{Y}_0)}$ and $\mu^n_{y,t} = \frac{1}{n}\sum_{i=1}^n \delta_{\Phi^{(i)}_{y,t}(\mathbf{X}_0,\mathbf{Y}_0)}$. For all $\psi_x \in C(\mathcal{X})$,

$$\begin{aligned}
\frac{d}{dt}\int_{\mathcal{X}} \psi_x(x)\, d\mu^n_{x,t}(x) &= \frac{1}{n}\sum_{i=1}^{n} \frac{d}{dt}\psi_x(\Phi^{(i)}_{x,t}(\mathbf{X}_0, \mathbf{Y}_0)) \\
&= \frac{1}{n}\sum_{i=1}^{n} \nabla_x \psi_x(\Phi^{(i)}_{x,t}(\mathbf{X}_0, \mathbf{Y}_0)) \cdot \left(-\frac{1}{n}\sum_{j=1}^{n} \nabla_x \ell(\Phi^{(i)}_{x,t}(\mathbf{X}_0, \mathbf{Y}_0), \Phi^{(j)}_{y,t}(\mathbf{X}_0, \mathbf{Y}_0))\right) \\
&= -\frac{1}{n}\sum_{i=1}^{n} \nabla_x \psi_x(\Phi^{(i)}_{x,t}(\mathbf{X}_0, \mathbf{Y}_0)) \cdot \nabla_x V_x(\mu^n_{y,t}, \Phi^{(i)}_{x,t}(\mathbf{X}_0, \mathbf{Y}_0)) \\
&= -\int_{\mathcal{X}} \nabla_x \psi_x(x) \cdot \nabla_x V_x(\mu^n_{y,t}, x)\, d\mu^n_{x,t}(x),
\end{aligned}$$

which is the first line of (2.6). The second line follows analogously.

A.7.4 Minimax problems and Stackelberg equilibria

Several machine learning problems, including GANs, are framed as a minimax problem

$$\min_{x\in\mathcal{X}} \max_{y\in\mathcal{Y}} \ell(x, y).$$

A minimax point (also known as a Stackelberg equilibrium or sequential equilibrium) is a pair $(\tilde{x}, \tilde{y})$ at which the minimum and maximum of the problem are attained, i.e.

$$\begin{cases} \min_{x \in \mathcal{X}} \max_{y \in \mathcal{Y}} \ell(x, y) = \max_{y \in \mathcal{Y}} \ell(\tilde{x}, y) \\ \max_{y \in \mathcal{Y}} \ell(\tilde{x}, y) = \ell(\tilde{x}, \tilde{y}) \end{cases}.$$

We consider the lifted version of the minimax problem (A.7.4) in the space of probability measures.

$$\min_{\mu_x \in \mathcal{P}(\mathcal{X})} \max_{\mu_y \in \mathcal{P}(\mathcal{Y})} \mathcal{L}(\mu_x, \mu_y). \tag{A.51}$$

By the generalized Von Neumann's minimax theorem, a Nash equilibrium of the game given by $\mathcal{L}$ is a solution of the lifted minimax problem (A.51) (see Theorem A.25 in the case $\epsilon = 0$).

The converse is not true: minimax points (solutions of (A.51)) are not necessarily mixed Nash equilibria even in the case where the loss function is convex-concave. An example is $\mathcal{L} : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ given by $\mathcal{L}(\mu_x, \mu_y) = \iint (x^2 + 2xy)\, d\mu_x\, d\mu_y$. Let $\mathcal{M}$ be the set of measures $\mu \in \mathcal{P}(\mathbb{R})$ such that $\int x\, d\mu = 0$. Notice that any pair (δ_0, μ_y) with $\mu_y \in \mathcal{P}(\mathbb{R})$ is a minimax point. That is because

$$\max_{\mu_y \in \mathcal{P}(\mathbb{R})} \mathcal{L}(\mu_x, \mu_y) = \begin{cases} +\infty & \text{if } \mu_x \notin \mathcal{M} \\ \text{positive} & \text{if } \mu_x \in \mathcal{M} \setminus \{\delta_0\} \\ 0 & \text{if } \mu_x = \delta_0, \end{cases}$$

and hence $\delta_0 = \arg\min_{\mu_x \in \mathcal{P}(\mathbb{R})} \max_{\mu_y \in \mathcal{P}(\mathbb{R})} \mathcal{L}(\mu_x, \mu_y)$. But in the case that $\mu_x = \delta_0$, we obtain that $\arg\max_{\mu_y \in \mathcal{P}(\mathbb{R})} \mathcal{L}(\mu_x, \mu_y) = \mathcal{P}(\mathbb{R})$, because for all measures $\mu_y \in \mathcal{P}(\mathbb{R})$, $\mathcal{L}(\delta_0, \mu_y) = 0$. However, for $\mu_y \notin \mathcal{M}$, $\mathcal{L}(\mu_x, \mu_y)$ as a function of μ_x does not have a minimum at δ_0, but at $\delta_{-\int y\, d\mu_y}$. Hence, the only mixed Nash equilibria are of the form (δ_0, μ_y), with $\mu_y \in \mathcal{M}$.

The intuition behind the counterexample is that minimax points only require the minimizing player to be non-exploitable, but the maximizing player is only subject to a weaker condition.

We define a ε-minimax point (or ϵ-Stackelberg equilibrium) of an objective $\mathcal{L}(\mu_x, \mu_y)$ as a couple $(\tilde{\mu}_x, \tilde{\mu}_y)$ such that

$$\begin{cases} \min_{\mu_x \in \mathcal{P}(\mathcal{X})} \max_{\mu_y \in \mathcal{P}(\mathcal{Y})} \mathcal{L}(\mu_x, \mu_y) \geq \max_{\mu_y \in \mathcal{P}(\mathcal{Y})} \mathcal{L}(\tilde{\mu}_x, \mu_y) - \varepsilon \\ \max_{\mu_y \in \mathcal{P}(\mathcal{Y})} \mathcal{L}(\tilde{\mu}_x, \mu_y) \leq \mathcal{L}(\tilde{\mu}_x, \tilde{\mu}_y) + \varepsilon \end{cases}.$$

Lemma A.25. *An ε-Nash equilibrium is a 2ε-minimax point, and it holds that*

$$\min_{\mu_x \in \mathcal{P}(\mathcal{X})} \max_{\mu_y \in \mathcal{P}(\mathcal{Y})} \mathcal{L}(\mu_x, \mu_y) - \epsilon \leq \mathcal{L}(\hat{\mu}_x, \hat{\mu}_y) \leq \max_{\mu_y \in \mathcal{P}(\mathcal{Y})} \min_{\mu_x \in \mathcal{P}(\mathcal{X})} \mathcal{L}(\mu_x, \hat{\mu}_y) + \varepsilon$$

Proof. Let $(\hat{\mu}_x, \hat{\mu}_y)$ be an ε-Nash equilibrium. Notice that $\max_{\mu_y \in \mathcal{P}(\mathcal{Y})} \min_{\mu_x \in \mathcal{P}(\mathcal{X})} \mathcal{L}(\tilde{\mu}_x, \mu_y) \leq \min_{\mu_x \in \mathcal{P}(\mathcal{X})} \max_{\mu_y \in \mathcal{P}(\mathcal{Y})} \mathcal{L}(\tilde{\mu}_x, \mu_y)$. Also,

$$\begin{aligned} \min_{\mu_x \in \mathcal{P}(\mathcal{X})} \max_{\mu_y \in \mathcal{P}(\mathcal{Y})} \mathcal{L}(\mu_x, \mu_y) &\leq \max_{\mu_y \in \mathcal{P}(\mathcal{Y})} \mathcal{L}(\hat{\mu}_x, \mu_y) \leq \mathcal{L}(\hat{\mu}_x, \hat{\mu}_y) + \varepsilon \leq \min_{\mu_x \in \mathcal{P}(\mathcal{X})} \mathcal{L}(\mu_x, \hat{\mu}_y) + 2\varepsilon \\ &\leq \max_{\mu_y \in \mathcal{P}(\mathcal{Y})} \min_{\mu_x \in \mathcal{P}(\mathcal{X})} \mathcal{L}(\mu_x, \hat{\mu}_y) + 2\varepsilon \end{aligned} \tag{A.52}$$

and this yields the chain of inequalities in the statement of the lemma. Also, the condition $\max_{\mu_y \in \mathcal{P}(\mathcal{Y})} \mathcal{L}(\tilde{\mu}_x, \mu_y) \leq \mathcal{L}(\tilde{\mu}_x, \tilde{\mu}_y) + \varepsilon$ of the definition of ε-minimax point follows directly from the definition of an ϵ-Nash equilibrium. Using part of (A.52), we get

$$\max_{\mu_y \in \mathcal{P}(\mathcal{Y})} \mathcal{L}(\hat{\mu}_x, \mu_y) - 2\epsilon \leq \max_{\mu_y \in \mathcal{P}(\mathcal{Y})} \min_{\mu_x \in \mathcal{P}(\mathcal{X})} \mathcal{L}(\mu_x, \hat{\mu}_y) \leq \min_{\mu_x \in \mathcal{P}(\mathcal{X})} \max_{\mu_y \in \mathcal{P}(\mathcal{Y})} \mathcal{L}(\tilde{\mu}_x, \mu_y),$$

which is the first condition of a 2ϵ-minimax. □

Theorem A.25 provides the link between approximate Nash equilibria and approximate Stackelberg equilibria, and it allows to translate our convergence results into minimax problems such as GANs.

A.7.5 Itô SDEs on Riemannian manifolds: a parametric approach

We provide a brief summary on how to deal with SDEs on Riemannian manifolds and their corresponding Fokker-Planck equations (see Chapter 8 of Chirikjian [2009]). While ODEs have a straightforward translation into manifolds, the same is not true for SDEs. Recall that the definitions of the gradient and divergence for Riemannian manifolds are

$$\nabla \cdot X = |g|^{-1/2}\partial_i(|g|^{1/2}X^i), \quad (\nabla f)^i = g^{ij}\partial_j f,$$

where g_{ij} is the metric tensor, $g^{ij} = (g_{ij})^{-1}$ and $|g| = \det(g_{ij})$. We use the Einstein convention for summing repeated indices.

The parametric approach to SDEs in manifolds is to define the SDE for the variables $\mathbf{q} = (q_1, \cdots, q_d)$ of a patch of the manifold:

$$d\mathbf{q} = \mathbf{h}(\mathbf{q}, t)dt + H(\mathbf{q}, t)d\mathbf{w}. \tag{A.53}$$

The corresponding forward Kolmogorov equation is

$$\frac{\partial f}{\partial t} + |g|^{-1/2}\sum_{i=1}^{d}\frac{\partial}{\partial q_i}\left(|g|^{1/2}h_i f\right) = \frac{1}{2}|g|^{-1/2}\sum_{i,j=1}^{d}\frac{\partial^2}{\partial q_i \partial q_j}\left(|g|^{1/2}\sum_{k=1}^{D}H_{ik}H_{kj}^{\top}f\right), \tag{A.54}$$

which is to be understood in the weak form.

Assume that the manifold $\mathcal{M}$ embedded in $\mathbb{R}^D$. If $\psi : \mathcal{U}_{\mathbb{R}^d} \subseteq \mathbb{R}^d \to \mathcal{U} \subseteq \mathcal{M} \subseteq \mathbb{R}^D$ is the mapping corresponding to the patch $\mathcal{U}$ and (A.53) is defined on $\mathcal{U}_{\mathbb{R}^d}$, let us set $H(\mathbf{q}) = (D\psi(\mathbf{q}))^{-1}$. In this

case, $\sum_k H_{ik}H^\top_{kj} = \sum_k (D\psi)^{-1}_{ik}((D\psi)^{-1}_{kj})^\top = g^{ij}(\mathbf{q})$. Hence, the right hand side of (A.54) becomes

$$\begin{aligned}
&\frac{1}{2}|g|^{-1/2}\sum_{i,j=1}^{d}\frac{\partial^2}{\partial q_i \partial q_j}\left(|g|^{1/2}g^{ij}f\right)\\
&= |g|^{-1/2}\sum_{i=1}^{d}\frac{\partial}{\partial q_i}\left(|g|^{1/2}\tilde{h}_i f\right) + \frac{1}{2}|g|^{-1/2}\sum_{i,j=1}^{d}\frac{\partial}{\partial q_i}\left(|g|^{1/2}g^{ij}\frac{\partial}{\partial q_j}f\right)\\
&= |g|^{-1/2}\sum_{i=1}^{d}\frac{\partial}{\partial q_i}\left(|g|^{1/2}\tilde{h}_i f\right) + \frac{1}{2}|g|^{-1/2}\sum_{i,j=1}^{d}\frac{\partial}{\partial q_i}\left(|g|^{1/2}g^{ij}\frac{\partial}{\partial q_j}f\right)\\
&= \nabla\cdot(\tilde{\mathbf{h}}f) + \frac{1}{2}\nabla\cdot\nabla f
\end{aligned}$$

where

$$\tilde{h}_i(\mathbf{q}) = \frac{1}{2}\sum_{j=1}^{d}\left(|g(\mathbf{q})|^{-1/2}g^{ij}(\mathbf{q})\frac{\partial|G(\mathbf{q})|^{1/2}}{\partial q_j} + \frac{\partial g^{ij}(\mathbf{q})}{\partial q_j}\right)$$

Hence, we can rewrite (A.54) as

$$\frac{\partial f}{\partial t} = \nabla\cdot((-\mathbf{h}+\tilde{\mathbf{h}})f) + \frac{1}{2}\nabla\cdot\nabla f$$

For this equation to be a Fokker-Planck equation with potential E (i.e. with a Gibbs equilibrium solution), we need $-\mathbf{h}+\tilde{\mathbf{h}} = \nabla E$, which implies $\mathbf{h} = -\nabla E + \tilde{\mathbf{h}}$.

We can convert an SDE in parametric form like (A.53) into an SDE on $\mathbb{R}^D$ by using Ito's lemma on $X = \psi(\mathbf{q})$:

$$dX_i = d\psi_i(\mathbf{q}) = \left(D\psi_i(\mathbf{q})\mathbf{h}(\mathbf{q}) + \frac{1}{2}\mathrm{Tr}(H(\mathbf{q},t)^\top(H\psi_i)(\mathbf{q})H(\mathbf{q},t))\right)dt + D\psi_i(\mathbf{q})H(\mathbf{q},t)d\mathbf{w} \quad \text{(A.55)}$$

If we set $H(\mathbf{q}) = (D\psi(\mathbf{q}))^{-1}$ as before, $D\psi(\mathbf{q})H(\mathbf{q},t)$ is the projection onto the tangent space of the manifold, i.e. $D\psi(\mathbf{q})H(\mathbf{q},t)v = \mathrm{Proj}_{T_{\psi(\mathbf{q})}M}v,\ \forall v \in \mathbb{R}^D$. In the case $\mathbf{h} = \nabla E + \tilde{\mathbf{h}}$, $D\psi_i(\mathbf{q})\mathbf{h}(\mathbf{q}) = D\psi_i(\mathbf{q})\nabla E(\mathbf{q}) + D\psi_i(\mathbf{q})\tilde{\mathbf{h}}(\mathbf{q})$. It is very convenient to abuse the notation and denote $D\psi(\mathbf{q})\nabla E(\mathbf{q})$

by $\nabla E(\psi(\mathbf{q}))$. We also use $\hat{\mathbf{h}}(\psi(\mathbf{q})) := D\psi(\mathbf{q})\tilde{\mathbf{h}}(\mathbf{q}) + \frac{1}{2}\mathrm{Tr}(((D\psi(\mathbf{q}))^{-1})^\top (H\psi)(\mathbf{q})(D\psi(\mathbf{q}))^{-1})$. Both definitions are well-defined because the variables are invariant by changes of coordinates. Hence, under these assumptions (A.55) becomes

$$dX = (-\nabla E(X) + \hat{\mathbf{h}}(X))\, dt + \mathrm{Proj}_{T_X M}(d\mathbf{w}) \tag{A.56}$$

In short that means that we can treat SDEs on embedded manifolds as SDEs on the ambient space by projecting the Brownian motions to the tangent space and adding a drift term $\hat{\mathbf{h}}$ that depends on the geometry of the manifold. Notice that for ODEs on manifolds the additional drift term does not appear and (A.56) reads simply $dX = \nabla E(X)dt$.

Notice that the forward Kolmogorov equation for (A.56) on $\mathbb{R}^D$ reads

$$\frac{d}{dt}\int f(x)\, d\mu_t(x) = \int (\nabla E(x) - \hat{\mathbf{h}}(x)) \cdot \nabla_x f(x) + \frac{1}{2}\mathrm{Tr}((\mathrm{Proj}_{T_x M})^\top Hf(x)\mathrm{Proj}_{T_x M})\, d\mu_t(x), \tag{A.57}$$

for an arbitrary f.

B Appendix: Multisample Flow Matching

B.1 Coupling algorithms

Multisample FM makes use of batch coupling algorithms to construct an implicit joint distribution satisfying the marginal constraints. While BatchOT coupling is motivated by approximating the OT map, we consider other lower complexity coupling algorithms which produce coupling that satisfy some desired property of optimal couplings. In Table B.1 we summarize the runtime complexities for the different algorithms used in this work. We will now describe in detail the Stable and Heuristic coupling algorithms.

	CondOT	BatchOT	BatchEOT	Stable	Heuristic
Runtime Complexity	$O(1)$	$O(k^3)$	$\tilde{O}(k^2/\varepsilon)$	$O(k^2 \log(k))$	$O(k^2 \log(k))$

Table B.1: Runtime complexities of the different coupling algorithms as a function of the batch size k.

B.1.1 Stable couplings

[Wolansky 2020] surveys discrete optimal transport from a stable coupling perspective proving that stability is a necessary condition for OT couplings. Although stable couplings are not OT, they are cheaper to compute and are therefore an appealing approach to pursue. For completeness

we formulate the Gale Shapely Algorithm in our setting in Algorithm 1. The rankings R_0, R_1 hold the preferences of the samples in $\{x_0^{(i)}\}_{i=1}^k$ and $\{x_1^{(i)}\}_{i=1}^k$ respectively. Where $R_0(i, j)$ is the rank of $x_1^{(j)}$ in $x_0^{(i)}$'s preferences and $R_1(i, j)$ is the rank of $x_0^{(j)}$ in $x_1^{(i)}$'s preferences.

B.1.2 Heuristic couplings

The stable coupling is agnostic to the cost of pairing samples and only takes into account the ranks. Therefore, reassignments during the Gale Shapely algorithms might increase the total cost although the rankings of assigned samples are improved. We draw inspiration from the cyclic monotonicity of OT couplings [Villani 2008] and from the marriage with sharing formulation in [Wolansky 2020] and modify the reassignment condition in the Gale Shapely algorithm (see Algorithm 2). The modified condition encourages "local" monotonicity between the reassigned pairs only, reassigning a pair only if the potentially newly assigned pairs have a lower cost.

Algorithm 5 Stable Coupling (Gale Shapley)

Result: assignment σ

Data: $\{x_0^{(i)}\}_{i=1}^k \sim q_0(x_0)$, $\{x_1^{(i)}\}_{i=1}^k \sim q_1(x_1)$, rankings R_0, R_1

initialization: σ empty assignment

while $\exists\ i \in [k]$ `s.t.` $\sigma(i)$ *is empty* **do**

  $j \leftarrow$ first sample in $R_0(i, \cdot)$ whom $x_0^{(i)}$ has not tried to match with yet

  if $\exists\ i'$ `s.t.` $\sigma(i') = j$ **then**

    if $R_1(j, i) < R_1(j, i')$ **then**

      $\sigma(i') \leftarrow$ empty

      $\sigma(i) \leftarrow j$

    end

  else

    $\sigma(i) \leftarrow j$

  end

end

Algorithm 6 Heuristic Coupling

Result: assignment σ

Data: $\{x_0^{(i)}\}_{i=1}^k \sim q_0(x_0)$, $\{x_1^{(i)}\}_{i=1}^k \sim q_1(x_1)$, rankings R_0, R_1, cost matrix C

initialization: σ empty assignment

while $\exists\ i \in [k]$ `s.t.` $\sigma(i)$ *is empty* **do**

  $j \leftarrow$ first sample in $R_0(i, \cdot)$ whom $x_0^{(i)}$ has not tried to match with yet

  if $\exists\ i'$ `s.t.` $\sigma(i') = j$ **then**

    $j' \leftarrow$ first sample in $R_0(i', \cdot)$ whom $x_0^{(i')}$ has not tried to match with yet

    $l \leftarrow$ second sample in $R_0(i, \cdot)$ whom $x_0^{(i)}$ has not tried to match with yet

    if $C(i, j) + C(i', j') < C(i, l) + C(i', j)$ **then**

      $\sigma(i') \leftarrow$ empty

      $\sigma(i) \leftarrow j$

    end

  else

    $\sigma(i) \leftarrow j$

  end

end

B.2 Additional tables and figures

B.2.1 Full results on ImageNet data

Model	ImageNet 32×32				ImageNet 64×64			
	NLL	FID	NFE	Var(u_t)	NLL	FID	NFE	Var(u_t)
Ablations[†]								
DDPM [Ho et al. 2020]	3.61	5.72	330		3.27	13.80	323	
ScoreSDE [Song et al. 2021b]	3.61	6.84	198		3.30	26.64	365	
ScoreFlow [Song et al. 2021a]	3.61	9.53	189		3.34	32.78	554	
Flow Matching [W]/ Diffusion [Lipman et al. 2023]	3.60	6.36	165		3.35	15.11	162	
Rectified Flow [Liu et al. 2022]	3.59	5.55	111		3.31	13.02	129	
Flow Matching [W]/ CondOT [Lipman et al. 2023]	3.58	5.04	139	594	3.27	13.93	131	1880
Ours								
Multisample Flow Matching [W]/ StableCoupling	3.59	5.79	148	523	3.27	11.82	132	1782
Multisample Flow Matching [W]/ HeuristicCoupling	3.58	5.29	133	555	3.26	13.37	110	1816
Multisample Flow Matching [W]/ BatchEOT	3.58	6.14	132	508	3.26	14.92	141	1736
Multisample Flow Matching [W]/ BatchOT	3.58	4.68	146	507	3.27	12.37	135	1733

Table B.2: Multisample Flow Matching improves on sample quality and sample efficiency while not trading off performance at all compared to Flow Matching. [†]Reproduction using the same training hyperparameters (architecture, optimizer, training iterations) as our methods.

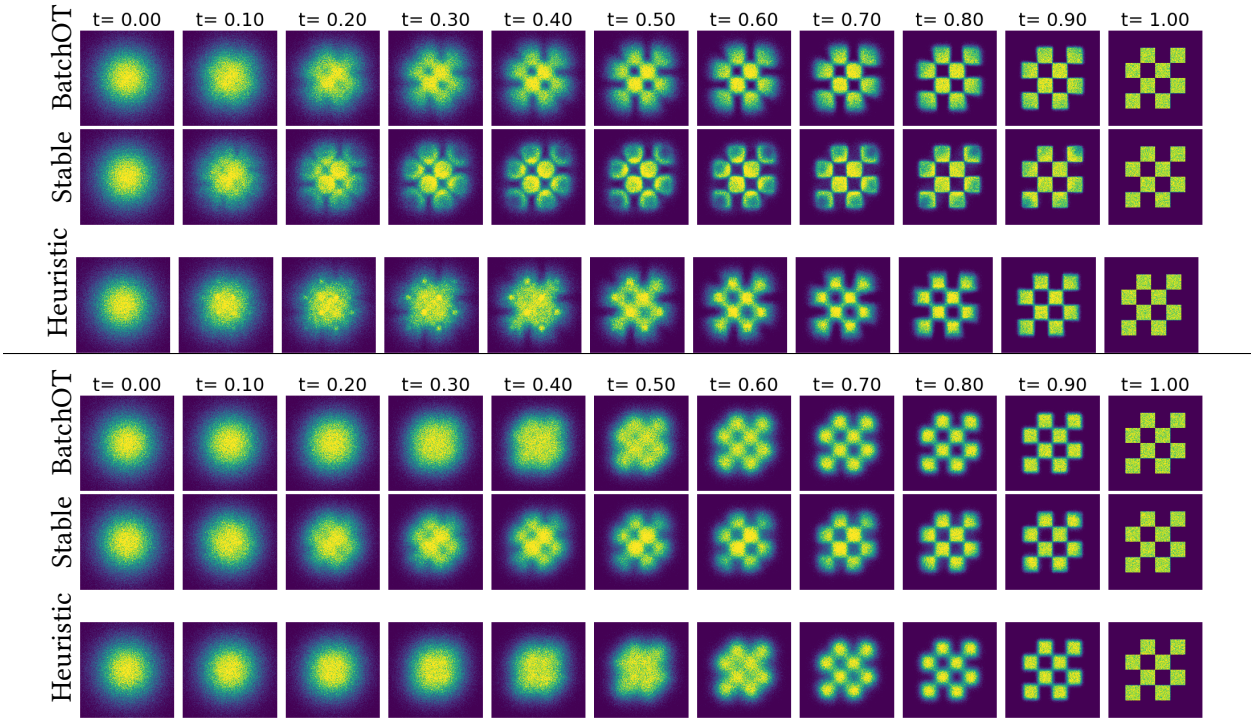


Figure B.1: Marginal probability paths. (*Top*) Batch size 64. (*Bottom*) Batch size 8.

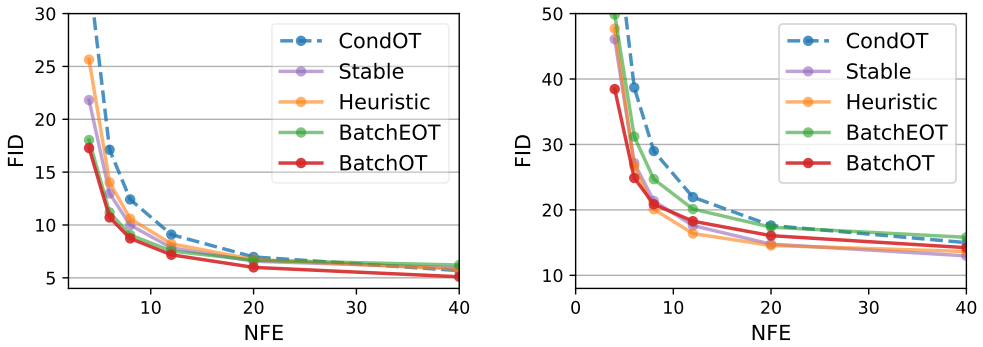
B.2.2 How batch size affects the marginal probability paths on 2D checkerboard data

B.2.3 FID vs NFE using midpoint discretization scheme

B.2.4 Comparison of FID vs NFE for baseline methods DDPM and ScoreSDE

B.2.5 Runtime per iteration is not significantly affected by solving for couplings

B.2.6 Convergence improves when using larger coupling sizes



ImageNet 32×32 ImageNet 64×64

Figure B.2: Sample quality (FID) vs compute cost (NFE); midpoint discretization.

ImageNet32 FID (Euler)					ImageNet32 FID (Midpoint)				
NFE	DDPM	ScoreSDE	BatchOT	Stable	NFE	DDPM	ScoreSDE	BatchOT	Stable
Adaptive	5.72	6.84	**4.68**	5.76	Adaptive	5.72	6.84	**4.68**	5.79
40	19.56	16.96	**5.94**	7.02	40	6.68	6.48	**5.09**	5.94
20	63.08	58.02	**7.71**	8.66	20	7.80	8.96	**5.98**	6.57
12	152.59	140.95	**10.72**	11.10	12	14.87	16.22	**7.18**	7.84
8	232.97	218.66	15.64	**14.89**	8	56.41	56.73	**8.73**	9.99
6	275.28	266.76	22.08	**19.88**	6	188.08	168.99	**10.71**	12.98
4	362.37	340.17	38.86	**33.92**	4	319.41	279.06	**17.28**	21.82

Table B.3: Comparing the FID vs. NFE on ImageNet32 for two baselines and two of our methods.

B.3 Generated samples

ImageNet64 FID (Euler)					ImageNet64 FID (Midpoint)				
NFE	DDPM	ScoreSDE	BatchOT	Stable	NFE	DDPM	ScoreSDE	BatchOT	Stable
Adaptive	13.80	26.64	12.37	**11.82**	Adaptive	13.80	26.64	12.37	**11.82**
40	25.83	44.16	14.79	**13.39**	39	15.3	26.67	14.22	**12.97**
20	66.42	82.97	17.06	**15.15**	15	15.05	25.73	16.05	**14.76**
12	158.46	141.79	20.94	**18.21**	11	18.91	29.99	18.27	**17.60**
8	258.49	210.29	27.56	**26.38**	7	53.15	67.83	**20.85**	21.36
6	321.04	262.20	**36.17**	37614	5	179.79	155.91	**24.87**	27.15
4	373.08	335.54	**56.75**	63425	3	330.53	279.00	**38.45**	46.08

Table B.4: Comparing the FID vs. NFE on ImageNet64 for two baselines and two of our methods.

	ImageNet 32×32		**ImageNet 64×64**	
	It./s	Rel. increase	It./s	Rel. increase
CondOT (reference)	1.16	—	1.31	—
BatchOT	1.15	0.8%	1.26	3.9%
Stable	1.15	0.8%	1.26	3.9%

Table B.5: Absolute and relative runtime comparisons between CondOT, BatchOT and Stable matching. “It./s" denotes the number of iterations per second, and “Rel. increase" is the relative increase with respect to CondOT. Note that these are on relatively standard batch sizes (refer to section B.5 for exact batch sizes).

B.4 Theorems and proofs

B.4.1 Proof of Lemma 3.1

We need only prove that the marginal probability path interpolates between q_0 and q_1.

$$p_0(x) = \int p_0(x|x_1)q_1(x_1)dx_1 = \int q(x|x_1)q_1(x_1)dx_1 = q_0(x).$$

Then since $u_t(x|x_1)$ transports all points $x \in \mathbb{R}^D$ to x_1 at time $t = 1$, we satisfy $p_{t=1}(x|x_1) = \delta(x - x_1)$.

$$p_1(x) = \int p_1(x|x_1)q_1(x_1)dx = \int \delta(x - x_1)q_1(x_1)dx_1 = q_1(x).$$

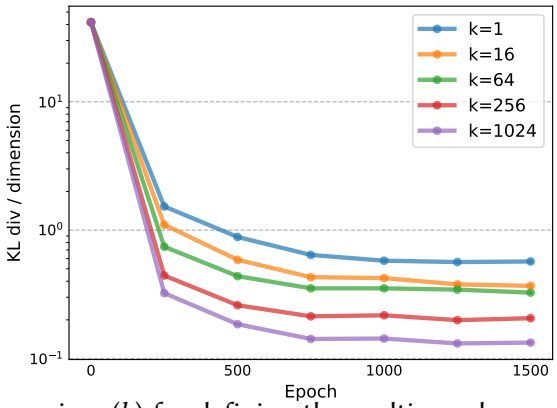


Figure B.3: Larger couplings sizes (k) for defining the multisample coupling results in faster and more stable convergence. This is done on the 64-D experiments in subsection 3.6.3. The batch size (number of samples) for training is kept thestr same and only k is varied for solving the couplings.

Theorems 1 and 2 of Lipman et al. [2023] can then be used to prove that (i) the marginal vector field $u_t(x)$ transports between $p_0 = q_0$ and $p_1 = q_1$, and (ii) the Joint CFM objective has the same gradient in expectation as the Flow Matching objective and is uniquely minimized by $v_t(x;\theta) = u_t(x)$.

B.4.2 Proof of Lemma 3.2

Note that

$$\begin{aligned}\mathrm{Cov}_{p_t(x_1|x)}\left(\nabla_\theta \left\|v_t(x;\theta) - u_t(x|x_1)\right\|^2\right) &= \mathrm{Cov}_{p_t(x_1|x)}\left(\nabla_\theta \left\|v_t(x;\theta)\right\|^2 - \left(\nabla_\theta v_t(x;\theta)\right)^\top u_t(x|x_0,x_1)\right) \\ &= \left(\nabla_\theta v_t(x;\theta)\right)^\top \mathrm{Cov}_{p_t(x_1|x)}\left(u_t(x|x_1)\right)\left(\nabla_\theta v_t(x;\theta)\right),\end{aligned} \tag{B.1}$$

and that

$$\mathrm{Cov}_{p_t(x_1|x)}\left(u_t(x|x_1)\right) = \mathbb{E}_{p_t(x_1|x)}\left(u_t(x|x_1) - u_t(x)\right)\left(u_t(x|x_1) - u_t(x)\right)^\top. \tag{B.2}$$

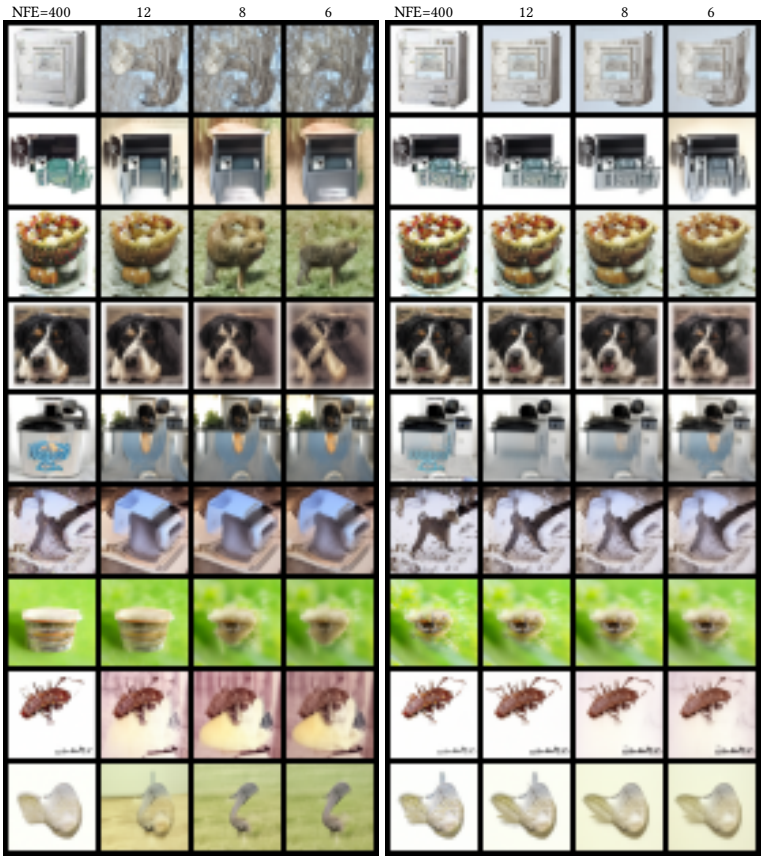


Figure B.4: Multisample Flow Matching trained with batch optimal couplings produces more consistent samples across varying NFEs on ImageNet32. From left to right, the NFEs used to generate these samples are 200, 12, 8, and 6 using a midpoint discretization. Note that both flows on each row start from the same noise sample.

Here, we used that $u_t(x) = \mathbb{E}_{p_t(x_1|x)}\left[u_t(x|x_1)\right]$ by (3.4). If we take the trace on both sides of (B.1), we get

$$
\begin{aligned}
&\mathrm{Tr}\left[\mathrm{Cov}_{p_t(x_1|x)}\left(\nabla_\theta \left\|v_t(x;\theta) - u_t(x|x_1)\right\|^2\right)\right] = \mathrm{Tr}\left[\left(\nabla_\theta v_t(x;\theta)\right)^{\mathsf{T}} \mathrm{Cov}_{p_t(x_1|x)}\left(u_t(x|x_1)\right)\left(\nabla_\theta v_t(x;\theta)\right)\right] \\
&= \mathrm{Tr}\left[\mathrm{Cov}_{p_t(x_1|x)}\left(u_t(x|x_1)\right)\left(\nabla_\theta v_t(x;\theta)\right)\left(\nabla_\theta v_t(x;\theta)\right)\right] \\
&= \left\langle \mathrm{Cov}_{p_t(x_1|x)}\left(u_t(x|x_1)\right), \left(\nabla_\theta v_t(x;\theta)\right)\left(\nabla_\theta v_t(x;\theta)\right)^{\mathsf{T}} \right\rangle_F \\
&\leq \left\|\mathrm{Cov}_{p_t(x_1|x)}\left(u_t(x|x_1)\right)\right\|_F \left\|\left(\nabla_\theta v_t(x;\theta)\right)\left(\nabla_\theta v_t(x;\theta)\right)^{\mathsf{T}}\right\|_F \\
&\leq \mathbb{E}_{p_t(x_1|x)} \left\|\left(u_t(x|x_1) - u_t(x)\right)\left(u_t(x|x_1) - u_t(x)\right)^{\mathsf{T}}\right\|_F \left\|\left(\nabla_\theta v_t(x;\theta)\right)\left(\nabla_\theta v_t(x;\theta)\right)^{\mathsf{T}}\right\|_F \\
&= \left\|\nabla_\theta v_t(x;\theta)\right\|^2 \mathbb{E}_{p_t(x_1|x)} \left\|u_t(x|x_1) - u_t(x)\right\|^2.
\end{aligned}
$$

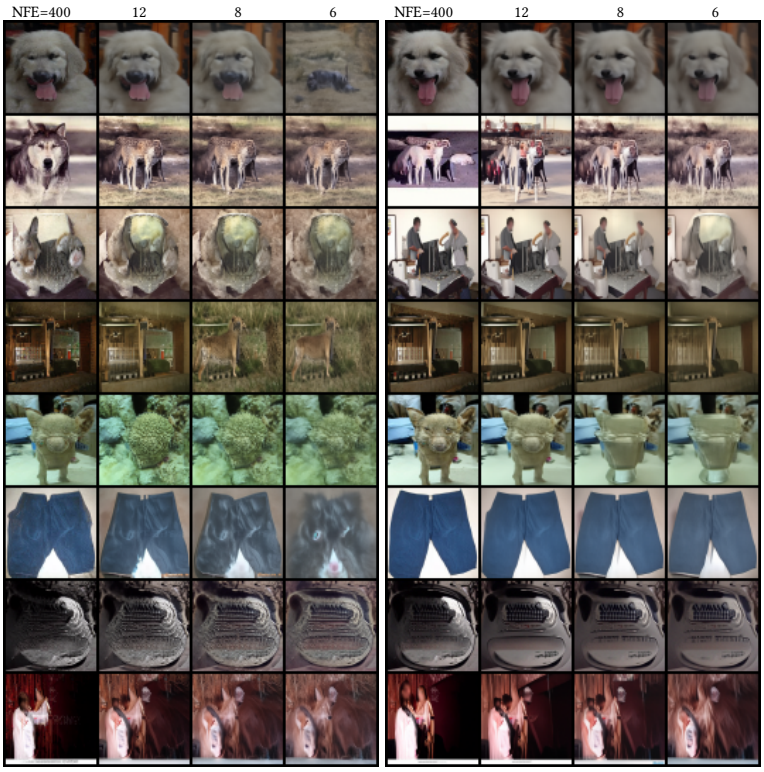


Figure B.5: Multisample Flow Matching trained with batch optimal couplings produces more consistent samples across varying NFEs on ImageNet64. From left to right, the NFEs used to generate these samples are 200, 12, 8, and 6 using a midpoint discretization. Note that both flows on each row start from the same noise sample.

The second equality holds because $\mathrm{Tr}(AB) = \mathrm{Tr}(BA)$ when both expressions are well defined, and the third equality holds by the definition of the Frobenius inner product $\langle \cdot, \cdot \rangle_F$. The first inequality holds by the Cauchy-Schwarz inequality. The second inequality holds by equation (B.2) and by the triangle inequality. In the last equality we used that for any vector v, $\|vv^\top\|_F = (\mathrm{Tr}(vv^\top, vv^\top))^{1/2} = \|v\|^2$. This proves (3.13).

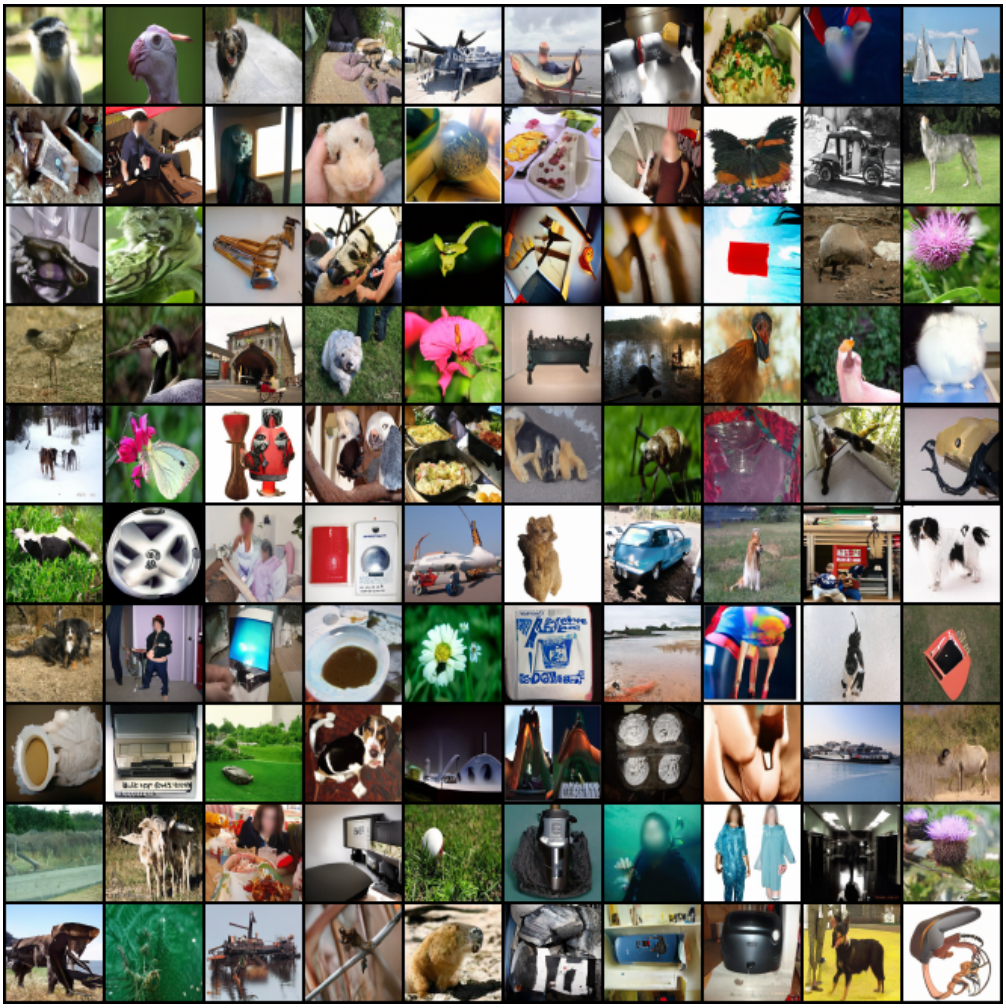


Figure B.6: Non-curated generated images for ImageNet64 using Multisample Flow Matching with BatchOT coupling.

To prove (3.14), we write:

$$
\begin{aligned}
&\mathbb{E}_{t,p_t(x)}[\sigma_{t,x}^2] \\
&\leq \mathbb{E}_{t,p_t(x)}[\|\nabla_\theta v_t(x;\theta)\|^2 \mathbb{E}_{p_t(x_1|x)}\|u_t(x|x_1) - u_t(x)\|^2] \\
&\leq \max_{x,t} \|\nabla_\theta v_t(x;\theta)\|^2 \times \mathbb{E}_{t,p_t(x)}[\mathbb{E}_{p_t(x_1|x)}\|u_t(x|x_1) - u_t(x)\|^2] \\
&= \max_{x,t} \|\nabla_\theta v_t(x;\theta)\|^2 \times \mathbb{E}_{t,q(x_0,x_1)}[\|u_t(x_t|x_1) - v_t(x_t;\theta)\|^2] \leq \max_{t,x} \|\nabla_\theta v_t(x;\theta)\|^2 \times \mathcal{L}_{\text{JCFM}}
\end{aligned}
$$

Here, the first inequality holds by (3.13), and the last inequality holds because $u_t(x)$ is the mini-

mizer of $\mathcal{L}_{\text{JCFM}}$.

B.4.3 Proof of Lemma 3.3

For an arbitrary test function f, by the construction of q we write

$$\mathbb{E}_{q(x_0,x_1)} f(x_0) = \mathbb{E}_{\{x_0^{(i)}\}_{i=1}^k \sim q_0, \{x_1^{(i)}\}_{i=1}^k \sim q_1} \mathbb{E}_{q^k(x_0,x_1)} f(x_0).$$

Since q^k has marginal $\frac{1}{k}\sum_{i=1}^k \delta(x_0 - x_0^{(i)})$ because π is a doubly stochastic matrix, we obtain that $\mathbb{E}_{q^k(x_0,x_1)} f(x_0) = \frac{1}{k}\sum_{i=1}^k f(x_0^{(i)})$ and then the right-hand side is equal to

$$\mathbb{E}_{\{x_0^{(i)}\}_{i=1}^k \sim q_0, \{x_1^{(i)}\}_{i=1}^k \sim q_1} \frac{1}{k}\sum_{i=1}^k f(x_0^{(i)}) = \mathbb{E}_{q_0(x_0)} f(x_0),$$

which proves that the marginal of q for x_0 is q_0. The same argument works for the x_1 marginal.

B.4.4 Proof of Theorem 3.4

Notation We begin by recalling and introducing some additional notation. Let $\boldsymbol{X}_0 = (x_0^i)_{i=1}^{+\infty}$, $\boldsymbol{X}_1 = (x_1^i)_{i=1}^{+\infty}$ be sequences of i.i.d. samples from the distributions q_0 and q_1, and denote by $\boldsymbol{X}_0^k = (x_0^i)_{i=1}^k$, $\boldsymbol{X}_1^k = (x_1^i)_{i=1}^k$ the finite sequences containing the initial k samples. We denote by q_0^k and q_1^k the empirical distributions corresponding to $\boldsymbol{X}_0^k$ and $\boldsymbol{X}_1^k$, i.e. $q_0^k = \frac{1}{k}\sum_{i=1}^k \delta_{x_0^i}$, $q_1^k = \frac{1}{k}\sum_{i=1}^k \delta_{x_1^i}$. Let q^k be the distribution over $\mathbb{R}^d \times \mathbb{R}^d$ which is output by the matching algorithm; q^k has marginals that are equal to q_0^k and q_1^k. Let q^* be the optimal transport plan between q_0 and q_1, and let $\tilde{q}^k$ be the optimal transport plan between q^k and q under the quadratic cost. Using this additional notation, we rewrite some of the objects that were defined in the main text in a lengthier, more precise way:

(i) The marginal vector field corresponding to sample size k:

$$u_t^k(x) = \mathbb{E}_{\boldsymbol{X}_0^k \overset{\text{iid}}{\sim} q_0, \boldsymbol{X}_1^k \overset{\text{iid}}{\sim} q_1, (x_0,x_1)\sim q^k}[x_1 - x_0 | x = tx_1 + (1-t)x_0], \qquad \forall t \in [0,1]. \tag{B.3}$$

We made the dependency on k explicit, and we used that $\psi_t(x_0|x_1) = tx_1 + (1-t)x_0$. Note that equivalently, we can write u_t^k as the solution of a simple variational problem.

$$u_t^k = \arg\min_{u_t} \mathbb{E}_{\boldsymbol{X}_0^k \overset{\text{iid}}{\sim} q_0, \boldsymbol{X}_1^k \overset{\text{iid}}{\sim} q_1, (x_0,x_1)\sim q^k} \|x_1 - x_0 - u_t(tx_1 + (1-t)x_0)\|^2, \qquad \forall t \in [0,1]. \tag{B.4}$$

(ii) The flow $\psi_t^k(x_0)$ corresponding to u_t^k, i.e. the solution of $\frac{dx_t}{dt} = u_t^k(x_t)$ with initial condition x_0. We made the dependency on k explicit.

(iii) The straightness of the flow ψ_t^k:

$$S^k = \mathbb{E}_{t\sim \mathrm{U}(0,1), x_0 \sim q_0}\left[\|u_t^k(\psi_t^k(x_0))\|^2 - \|\psi_1^k(x_0) - x_0\|^2\right]. \tag{B.5}$$

Assumptions We will use the following three assumptions, which allow us to potentially extend our result beyond BatchOT:

(A1) The distributions q_0 and q_1 over $\mathbb{R}^d$ have bounded supports, i.e. there exists $C > 0$ such that for any $x \in \mathrm{supp}(q_0) \cup \mathrm{supp}(q_1)$, $\|x\| \leq C$.

(A2) q_0 admits a density and the optimal transport map T between q_0 and q_1 under the quadratic cost is continuous.

(A3) We assume that almost surely w.r.t. the draw of $\boldsymbol{X}_0$ and $\boldsymbol{X}_1$, q^k converges weakly to q as $k \to \infty$.

Some comments are in order as to when assumptions **(A2)**, **(A3)** hold, since they are not directly verifiable. By the Caffarelli regularity theorem (see Villani [2008], Ch. 12, originally in Caffarelli [1992]), a sufficient condition for **(A2)** to hold is the following:

(A2') q_0 and q_1 have a common support Ω which is compact and convex, have α-Hölder densities, and they satisfy the lower bound $q_0, q_1 > \gamma$ for some $\gamma > 0$.

Assumption **(A3)** holds when the matching algorithm is BatchOT, that is, when q^k is the optimal transport plan between q_0^k and q_1^k, as shown by the following proposition, which is proven in App. B.4.4.3.

Proposition B.1. *Let q^k be the optimal transport plan between q_0^k and q_1^k under the quadratic cost (i.e. the result of Steps [1-3] under BatchOT). We have that almost surely w.r.t. the draws of $\boldsymbol{X}_0$ and $\boldsymbol{X}_1$, the sequence $(q_k)_{k\geq 0}$ converges weakly to q^*, i.e. assumption **(A3)** holds.*

Proof structure We split the proof of Theorem 3.4 into two parts: in Subsubsec. B.4.4.1 we prove that the optimal value of the Joint CFM objective (3.12) converges to zero as $k \to \infty$. In Subsubsec. B.4.4.2, we prove that the straightness converges to zero and the transport cost converges to the optimal transport cost as $k \to \infty$.

B.4.4.1 Convergence of the optimal value of the CFM objective

Theorem B.2. *Suppose that assumptions **(A1)**, **(A2)** and **(A3)** hold. We have that*

$$\lim_{k\to\infty} \mathbb{E}_{t\sim \mathrm{U}(0,1), \boldsymbol{X}_0^k \overset{\text{iid}}{\sim} q_0, \boldsymbol{X}_1^k \overset{\text{iid}}{\sim} q_1, (x_0,x_1)\sim q^k} \|x_1 - x_0 - u_t^k(tx_1 + (1-t)x_0)\|^2 = 0,$$

where u_t^k is the marginal vector field as defined in (B.3).

Proof. The transport plan q^* satisfies the non-crossing paths property, that is, for each $x \in \mathbb{R}^d$ and $t \in [0,1]$, there exists at most one pair (x_0, x_1) such that $x = tx_1 + (1-t)x_0$ [Nurbekyan et al. 2020; Villani 2003]. Consequently, when such a pair (x_0', x_1') exists, we have that the analogue of the vector field in (B.3) admits a simple expression:

$$u_t^*(x) := \mathbb{E}_{(x_0,x_1)\sim q^*}[x_1 - x_0 | x = tx_1 + (1-t)x_0] = x_1' - x_0' \tag{B.6}$$

This directly implies that

$$\mathbb{E}_{(x_0,x_1)\sim q^*}\|x_1 - x_0 - u_t^*(tx_1 + (1-t)x_0)\|^2 = 0.$$

Applying this, we can write

$$\begin{aligned}
&\mathbb{E}_{t\sim\mathrm{U}(0,1),(x_0,x_1)\sim q^k}\|x_1 - x_0 - u_t^*(tx_1 + (1-t)x_0)\|^2 \\
&= |\mathbb{E}_{(x_0,x_1)\sim q^k}[\mathbb{E}_{t\sim\mathrm{U}(0,1)}\|x_1 - x_0 - u_t^*(tx_1 + (1-t)x_0)\|^2] \\
&\quad - \mathbb{E}_{(x_0,x_1)\sim q^*}[\mathbb{E}_{t\sim\mathrm{U}(0,1)}\|x_1 - x_0 - u_t^*(tx_1 + (1-t)x_0)\|^2]|. \qquad \text{(B.7)}
\end{aligned}$$

Now, define the function $f : \mathrm{supp}(q_0) \times \mathrm{supp}(q_1) \to \mathbb{R}$ as

$$f(x_0, x_1) = \mathbb{E}_{t\sim\mathrm{U}(0,1)}\|x_1 - x_0 - u_t^*(tx_1 + (1-t)x_0)\|^2. \qquad \text{(B.8)}$$

By Lemma B.3, which holds under **(A1)** and **(A2)**, we have that f is bounded and continuous. Assumption **(A3)** states that almost surely w.r.t. the draws of $\boldsymbol{X}_0$ and $\boldsymbol{X}_1$, the measure q_k converges weakly to q^*. We apply the definition of weak convergence of measures, which implies that almost surely,

$$\lim_{k\to\infty} \mathbb{E}_{(x_0,x_1)\sim q^k}[f(x_0, x_1)] = \mathbb{E}_{(x_0,x_1)\sim q}[f(x_0, x_1)].$$

Equivalently, the right-hand side of (B.7) converges to zero as k tends to infinity. Hence, we obtain that $\mathbb{E}_{t\sim\mathrm{U}(0,1),(x_0,x_1)\sim q^k}\|x_1 - x_0 - u_t^*(tx_1 + (1-t)x_0)\|^2 \to 0$ almost surely. Almost sure convergence implies convergence in probability, which means that

$$\Pr(\mathbb{E}_{t\sim\mathrm{U}(0,1),(x_0,x_1)\sim q^k}\|x_1 - x_0 - u_t^*(tx_1 + (1-t)x_0)\|^2 > \epsilon) \xrightarrow{k\to\infty} 0, \quad \forall\epsilon > 0.$$

Here, the randomness comes only from drawing the random variables $\boldsymbol{X}_0^k, \boldsymbol{X}_1^k$. Also, using again that f is bounded, say by the constant $C > 0$, we can write $\mathbb{E}_{t\sim \mathrm{U}(0,1),(x_0,x_1)\sim q^k}\|x_1 - x_0 - u_t^*(tx_1 + (1-t)x_0)\|^2 \leq C$, for all $k \geq 0$. A crude bound yields

$$\begin{aligned}
&\mathbb{E}_{t\sim \mathrm{U}(0,1),\boldsymbol{X}_0^k,\boldsymbol{X}_1^k,(x_0,x_1)\sim q^k}\|x_1 - x_0 - u_t^*(tx_1 + (1-t)x_0)\|^2 \\
&\leq \epsilon + C\mathrm{Pr}(\mathbb{E}_{t\sim \mathrm{U}(0,1),(x_0,x_1)\sim q^k}\|x_1 - x_0 - u_t^*(tx_1 + (1-t)x_0)\|^2 > \epsilon).
\end{aligned}$$

In this equation and from now, we write $\boldsymbol{X}_0^k, \boldsymbol{X}_1^k$ instead of $\boldsymbol{X}_0^k \overset{\text{iid}}{\sim} q_0, \boldsymbol{X}_1^k \overset{\text{iid}}{\sim} q_1$ for shortness. We can take ϵ arbitrarily small, and for a given ϵ we can make the second term in the right-hand side arbitrarily small by taking k large enough. Hence, we obtain that

$$\lim_{k\to\infty} \mathbb{E}_{t\sim \mathrm{U}(0,1),\boldsymbol{X}_0^k,\boldsymbol{X}_1^k,(x_0,x_1)\sim q^k}\|x_1 - x_0 - u_t^*(tx_1 + (1-t)x_0)\|^2 = 0.$$

To conclude the proof, we use the variational characterization of u_t^k given in (B.4), which implies that

$$\begin{aligned}
&\mathbb{E}_{t\sim \mathrm{U}(0,1),\boldsymbol{X}_0^k,\boldsymbol{X}_1^k,(x_0,x_1)\sim q^k}\|x_1 - x_0 - u_t^k(tx_1 + (1-t)x_0)\|^2 \\
&\leq \mathbb{E}_{t\sim \mathrm{U}(0,1),\boldsymbol{X}_0^k,\boldsymbol{X}_1^k,(x_0,x_1)\sim q^k}\|x_1 - x_0 - u_t^*(tx_1 + (1-t)x_0)\|^2 \to 0.
\end{aligned}$$

□

Lemma B.3. *Let f be the function defined in equation (B.8). Suppose that assumptions **(A1)** and **(A2)** hold. Then, f is bounded and continuous.*

Proof. First, we show that the function u_t^* defined in equation (B.6) is bounded and continuous wherever it is defined. It is bounded because $u_t^*(x) = x_1' - x_0'$ for some x_0' in $\mathrm{supp}(q_0)$ and x_1' in $\mathrm{supp}(q_1)$, which are both bounded by assumption.

To show that u_t^* is continuous, we use that q_0 is absolutely continuous and that consequently a

transport map T exists. Moreover, we have that $x_1' = T(x_0')$. Consider the transport map T_t at time t, defined as $T_t(x) = tT(x) + (1-t)x$. Thus, we can write that $u_t^*(T_t(x_0)) = T(x_0) - x_0$. The non-crossing paths property implies that T_t is invertible, which means that an inverse T_t^{-1} exists. We can write

$$u_t^*(x) = T(T_t^{-1}(x)) - T_t^{-1}(x). \tag{B.9}$$

By assumption **(A2)**, the transport map T is continuous, and so is T_t. It is well-known fact that if E, E' are metric spaces, E is compact, and $f : E \to E'$ a continuous bijective function, then $f^{-1} : E' \to E$ is continuous. Thus, T_t^{-1} is also continuous. From equation (B.9), we conclude that u_t^* is continuous.

The rest of the proof is straightforward: $(x_1, x_0) \mapsto \|x_1 - x_0 - u_t^*(tx_1 + (1-t)x_0)\|^2$ is bounded and continuous on the bounded supports of q_0 and q_1 for all $t \in [0,1]$, and then f is also continuous and bounded since it is an average of continuous bounded functions, applying the dominated convergence theorem. □

B.4.4.2 Convergence of the straightness and the transport cost

Theorem B.4. *Suppose that assumptions **(A1)** and **(A3)** hold. Then,*

(i) We have that $\lim_{k\to\infty} S^k = 0$, *where* S^k *is the straightness defined in* (B.5).

(ii) We also have that

$$\mathbb{E}_{t\sim \mathrm{U}(0,1), x_0\sim q_0} \|u_t^k(\psi_t^k(x_0))\|^2 \geq \mathbb{E}_{x_0\sim q_0} \|\psi_1^k(x_0) - x_0\|^2 \geq W_2^2(q_0, q_1), \tag{B.10}$$

$$\lim_{k\to\infty} \mathbb{E}_{t\sim \mathrm{U}(0,1), x_0\sim q_0} \|u_t^k(\psi_t^k(x_0))\|^2 = \lim_{k\to\infty} \mathbb{E}_{x_0\sim q_0} \|\psi_1^k(x_0) - x_0\|^2 = W_2^2(q_0, q_1).$$

Proof. We begin with the proof of (i). We introduce some additional notation. We define the

quantity S^* in analogy with S^k:

$$S^* = \mathbb{E}_{t\sim \mathrm{U}(0,1), x_0\sim q_0}\left[\|u_t^*(\psi_t^*(x_0))\|^2 - \|\psi_1^*(x_0) - x_0\|^2\right],$$

and $\psi_t^*(x_0)$ as the solution of the ODE $\frac{dx_t}{dt} = u_t^*(x_t)$. Since the trajectories for the optimal transport vector field are straight lines, we deduce from the alternative expression of the straightness (equation (3.16)) that $S^* = 0$. An alternative way to see this is by the Benamou-Brenier theorem [Benamou and Brenier 2000], which states that the dynamic optimal transport cost $\mathbb{E}_{t\sim \mathrm{U}(0,1), x_0\sim q_0}\|u_t^*(\psi_t^*(x_0))\|^2$ is equal to the static optimal transport cost $\mathbb{E}_{t\sim \mathrm{U}(0,1), x_0\sim q_0}\|\psi_1^*(x_0) - x_0\|^2$.

We will first prove that $\mathbb{E}_{t\sim \mathrm{U}(0,1), x_0\sim q_0}\|u_t^k(\psi_t^k(x_0))\|^2$ converges to $\mathbb{E}_{t\sim \mathrm{U}(0,1), x_0\sim q_0}\|u_t^*(\psi_t^*(x_0))\|^2$ and then that $\mathbb{E}_{t\sim \mathrm{U}(0,1), x_0\sim q_0}\|\psi_1^k(x_0) - x_0\|^2$ converges to $\mathbb{E}_{t\sim \mathrm{U}(0,1), x_0\sim q_0}\|\psi_1^*(x_0) - x_0\|^2$.

For given instances of $\boldsymbol{X}_0^k$ and $\boldsymbol{X}_1^k$, let $\tilde{q}_k$ be the optimal transport plan between the optimal transport plans q and q^k. In other words, $\tilde{q}_k$ is a measure over the variables x_0, x_1, x_0', x_1', and is such that its marginal w.r.t. x_0, x_1 is q, while its marginal w.r.t. x_0', x_1' is q^k.

That is, we will use that for all $t \in [0,1]$, the random variable $tx_1 + (1-t)x_0$, with $(x_0, x_1) \sim q^k$, and q^k built randomly from $\boldsymbol{X}_0^k \overset{\text{iid}}{\sim} q_0, \boldsymbol{X}_1^k \overset{\text{iid}}{\sim} q_1$, has the same distribution as the random variable $\psi_t^k(x_0)$, with $x_0 \sim q_0$. This is a direct consequence of Lemma 3.1, i.e. the marginal vector field u_t generates the marginal probability path p_t. An analogous statement holds for q, i.e. the random variable $tx_1 + (1-t)x_0$, with $(x_0, x_1) \sim q$, has the same distribution as the random variable $\psi_t^*(x_0)$, with $x_0 \sim q_0$. However, in this case it can be obtained immediately by the non-crossing paths property of the optimal transport plan. Hence,

$$\mathbb{E}_{t\sim \mathrm{U}(0,1), x_0\sim q_0}\|u_t^*(\psi_t^*(x_0))\|^2 = \mathbb{E}_{t\sim \mathrm{U}(0,1), (x_0,x_1)\sim q}\|u_t^*(tx_1 + (1-t)x_0)\|^2,$$
$$\mathbb{E}_{t\sim \mathrm{U}(0,1), x_0\sim q_0}\|u_t^k(\psi_t^k(x_0))\|^2 = \mathbb{E}_{t\sim \mathrm{U}(0,1), \boldsymbol{X}_0^k, \boldsymbol{X}_1^k, (x_0,x_1)\sim q^k}\|u_t^k(tx_1 + (1-t)x_0)\|^2.$$

Using this and the definition of $\tilde{q}^k$, and applying Jensen's inequality, the Cauchy-Schwarz inequality and the triangle inequality, we can write

$$
\begin{aligned}
&\left|\mathbb{E}_{t\sim\mathrm{U}(0,1),x_0\sim q_0}\|u_t^*(\psi_t^*(x_0))\|^2 - \mathbb{E}_{t\sim\mathrm{U}(0,1),x_0\sim q_0}\|u_t^k(\psi_t^k(x_0))\|^2\right| \\
&= \left|\mathbb{E}_{t\sim\mathrm{U}(0,1),(x_0,x_1)\sim q}\|u_t^*(tx_1+(1-t)x_0)\|^2 - \mathbb{E}_{t\sim\mathrm{U}(0,1),X_0^k,X_1^k,(x_0',x_1')\sim q^k}\|u_t^k(tx_1'+(1-t)x_0')\|^2\right| \\
&= \left|\mathbb{E}_{t\sim\mathrm{U}(0,1),X_0^k,X_1^k,(x_0,x_1,x_0',x_1')\sim\tilde{q}^k}\left[\|u_t^*(tx_1+(1-t)x_0)\|^2 - \|u_t^k(tx_1'+(1-t)x_0')\|^2\right]\right| \\
&= \Big|\mathbb{E}_{t\sim\mathrm{U}(0,1),X_0^k,X_1^k,(x_0,x_1,x_0',x_1')\sim\tilde{q}^k}\Big[\left(\|u_t^*(tx_1+(1-t)x_0)\| - \|u_t^k(tx_1'+(1-t)x_0')\|\right) \\
&\qquad\qquad \times\left(\|u_t^*(tx_1+(1-t)x_0)\| + \|u_t^k(tx_1'+(1-t)x_0')\|\right)\Big]\Big| \\
&\le \left(\mathbb{E}_{t\sim\mathrm{U}(0,1),X_0^k,X_1^k,(x_0,x_1,x_0',x_1')\sim\tilde{q}^k}\left[\left(\|u_t^*(tx_1+(1-t)x_0)\| - \|u_t^k(tx_1'+(1-t)x_0')\|\right)^2\right]\right)^{1/2} \\
&\qquad \times\left(\mathbb{E}_{t\sim\mathrm{U}(0,1),X_0^k,X_1^k,(x_0,x_1,x_0',x_1')\sim\tilde{q}^k}\left[\left(\|u_t^*(tx_1+(1-t)x_0)\| + \|u_t^k(tx_1'+(1-t)x_0')\|\right)^2\right]\right)^{1/2} \\
&\le \left(\mathbb{E}_{t\sim\mathrm{U}(0,1),X_0^k,X_1^k,(x_0,x_1,x_0',x_1')\sim\tilde{q}^k}\|u_t^*(tx_1+(1-t)x_0) - u_t^k(tx_1'+(1-t)x_0')\|^2\right)^{1/2} \\
&\qquad \times\left(\mathbb{E}_{t\sim\mathrm{U}(0,1),X_0^k,X_1^k,(x_0,x_1,x_0',x_1')\sim\tilde{q}^k}\left[\left(\|u_t^*(tx_1+(1-t)x_0)\| + \|u_t^k(tx_1'+(1-t)x_0')\|\right)^2\right]\right)^{1/2}.
\end{aligned}
$$

Remark that the second factor in the right-hand side is bounded because u_t^* and u_t^k are bounded. Using Lemma B.5, we obtain that the first factor in the right-hand side tends to zero as k grows. Thus,

$$
\left|\mathbb{E}_{t\sim\mathrm{U}(0,1),x_0\sim q_0}\|u_t^*(\psi_t^*(x_0))\|^2 - \mathbb{E}_{t\sim\mathrm{U}(0,1),x_0\sim q_0^k}\|u_t^k(\psi_t^k(x_0))\|^2\right| \xrightarrow{k\to\infty} 0. \tag{B.11}
$$

Now, since $\mathbb{E}_{x_0\sim q_0}\|\psi_1^*(x_0)-x_0\|^2$ is the optimal cost and $S^*=0$, we write

$$
\begin{aligned}
&\left|\mathbb{E}_{x_0\sim q_0}\|\psi_1^*(x_0)-x_0\|^2 - \mathbb{E}_{x_0\sim q_0}\|\psi_1^k(x_0)-x_0\|^2\right| \\
&= \mathbb{E}_{x_0\sim q_0}\|\psi_1^k(x_0)-x_0\|^2 - \mathbb{E}_{x_0\sim q_0}\|\psi_1^*(x_0)-x_0\|^2 \\
&= \mathbb{E}_{x_0\sim q_0}\|\psi_1^k(x_0)-x_0\|^2 - \mathbb{E}_{t\sim\mathrm{U}(0,1),x_0\sim q_0}\|u_t^*(\psi_t^*(x_0))\|^2.
\end{aligned} \tag{B.12}
$$

Since ψ_t^k is the flow of u_t^k and by Jensen's inequality, we have that

$$\begin{aligned}\mathbb{E}_{x_0\sim q_0}\|\psi_1^k(x_0)-x_0\|^2 &= \mathbb{E}_{x_0\sim q_0}\left\|\int_0^1 u_s^k(\psi_s^k(x_0'))\,ds\right\|^2 \\ &\le \mathbb{E}_{x_0\sim q_0}\int_0^1 \|u_s^k(\psi_s^k(x_0))\|^2\,ds = \mathbb{E}_{t\sim U(0,1),x_0\sim q_0}\|u_t^k(\psi_t^k(x_0))\|^2.\end{aligned}$$

Plugging this into (B.12), we get that

$$\begin{aligned}&\left|\mathbb{E}_{x_0\sim q_0}\|\psi_1^*(x_0)-x_0\|^2-\mathbb{E}_{x_0\sim q_0}\|\psi_1^k(x_0)-x_0\|^2\right| \\ &\le \mathbb{E}_{t\sim U(0,1),x_0\sim q_0}\|u_t^k(\psi_t^k(x_0))\|^2-\mathbb{E}_{t\sim \mathrm{U}(0,1),x_0\sim q_0}\|u_t^*(\psi_t^*(x_0))\|^2 \xrightarrow{k\to\infty} 0,\end{aligned} \tag{B.13}$$

where the limit holds by . Putting together (B.11) and (B.4.4.2), we end up with $S^k = |S^*-S^k| \xrightarrow{k\to\infty}$ 0, which proves (i).

We prove (ii). The first inequality in (B.10) holds because $S^k \ge 0$ since it can be written in a form analogous to (3.16). The second inequality in (B.10) holds because $\mathbb{E}_{x_0\sim q_0}\|\psi_1^k(x_0)-x_0\|^2$ is the squared transport cost for the map $x \mapsto \psi_1^k(x)$, which must be at least as large as the optimal cost. The first equality in (B.10) is a direct consequence of (i). To prove the second equality in (B.10), we remark that $W_2^2(q_0,q_1) = \mathbb{E}_{x_0\sim q_0}\|\psi_1^*(x_0)-x_0\|^2$. Then, equation (B.4.4.2) readily implies that $|\mathbb{E}_{x_0\sim q_0^k}\|\psi_1^k(x_0)-x_0\|^2 - W_2^2(q_0,q_1)| \xrightarrow{k\to\infty} 0$. □

Lemma B.5. *Suppose that assumptions **(A1)** and **(A3)** hold. Let $\tilde{q}^k$ be the optimal transport plan between the optimal transport plans q and q^k. We have that*

$$\lim_{k\to\infty}\mathbb{E}_{t\sim\mathrm{U}(0,1),\boldsymbol{X}_0^k\overset{\mathrm{iid}}{\sim}q_0,\boldsymbol{X}_1^k\overset{\mathrm{iid}}{\sim}q_1,(x_0,x_1,x_0',x_1')\sim\tilde{q}^k}\left[\|u_t^*(tx_1+(1-t)x_0)-u_t^k(tx_1'+(1-t)x_0')\|^2\right]=0$$

Proof. For given instances of $\boldsymbol{X}_0^k$ and $\boldsymbol{X}_1^k$, we can write

$$
\begin{aligned}
&\mathbb{E}_{t\sim\mathrm{U}(0,1),(x_0,x_1,x_0',x_1')\sim\tilde{q}^k}\left[\|u_t^*(tx_1+(1-t)x_0)-u_t^k(tx_1'+(1-t)x_0')\|^2\right]\\
&=\mathbb{E}_{t\sim\mathrm{U}(0,1),(x_0,x_1,x_0',x_1')\sim\tilde{q}^k}\Big[\|\mathbb{E}_{\tilde{x}_0,\tilde{x}_1\sim q}[\tilde{x}_1-\tilde{x}_0|tx_1+(1-t)x_0=t\tilde{x}_1+(1-t)\tilde{x}_0]\\
&\qquad\qquad-\mathbb{E}_{\tilde{x}_0',\tilde{x}_1'\sim q^k}[\tilde{x}_1'-\tilde{x}_0'|tx_1'+(1-t)x_0'=t\tilde{x}_1'+(1-t)\tilde{x}_0']\|^2\Big]\\
&\le\mathbb{E}_{(x_0,x_1,x_0',x_1')\sim\tilde{q}^k}\left[\left\|x_1-x_0-(x_1'-x_0')\right\|^2\right]\le 2\mathbb{E}_{(x_0,x_1,x_0',x_1')\sim\tilde{q}^k}\left[\left\|x_1-x_1'\right\|^2+\left\|x_0-x_0'\right\|^2\right]\\
&=2\mathbb{E}_{(x_0,x_1,x_0',x_1')\sim\tilde{q}^k}\left[\left\|(x_0,x_1)-(x_0',x_1')\right\|^2\right]=2W_2^2(q,q^k)
\end{aligned}
$$

Assumption **(A3)** implies that almost surely, q^k converges to q weakly. For distributions on a bounded domain, weak convergence is equivalent to convergence in the Wasserstein distance [Villani 2008, Thm. 6.8], and this means that $W_2^2(q,q^k)\xrightarrow{k\to\infty}0$ almost surely. Almost sure convergence implies convergence in probability, which means that

$$
\Pr(W_2^2(q,q^k)>\epsilon)\xrightarrow{k\to\infty}0,\quad\forall\epsilon>0.
$$

Note that $W_2^2(q,q^k)$ is a bounded random variable because q and q^k have bounded support as q_0,q_1,q_0^k and q_1^k have bounded support. Suppose that $W_2^2(q,q^k)$ is bounded by the constant C. Hence, we can write

$$
\begin{aligned}
&\mathbb{E}_{\boldsymbol{X}_0^k,\boldsymbol{X}_1^k}\mathbb{E}_{t\sim\mathrm{U}(0,1),(x_0,x_1,x_0',x_1')\sim\tilde{q}^k}\left[\|u_t^*(tx_1+(1-t)x_0)-u_t^k(tx_1'+(1-t)x_0')\|^2\right]\\
&\le 2\mathbb{E}_{\boldsymbol{X}_0^k,\boldsymbol{X}_1^k}W_2^2(q,q^k)\le 2\big(\epsilon+C\Pr(W_2^2(q,q^k)>\epsilon)\big).
\end{aligned}
$$

We can take ϵ arbitrarily small, and for a given ϵ we can make the second term in the right-hand side arbitrarily small by taking k large enough. The final result follows. □

B.4.4.3 Proof of Proposition B.1

We have that almost surely, the empirical distributions q_0^k, resp. q_1^k, converge weakly to q_0, resp. q_1 [Varadarajan 1958]. Hence, we can apply Theorem B.6. Since convergence in distribution of random variables is equivalent to weak convergence of their laws, and the law of an optimal coupling is the optimal transport plan, we conclude that $(q_k)_{k\geq 0}$ converges weakly to q^*.

Theorem B.6 ([Cuesta-Albertos et al. 1997], Theorem 3.2)**.** *Let $(P_n)_n$, $(Q_n)_n$, P, Q be probability measures in $\mathcal{P}_2$ (the space of Borel probability measures with bounded second order moment) such that $P \ll \lambda_p$ (P is absolutely continuous with respect to the Lebesgue measure) and $P_n \xrightarrow{w} P$, $Q_n \xrightarrow{w} Q$, where $\xrightarrow{w}$ denotes weak convergence of probability measures. Let (X_n, Y_n) be an optimal coupling between P_n and Q_n, $n \in \mathbb{N}$, and (X, Y) an optimal coupling between P and Q. Then, $(X_n, Y_n) \xrightarrow{\mathcal{L}} (X, Y)$, where $\xrightarrow{\mathcal{L}}$ denotes convergence of random variables in distribution.*

B.4.5 Bounds on the transport cost and monotone convergence results

The following result shows that for an arbitrary joint distribution $q(x_0, x_1)$, we can upper-bound the transport cost associated to the marginal vector field u_t to a quantity that depends only $q(x_0, x_1)$.

Proposition B.7. *For an arbitrary joint distribution $q(x_0, x_1)$ with marginals $q_0(x_0)$ and $q_1(x_1)$, let ψ_t be the flow corresponding to the marginal vector field u_t. We have that*

$$\mathbb{E}_{q_0(x_0)} \|\psi_1(x_0) - x_0\|^2 \leq \mathbb{E}_{q(x_0,x_1)} \|x_1 - x_0\|^2, \tag{B.14}$$

Proof. We make use of the notation introduced in App. B.4.4. We will rely on the fact that for all $t \in [0, 1]$, the random variable $tx_1 + (1-t)x_0$, with $(x_0, x_1) \sim q$ has the same distribution as the random variable $\psi_t(x_0)$, with $x_0 \sim q_0$. This is a direct consequence of Lemma 3.1. Using that ψ_t

is the flow for u_t and Jensen's inequality twice, we have that

$$
\begin{aligned}
&\mathbb{E}_{x_0\sim q_0}\|\psi_1(x_0) - x_0\|^2 \\
&= \mathbb{E}_{x_0\sim q_0}\left\|\int_0^1 u_s(\psi_s(x_0))\,ds\right\|^2 \leq \mathbb{E}_{t\sim \mathrm{U}(0,1),x_0\sim q_0}\|u_t(\psi_t(x_0))\|^2 \\
&= \mathbb{E}_{t\sim \mathrm{U}(0,1),(x_0,x_1)\sim q}\|u_t(tx_1 + (1-t)x_0)\|^2 \\
&= \mathbb{E}_{t\sim \mathrm{U}(0,1),(x_0,x_1)\sim q}\|\mathbb{E}_{(x_0',x_1')\sim q}\left[u_t(tx_1 + (1-t)x_0|x_0', x_1')|tx_1 + (1-t)x_0 = tx_1' + (1-t)x_0'\right]\|^2 \\
&\leq \mathbb{E}_{t\sim \mathrm{U}(0,1),(x_0,x_1)\sim q}\mathbb{E}_{(x_0',x_1')\sim q}\left[\|u_t(tx_1 + (1-t)x_0|x_0', x_1')\|^2|tx_1 + (1-t)x_0 = tx_1' + (1-t)x_0'\right] \\
&= \mathbb{E}_{t\sim \mathrm{U}(0,1),(x_0,x_1)\sim q}\mathbb{E}_{(x_0',x_1')\sim q}\left[\|x_1' - x_0'\|^2|tx_1 + (1-t)x_0 = tx_1' + (1-t)x_0'\right] \\
&= \mathbb{E}_{t\sim \mathrm{U}(0,1),(x_0,x_1)\sim q}\|x_1 - x_0\|^2
\end{aligned}
$$

as needed. □

Note that that the statement and proof of this proposition is equivalent to Theorem 3.5 of [Liu et al. 2022], although the language and notation that we use is different, which is why we though convenient to include it.

For the case of BatchOT, the following theorem shows that the quantity in the upper bound of (B.14) is monotonically decreasing in k. The combination of Proposition B.7 and Theorem B.8 provides a weak guarantee that for BatchOT, the transport cost should not get much higher when k increases.

Theorem B.8. *Suppose that Multisample Flow Matching is run with BatchOT. For clarity, we make the dependency on the sample size k explicit and let*[1] $q^{(k)}(x_0, x_1) := q(x_0, x_1)$, *and* $\psi_t^k(x_0) := \psi_t(x_0)$.

[1]Note that here $q^{(k)} := q$ is a marginalized distribution and is different from q^k defined in Step 3.

Then, for any $k \geq 1$, we have that

$$\mathbb{E}_{q_0(x_0)} \|\psi_1^k(x_0) - x_0\|^2 \leq \mathbb{E}_{q^{(k)}(x_0,x_1)} \|x_1 - x_0\|^2,$$
$$\mathbb{E}_{q^{(k+1)}(x_0,x_1)} \|x_1 - x_0\|^2 \leq \mathbb{E}_{q^{(k)}(x_0,x_1)} \|x_1 - x_0\|^2.$$

Proof. We write

$$\mathbb{E}_{t\sim \mathrm{U}(0,1), X_0^{k+1}, X_1^{k+1}} \mathbb{E}_{(x_0,x_1)\sim q^{k+1}} \|x_1 - x_0\|^2 = \frac{1}{k} \mathbb{E}_{t\sim \mathrm{U}(0,1), X_0^{k+1}, X_1^{k+1}} \left[\sum_{i=1}^{k} \|x_1^{(i)} - x_0^{(\sigma_{k+1}(i))}\|^2 \right]$$
$$= \frac{1}{k}\frac{1}{k+1} \mathbb{E}_{t\sim \mathrm{U}(0,1), X_0^{k+1}, X_1^{k+1}} \left[\sum_{j=1}^{k+1} \sum_{i\in[k+1]\setminus\{j\}} \|x_1^{(i)} - x_0^{(\sigma_{k+1}(i))}\|^2 \right]$$
$$\leq \frac{1}{k}\frac{1}{k+1} \mathbb{E}_{t\sim \mathrm{U}(0,1), X_0^{k+1}, X_1^{k+1}} \left[\sum_{j=1}^{k+1} \sum_{i\in[k+1]\setminus\{j\}} \|x_1^{(i)} - x_0^{(\sigma_k^{-j}(i))}\|^2 \right]$$
$$= \mathbb{E}_{t\sim \mathrm{U}(0,1), X_0^{k}, X_1^{k}} \left[\frac{1}{k} \sum_{j=1}^{k} \|x_1^{(i)} - x_0^{(\sigma_k(i))}\|^2 \right] = \mathbb{E}_{t\sim \mathrm{U}(0,1), X_0^{k}, X_1^{k}} \mathbb{E}_{(x_0,x_1)\sim q^k} \|x_1 - x_0\|^2.$$

In the first equality, we used that the optimal transport map between the empirical distributions q_0^k and q_1^k can be encoded as a permutation, which we denote by σ_{k+1}. In the inequality, we introduced the notation σ_k^{-j} to denote the optimal permutation within $\{x_0^{(i)}\}_{i\in[k+1]\setminus\{j\}}$. The inequality holds because using the optimality of σ_{k+1}:

$$\sum_{j=1}^{k+1} \sum_{i\in[k+1]\setminus\{j\}} \|x_1^{(i)} - x_0^{(\sigma_{k+1}(i))}\|^2 \leq \sum_{j=1}^{k+1} \sum_{i\in[k+1]} \|x_1^{(i)} - x_0^{(\sigma_{k+1}(i))}\|^2$$
$$\leq \sum_{j=1}^{k+1} \left(\sum_{i\in[k+1]\setminus\{j\}} \|x_1^{(i)} - x_0^{(\sigma_k^{-j}(i))}\|^2 + \|x_1^{(j)} - x_0^{(j)}\|^2 \right) \leq \sum_{j=1}^{k+1} \sum_{i\in[k+1]\setminus\{j\}} \|x_1^{(i)} - x_0^{(\sigma_k^{-j}(i))}\|^2.$$

□

B.5 Experimental & evaluation details

	ImageNet-32	ImageNet-64
Channels	256	192
Depth	3	3
Channels multiple	1,2,2,2	1,2,3,4
Heads	4	4
Heads Channels	64	64
Attention resolution	4	8
Dropout	0.0	0.1
Batch size / GPU	256	50
GPUs	4	16
Effective Batch size	1024	800
Epochs	350	575
Effective Iterations	438k	957k
Learning Rate	1e-4	1e-4
Learning Rate Scheduler	Polynomial Decay	Constant
Warmup Steps	20k	-

Table B.6: Hyper-parameters used for training each model.

B.5.1 Image datasets

We report the hyper-parameters used in Table B.6. We use the architecture from Dhariwal and Nichol [2021] but with much lower attention resolution. We use full 32 bit-precision for training ImageNet-32 and 16-bit mixed precision for training ImageNet-64. All models are trained using the Adam optimizer with the following parameters: $\beta_1 = 0.9$, $\beta_2 = 0.999$, weight decay = 0.0, and $\epsilon = 1e{-}8$. All methods we trained using identical architectures, with the same parameters for the the same number of epochs (see Table B.6 for details), with the exception of Rectified Flow, which we trained for much longer starting from the fully trained CondOT model. We use either a constant learning rate schedule or a polynomial decay schedule (see Table B.6). The polynomial decay learning rate schedule includes a warm-up phase for a specified number of training steps.

In the warm-up phase, the learning rate is linearly increased from $1e{-}8$ to the peak learning rate (specified in Table B.6). Once the peak learning rate is achieved, it linearly decays the learning rate down to $1e{-}8$ until the final training step.

When reporting negative log-likelihood, we dequantize using the standard uniform dequantization [Dinh et al. 2016]. We report an importance-weighted estimate using

$$\text{BPD}(K) = -\frac{1}{D}\log_2 \frac{1}{K}\sum_{k=1}^{K} p_t(x + u_k), \text{ where } u_k \sim [U(0,1)]^D,$$

with x is in $\{0, \ldots, 255\}^D$. We solve for p_t at exactly $t = 1$ with an adaptive step size solver `dopri5` with `atol=rtol=1e-5` using the `torchdiffeq` [Chen 2018] library. We used K=15 for ImageNet32 and K=10 for ImageNet64.

When computing FID, we use the TensorFlow-GAN library `https://github.com/tensorflow/gan`.

We run coupling algorithms only within each GPU. We also ran coupling algorithms across all GPUs (using the "Effective Batch Size") in preliminary experiments, but did not see noticeable gains in sample efficiency while obtaining slightly worse performance and sample quality, so we stuck to the smaller batch sizes for running our coupling algorithms.

For Rectified Flow, we use the finalized FM-CondOT model, generate 50000 noise and sample pairs, then train using the same FM-CondOT algorithm and hyperparameters on these sampled pairs. This is equivalent to their 2-Rectified Flow approach [Liu et al. 2022]. For the rectification process, we train for 300 epochs.

B.5.2 Improved batch optimal couplings

Datasets. We experimented with 3 datasets in dimensions $\{2, 32, 64\}$ consisting of $50K$ samples. Both q_0 and q_1 were Gaussian mixtures with number of centers described in Table B.7.

Neural Networks Architectures. For B-ST we used stacked blocks of Convex Potential Flows

[Huang et al. 2020] as an invertible neural network parametrizing the map, which also allowed us to estimate KL divergence:

$$\mathrm{KL}(q_1 || (\psi_1)_\sharp\, q_0) = \mathbb{E}_{x\sim q_1}\left[\log q_1(x) - \log\left((\psi_1)_\sharp q_0\right)(x)\right].$$

For B-FM we used a simple MLP with Swish activation. For each dataset we built architectures with roughly the same number of parameters.

Hyperparameter Search. For each dataset and each cost we swept over learning rates $\{0.005, 0.001, 0.0005\}$ and chose the best setting.

	2-D	32-D	64-D
q_0 #centers	1	50	100
q_1 #centers	8	50	100
#params	50K	800K	800K
batch size	128	1024	1024
epochs	100	1000	1000

Table B.7: Hyperparameters for experiments on synthetic datasets.

C | Appendix: Stochastic Optimal Control Matching

C.1 Technical assumptions

Throughout our work, we make the same assumptions as [Nüsken and Richter 2021], which are needed for all the objects considered to be well-defined. Namely, we assume that:

(i) The set $\mathcal{U}$ of *admissible controls* is given by

$$\mathcal{U} = \{u \in C^1(\mathbb{R}^d \times [0,T]; \mathbb{R}^d) \mid \exists C > 0,\ \forall (x,s) \in \mathbb{R}^d \times [0,T],\ b(x,s) \leq C(1+|x|)\}.$$

(ii) The coefficients b and σ are continuously differentiable, σ has bounded first-order spatial derivatives, and $(\sigma\sigma^\top)(x,s)$ is positive definite for all $(x,s) \in \mathbb{R}^d[0,T]$. Furthermore, there exist constants $C, c_1, c_2 > 0$ such that

$$\begin{aligned} \|b(x,s)\| &\leq C(1+\|x\|), && \text{(linear growth)} \\ c_1\|\xi\|^2 \leq \xi^\top(\sigma\sigma^\top)(x,s)\xi &\leq c_2\|\xi\|^2, && \text{(ellipticity)} \end{aligned}$$

for all $(x,s) \in \mathbb{R}^d \times [0,T]$ and $\xi \in \mathbb{R}^d$.

C.2 Proofs of section 4.2

Proof of (4.5) By Itô's lemma, we have that

$$V(X_T^u, T) - V(X_t^u, t) = \int_t^T \big(\partial_s V(X_s^u, s) + \langle b(X_s^u, s) + \sigma(X_s^u, s)u(X_s^u, s), \nabla V(X_s^u, s)\rangle + \frac{\lambda}{2}\sum_{i,j=1}^{d} (\sigma\sigma^\top)_{ij}(X_s^u, s)\partial_{x_i}\partial_{x_j} V(X_s^u, s)\big)\,\mathrm{d}s + S_t^u,$$

where $S_t^u = \sqrt{\lambda}\int_t^T \nabla V(X_s^u, s)^\top \sigma(X_s^u, s)\,\mathrm{d}B_s$. Note that by (4.4),

$$\begin{aligned}
&\partial_s V(X_s^u, s) + \langle b(X_s^u, s) + \sigma(X_s^u, s)u(X_s^u, s), \nabla V(X_s^u, s)\rangle \\
&\qquad + \frac{\lambda}{2}\sum_{i,j=1}^{d} (\sigma\sigma^\top)_{ij}(X_s^u, s)\partial_{x_i}\partial_{x_j} V(X_s^u, s) \\
&= \frac{1}{2}\|(\sigma^\top \nabla V)(X_s^u, s)\|^2 - f(X_s^u, s) + \langle \sigma(X_s^u, s)u(X_s^u, s), \nabla V(X_s^u, s)\rangle \\
&= \frac{1}{2}\|(\sigma^\top \nabla V)(X_s^u, s) + u(X_s^u, s)\|^2 - \frac{1}{2}\|u(X_s^u, s)\|^2 - f(X_s^u, s),
\end{aligned}$$

and this implies that

$$g(X_T^u) - V(X_t^u, t) = \int_t^T \big(\frac{1}{2}\|(\sigma^\top \nabla V)(X_s^u, s) + u(X_s^u, s)\|^2 - \frac{1}{2}\|u(X_s^u, s)\|^2 - f(X_s^u, s)\big)\,\mathrm{d}s + S_t^u \quad \text{(C.1)}$$

Since $\mathbb{E}[S_t^u \mid X_t^u = x] = 0$, rearranging (C.1) and taking the conditional expectation with respect to X_t^u yields the final result.

Proof of (4.6)-(4.7) By Itô's lemma, we have that

$$dV(X_s,s) = \big(\partial_s V(X_s,s) + \langle b(X_s,s), \nabla V(X_s,s)\rangle + \frac{\lambda}{2}\sum_{i,j=1}^{d}(\sigma\sigma^\top)_{ij}(X_s,s)\partial_{x_i}\partial_{x_j}V(X_s,s)\big)\,\mathrm{d}s + \sqrt{\lambda}\nabla V(X_s^u,s)^\top\sigma(X_s^u,s)\,\mathrm{d}B_s, \quad \text{(C.2)}$$

Note that by (4.4),

$$\partial_s V(X_s,s) + \langle b(X_s,s), \nabla V(X_s,s)\rangle + \frac{\lambda}{2}\sum_{i,j=1}^{d}(\sigma\sigma^\top)_{ij}(X_s,s)\partial_{x_i}\partial_{x_j}V(X_s,s)$$
$$= \frac{1}{2}\|(\sigma^\top\nabla V)(X_s,s)\|^2 - f(X_s,s).$$

Plugging this into (C.2) concludes the proof.

Proof of (4.8) Since $Y_s = V(X_s,s)$ and $Z_s = \sigma^\top(s)\nabla V(X_s,s) = -u^*(X_s,s)$ satisfy (4.7), we have that

$$g(X_T) = Y_T = Y_t - \int_t^T (f(X_s,s) - \frac{1}{2}\|u^*(X_s,s)\|^2)\,\mathrm{d}s - \sqrt{\lambda}\int_t^T \langle u^*(X_s,s), \mathrm{d}B_s\rangle.$$

Hence, recalling the definition of the work functional in (4.10), we have that

$$\mathcal{W}(X,t) = Y_t + \frac{1}{2}\int_t^T \|u^*(X_s,s)\|^2\,\mathrm{d}s - \sqrt{\lambda}\int_t^T \langle u^*(X_s,s), \mathrm{d}B_s\rangle. \quad \text{(C.3)}$$

By Novikov's theorem (Thm. C.1), we have that

$$\mathbb{E}[\exp(-\lambda^{-1}\mathcal{W}(X,t))|X_t]$$
$$= e^{-\lambda^{-1}Y_t}\mathbb{E}\big[\exp\big(\lambda^{-1/2}\int_t^T \langle u^*(X_s,s), \mathrm{d}B_s\rangle - \frac{\lambda^{-1}}{2}\int_t^T \|u^*(X_s,s)\|^2\,\mathrm{d}s\big)\big|X_t\big] = e^{-\lambda^{-1}Y_t},$$

which concludes the proof of (4.8).

Theorem C.1 (Novikov's theorem). *Let θ_s be a locally-$\mathcal{H}_2$ process which is adapted to the natural filtration of the Brownian motion $(B_t)_{t\geq 0}$. Define*

$$Z(t) = \exp\Big(\int_0^t \theta_s \, \mathrm{d}B_s - \frac{1}{2}\int_0^t \|\theta_s\|^2 \, \mathrm{d}s\Big). \tag{C.4}$$

If for each $t \geq 0$,

$$\mathbb{E}\Big[\exp\Big(\int_0^t \|\theta_s\|^2 \, \mathrm{d}s\Big)\Big] < +\infty,$$

then for each $t \geq 0$,

$$\mathbb{E}[Z(t)] = 1. \tag{C.5}$$

Moreover, the process $Z(t)$ is a positive martingale, i.e. if $(\mathcal{F}_t)_{t\geq 0}$ is the filtration associated to the Brownian motion $(B_t)_{t\geq 0}$, then for $t \geq s$, $\mathbb{E}[Z_t|\mathcal{F}_s] = Z_s$.

Theorem C.2 (Girsanov theorem). *Let $W = (W_t)_{t\in[0,T]}$ be a standard Wiener process, and let $\mathbb{P}$ be its induced probability measure over $C([0,T];\mathbb{R}^d)$, known as the Wiener measure. Let $Z(t)$ be as defined in (C.4) and suppose that the assumptions of Theorem C.1 hold. Let $(\Omega, \mathcal{F})$ be the σ-algebra associated to B_T. For any $F \in \mathcal{F}$, define the measure*

$$\mathbb{Q}(F) = \mathbb{E}_{\mathbb{P}}[Z(T)\mathbf{1}_F] \ .$$

$\mathbb{Q}$ is a probability measure because of (C.5). Under the probability measure $\mathbb{Q}$, the stochastic process $\{\tilde{W}(t)\}_{0\leq t\leq T}$ defined as

$$\tilde{W}(t) = W(t) - \int_0^t \theta_s \, \mathrm{d}s$$

is a standard Wiener process. That is, for any $n \geq 0$ and any $0 = t_0 < t_1 < \cdots < t_n$, the increments $\{\tilde{W}(t_{i+1}) - \tilde{W}(t_i)\}_{i=0}^{n-1}$ are independent and $\mathbb{Q}$-Gaussian distributed with mean zero and covariance $(t_{i+1} - t_i)\mathrm{I}$, which means that for any $\alpha \in \mathbb{R}^d$, the moment generating function of $\tilde{W}(t_{i+1}) - \tilde{W}(t_i)$ with respect to $\mathbb{Q}$ is as follows:

$$\begin{aligned}&\mathbb{E}_{\mathbb{Q}}[\exp(\langle \alpha, \tilde{W}(t_{i+1}) - \tilde{W}(t_i)\rangle)] \\ &:= \mathbb{E}_{\mathbb{P}}\Big[\exp\Big(\Big\langle \alpha, W(t_{i+1}) - \int_0^{t_{i+1}} \theta_s \,\mathrm{d}s - W(t_i) + \int_0^{t_i} \theta_s \,\mathrm{d}s\Big\rangle\Big) Z(T)\Big] = \exp\Big(\frac{(t_{i+1} - t_i)\|\alpha\|^2}{2}\Big).\end{aligned}$$

Corollary C.3 (Girsanov theorem for SDEs)**.** *If the two SDEs*

$$\begin{aligned}\mathrm{d}X_t &= b_1(X_t, t)\,\mathrm{d}t + \sigma(X_t, t)\,\mathrm{d}B_t, \qquad X_0 = x_{\text{init}} \\ dY_t &= (b_1(Y_t, t) + b_2(Y_t, t))\,\mathrm{d}t + \sigma(Y_t, t)\,\mathrm{d}B_t, \qquad Y_0 = x_{\text{init}}\end{aligned}$$

admit unique strong solutions on $[0, T]$, then for any bounded continuous functional Φ on $C([0, T])$, we have that

$$\begin{aligned}\mathbb{E}[\Phi(X)] &= \mathbb{E}\Big[\Phi(Y)\exp\Big(-\int_0^T \sigma(Y_t, t)^{-1} b_2(Y_t, t)\,\mathrm{d}B_t - \frac{1}{2}\int_0^T \|\sigma(Y_t, t)^{-1} b_2(Y_t, t)\|^2\,\mathrm{d}t\Big)\Big] \\ &= \mathbb{E}\Big[\Phi(Y)\exp\Big(-\int_0^T \sigma(Y_t, t)^{-1} b_2(Y_t, t)\,d\tilde{B}_t + \frac{1}{2}\int_0^T \|\sigma(Y_t, t)^{-1} b_2(Y_t, t)\|^2\,\mathrm{d}t\Big)\Big],\end{aligned}$$

where $\tilde{B}_t = B_t + \int_0^t \sigma(Y_s, s)^{-1} b_2(Y_s, s)\,\mathrm{d}s$. More generally, b_1 and b_2 can be random processes that are adapted to filtration of B.

Lemma C.4. *For an arbitrary $v \in \mathcal{U}$, let $\mathbb{P}^v$ and $\mathbb{P}$ be respectively the laws of the SDEs*

$$\begin{aligned}\mathrm{d}X_t^v &= (b(X_t^v, t) + \sigma(t)v(X_t^v, t))\,\mathrm{d}t + \sqrt{\lambda}\sigma(t)\mathrm{d}B_t, \qquad X_0^v \sim p_0, \\ \mathrm{d}X_t &= b(X_t, t)\,\mathrm{d}t + \sqrt{\lambda}\sigma(t)\mathrm{d}B_t, \qquad X_0 \sim p_0.\end{aligned}$$

We have that

$$\frac{d\mathbb{P}}{d\mathbb{P}^v}(X^v) = \exp\Big(-\lambda^{-1/2}\int_0^T \langle v(X_t^v,t), \mathrm{d}B_t^v\rangle + \frac{\lambda^{-1}}{2}\int_0^T \|v(X_t^v,t)\|^2\,\mathrm{d}t\Big) \tag{C.6}$$

$$= \exp\Big(-\lambda^{-1/2}\int_0^T \langle v(X_t^v,t), \mathrm{d}B_t\rangle - \frac{\lambda^{-1}}{2}\int_0^T \|v(X_t^v,t)\|^2\,\mathrm{d}t\Big),$$

$$\frac{d\mathbb{P}^v}{d\mathbb{P}}(X) = \exp\big(\lambda^{-1/2}\int_0^T \langle v(X_t,t), \mathrm{d}B_t\rangle - \frac{\lambda^{-1}}{2}\int_0^T \|v(X_t,t)\|^2\,\mathrm{d}t\big). \tag{C.7}$$

where $B_t^v := B_t - \lambda^{-1/2}\int_0^t v(X_s^v,s)\,\mathrm{d}s$. *For the optimal control* u^*, *we have that*

$$\frac{d\mathbb{P}}{d\mathbb{P}^{u^*}}(X^{u^*}) = \exp\big(\lambda^{-1}\big(-V(X_0^{u^*},0) + \mathcal{W}(X^{u^*},0)\big)\big), \tag{C.8}$$

$$\frac{d\mathbb{P}^{u^*}}{d\mathbb{P}}(X) = \exp\big(\lambda^{-1}\big(V(X_0,0) - \mathcal{W}(X,0)\big)\big), \tag{C.9}$$

where the functional $\mathcal{W}$ *is defined in* (4.10).

Proof. The proof of (C.6)-(C.7) follows directly from Theorem C.3. To prove (C.9), we use that by (C.3),

$$\mathcal{W}(X,0) = V(X_0,0) + \frac{1}{2}\int_0^T \|u^*(X_s,s)\|^2\,\mathrm{d}s - \sqrt{\lambda}\int_0^T \langle u^*(X_s,s), \mathrm{d}B_s\rangle, \tag{C.10}$$

which implies that

$$\frac{d\mathbb{P}^{u^*}}{d\mathbb{P}}(X) = \exp\big(\lambda^{-1/2}\int_0^T \langle u^*(X_t,t), \mathrm{d}B_t\rangle - \frac{\lambda^{-1}}{2}\int_0^T \|u^*(X_t,t)\|^2\,\mathrm{d}t\big)$$

$$= \exp\big(\lambda^{-1}\big(V(X_0,0) - \mathcal{W}(X,0)\big)\big).$$

To prove (C.8), we use that since $\mathrm{d}X_t^{u^*} = b(X_t^{u^*},t)\,\mathrm{d}t + \sqrt{\lambda}\sigma(t)\mathrm{d}B_t^{u^*}$, equation (C.10) holds if we

replace X and B by X^{u^*} and B^{u^*}, which reads

$$\mathcal{W}(X^{u^*},0) = V(X_0^{u^*},0) + \frac{1}{2}\int_t^T \|u^*(X_s^{u^*},s)\|^2\,\mathrm{d}s - \sqrt{\lambda}\int_t^T \langle u^*(X_s^{u^*},s),\mathrm{d}B_s^v\rangle.$$

Hence,

$$\begin{aligned}\frac{d\mathbb{P}}{d\mathbb{P}^{u^*}}(X^{u^*}) &= \exp\big(-\lambda^{-1/2}\int_0^T \langle u^*(X_t^{u^*},t),\mathrm{d}B_t^{u^*}\rangle + \frac{\lambda^{-1}}{2}\int_0^T \|u^*(X_t^{u^*},t)\|^2\,\mathrm{d}t\big)\\ &= \exp\big(\lambda^{-1}\big(-V(X_0^{u^*},0) + \mathcal{W}(X^{u^*},0)\big)\big).\end{aligned}$$

□

Lemma C.5. *The following expression holds:*

$$\mathbb{E}_{\mathbb{P}^u}\Big[\log\frac{d\mathbb{P}^u}{d\mathbb{P}^{u^*}}\Big] = \lambda^{-1}\mathbb{E}\Big[\int_0^T \Big(\frac{1}{2}\|u(X_t^u,t)\|^2 + f(X_t^u,t)\Big)\,\mathrm{d}t + g(X_T^u) - V(X_0^u,0)\Big], \tag{C.11}$$

Proof. To prove (C.11), we write

$$\begin{aligned}\log\frac{d\mathbb{P}^{u^*}}{d\mathbb{P}^u}(X^u) &= \log\Big(\frac{d\mathbb{P}^{u^*}}{d\mathbb{P}}(X^u)\frac{d\mathbb{P}}{d\mathbb{P}^u}(X^u)\Big) = \log\frac{d\mathbb{P}^{u^*}}{d\mathbb{P}}(X^u) + \log\frac{d\mathbb{P}}{d\mathbb{P}^u}(X^u)\\ &= \lambda^{-1}\big(V(X_0^u,0) - \int_0^T f(X_t^u,t)\,\mathrm{d}t - g(X_T^u)\big)\\ &\quad - \lambda^{-1/2}\int_0^T \langle u(X_t^u,t),\mathrm{d}B_t\rangle - \frac{\lambda^{-1}}{2}\int_0^T \|u(X_t^u,t)\|^2\,\mathrm{d}t\,.\end{aligned}$$

Since $\mathbb{E}_{\mathbb{P}^u}\big[\log\frac{d\mathbb{P}^u}{d\mathbb{P}^{u^*}}\big] = -\mathbb{E}_{\mathbb{P}^u}\big[\log\frac{d\mathbb{P}^{u^*}}{d\mathbb{P}^u}\big]$, and $\mathbb{E}_{\mathbb{P}^u}\big[\int_0^T \langle u(X_t^u,t),\mathrm{d}B_t\rangle\big] = 0$, the result follows. □

Proposition C.6. *(i) The following two expressions hold for arbitrary controls u, v in the class $\mathcal{U}$ of*

admissible controls:

$$\tilde{\mathcal{L}}_{\mathrm{CE}}(u) = \mathbb{E}_{\mathbb{P}^{u^*}}\Big[\log\frac{d\mathbb{P}^{u^*}}{d\mathbb{P}^{u}}\Big] = \mathbb{E}\Big[\Big(-\lambda^{-1/2}\int_0^T \langle u(X_t^v,t), \mathrm{d}B_t\rangle - \lambda^{-1}\int_0^T \langle u(X_t^v,t), v(X_t^v,t)\rangle\,\mathrm{d}t \tag{C.12}$$
$$+\frac{\lambda^{-1}}{2}\int_0^T \|u(X_t^v,t)\|^2\,\mathrm{d}t + \lambda^{-1}\big(V(X_0^v,0) - \mathcal{W}(X^v,0)\big)\Big)$$
$$\times \exp\big(\lambda^{-1}\big(V(X_0^v,0) - \mathcal{W}(X^v,0)\big)$$
$$-\lambda^{-1/2}\int_0^T \langle v(X_t^v,t), \mathrm{d}B_t\rangle - \frac{\lambda^{-1}}{2}\int_0^T \|v(X_t^v,t)\|^2\,\mathrm{d}t\big)\Big],$$

$$\tilde{\mathcal{L}}_{\mathrm{CE}}(u) = \frac{\lambda^{-1}}{2}\mathbb{E}\Big[\int_0^T \|u^*(X_t^{u^*},t) - u(X_t^{u^*},t)\|^2\,\mathrm{d}t\Big]. \tag{C.13}$$

When p_0 is concentrated at a single point x_{init}, the terms $V(x_{\mathrm{init}},0)$ are constant and can be removed without modifying the landscape. In other words, $\tilde{\mathcal{L}}_{\mathrm{CE}}$ and $\mathcal{L}_{\mathrm{CE}}$ are equal up to constant terms and constant factors.

(ii) When p_0 is a generic probability measure, $\tilde{\mathcal{L}}_{\mathrm{CE}}$ and $\mathcal{L}_{\mathrm{CE}}$ have different landscapes, and $\mathcal{L}_{\mathrm{CE}}(u) = \mathbb{E}_{\mathbb{P}^{u^*}}\big[\log\frac{d\mathbb{P}^{u^*}}{d\mathbb{P}^{u}}\exp\big(-\lambda^{-1}V(X_0^{u^*},0)\big)\big]$. u^* is still the only minimizer of the loss $\mathcal{L}_{\mathrm{CE}}$, and for some constant K, we have that

$$\mathcal{L}_{\mathrm{CE}}(u,0) = \frac{\lambda^{-1}}{2}\mathbb{E}\Big[\int_0^T \|u^*(X_t^{u^*},t) - u(X_t^{u^*},t)\|^2\,\mathrm{d}t\exp\big(-\lambda^{-1}V(X_0^{u^*},0)\big)\Big] + K. \tag{C.14}$$

Proof. We begin with the proof of (i), and prove (C.12) first. Note that by the Girsanov theorem (Theorem C.2),

$$\mathbb{E}_{\mathbb{P}^{u^*}}\Big[\log\frac{d\mathbb{P}^{u^*}}{d\mathbb{P}^{u}}(X^{u^*})\Big] = -\mathbb{E}_{\mathbb{P}^{u^*}}\Big[\log\frac{d\mathbb{P}^{u}}{d\mathbb{P}^{u^*}}(X^{u^*})\Big] = -\mathbb{E}_{\mathbb{P}^{u^*}}\Big[\log\frac{d\mathbb{P}^{u}}{d\mathbb{P}}(X^{u^*}) + \log\frac{d\mathbb{P}}{d\mathbb{P}^{u^*}}(X^{u^*})\Big]$$
$$= -\mathbb{E}_{\mathbb{P}^{v}}\Big[\Big(\log\frac{d\mathbb{P}^{u}}{d\mathbb{P}}(X^{v}) + \log\frac{d\mathbb{P}}{d\mathbb{P}^{u^*}}(X^{v})\Big)\frac{d\mathbb{P}^{u^*}}{d\mathbb{P}}(X^{v})\frac{d\mathbb{P}}{d\mathbb{P}^{v}}(X^{v})\Big] \tag{C.15}$$

Note that by equations (C.7) and (C.9),

$$\begin{aligned}
\log \frac{d\mathbb{P}^u}{d\mathbb{P}}(X^v) &= \lambda^{-1/2} \int_0^T \langle u(X_t^v, t), \mathrm{d}B_t^v \rangle - \frac{\lambda^{-1}}{2} \int_0^T \|u(X_t^v, t)\|^2 \,\mathrm{d}t, \\
&= \lambda^{-1/2} \int_0^T \langle u(X_t^v, t), \mathrm{d}B_t \rangle + \lambda^{-1} \int_0^T \langle u(X_t^v, t), v(X_t^v, t) \rangle \,\mathrm{d}t - \frac{\lambda^{-1}}{2} \int_0^T \|u(X_t^v, t)\|^2 \,\mathrm{d}t, \\
\log \frac{d\mathbb{P}}{d\mathbb{P}^{u^*}}(X^v) &= \lambda^{-1} \big(- V(X_0^v, 0) + \mathcal{W}(X^v, 0) \big). \qquad \text{(C.16)}
\end{aligned}$$

where $B_t^v := B_t + \lambda^{-1/2} \int_0^t v(X_s^v, s) \,\mathrm{d}s$. Also,

$$\begin{aligned}
\frac{d\mathbb{P}^{u^*}}{d\mathbb{P}}(X^v) &= \exp\big(\lambda^{-1}\big(V(X_0^v, 0) - \mathcal{W}(X^v, 0)\big)\big), \\
\frac{d\mathbb{P}}{d\mathbb{P}^v}(X^v) &= \exp\Big(- \lambda^{-1/2} \int_0^T \langle v(X_t^v, t), \mathrm{d}B_t \rangle - \frac{\lambda^{-1}}{2} \int_0^T \|v(X_t^v, t)\|^2 \,\mathrm{d}t \Big). \qquad \text{(C.17)}
\end{aligned}$$

If we plug (C.16) and (C.17) into the right-hand side of (C.15), we obtain

$$\begin{aligned}
&\mathbb{E}_{\mathbb{P}^{u^*}}\Big[\log \frac{d\mathbb{P}^{u^*}}{d\mathbb{P}^u}(X^{u^*}) \Big] = -\mathbb{E}_{\mathbb{P}^{u^*}}\Big[(\log \frac{d\mathbb{P}^u}{d\mathbb{P}}(X^v) + \log \frac{d\mathbb{P}}{d\mathbb{P}^{u^*}}(X^v)) \frac{d\mathbb{P}^{u^*}}{d\mathbb{P}}(X^v) \frac{d\mathbb{P}}{d\mathbb{P}^u}(X^v) \Big] \\
&= -\mathbb{E}\Big[\big(\lambda^{-1/2} \int_0^T \langle u(X_t^v, t), \mathrm{d}B_t \rangle + \lambda^{-1} \int_0^T \langle u(X_t^v, t), v(X_t^v, t) \rangle \,\mathrm{d}t \\
&\qquad - \frac{\lambda^{-1}}{2} \int_0^T \|u(X_t^v, t)\|^2 \,\mathrm{d}t + \lambda^{-1}\big(- V(X_0^v, 0) + \mathcal{W}(X^v, 0) \big) \big) \\
&\qquad \times \exp\big(\lambda^{-1}\big(V(X_0^v, 0) - \mathcal{W}(X^v, 0)\big) - \lambda^{-1/2} \int_0^T \langle v(X_t^v, t), \mathrm{d}B_t \rangle - \frac{\lambda^{-1}}{2} \int_0^T \|v(X_t^v, t)\|^2 \,\mathrm{d}t \big) \Big],
\end{aligned}$$

which concludes the proof.

To show (C.13), we use that by Theorem C.3,

$$\frac{d\mathbb{P}^u}{d\mathbb{P}^{u^*}}(X^{u^*}) = \exp\Big(-\lambda^{-1/2} \int_0^T \langle u^*(X_t^{u^*}, t) - u(X_t^{u^*}, t), \mathrm{d}B_t \rangle - \frac{\lambda^{-1}}{2} \int_0^T \|u^*(X_t^{u^*}, t) - u(X_t^{u^*}, t)\|^2 \,\mathrm{d}t \Big).$$

Hence,

$$\mathbb{E}_{\mathbb{P}^{u^*}}\Big[\log\frac{d\mathbb{P}^{u^*}}{d\mathbb{P}^{u}}\Big] = -\mathbb{E}_{\mathbb{P}^{u^*}}\Big[\log\frac{d\mathbb{P}^{u}}{d\mathbb{P}^{u^*}}\Big] = \frac{\lambda^{-1}}{2}\mathbb{E}\Big[\int_0^T \|u^*(X_t^{u^*},t) - u(X_t^{u^*},t)\|^2\,\mathrm{d}t\Big].$$

Next, we prove (ii). The first instance of $V(X_0^v, 0)$ in (C.12) can be removed without modifying the landscape of the loss. Hence, we are left with

$$\begin{aligned}\bar{\mathcal{L}}_{\mathrm{CE}}(u) = \mathbb{E}\Big[\Big(&-\lambda^{-1/2}\int_0^T \langle u(X_t^v,t), \mathrm{d}B_t^v\rangle - \lambda^{-1}\int_0^T \langle u(X_t^v,t), v(X_t^v,t)\rangle\,\mathrm{d}t \\ &+ \frac{\lambda^{-1}}{2}\int_0^T \|u(X_t^v,t)\|^2\,\mathrm{d}t - \lambda^{-1}\mathcal{W}(X^v,0)\Big) \\ &\times \exp\Big(\lambda^{-1}\big(V(X_0^v,0) - \mathcal{W}(X^v,0)\big) - \lambda^{-1/2}\int_0^T \langle v(X_t^v,t), \mathrm{d}B_t\rangle - \frac{\lambda^{-1}}{2}\int_0^T \|v(X_t^v,t)\|^2\,\mathrm{d}t\Big)\Big]\end{aligned} \tag{C.18}$$

And this can be expressed as

$$\bar{\mathcal{L}}_{\mathrm{CE}}(u) = \mathbb{E}\Big[g(u;X_0^v)\exp\big(\lambda^{-1}V(X_0^v,0)\big)\Big],$$

where

$$\begin{aligned}g(u;x) = \mathbb{E}\Big[\Big(&-\lambda^{-1/2}\int_0^T \langle u(X_t^v,t), \mathrm{d}B_t^v\rangle - \lambda^{-1}\int_0^T \langle u(X_t^v,t), v(X_t^v,t)\rangle\,\mathrm{d}t \\ &+ \frac{\lambda^{-1}}{2}\int_0^T \|u(X_t^v,t)\|^2\,\mathrm{d}t - \lambda^{-1}\mathcal{W}(X^v,0)\Big) \\ &\times \exp\Big(-\lambda^{-1}\mathcal{W}(X^v,0) - \lambda^{-1/2}\int_0^T \langle v(X_t^v,t), \mathrm{d}B_t\rangle - \frac{\lambda^{-1}}{2}\int_0^T \|v(X_t^v,t)\|^2\,\mathrm{d}t\Big)\Big|X_0^v = x\Big].\end{aligned}$$

If we consider $g(u;x)$ as a loss function for u, note that it is equivalent to the loss $\bar{L}_{\mathrm{CE}}(u)$ equation in (C.18) for the choice $p_0 = \delta_x$, i.e., p_0 concentrated at x. Since the optimal control u^* is independent of the starting distribution p_0, we deduce that u^* is the unique minimizer of $g(u;x)$,

for all $x \in \mathbb{R}^d$. In consequence, u^* is the unique minimizer of $\mathcal{L}_{\mathrm{CE}}(u) = \mathbb{E}[g(u; X_0^v)]$.

To prove (C.14), note that up to a constant term, the only difference between $\bar{\mathcal{L}}_{\mathrm{CE}}(u)$ and $\mathcal{L}_{\mathrm{CE}}(u)$ is the expectation is reweighted importance weight $\exp\big(-\lambda^{-1}V(X_0^v, 0)\big)$. □

Lemma C.7. *(i) We can rewrite*

$$\tilde{\mathcal{L}}_{\mathrm{Var}_v}(u) = \mathrm{Var}\big(\exp\big(\tilde{Y}_T^{u,v} - \lambda^{-1}g(X_T^v) + \lambda^{-1}V(X_0^v, 0)\big)\big),$$
$$\tilde{\mathcal{L}}^{\log}_{\mathrm{Var}_v}(u) = \mathrm{Var}\big(\tilde{Y}_T^{u,v} - \lambda^{-1}g(X_T^v) + \lambda^{-1}V(X_0^v, 0)\big).$$

When p_0 is concentrated at x_{init}, the terms $V(x_{\mathrm{init}}, 0)$ are constants and can be removed without modifying the landscape. In other words, $\tilde{\mathcal{L}}_{\mathrm{Var}_v}$ and $\tilde{\mathcal{L}}^{\log}_{\mathrm{Var}_v}$ are equal to $\mathcal{L}_{\mathrm{Var}_v}$ and $\mathcal{L}^{\log}_{\mathrm{Var}_v}$ up to a constant term and a constant factor, respectively.

(ii) When p_0 is general, $\tilde{\mathcal{L}}_{\mathrm{Var}_v}$ and $\mathcal{L}_{\mathrm{Var}_v}$ have a different landscape, and the optimum of $\mathcal{L}_{\mathrm{Var}_v}$ may be different from u^*. A related loss that does preserve the optimum is:

$$\begin{aligned}\bar{\mathcal{L}}_{\mathrm{Var}_v}(u) &= \mathbb{E}[\mathrm{Var}_{\mathbb{P}^v}(\frac{d\mathbb{P}^{u^*}}{d\mathbb{P}^u}(X^v)|X_0^v)\exp(-\lambda^{-1}V(X_0^v, 0))] \\ &= \mathbb{E}[\mathrm{Var}\big(\exp(\tilde{Y}_T^{u,v} - \lambda^{-1}g(X_T^v))|X_0^v\big)].\end{aligned}$$

In practice, this is implemented by sampling the m trajectories in one batch starting at the same point X_0^v.

(iii) Also, $\tilde{\mathcal{L}}^{\log}_{\mathrm{Var}_v}$ and $\mathcal{L}^{\log}_{\mathrm{Var}_v}$ have a different landscape, and the optimum of $\mathcal{L}^{\log}_{\mathrm{Var}_v}$ may be different from u^*. In particular, $\mathcal{L}^{\log}_{\mathrm{Var}_v}(u) = \mathrm{Var}_{\mathbb{P}^v}(\log\frac{d\mathbb{P}^{u^*}}{d\mathbb{P}^u}(X^v)\exp(-\lambda^{-1}V(X_0^v, 0)))$. A loss that does preserve the optimum u^* is

$$\begin{aligned}\bar{\mathcal{L}}^{\log}_{\mathrm{Var}_v}(u) &= \mathbb{E}[\mathrm{Var}_{\mathbb{P}^v}(\log\frac{d\mathbb{P}^{u^*}}{d\mathbb{P}^u}(X^v)|X_0^v)\exp(-\lambda^{-1}V(X_0^v, 0))] \\ &= \mathbb{E}[\mathrm{Var}\big(\tilde{Y}_T^{u,v} - \lambda^{-1}g(X_T^v)|X_0^v\big)].\end{aligned}$$

Proof. Using (C.9) and (C.6), we have that

$$\begin{aligned}
\frac{d\mathbb{P}^{u^*}}{d\mathbb{P}}(X^v) &= \exp\big(\lambda^{-1}\big(V(X_0^v,0) - \mathcal{W}(X^v,0)\big)\big),\\
\frac{d\mathbb{P}}{d\mathbb{P}^u}(X^v) &= \exp\Big(-\lambda^{-1/2}\int_0^T \langle u(X_t^v,t), \mathrm{d}B_t^v\rangle + \frac{\lambda^{-1}}{2}\int_0^T \|u(X_t^v,t)\|^2\,\mathrm{d}t\Big)\\
&= \exp\Big(-\lambda^{-1/2}\int_0^T \langle u(X_t^v,t), \mathrm{d}B_t\rangle - \lambda^{-1}\int_0^T \langle u(X_t^v,t), v(X_t^v,t)\rangle\,\mathrm{d}t\\
&\qquad + \frac{\lambda^{-1}}{2}\int_0^T \|u(X_t^v,t)\|^2\,\mathrm{d}t\Big).
\end{aligned}$$

Hence,

$$\log\frac{d\mathbb{P}^{u^*}}{d\mathbb{P}^u}(X^v) = \log\frac{d\mathbb{P}^{u^*}}{d\mathbb{P}}(X^v) + \log\frac{d\mathbb{P}}{d\mathbb{P}^u}(X^v) = \tilde{Y}_T^{u,v} - \lambda^{-1}g(X_T^v) + \lambda^{-1}V(X_0^v,0).$$

Since $\tilde{\mathcal{L}}_{\mathrm{Var}_v}(u) = \mathrm{Var}_{\mathbb{P}^v}(\frac{d\mathbb{P}^{u^*}}{d\mathbb{P}^u})$ and $\tilde{\mathcal{L}}^{\log}_{\mathrm{Var}_v}(u) = \mathrm{Var}_{\mathbb{P}^v}(\log\frac{d\mathbb{P}^{u^*}}{d\mathbb{P}^u})$, this concludes the proof of (i).

To prove (ii), note that for general p_0, $V(X_0^v,0)$ is no longer a constant, but it is if we condition on X_0^v. The proof of (iii) is analogous. □

C.3 Proofs of section 4.3

C.3.1 Proof of Theorem 4.3 and Theorem 4.5

We prove Theorem 4.3 and Theorem 4.5 at the same time. Recall that by (4.9), the optimal control is of the form $u^*(x,t) = -\sigma(t)^\top \nabla V(x,t)$. Consider the loss

$$\tilde{\mathcal{L}}(u) = \mathbb{E}\Big[\frac{1}{T}\int_0^T \big\|u(X_t,t) + \sigma(t)^\top\nabla V(X_t,t)\big\|^2\,\mathrm{d}t\,\exp\Big(-\lambda^{-1}\int_0^T f(X_t,t)\,\mathrm{d}t - \lambda^{-1}g(X_T)\Big)\Big].$$

Clearly, the unique optimum of $\tilde{\mathcal{L}}$ is $-\sigma(t)^\top \nabla V$. We can rewrite $\tilde{\mathcal{L}}$ as

$$\tilde{\mathcal{L}}(u) = \mathbb{E}\Big[\frac{1}{T}\int_0^T \big(\big\|u(X_t,t)\big\|^2 + 2\langle u(X_t,t), \sigma(t)^\top \nabla V(X_t,t)\rangle + \big\|\sigma(t)^\top \nabla V(X_t,t)\big\|^2\big)\,\mathrm{d}t \quad \text{(C.19)}$$
$$\times \exp\big(-\lambda^{-1}\int_0^T f(X_t,t)\,\mathrm{d}t - \lambda^{-1} g(X_T)\big)\Big]\,.$$

Hence, we can express $\tilde{\mathcal{L}}$ as a sum of three terms: one involving $\|u(X_t,t)\|^2$, another involving $\langle u(X_t,t), \sigma(t)^\top V(X_t,t)\rangle$, and a third one, which is constant with respect to u, involving $\|\nabla V(X_t,t)\|^2$. The following lemma provides an alternative expression for the cross term:

Lemma C.8. *The following equality holds:*

$$\begin{aligned}
&\mathbb{E}\Big[\frac{1}{T}\int_0^T \langle u(X_t,t), \sigma(t)^\top \nabla V(X_t,t)\rangle\,\mathrm{d}t\ \exp\big(-\lambda^{-1}\int_0^T f(X_t,t)\,\mathrm{d}t - \lambda^{-1} g(X_T)\big)\Big] \\
&= -\lambda \mathbb{E}\Big[\frac{1}{T}\int_0^T \Big\langle u(X_t,t), \sigma(t)^\top \nabla_x \mathbb{E}\Big[\exp\big(-\lambda^{-1}\int_t^T f(X_s,s)\,\mathrm{d}s - \lambda^{-1} g(X_T)\big)\Big| X_t = x\Big]\Big\rangle\,\mathrm{d}t \\
&\qquad \times \exp\big(-\lambda^{-1}\int_0^t f(X_s,s)\,\mathrm{d}s\big)\Big]. \qquad \text{(C.20)}
\end{aligned}$$

Proof. Recall the definition of $\mathcal{W}(X,t)$ in (C.10), which means that

$$\mathcal{W}(X,0) = \mathcal{W}(X,t) + \int_0^t f(X_s,s)\,\mathrm{d}s. \qquad \text{(C.21)}$$

Let $\{\mathcal{F}_t\}_{t\in[0,T]}$ be the filtration generated by the Brownian motion B. Then, equation (4.9) implies that

$$\sigma(t)^\top \nabla V(X_t,t) = -\frac{\lambda \sigma(t)^\top \nabla_x \mathbb{E}\big[\exp\big(-\lambda^{-1}\mathcal{W}(X,t)\big)\big|\mathcal{F}_t\big]}{\mathbb{E}\big[\exp\big(-\lambda^{-1}\mathcal{W}(X,t)\big)\big|\mathcal{F}_t\big]} \qquad \text{(C.22)}$$

We proceed as follows:

$$
\begin{aligned}
&\mathbb{E}\Big[\frac{1}{T}\int_0^T \langle u(X_t,t), \sigma(t)^\top \nabla V(X_t,t)\rangle \,\mathrm{d}t \,\exp\big(-\lambda^{-1}\mathcal{W}(X,0)\big)\Big] \\
&\overset{\text{(i)}}{=} -\lambda\mathbb{E}\Big[\frac{1}{T}\int_0^T \Big\langle u(X_t,t), \frac{\sigma(t)^\top \nabla_x \mathbb{E}\big[\exp\big(-\lambda^{-1}\mathcal{W}(X,t)\big)\big|\mathcal{F}_t\big]}{\mathbb{E}\big[\exp\big(-\lambda^{-1}\mathcal{W}(X,t)\big)\big|\mathcal{F}_t\big]}\Big\rangle \,\mathrm{d}t \\
&\qquad\qquad \times \mathbb{E}\big[\exp\big(-\lambda^{-1}\mathcal{W}(X,t)\big)\big|\mathcal{F}_t\big]\exp\Big(-\lambda^{-1}\int_0^t f(X_t,t)\,\mathrm{d}t\Big)\Big] \\
&= -\lambda\mathbb{E}\Big[\frac{1}{T}\int_0^T \Big\langle u(X_t,t), \sigma(t)^\top \nabla_x \mathbb{E}\big[\exp\big(-\lambda^{-1}\mathcal{W}(X,t)\big)\big|\mathcal{F}_t\big]\Big\rangle \,\mathrm{d}t \exp\Big(-\lambda^{-1}\int_0^t f(X_s,s)\,\mathrm{d}s\Big)\Big] \\
&\overset{\text{(ii)}}{=} -\lambda\mathbb{E}\Big[\frac{1}{T}\int_0^T \Big\langle u(X_t,t), \sigma(t)^\top \nabla_x \mathbb{E}\big[\exp\big(-\lambda^{-1}\mathcal{W}(X,t)\big)\big|X_t=x\big]\Big\rangle \,\mathrm{d}t \exp\Big(-\lambda^{-1}\int_0^t f(X_s,s)\,\mathrm{d}s\Big)\Big].
\end{aligned}
$$

Here, (i) holds by equation (C.22), the law of total expectation and equation (C.21), and (ii) holds by the Markov property of the solution of an SDE. □

The following proposition, which we prove in subsection C.3.2, provides an alternative expression for $\nabla_x \mathbb{E}\big[\exp\big(-\lambda^{-1}\int_t^T f(X_s,s)\,\mathrm{d}s - \lambda^{-1}g(X_T)\big)\big|X_t = x\big]$. The technique, which is novel and we denote by *path-wise reparamaterization trick*, is of independent interest and may be applied in other settings, as we discuss in section 4.5.

Proposition C.9 (Path-wise reparameterization trick for stochastic optimal control)**.** *For each $t \in [0,T]$, let $M_t : [t,T] \to \mathbb{R}^{d\times d}$ be an arbitrary continuously differentiable function matrix-valued function such that $M_t(t) = \mathrm{Id}$. We have that*

$$
\begin{aligned}
&\nabla_x \mathbb{E}\big[\exp\big(-\lambda^{-1}\int_t^T f(X_s,s)\,\mathrm{d}s - \lambda^{-1}g(X_T)\big)\big|X_t=x\big] \\
&= \mathbb{E}\Big[\Big(-\lambda^{-1}\int_t^T M_t(s)\nabla_x f(X_s,s)\,\mathrm{d}s - \lambda^{-1}M_t(T)\nabla g(X_T) \\
&\qquad + \lambda^{-1/2}\int_t^T (M_t(s)\nabla_x b(X_s,s) - \partial_s M_t(s))(\sigma^{-1})^\top(s)\mathrm{d}B_s\Big) \\
&\qquad \times \exp\Big(-\lambda^{-1}\int_t^T f(X_s,s)\,\mathrm{d}s - \lambda^{-1}g(X_T)\Big)\big|X_t=x\Big]. \qquad (4.22)
\end{aligned}
$$

Plugging (4.22) into the right-hand side of (C.20), we obtain that

$$\begin{aligned}
&\mathbb{E}\Big[\frac{1}{T}\int_0^T \langle u(X_t,t), \sigma(t)^\top \nabla V(X_t,t)\rangle \,\mathrm{d}t \,\exp\Big(-\lambda^{-1}\int_0^T f(X_t,t)\,\mathrm{d}t - \lambda^{-1} g(X_T)\Big)\Big] \\
&= \mathbb{E}\Big[\frac{1}{T}\int_0^T \Big\langle u(X_t,t), \sigma(t)^\top \Big(\int_t^T M_t(s)\nabla_x f(X_s,s)\,\mathrm{d}s + M_t(T)\nabla g(X_T) \\
&\qquad - \lambda^{1/2}\int_t^T (M_t(s)\nabla_x b(X_s,s) - \partial_s M_t(s))(\sigma^{-1})^\top(s)\mathrm{d}B_s\Big)\Big\rangle \,\mathrm{d}t \\
&\qquad \times \exp\Big(-\lambda^{-1}\int_0^T f(X_t,t)\,\mathrm{d}t - \lambda^{-1} g(X_T)\Big)\Big].
\end{aligned}$$

If we plug this into the right-hand side of (C.19) and complete the squared norm, we get that

$$\begin{aligned}
\tilde{\mathcal{L}}(u) = \mathbb{E}\Big[\frac{1}{T}\int_0^T &\big(\big\|u(X_t,t) - \tilde{w}(t,X,B,M_t)\big\|^2 \\
&- \big\|\tilde{w}(t,X,B,M_t)\big\|^2 + \big\|u^*(X_t,t)\big\|^2\big)\,\mathrm{d}t \,\exp\Big(-\lambda^{-1}\mathcal{W}(X,0)\Big)\Big]
\end{aligned}$$

where $\tilde{w}$ is defined in equation (4.25). We also define $\Phi(u;X,B)$ as

$$\Phi(u;X,B) = \frac{1}{T}\int_0^T \big(\big\|u(X_t,t) - \tilde{w}(t,X,B,M_t)\big\|^2\big)\,\mathrm{d}t.$$

Now, by the Girsanov theorem (Theorem C.2), we have that for an arbitrary control $v \in \mathcal{U}$,

$$\begin{aligned}
&\mathbb{E}[\Phi(u;X,B)\exp\Big(-\lambda^{-1}\mathcal{W}(X,0)\Big)] \\
&= \mathbb{E}\Big[\Phi(u;X^v,B^v)\exp\Big(-\lambda^{-1}\mathcal{W}(X^v,0) - \lambda^{-1/2}\int_0^T \langle v(X_t^v,t), \mathrm{d}B_t^v\rangle + \frac{\lambda^{-1}}{2}\int_0^T \|v(X_t^v,t)\|^2\,\mathrm{d}t\Big)\Big] \\
&= \mathbb{E}\Big[\Phi(u;X^v,B^v)\exp\Big(-\lambda^{-1}\mathcal{W}(X^v,0) - \lambda^{-1/2}\int_0^T \langle v(X_t^v,t), \mathrm{d}B_t\rangle - \frac{\lambda^{-1}}{2}\int_0^T \|v(X_t^v,t)\|^2\,\mathrm{d}t\Big)\Big],
\end{aligned}$$

where $B_t^v := B_t + \lambda^{-1/2}\int_0^t v(X_s^v,s)\,\mathrm{d}s$. Reexpressing B^v in terms of B, we can rewrite $\Phi(u;X^v,B^v)$

and $\tilde{w}(t, X^v, B^v, M_t)$ as follows:

$$\Phi(u; X^v, B^v) = \frac{1}{T}\int_0^T \left\|u(X_t^v, t) - \tilde{w}(t, X^v, B^v, M_t)\right\|^2 \mathrm{d}t,$$

$$\begin{aligned}\tilde{w}(t, X^v, B^v, M_t) = \sigma(t)^\top \Big(&- \int_t^T M_t(s)\nabla_x f(X_s^v, s)\,\mathrm{d}s - M_t(T)\nabla g(X_T^v) \\ &+ \lambda^{1/2}\int_t^T (M_t(s)\nabla_x b(X_s^v, s) - \partial_s M_t(s))(\sigma^{-1})^\top(X_s^v, s)\mathrm{d}B_s \\ &+ \int_t^T (M_t(s)\nabla_x b(X_s^v, s) - \partial_s M_t(s))(\sigma^{-1})^\top(X_s^v, s)v(X_s^v, s)\mathrm{d}s\Big).\end{aligned}$$

Putting everything together, we obtain that

$$\tilde{\mathcal{L}}(u) = \mathcal{L}_{\mathrm{SOCM}}(u, M) - K,$$

where $\mathcal{L}(u, M)$ is the loss defined in (C.31) (note that $w(t, v, X^v, B, M_t) := \tilde{w}(t, X^v, B^v, M_t)$), and

$$K = \mathbb{E}\Big[\frac{1}{T}\int_0^T \big(\left\|\tilde{w}(t, X, B, M_t)\right\|^2 - \left\|u^*(X_t, t)\right\|^2\big)\,\mathrm{d}t\ \exp\big(-\lambda^{-1}\mathcal{W}(X, 0)\big)\Big]$$

To complete the proof of equation (4.23), remark that $\tilde{\mathcal{L}}(u)$ can be rewritten as

$$\begin{aligned}\tilde{\mathcal{L}}(u) &= \mathbb{E}\Big[\frac{1}{T}\int_0^T \left\|u(X_t, t) - u^*(X_t, t)\right\|^2 \mathrm{d}t\ \exp\big(-\lambda^{-1}\mathcal{W}(X, 0)\big)\Big] \\ &= \mathbb{E}\Big[\frac{1}{T}\int_0^T \left\|u(X_t, t) - u^*(X_t, t)\right\|^2 \mathrm{d}t\ \frac{d\mathbb{P}^{u^*}}{d\mathbb{P}}(X)\exp(-\lambda^{-1}V(X_0, 0))\Big] \\ &= \mathbb{E}\Big[\frac{1}{T}\int_0^T \left\|u(X_t^{u^*}, t) - u^*(X_t^{u^*}, t)\right\|^2 \mathrm{d}t\ \exp(-\lambda^{-1}V(X_0^{u^*}, 0))\Big].\end{aligned}$$

It only remains to reexpress K. Note that by Theorem 4.4, we have that

$$\begin{aligned} u^*(X_t,t) &= \frac{\mathbb{E}\big[\tilde{w}(t,X,B,M_t)\exp\big(-\lambda^{-1}\mathcal{W}(X,0)\big)|\mathcal{F}_t\big]}{\mathbb{E}\big[\exp\big(-\lambda^{-1}\mathcal{W}(X,0)\big)|\mathcal{F}_t\big]} \\ &= \frac{\mathbb{E}\big[\tilde{w}(t,X,B,M_t)\frac{d\mathbb{P}^{u^*}}{d\mathbb{P}}(X)|\mathcal{F}_t\big]\exp(-\lambda^{-1}V(X_0,0))}{\mathbb{E}\big[\frac{d\mathbb{P}^{u^*}}{d\mathbb{P}}(X)|\mathcal{F}_t\big]\exp(-\lambda^{-1}V(X_0,0))} = \frac{\mathbb{E}\big[\tilde{w}(t,X,B,M_t)\frac{d\mathbb{P}^{u^*}}{d\mathbb{P}}(X)|\mathcal{F}_t\big]}{\mathbb{E}\big[\frac{d\mathbb{P}^{u^*}}{d\mathbb{P}}(X)|\mathcal{F}_t\big]} \\ &= \mathbb{E}\big[\tilde{w}(t,X^{u^*},B^{u^*},M_t)|X_t^{u^*}=X_t\big] \end{aligned}$$

Hence, using the Girsanov theorem (Theorem C.2) several times, we have that

$$\begin{aligned} K &= \mathbb{E}\Big[\frac{1}{T}\int_0^T \big\|\tilde{w}(t,X^{u^*},B^{u^*},M_t)\|^2 - \|\mathbb{E}\big[\tilde{w}(t,X^{u^*},B^{u^*},M_t)|X_t^{u^*}\big]\big\|^2 \,\mathrm{d}t\ \exp(-\lambda^{-1}V(X_0^{u^*},0))\Big] \\ &= \mathbb{E}\Big[\frac{1}{T}\int_0^T \big\|\tilde{w}(t,X^{u^*},B^{u^*},M_t) - \mathbb{E}\big[\tilde{w}(t,X^{u^*},B^{u^*},M_t)|X_t^{u^*}\big]\big\|^2 \,\mathrm{d}t\ \exp(-\lambda^{-1}V(X_0^{u^*},0))\Big] \\ &= \mathbb{E}\Big[\frac{1}{T}\int_0^T \Big\|\tilde{w}(t,X,B,M_t) - \frac{\mathbb{E}[\tilde{w}(t,X,B,M_t)\exp(-\lambda^{-1}\mathcal{W}(X,0))|X_t]}{\mathbb{E}[\exp(-\lambda^{-1}\mathcal{W}(X,0))|X_t]}\Big\|^2 \mathrm{d}t\exp(-\lambda^{-1}\mathcal{W}(X,0))\Big] \\ &= \mathbb{E}\Big[\frac{1}{T}\int_0^T \Big\|w(t,v,X^v,B,M_t) - \frac{\mathbb{E}[w(t,v,X^v,B,M_t)\alpha(v,X^v,B)|X_t^v]}{\mathbb{E}[\alpha(v,X^v,B)|X_t^v]}\Big\|^2 \mathrm{d}t\,\alpha(v,X^v,B)\Big], \end{aligned}$$

which concludes the proof, noticing that $K = \mathrm{Var}(w;M)$.

C.3.2 Proof of the path-wise reparameterization trick (Theorem 4.4)

We prove a more general statement (Theorem C.10), and show that Theorem 4.4 is a particular case of it.

Proposition C.10 (Path-wise reparameterization trick)**.** *Let $(\Omega,\mathcal{F},\mathbb{P})$ be a probability space, and $B:\Omega\times[0,T]\to\mathbb{R}^d$ be a Brownian motion. Let $X:\Omega\times[0,T]\to\mathbb{R}^d$ be the uncontrolled process given by (4.6), and let $\psi:\Omega\times\mathbb{R}^d\times[0,T]\to\mathbb{R}^d$ be an arbitrary random process such that:*

- *For all $z\in\mathbb{R}^d$, the process $\psi(\cdot,z,\cdot):\Omega\times[0,T]\to\mathbb{R}^d$ is adapted to the filtration $(\mathcal{F}_s)_{s\in[0,T]}$ of the Brownian motion B.*

- *For all $\omega \in \Omega$, $\psi(\omega, \cdot, \cdot) : \mathbb{R}^d \times [0,T] \to \mathbb{R}^d$ is a twice-continuously differentiable function such that $\psi(\omega, z, 0) = z$ for all $z \in \mathbb{R}^d$, and $\psi(\omega, 0, s) = 0$ for all $s \in [0,T]$.*

Let $F : C([0,T]; \mathbb{R}^d) \to \mathbb{R}$ be a Fréchet-differentiable functional. We use the notation $X + \psi(z, \cdot) = (X_s(\omega) + \psi(\omega, z, s))_{s\in[0,T]}$ to denote the shifted process, and we will omit the dependency of ψ on ω in the proof. Then,

$$\begin{aligned}
&\nabla_x \mathbb{E}\big[\exp\big(-F(X)\big)\big|X_0 = x\big] \qquad (C.23)\\
&= \mathbb{E}\Big[\big(-\nabla_z F(X+\psi(z,\cdot))|_{z=0} + \lambda^{-1/2}\int_0^T (\nabla_z\psi(0,s)\nabla_x b(X_s,s) - \nabla_z\partial_s\psi(0,s))(\sigma^{-1})^\top(s)\mathrm{d}B_s\big)\\
&\qquad \times \exp\big(-F(X)\big)\big|X_0 = x\Big]
\end{aligned}$$

Proof of Theorem 4.4. Given a family of functions $(M_t)_{t\in[0,T]}$ satisfying the conditions in Theorem 4.4, we can define a family $(\psi_t)_{t\in[0,T]}$ of functions $\psi_t : \mathbb{R}^d \times [t,T] \to \mathbb{R}^d$ as $\psi_t(z,s) = M_t(s)^\top z$. Note that $\psi_t(z,t) = z$ for all $z \in \mathbb{R}^d$ and $\psi_t(0,s) = 0$ for all $s \in [t,T]$, and that $\nabla_z\psi_t(z,s) = M_t(s)$. Hence, ψ_t can be seen as a random process which is constant with respect to $\omega \in \Omega$, and which fulfills the conditions in Theorem C.10 up to a trivial time change of variable from $[t,T]$ to $[0,T]$. We also define the family $(F_t)_{t\in[0,T]}$ of functionals $F_t : C([t,T]; \mathbb{R}^d) \to \mathbb{R}$ as $F_t(X) = \lambda^{-1}\int_t^T f(X_s,s)\,\mathrm{d}s + \lambda^{-1}g(X_T)$. We have that

$$\begin{aligned}
&\nabla_z F_t(X+\psi_t(z,\cdot))\\
&= \nabla_z\big(\lambda^{-1}\int_t^T f(X_s+\psi_t(z,s),s)\,\mathrm{d}s + \lambda^{-1}g(X_T+\psi_t(z,T))\big)\\
&\overset{\text{(i)}}{=} \lambda^{-1}\int_t^T \nabla_z\psi_t(z,s)\nabla f(X_s+\psi_t(z,s),s)\,\mathrm{d}s + \lambda^{-1}\nabla_z\psi_t(z,T)\nabla g(X_T+\psi_t(z,T))\\
&= \lambda^{-1}\int_t^T M_t(s)\nabla f(X_s+\psi_t(z,s),s)\,\mathrm{d}s + \lambda^{-1}M_t(T)\nabla g(X_T+\psi_t(z,T)),
\end{aligned}$$

where equality (i) holds by the Leibniz rule. Using that $\psi_t(0,s) = 0$, we obtain that:

$$\nabla_z F_t(X + \psi_t(z,\cdot))\big|_{z=0} = \lambda^{-1} \int_t^T \nabla_z \psi_t(0,s) \nabla f(X_s, s)\, \mathrm{d}s + \lambda^{-1} \nabla_z \psi_t(T,0) \nabla g(X_T),$$

Up to a trivial time change of variable from $[t,T]$ to $[0,T]$, Theorem 4.4 follows from plugging these choices into equation (C.23).

Remark 2. *We can use matrices $M_t(s)$ that depend on the process X up to time s, since the resulting processes $\psi_t(\cdot, z, \cdot)$ are adapted to the filtration of the Brownian motion B. More specifically, if we let $M_t : \mathbb{R}^d \times [t,T] \longrightarrow \mathbb{R}^{d\times d}$ be an arbitrary continuously differentiable function matrix-valued function such that $M_t(x,t) = \mathrm{Id}$ for all $x \in \mathbb{R}^d$, and we define the exponential moving average of X as the process $X^{(\upsilon)}$ given by*

$$X_t^{(\upsilon)} = \upsilon \int_0^t e^{-\upsilon(t-s)} X_s \, ds,$$

we have that

$$\begin{aligned}\frac{d}{ds} M_t(X_s^{(\upsilon)}, s) &= \langle \nabla M_t(X_s^{(\upsilon)}, s), \frac{dX_s^{(\upsilon)}}{ds} \rangle + \partial_s M_t(X_s^{(\upsilon)}, s) \\ &= \lambda \langle \nabla_x M_t(X_s^{(\upsilon)}, s), X_s - X_s^{(\upsilon)} \rangle + \partial_s M_t(X_s^{(\upsilon)}, s),\end{aligned}$$

and we can write

$$\begin{aligned}&\nabla_x \mathbb{E}\big[\exp\big(-\lambda^{-1} \int_t^T f(X_s,s)\, \mathrm{d}s - \lambda^{-1} g(X_T)\big) \big| X_t = x\big] \\ &= \mathbb{E}\Big[\big(-\lambda^{-1} \int_t^T M_t(X_s^{(\upsilon)}, s) \nabla_x f(X_s, s)\, \mathrm{d}s - \lambda^{-1} M_t(X_T^{(\upsilon)}, T) \nabla g(X_T) \\ &\qquad + \lambda^{-1/2} \int_t^T (M_t(X_s^{(\upsilon)}, s) \nabla_x b(X_s, s) - \frac{d}{ds} M_t(X_s^{(\upsilon)}, s))(\sigma^{-1})^\top(s) \mathrm{d}B_s\big) \\ &\qquad \times \exp\big(-\lambda^{-1} \int_t^T f(X_s,s)\, \mathrm{d}s - \lambda^{-1} g(X_T)\big) \big| X_t = x\Big].\end{aligned}$$

Plugging this into the proof of Theorem 4.3, we would obtain a variant of SOCM (Alg. 4) where the matrix-valued neural network M_ω takes inputs (t, s, x) instead of (t, s). Since the optimization class is larger, from the bias-variance in Theorem 4.5 we deduce that this variant would yield a lower variance of the vector field w, and likely an algorithm with lower error. This is at the expense of an increased number of function evaluations (NFE) of M_ω; one would need $\frac{K(K+1)m}{2}$ NFE per batch instead of only $\frac{K(K+1)}{2}$, which may be too expensive if the architecture of M_ω is large. A way to speed up the computation per batch is to parameterize M_t using cubic splines.

□

Proof of Theorem C.10. Recall that

$$\mathrm{d}X_s = b(X_s, s)\,\mathrm{d}s + \sqrt{\lambda}\sigma(s)\,\mathrm{d}B_s, \qquad X_0 \sim p_0,$$

is the SDE for the uncontrolled process. For arbitrary $x, z \in \mathbb{R}^d$, we consider the following SDEs conditioned on the initial points:

$$\mathrm{d}X_s^{(x+z)} = b(X_s^{(x+z)}, s)\,\mathrm{d}s + \sqrt{\lambda}\sigma(s)\,\mathrm{d}B_s, \qquad X_0^{(x+z)} = x + z, \tag{C.24}$$

$$\mathrm{d}X_s^{(x)} = b(X_s^{(x)}, s)\,\mathrm{d}s + \sqrt{\lambda}\sigma(s)\,\mathrm{d}B_s, \qquad X_0^{(x)} = x. \tag{C.25}$$

Suppose that $\psi : \mathbb{R}^d \times [0, T] \to \mathbb{R}^d$ satisfies the properties in the statement of Theorem C.10. If $\tilde{X}^{(x)}$ is a solution of

$$\mathrm{d}\tilde{X}_s^{(x)} = (b(\tilde{X}_s^{(x)} + \psi(z, s), s) - \partial_s\psi(z, s))\,\mathrm{d}s + \sqrt{\lambda}\sigma(s)\,\mathrm{d}B_s, \qquad \tilde{X}_0^{(x)} = x,$$

then $X^{(x+z)} = \tilde{X}^{(x)} + \psi(z, \cdot)$ is a solution of (C.24). This is because $X_0^{(x+z)} = \tilde{X}_0^{(x)} + \psi(z, 0) =$

$\tilde{X}_0^{(x)} + z = x + z$, and

$$\begin{aligned}
\mathrm{d}X_s^{(x+z)} &= \mathrm{d}\tilde{X}_s^{(x)} + \partial_s\psi(z,s)\,\mathrm{d}s \\
&= (b(\tilde{X}_s^{(x)} + \psi(z,s), s) - \partial_s\psi(z,s))\,\mathrm{d}s + \sqrt{\lambda}\sigma(s)\,\mathrm{d}B_s + \partial_s\psi(z,s)\,\mathrm{d}s \\
&= b(X_s^{(x+z)}, s)\,\mathrm{d}s + \sqrt{\lambda}\sigma(s)\,\mathrm{d}B_s,
\end{aligned}$$

Note that we may rewrite (C.25) as

$$\begin{aligned}
\mathrm{d}X_s^{(x)} = (b(X_s^{(x)} + \psi(z,s), s) - \partial_s\psi(z,s))\,\mathrm{d}s \\
+ (b(X_s^{(x)}, s) - b(X_s^{(x)} + \psi(z,s), s) + \partial_s\psi(z,s))\,\mathrm{d}s + \sqrt{\lambda}\sigma(s)\,\mathrm{d}B_s, \qquad X_t^{(x)} \sim p_0.
\end{aligned}$$

Hence, since $\psi(z,s)$ is a random process adapted to the filtration of B, we can apply the Girsanov theorem for SDEs (Corollary C.3) on $\tilde{X}^{(x)}$ and $X^{(x)}$, and we have that for any bounded continuous functional Φ,

$$\begin{aligned}
&\mathbb{E}[\Phi(\tilde{X}^{(x)})] \\
&= \mathbb{E}\Big[\Phi(X^{(x)}) \exp\Big(\int_0^T \lambda^{-1/2}\sigma(s)^{-1}(b(X_s^{(x)} + \psi(z,s), s) - b(X_s^{(x)}, s) - \partial_s\psi(z,s))\,\mathrm{d}B_s \\
&\qquad - \frac{1}{2}\int_0^T \|\lambda^{-1/2}\sigma(s)^{-1}(b(X_s^{(x)} + \psi(z,s), s) - b(X_s^{(x)}, s) - \partial_s\psi(z,s))\|^2\,\mathrm{d}s\Big)\Big]
\end{aligned} \tag{C.26}$$

We can write

$$\begin{aligned}
&\mathbb{E}\big[\exp\big(-F(X)\big)\big|X_0 = x+z\big] \overset{\text{(i)}}{=} \mathbb{E}\big[\exp\big(-F(X^{(x+z)})\big)\big] \overset{\text{(ii)}}{=} \mathbb{E}\big[\exp\big(-F(\tilde{X}^{(x)}+\psi(z,\cdot))\big)\big] \\
&\overset{\text{(iii)}}{=} \mathbb{E}\big[\exp\big(-F(X^{(x)}+\psi(z,\cdot))\big) \\
&\qquad \times \exp\Big(\int_0^T \lambda^{-1/2}\sigma(s)^{-1}(b(X_s^{(x)}+\psi(z,s),s) - b(X_s^{(x)},s) - \partial_s\psi(z,s))\,\mathrm{d}B_s \\
&\qquad\qquad - \frac{1}{2}\int_0^T \|\lambda^{-1/2}\sigma(s)^{-1}(b(X_s^{(x)}+\psi(z,s),s) - b(X_s^{(x)},s) - \partial_s\psi(z,s))\|^2\,\mathrm{d}s\Big)\Big] \\
&\overset{\text{(iv)}}{=} \mathbb{E}\Big[\exp\Big(-F(X+\psi(z,\cdot)) + \int_0^T \lambda^{-1/2}\sigma(s)^{-1}(b(X_s+\psi(z,s),s) - b(X_s,s) - \partial_s\psi(z,s))\,\mathrm{d}B_s \\
&\qquad\qquad - \frac{1}{2}\int_0^T \|\lambda^{-1/2}\sigma(s)^{-1}(b(X_s+\psi(z,s),s) - b(X_s,s) - \partial_s\psi(z,s))\|^2\,\mathrm{d}s\Big)\big|X_0 = x\Big]
\end{aligned} \tag{C.27}$$

Equality (i) holds by the definition of $X^{(x+z)}$, equality (ii) holds by the fact $X_s^{(x+z)} = \tilde{X}_s^{(x)} + \psi(z,s)$, equality (iii) holds by equation (C.26), and equality (iv) holds by the definition of $X_s^{(x)}$. We conclude the proof by differentiating the right-hand side of (C.27) with respect to z. Namely,

$$\begin{aligned}
&\nabla_x \mathbb{E}\big[\exp\big(-F(X)\big)\big|X_0 = x\big] = \nabla_z \mathbb{E}\big[\exp\big(-F(X)\big)\big|X_0 = x+z\big]\big|_{z=0} \\
&\overset{\text{(i)}}{=} \mathbb{E}\Big[\Big(-\nabla_z F(X+\psi(z,\cdot)) + \lambda^{-1/2}\int_0^T (\nabla_z\psi(0,s)\nabla_x b(X_s,s) - \nabla_z\partial_s\psi(0,s))(\sigma^{-1})^\top(s)\mathrm{d}B_s\Big) \\
&\qquad \times \exp\big(-F(X)\big)\big|X_0 = x\Big]
\end{aligned}$$

In equality (i) we used (C.27), and that:

- by the Leibniz rule,

$$\begin{aligned}
&\nabla_z \int_0^T \|\sigma(s)^{-1}(b(X_s+\psi(z,s),s) - b(X_s,s) - \partial_s\psi(z,s))\|^2\,\mathrm{d}s\Big|_{z=0} \\
&= \int_0^T \nabla_z \|\sigma(s)^{-1}(b(X_s+\psi(z,s),s) - b(X_s,s) - \partial_s\psi(z,s))\|^2\Big|_{z=0}\,\mathrm{d}s = 0.
\end{aligned}$$

- and by the Leibniz rule for stochastic integrals (see [Hutton and Nelson 1984]),

$$\begin{aligned}&\nabla_z\Big(\int_0^T \sigma(s)^{-1}(b(X_s+\psi(z,s),s)-b(X_s,s)-\partial_s\psi(z,s))\,\mathrm{d}B_s\Big)\Big|_{z=0}\\&=\int_0^T(\nabla_z\psi(0,s)\nabla_x b(X_s,s)-\nabla_z\partial_s\psi(0,s))(\sigma^{-1})^\top(s)\,\mathrm{d}B_s.\end{aligned}$$

□

C.3.3 Informal derivation of the path-wise reparameterization trick

In this subsection, we provide an informal, intuitive derivation of the path-wise reparameterization trick as stated in Theorem C.10. For simplicity, we particularize the functional F to $F(X) = \lambda^{-1}\int_0^T f(X_s,s)\,\mathrm{d}s + \lambda^{-1}g(X_T)$. Consider the Euler-Maruyama discretization of the uncontrolled process X defined in (4.6), with $K+1$ time steps (let $\delta = T/K$ be the step size). This is a family of random variables $\hat{X} = (\hat{X}_k)_{k=0:K}$ defined as

$$\hat{X}_0 \sim p_0, \qquad \hat{X}_{k+1} = \hat{X}_k + \delta b(\hat{X}_k, k\delta) + \sqrt{\delta\lambda}\sigma(k\delta)\varepsilon_k, \qquad \varepsilon_k \sim N(0,I).$$

Note that we can approximate

$$\begin{aligned}&\mathbb{E}\Big[\exp\Big(-\lambda^{-1}\int_0^T f(X_s,s)\,\mathrm{d}s - \lambda^{-1}g(X_T)\Big)\Big|X_0 = x\Big]\\&\approx \mathbb{E}\Big[\exp\Big(-\lambda^{-1}\delta\sum_{k=0}^{K-1} f(\hat{X}_k,s) - \lambda^{-1}g(\hat{X}_K)\Big)\Big|\hat{X}_0 = x\Big],\end{aligned}$$

and that this is an equality in the limit $K \to \infty$, as the interpolation of the Euler-Maruyama discretization $\hat{X}^{(x)}$ converges to the process $X^{(x)}$. Now, remark that for $k \in \{0,\dots,K-1\}$,

$\hat{X}_{k+1}|\hat{X}_k \sim N(\hat{X}_k + \delta b(\hat{X}_k, k\delta), \delta\lambda(\sigma\sigma^\top)(k\delta))$. Hence,

$$
\begin{aligned}
&\mathbb{E}\big[\exp\big(-\lambda^{-1}\delta\sum_{k=0}^{K-1} f(\hat{X}_k, s) - \lambda^{-1} g(\hat{X}_K)\big)\big|\hat{X}_0 = x\big] \\
&= C^{-1}\iint_{(\mathbb{R}^d)^K} \exp\big(-\lambda^{-1}\delta\sum_{k=0}^{K-1} f(\hat{x}_k, s) - \lambda^{-1} g(\hat{x}_K) \\
&\qquad\qquad -\frac{1}{2\delta\lambda}\sum_{k=1}^{K-1}\|\sigma^{-1}(k\delta)(\hat{x}_{k+1}-\hat{x}_k-\delta b(\hat{x}_k, k\delta))\|^2 \\
&\qquad\qquad -\frac{1}{2\delta\lambda}\|\sigma^{-1}(0)(\hat{x}_1-x-\delta b(x,0))\|^2\big)\,\mathrm{d}\hat{x}_1\cdots\mathrm{d}\hat{x}_K,
\end{aligned}
$$

where $C = \sqrt{(2\pi\delta\lambda)^K\prod_{k=0}^{K-1}\det((\sigma\sigma^\top)(k\delta))}$. Now, let $\psi : \mathbb{R}^d \times [0,T] \to \mathbb{R}^d$ be an arbitrary twice differentiable function such that $\psi(z,0) = z$ for all $z \in \mathbb{R}^d$, and $\psi(0,s) = 0$ for all $s \in [0,T]$. We can write

$$
\begin{aligned}
&\nabla_x\mathbb{E}\big[\exp\big(-\lambda^{-1}\delta\sum_{k=0}^{K-1} f(\hat{X}_k, s) - \lambda^{-1} g(\hat{X}_K)\big)|\hat{X}_0 = x\big] \\
&= \nabla_z\mathbb{E}\big[\exp\big(-\lambda^{-1}\delta\sum_{k=0}^{K-1} f(\hat{X}_k, s) - \lambda^{-1} g(\hat{X}_K)\big)|\hat{X}_0 = x+z\big]|_{z=0} \\
&= C^{-1}\nabla_z\Big(\iint_{(\mathbb{R}^d)^K} \exp\big(-\lambda^{-1}\delta\sum_{k=0}^{K-1} f(\hat{x}_k, s) - \lambda^{-1} g(\hat{x}_K) \\
&\qquad\qquad -\frac{1}{2\delta\lambda}\sum_{k=1}^{K-1}\|\sigma^{-1}(k\delta)(\hat{x}_{k+1}-\hat{x}_k-\delta b(\hat{x}_k, k\delta))\|^2 \\
&\qquad\qquad -\frac{1}{2\delta\lambda}\|\sigma^{-1}(0)(\hat{x}_1-(x+z)-\delta b(x+z,0))\|^2\big)\,\mathrm{d}\hat{x}_1\cdots\mathrm{d}\hat{x}_K\Big)|_{z=0} \\
&= C^{-1}\nabla_z\Big(\iint_{(\mathbb{R}^d)^K} \exp\big(-\lambda^{-1}\delta\sum_{k=0}^{K-1} f(\hat{x}_k + \psi(z,k\delta), s) - \lambda^{-1} g(\hat{x}_K + \psi(z,K\delta)) \\
&\qquad -\frac{1}{2\delta\lambda}\sum_{k=1}^{K-1}\|\sigma^{-1}(k\delta)(\hat{x}_{k+1}+\psi(z,(k+1)\delta)-\hat{x}_k-\psi(z,k\delta)-\delta b(\hat{x}_k+\psi(z,k\delta),k\delta))\|^2 \\
&\qquad -\frac{1}{2\delta\lambda}\|\sigma^{-1}(0)(\hat{x}_1+\psi(z,\delta)-(x+\psi(z,0))-\delta b(x+\psi(z,0),0))\|^2\big)\,\mathrm{d}\hat{x}_1\cdots\mathrm{d}\hat{x}_K\Big)|_{z=0}
\end{aligned}
\tag{C.28}
$$

In the last equality, we used that for $k \in \{1, \dots, K\}$, the variables $\hat{x}_k$ are integrated over $\mathbb{R}^d$, which means that adding an offset $\psi(z, k\delta)$ does not change the value of the integral. We also used that $\psi(z, 0) = z$. Now, for fixed values of $\hat{x} = (\hat{x}_1, \dots, \hat{x}_K)$, and letting $\hat{x}_0 = x$, we define

$$G_{\hat{x}}(z) = \lambda^{-1}\delta \sum_{k=0}^{K-1} f(\hat{x}_k + \psi(z, k\delta), s) + \lambda^{-1} g(\hat{x}_K + \psi(z, K\delta)) \\ + \frac{1}{2\delta\lambda} \sum_{k=0}^{K-1} \|\sigma^{-1}(k\delta)(\hat{x}_{k+1} + \psi(z, (k+1)\delta) - \hat{x}_k - \psi(z, k\delta) - \delta b(\hat{x}_k + \psi(z, k\delta), k\delta))\|^2.$$

Using that $\psi(0, s) = 0$ for all $s \in [0, T]$, we have that:

$$G_{\hat{x}}(0) = \lambda^{-1}\delta \sum_{k=0}^{K-1} f(\hat{x}_k, s) + \lambda^{-1} g(\hat{x}_K) + \frac{1}{2\delta\lambda} \sum_{k=0}^{K-1} \|\sigma^{-1}(k\delta)(\hat{x}_{k+1} - \hat{x}_k - \delta b(\hat{x}_k, k\delta))\|^2.$$

$$\nabla G_{\hat{x}}(z)|_{z=0} = \lambda^{-1}\delta \sum_{k=0}^{K-1} \nabla\psi(0, k\delta) \nabla f(\hat{x}_k, s) + \lambda^{-1} \nabla\psi(0, K\delta) \nabla g(\hat{x}_K) \\ + \frac{1}{\delta\lambda} \sum_{k=0}^{K-1} (\nabla_z \psi(0, (k+1)\delta) - \nabla_z \psi(0, k\delta) - \delta \nabla\psi(0, k\delta) \nabla b(\hat{x}_k, k\delta)) \\ \times ((\sigma^{-1})^\top \sigma^{-1})(k\delta)(\hat{x}_{k+1} - \hat{x}_k - \delta b(\hat{x}_k, k\delta))$$

And we can express the right-hand side of (C.28) in terms of $G_{\hat{x}}(0)$ and $\nabla G_{\hat{x}}(z)|_{z=0}$:

$$\nabla_z \Big(C^{-1} \iint_{(\mathbb{R}^d)^K} \exp\big(-G_{\hat{x}}(z)\big)\, \mathrm{d}y_1 \cdots \mathrm{d}y_K \Big) \\ = -C^{-1} \iint_{(\mathbb{R}^d)^K} \nabla G_{\hat{x}}(z)|_{z=0} \exp\big(-G_{\hat{x}}(0)\big)\, \mathrm{d}y_1 \cdots \mathrm{d}y_K$$

We define $\epsilon_k = \frac{1}{\sqrt{\delta\lambda}}\sigma^{-1}(k\delta)(\hat{x}_{k+1} - \hat{x}_k - \delta b(\hat{x}_k, k\delta))$, and then, we are able to write

$$\hat{x}_{k+1} = \hat{x}_k + \delta b(\hat{x}_k, k\delta) + \sqrt{\delta\lambda}\sigma(k\delta)\epsilon_k, \qquad \hat{x}_0 = x \tag{C.29}$$

$$\begin{aligned} G_{\hat{x}}(0) &= \lambda^{-1}\delta \sum_{k=0}^{K-1} f(\hat{x}_k, s) + \lambda^{-1} g(\hat{x}_K) + \frac{1}{2}\sum_{k=0}^{K-1} \|\epsilon_k\|^2, \\ \nabla G_{\hat{x}}(z)|_{z=0} &= \lambda^{-1}\delta \sum_{k=0}^{K-1} \nabla\psi(0, k\delta)\nabla f(\hat{x}_k, s) + \lambda^{-1}\nabla\psi(0, K\delta)\nabla g(\hat{x}_K) \\ &\quad + \sqrt{\delta\lambda^{-1}} \sum_{k=0}^{K-1} (\partial_s \nabla_z \psi(0, k\delta) + O(\delta) - \nabla\psi(0, k\delta)\nabla b(\hat{x}_k, k\delta))(\sigma^{-1})^\top(k\delta)\epsilon_k. \end{aligned} \tag{C.30}$$

Then, taking the limit $K \to \infty$ (i.e. $\delta \to 0$), we recognize (C.29) as Euler-Maruyama discretization of the uncontrolled process X in equation (4.6) conditioned on $X_0 = x$, and the last term in (C.30) as the Euler-Maruyama discretization of the stochastic integral $\lambda^{-1/2}\int_0^T (\partial_s \nabla_z \psi(0, s) - \nabla\psi(0, s)\nabla b(X_s^{(x)}, s))(\sigma^{-1})^\top(s)\, dB_s$. Thus,

$$\begin{aligned} &\lim_{K\to\infty} \nabla_x \mathbb{E}\Big[\exp\Big(-\lambda^{-1}\delta \sum_{k=0}^{K-1} f(\hat{X}_k, s) - \lambda^{-1} g(\hat{X}_K)\Big)\Big] \\ &= \mathbb{E}\Big[\Big(-\lambda^{-1}\int_0^T \nabla\psi(0, s)\nabla_x f(X_s, s)\, \mathrm{d}s - \lambda^{-1}\nabla\psi(0, T)\nabla g(X_T) \\ &\qquad + \lambda^{-1/2}\int_0^T (\nabla\psi(0, s)\nabla_x b(X_s, s) - \partial_s\nabla\psi(0, s))(\sigma^{-1})^\top(s)\, \mathrm{d}B_s\Big) \\ &\qquad \times \exp\Big(-\lambda^{-1}\int_0^T f(X_s, s)\, \mathrm{d}s - \lambda^{-1} g(X_T)\Big)\Big| X_0 = x\Big], \end{aligned}$$

which concludes the derivation.

C.3.4 SOCM with the adjoint method for SDEs

Proposition C.11. *Let $\mathcal{L}_{\mathrm{SOCM}} : L^2(\mathbb{R}^d \times [0,T]; \mathbb{R}^d) \times L^2([0,T]^2; \mathbb{R}^{d\times d}) \to \mathbb{R}$ be the loss function defined as*

$$\mathcal{L}_{\mathrm{SOCM}}(u, M) := \mathbb{E}\Big[\frac{1}{T}\int_0^T \left\|u(X_t^v, t) + \sigma(t)^\top a(t, X^v)\right\|^2 \mathrm{d}t \times \alpha(v, X^v, B)\Big], \tag{C.31}$$

where X^v is the process controlled by v (i.e., $dX_t = (b(X_t,t) + \sigma(t)v(X_t,t))\,\mathrm{d}t + \sqrt{\lambda}\sigma(X_t,t)\,\mathrm{d}B_t$ and $X_0 \sim p_0$), $\alpha(v, X^v, B)$ is the importance weight defined in (4.19), and $a(t, X^v)$ is the solution of the ODE

$$\begin{aligned}\frac{da(t)}{dt} &= -\nabla_x b(X_t,t)a(t) - \nabla_x f(X_t,t),\\ a(T) &= \nabla g(X_T),\end{aligned}$$

$\mathcal{L}_{\mathrm{SOCM}}$ has a unique optimum (u^*, M^*), where u^* is the optimal control.

Proof. The proof follows the same structure as that of Theorem 4.3. Instead of plugging the pathwise reparameterization trick (Theorem 4.4) in the right-hand side of (4.21), we make use of (C.12) to evaluate $\nabla_x \mathbb{E}\big[\exp\big(-\lambda^{-1}\int_0^T f(X_t,t)\,dt - \lambda^{-1}g(X_T)\big)|X_0 = x\big]$. □

Lemma C.12 (Adjoint method for SDEs, [Li et al. 2020; Kidger et al. 2021])**.** *Let $X : \Omega \times [0,T] \to \mathbb{R}^d$ be the uncontrolled process defined in (4.6), with initial condition $X_0 = x$. We define the random process $a : \Omega \times [0,T] \to \mathbb{R}^d$ such that for all $\omega \in \Omega$, using the short-hand $a(t) := a(\omega, t)$,*

$$\begin{aligned}\frac{da(t)}{dt} &= -\nabla_x b(X_t,t)a(t) - \nabla_x f(X_t,t),\\ a(T) &= \nabla g(X_T),\end{aligned}$$

Then, we have that

$$\nabla_x \mathbb{E}[\int_0^T f(X_t, t)\, dt + g(X_T)|X_0 = x] = \mathbb{E}[a(0)|X_0 = x],$$
$$\nabla_x \mathbb{E}\big[\exp\big(-\lambda^{-1}\int_0^T f(X_t, t)\, dt - \lambda^{-1} g(X_T)\big)|X_0 = x\big]$$
$$= -\lambda^{-1}\mathbb{E}\big[a(0)\exp\big(-\lambda^{-1}\int_0^T f(X_t, t)\, dt - \lambda^{-1} g(X_T)\big)|X_0 = x\big].$$

Proof. We will use an approach based on Lagrangian multipliers. Define a process $a : \Omega \times [0, T] \to \mathbb{R}^d$ such that for any $\omega \in \Omega$, $a(\omega, \cdot)$ is differentiable. For a given $\omega \in \Omega$, we can write

$$\int_0^T f(X_t(\omega), t)\, dt + g(X_T(\omega))$$
$$= \int_0^T f(X_t(\omega), t)\, dt + g(X_T(\omega)) - \int_0^T \langle a_t(\omega), (dX_t(\omega) - b(X_t(\omega), t)\, dt - \sqrt{\lambda}\sigma(t)\, dB_t)\rangle.$$

By Theorem C.13, we have that

$$\int_0^T \langle a_t(\omega), dX_t(\omega)\rangle = \langle a_T(\omega), X_T(\omega)\rangle - \langle a_0(\omega), X_0(\omega)\rangle - \int_0^T \langle X_t(\omega), \frac{da_t}{dt}(\omega)\rangle\, dt$$

Hence,

$$
\begin{aligned}
&\nabla_x\Big(\int_0^T f(X_t(\omega),t)\,dt + g(X_T(\omega))\Big)\\
&= \nabla_x\Big(\int_0^T f(X_t(\omega),t)\,dt + g(X_T(\omega)) - \langle a_T(\omega), X_T(\omega)\rangle + \langle a_0(\omega), X_0(\omega)\rangle\\
&\qquad + \int_0^T \Big(\langle a_t(\omega), b(X_t(\omega),t)\rangle + \langle \frac{da_t}{dt}(\omega), X_t(\omega)\rangle\Big)\,dt + \sqrt{\lambda}\int_0^T \langle a_t(\omega), \sigma(t)\,dB_t\rangle\Big)\\
&= \int_0^T \nabla_x X_t(\omega)\nabla_x f(X_t(\omega),t)\,dt + \nabla_x X_T(\omega)\nabla_x g(X_T(\omega)) - \nabla_x X_T(\omega)a_T(\omega) + \nabla_x X_0(\omega)a_0(\omega)\\
&\qquad + \int_0^T \Big(\nabla_x X_t(\omega)\nabla_x b(X_t(\omega),t)a_t(\omega) + \nabla_x X_t(\omega)\frac{da_t}{dt}(\omega)\Big)\,dt\\
&= \int_0^T \nabla_x X_t(\omega)\Big(\nabla_x f(X_t(\omega),t) + \nabla_x b(X_t(\omega),t)a_t(\omega) + \frac{da_t}{dt}(\omega)\Big)\,dt\\
&\qquad + \nabla_x X_T(\omega)\Big(\nabla_x g(X_T(\omega)) - a_T(\omega)\Big) + a_0(\omega).
\end{aligned}
$$

In the last line we used that $\nabla_x X_0(\omega) = \nabla_x x = \mathrm{I}$. If choose a such that

$$
\begin{aligned}
da_t(\omega) &= \big(-\nabla_x b(X_t(\omega),t)a_t(\omega) - \nabla_x f(X_t(\omega),t)\big)\,dt - \nabla_x h(X_t(\omega),t)\,dB_t,\\
a_T(\omega) &= \nabla_x g(X_T(\omega)),
\end{aligned}
$$

then we obtain that

$$
\nabla_x\Big(\int_0^T f(X_t(\omega),t)\,dt + g(X_T(\omega))\Big) = a_0(\omega),
$$

and by the Leibniz rule,

$$
\nabla_x \mathbb{E}\Big[\int_0^T f(X_t(\omega),t)\,dt + g(X_T(\omega))\Big] = \mathbb{E}\Big[\nabla_x\Big(\int_0^T f(X_t(\omega),t)\,dt + g(X_T(\omega))\Big)\Big] = \mathbb{E}\Big[a_0(\omega)\Big].
$$

and

$$\begin{aligned}
&\nabla_x \mathbb{E}\Big[\exp\Big(-\lambda^{-1}\int_0^T f(X_t(\omega),t)\,dt - \lambda^{-1} g(X_T(\omega))\Big)\Big] \\
&= -\lambda^{-1}\mathbb{E}\Big[\nabla_x\Big(\int_0^T f(X_t(\omega),t)\,dt + g(X_T(\omega))\Big)\exp\Big(-\lambda^{-1}\int_0^T f(X_t(\omega),t)\,dt - \lambda^{-1} g(X_T(\omega))\Big)\Big] \\
&= -\lambda^{-1}\mathbb{E}\Big[a_0(\omega)\exp\Big(-\lambda^{-1}\int_0^T f(X_t(\omega),t)\,dt - \lambda^{-1} g(X_T(\omega))\Big)\Big].
\end{aligned}$$

□

Lemma C.13 (Stochastic integration by parts, [Oksendal 2013]). *Let*

$$\begin{aligned}
dX_t &= a_t\,dt + b_t\,dW_t^1, \\
dY_t &= f_t\,dt + g_t\,dW_t^2.
\end{aligned}$$

where a_t, b_t, f_t, g_t are continuous square integrable processes adapted to a filtration $(\mathcal{F}_t)_{t\in[0,T]}$, and W^1, W^2 are Brownian motions adapted to the same filtration. Then,

$$X_t Y_t - X_0 Y_0 = \int_0^t X_s\,dY_s + \int_0^t Y_s\,dX_s + \int_0^t b_s g_s \langle dX_s, dY_s\rangle.$$

C.3.5 Proof of Theorem 4.6

Proof. Since the equality (C.3) holds almost surely for the pair (X, B), it must also hold almost surely for (X^v, B^v), which satisfy the same SDE. That is

$$\mathcal{W}(X^v, 0) = V(X_0^v, 0) + \frac{1}{2}\int_0^T \|u^*(X_s^v, s)\|^2\,\mathrm{d}s - \sqrt{\lambda}\int_0^T \langle u^*(X_s^v, s), \mathrm{d}B_s^v\rangle,$$

Thus, we obtain that

$$\begin{aligned}\alpha(v, X^v, B) &= \exp\Big(-\lambda^{-1}\mathcal{W}(X^v, 0) - \lambda^{-1/2}\int_0^T \langle v(X_t^v, t), \mathrm{d}B_t^v\rangle + \frac{\lambda^{-1}}{2}\int_0^T \|v(X_t^v, t)\|^2\,\mathrm{d}t\Big) \\ &= \exp\Big(-\lambda^{-1}V(X_0^v, 0) - \frac{\lambda^{-1}}{2}\int_0^T \|u^*(X_s^v, s)\|^2\,\mathrm{d}s + \lambda^{-1/2}\int_0^T \langle u^*(X_s^v, s), \mathrm{d}B_s^v\rangle \\ &\qquad - \lambda^{-1/2}\int_0^T \langle v(X_t^v, t), \mathrm{d}B_t^v\rangle + \frac{\lambda^{-1}}{2}\int_0^T \|v(X_t^v, t)\|^2\,\mathrm{d}t\Big),\end{aligned}$$

and this is equal to $\exp\big(-V(X_0^v, 0)\big)$ when $v = u^*$. Since we condition on $X_0^v = x_{\text{init}}$, we have obtained that the random variable takes constant value $\exp\big(-V(x_{\text{init}}, 0)\big)$ almost surely, which means that its variance is zero. □

C.3.6 Proof of Theorem 4.7

The proof of (4.24) shows that minimizing $\mathrm{Var}(w; M)$ is equivalent to minimizing

$$\mathbb{E}\Big[\frac{1}{T}\int_0^T \big\|w(t, v, X^v, B, M_t)\big\|^2\,\mathrm{d}t\,\alpha(v, X^v, B)\Big]. \tag{C.32}$$

To optimize with respect to M, it is convenient to reexpress it in terms of $\dot{M} = (\dot{M}_t)_{t\in[0,T]}$ as $M_t(s) = I + \int_t^s \dot{M}_t(s')\,\mathrm{d}s'$. By Fubini's theorem, we have that

$$\begin{aligned}\int_t^T M_t(s)\nabla_x f(X_s^v, s)\,\mathrm{d}s &= \int_t^T \big(I + \int_t^s \dot{M}_t(s')\,\mathrm{d}s'\big)\nabla_x f(X_s^v, s)\,\mathrm{d}s \\ &= \int_t^T \nabla_x f(X_s^v, s)\,\mathrm{d}s + \int_t^T \dot{M}_t(s)\int_s^T \nabla_x f(X_{s'}^v, s')\,\mathrm{d}s'\,\mathrm{d}s,\end{aligned}$$

$$\begin{aligned}&-\int_t^T (M_t(s)\nabla_x b(X_s^v, s) - \dot{M}_t(s))(\sigma^{-1})^\top(s)v(X_s^v, s)\,\mathrm{d}s \\ &= \int_t^T \dot{M}_t(s)(\sigma^{-1})^\top(s)v(X_s^v, s)\,\mathrm{d}s - \int_t^T \dot{M}_t(s)\int_s^T \nabla_x b(X_{s'}^v, s')(\sigma_{s'}^{-1})^\top(s')v(X_s^v, s)\,\mathrm{d}s'\,\mathrm{d}s,\end{aligned}$$

$$
\begin{aligned}
&- \lambda^{1/2} \int_t^T (M_t(s)\nabla_x b(X_s^v, s) - \dot{M}_t(s))(\sigma^{-1})^\top(s)\, \mathrm{d}B_s \\
&= \lambda^{1/2} \Big(\int_t^T \dot{M}_t(s)(\sigma^{-1})^\top(s) v(X_s^v, s)\, \mathrm{d}s - \int_t^T \dot{M}_t(s) \int_s^T \nabla_x b(X_{s'}^v, s')(\sigma^{-1})^\top(s')\, \mathrm{d}B_{s'}\, \mathrm{d}s \Big).
\end{aligned}
$$

Hence, we can rewrite (C.32) as

$$
\begin{aligned}
\mathcal{G}(\dot{M}) = \mathbb{E}\Big[\frac{1}{T} \int_0^T \Big\| \sigma(t)^\top \Big(& \int_t^T \nabla_x f(X_s^v, s)\, \mathrm{d}s + \nabla g(X_T^v) \\
& + \int_t^T \dot{M}_t(s) \Big(\int_s^T \nabla_x f(X_{s'}^v, s')\, \mathrm{d}s' + \nabla g(X_T^v) + (\sigma^{-1})^\top(s) v(X_s^v, s) \\
& - \int_s^T \nabla_x b(X_{s'}^v, s')(\sigma_{s'}^{-1})^\top(s') v(X_s^v, s)\, \mathrm{d}s' - \int_s^T \nabla_x b(X_{s'}^v, s')(\sigma_{s'}^{-1})^\top(s')\, \mathrm{d}B_{s'} \Big) \mathrm{d}s \Big) \Big\|^2 \mathrm{d}t \\
& \times \alpha(v, X^v, B) \Big]
\end{aligned}
$$

The first variation $\frac{\delta \mathcal{G}}{\delta \dot{M}}(\dot{M})$ of $\mathcal{G}$ at $\dot{M}$ is defined as the family $Q = (Q_t)_{t\in[0,T]}$ of matrix-valued functions such that for any collection of matrix-valued functions $P = (P_t)_{t\in[0,T]}$,

$$
\partial_\epsilon \mathcal{V}(\dot{M} + \epsilon P)|_{\epsilon=0} = \lim_{\epsilon \to 0} \frac{\mathcal{V}(\dot{M} + \epsilon P) - \mathcal{V}(M)}{\epsilon} = \langle P, Q \rangle := \int_0^T \int_t^T \langle P_t(s), Q_t(s) \rangle_F \, \mathrm{d}s \, \mathrm{d}t,
$$

where $\dot{M} + \epsilon P := (\dot{M}_t + \epsilon P_t)_{t\in[0,T]}$. Now, note that

$$\begin{aligned}
\partial_\epsilon \mathcal{V}(\dot{M}+\epsilon P)|_{\epsilon=0} &= \partial_\epsilon \mathbb{E}\Big[\frac{1}{T}\int_0^T \Big\|\sigma(t)^\top\Big(\int_t^T \nabla_x f(X_s^v, s)\,\mathrm{d}s + \nabla g(X_T^v) \\
&\quad + \int_t^T (\dot{M}_t(s)+\epsilon P_t(s))\Big(\int_s^T \nabla_x f(X_{s'}^v, s')\,\mathrm{d}s' + \nabla g(X_T^v) + (\sigma^{-1})^\top(s) v(X_s^v, s) \\
&\qquad - \int_s^T \nabla_x b(X_{s'}^v, s')(\sigma_{s'}^{-1})^\top(s') v(X_s^v, s)\,\mathrm{d}s' - \int_s^T \nabla_x b(X_{s'}^v, s')(\sigma_{s'}^{-1})^\top(s')\,\mathrm{d}B_{s'}\Big)\,\mathrm{d}s\Big)\Big\|^2 \mathrm{d}t \\
&\quad \times \alpha(v, X^v, B)\Big]\Big|_{\epsilon=0} \\
&= \mathbb{E}\Big[\frac{2}{T}\int_0^T \Big\langle \sigma(t)\sigma(t)^\top\Big(\int_t^T \nabla_x f(X_s^v, s)\,\mathrm{d}s + \nabla g(X_T^v) \\
&\quad + \int_t^T \dot{M}_t(s)\Big(\int_s^T \nabla_x f(X_{s'}^v, s')\,\mathrm{d}s' + \nabla g(X_T^v) + (\sigma^{-1})^\top(s) v(X_s^v, s) \\
&\qquad - \int_s^T \nabla_x b(X_{s'}^v, s')(\sigma_{s'}^{-1})^\top(s') v(X_s^v, s)\,\mathrm{d}s' - \int_s^T \nabla_x b(X_{s'}^v, s')(\sigma_{s'}^{-1})^\top(s')\,\mathrm{d}B_{s'}\Big)\,\mathrm{d}s\Big), \\
&\qquad \int_t^T P_t(s)\Big(\int_s^T \nabla_x f(X_{s'}^v, s')\,\mathrm{d}s' + \nabla g(X_T^v) + (\sigma^{-1})^\top(s) v(X_s^v, s) \\
&\qquad - \int_s^T \nabla_x b(X_{s'}^v, s')(\sigma_{s'}^{-1})^\top(s') v(X_s^v, s)\,\mathrm{d}s' - \int_s^T \nabla_x b(X_{s'}^v, s')(\sigma_{s'}^{-1})^\top(s')\,\mathrm{d}B_{s'}\Big)\,\mathrm{d}s\Big\rangle\,\mathrm{d}t \\
&\quad \times \alpha(v, X^v, B)\Big]. \qquad \text{(C.33)}
\end{aligned}$$

If we define

$$\begin{aligned}
\chi(s, X^v, B) &:= \int_s^T \nabla_x f(X_{s'}^v, s')\,\mathrm{d}s' + \nabla g(X_T^v) + (\sigma^{-1})^\top(s) v(X_s^v, s) \\
&\qquad - \int_s^T \nabla_x b(X_{s'}^v, s')(\sigma_{s'}^{-1})^\top(s') v(X_s^v, s)\,\mathrm{d}s' - \int_s^T \nabla_x b(X_{s'}^v, s')(\sigma_{s'}^{-1})^\top(s')\,\mathrm{d}B_{s'},
\end{aligned}$$

we can rewrite (C.33) as

$$\begin{aligned}
&\partial_\epsilon \mathcal{V}(\dot{M}+\epsilon P)|_{\epsilon=0} \\
&= \mathbb{E}\Big[\frac{1}{T}\int_0^T \Big\langle \sigma(t)\sigma(t)^\top \Big(\int_t^T \nabla_x f(X_s^v, s)\mathrm{d}s + \nabla g(X_T^v) + \int_t^T M_t(s)\chi(s, X^v, B)\,\mathrm{d}s\Big), \\
&\qquad \int_t^T P_t(s)\chi(s, X^v, B)\mathrm{d}s\Big\rangle \,\mathrm{d}s \times \alpha(v, X^v, B)\Big]
\end{aligned} \tag{C.34}$$

Now let us reexpress equation (C.34) as:

$$\begin{aligned}
&\mathbb{E}\Big[\frac{1}{T}\int_0^T \Big\langle \sigma\sigma^\top(t)\big(\nabla g(X_T^v) + \int_t^T \big(\nabla_x f(X_s^v, s) + \dot{M}_t(s)\chi(s, X^v, B)\big)\,\mathrm{d}s\big), \\
&\qquad \int_t^T P_t(s)\chi(s, X^v, B)\,\mathrm{d}s\Big\rangle \,\mathrm{d}t \times \alpha(v, X^v, B)\Big] \\
&\overset{\text{(i)}}{=} \mathbb{E}\Big[\frac{1}{T}\int_0^T \int_0^s \Big\langle P_t(s)\chi(s, X^v, B), \\
&\qquad \sigma\sigma^\top(t)\big(\nabla g(X_T^v) + \int_t^T (\nabla_x f(X_{s'}^v, s') + \dot{M}_t(s')\chi(s', X^v, B))\,\mathrm{d}s'\big)\Big\rangle \,\mathrm{d}t\,\mathrm{d}s \times \alpha(v, X^v, B)\Big] \\
&\overset{\text{(ii)}}{=} \mathbb{E}\Big[\frac{1}{T}\int_0^T \int_0^s \Big\langle \sigma\sigma^\top(t)\big(\nabla g(X_T^v) + \int_t^T (\nabla_x f(X_{s'}^v, s') + \dot{M}_t(s')\chi(s', X^v, B))\,\mathrm{d}s'\big)\chi(X^v, s, B)^\top, \\
&\qquad P_t(s)\Big\rangle_F \,\mathrm{d}t\,\mathrm{d}s \times \alpha(v, X^v, B)\Big] \\
&= \int_0^T \int_0^s \Big\langle \frac{1}{T}\sigma\sigma^\top(t)\mathbb{E}\Big[\big(\nabla g(X_T^v) + \int_t^T (\nabla_x f(X_{s'}^v, s') + \dot{M}_t(s')\chi(X^v, s', B))\,\mathrm{d}s'\big)\chi(X^v, s, B)^\top \alpha(v, X^v, B)\Big], \\
&\qquad P_t(s)\Big\rangle_F \,\mathrm{d}t\,\mathrm{d}s.
\end{aligned} \tag{C.35}$$

Here, equality (i) holds by Lemma C.14, making the choices $\alpha(t,s) = P_t(s)\chi(X^v, s, B)$, and $\gamma(t) = \sigma\sigma^\top(t)\big(\nabla g(X_T^v) + \int_t^T \big(\nabla_x f(X_s^v, s) + \dot{M}_t(s)\chi(X^v, s, B)\big)\,\mathrm{d}s$. Equality (ii) follows from the fact that for any matrix A and vectors b, c, $\langle Ab, c\rangle = c^\top Ab = \mathrm{Tr}(c^\top Ab) = \mathrm{Tr}(Abc^\top) = \langle B, cb^\top\rangle_F$, where $\langle\cdot,\cdot\rangle_F$ denotes the Frobenius inner product. The first-order necessary condition for optimality states that at the optimal $\dot{M}^*$, the first variation $\frac{\delta \mathcal{G}}{\delta \dot{M}}(\dot{M}^*)$ is zero. In other words, $\partial_\epsilon \mathcal{V}(\dot{M}+\epsilon P)|_{\epsilon=0}$ is zero for any P. Hence, the right-hand side of (C.35) must be zero for any P, which implies that almost

everywhere with respect to $t \in [0,T]$, $s \in [s,T]$,

$$\mathbb{E}\Big[\big(\nabla g(X_T^v) + \int_t^T (\nabla_x f(X_{s'}^v, s') + \dot{M}_t(s')\chi(X^v, s', B))\,\mathrm{d}s'\big)\chi(X^v, s, B)^\top \alpha(v, X^v, B)\Big] = 0.$$

To derive this, we also used that $\sigma(t)$ is invertible by assumption.

Define the integral operator $\mathcal{T}_t : L^2([t,T];\mathbb{R}^{d\times d}) \to L^2([t,T];\mathbb{R}^{d\times d})$ as

$$[\mathcal{T}_t(\dot{M}_t)](s) = \int_t^T \dot{M}_t(s')\mathbb{E}\Big[\chi(X^v, s', B)\chi(X^v, s, B)^\top \times \alpha(v, X^v, B)\Big]\,\mathrm{d}s'$$

If we define $N_t(s) = -\mathbb{E}\Big[\big(\nabla g(X_T^v) + \int_t^T \nabla_x f(X_{s'}^v, s')\,\mathrm{d}s'\big)\chi(X^v, s, B)^\top \times \alpha(v, X^v, B)\Big]$, the problem that we need to solve to find the optimal $\dot{M}_t$ is

$$\mathcal{T}_t(\dot{M}_t) = N_t.$$

This is a Fredholm equation of the first kind.

Lemma C.14. *If $\alpha, \beta : [0,T]\times[0,T] \to \mathbb{R}^d$, $\gamma : [0,T] \to \mathbb{R}^d$, $\delta : [0,T] \to \mathbb{R}^{d\times d}$ are arbitrary integrable functions, we have that*

$$\int_0^T \Big\langle \int_t^T \alpha(t,s)\,\mathrm{d}s, \gamma(t)\Big\rangle\,\mathrm{d}t = \int_0^T \int_0^s \big\langle \alpha(t,s), \gamma(t)\big\rangle\,\mathrm{d}t\,\mathrm{d}s,$$

Proof. We have that:

$$\begin{aligned}
&\int_0^T \int_t^T \big\langle \alpha(t,s), \gamma(t)\big\rangle\,\mathrm{d}s\,\mathrm{d}t \overset{\text{(i)}}{=} \int_0^T \int_0^{T-t} \big\langle \alpha(t, T-s), \gamma(t)\big\rangle\,\mathrm{d}s\,\mathrm{d}t \\
&\overset{\text{(ii)}}{=} \int_0^T \int_0^t \big\langle \alpha(T-t, T-s), \gamma(T-t)\big\rangle\,\mathrm{d}s\,\mathrm{d}t \overset{\text{(iii)}}{=} \int_0^T \int_s^T \big\langle \alpha(T-t, T-s), \gamma(T-t)\big\rangle\,\mathrm{d}t\,\mathrm{d}s \\
&\overset{\text{(iv)}}{=} \int_0^T \int_{T-s}^T \big\langle \alpha(T-t, s), \gamma(T-t)\big\rangle\,\mathrm{d}t\,\mathrm{d}s \overset{\text{(v)}}{=} \int_0^T \int_0^s \big\langle \alpha(t,s), \gamma(t)\big\rangle\,\mathrm{d}t\,\mathrm{d}s
\end{aligned}$$

Here, in equalities (i), (ii), (iv) and (v) we make changes of variables of the form $t \mapsto T - t$,

$s \mapsto T - s$, $s' \mapsto T - s'$. In equality (iii) we use Fubini's theorem. □

C.4 Control warm-starting

We introduce the *Gaussian warm-start*, a control warm-start strategy that we adapt from [Liu et al. 2023a], and that we use in our experiments in Figure 4.3. Their work tackles generalized Schrödinger bridge problems, which are different from the control setting in that the final distribution is known and there is no terminal cost. The following proposition, that provides an analytic expression of the control needed for the density of the process to be Gaussian at all times, is the foundation of our method.

Proposition C.15. *Given $Z \sim N(0, I)$ define the random process Y as*

$$Y_t = \mu(t) + \tilde{\Gamma}(t)Z, \qquad \text{where } \mu(t) \in \mathbb{R}^d,\ \tilde{\Gamma}(t) = \sqrt{t}\Gamma(t) \in \mathbb{R}^{d\times d}. \tag{C.36}$$

Define the control $u : \mathbb{R}^d \times [0, T] \to \mathbb{R}^d$ as

$$u(x,t) = \sigma(t)^{-1}\big(\partial_t\mu(t) + \big(\big(\partial_t\Gamma(t)\big)\Gamma(t)^{-1} + \frac{I - (\sigma\sigma^\top)(t)(\Sigma\Sigma^\top)^{-1}(t)}{2t}\big)(x - \mu(t)) - b(x,t)\big). \tag{C.37}$$

Then, if $\Gamma_0 = \sigma(0)$, the controlled process X^u defined in equation (4.2) has the same marginals as Y. That is, for all $t \in [0, T]$, $\mathrm{Law}(Y_t) = \mathrm{Law}(X^u_t)$.

Proof. Following [Liu et al. 2023a], we have that

$$\partial_t X_t = \partial_t\mu_t + \partial_t\tilde{\Gamma}(t)Z = \partial_t\mu(t) + (\partial_t\tilde{\Gamma}(t))\tilde{\Gamma}(t)^{-1}(X_t - \mu(t)),$$
$$\nabla \log p_t(x) = -\tilde{\Sigma}(t)^{-1}(x - \mu(t)), \qquad \tilde{\Sigma}(t) = \tilde{\Gamma}(t)\tilde{\Gamma}(t)^\top.$$

Now, p_t satisfies the continuity equation equation

$$\partial_t p_t = -\nabla \cdot ((\partial_t \mu(t) + (\partial_t \tilde{\Gamma}(t))\tilde{\Gamma}(t)^{-1}(x - \mu(t)))p_t) \tag{C.38}$$

Let $D(t) = \frac{1}{2}\sigma(t)\sigma(t)^\top$. We want to reexpress (C.38) as a Fokker-Planck equation of the form

$$\begin{aligned}
\partial_t p_t &= -\nabla \cdot (v(x,t)p_t) + \sum_{i=1}^{d}\sum_{j=1}^{d} \partial_i \partial_j (D_{ij}(t)p_t) = -\nabla \cdot (v(x,t)p_t) + \sum_{i=1}^{d} \partial_i \sum_{j=1}^{d} (D_{ij}(t)\partial_j p_t) \\
&= -\nabla \cdot (v(x,t)p_t) + \nabla \cdot (D(t)\nabla p_t) = -\nabla \cdot (v(x,t)p_t) + \nabla \cdot (D(t)\nabla \log p_t(x) p_t) \\
&= -\nabla \cdot ((v(x,t) - D(t)\nabla \log p_t(x))p_t).
\end{aligned}$$

Hence, we need that

$$\begin{aligned}
v(x,t) - D(t)\nabla \log p_t &= \partial_t \mu(t) + (\partial_t \tilde{\Gamma}(t))\tilde{\Gamma}(t)^{-1}(x - \mu(t)), \\
\implies v_t(x) &= \partial_t \mu(t) + ((\partial_t \tilde{\Gamma}(t))\tilde{\Gamma}(t)^{-1}(x - \mu(t)) + \frac{(\sigma\sigma^\top)(t)}{2}\nabla \log p_t(x) \\
&= \partial_t \mu(t) + (\partial_t \tilde{\Gamma}(t))\tilde{\Gamma}(t)^{-1}(x - \mu(t)) - \frac{(\sigma\sigma^\top)(t)}{2}\Sigma(t)^{-1}(x - \mu(t)).
\end{aligned}$$

If we let $\tilde{\Gamma}(t) = \Gamma(t)\sqrt{t}$, then $\tilde{\Sigma}(t) = t\Gamma(t)\Gamma(t)^\top = t\Sigma(t)$ and $\partial_t \tilde{\Gamma}(t) = \partial_t \Gamma(t)\sqrt{t} + \frac{\Gamma(t)}{2\sqrt{t}}$. That is,

$$\begin{aligned}
v(x,t) &= \partial_t \mu(t) + \Big(\partial_t \Gamma(t)\sqrt{t} + \frac{\Gamma(t)}{2\sqrt{t}}\Big)\frac{\Gamma(t)^{-1}}{\sqrt{t}}(x - \mu(t)) - \frac{(\sigma\sigma^\top)(t)}{2}\frac{\Sigma(t)^{-1}}{t}(x - \mu(t)) \\
&= \partial_t \mu(t) + \big(\partial_t \Gamma(t)\big)\Gamma(t)^{-1}(x - \mu(t)) + \frac{1}{2t}(x - \mu(t)) - \frac{(\sigma\sigma^\top)(t)\Sigma(t)^{-1}}{2t}(x - \mu(t))
\end{aligned}$$

For v to be finite at $t = 0$, we need that $(\sigma\sigma^\top)(0)\Sigma(0)^{-1} = I$, which holds, for example, if $\Gamma(0) =$

$\sigma(0)$. Also, to match the form of (4.2), we need that

$$
\begin{aligned}
& v(x,t) = b(x,t) + \sigma(t)u(x,t), \\
\implies & u(x,t) = \sigma(t)^{-1}\big(\partial_t \mu_t + \big(\big(\partial_t \Gamma(t)\big)\Gamma(t)^{-1} + \frac{I - (\sigma\sigma^\top)(t)\Sigma(t)^{-1}}{2t}\big)(x - \mu_t) - b(x,t)\big).
\end{aligned}
$$

□

The warm-start control is computed as the solution of a *Restricted Gaussian Stochastic Optimal Control* problem, where we constrain the space of controls to those that induce Gaussian paths as described in Theorem C.15. In practice, we learn a linear spline $\mu = (\mu^{(b)})_{b=0}^{\mathcal{B}}$, where $\mu^{(b)} \in \mathbb{R}^d$, and a linear spline $\Gamma = (\Gamma^{(b)})_{b=0}^{\mathcal{B}}$, where $\Gamma^{(b)} \in \mathbb{R}^{d\times d}$. These linear splines take the role of $\mu(t)$ and $\Sigma(t)$ in (C.36). Given splines μ and Γ, we obtain the warm-start control using (C.37); for a given $t \in [0,T)$, if we let $b_- = \lfloor \mathcal{B}t/T \rfloor$, $b_+ = b_- + 1$, $\Delta = T/\mathcal{B}$, we have that

$$
\widehat{\mu}(t) = \frac{(t - b_-\Delta)\mu^{(b_+)} + (b_+\Delta - t)\mu^{(b_-)}}{\Delta}, \qquad \widehat{\partial_t \mu}(t) = \frac{\mu^{(b_+)} - \mu^{(b_-)}}{\Delta}, \tag{C.39}
$$

$$
\widehat{\Gamma}(t) = \frac{(t - b_-\Delta)\Gamma^{(b_+)} + (b_+\Delta - t)\Gamma^{(b_-)}}{\Delta}, \qquad \widehat{\partial_t \Gamma}(t) = \frac{\Gamma^{(b_+)} - \Gamma^{(b_-)}}{\Delta}, \tag{C.40}
$$

$$
\hat{u}(x,t) = \sigma(t)^{-1}\big(\widehat{\partial_t \mu}(t) + \big(\widehat{\partial_t \Gamma}(t)\widehat{\Gamma}(t)^{-1} + \frac{I - (\sigma\sigma^\top)(t)(\widehat{\Sigma\Sigma}^\top)^{-1}(t)}{2t}\big)(x - \widehat{\mu}(t)) - b(x,t)\big). \tag{C.41}
$$

Algorithm 7 provides a method to learn the splines μ, Γ. It is a stochastic optimization algorithms in which the spline parameters are updated by sampling Y_t in (C.36) at different times, computing the control cost relying on (C.41), and taking its gradient.

Once we have access to the restricted control $\hat{u}_N$, we can warm-start the control in Algorithms 3 and 4 by introducing $\hat{u}_N$ as an offset. That is, we parameterize the control as $u_\theta = \hat{u}_N + \tilde{u}_\theta$.

Algorithm 7 Restricted Gaussian Stochastic Optimal Control

Input: State cost $f(x,t)$, terminal cost $g(x)$, covariance matrix $\sigma(t)$, base drift $b(x,t)$, noise level λ, number of iterations N, batch size m, number of time steps K, number of spline knots $\mathcal{B}$, initial mean spline knots $\mu_0 = (\mu_0^{(b)})_{b=0}^{\mathcal{B}}$, initial noise spline knots $\Gamma_0 = (\Gamma_0^{(b)})_{b=0}^{\mathcal{B}}$.

for $n = 0 : (N-1)\}$ **do**
  Sample m i.i.d. variables $(Z_i)_{i=1}^n \sim N(0,I)$ and m times $(t_i)_{i=1}^n \sim \mathrm{Unif}([0,T])$.
  for $j = 0 : K$ **do**
    Set $t_j = jT/K$, and compute $\widehat{\mu}_n(t_j)$, $\widehat{\partial_t \mu}_n(t_j)$, $\widehat{\Gamma}_n(t_j)$, $\widehat{\partial_t \Gamma}_n(t_j)$ according to (C.39), (C.40) using μ_n, Γ_n
    for $i = 1 : m$ **do** compute $Y_{ij} = \hat{\mu}(t_j) + \sqrt{t_j}\widehat{\Gamma}(t_j) Z_i$ and $\hat{u}_n(Y_{ij}, t_j)$ using (C.41);
  end
  Compute $\hat{\mathcal{L}}_{\mathrm{RGSOC}}(\mu_n, \Gamma_n) = \frac{1}{m}\sum_{i=1}^m \left(\frac{T}{K}\sum_{j=0}^{K-1}\left(\frac{1}{2}\|\hat{u}(Y_{ij},t_j)\|^2 + f(Y_{ij},t_j)\right) + g(Y_{iK})\right)$
  Compute the gradient of $\hat{\mathcal{L}}_{\mathrm{RGSOC}}(\mu_n,\Gamma_n)$ with respect to the spline parameters (μ_n,Γ_n).
  Obtain μ_{n+1}, Γ_{n+1} with via an Adam update on μ_n, Γ_n resp. (or another stochastic algorithm)
end

Output: Learned splines μ_N, Γ_N, control $\hat{u}_N$

C.5 Experimental details and additional plots

C.5.1 Experimental details

The control L^2 error curves show the following quantity:

$$\begin{aligned}&\mathbb{E}_{t,\mathbb{P}^{u^*}}[\|u^*(X_t^{u^*},t) - u(X_t^{u^*},t)\|^2 e^{-\lambda^{-1}V(X_0^{u^*},0)}]/\mathbb{E}_{t,\mathbb{P}^{u^*}}[e^{-\lambda^{-1}V(X_0^{u^*},0)}]\\ &= \mathbb{E}_{t,\mathbb{P}^v}[\|u^*(X_t^v,t) - u(X_t^v,t)\|^2 \alpha(v,X^v,B)]/\mathbb{E}_{t,\mathbb{P}^v}[\alpha(v,X^v,B)].\end{aligned}$$

The equality holds by the Girsanov theorem. In practice, we use Monte Carlo estimates of the right-hand side, which we can easily compute during training because we sample trajectories controlled by X^v.

For all losses and all settings, we train the control using Adam with learning rate 1×10^{-4}. For SOCM, we train the reparametrization matrices using Adam with learning rate 1×10^{-2}. We use batch size $m = 128$ unless otherwise specified. When used, we run the warm-start algorithm (Algorithm 7) with $\mathcal{B} = 20$ knots, $K = 200$ time steps, and batch size $m = 512$, and we use Adam

with learning rate 3×10^{-4} for $N = 60000$ iterations.

Quadratic Ornstein-Uhlenbeck The choices for the functions of the control problem are:

$$b(x,t) = Ax, \quad f(x,t) = x^\top P x, \quad g(x) = x^\top Q x, \quad \sigma(t) = \sigma_0.$$

where Q is a positive definite matrix. Control problems of this form are better known as linear quadratic regulator (LQR) and they admit a closed form solution [Van Handel 2007, Thm. 6.5.1]. The optimal control is given by:

$$u_t^*(x) = -2\sigma_0^\top F_t x,$$

where F_t is the solution of the Ricatti equation

$$\frac{dF_t}{dt} + A^\top F_t + F_t A - 2F_t \sigma_0 \sigma_0^\top F_t + P = 0$$

with the final condition $F_T = Q$. Within the Quadratic OU class, we consider two settings:

- Easy: We set $d = 20$, $A = 0.2I$, $P = 0.2I$, $Q = 0.1I$, $\sigma_0 = I$, $\lambda = 1$, $T = 1$, $x_{\text{init}} = 0.5N(0, I)$. We do not use warm-start for any algorithm. We take $K = 50$ time discretization steps, and we use random seed 0.
- Hard: We set $d = 20$, $A = I$, $P = I$, $Q = 0.5I$, $\sigma_0 = I$, $\lambda = 1$, $T = 1$, $x_{\text{init}} = 0.5N(0, I)$. We use the *Gaussian warm-start* (section C.4). We take batch size $m = 64$ and $K = 150$ time discretization steps, and we use random seed 0.

Linear Ornstein-Uhlenbeck The functions of the control problem are chosen as follows:

$$b(x,t) = Ax, \quad f(x,t) = 0, \quad g(x) = \langle \gamma, x \rangle, \quad \sigma(t) = \sigma_0.$$

The optimal control for this class of problems is given by [Nüsken and Richter 2021, Sec. A.4]:

$$u_t^*(x) = -\sigma_0^\top e^{A^\top (T-t)} \gamma.$$

We use exactly the same functions as [Nüsken and Richter 2021]: we sample $(\xi_{ij})_{1\le i,j\le d}$ once at the beginning of the simulation, and set:

$$d = 10, \quad A = -I + (\xi_{ij})_{1\le i,j\le d}, \quad \gamma = \mathbb{1}, \quad \sigma_0 = I + (\xi_{ij})_{1\le i,j\le d},$$
$$T = 1, \quad \lambda = 1, \quad x_{\text{init}} = 0.5N(0, I).$$

We take $K = 100$ time discretization steps, and we use random seed 0.

Double Well We also use exactly the same functions as [Nüsken and Richter 2021], which are the following:

$$b(x,t) = -\nabla\Psi(x), \quad \Psi(x) = \sum_{i=1}^d \kappa_i (x_i^2 - 1)^2, \quad g(x) = \sum_{i=1}^d \nu_i (x_i^2 - 1)^2, \quad \sigma_0 = \mathrm{I},$$

where $d = 10$, and $\kappa_i = 5$, $\nu_i = 3$ for $i \in \{1, 2, 3\}$ and $\kappa_i = 1$, $\nu_i = 1$ for $i \in \{4, \dots, 10\}$. We set $T = 1$, $\lambda = 1$ and $x_{\text{init}} = 0$. We take $K = 200$ time discretization steps, and we use random seed 1.

C.5.2 Additional plots

Figure C.1 shows the control objective (4.1) for the four settings. The error bars for the control objective plots show the confidence intervals for ± one standard deviation. As expected, SOCM also obtains the lowest values for the control objective, up to the estimation error.

Figure C.2 shows the normalized standard deviation of the importance weight for the learned control u: $\sqrt{\mathrm{Var}[\alpha(u, X^u, B)]}/\mathbb{E}[\alpha(u, X^u, B)]$. By Theorem 4.6, when $X_0^u = x_{\text{init}}$ for an arbitrary x_{init} (which is the case for all our experiments), this quantity is zero for the optimal control u^*.

Hence, the normalized standard deviation of α is an alternative metric to measure the optimality of the learned control.

Figure C.3 shows an exponential moving average of the norm squared of the gradient for Linear OU and Double Well. For Linear OU, the minimum gradient norm is achieved by the adjoint method, while for Double Well it is achieved by the cross entropy loss. The training instabilities of the adjoint method become apparent as well. Interestingly, in both settings the algorithms with smallest gradients are not SOCM, which is the algorithm with smallest error as shown in Figure 4.2. Understanding this phenomenon is outside of the scope of this paper.

Figure C.4 shows that the instabilities of the adjoint method are inherent to the loss, because they also appear at small learning rates: 3×10^{-5} is smaller than the learning rates typically used for Adam, which hover from 1×10^{-4} to 1×10^{-3}.

Figure C.5 shows plots of the control L^2 error, the norm squared of the gradient, and the control objective for the Quadratic OU (hard) setting, without using warm-start, i.e., with the same algorithms plotted in Figure 4.1 and Figure 4.2. For over 30000 iterations, SOCM and cross entropy have large gradient variance and substantially larger control objective than the adjoint, log-variance and moment losses. This can be attributed to the large variance of the factor $\alpha(v, X^v, B)$, which is present in the SOCM and the cross entropy losses. Eventually, both the gradient variance and the error of SOCM drop below those of existing losses.

Figure C.6 shows the value of the training loss for SOCM and its two ablations: SOCM with constant $M_t = I$, and SOCM-Adjoint. For all such algorithms, the training loss is the sum of the L^2 error of the learned control u, and the expected conditional variance of the matching vector field w. Thus, the difference between the training loss plots and the L^2 error plots is the expected conditional variance of w. We observe that the expected conditional variance in the Quadratic OU setting is orders of magnitude smaller for SOCM than for its two ablations. For Linear OU, SOCM and SOCM-adjoint have similar expected conditional variance, and a possible explanation is that the Linear OU setting is very simple. In the Double Well setting, the SOCM-adjoint training

loss curve has spikes that are probably caused by instabilities of the adjoint method. These spikes can be attributed mostly to the expected conditional variance term, since the corresponding L^2 error curve in Figure 4.2 does not present them.

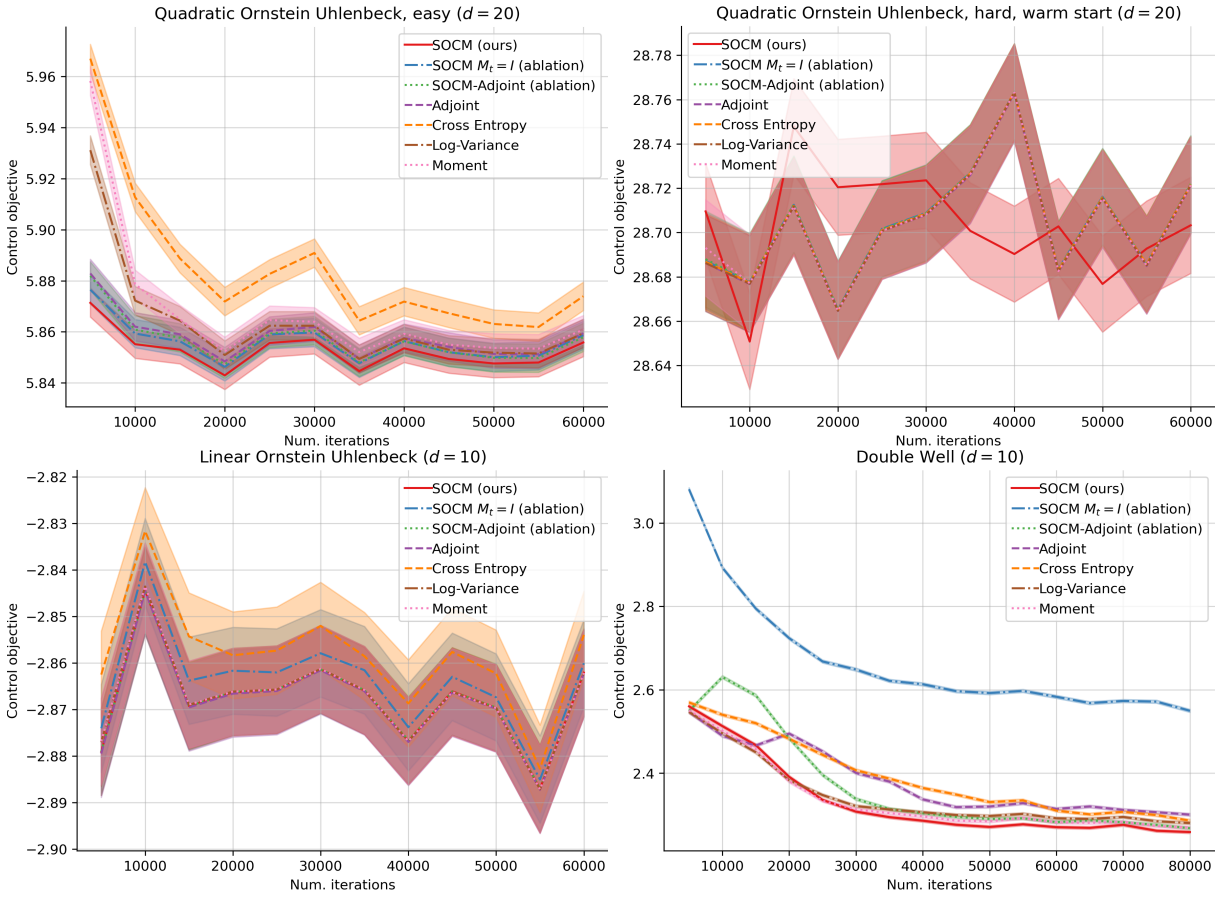


Figure C.1: Plots of the control objective for the four settings.

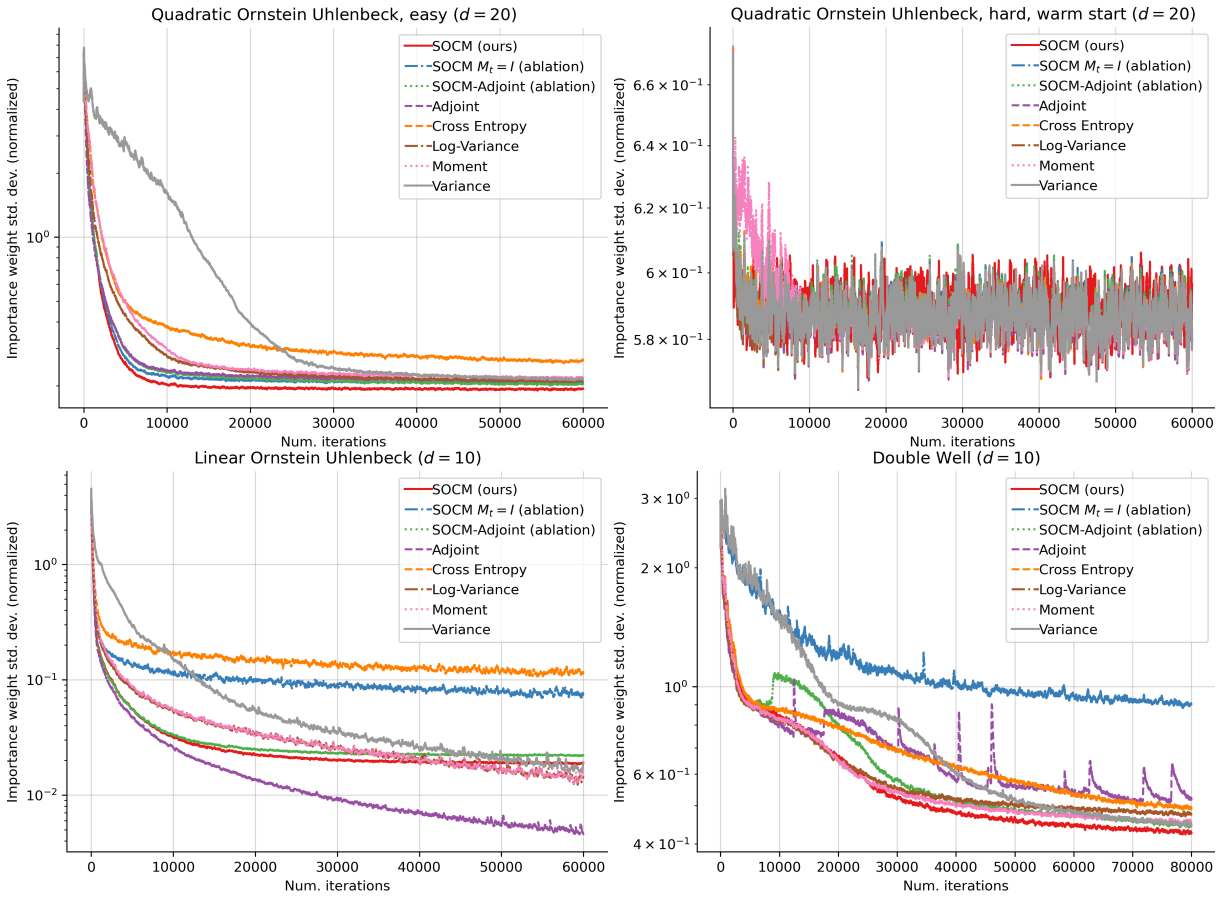


Figure C.2: Plots of the normalized standard deviation of the importance weights, which is given by $\sqrt{\mathrm{Var}[\alpha(u, X^u, B)]}/\mathbb{E}[\alpha(u, X^u, B)]$.

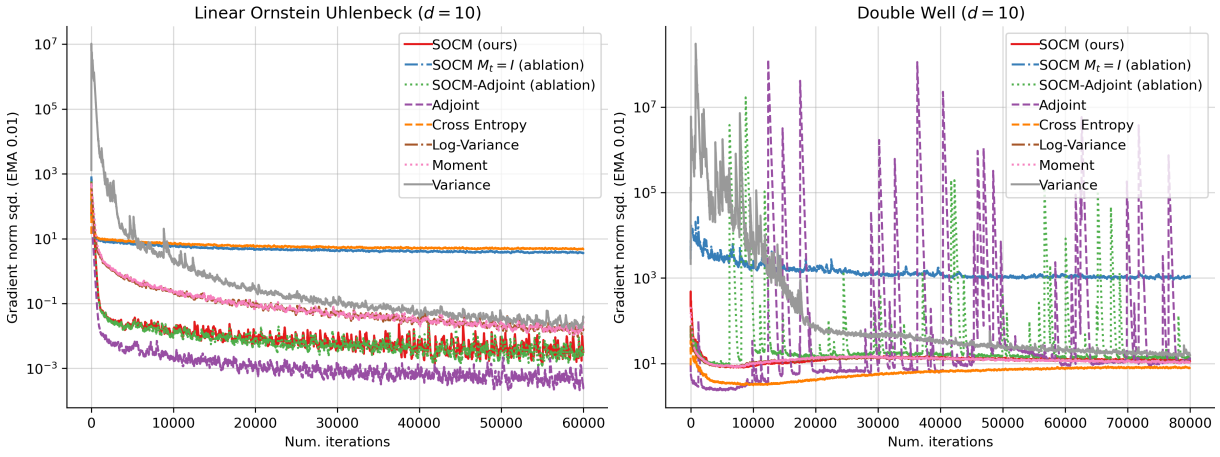


Figure C.3: Plots of the norm squared of the gradient for the Linear Ornstein Uhlenbeck and Double Well settings.

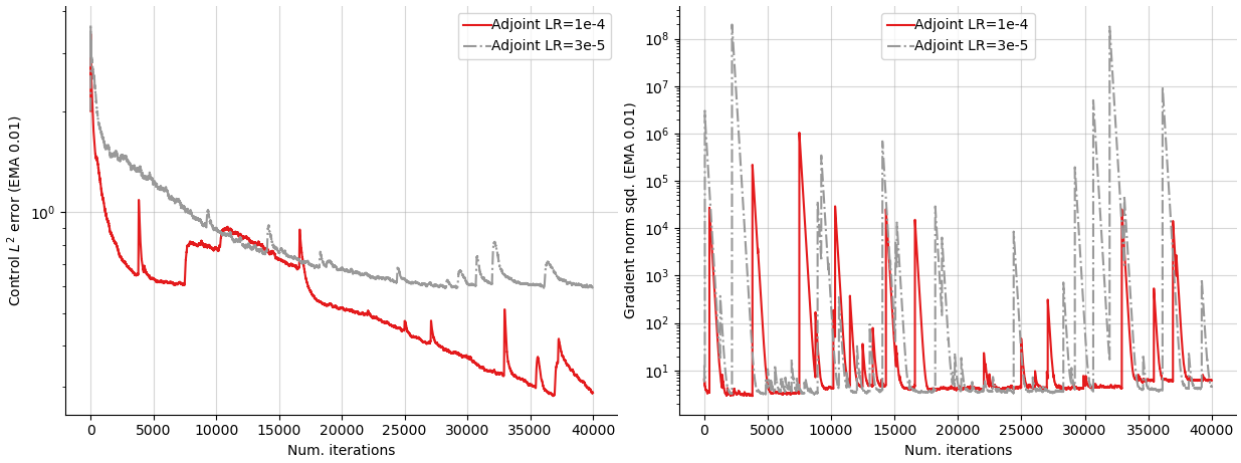


Figure C.4: Plots of the control L^2 error and the norm squared of the gradient for the adjoint method on Double Well, for two different values of the Adam learning rate. The instabilities of the adjoint method persist for small learning rates, signaling an inherent issue with the loss.

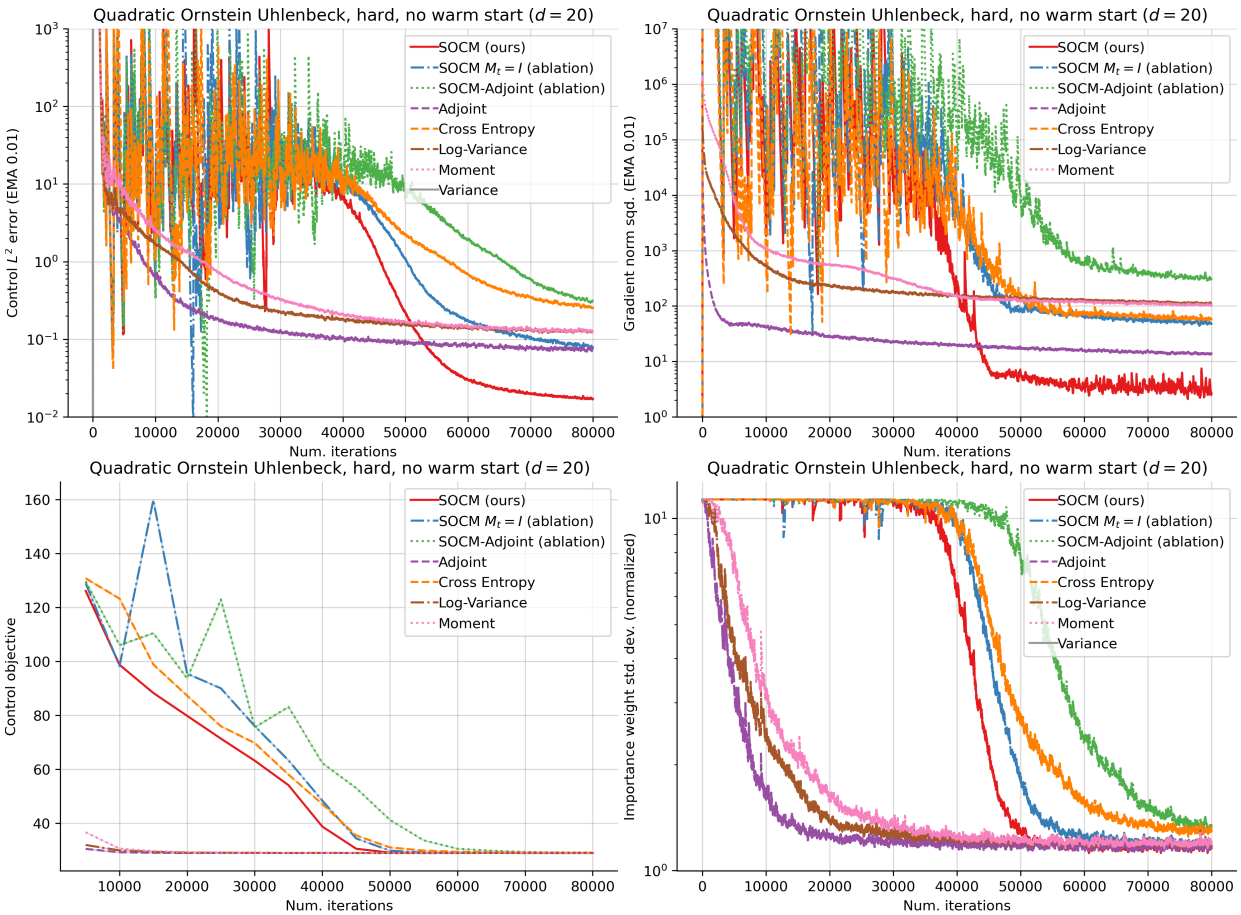


Figure C.5: Plots of the control L^2 error, the norm squared of the gradient, and the control objective for the Quadratic Ornstein-Uhlenbeck (hard) setting, without using warm-start.

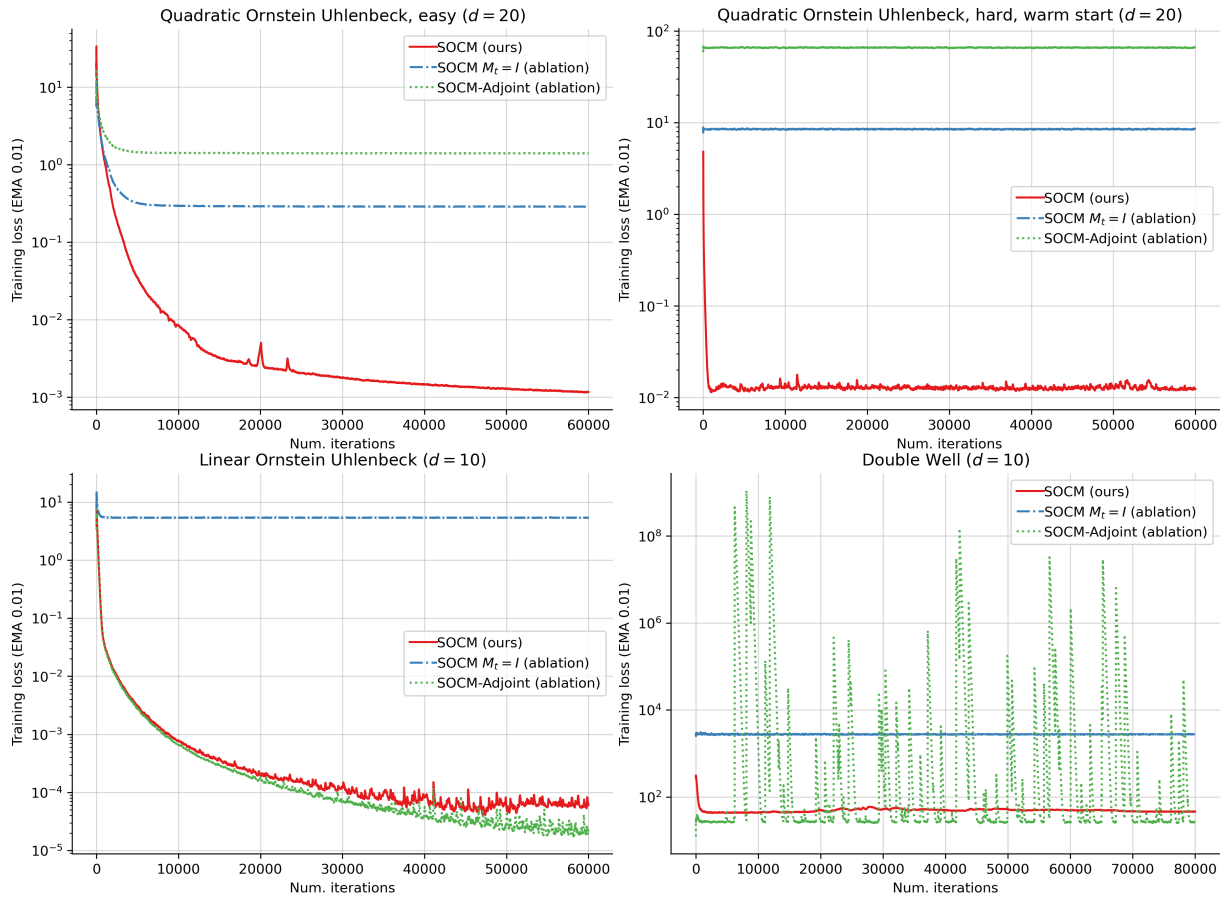


Figure C.6: Plots of the training loss for SOCM and its two ablations: SOCM with constant $M_t = I$, and SOCM-Adjoint.

Bibliography

Adolphs, L., Daneshmand, H., Lucchi, A., and Hofmann, T. (2018). Local saddle point optimization: A curvature exploitation approach. *arXiv preprint arXiv:1805.05751*.

Albergo, M. S., Boffi, N. M., and Vanden-Eijnden, E. (2023). Stochastic interpolants: A unifying framework for flows and diffusions. *arXiv preprint arXiv:2303.08797*.

Albergo, M. S. and Vanden-Eijnden, E. (2023). Building normalizing flows with stochastic interpolants. *International Conference on Learning Representations*.

Altschuler, J., Weed, J., and Rigollet, P. (2017). Near-linear time approximation algorithms for optimal transport via Sinkhorn iteration. In *Advances in Neural Information Processing Systems 30*.

Amos, B. (2023). On amortizing convex conjugates for optimal transport. *International Conference on Learning Representations*.

Arjovsky, M., Chintala, S., and Bottou, L. (2017). Wasserstein generative adversarial networks. In *International conference on machine learning*, pages 214–223. PMLR.

Arora, S., Ge, R., Liang, Y., Ma, T., and Zhang, Y. (2017). Generalization and equilibrium in generative adversarial nets (GANs). In *Proceedings of the 34th International Conference on Machine Learning-Volume 70*, pages 224–232. JMLR. org.

Asadulaev, A., Korotin, A., Egiazarian, V., and Burnaev, E. (2022). Neural optimal transport with general cost functionals. *arXiv preprint arXiv:2205.15403*.

Balandat, M., Krichene, W., Tomlin, C., and Bayen, A. (2016). Minimizing regret on reflexive banach spaces and nash equilibria in continuous zero-sum games. In *Advances in Neural Information Processing Systems*, pages 154–162.

Balduzzi, D., Racanière, S., Martens, J., Foerster, J., Tuyls, K., and Graepel, T. (2018). The mechanics of *n*-player differentiable games. In *Proceedings of the International Conference on Machine Learning*.

Baňas, L., Dawid, H., Randrianasolo, T. A., Storn, J., and Wen, X. (2022). Numerical approximation of a system of hamilton–jacobi–bellman equations arising in innovation dynamics. *Journal of Scientific Computing*, 92.

Beck, C., Becker, S., Grohs, P., Jaafari, N., and Jentzen, A. (2018). Solving stochastic differential equations and Kolmogorov equations by means of deep learning. *arXiv:1806.00421*.

Beck, C., Hornung, F., Hutzenthaler, M., Jentzen, A., and Kruse, T. (2019). Overcoming the curse of dimensionality in the numerical approximation of Allen-Cahn partial differential equations via truncated full-history recursive multilevel Picard approximations. *arXiv:1907.06729*.

Becker, S., Cheridito, P., and Jentzen, A. (2019). Deep optimal stopping. *Journal of Machine Learning Research*, 20.

Belloni, A., Piroddi, L., and Prandini, M. (2016). A stochastic optimal control solution to the energy management of a microgrid with storage and renewables. In *2016 American Control Conference (ACC)*, pages 2340–2345.

Benamou, J.-D. and Brenier, Y. (2000). A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem. *Numerische Mathematik*, 84(3):375–393.

Bierkens, J. and Kappen, H. J. (2014). Explicit solution of relative entropy weighted control. *Systems & Control Letters*, 72:36–43.

Bogachev, V. I., Krylov, N. V., and Röckner, M. (2001). On regularity of transition probabilities and invariant measures of singular diffusions under minimal conditions. *Communications in Partial Differential Equations*, 26(11-12):2037–2080.

Bonnans, J., Ottenwaelter, E., and Zidani, H. (2004). A fast algorithm for the two dimensional hjb equation of stochastic control. *M2AN. Mathematical Modelling and Numerical Analysis. ESAIM, European Series in Applied and Industrial Mathematics*, 38.

Bu, L., Babu, R., De Schutter, B., et al. (2008). A comprehensive survey of multi-agent reinforcement learning. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 38(2):156–172.

Bunne, C., Krause, A., and Cuturi, M. (2022). Supervised training of conditional Monge maps. *arXiv preprint arXiv:2206.14262*.

Bunne, C., Stark, S. G., Gut, G., del Castillo, J. S., Lehmann, K.-V., Pelkmans, L., Krause, A., and Rätsch, G. (2021). Learning single-cell perturbation responses using neural optimal transport. *bioRxiv*.

Caffarelli, L. (1992). The regularity of mappings with a convex potential. *Journal of the American Mathematical Society*, 5:99–104.

Calzola, E., Carlini, E., Dupuis, X., and Silva, F. (2022). A semi-Lagrangian scheme for Hamilton–Jacobi–Bellman equations with oblique derivatives boundary conditions. *Numerische Mathematik*, page 153.

Carlini, E., Festa, A., and Forcadel, N. (2020). A semi-Lagrangian scheme for Hamilton–Jacobi–Bellman equations on networks. *SIAM J. Numer. Anal.*, 58(6):3165–3196.

Carmona, R. (2016). *Lectures on BSDEs, stochastic control, and stochastic differential games with financial applications*, volume 1. SIAM.

Carmona, R., Delarue, F., et al. (2018). *Probabilistic Theory of Mean Field Games with Applications I-II*. Springer.

Carmona, R. and Laurière, M. (2021). Convergence analysis of machine learning algorithms for the numerical solution of mean field control and games i: The ergodic case. *SIAM Journal on Numerical Analysis*, 59(3):1455–1485.

Carmona, R. and Laurière, M. (2022). Convergence analysis of machine learning algorithms for the numerical solution of mean field control and games: Ii—the finite horizon case. *The Annals of Applied Probability*, 32(6):4065–4105.

Chan-Wai-Nam, Q., Mikael, J., and Warin, X. (2019). Machine learning for semilinear PDEs. *Journal of Scientific Computing*, 79(3):1667–1712.

Chaudhari, P., Oberman, A., Osher, S., Soatto, S., and Carlier, G. (2018). Deep relaxation: partial differential equations for optimizing deep neural networks. *Research in the Mathematical Sciences*, 5(3):30.

Chen, R. T. Q. (2018). torchdiffeq.

Chen, R. T. Q. and Lipman, Y. (2023). Riemannian flow matching on general geometries. *International Conference on Machine Learning*.

Chen, R. T. Q., Rubanova, Y., Bettencourt, J., and Duvenaud, D. K. (2018). Neural ordinary differential equations. *Advances in neural information processing systems*, 31.

Chirikjian, G. (2009). *Stochastic models, information theory, and Lie groups*. Number v. 1 in Applied and numerical harmonic analysis. Birkhäuser.

Chizat, L. (2019). Sparse optimization on measures with over-parameterized gradient descent. *arXiv preprint arXiv:1907.10300v1*.

Chizat, L. and Bach, F. (2018). On the global convergence of gradient descent for over-parameterized models using optimal transport. In *Advances in neural information processing systems*, pages 3036–3046.

Chizat, L., Peyré, G., Schmitzer, B., and Vialard, F.-X. (2015). Unbalanced optimal transport: Dynamic and Kantorovich formulation. *arXiv preprint arXiv:1508.05216*.

Chizat, L., Peyré, G., Schmitzer, B., and Vialard, F.-X. (2018). An interpolating distance between optimal transport and Fisher–Rao metrics. 18(1):1–44. tex.day: 01.

Chrabaszcz, P., Loshchilov, I., and Hutter, F. (2017). A downsampled variant of ImageNet as an alternative to the Cifar datasets. *arXiv preprint arXiv:1707.08819*.

Cuesta-Albertos, J., Matrán, C., and Tuero-Diaz, A. (1997). Optimal transportation plans and convergence in distribution. *Journal of Multivariate Analysis*, 60(1):72–83.

Cuturi, M. (2013). Sinkhorn distances: Lightspeed computation of optimal transport. *Advances in neural information processing systems*, 26.

Daskalakis, C., Goldberg, P. W., and Papadimitriou, C. H. (2009). The complexity of computing a Nash equilibrium. *SIAM Journal on Computing*, 39(1):195–259.

Debrabant, K. and Jakobsen, E. R. (2013). Semi-lagrangian schemes for linear and fully non-linear diffusion equations. *Mathematics of Computation*, 82(283):1433–1462.

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., and Fei-Fei, L. (2009). Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee.

Dhariwal, P. and Nichol, A. (2021). Diffusion models beat gans on image synthesis. *Advances in Neural Information Processing Systems*, 34:8780–8794.

Dinh, L., Sohl-Dickstein, J., and Bengio, S. (2016). Density estimation using real nvp. *arXiv preprint arXiv:1605.08803*.

Domingo-Enrich, C., Han, J., Amos, B., Bruna, J., and Chen, R. T. Q. (2023). Stochastic optimal control matching.

Domingo-Enrich, C., Jelassi, S., Mensch, A., Rotskoff, G., and Bruna, J. (2020). A mean-field analysis of two-player zero-sum games. In *Advances in Neural Information Processing Systems*, volume 33, pages 20215–20226. Curran Associates, Inc.

E, W., Han, J., and Jentzen, A. (2017). Deep learning-based numerical methods for high-dimensional parabolic partial differential equations and backward stochastic differential equations. *Communications in Mathematics and Statistics*, 5(4):349–380.

E, W., Han, J., and Jentzen, A. (2021). Algorithms for solving high dimensional pdes: from nonlinear monte carlo to machine learning. *Nonlinearity*, 35(1):278.

Eberle, A., Guillin, A., and Zimmer, R. (2019). Quantitative Harris-type theorems for diffusions and McKean-Vlasov processes. *Trans. Amer. Math. Soc.*, 371:7135–7173.

Fan, J., Liu, S., Ma, S., Chen, Y., and Zhou, H. (2021). Scalable computation of monge maps with general costs. *arXiv preprint arXiv:2106.03812*, page 4.

Fatras, K., Zine, Y., Flamary, R., Gribonval, R., and Courty, N. (2019). Learning with minibatch Wasserstein: asymptotic and gradient properties. *arXiv preprint arXiv:1910.04091*.

Fatras, K., Zine, Y., Majewski, S., Flamary, R., Gribonval, R., and Courty, N. (2021). Minibatch optimal transport distances; analysis and applications. *arXiv preprint arXiv:2101.01792*.

Feng, J. and Kurtz, T. G. (2006). *Large deviations for stochastic processes*. Number 131. American Mathematical Soc.

Ferradans, S., Papadakis, N., Peyré, G., and Aujol, J.-F. (2014). Regularized discrete optimal transport. *SIAM Journal on Imaging Sciences*, 7(3):1853–1882.

Feydy, J., Charlier, B., Vialard, F.-X., and Peyré, G. (2017). Optimal transport for diffeomorphic registration. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 291–299. Springer.

Fiez, T., Chasnov, B., and Ratliff, L. J. (2019). Convergence of learning dynamics in stackelberg games.

Finlay, C., Gerolin, A., Oberman, A. M., and Pooladian, A.-A. (2020a). Learning normalizing flows from Entropy-Kantorovich potentials. *arXiv preprint arXiv:2006.06033*.

Finlay, C., Jacobsen, J.-H., Nurbekyan, L., and Oberman, A. (2020b). How to train your neural ode: the world of jacobian and kinetic regularization. In *International Conference on Machine Learning*, pages 3154–3164. PMLR.

Flamary, R., Courty, N., Gramfort, A., Alaya, M. Z., Boisbunon, A., Chambon, S., Chapel, L., Corenflos, A., Fatras, K., Fournier, N., Gautheron, L., Gayraud, N. T., Janati, H., Rakotomamonjy, A., Redko, I., Rolet, A., Schutz, A., Seguy, V., Sutherland, D. J., Tavenard, R., Tong, A., and Vayer, T. (2021). Pot: Python optimal transport. *Journal of Machine Learning Research*, 22(78):1–8.

Fleming, W. H. and Stein, J. L. (2004). Stochastic optimal control, international finance and debt. *Journal of Banking & Finance*, 28(5):979–996.

Gafni, O., Polyak, A., Ashual, O., Sheynin, S., Parikh, D., and Taigman, Y. (2022). Make-a-scene: Scene-based text-to-image generation with human priors. *arXiv preprint arXiv:2203.13131*.

Gale, D. and Shapley, L. S. (1962). College admissions and the stability of marriage. *The American Mathematical Monthly*, 69(1):9–15.

Gallouët, T. O. and Monsaingeon, L. (2016). A jko splitting scheme for kantorovich-fisher-rao gradient flows. *SIAM J. Math. Analysis*, 49:1100–1130.

Gatys, L. A., Ecker, A. S., and Bethge, M. (2015). A neural algorithm of artistic style. *arXiv preprint arXiv:1508.06576*.

Genevay, A., Peyré, G., and Cuturi, M. (2018). Learning generative models with Sinkhorn divergences. In *International Conference on Artificial Intelligence and Statistics*, pages 1608–1617. PMLR.

Ghosh, A., Kulharia, V., Namboodiri, V., Torr, P. H. S., and Dokania, P. K. (2018). Multi-agent diverse generative adversarial networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*.

Gidel, G., Berard, H., Vignoud, G., Vincent, P., and Lacoste-Julien, S. (2019). A variational inequality perspective on generative adversarial networks. In *International Conference on Learning Representations*.

Glicksberg, I. L. (1952). A further generalization of the Kakutani fixed point theorem, with application to Nash equilibrium points. *Proceedings of the American Mathematical Society*, 3(1):170–174.

Gobet, E. (2016). *Monte-Carlo methods and stochastic processes: from linear to non-linear*. CRC Press.

Gobet, E., Lemor, J.-P., Warin, X., et al. (2005). A regression-based Monte Carlo method to solve backward stochastic differential equations. *The Annals of Applied Probability*, 15(3):2172–2202.

Gómez, V., Kappen, H. J., Peters, J., and Neumann, G. (2014). Policy search for path integral control. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pages 482–497. Springer.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y. (2014). Generative adversarial nets. In *Advances in Neural Information Processing Systems*, pages 2672–2680.

Gorodetsky, A., Karaman, S., and Marzouk, Y. (2018). High-dimensional stochastic optimal control using continuous tensor decompositions. *International Journal of Robotics Research*, 37(2-3).

Greif, C. (2017). Numerical methods for hamilton-jacobi-bellman equations.

Grnarova, P., Levy, K. Y., Lucchi, A., Hofmann, T., and Krause, A. (2017). An online learning approach to generative adversarial networks. *arXiv preprint arXiv:1706.03269*.

Han, J. and Hu, R. (2020). Deep fictitious play for finding markovian nash equilibrium in multi-agent games. In *Mathematical and scientific machine learning*, pages 221–245. PMLR.

Han, J., Jentzen, A., and E, W. (2018). Solving high-dimensional partial differential equations using deep learning. *Proceedings of the National Academy of Sciences*, 115(34):8505–8510.

Hartmann, C., Banisch, R., Sarich, M., Badowski, T., and Schütte, C. (2014). Characterization of rare events in molecular dynamics. *Entropy*, 16(1):350–376.

Hartmann, C., Kebiri, O., Neureither, L., and Richter, L. (2019). Variational approach to rare event simulation using least-squares regression. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 29(6):063107.

Hartmann, C., Richter, L., Schütte, C., and Zhang, W. (2017). Variational characterization of free energy: Theory and algorithms. *Entropy*, 19(11).

Hartmann, C. and Schütte, C. (2012). Efficient rare event simulation by optimal nonequilibrium forcing. *Journal of Statistical Mechanics: Theory and Experiment*, 2012(11):P11004.

Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B., and Hochreiter, S. (2017). Gans trained by a two time-scale update rule converge to a local nash equilibrium. *Advances in neural information processing systems*, 30.

Ho, J., Jain, A., and Abbeel, P. (2020). Denoising diffusion probabilistic models. *Advances in Neural Information Processing Systems*, 33:6840–6851.

Holdijk, L., Du, Y., Hooft, F., Jaini, P., Ensing, B., and Welling, M. (2023). Stochastic optimal control for collective variable free sampling of molecular transition paths.

Hsieh, Y.-P., Liu, C., and Cevher, V. (2019). Finding mixed Nash equilibria of generative adversarial networks. In Chaudhuri, K. and Salakhutdinov, R., editors, *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pages 2810–2819, Long Beach, California, USA. PMLR.

Huang, C.-W., Chen, R. T., Tsirigotis, C., and Courville, A. (2020). Convex potential flows: Universal probability distributions with optimal transport and convex optimization. In *International Conference on Learning Representations*.

Huang, W., Ji, M., Liu, Z., and Yi, Y. (2015). Steady states of fokker–planck equations: I. existence. *Journal of Dynamics and Differential Equations*, 27:721–742.

Hutton, J. E. and Nelson, P. I. (1984). Interchanging the order of differentiation and stochastic integration. *Stochastic Processes and their Applications*, 18(2):371–377.

Hutzenthaler, M., Jentzen, A., and Kruse, T. (2019). Overcoming the curse of dimensionality in the numerical approximation of parabolic partial differential equations with gradient-dependent nonlinearities. *arXiv:1912.02571*.

Hutzenthaler, M., Jentzen, A., Kruse, T., et al. (2016). Multilevel picard iterations for solving smooth semilinear parabolic heat equations. *arXiv preprint arXiv:1607.03295*.

Hutzenthaler, M., Jentzen, A., Kruse, T., Nguyen, T. A., and von Wurstemberger, P. (2018). Overcoming the curse of dimensionality in the numerical approximation of semilinear parabolic partial differential equations. *arXiv:1807.01212*.

Hutzenthaler, M. and Kruse, T. (2020). Multilevel picard approximations of high-dimensional semilinear parabolic differential equations with gradient-dependent nonlinearities. *SIAM Journal on Numerical Analysis*, 58(2):929–961.

Jensen, M. and Smears, I. (2013). On the convergence of finite element methods for hamilton–jacobi–bellman equations. *SIAM Journal on Numerical Analysis*, 51(1):137–162.

Jin, C., Netrapalli, P., and Jordan, M. I. (2019). Minmax optimization: Stable limit points of gradient descent ascent are locally optimal. *arXiv preprint arXiv:1902.00618*.

Johnson, J., Alahi, A., and Fei-Fei, L. (2016). Perceptual losses for real-time style transfer and super-resolution. In *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part II 14*, pages 694–711. Springer.

Juditsky, A., Nemirovski, A., and Tauvel, C. (2011). Solving variational inequalities with stochastic mirror-prox algorithm. *Stochastic Systems*, 1(1):17–58.

Kallenberg, O. (2002). *Foundations of Modern Probability*. Probability and Its Applications. Springer New York.

Kantorovitch, L. (1942). On the translocation of masses. *C. R. (Doklady) Acad. Sci. URSS (N.S.)*, 37:199–201.

Kappen, H. J. (2005). Path integrals and symmetry breaking for optimal control theory. *Journal of Statistical Mechanics: Theory and Experiment*, 2005(11).

Kappen, H. J., Gómez, V., and Opper, M. (2012). Optimal control as a graphical model inference problem. *Machine learning*, 87(2):159–182.

Kappen, H. J. and Ruiz, H. C. (2016). Adaptive importance sampling for control and inference. *Journal of Statistical Physics*, 162(5):1244–1266.

Karatzas, I. and Shreve, S. (1991). *Brownian Motion and Stochastic Calculus*. Graduate Texts in Mathematics (113) (Book 113). Springer New York.

Kidger, P., Foster, J., Li, X., Oberhauser, H., and Lyons, T. (2021). Neural sdes as infinite-dimensional gans. In *International Conference on Machine Learning*.

Kondratyev, S., Monsaingeon, L., Vorotnikov, D., et al. (2016). A new optimal transport distance on the space of finite Radon measures. *Advances in Differential Equations*, 21(11/12):1117–1164.

Lacker, D. (2018). Mean field games and interacting particle systems. *Preprint*.

Lee, S., Kim, B., and Ye, J. C. (2023). Minimizing trajectory curvature of ode-based generative models. *arXiv preprint arXiv:2301.12003*.

Lei, Q., Lee, J. D., Dimakis, A. G., and Daskalakis, C. (2019). SGD learns one-layer networks in WGANs. *arXiv preprint arXiv:1910.07030*.

Li, C.-L., Chang, W.-C., Cheng, Y., Yang, Y., and Póczos, B. (2017). MMD GAN: Towards deeper understanding of moment matching network. *Advances in neural information processing systems*, 30.

Li, X., Wong, T.-K. L., Chen, R. T., and Duvenaud, D. (2020). Scalable gradients for stochastic differential equations. In *International Conference on Artificial Intelligence and Statistics*, pages 3870–3882. PMLR.

Liero, M., Mielke, A., and Savaré, G. (2018). Optimal entropy-transport problems and a new Hellinger–Kantorovich distance between positive measures. *Inventiones mathematicae*, 211(3):969–1117.

Lin, Q., Liu, M., Rafique, H., and Yang, T. (2018). Solving weakly-convex-weakly-concave saddle-point problems as weakly-monotone variational inequality. *arXiv preprint arXiv:1810.10207*.

Lipman, Y., Chen, R. T., Ben-Hamu, H., Nickel, M., and Le, M. (2023). Flow matching for generative modeling. *International Conference on Learning Representations*.

Liu, G.-H., Lipman, Y., Nickel, M., Karrer, B., Theodorou, E. A., and Chen, R. T. Q. (2023a). Generalized schrödinger bridge matching.

Liu, G.-H., Vahdat, A., Huang, D.-A., Theodorou, E. A., Nie, W., and Anandkumar, A. (2023b). I^2sb: Image-to-image Schr\" odinger bridge. *International Conference on Machine Learning*.

Liu, H., Gu, X., and Samaras, D. (2019). Wasserstein gan with quadratic transport cost. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 4832–4841.

Liu, Q. (2022). Rectified flow: A marginal preserving approach to optimal transport.

Liu, X., Gong, C., and Liu, Q. (2022). Flow straight and fast: Learning to generate and transfer data with rectified flow. *arXiv preprint arXiv:2209.03003*.

Lübeck, F., Bunne, C., Gut, G., del Castillo, J. S., Pelkmans, L., and Alvarez-Melis, D. (2022). Neural unbalanced optimal transport via cycle-consistent semi-couplings. *arXiv preprint arXiv:2209.15621*.

Ma, J. and Ma, J. (2020). Finite difference methods for the hamilton-jacobi-bellman equations arising in regime switching utility maximization. *J. Sci. Comput.*, 85(3):55.

Makkuva, A., Taghvaei, A., Oh, S., and Lee, J. (2020). Optimal transport mapping via input convex neural networks. In *International Conference on Machine Learning*, pages 6672–6681. PMLR.

Markowich, P. A. and Villani, C. (1999). On the trend to equilibrium for the fokker-planck equation: An interplay between physics and functional analysis. In *Physics and Functional Analysis, Matematica Contemporanea (SBM) 19*, pages 1–29.

Mazumdar, E. V., Jordan, M. I., and Sastry, S. S. (2019). On finding local Nash equilibria (and only local Nash equilibria) in zero-sum games. *arXiv:1901.00838*.

McCann, R. J. (1997). A convexity principle for interacting gases. *Advances in mathematics*, 128(1):153–179.

Mei, S., Montanari, A., and Nguyen, P.-M. (2018). A mean field view of the landscape of two-layer neural networks. *Proceedings of the National Academy of Sciences*, 115(33):E7665–E7671.

Mertikopoulos, P., Lecouat, B., Zenati, H., Foo, C.-S., Chandrasekhar, V., and Piliouras, G. (2019). Optimistic mirror descent in saddle-point problems: Going the extra (gradient) mile. In *International Conference on Learning Representations*.

Mitter, S. K. (1996). Filtering and stochastic control: A historical perspective. *IEEE Control Systems Magazine*, 16(3):67–76.

Munos, R., Stepleton, T., Harutyunyan, A., and Bellemare, M. (2016). Safe and efficient off-policy reinforcement learning. In *Advances in Neural Information Processing Systems*, volume 29. Curran Associates, Inc.

Nash, J. (1951). Non-cooperative games. *Annals of Mathematics*, pages 286–295.

Neklyudov, K., Severo, D., and Makhzani, A. (2022). Action matching: A variational method for learning stochastic dynamics from samples. *arXiv preprint arXiv:2210.06662*.

Nguyen, K., Nguyen, D., Pham, T., Ho, N., et al. (2022). Improving mini-batch optimal transport via partial transportation. In *International Conference on Machine Learning*, pages 16656–16690. PMLR.

Nikaidô, H. and Isoda, K. (1955). Note on non-cooperative convex games. *Pacific Journal of Mathematics*, 5(Suppl. 1):807–815.

Nouiehed, M., Sanjabi, M., Huang, T., Lee, J. D., and Razaviyayn, M. (2019). Solving a class of non-convex min-max games using iterative first order methods. In *Advances in Neural Information Processing Systems*, pages 14905–14916.

Nurbekyan, L., Iannantuono, A., and Oberman, A. M. (2020). No-collision transportation maps. *Journal of Scientific Computing*, 82(2).

Nüsken, N. and Richter, L. (2021). Solving high-dimensional Hamilton–Jacobi–Bellman pdes using neural networks: perspectives from the theory of controlled diffusions and measures on path space. *Partial differential equations and applications*, 2:1–48.

Oksendal, B. (2013). *Stochastic differential equations: an introduction with applications*. Springer Science & Business Media.

Onken, D., Fung, S. W., Li, X., and Ruthotto, L. (2021). Ot-flow: Fast and accurate continuous normalizing flows via optimal transport. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 9223–9232.

Onken, D., Nurbekyan, L., Li, X., Fung, S. W., Osher, S., and Ruthotto, L. (2023). A neural network approach for high-dimensional optimal control applied to multiagent path finding. *IEEE Transactions on Control Systems Technology*, 31(1):235–251.

Pavliotis, G. A. (2014). *Stochastic processes and applications: diffusion processes, the Fokker-Planck and Langevin equations*, volume 60. Springer.

Peyré, G. and Cuturi, M. (2019). Computational optimal transport. *Foundations and Trends® in Machine Learning*, 11(5-6):355–607.

Pham, H. (2009). *Continuous-time stochastic control and optimization with financial applications*, volume 61. Springer Science & Business Media.

Pontryagin, L. (1962). *The Mathematical Theory of Optimal Processes*. Interscience Publishers.

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

Pooladian, A.-A. and Niles-Weed, J. (2021). Entropic estimation of optimal transport maps. *arXiv preprint arXiv:2109.12004*.

Posner, E. C. (1975). Random coding strategies for minimum entropy. *IEEE Transations on Information Theory*, 21(4):388–391.

Powell, W. B. and Meisel, S. (2016). Tutorial on stochastic optimization in energy—part i: Modeling and policies. *IEEE Transactions on Power Systems*, 31(2):1459–1467.

Racanière, S., Weber, T., Reichert, D., Buesing, L., Guez, A., Rezende, D. J., Badia, A. P., Vinyals, O., Heess, N., Li, Y., et al. (2017). Imagination-augmented agents for deep reinforcement learning. In *Advances in Neural Information Processing Systems*, pages 5690–5701.

Ramesh, A., Dhariwal, P., Nichol, A., Chu, C., and Chen, M. (2022). Hierarchical text-conditional image generation with clip latents. *arXiv preprint arXiv:2204.06125*.

Rawlik, K., Toussaint, M., and Vijayakumar, S. (2013). On stochastic optimal control and reinforcement learning by approximate inference. In *Twenty-Third International Joint Conference on Artificial Intelligence*.

Reich, S. (2019). Data assimilation: The Schrödinger perspective. *Acta Numerica*, 28:635–711.

Rezende, D. and Mohamed, S. (2015). Variational inference with normalizing flows. In *Proceedings of the 32nd International Conference on Machine Learning*.

Ronneberger, O., Fischer, P., and Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*, pages 234–241. Springer.

Rosen, J. B. (1965). Existence and uniqueness of equilibrium points for concave n-person games. *Econometrica*, 33(3):520–534.

Rotskoff, G. M., Jelassi, S., Bruna, J., and Vanden-Eijnden, E. (2019). Global convergence of neuron birth-death dynamics. In *Proceedings of the 36th International Conference on International Conference on Machine Learning*, Long Beach, CA, USA.

Rotskoff, G. M. and Vanden-Eijnden, E. (2018). Neural networks as interacting particle systems: Asymptotic convexity of the loss landscape and universal scaling of the approximation error. *arXiv preprint arXiv:1805.00915*.

Rubinstein, R. Y. and Kroese, D. P. (2013). *The cross-entropy method: a unified approach to combinatorial optimization, Monte-Carlo simulation and machine learning*. Springer Science & Business Media.

Saharia, C., Chan, W., Saxena, S., Li, L., Whang, J., Denton, E., Ghasemipour, S. K. S., Ayan, B. K., Mahdavi, S. S., Lopes, R. G., et al. (2022). Photorealistic text-to-image diffusion models with deep language understanding. *arXiv preprint arXiv:2205.11487*.

Sanjabi, M., Razaviyayn, M., and Lee, J. D. (2018). Solving non-convex non-concave min-max games under polyaklojasiewicz condition. *arXiv preprint arXiv:1812.02878*.

Santambrogio, F. (2015). Optimal transport for applied mathematicians.

Schiebinger, G., Shu, J., Tabaka, M., Cleary, B., Subramanian, V., Solomon, A., Gould, J., Liu, S., Lin, S., Berube, P., et al. (2019). Optimal-transport analysis of single-cell gene expression identifies developmental trajectories in reprogramming. *Cell*, 176(4):928–943.

Seguy, V., Damodaran, B. B., Flamary, R., Courty, N., Rolet, A., and Blondel, M. (2017). Large-scale optimal transport and mapping estimation. *arXiv preprint arXiv:1711.02283*.

Sirignano, J. and Spiliopoulos, K. (2019). Mean field analysis of neural networks: A central limit theorem. *Stochastic Processes and their Applications*.

Solomon, J., De Goes, F., Peyré, G., Cuturi, M., Butscher, A., Nguyen, A., Du, T., and Guibas, L. (2015). Convolutional Wasserstein distances: Efficient optimal transportation on geometric domains. *ACM Transactions on Graphics (TOG)*, 34(4):66.

Solomon, J., Peyré, G., Kim, V. G., and Sra, S. (2016). Entropic metric alignment for correspondence problems. *ACM Trans. Graph.*, 35(4):72:1–72:13.

Song, Y., Durkan, C., Murray, I., and Ermon, S. (2021a). Maximum likelihood training of score-based diffusion models. *Advances in Neural Information Processing Systems*, 34:1415–1428.

Song, Y. and Ermon, S. (2019). Generative modeling by estimating gradients of the data distribution. *arXiv preprint arXiv:1907.05600*.

Song, Y., Sohl-Dickstein, J., Kingma, D. P., Kumar, A., Ermon, S., and Poole, B. (2021b). Score-based generative modeling through stochastic differential equations. *International Conference on Learning Representations*.

Song, Y., Sohl-Dickstein, J., Kingma, D. P., Kumar, A., Ermon, S., and Poole, B. (2021c). Score-based generative modeling through stochastic differential equations. In *International Conference on Learning Representations (ICLR 2021)*.

Sznitman, A.-S. (1991). Topics in propagation of chaos. In Hennequin, P.-L., editor, *Ecole d'Eté de Probabilités de Saint-Flour XIX — 1989*, pages 165–251, Berlin, Heidelberg. Springer Berlin Heidelberg.

Theodorou, E., Stulp, F., Buchli, J., and Schaal, S. (2011). An iterative path integral stochastic optimal control approach for learning robotic tasks. *IFAC Proceedings Volumes*, 44(1):11594–11601. 18th IFAC World Congress.

Tong, A., Huang, J., Wolf, G., Van Dijk, D., and Krishnaswamy, S. (2020). Trajectorynet: A dynamic optimal transport network for modeling cellular dynamics. In *International conference on machine learning*, pages 9526–9536. PMLR.

Tong, A., Malkin, N., Huguet, G., Zhang, Y., Rector-Brooks, J., Fatras, K., Wolf, G., and Bengio, Y. (2023). Conditional flow matching: Simulation-free dynamic optimal transport. *arXiv preprint arXiv:2302.00482*.

Van Handel, R. (2007). Stochastic calculus, filtering, and stochastic control. *Course notes, URL http://www. prince- ton. edu/rvan/acm217/ACM217*.

Varadarajan, V. S. (1958). On the convergence of sample probability distributions. *Sankhyā: The Indian Journal of Statistics (1933-1960)*, 19(1/2):23–26.

Villani, C. (2003). *Topics in Optimal Transportation*. Graduate studies in mathematics. American Mathematical Society.

Villani, C. (2008). *Optimal Transport: Old and New*. Grundlehren der mathematischen Wissenschaften. Springer Berlin Heidelberg.

Wayne, G. and Abbott, L. (2014). Hierarchical control using networks trained with higher-level forward models. *Neural Computation*, 26(10):2163–2193.

Wolansky, G. (2020). Semi-discrete optimal transport. *arXiv preprint arXiv:1911.04348*.

Zhang, J. et al. (2004). A numerical scheme for BSDEs. *The annals of applied probability*, 14(1):459–488.

Zhang, Q. and Chen, Y. (2022). Path integral sampler: A stochastic control approach for sampling. In *International Conference on Learning Representations*.

Zhang, W., Wang, H., Hartmann, C., Weber, M., and Schütte, C. (2014). Applications of the cross-entropy method to importance sampling and optimal control of diffusions. *SIAM Journal on Scientific Computing*, 36(6):A2654–A2672.

Zhou, M., Han, J., and Lu, J. (2021). Actor-critic method for high dimensional static Hamilton–Jacobi–Bellman partial differential equations based on neural networks. *SIAM Journal on Scientific Computing*, 43(6):A4043–A4066.