Sampling From a Schrödinger Bridge

Austin J. Stromme
Department of EECS,
Massachusetts Institute of Technology

Abstract

The Schrödinger bridge is a stochastic process that finds the most likely coupling of two measures wrt Brownian motion, and is equivalent to the popular entropically regularized optimal transport problem. Motivated by recent applications of the SB to trajectory reconstruction problems, we study the problem of sampling from a SB in high dimensions. We assume sample access to the marginals of the SB process and prove that the natural plug-in sampler achieves a fast statistical rate of estimation for the population bridge in terms of relative entropy. This sampling procedure is given by computing the entropic OT plan between samples from each marginal, and joining a draw from this plan with a Brownian bridge. We apply this result to construct a new and computationally feasible estimator that yields improved rates for entropic optimal transport map estimation.

1 INTRODUCTION

Finding a meaningful and computationally tractable method of comparing and interpolating high-dimensional probability distributions is a problem of major scientific significance which arises in generative modeling, transfer learning, cellular biology, and beyond Arjovsky et al. (2017); Wilson and Cook (2020); Schiebinger et al. (2019). A popular approach to this problem is called optimal transport (OT), and seeks to find a coupling of two distributions that minimizes a given energy criterion Villani (2003, 2008). OT offers practitioners a geometrically meaningful and flexible means of working with complex data and has been been applied in a wide array of scientific disciplines including economics, graphics, and the aforementioned fields Peyré and Cuturi (2018).

To introduce the OT problem let $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$ be probability measures on $\mathbb{R}^d$ with finite second moment. Then the OT problem is

$$\min_{\pi \in \Pi(\mu,\nu)} \mathbb{E}_\pi[\|x-y\|^2], \tag{1.1}$$

where $\|\cdot\|$ denotes the Euclidean ℓ_2 norm, and $\Pi(\mu,\nu)$ denotes the set of all probability measures on $\mathbb{R}^d \times \mathbb{R}^d$ with marginals μ and ν. A rich theory exists for this problem, including precise characterizations of the optimizer as supported on the graph of a function $T(x)$, called the OT map, and the Riemannian-like geometry of the induced metric space, and has led to major developments in pure mathematics Villani (2008); Ambrosio et al. (2008).

1.1 Curse of Dimensionality for un-Regularized OT

Unfortunately, recent work has shown that OT has limited applicability in practice for high-dimensional data. Namely, to estimate the value of (1.1) in general requires a number of samples exponential in the dimension d, and similar rates hold for the optimizer itself Weed and Berthet (2019); Hütter and Rigollet (2019). While this curse of dimensionality can be mitigated somewhat by making smoothness assumptions on μ and ν, developing computationally feasible methods which achieve improved rates is an area of ongoing research Pooladian and Niles-Weed (2021); Muzellec et al. (2021).

1.2 Entropic Optimal Transport

When solved in practice, however, the OT problem is typically regularized with a relative entropy term, and dubbed entropic OT. **The entropic OT problem**:

$$\min_{\pi \in \Pi(\mu,\nu)} \mathbb{E}_\pi[\|x-y\|^2] + \frac{1}{\eta}\mathrm{KL}(\pi \,\|\, \mu \otimes \nu), \tag{1.2}$$

Rather than use entropic OT as an approximation for un-regularized OT, in this work we instead study entropic OT

Proceedings of the 26[th] International Conference on Artificial Intelligence and Statistics (AISTATS) 2023, Valencia, Spain. PMLR: Volume 206.

on its own. In particular, we treat $\eta > 0$ as a fixed parameter throughout the paper. We denote the optimum and optimizer of (1.2) by $S_\star$ and $\pi_\star$, respectively. In fact, the entropic OT problem is of great interest on its own Chen et al. (2021b), even beyond its use as an approximation of OT, because it possesses a number of attractive properties and connections with diverse areas:

1, this entropy-regularized OT problem is practical to solve: the famous Sinkhorn 's algorithm is a simple and easy to implement algorithm which converges in time linear in the input size and inverse linear in the error Sinkhorn (1964); Sinkhorn and Knopp (1967); Dvurechensky et al. (2018). Indeed , the usual purpose of introducing entropic regularization into the OT problem is just that it permits the use of Sinkhorn's algorithm Cuturi (2013).

2, entropic OT has much better statistical performance than un-regularized OT in high dimension. A recent line of work has shown that entropic OT has statistical rates of estimation of the form $\eta^{d/2}/\sqrt{n}$, transferring the curse of dimensionality from n, the sample size to η, the regularization parameter Genevay et al. (2019); Mena and Niles-Weed (2019); Luise et al. (2019), and in fact that fully-dimension free rates hold albeit with exponential dependence on the regularization parameter Rigollet and Stromme (2022).

3, the entropic OT problem has connections to stochastic processes and optimal control through an equivalent dynamical formulation known as the Schrodinger ¨ bridge Schrodinger ¨ (1931); Schrodinger ¨ (1932). Let R_η be the law of reversible Brownian motion with diffusion $1/2\eta$(namely the variance of each increment is rescaled by a factor of $1/2\eta$). Then the SB is the solution of

$$\min_{D:\ D_0=\mu,\ D_1=\nu} \mathrm{KL}(D \,\|\, R_\eta), \tag{1.3}$$

where the minimization is over probability measures on the Wiener space $D \in \mathcal{P}(C([0,1];\mathbb{R}^d))$, and for such a stochastic process D the notation D_0, D_1 refers to its marginals at times $0, 1$, respectively. (Additional details are provided in section 2.) We write the optimizer of (1.3) as $D_\star$. The solution $D_\star$ of (1.3) is connected to the solution $\pi_\star$ of the entropic OT problem (1.2) in the following manner: take $(x, y) \sim \pi_\star$ and draw a Brownian bridge R_η^{xy} joining x to y; then the law of R_η^{xy} is $D_\star$ Léonard (2014). In other words, $\pi_\star$ is the joint distribution on the endpoints of $D_\star$, and the full stochastic process $D_\star$ is generated by joining endpoints with a Brownian bridge. In the $\eta \to \infty$ limit, this stochastic process becomes deterministic and recovers the dynamical form of un-regularized optimal transport Léonard (2012); Chen et al. (2016).

1.3 Trajectory Reconstruction

There has recently been a great deal of interest in applying the dynamical form of entropic OT, as well as its unregularized cousin, to the problem of trajectory reconstruction Pavon et al. (2021); Lavenant et al. (2021); Bunne et al. (2022b). In the trajectory reconstruction problem, one observes iid samples $x_1, \ldots, x_n \sim \mu$ and $y_1, \ldots, y_n \sim \nu$, thought to be snapshots of a population evolving in time, and wishes to infer the population trajectory at intermediate times, but with the crucial condition that the observations at initial and final times are un-coupled, so one does not know which particle went to which location.

When applying OT-based methods to trajectory reconstruction, one generally uses OT to couple the observations and then interpolates between coupled points. This procedure was famously applied to the problem of single-cell RNA sequencing, where practitioners wish to track the evolution of a population of cells but must destroy a cell to observe its RNA Schiebinger et al. (2019). However, the direct use of un-regularized OT for trajectory reconstruction is difficult for high-dimensional data such as that in scRNA sequencing, due to un-regularized OT's curse of dimensionality. Most approaches to trajectory reconstruction with OT use some form of regularization, most often entropic regularization Schiebinger et al. (2019); Chizat et al. (2022); Bunne et al. (2022a); Scarvelis and Solomon (2022).

1.4 Contributions

Motivated by the problem of trajectory reconstruction and the growing body of work that uses the Schrödinger bridge to solve it, we consider the problem of sampling from the Schrödinger bridge. To isolate the difficulty of sampling the Schrödinger bridge from the difficulty of sampling from its marginals μ and ν, we assume sample access to μ and ν, and study the most natural means of solving this problem: take samples of size n from each of μ and ν, form the entropic OT plan between these empirical measures, take a sample from this plan, and then connect this sample with a Brownian bridge.

The resulting process is denoted $\hat{D}_n$, and our main result controls the relative entropy between $\hat{D}_n$ and the true Schrödinger bridge $D_\star$. To state this result, we suppress constants that only depend on the dimension d with the notation $A \lesssim B$. Corollary 6 states that, if μ and ν are σ^2-sub-Gaussian and have finite entropy, then

$$\mathrm{KL}(\hat{D}_n \,\|\, D_\star) \lesssim \left(1 + \sigma^{\lceil 5d/2 \rceil + 6} \cdot \eta^{\lceil 5d/4 \rceil + 3}\right) \cdot \frac{1}{\sqrt{n}}$$

We apply this result to propose and analyze an estimator $\bar{T}_n$ for the entropic OT map $T_\star(x) := \mathbb{E}_{\pi_\star}[y \,|\, x]$. We assume that μ, ν have support of diameter at most R and finite entropy, and show in Theorem 7 that given samples $\mathcal{X}$ and

$\mathcal{Y}$ of size n from μ and ν respectively, $\bar{T}_n$ achieves sub-exponential dependence on the regularization parameter η and the statistical rate $n^{-1/3}$, namely

$$\mathbb{E}_{\mathcal{X},\mathcal{Y}}[\|\bar{T}_n - T_\star\|^2_{L^2(\mu)}] \lesssim R^2\big(1 + R^{\lceil 5d/2\rceil+6} \cdot \eta^{\lceil 5d/4\rceil+3}\big) \cdot \frac{1}{n^{1/3}}.$$

We remark here that for entropic OT map estimation, the dependence η^{Cd} for a constant C is necessary, since otherwise we could use this estimator to estimate un-regularized OT maps faster than known minimax lower bounds Hütter and Rigollet (2019).

1.5 Related Work

Various methods have been proposed for trajectory reconstruction with OT. An early work fits a deep neural network to predict the trajectory with entropic OT loss Hashimoto et al. (2016). Several methods directly use the OT couplings from empirical data to couple and then interpolate either deterministically Chewi et al. (2021); Schiebinger et al. (2019), or stochastically Lavenant et al. (2021); Chizat et al. (2022). Another work jointly optimizes the dual OT objective with respect to a Riemannian manifold and the metric of the Riemannian manifold Scarvelis and Solomon (2022). The work Bunne et al. (2022b) proposed using the JKO scheme parametrized by an input-convex neural network. And performing the entropic OT coupling with respect to a data-driven reference process other than Brownian motion was studied in Bunne et al. (2022a). Other methods based on iterating optimal control formulations were proposed in Vargas et al. (2021); Pavon et al. (2021); Chen et al. (2021a). There have been some works using Schrödinger bridge to sample from one of its marginals Bernton et al. (2019); Huang et al. (2021). Schrödinger bridge has also been applied to generative modeling, see for example De Bortoli et al (2021).

The statistical theory of un-regularized optimal transport is largely established, with nearly optimal minimax rates known for cost Niles-Weed and Rigollet (2019) and map estimation Hütter and Rigollet (2019), both of which have a curse of dimensionality. For the entropic OT problem, non-parametric statistical rates were first shown for the cost Genevay et al. (2019); Mena and Niles-Weed (2019). The entropic problem has been the subject of several central limit theorems del Barrio et al. (2022); Goldfeld et al. (2022a,b); Gonzalez-Sanz et al. (2022).

The most relevant works to this paper are finite sample results for the entropic OT map, density, and dual potentials Luise et al. (2019); del Barrio et al. (2022); Rigollet and Stromme (2022); Masud et al. (2021). However, these works do not apply to our situation because we consider producing a sample rather than estimation of population quantities. Other works which are closely related study the stability of the entropic OT plan to changes in its marginal are given asymptotically Ghosal et al. (2021) and quantitatively in W_2 Carlier et al. (2022) and W_1 Deligiannidis et al. (2021). Because these results measure stability with respect to Wasserstein distances they are inadequate for yielding statistical rates for plug-in estimators without a curse of dimensionality in the sample size n.

2 PRELIMINARIES

2.1 Notation and Definitions

Given probability measures P, Q on some measure space $(X, \mathcal{A})$

We also consider the KL divergence when Q is not a probability measure, in particular when Q is the law of reversible Brownian motion. In that case we use the same formula and refer to Leonard (2014) for a precise definition of its meaning. We let $\mathcal{L}^d$ denote the Lebesgue measure on $\mathbb{R}^d$.

We use the notation $A \lesssim B$ to indicate that there exists a constant $C = C(d)$ depending only on the dimension such that $A \leqslant CB$. Given a Borel-measurable $f : \mathbb{R}^k \to \mathbb{R}^l$ and Borel probability measure β on $\mathbb{R}^k$, the L^2 norm is defined as

$$\|f\|_{L^2(\beta)} := \mathbb{E}_{z\sim\beta}[\|f(z)\|^2]^{1/2},$$

and the set $L^2(\beta)$ is the set of those f for which $\|f\|_{L^2(\beta)} < \infty$ (the co-dimension will be clear from context), modulo β-almost everywhere equivalence. For a function $f \in L^2(\beta)$, we let $\beta(f) := \mathbb{E}_{x\sim\beta}[f(x)]$. The set $[k]$ for $k \in \mathbb{N}$ is the set of positive integers at least 1 and no more than k. By $\mu \otimes \nu$ we mean the joint law of (x, y) when $x \sim \mu$ and $y \sim \nu$. We frequently work with iid samples $x_1, \ldots, x_n \sim \mu$ and $y_1, \ldots, y_n \sim \nu$, and for convenience write the entire samples $\mathcal{X} := (x_1, \ldots, x_n)$ and $\mathcal{Y} := (y_1, \ldots, y_n)$. The empirical distributions on $\mathcal{X}$ and $\mathcal{Y}$ are written $\mu\hat{}_n$ and $\nu\hat{}_n$,

We always fix a regularization parameter $\eta > 0$, and the corresponding optimum of (1.2) and its optimizer are written $S_\star$ and $\pi_\star$, respectively. Existence of $S_\star$ and existence and uniqueness of $\pi_\star$ follows from our assumption of finite second moment (see below for a formal statement of our assumptions) Csiszar (1975). The empirical analog to (1.2) is

$$\min_{\pi\in\Pi(\hat{\mu}_n,\hat{\nu}_n)} \mathbb{E}_\pi[\|x-y\|^2] + \frac{1}{\eta}\mathrm{KL}(\pi \,\|\, \hat{\mu}_n \otimes \hat{\nu}_n), \quad (2.1)$$

and its optimum and optimizer are denoted $\hat{S}_n$ and $\hat{\pi}_n$, respectively, and similarly uniquely exist due to their finite second moment.

For a stochastic process $D \in \mathcal{P}(C([0,1];\mathbb{R}^d))$ its endpoint distribution, namely the distribution of $(\omega(0), \omega(1))$ for $\omega \sim D$ is written D_{01}. The distribution of $\omega(0)$ for $\omega \sim D$ is written D_0, and likewise for D_1. Its distribution conditional on paths which start at x at time $t = 0$ and end at y at time $t = 1$ is written as D^{xy}.

2.2 Duality for Entropic Optimal Transport

For background on OT, we refer the reader to the book Villani (2008). As in un-regularized OT, the entropic OT problem has a dual which plays an important role in its theory. For the population problem, the dual is

$$\max_{(f,g)} \mu(f) + \nu(g) - \frac{1}{\eta}(\mu \otimes \nu)(e^{\eta(f+g-\|x-y\|^2)}), \quad (2.2)$$

where the maximum runs over $(f, g) \in L^1(\mu) \times L^1(\nu)$. Under our assumption that μ, ν have finite second moment, the optimum is attained Csiszar (1975) by $(f_\star, g_\star) \in L^1(\mu) \times L^1(\nu)$ unique up to the translation $(f_\star, g_\star) \mapsto (f_\star - c, g_\star + c)$ for $c \in \mathbb{R}$. Moreover, the primal and dual solutions are related in the following manner

$$p_\star(x, y) := \frac{\mathrm{d}\pi_\star}{\mathrm{d}(\mu \otimes \nu)}(x, y) = e^{\eta(f_\star(x)+g_\star(y)-\|x-y\|^2)}. \quad (2.3)$$

And the value of the primal problem has the following relationship with the value of the dual problem

$$S_\star = \mu(f_\star) + \nu(g_\star). \quad (2.4)$$

The empirical dual problem is defined analogously by

$$\max_{(f,g)} \hat{\mu}_n(f) + \hat{\nu}_n(g) - \frac{1}{\eta}(\hat{\mu}_n \otimes \hat{\nu}_n)(e^{-\eta(f+g-\|x-y\|^2)}), \quad (2.5)$$

where the maximum runs over $(f, g) \in L^1(\hat{\mu}_n) \times L^1(\hat{\nu}_n)$. We denote the optimizers of (2.5) by $(\hat{f}_n, \hat{g}_n)$ which are again unique up to the translation $(\hat{f}_n, \hat{g}_n) \mapsto (\hat{f}_n + c, \hat{g}_n - c)$ for $c \in \mathbb{R}$. Also, the primal and dual solutions to the empirical problem are related in the following manner

$$p_n(x, y) := \frac{\mathrm{d}\hat{\pi}_n}{\mathrm{d}(\hat{\mu}_n \otimes \hat{\nu}_n)}(x, y) = e^{\eta(\hat{f}_n(x)+\hat{g}_n(y)-\|x-y\|^2)}. \quad (2.6)$$

And the value of the empirical primal problem has the following relationship with the value of the dual problem

$$\hat{S}_n = \hat{\mu}_n(f_\star) + \hat{\nu}_n(g_\star). \quad (2.7)$$

Finally, we use the marginal constraints $\hat{\pi}_n \in \Pi(\hat{\mu}_n, \hat{\nu}_n)$ to write the following constraints for p_n: for all $i, j \in [n]$, we have

$$\frac{1}{n}\sum_{k=1}^{n} p_n(x_i, y_k) = 1, \qquad \frac{1}{n}\sum_{k=1}^{n} p_n(x_k, y_j) = 1. \quad (2.8)$$

2.3 Background on the Schrödinger Bridge

For background on Brownian motion we refer to Durrett (2019), and for background on the the Schrödinger bridge and its connections with entropic OT, we refer to the survey Leonard (2014). We let $W(t)$ denote the standard Wiener process on $\mathbb{R}^d$, and let $W_\eta(t) := W(t/\sqrt{2\eta})$. A Brownian bridge R_η^{xy} from x to y is a sample from Brownian motion (with appropriate variance) conditioned to be x at time $t = 0$ and y at time $t = 1$. It can be written in terms of the Wiener process as

$$R_\eta^{xy}(t) = (1-t)x + ty + W_\eta(t) - tW_\eta(1). \quad (2.9)$$

We let R_η be the law of reversible Brownian motion, namely the Wiener process W_η but with the Lebesgue measure as its initial distribution. The Schrödinger bridge $D_\star$ is then the solution of (1.3), which is guaranteed to exist under our assumptions (stated formally below) that μ, ν have have finite second moments and finite entropy (Léonard, 2014, Prop. 2.5).

We use the following basic facts from Leonard (2014).

Proposition 1 (Equivalence of Schrödinger bridge and entropic OT)**.** *The Schrödinger bridge process $D_\star$ and the entropic OT plan $\pi_\star$ from μ to ν are related in the following manner. First, the endpoint marginal distribution of $D_\star$ is $\pi_\star$, namely $(D_\star)_{01} = \pi_\star$. Second, $D_\star$ can be described as Brownian bridges connecting draws from $\pi_\star$, namely for all $\omega \in C([0,1];\mathbb{R}^d)$,*

$$D_\star(\omega) = \int R_\eta^{xy}(\omega)\mathrm{d}\pi_\star(x, y).$$

Proposition 2 (Entropy additivity formula)**.** *Let $P, Q \in \mathcal{P}(C([0,1];\mathbb{R}^d))$. Then*

$$\mathrm{KL}(P \,\|\, Q) = \mathrm{KL}(P_{01} \,\|\, Q_{01}) + \int \mathrm{KL}(P^{xy} \,\|\, Q^{xy})\mathrm{d}Q_{01}(x, y).$$

2.4 Previous Result on Empirical Entropic OT Cost

The following result gives the best known rates for the error of using the empirical entropic OT cost $\hat{S}_n$ for its population counterpart $S_\star$, and will be used as a main step in our arguments. Recall that a probability distribution μ is said to be σ^2-subGaussian Vershynin (2018), if for all unit vectors $v \in \mathbb{R}^d$ and $t \geqslant 0$, $\mathbb{E}_{x\sim\mu}[\exp(t\langle x - \mathbb{E}[x], v\rangle)] \leqslant \exp(\sigma^2 t^2/2)$,

Theorem 3 (Corollary 1 in Mena and Niles-Weed (2019))**.** *Suppose μ and ν are both σ^2-subGaussian, then*

$$\mathbb{E}_{\mathcal{X},\mathcal{Y}}[|\hat{S}_n - S_\star|] \lesssim \frac{1}{\eta}\left(1 + \sigma^{\lceil 5d/2\rceil+6} \cdot \eta^{\lceil 5d/4\rceil+3}\right) \cdot \frac{1}{\sqrt{n}}.$$

2.5 Assumptions

We assume throughout that μ, ν each have finite second moment, and each have finite entropy:

$$\mathrm{KL}(\mu \,\|\, \mathcal{L}^d),\ \mathrm{KL}(\nu \,\|\, \mathcal{L}^d) < \infty.$$

3 MAIN RESULTS

3.1 Sampling From a Schrödinger Bridge

In this section, we propose an estimator $\hat{D}_n$ for the Schrödinger bridge $D_\star$ from μ to ν.

Algorithm 1 Schrödinger Bridge Sampler.

```
Input: Sample access to probability distributions μ, ν.
       Sample size n and regularization parameter η.
Output: Random path whose law is approximately
        that of Schrödinger bridge.
procedure SchrödingerSampler(μ, ν, n, η)
    draw 𝒳 ~ μ^{⊗n}
    draw 𝒴 ~ ν^{⊗n}
    μ̂_n ← (1/n) Σ_{i=1}^n δ_{x_i}
    ν̂_n ← (1/n) Σ_{j=1}^n δ_{y_j}
    π̂_n ← Sinkhorn(μ̂_n, ν̂_n, η)
    draw (x, y) ~ π̂_n
    return Brownian bridge joining x to y, R_η^{xy}
end procedure
```

Definition 4 (Estimator $\hat{D}_n$)**.** *We let $\hat{D}_n$ be the law of the output of Algorithm 1.*

We emphasize that each draw of $\hat{D}_n$ uses fresh samples $\mathcal{X}, \mathcal{Y}$ from μ, ν; indeed, if this were not the case, the term $\mathrm{KL}(\hat{D}_n \,\|\, D_\star)$ would be infinite because $\hat{D}_n$ would have a finitely supported endpoint distribution. In fact, it is not hard to see that $(\hat{D}_n)_0 = \mu$ and $(\hat{D}_n)_1 = \nu$.

We also remark that $\hat{D}_n$ is practical to compute, given sample access to μ and ν. Indeed, given the output of Algorithm 1, for any $N \in \mathbb{N}$ and set of times $t_1, \dots, t_N \in [0,1]$, the joint distribution of $(\omega(t_1), \dots \omega(t_N))$ for $\omega \sim R_\eta^{x_i y_j}$ is Gaussian with mean and covariance given by equation (2.9), and so is practical to sample from.

Our main result bounds the KL divergence between this process $\hat{D}_n$ and the true Schrödinger bridge $D_\star$, by the expected bias of the entropic OT cost when using the empirical measures $\hat{\mu}_n, \hat{\nu}_n$ for samples $\mathcal{X}, \mathcal{Y}$.

Theorem 5. *Recall that $\hat{S}_n$ and $S_\star$ are the empirical and population entropic OT cost, namely the values of (2.1) and (1.2), respectively. We have the bound*

$$\mathrm{KL}(\hat{D}_n \,\|\, D_\star) \leqslant \eta \mathbb{E}_{\mathcal{X},\mathcal{Y}}[\hat{S}_n - S_\star].$$

Using the bias bound given in Mena and Niles-Weed (2019), which we include herein as Theorem 3, we get the following Corollary.

Corollary 6. *Suppose μ, ν are σ^2-subGaussian. Then*

$$\mathrm{KL}(\hat{D}_n \,\|\, D_\star) \lesssim \left(1 + \sigma^{\lceil 5d/2 \rceil + 6} \cdot \eta^{\lceil 5d/4 \rceil + 3}\right) \cdot \frac{1}{\sqrt{n}}.$$

This result states that if we have sample access to distributions μ, ν, which are σ^2-subGaussian and have finite entropy, then we can nearly sample from $D_\star$ using $\hat{D}_n$ and incur error of order $n^{-1/2}$.

We remark that we can also derive a $1/n$ rate from Theorem 5 for compactly supported μ, ν by employing the improved bias bounds from the works Rigollet and Stromme (2022) and del Barrio et al. (2022), but we avoid this consequence because it introduces an exponential dependence on the regularization parameter η.

In the next section, we apply this result to give improved estimators for entropic OT maps.

3.2 Improved Estimators for Entropic OT Maps

We remind the reader that for the un-regularized OT problem from μ to ν, the OT plan is a deterministic coupling $(x, T(x))$, where $T(x)$ is known as the OT map from μ to ν Villani (2008). In many applications of OT such as domain adaptation or trajectory reconstruction, the OT map T is more important than the OT cost, since it provides a means of coupling one distribution to another. Unfortunately, recent work has established that the OT map T suffers from a curse of dimensionality Hütter and Rigollet (2019).

The entropic OT map is the entropically-regularized analog of the OT map T, and defined to be $T_\star(x) := \mathbb{E}_{\pi_\star}[y \,|\, x]$, also known as the barycentric projection of $\pi_\star$. In the un-regularized case the OT map encodes all the information of the OT plan, but in contrast, the entropic OT plan $\pi_\star$ is not deterministic given x: as can be seen from equation (2.3), $\pi_\star(\cdot \,|\, x)$ is not a point mass. While $T_\star$ does not push forward μ to ν as its un-regularized analog does, it still provides a useful summary of $\pi_\star(\cdot \,|\, x)$, and was recently studied as a computationally efficient estimator of the un-regularized OT map Pooladian and Niles-Weed (2021). In this section, we consider the problem of estimating $T_\star$ itself.

Given samples $\mathcal{X}, \mathcal{Y}$ from μ, ν of size n each, the most natural way of estimating $T_\star$ is through its plug-in estimator, for $x \in \mathcal{X}$,

$$\hat{T}_n(x) := \frac{1}{n} \sum_{j=1}^{n} y_j p_n(x, y_j).$$

It was recently shown that, unlike the un-regularized OT map, the entropic OT map $T_\star$ can be estimated by its empirical plug-in estimator with no curse of dimensionality Rigollet and Stromme (2022). In particular, it was shown that when μ, ν are compactly supported, a certain canonically extended version of $\hat{T}_n$ achieves the rate

$$\mathbb{E}_{\mathcal{X},\mathcal{Y}}[\|\hat{T}_n - T_\star\|^2_{L^2(\mu)}] \lesssim e^{C\eta} \cdot \frac{1}{n},$$

where C is a constant depending on the diameter of the support of μ, ν. While this result provides a compelling $1/n$ rate of estimation for $T_\star$, it has limited practical meaning because of the exponential dependence on the regularization parameter η.

In this section we propose and analyze a new estimator, distinct from the plug-in estimator $\hat{T}_n$, which removes the exponential dependence on η, at the cost of a worse rate in n. Our estimator is based on solving the empirical entropic OT problem on many small problem instances and then recombining those instances to obtain an estimator of $T_\star$.

Suppose we are given samples $\mathcal{X} = (x_1, \dots, x_n) \sim \mu^{\otimes n}$ and $\mathcal{Y} = (y_1, \dots, y_n) \sim \nu^{\otimes n}$. Our goal is to create a function $\bar{T}_n$ which maps $x \mapsto \bar{T}_n(x)$ such that $\mathbb{E}_{\mathcal{X},\mathcal{Y}}[\|\bar{T}_n - T_\star\|_{L^2(\mu)}]$ is small. To that end, fix an integer $m \in [n]$ such that m divides n evenly, and let $x \in \mathbb{R}^d$. For $k \in [n/m]$, let

$$\mathcal{X}_k(x) := (x, x_{(k-1)m+1}, x_{(k-1)m+2}, \dots, x_{km-1}),$$

and

$$\mathcal{Y}_k := (y_{(k-1)m}, \dots, y_{km-1}).$$

For $k \in [n/m]$, let $\hat{\pi}^x_{n,k}$ be the entropic OT plan between the empirical distributions supported on $\mathcal{X}_k(x)$ and $\mathcal{Y}_k$, and put $\bar{y}_k(x) := E_{\hat{\pi}^x_{n,k}}[y \,|\, x]$. Then set

$$\bar{T}_n(x) := \frac{m}{n} \sum_{k=1}^{n/m} \bar{y}_k(x).$$

In brief, $\bar{T}_n(x)$ is defined to be the average of independent draws of the plug-in estimator $\hat{T}_m$, or equivalently, $\mathbb{E}_{\hat{\pi}_m}[y \,|\, x]$, where the sample from μ has x as its first entry. The purpose of constructing $\bar{T}_n$ is to yield an estimator for $T_\star$ with sub-exponential dependence on the regularization parameter η, but an incidental benefit of $\bar{T}_n$ is that each $\bar{y}_k$ can be computed in parallel. Even without parallelization, computing $\bar{T}_n(x)$ involves solving n/m entropic OT problems on discrete measures with support each size m, so (ignoring approximations resulting from partially solving for $\hat{\pi}_n$) takes time $O((n/m)m^2) = O(nm)$. Our main result on estimation $T_\star$ is as follows.

Theorem 7. *Suppose μ, ν have compact supports of diameter no more than R. Let $\bar{T}_n$ be the previously described estimator with parameter $m = n^{2/3}$ and suppose m is an integer. Then*

$$\mathbb{E}_{\mathcal{X},\mathcal{Y}}[\|\bar{T}_n - T_\star\|^2_{L^2(\mu)}] \lesssim R^2\big(1 + R^{\lceil 5d/2\rceil+6} \cdot \eta^{\lceil 5d/4\rceil+3}\big) \cdot \frac{1}{n^{1/3}}.$$

Note that for simplicity we have not fully optimized the choice of m to take into account the regularization parameter, which would reduce the exponent of η in the bound to slightly less than η^d.

4 PROOF OF THEOREM 5

For ease of notation, put $\tilde{\pi}_n := (\hat{D}_n)_{01}$, namely its joint distribution on endpoints.

By Propositions 1, 2, and 9, as well as the definition of $\hat{D}_n$, we have

$$\begin{aligned}\mathrm{KL}(\hat{D}_n \,\|\, D_\star) &= \mathrm{KL}(\tilde{\pi}_n \,\|\, \pi_\star) \\ &\quad + \int \mathrm{KL}((\hat{D}_n)^{xy} \,\|\, D_\star^{xy}) \mathrm{d}\pi_\star(x,y) \\ &= \mathrm{KL}(\tilde{\pi}_n \,\|\, \pi_\star).\end{aligned}$$

Thus, to prove Theorem 5, it suffices to show the following Lemma.

Lemma 8. *Let $\tilde{\pi}_n := (\hat{D}_n)_{01}$. Then*

$$\mathrm{KL}(\tilde{\pi}_n \,\|\, \pi_\star) \leqslant \eta \mathbb{E}_{\mathcal{X},\mathcal{Y}}[\hat{S}_n - S_\star].$$

To this end, we give a precise description of $\tilde{\pi}_n$ here.

Proposition 9. *The measure $\tilde{\pi}_n$ is absolutely continuous with respect to $\mu \otimes \nu$, and has Radon-Nikodym derivative*

$$\frac{\mathrm{d}\tilde{\pi}_n}{\mathrm{d}(\mu\otimes\nu)}(x,y) = \frac{1}{n^2}\sum_{i,j=1}^{n} \mathbb{E}_{\mathcal{X},\mathcal{Y}}[p_n(x_i, y_j) \,|\, (x_i, y_j) = (x, y)].$$

We prove this Proposition below, but first use it in our proof of Lemma 8.

For ease of notation, recall our definitions of the density $p_\star$ of $\pi_\star$ with respect to $\mu\otimes\nu$ in (2.3) and p_n of $\hat{\pi}_n$ with respect to $\hat{\mu}_n \otimes \hat{\nu}_n$ in (2.3). We similarly introduce the notation

$$\tilde{p}_n(x,y) := \frac{\mathrm{d}\tilde{\pi}_n}{\mathrm{d}(\mu\otimes\nu)}(x,y).$$

Applying this notation, Proposition 9, and using Jensen's inequality twice yields

$$\begin{aligned}\mathrm{KL}(\tilde{\pi}_n \,\|\, \pi_\star) &= \mathbb{E}_{(x,y)\sim\mu\otimes\nu}\big[\tilde{p}_n(x,y) \cdot \ln \frac{\tilde{p}_n(x,y)}{p_\star(x,y)}\big] \\ &\leqslant \frac{1}{n^2}\sum_{i,j=1}^{n} \mathbb{E}_{(x,y)\sim\mu\otimes\nu}\big[\\ &\quad \mathbb{E}_{\mathcal{X},\mathcal{Y}}[p_n(x_i,y_j) \ln \frac{p_n(x_i,y_j)}{p_\star(x,y)} \,|\, (x_i,y_j) = (x,y)]\big].\end{aligned}$$

We then observe that because $\mathcal{X}$ and $\mathcal{Y}$ are iid samples from μ and ν, this double expectation can be re-written as

$$\mathbb{E}_{\mathcal{X},\mathcal{Y}}\big[p_n(x_i,y_j)\ln\frac{p_n(x_i,y_j)}{p_\star(x_i,y_j)}\big].$$

Applying the definitions of $p_\star$ and p_n, from (2.3) and (2.6) respectively, we can write this as

$$\mathrm{KL}(\tilde{\pi}_n\,\|\,\pi_\star) \leqslant \frac{\eta}{n^2}\sum_{i,j=1}^n \mathbb{E}_{\mathcal{X},\mathcal{Y}}\big[p_n(x_i,y_j)(\hat{f}_n(x_i) - f_\star(x_i) + \hat{g}_n(y_j) - g_\star(y_j))\big].$$

We now apply the marginal constraints from (2.8) to simplify the previous equation to

$$\mathrm{KL}(\tilde{\pi}_n\,\|\,\pi_\star) \leqslant \frac{\eta}{n}\mathbb{E}_{\mathcal{X},\mathcal{Y}}\Big[\sum_{i=1}^n \hat{f}_n(x_i) - f_\star(x_i) + \sum_{j=1}^n \hat{g}_n(y_j) - g_\star(y_j)\Big].$$

Finally, we observe that the sum of $\hat{f}_n$ and $\hat{g}_n$ terms are exactly $\hat{S}_n$ by (2.7). Also, for any $i,j\in[n]$ we have $\mathbb{E}_{\mathcal{X},\mathcal{Y}}[f_\star(x_i)] = \mu(f_\star)$ and $\mathbb{E}_{\mathcal{X},\mathcal{Y}}[g_\star(y_j)] = \nu(g_\star)$, since $\mathcal{X},\mathcal{Y}$ are iid samples from μ,ν. Therefore, the $f_\star, g_\star$ terms can be written as $S_\star$ by (2.4), so that in fact

$$\mathrm{KL}(\tilde{\pi}_n\,\|\,\pi_\star) \leqslant \eta\mathbb{E}_{\mathcal{X},\mathcal{Y}}[\hat{S}_n - S_\star].$$

This concludes the proof of Lemma 8.

The only remaining element is to prove Proposition 9.

Proof of Proposition 9. To this end, fix a Borel measurable function ϕ on $\mathbb{R}^d\times\mathbb{R}^d$. Then

$$\begin{aligned}\tilde{\pi}_n(\phi) &= \mathbb{E}_{\mathcal{X},\mathcal{Y},(x_{i_0},y_{j_0})\sim\hat{\pi}_n}[\phi(x_{i_0},y_{j_0})]\\ &= \sum_{i,j=1}^n \mathbb{E}_{\mathcal{X},\mathcal{Y},(x_{i_0},y_{j_0})\sim\hat{\pi}_n}[\phi(x_i,y_j)\mathbb{1}[(i_0,j_0)=(i,j)]]\\ &= \frac{1}{n^2}\sum_{i,j=1}^n \mathbb{E}_{\mathcal{X},\mathcal{Y}}[\phi(x_i,y_j)p_n(x_i,y_j)]\\ &= \mathbb{E}_{(x,y)\sim\mu\otimes\nu}\big[\phi(x,y)\\ &\cdot\frac{1}{n^2}\sum_{i,j=1}^n \mathbb{E}_{\mathcal{X},\mathcal{Y}}[p_n(x_i,y_j)\,|\,(x_i,y_j)=(x,y)]\big].\end{aligned}$$

Since this is true for any such ϕ, Proposition 9 follows by definition of the Radon-Nikodym derivative. □

5 PROOF OF THEOREM 7

Throughout this section, we work in the setting, and with the notation, of subsection 3.2. To prove Theorem 7, we argue that the expectation of $\bar{y}_k(x)$ on each sub-problem is $1/\sqrt{m}$ from to $T_\star(x)$, a result encapsulated in the following Lemma.

Lemma 10. *Let $\tilde{\pi}_m$ be the measure $(\hat{D}_m)_{01}$. For each $k\in[n/m]$, we have*

$$\|\mathbb{E}_{\mathcal{X}_k(x),\mathcal{Y}_k}[\bar{y}_k] - T_\star\|^2_{L^2(\mu)} \leqslant R^2\,\mathrm{KL}(\tilde{\pi}_m\,\|\,\pi_\star). \quad (5.1)$$

Given this Lemma, the proof of Theorem 7 follows quickly. Indeed, we can expand the error as follows

$$\mathbb{E}_{\mathcal{X},\mathcal{Y}}[\|\bar{T}_n - T_\star\|^2_{L^2(\mu)}] = \frac{m^2}{n^2}\sum_{k=1}^{n/m}\mathbb{E}_{\mathcal{X},\mathcal{Y}}[\|\bar{y}_k - T_\star\|^2_{L^2(\mu)}] + \frac{m^2}{n^2}\sum_{k\neq\ell}\mathbb{E}_{\mathcal{X},\mathcal{Y}}[\langle\bar{y}_k - T_\star, \bar{y}_\ell - T_\star\rangle_{L^2(\mu)}].$$

For the first term, we bound it using the assumption that μ,ν are supported in a set of diameter at most R, so that

$$\frac{m^2}{n^2}\sum_{k=1}^{n/m}\mathbb{E}_{\mathcal{X},\mathcal{Y}}[\|\bar{y}_k - T_\star\|^2_{L^2(\mu)}] \leqslant \frac{m}{n}R^2.$$

For the second term, we recognize that the inner product terms are independent, so we can pass the expectation inside and observe that

$$\begin{aligned}&\frac{m^2}{n^2}\sum_{k\neq\ell}\mathbb{E}_{\mathcal{X},\mathcal{Y}}[\langle\bar{y}_k - T_\star, \bar{y}_\ell - T_\star\rangle_{L^2(\mu)}]\\ &= \frac{m^2}{n^2}\sum_{k\neq\ell}\langle\mathbb{E}_{\mathcal{X}_k(x),\mathcal{Y}_k}[\bar{y}_k] - T_\star, \mathbb{E}_{\mathcal{X}_\ell(x),\mathcal{Y}_\ell}[\bar{y}_\ell] - T_\star\rangle_{L^2(\mu)}.\end{aligned}$$

Because the samples $\mathcal{X}$ and $\mathcal{Y}$ are iid, each of these terms is equivalent to the LHS of (5.1), so we can apply Lemma 10 to yield

$$\frac{m^2}{n^2}\sum_{k\neq\ell}\mathbb{E}_{\mathcal{X},\mathcal{Y}}[\langle\bar{y}_k - T_\star, \bar{y}_\ell - T_\star\rangle_{L^2(\mu)}] \leqslant R^2\,\mathrm{KL}(\tilde{\pi}_m\,\|\,\pi_\star).$$

Since $x\sim\mu$ and $y\sim\nu$ are contained in a set of diameter at most R, they are sub-Gaussian with variance proxy $\lesssim R^2$ Vershynin (2018), and thus we can apply Corollary 6 to give

$$\begin{aligned}&\frac{m^2}{n^2}\sum_{k\neq\ell}\mathbb{E}_{\mathcal{X},\mathcal{Y}}[\langle\bar{y}_k - T_\star, \bar{y}_\ell - T_\star\rangle_{L^2(\mu)}]\\ &\lesssim R^2(1 + R^{\lceil 5d/2\rceil+6}\eta^{\lceil 5d/4\rceil+3})\cdot\frac{1}{\sqrt{m}}.\end{aligned}$$

Combining this with the bound on the first term and plugging in our choice of $m = n^{2/3}$ yields Theorem 7.

Hence Lemma 10 is sufficient to prove Theorem 7, and we now turn to its proof.

Proof of Lemma 10. Fix an $x\in\mathbb{R}^d$, let y_0 be distributed according to the conditional distribution $\tilde{\pi}_m(\cdot\,|\,x)$, and let y_1 be distributed according to the conditional distribution $\pi_\star(\cdot\,|\,x)$. We first observe that $\mathbb{E}_{\mathcal{X},\mathcal{Y}}[\bar{y}_k(x)] =$

$\mathbb{E}_{\tilde{\pi}_m(\cdot\,|\,x)}[y_0]$, by definition of $\bar{y}_k$ and $\tilde{\pi}_m$. Also, $\mathbb{E}_{\pi_\star(\cdot\,|x)}[y_1] = T_\star(x)$, by definition. Thus,

$$\begin{aligned}\|\mathbb{E}_{\mathcal{X}_k(x),\mathcal{Y}_k}[\bar{y}_k(x)] - T_\star(x)\|^2 \\ = \|\mathbb{E}_{\tilde{\pi}_m(\cdot\,|\,x)}[y_0] - \mathbb{E}_{\pi_\star(\cdot\,|,x)}[y_1]\|^2.\end{aligned} \tag{5.2}$$

We now make the following observation about bounded distributions, proved at the end of this section.

Proposition 11. *Suppose P and Q are distributions with support contained in a set of diameter at most R. Then*

$$\|\mathbb{E}_{z_0\sim P}[z_0] - \mathbb{E}_{z_1\sim Q}[z_1]\|^2 \leqslant \frac{R^2}{2}\,\mathrm{KL}(P\,\|\,Q).$$

Applying Proposition 11 to (5.2) yields

$$\begin{aligned}&\|\mathbb{E}_{\mathcal{X}_k(x),\mathcal{Y}_k}[\bar{y}_k(x)] - T_\star(x)\|^2 \\ &\qquad \leqslant \frac{R^2}{2}\,\mathrm{KL}(\tilde{\pi}_m(\cdot\,|\,x)\,\|\,\pi_\star(\cdot\,|\,x)).\end{aligned}$$

Now, note that $\pi_\star \in \Pi(\mu,\nu)$, so has x-marginal μ. But we claim that, additionally, $\tilde{\pi}_m \in \Pi(\mu,\nu)$. This is because $\tilde{\pi}_m$ always outputs a draw from μ in its first coordinate and a draw from ν in its second coordinate. Therefore,

$$\begin{aligned}&\|\mathbb{E}_{\mathcal{X}_k(x),\mathcal{Y}_k}[\bar{y}_k] - T_\star\|^2_{L^2(\mu)} \\ &\qquad \leqslant \frac{R^2}{2}\mathbb{E}_{x\sim\mu}[\mathrm{KL}(\tilde{\pi}_m(\cdot\,|\,x)\,\|\,\pi_\star(\cdot\,|\,x))] \\ &\qquad = \frac{R^2}{2}\,\mathrm{KL}(\tilde{\pi}_m\,\|\,\pi_\star).\end{aligned}$$

This concludes the proof of Lemma 10. □

Proof of Proposition 11. This proposition is proved by letting π_0 be a W_1-optimal coupling for P to Q, and observing that

$$\begin{aligned}\|\mathbb{E}_{z_0\sim P}[z_0] - \mathbb{E}_{z_1\sim Q}[z_1]\| &= \|\mathbb{E}_{(z_0,z_1)\sim\pi_0}[z_0 - z_1]\| \\ &\leqslant W_1(P,Q).\end{aligned}$$

We then use Villani Theorem 6.15 (2008) to compare to total variation distance and Pinsker's inequality to conclude. □

6 CONCLUSION

In this work, we studied the problem of sampling from the Schrödinger bridge process given sample access to its marginals. We showed that the natural plug-in process achieves the rate $n^{-1/2}$ in Corollary 6. We applied this result to analyze a novel estimator for the entropic OT map, and prove that it achieves an $n^{-1/3}$ rate with a sub-exponential dependence on the regularization parameter in Theorem 7. We emphasize that both of our estimation procedures are computationally practical.

While these statistical rates show that the Schrödinger bridge can be efficiently learned in high dimensions, there remain questions about scalability of specific estimators used for trajectory reconstruction, and in particular our results crucially use the fact that we are comparing to Brownian motion, rather than another reference process as is sometimes considered in the literature Bunne et al. (2022a); De Bortoli et al. (2021). We thus propose that future work consider generalizing these results to reference processes beyond Brownian motion. We also propose that future work study how these results, in particular Theorem 7, can be extended to general Monte Carlo estimation for integrals of the form $\mathbb{E}_{\pi_\star}[h(y)\,|\,x]$ for $h\colon \mathbb{R}^d \to \mathbb{R}^k$.

Acknowledgements. The author gratefully acknowledges partial support from NSF award IIS-1838071.

References

Ambrosio, L., Gigli, N., and Savaré, G. (2008). *Gradient flows: in metric spaces and in the space of probability measures*. Springer Science & Business Media. 1

Arjovsky, M., Chintala, S., and Bottou, L. (2017). Wasserstein generative adversarial networks. In *International conference on machine learning*, pages 214–223. PMLR. 1

Bernton, E., Heng, J., Doucet, A., and Jacob, P. E. (2019). Schr\" odinger bridge samplers. *arXiv preprint arXiv:1912.13170*. 3

Bunne, C., Hsieh, Y.-P., Cuturi, M., and Krause, A. (2022a). Recovering stochastic dynamics via gaussian schr\" odinger bridges. *arXiv preprint arXiv:2202.05722*. 2, 3, 8

Bunne, C., Papaxanthos, L., Krause, A., and Cuturi, M. (2022b). Proximal optimal transport modeling of population dynamics. In *International Conference on Artificial Intelligence and Statistics*, pages 6511–6528. PMLR. 2, 3

Carlier, G., Chizat, L., and Laborde, M. (2022). Lipschitz continuity of the schr\" odinger map in entropic optimal transport. *arXiv preprint arXiv:2210.00225*. 3

Chen, T., Liu, G.-H., and Theodorou, E. A. (2021a). Likelihood training of schr\" odinger bridge using forward-backward sdes theory. *arXiv preprint arXiv:2110.11291*. 3

Chen, Y., Georgiou, T. T., and Pavon, M. (2016). On the relation between optimal transport and schrödinger bridges: A stochastic control viewpoint. *Journal of Optimization Theory and Applications*, 169:671–691. 2

Chen, Y., Georgiou, T. T., and Pavon, M. (2021b). Stochastic control liaisons: Richard sinkhorn meets gaspard monge on a schrodinger bridge. *Siam Review*, 63(2):249–313. 2

Chewi, S., Clancy, J., Le Gouic, T., Rigollet, P., Stepaniants, G., and Stromme, A. (2021). Fast and smooth interpolation on wasserstein space. In *International Conference on Artificial Intelligence and Statistics*, pages 3061–3069. PMLR. 3

Chizat, L., Zhang, S., Heitz, M., and Schiebinger, G. (2022). Trajectory inference via mean-field langevin in path space. *arXiv preprint arXiv:2205.07146*. 2, 3

Csiszar, I. (1975). I-Divergence Geometry of Probability Distributions and Minimization Problems. *The Annals of Probability*, 3(1):146 – 158. 3, 4

Cuturi, M. (2013). Sinkhorn distances: Lightspeed computation of optimal transport. In *Advances in neural information processing systems*, pages 2292–2300. 2

De Bortoli, V., Thornton, J., Heng, J., and Doucet, A. (2021). Diffusion schrödinger bridge with applications to score-based generative modeling. *Advances in Neural Information Processing Systems*, 34:17695–17709. 3, 8

del Barrio, E., Gonzalez-Sanz, A., Loubes, J.-M., and Niles-Weed, J. (2022). An improved central limit theorem and fast convergence rates for entropic transportation costs. *arXiv preprint arXiv:2204.09105*. 3, 5

Deligiannidis, G., De Bortoli, V., and Doucet, A. (2021). Quantitative uniform stability of the iterative proportional fitting procedure. *arXiv preprint arXiv:2108.08129*. 3

Durrett, R. (2019). *Probability: theory and examples*, volume 49. Cambridge university press. 4

Dvurechensky, P., Gasnikov, A., and Kroshnin, A. (2018). Computational optimal transport: Complexity by accelerated gradient descent is better than by sinkhorn's algorithm. In Dy, J. and Krause, A., editors, *Proceedings of the 35th International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*, pages 1367–1376. PMLR. 2

Genevay, A., Chizat, L., Bach, F., Cuturi, M., and Peyré, G. (2019). Sample complexity of sinkhorn divergences. In *The 22nd international conference on artificial intelligence and statistics*, pages 1574–1583. PMLR. 2, 3

Ghosal, P., Nutz, M., and Bernton, E. (2021). Stability of entropic optimal transport and Schrödinger bridges. *arXiv preprint arXiv:2106.03670*. 3

Goldfeld, Z., Kato, K., Rioux, G., and Sadhu, R. (2022a). Limit theorems for entropic optimal transport maps and the sinkhorn divergence. *arXiv preprint arXiv:2207.08683*. 3

Goldfeld, Z., Kato, K., Rioux, G., and Sadhu, R. (2022b). Statistical inference with regularized optimal transport. *arXiv preprint arXiv:2205.04283*. 3

Gonzalez-Sanz, A., Loubes, J.-M., and Niles-Weed, J. (2022). Weak limits of entropy regularized optimal transport; potentials, plans and divergences. *arXiv preprint arXiv:2207.07427*. 3

Hashimoto, T., Gifford, D., and Jaakkola, T. (2016). Learning population-level diffusions with generative rnns. In *International Conference on Machine Learning*, pages 2417–2426. PMLR. 3

Huang, J., Jiao, Y., Kang, L., Liao, X., Liu, J., and Liu, Y. (2021). Schr {\" o} dinger-f {\" o} llmer sampler: Sampling without ergodicity. *arXiv preprint arXiv:2106.10880*. 3

Hütter, J.-C. and Rigollet, P. (2019). Minimax rates of estimation for smooth optimal transport maps. *arXiv preprint arXiv:1905.05828*. 1, 3, 5

Lavenant, H., Zhang, S., Kim, Y.-H., and Schiebinger, G. (2021). Towards a mathematical theory of trajectory inference. *arXiv preprint arXiv:2102.09204*. 2, 3

Léonard, C. (2012). From the schrödinger problem to the monge–kantorovich problem. *Journal of Functional Analysis*, 262(4):1879–1920. 2

Léonard, C. (2014). A survey of the schrödinger problem and some of its connections with optimal transport. *Discrete and Continuous Dynamical Systems-Series A*, 34(4):1533–1574. 2, 3, 4

Luise, G., Salzo, S., Pontil, M., and Ciliberto, C. (2019). Sinkhorn barycenters with free support via Frank-Wolfe algorithm. *Advances in Neural Information Processing Systems*, 32:9322–9333. 2, 3

Masud, S. B., Werenski, M., Murphy, J. M., and Aeron, S. (2021). Multivariate rank via entropic optimal transport: sample efficiency and generative modeling. 3

Mena, G. and Niles-Weed, J. (2019). Statistical bounds for entropic optimal transport: sample complexity and the central limit theorem. *Advances in Neural Information Processing Systems*, 32. 2, 3, 4, 5

Muzellec, B., Vacher, A., Bach, F., Vialard, F.-X., and Rudi, A. (2021). Near-optimal estimation of smooth transport maps with kernel sums-of-squares. *arXiv preprint arXiv:2112.01907*. 1

Niles-Weed, J. and Rigollet, P. (2019). Estimation of wasserstein distances in the spiked transport model. *arXiv preprint arXiv:1909.07513*. 3

Pavon, M., Trigila, G., and Tabak, E. G. (2021). The data-driven schrödinger bridge. *Communications on Pure and Applied Mathematics*, 74(7):1545–1573. 2, 3

Peyré, G. and Cuturi, M. (2018). Computational Optimal Transport. *ArXiv e-prints*. 1

Pooladian, A.-A. and Niles-Weed, J. (2021). Entropic estimation of optimal transport maps. *arXiv preprint arXiv:2109.12004*. 1, 5

Rigollet, P. and Stromme, A. J. (2022). On the sample complexity of entropic optimal transport. *arXiv preprint arXiv:2206.13472*. 2, 3, 5, 6

Scarvelis, C. and Solomon, J. (2022). Riemannian metric learning via optimal transport. *arXiv preprint arXiv:2205.09244*. 2, 3

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. 1, 2, 3

Schrödinger, E. (1931). Über die Umkehrung der Naturgesetze. *Angewandte Chemie*, 44(30):636–636. 2

Schrödinger, E. (1932). Sur la théorie relativiste de l'électron et l'interprétation de la mécanique quantique. *Annales de l'institut Henri Poincaré*, 2(4):269–310. 2

Sinkhorn, R. (1964). A relationship between arbitrary positive matrices and doubly stochastic matrices. *The annals of mathematical statistics*, 35(2):876–879. 2

Sinkhorn, R. and Knopp, P. (1967). Concerning nonnegative matrices and doubly stochastic matrices. *Pacific Journal of Mathematics*, 21(2):343–348. 2

Vargas, F., Thodoroff, P., Lamacraft, A., and Lawrence, N. (2021). Solving schrödinger bridges via maximum likelihood. *Entropy*, 23(9):1134. 3

Vershynin, R. (2018). *High-dimensional probability: An introduction with applications in data science*, volume 47. Cambridge University Press. 4, 7

Villani, C. (2003). *Topics in optimal transportation*, volume 58 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI. 1

Villani, C. (2008). *Optimal transport: old and new*, volume 338. Springer Science & Business Media. 1, 4, 5, 8

Weed, J. and Berthet, Q. (2019). Estimation of smooth densities in wasserstein distance. In *Conference on Learning Theory*, pages 3118–3119. PMLR. 1

Wilson, G. and Cook, D. J. (2020). A survey of unsupervised deep domain adaptation. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 11(5):1–46. 1