Discrete Schrödinger Bridges with Applications to Two-Sample Homogeneity Testing

Zaid Harchaoui, Lang Liu, Soumik Pal

University of Washington

December 13, 2021

Institute for Foundations of Data Science

Team

Zaid Harchaoui

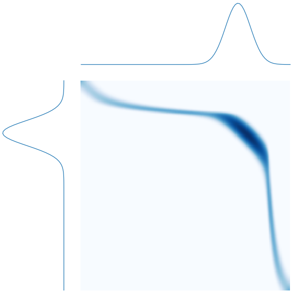
Lang Liu

Soumik Pal

Monge-Kantorovich Optimal Transport

- c nonnegative cost function such that $c(x, y) = 0$ iff $x = y$.
- $\mathrm{CP}(P, Q)$ the set of couplings (joint distributions) with marginals P and Q.

$$C_{\mathrm{OT}}(P, Q) := \inf_{\gamma \in \mathrm{CP}(P,Q)} \int c(x, y) \mathrm{d}\gamma(x, y).$$



Empirical Optimal Transport

- $\{X_i\}_{i=1}^n$ and $\{Y_i\}_{i=1}^n$ two i.i.d. samples from P and Q.
- $P_n := \frac{1}{n}\sum_{i=1}^n \delta_{X_i}$ and $Q_n := \frac{1}{n}\sum_{i=1}^n \delta_{Y_i}$ *empirical distributions.*

$$\hat{C}_{\mathrm{OT}}(P_n, Q_n) := \min_{\gamma \in \mathrm{CP}(P_n, Q_n)} \int c(x, y) \mathrm{d}\gamma(x, y).$$

- $\hat{C}_{\mathrm{OT}}$ converges to C_{OT} (Dudley '69, Weed & Bach '17, Sommerfeld & Munk '18, etc.)

Two challenges:

- The curse of dimensionality $\mathbb{E}\left|\hat{C}_{\mathrm{OT}} - C_{\mathrm{OT}}\right| = O(n^{-1/d})$.
- Computational complexity $O(n^3)$.

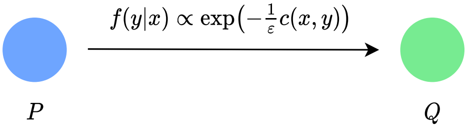
The Schrödinger Bridge Problem and Entropy-Regularized OT

The Schrödinger bridge problem in continuum (Föllmer '88, Léonard '12)

- Assume P and Q are densities,

$$\mu_{\mathrm{SB}} := \operatorname*{arg\,min}_{\gamma \in \mathrm{CP}(P,Q)} \left[\int c(x,y)\,\mathrm{d}\gamma(x,y) + \varepsilon H(\gamma) \right],$$

  where $H(\gamma) = \int \log \gamma(x,y)\,\mathrm{d}\gamma(x,y)$ if γ has a density and ∞ otherwise.
- Easier to estimate both statistically and computationally.



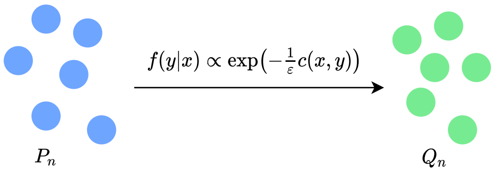
Discrete Schrödinger Bridge

Discrete Schrödinger bridge (DSB)

$$\hat{\mu}_{\mathrm{SB}} := \sum_{\sigma \in \mathrm{Perm}} q_{\mathrm{SB}}(\sigma) \frac{1}{n} \sum_{i=1}^{n} \delta_{(X_i, Y_{\sigma_i})},$$

where

$$q_{\mathrm{SB}}(\sigma) \propto \exp\left(-\frac{1}{\varepsilon} \sum_{i=1}^{n} c(X_i, Y_{\sigma_i})\right).$$



Main Results

Let $\hat{\mu}_{\mathrm{SB}} := \sum_\sigma q_{\mathrm{SB}}(\sigma) \frac{1}{n} \sum_{i=1}^n \delta_{(X_i, Y_{\sigma_i})}$ and $\hat{C}_{\mathrm{SB}} := \sum_\sigma q_{\mathrm{SB}}(\sigma) \frac{1}{n} \sum_{i=1}^n c(X_i, Y_{\sigma_i})$.

Main Results

Let $\hat{\mu}_{\mathrm{SB}} := \sum_\sigma q_{\mathrm{SB}}(\sigma)\frac{1}{n}\sum_{i=1}^n \delta_{(X_i, Y_{\sigma_i})}$ and $\hat{C}_{\mathrm{SB}} := \sum_\sigma q_{\mathrm{SB}}(\sigma)\frac{1}{n}\sum_{i=1}^n c(X_i, Y_{\sigma_i})$.

Theorem 1

Take any function $\eta \in \mathbf{L}^1(\mu_{SB})$. Under appropriate assumptions,

$$T_n(\eta) := \int \eta \mathrm{d}\hat{\mu}_{SB} = \sum_\sigma q_{SB}(\sigma)\frac{1}{n}\sum_{i=1}^n \eta(X_i, Y_{\sigma_i}) = \int \eta \mathrm{d}\mu_{SB} + \mathcal{L}_n + o_p(n^{-1/2}).$$

Main Results

Let $\hat{\mu}_{\text{SB}} := \sum_\sigma q_{\text{SB}}(\sigma)\frac{1}{n}\sum_{i=1}^n \delta_{(X_i, Y_{\sigma_i})}$ and $\hat{C}_{\text{SB}} := \sum_\sigma q_{\text{SB}}(\sigma)\frac{1}{n}\sum_{i=1}^n c(X_i, Y_{\sigma_i})$.

Theorem 1

Take any function $\eta \in \mathbf{L}^1(\mu_{SB})$. Under appropriate assumptions,

$$T_n(\eta) := \int \eta \mathrm{d}\hat{\mu}_{SB} = \sum_\sigma q_{SB}(\sigma)\frac{1}{n}\sum_{i=1}^n \eta(X_i, Y_{\sigma_i}) = \int \eta \mathrm{d}\mu_{SB} + \mathcal{L}_n + o_p(n^{-1/2}).$$

- $\mathcal{L}_n = O_p(n^{-1/2}) \longrightarrow$ *weak convergence of* $\hat{\mu}_{SB}$.

Main Results

Let $\hat{\mu}_{\text{SB}} := \sum_\sigma q_{\text{SB}}(\sigma) \frac{1}{n} \sum_{i=1}^n \delta_{(X_i, Y_{\sigma_i})}$ and $\hat{C}_{\text{SB}} := \sum_\sigma q_{\text{SB}}(\sigma) \frac{1}{n} \sum_{i=1}^n c(X_i, Y_{\sigma_i})$.

Theorem 1

Take any function $\eta \in \mathbf{L}^1(\mu_{SB})$. Under appropriate assumptions,

$$T_n(\eta) := \int \eta \mathrm{d}\hat{\mu}_{SB} = \sum_\sigma q_{SB}(\sigma) \frac{1}{n} \sum_{i=1}^n \eta(X_i, Y_{\sigma_i}) = \int \eta \mathrm{d}\mu_{SB} + \mathcal{L}_n + o_p(n^{-1/2}).$$

- $\mathcal{L}_n = O_p(n^{-1/2}) \longrightarrow$ *weak convergence of* $\hat{\mu}_{SB}$.
- $\sqrt{n}\mathcal{L}_n \asymp Z$ *where* Z *is mean-zero normal* $\longrightarrow$ *limit law of* $\sqrt{n}(\hat{C}_{SB} - C_{SB})$.

Main Results

Let $\hat{\mu}_{\mathrm{SB}} := \sum_\sigma q_{\mathrm{SB}}(\sigma)\frac{1}{n}\sum_{i=1}^n \delta_{(X_i, Y_{\sigma_i})}$ and $\hat{C}_{\mathrm{SB}} := \sum_\sigma q_{\mathrm{SB}}(\sigma)\frac{1}{n}\sum_{i=1}^n c(X_i, Y_{\sigma_i})$.

Theorem 2

Take any function $\eta \in \mathbf{L}^1(\mu_{SB})$. Under appropriate assumptions,

$$T_n(\eta) := \int \eta \mathrm{d}\hat{\mu}_{SB} = \sum_\sigma q_{SB}(\sigma)\frac{1}{n}\sum_{i=1}^n \eta(X_i, Y_{\sigma_i}) = \int \eta \mathrm{d}\mu_{SB} + \mathcal{L}_n + \mathcal{Q}_n + o_p(n^{-1}).$$

- *$\{Z_k\}$ and $\{Z_k'\}$ are independent standard normals.*
- $n\mathcal{Q}_n \asymp \sum_{k,l\geq 1}[a_{kl}Z_kZ_l' + b_{kl}(Z_kZ_l - \mathbb{1}\{k=l\}) + c_{kl}(Z_k'Z_l' - \mathbb{1}\{k=l\})]$.
- *Limit law of $n(\hat{C}_{SB} - C_{SB} - \mathcal{L}_n)$.*

Discrete Entropy-Regularized Optimal Transport

Sinkhorn distance (Cuturi '13, Ferradans et al. '14)

$$\hat{\mu}_{\mathrm{EOT}} = \underset{\gamma \in \mathrm{CP}(P_n, Q_n)}{\arg\min} \left[\int c(x, y) d\gamma(x, y) + \varepsilon \mathrm{Ent}(\gamma) \right],$$

where $\mathrm{Ent}(\gamma) := \sum_{i,j=1}^{n} \gamma(X_i, Y_j) \log \gamma(X_i, Y_j)$ (negative Shannon entropy).

Discrete Entropy-Regularized Optimal Transport

Sinkhorn distance

$$\min_{\gamma \in \mathrm{CP}(P_n, Q_n)} \left[\int c \, d\gamma + \varepsilon \, \mathrm{Ent}(\gamma) \right],$$

$\mathrm{Ent}(\gamma) := \sum_{i,j=1}^{n} \gamma(X_i, Y_j) \log \gamma(X_i, Y_j).$

Discrete Schrödinger bridge

$$\min_{q \in \mathcal{P}(\mathrm{Perm})} \left[\int c \, d\gamma_q + \frac{\varepsilon}{n} \mathrm{Ent}(q) \right],$$

$\gamma_q = \sum_{\sigma} q(\sigma) \frac{1}{n} \sum_{i=1}^{n} \delta_{(X_i, Y_{\sigma_i})}.$

Discrete Entropy-Regularized Optimal Transport

Sinkhorn distance

$$\min_{\gamma \in \mathrm{CP}(P_n, Q_n)} \left[\int c \, d\gamma + \varepsilon \, \mathrm{Ent}(\gamma) \right],$$

$$\mathrm{Ent}(\gamma) := \sum_{i,j=1}^{n} \gamma(X_i, Y_j) \log \gamma(X_i, Y_j).$$

Discrete Schrödinger bridge

$$\min_{q \in \mathcal{P}(\mathrm{Perm})} \left[\int c \, d\gamma_q + \frac{\varepsilon}{n} \mathrm{Ent}(q) \right],$$

$$\gamma_q = \sum_{\sigma} q(\sigma) \frac{1}{n} \sum_{i=1}^{n} \delta_{(X_i, Y_{\sigma_i})}.$$

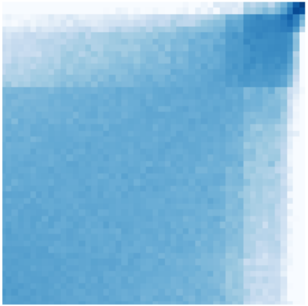
	Consistency		Limit Law		Computation
	Transport cost	**Transport plan**	**First order**	**Second order**	**Stable for small ε**
Sinkhorn	Yes	Pooladian & Weed ’21	Yes	Unknown	No
DSB	Yes	**Yes**	Yes	**Yes**	**Yes**

[Refs] Bigot et al. ’19, Klatt et al ’19, Genevay et al. ’19, Mena & Weed ’19, etc.

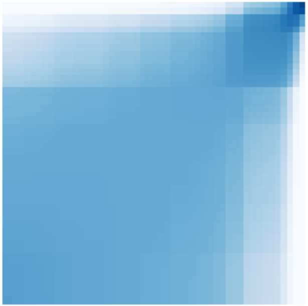
Comparison of the Optimal Couplings

- Fix $n = 100$ and consider $P = \text{Exp}(2)$ and $Q = \text{Exp}(3)$.
- Visualize the transport map as ε decreases.

DSB



Sinkhorn

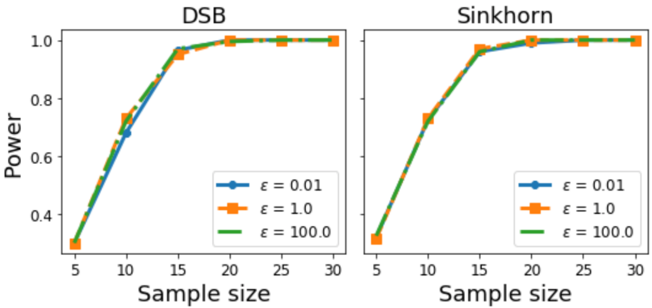


Two-Sample Testing

Two-sample testing on MNIST

- OT cost between two images.
- $\varepsilon \in \{0.01, 1, 100\}$.

DSB provides a powerful test that is robust to ε.



Thank you!

OTML paper: langliu95.github.io/files/OTML2021-eot.pdf

Longer version: arxiv.org/abs/2011.08963

Schrödinger's Lazy Gas Experiment

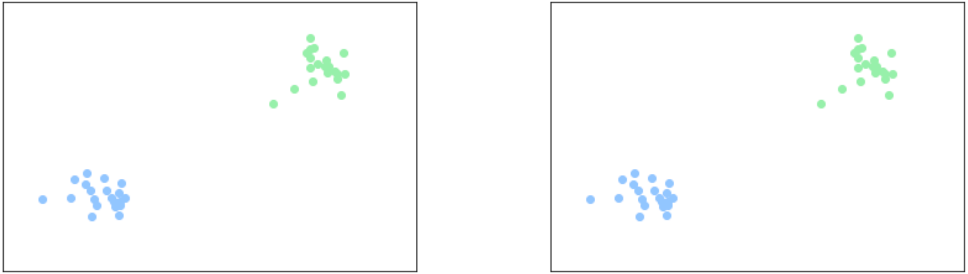


Figure: **Left**: high temperature; **Right**: low temperature.

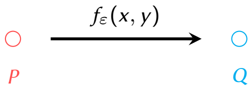
The Schrödinger Bridge Problem and Entropy-Regularized OT

The Schrödinger bridge problem in continuum (Föllmer ’88, Léonard ’12)

- A particle making jumps according to

$$f_\varepsilon(x, y) := \frac{1}{Z_\varepsilon(x)} \exp\left(-\frac{1}{\varepsilon} c(x, y)\right) .$$

- Observe initial and terminal configurations P and Q.
- What is the most likely coupling between P and Q?



The Schrödinger Bridge Problem and Entropy-Regularized OT

The Schrödinger bridge problem in continuum (Föllmer ’88, Léonard ’12)

- Consider a Markov chain with initial distribution P and transition probability f_ε.
- The joint distribution is

$$R_\varepsilon(x, y) := P(x) f_\varepsilon(x, y).$$

The Schrödinger Bridge Problem and Entropy-Regularized OT

The Schrödinger bridge problem in continuum (Föllmer '88, Léonard '12)

- Consider a Markov chain with initial distribution P and transition probability f_ε.
- The joint distribution is

$$R_\varepsilon(x, y) := P(x) f_\varepsilon(x, y).$$

- Conditioned on the initial and terminal configurations being P and Q,

$$\mu_{\mathrm{SB}} := \operatorname*{arg\,min}_{\gamma \in \mathrm{CP}(P,Q)} \mathrm{KL}(\gamma \| R_\varepsilon) = \operatorname*{arg\,min}_{\gamma \in \mathrm{CP}(P,Q)} \left[\int c(x, y) \mathrm{d}\gamma(x, y) + \varepsilon H(\gamma) \right], \tag{1}$$

where $H(\gamma) = \int \log \gamma(x, y) \mathrm{d}\gamma(x, y)$ if γ has a density and ∞ otherwise.

Gibbs Sampling for the Discrete Schrödinger Bridge

Algorithm 1 Gibbs sampling for the Schrödinger bridge statistic

Input: samples $\{X_i\}_{i=1}^n$ and $\{Y_i\}_{i=1}^n$, functions c and ξ, burn-in B and number of iterations L.
Initialization: $\sigma^{(0)} \leftarrow \text{id}$.
for $t = 0, \ldots, L-1$ **do**
    Randomly select $i \neq j \in [n]$.
    Compute $r \leftarrow \exp\left\{[c(X_i, Y_{\sigma_i^{(t)}}) + c(X_j, Y_{\sigma_j^{(t)}}) - c(X_i, Y_{\sigma_j^{(t)}}) - c(X_j, Y_{\sigma_i^{(t)}})]/\varepsilon\right\}$.
    Generate $a \sim \text{Bern}(r/(1+r))$.
    if a = 1 **then**
        Obtain $\sigma^{(t+1)}$ from $\sigma^{(t)}$ by swapping the entries $\sigma_i^{(t)}$ and $\sigma_j^{(t)}$.
    else
        Set $\sigma^{(t+1)} \leftarrow \sigma^{(t)}$.
    end if
end for
Output: $T \leftarrow \frac{1}{L-B} \sum_{t=B+1}^{L} \frac{1}{n} c(X, Y_{\sigma^{(t)}})$.

Examples

Set $c(x, y) = \|x - y\|^2$.

Example 3

Let $P = \mathcal{N}_d(\mu_1, \Sigma_1)$ and $Q = \mathcal{N}_d(\mu_2, \Sigma_2)$, then the cost of the population SCB reads

$$\|\mu_1 - \mu_2\|^2 + \mathbf{Tr}(\Sigma_1) + \mathbf{Tr}(\Sigma_2) - 2\,\mathbf{Tr}\left(\left(\Sigma_1^{1/2}\Sigma_2\Sigma_1^{1/2} + \frac{\varepsilon^2}{16} I_d\right)^{1/2} - \frac{\varepsilon}{4} I_d\right). \tag{2}$$

Example 4

Let P be a density on $\mathbb{R}^d$ and $Q = P * \mathcal{N}_d(0, \frac{\varepsilon}{2} I_d)$, then the population SCB and its cost read

$$\mu_{\mathrm{SB}}(x, y) = P(x)\frac{1}{(\pi\varepsilon)^{d/2}} \exp\left(-\frac{1}{\varepsilon}\|x - y\|^2\right) \quad \text{and} \quad C_{\mathrm{SB}} = \frac{\varepsilon d}{2}. \tag{3}$$

Convergence of the Transport Cost

Goal: explore the convergence empirically.

- Set $c(x, y) = \|x - y\|^2$ and $\varepsilon = 0.1$.
- Generate **independent** samples $\{X_i\}_{i=1}^n \overset{\text{i.i.d.}}{\sim} P$ and $\{Y_i\}_{i=1}^n \overset{\text{i.i.d.}}{\sim} Q$.

(a) $P = \mathcal{N}(0, 1)$ and $Q = \mathcal{N}(0, 1)$.
(b) $P = 0.5\mathcal{N}(-1, 0.3) + 0.5\mathcal{N}(1, 0.3)$ and $Q = P * \mathcal{N}(0, 0.5\varepsilon)$.
(c) $P = \text{Exp}(1)$ and $Q = P * \mathcal{N}(0, 0.5\varepsilon)$.

- Plot $\hat{C}_{\text{SB}} := \int c \mathrm{d}\hat{\mu}_{\text{SB}}$, $\hat{C}_{\text{EOT}} := \int c \mathrm{d}\hat{\mu}_{\text{EOT}}$, and $C_{\text{SB}} := \int c \mathrm{d}\mu_{\text{SB}}$.

Convergence of the Transport Cost

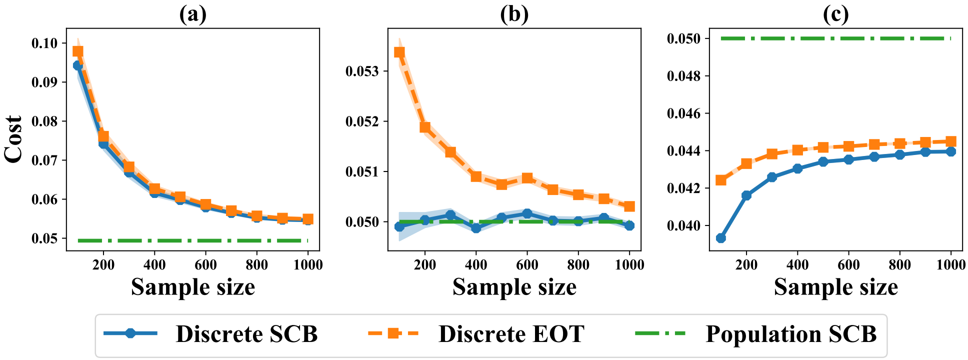


Figure: Cost versus sample size with $\varepsilon = 0.1$.

Measuring the Distance Between Probability Distributions

- *Optimal transport distance*:

$$C_{\mathrm{OT}}(P, Q) = \inf_{\nu \in \mathrm{CP}(P,Q)} \int c(x, y) \mathrm{d}\nu(x, y).$$

- *Transport cost of the Schrödinger bridge*:

$$C_{\mathrm{SB}}(P, Q) := \int c(x, y) \mathrm{d}\mu_{\mathrm{SB}}(x, y).$$

- *Centered cost of the Schrödinger bridge*:

$$\bar{C}_{\mathrm{SB}}(P, Q) := C_{\mathrm{SB}}(P, Q) - \frac{1}{2} C_{\mathrm{SB}}(P, P) - \frac{1}{2} C_{\mathrm{SB}}(Q, Q).$$