Asymptotics of Entropy-Regularized Optimal Transport via Chaos Decomposition

Lang Liu

Department of Statistics, University of Washington

November 23, 2020

Collaborators

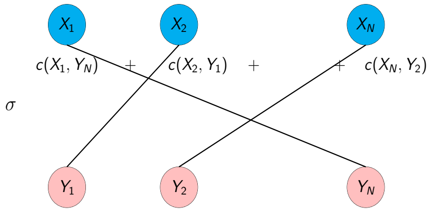
Zaid Harchaoui

Soumik Pal

Optimal matching

- ρ_0 and ρ_1 densities on $\mathbb{R}^d$.
- $\{X_i\}_{i=1}^N$ and $\{Y_i\}_{i=1}^N$ independent i.i.d. samples from ρ_0 and ρ_1.
- c nonnegative and continuous with $c(x,x) = 0$.
- σ permutation or matching.
- $\mathcal{S}_N$ set of permutations on $[N] := \{1, \ldots, N\}$.

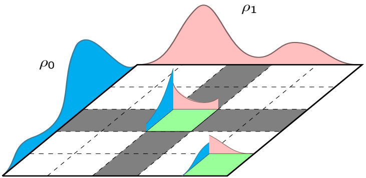
$$\hat{\mathbf{C}} := \min_{\sigma \in \mathcal{S}_N} \frac{1}{N} \sum_{i=1}^{N} c(X_i, Y_{\sigma_i}). \tag{1}$$



Monge-Kantorovich optimal transport (OT)

$$\mathbf{C} := \mathbf{C}(\rho_0, \rho_1) := \inf_{\nu \in \mathrm{CP}(\rho_0, \rho_1)} \int c(x, y) d\nu(x, y). \tag{2}$$

- $\mathrm{CP}(\rho_0, \rho_1)$ set of couplings (joint distributions) with marginals ρ_0 and ρ_1.



The limiting behavior of $(\hat{\mathbf{C}} - \mathbf{C})$ has been studied extensively (in special cases) in the literature.

- Rate of convergence.
- Limiting distributions when $\rho_0 \neq \rho_1$ and $\rho_0 = \rho_1$.

Previous work

Optimal matching:

- Combinatorics (Ajtai, Komlós, Tusnády '84).
- Probability and statistics (Talagrand '92, Fournier & Guillin '15, Weed & Bach '19, Lei '20).
- Economics (Kosowsky & Yuille '94, Galinchon & Salanié '09).

Discrete optimal transport (OT):

- On $\mathbb{R}$ (Munk & Czado '98, del Barrio, Giné, Matran '99, del Barrio, Giné, Utzet '05).
- On $\mathbb{R}^d$ (Ripple, Munk, Sturm '16, del Barrio & Loubes '19).
- On a finite metric space (Sommerfeld & Munk '18, Klatt, Munk, Zemel '20).
- On a countable metric space (Tameling, Sommerfeld, Munk '19)

Discrete optimal transport

Notice that

$$\hat{\mathbf{C}} := \min_{\sigma \in \mathcal{S}_N} \frac{1}{N} \sum_{i=1}^{N} c(X_i, Y_{\sigma_i}) = \min_{A_\sigma : \sigma \in \mathcal{S}_N} \frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{N} C_{ij} [A_\sigma]_{ij}.$$

- C cost matrix, i.e., $C_{ij} := c(X_i, Y_j)$.
- A_σ permutation matrix, i.e., $A_{i\sigma_i} = 1$ and 0 elsewhere.

Matrix relaxation (linear programming):

$$\hat{\mathbf{C}} = \min_{M \in \mathrm{DS}} \langle C, M \rangle. \tag{3}$$

- DS (normalized) doubly stochastic matrices (convex hull of $\{\frac{1}{N} A_\sigma\}$).
- $\langle C, M \rangle := \sum_{i=1}^{N} \sum_{j=1}^{N} C_{ij} M_{ij}$.

Discrete optimal transport

Let $\hat{\rho}_0 := \frac{1}{N}\sum_{i=1}^{N}\delta_{X_i}$ and $\hat{\rho}_1 := \frac{1}{N}\sum_{i=1}^{N}\delta_{Y_i}$ be *empirical measures*.

$$\min_{M\in \mathrm{DS}} \langle C, M\rangle = \min_{\nu\in \mathrm{CP}(\hat{\rho}_0,\hat{\rho}_1)} \int c(x,y)d\nu(x,y).$$

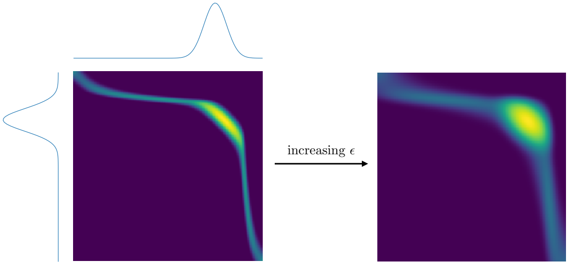
- $M_{ij} \longleftrightarrow \nu(X_i, Y_j)$.
- $\frac{1}{N}A_\sigma$ (permutation matrix) $\longleftrightarrow \frac{1}{N}\sum_{i=1}^{N}\delta_{(X_i, Y_{\sigma_i})}$ (Monge coupling).

Entropic regularization

Entropy-regularized optimal transport (EOT) (Cuturi '13, Ferradans, Papadakis, Peyré, Aujol '14)

$$\min_{M \in \mathrm{DS}} \left[\langle C, M \rangle + \epsilon \mathrm{Ent}_0(M) \right]. \tag{4}$$

- $\epsilon > 0$ regularization parameter.
- $\mathrm{Ent}_0(M) := \sum_{i,j=1}^{N} M_{ij} \log M_{ij}$ (negative Shannon entropy).



Asymptotics of entropy-regularized OT

Gaussian and chi-squares limits can be obtained (Bigot, Cazelles, Papadakis '19, Klatt, Tameling, Munk '20) for the EOT (4)

- ρ_0 and ρ_1 have **finite support** $\Rightarrow$ finite-dimensional simplex.
- $c(x, y) := \|x - y\|^p$ for $p \geq 1$.
- Core idea: calculus on finite-dimensional simplex.

Can we extend these results to densities and arbitrary cost functions?

Let $H(\nu) := \int \nu(x, y) \log \nu(x, y) dxdy$, if ν is a density, and infinity otherwise.

$$\mu := \mu_\epsilon := \underset{\nu \in \mathrm{CP}(\rho_0, \rho_1)}{\arg\min} \left[\int c(x, y) d\nu(x, y) + \epsilon H(\nu) \right]. \tag{5}$$

$\exists$ functions a and b (Csiszar '75, Rüschendorf & Thomsen '93) such that

$$\mu(x, y) \overset{\text{a.s.}}{=} \xi(x, y) \rho_0(x) \rho_1(y), \tag{6}$$

where $\xi(x, y) := \exp\left(-\frac{1}{\epsilon}(c(x, y) - a(x) - b(y))\right)$, and

$$\int \xi(x, y) \rho_1(y) dy \overset{\text{a.s.}}{=} 1 \quad \text{and} \quad \int \xi(x, y) \rho_0(x) dx \overset{\text{a.s.}}{=} 1. \tag{7}$$

Markov transition kernels.

Remark. μ is the *Schrödinger bridge* between ρ_0 and ρ_1 (Schrödinger '32, Föllmer '88, Léonard '12).

Reformulation with a Gibbs measure

How to estimate the Schrödinger bridge μ?

An explicit convex combination of all pairwise empirical measures:

$$\hat{\mu} := \sum_{\sigma \in \mathcal{S}_N} q^*(\sigma) \frac{1}{N} \sum_{i=1}^{N} \delta_{(X_i, Y_{\sigma_i})}, \tag{8}$$

where q^* is a Gibbs measure, with $c(X, Y_\sigma) := \sum_{i=1}^{N} c(X_i, Y_{\sigma_i})$,

$$q^*(\sigma) := \frac{\exp\left(-\frac{1}{\epsilon} c(X, Y_\sigma)\right)}{\sum_{\tau \in \mathcal{S}_N} \exp\left(-\frac{1}{\epsilon} c(X, Y_\tau)\right)}. \tag{9}$$

Remark. $\hat{\mu}$ can be viewed as the Schrödinger bridge between $\hat{\rho}_0$ and $\hat{\rho}_1$. We will call it the discrete Schrödinger bridge.

Reformulation with a Gibbs measure

Define

- $\mathcal{P}(\mathcal{S}_N)$ probability measures on the set of permutations $\mathcal{S}_N$.
- $\mathrm{Ent}(q) := \sum_{\sigma \in \mathcal{S}_N} q(\sigma) \log q(\sigma)$ for $q \in \mathcal{P}(\mathcal{S}_N)$.

$$q^* = \underset{q \in \mathcal{P}(\mathcal{S}_N)}{\arg\min} \left[\sum_{\sigma \in \mathcal{S}_N} q(\sigma) \frac{1}{N} c(X, Y_\sigma) + \frac{\epsilon}{N} \mathrm{Ent}(q) \right]. \tag{10}$$

$$\mathcal{P}(\mathcal{S}_N) \xrightarrow{M = \sum_{\sigma \in \mathcal{S}_N} q(\sigma) \frac{1}{N} A_\sigma} \mathrm{DS}$$

$$\sum_{\sigma \in \mathcal{S}_N} q(\sigma) \tfrac{1}{N} c(X, Y_\sigma) \quad = \quad \langle C, M \rangle$$

Recap

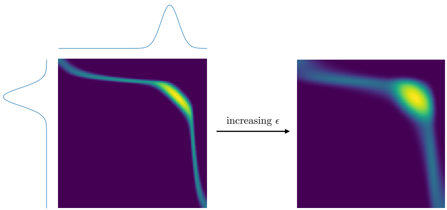
The discrete Schrödinger bridge $\hat{\mu} := \sum_{\sigma \in \mathcal{S}_N} q^*(\sigma) \frac{1}{N} \sum_{i=1}^N \delta_{(X_i, Y_{\sigma_i})}$, where

$$q^*(\sigma) := \frac{\exp\left(-\frac{1}{\epsilon} c(X, Y_\sigma)\right)}{\sum_{\tau \in \mathcal{S}_N} \exp\left(-\frac{1}{\epsilon} c(X, Y_\tau)\right)}.$$

The continuum Schrödinger bridge

$$\mu := \underset{\nu \in \mathrm{CP}(\rho_0, \rho_1)}{\arg\min} \left[\int c(x, y) d\nu(x, y) + \epsilon H(\nu) \right].$$

- Pal and Wong ('20) considered a particular cost and decaying $\epsilon = \epsilon_N$.
- We consider an arbitrary cost c and fixed ϵ.



Theorem 1 (Consistency, Harchaoui, L., Pal '20)

Under appropriate assumptions, the discrete Schrödinger bridge $\hat{\mu}$ converges weakly to the continuum Schrödinger bridge μ, in probability, as N $\to \infty$.

To get the CLT-type result, we consider $\int \eta d\hat{\mu} - \int \eta d\mu$.

- η arbitrary test function.
- $\int \eta d\hat{\mu}$ the *EOT statistic*.
- $\int \eta d\mu$ the population parameter.

Main results

Theorem 2 (First order chaos, Harchaoui, L., Pal '20)
Under appropriate assumptions, $\exists$ functions $f := f^\eta$ and $g := g^\eta$ such that

$$\int \eta(d\hat{\mu} - d\mu) = \mathcal{L}_1 + o_p(N^{-1/2}),$$

where

$$\mathcal{L}_1 := \frac{1}{N}\sum_{i=1}^{N}[f(X_i) + g(Y_i)].$$

We call $\mathcal{L}_1$ the first order chaos of the EOT statistic $\int \eta d\hat{\mu}$.

Corollary 3 (Functional CLT)
Under appropriate assumptions, $\sqrt{N}\int \eta(d\hat{\mu} - d\mu) \to_d \mathcal{N}(0, \varsigma^2)$, where

$$\varsigma^2 := \varsigma^2(\eta) := \int f^2(x)\rho_0(x)dx + \int g^2(y)\rho_1(y)dy.$$

Theorem 4 (Second order chaos, Harchaoui, L., Pal '20)

Under appropriate assumptions, there exists $C := C^\eta$, $f := f^\eta$, $g := g^\eta$ and $h := h^\eta$ such that

$$\int \eta \, d(\hat{\mu} - \mu) = \mathcal{L}_1 + \mathcal{L}_2 - \frac{C}{N} + o_p(N^{-1}),$$

where

$$\mathcal{L}_2 := \frac{1}{N(N-1)} \Big\{ \sum_{i \neq j} [f(X_i, X_j) + g(Y_i, Y_j)] + \sum_{i,j=1}^{N} h(X_i, Y_j) \Big\}.$$

We call $\mathcal{L}_2$ the second order chaos of the EOT statistic $\int \eta d\hat{\mu}$.

Corollary 5 (Second order functional convergence)

Under appropriate assumptions, there exists $\{\lambda_{kl}\}, \{s_k\}$ *such that*

$$N\left[\int \eta \, d(\hat{\mu} - \mu) - \mathcal{L}_1\right] + C \to_d Z,$$

where

$$Z := \sum_{k,l\geq 1} \lambda_{kl} \left\{U_k V_l + s_k s_l U_l V_k - s_l U_k U_l - s_k V_k V_l + (s_l + s_k)\mathbb{1}\{k = l\}\right\},$$

and $\{U_k\}, \{V_k\}$ *are two independent i.i.d. samples from* $\mathcal{N}(0, 1)$.

Remark. In particular, we can choose η to be the cost c.

Contiguity

Definition 6 (Le Cam '60)

Consider two sequences of probability measures $(P^N, N \geq 1)$ and $(Q^N, N \geq 1)$. We say P^N is contiguous w.r.t. Q^N, denoted by

$$P^N \lhd Q^N, \text{ if } Q^N(A_N) \to 0 \text{ implies } P^N(A_N) \to 0.$$

- An asymptotic version of absolute continuity:

$$P \ll Q \text{ if } Q(A) = 0 \text{ implies } P(A) = 0.$$

- If $P^N \lhd Q^N$, then $Z_N = o_p(1)$ under Q^N implies $Z_N = o_p(1)$ under P^N.

Contiguity

We will change the model $(X_i, Y_i) \overset{\text{i.i.d.}}{\sim} \rho_0 \otimes \rho_1$ to $(X_i, Y_i) \overset{\text{i.i.d.}}{\sim} \mu$.

- $(\rho_0 \otimes \rho_1)^\infty$ and μ^∞ are mutually singular.
- The empirical measures $\hat{\rho}_0 \to_{a.s.} \rho_0$ and $\hat{\rho}_1 \to_{a.s.} \rho_1$ under both models.
- Consider the law of $(\hat{\rho}_0, \hat{\rho}_1)$ under $(\rho_0 \otimes \rho_1)^N$ and under μ^N—P^N and Q^N.

P^N is contiguous[1] w.r.t. Q^N.

- $Z_N = o_p(1)$ under Q^N implies $Z_N = o_p(1)$ under P^N.
- If $Z_N := Z_N(\hat{\rho}_0, \hat{\rho}_1)$, then $o_p(1)$ under μ^N implies $o_p(1)$ under $(\rho_0 \otimes \rho_1)^N$.

How to characterize a function of $(\hat{\rho}_0, \hat{\rho}_1)$? Permutation symmetry.

$$f(X_1, \ldots, X_N, Y_1, \ldots, Y_N) = f(X_{\sigma_1}, \ldots, X_{\sigma_N}, Y_{\tau_1}, \ldots, Y_{\tau_N}).$$

[1] See Theorem 7 in (Harchaoui, L., Pal '20).

Theorem 1: consistency

Recall Theorem 1: $\hat{\mu}$ converges weakly to μ, in probability.

Proof sketch of Theorem 1.

- Change the measure to $(X_i, Y_i) \overset{\text{i.i.d.}}{\sim} \mu$ (Schrödinger bridge).
- $\int \eta d\hat{\mu} = \mu^N[\eta(X_1, Y_1) \mid \mathcal{G}_N]$ with

  $\mathcal{G}_N$ is the σ-algebra generated by empirical measures $\hat{\rho}_0$ and $\hat{\rho}_1$.

- By reverse martingale convergence theorem,

$$\int \eta d\hat{\mu} \to_{a.s.} \int \eta d\mu \text{ under } \mu^N.$$

- Contiguity to pull back the result to $(X_i, Y_i) \overset{\text{i.i.d.}}{\sim} \rho_0 \otimes \rho_1$.

$\square$

Remark. $\int \eta d\hat{\mu}$ is unbiased under μ^N, i.e., $\mu^N\left[\int \eta d\hat{\mu}\right] = \int \eta d\mu$.

$\mathbf{L}^2$ projection for paired samples

Recall $\int \eta d\mu = \mu^N \left[\int \eta d\hat{\mu} \right]$.

- $\mathbf{L}^2$ projection of $\int \eta d\hat{\mu}$ onto the space of constants.

What's the "best" $f(X_1) + g(Y_1)$ that approximates $\int \eta d(\hat{\mu} - \mu)$?

- $\mathbf{L}^2$ projection of $\int \eta d(\hat{\mu} - \mu)$ onto $\mathrm{Span}\{f(X_1) + g(Y_1)\}$.
- How to compute?

Matching conditional expectations:

$$\mu^N \left[\int \eta \, d(\hat{\mu} - \mu) \mid X_1 \right] = \mu[f(X_1) + g(Y_1) \mid X_1] = f(X_1) + \mu[g(Y_1) \mid X_1]$$
$$\mu^N \left[\int \eta \, d(\hat{\mu} - \mu) \mid Y_1 \right] = \mu[f(X_1) + g(Y_1) \mid Y_1] = \mu[f(X_1) \mid Y_1] + g(Y_1).$$

$\mathbf{L}^2$ projection for paired samples

How to compute $\mu^N\left[\int \eta d(\hat{\mu}-\mu) \mid X_1\right]$?

By the tower property and symmetry,

$$\mu^N\left[\int \eta \, d(\hat{\mu}-\mu) \mid X_1\right] = \frac{1}{N}\mu\left[\eta(X_1, Y_1) - \int \eta d\mu \mid X_1\right] =: \frac{1}{N}\kappa_{1,0}(X_1)$$
$$\mu^N\left[\int \eta \, d(\hat{\mu}-\mu) \mid Y_1\right] = \frac{1}{N}\mu\left[\eta(X_1, Y_1) - \int \eta d\mu \mid Y_1\right] =: \frac{1}{N}\kappa_{0,1}(Y_1).$$

How to compute $\mu[f(X_1) \mid Y_1]$?

- Direct computation.
- Markov operators induced by Markov transition kernels.

Projection with Markov operators

Recall $\mu(x, y) = \xi(x, y)\rho_0(x)\rho_1(y)$ and

$$\int \xi(x, y)\rho_0(x)dx \overset{\text{a.s.}}{=} 1 \quad \text{and} \quad \int \xi(x, y)\rho_1(y)dy \overset{\text{a.s.}}{=} 1.$$

Conditional probability densities

$$p_{X_1|Y_1}(x \mid y) = \xi(x, y)\rho_0(x) \quad \text{and} \quad p_{Y_1|X_1}(y \mid x) = \xi(x, y)\rho_1(y).$$

Markov operators $\mathcal{A} : \mathbf{L}^2(\rho_0) \to \mathbf{L}^2(\rho_1)$ and $\mathcal{A}^* : \mathbf{L}^2(\rho_1) \to \mathbf{L}^2(\rho_0)$,

$$\mathcal{A}f(y) := \int f(x)\xi(x, y)\rho_0(x)dx = \mu[f(X_1) \mid Y_1](y)$$

$$\mathcal{A}^*g(x) := \int g(y)\xi(x, y)\rho_1(y)dy = \mu[g(Y_1) \mid X_1](x).$$

Projection with Markov operators

Goal: compute the projection of $\int \eta d(\hat{\mu} - \mu)$ onto $\mathrm{Span}\{f(X_1) + g(Y_1)\}$.
Strategy: match conditional expectations:

$$\frac{1}{N}\kappa_{1,0}(x) = f(x) + \mathcal{A}^* g(x) \quad \text{and} \quad \frac{1}{N}\kappa_{0,1}(y) = g(y) + \mathcal{A} f(y).$$

Solutions:

$$\begin{aligned} f &= \frac{1}{N}(I - \mathcal{A}^*\mathcal{A})^{-1}(\kappa_{1,0} - \mathcal{A}^*\kappa_{0,1}) \\ g &= \frac{1}{N}(I - \mathcal{A}\mathcal{A}^*)^{-1}(\kappa_{0,1} - \mathcal{A}\kappa_{1,0}). \end{aligned}$$

$$Y \xrightarrow{\mathcal{A}} X \xrightarrow{\mathcal{A}^*} Y$$

$$X \xrightarrow{\mathcal{A}^*} Y \xrightarrow{\mathcal{A}} X$$

Theorem 2: first order chaos

What's the projection of $\int \eta d(\hat{\mu} - \mu)$ onto $\mathrm{Span}\{ \sum_{i=1}^{N} [f(X_i) + g(Y_i)] \}$?

First order chaos $\mathcal{L}_1$:

$$\frac{1}{N} \sum_{i=1}^{N} \Big[(I - \mathcal{A}^* \mathcal{A})^{-1} (\kappa_{1,0} - \mathcal{A}^* \kappa_{0,1})(X_i) + (I - \mathcal{A}\mathcal{A}^*)^{-1} (\kappa_{0,1} - \mathcal{A} \kappa_{1,0})(Y_i) \Big] .$$

It can be shown that[2]

$$\int \eta \; d(\hat{\mu} - \mu) - \mathcal{L}_1 = o_p(N^{-1/2}), \quad \text{under } \mu^N .$$

Remark. This is an extension of the Hoeffding decomposition in the U-statistics theory, where $(X_i, Y_i) \overset{\text{i.i.d.}}{\sim} \rho_0 \otimes \rho_1$.

[2]See Proposition 28 in (Harchaoui, L., Pal '20).

Theorem 3: second order chaos

What about second order terms?

- Projection onto

$$\mathrm{Span}\left\{\sum_{i\neq j}[f(X_i, X_j) + g(Y_i, Y_j)] + \sum_{i,j=1}^{N} h(X_i, Y_j)\right\}.$$

- *Tensor products* of operators, e.g., $\mathcal{A} \otimes \mathcal{A}^* : \mathbf{L}^2(\rho_0 \otimes \rho_1) \to \mathbf{L}^2(\rho_1 \otimes \rho_0)$,

$$(\mathcal{A} \otimes \mathcal{A}^*) f(Y_1, X_2) = \mu^2[f(X_1, Y_2) \mid Y_1, X_2].$$

Building block $\mathcal{C} := (I - \mathcal{A}^*\mathcal{A}) \otimes (I - \mathcal{A}\mathcal{A}^*)$.

- $\mathcal{C}^{-1}$ is well-defined on a proper domain.
- $\mathcal{C}^{-1} = (I - \mathcal{A}^*\mathcal{A})^{-1} \otimes (I - \mathcal{A}\mathcal{A}^*)^{-1}$.

Control of remainders

Let $R_1 := \int \eta d(\hat{\mu} - \mu) - \mathcal{L}_1$. By a change of measure,

$$\mu^N[|R_1|] \leq \sqrt{\mathbb{E}[U_N^2]},$$

where U_N, the numerator of R_1, reads

$$U_N := \frac{1}{N!} \sum_{\sigma \in \mathcal{S}_N} \frac{1}{N} \widetilde{\eta}(X, Y_\sigma) \xi^{\otimes}(X, Y_\sigma)$$

- $\widetilde{\eta}$ is the *degenerate* version of η.
- $\xi^{\otimes}(X, Y_\sigma) := \prod_{i=1}^N \xi(X_i, Y_{\sigma_i})$.

How to control $\mathbb{E}[U_N^2]$?

- Hoeffding decomposition up to the Nth order terms.
- Variance bound using a chain of Markov operators.

Schrödinger bridge in the continuum

Suppose $(Z(0), Z(1))$ is distributed according to a Markov transition kernel

$$p(x, y) := \frac{1}{\Lambda(x)} \exp\left(-\frac{1}{\epsilon} c(x, y)\right).$$

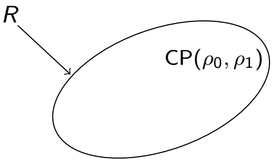
What is $\mathrm{Law}(Z(0), Z(1))$, conditioned on $Z(0) \sim \rho_0$ and $Z(1) \sim \rho_1$?

A first guess: $R(x, y) := \rho_0(x) p(x, y)$.

- A valid probability distribution.
- Marginal of x is ρ_0.
- Marginal of y may NOT be ρ_1.

How to fix it? Information projection.

Schrödinger bridge in the continuum



The joint law is the *I-projection* (Föllmer '88, Léonard '12)

$$\mu := \arg\min_{\nu \in \mathrm{CP}(\rho_0, \rho_1)} H(\nu \mid R),$$

where $H(\nu_1 \mid \nu_2) = \int \log\left(\frac{d\nu_1}{d\nu_2}\right) d\nu_1$, if $\nu_1 \ll \nu_2$, and infinity otherwise.

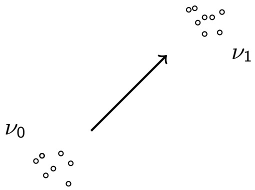
- μ is called the **Schrödinger bridge** connecting ρ_0 and ρ_1.
- A simple algebra shows

$$\mu = \arg\min_{\nu \in \mathrm{CP}(\rho_0, \rho_1)} \left[\int c(x, y) d\nu(x, y) + \epsilon H(\nu) \right].$$

The Schrödinger problem

Schrödinger's lazy gas experiment.

- Gas particles moving as BM(ϵ).
- Initial configuration $L_N(0) \approx \nu_0$.
- Question: conditioned on $\{L_N(1) \approx \nu_1\}$, what are the most likely paths of these particles?



How to get the paths?

- Solve for the Schrödinger bridge, i.e., Law($Z(0), Z(1)$).
- Connect $Z(0)$ and $Z(1)$ with a Brownian bridge.

Discrete Schrödinger bridge

Consider

- N i.i.d. particles $\{Z_i\}_{i=1}^N$.
- $(Z_i(0), Z_i(1))$ is distributed as a Markov transition kernel

$$p(x, y) := \frac{1}{\Lambda(x)} \exp\left(-\frac{1}{\epsilon} c(x, y)\right).$$

- Two discrete measures $\nu_0 := \frac{1}{N}\sum_{i=1}^N \delta_{x_i}$ and $\nu_1 := \frac{1}{N}\sum_{i=1}^N \delta_{y_i}$.

Conditioned on $L_N(0) = \nu_0$ and $L_N(1) = \nu_1$, what is the joint law

$$\frac{1}{N}\sum_{i=1}^N \delta_{(Z_i(0), Z_i(1))}?$$

$\sum_{\sigma\in\mathcal{S}_N} q^*(\sigma) \frac{1}{N}\sum_{i=1}^N \delta_{(x_i, y_{\sigma_i})}!$

Summary

We are interested in the Schrödinger bridge:

$$\mu := \underset{\nu \in \mathrm{CP}(\rho_0, \rho_1)}{\arg\min} \left[\int c(x,y) d\nu(x,y) + \epsilon H(\nu) \right].$$

We proposed the discrete Schrödinger bridge using the Gibbs measure q^*:

$$\hat{\mu} := \sum_{\sigma \in \mathcal{S}_N} q^*(\sigma) \frac{1}{N} \sum_{i=1}^{N} \delta_{(X_i, Y_{\sigma_i})}.$$

We proved that

- $\hat{\mu}$ converges weakly to μ in probability.
- Functional CLT for $\hat{\mu}$.
- Second order functional Gaussian chaos limit when the Gaussian limit is degenerate.

Thank you!

Paper: arxiv.org/abs/2011.08963

Webpage: langliu95.github.io

Change of measure

The Gibbs measure q^* can be rewritten as

$$q^*(\sigma) = \frac{\prod_{i=1}^N \mu(X_i, Y_{\sigma_i})}{\sum_{\tau \in \mathcal{S}_N} \prod_{i=1}^N \mu(X_i, Y_{\tau_i})}. \tag{11}$$

σ gets the largest probability if $\{(X_i, Y_{\sigma_i})\}_{i=1}^N$ is roughly i.i.d. from μ.

Remark. This change of measure is reminiscent of exponential tilting—we change the original product measure to μ by adding an exponential factor $\exp\left(-\frac{1}{\epsilon}(c(x, y) - a(x) - b(y))\right)$.

Discrete Schrödinger bridge

Recall that the discrete Schrödinger bridge is

$$\hat{\mu} := \sum_{\sigma \in \mathcal{S}_N} q^*(\sigma) \frac{1}{N} \sum_{i=1}^{N} \delta_{(X_i, Y_{\sigma_i})}.$$

To see this, assume $X_i = x_i$, leading to the initial law $\nu_0 := \frac{1}{N}\sum_{i=1}^N \delta_{x_i}$.

- Given $\nu_1 := \frac{1}{N}\sum_{i=1}^N \delta_{y_i}$, the event $\{L_N(1) = \nu_1\} = \cup_{\sigma \in \mathcal{S}_N} E_\sigma$, where

$$E_\sigma := \{Y_i = y_{\sigma_i} : i \in [N]\}.$$

- On E_σ, the joint law $\frac{1}{N}\sum_{i=1}^N \delta_{(X_i, Y_i)}$ reads $\frac{1}{N}\sum_{i=1}^N \delta_{(x_i, y_{\sigma_i})}$.
- The conditional probability of E_σ given $L_N(1) = \nu_1$ is

$$\frac{\prod_{i=1}^N \frac{1}{\Lambda_\epsilon(x_i)} \exp\left(-\frac{1}{\epsilon} c(x_i, y_{\sigma_i})\right) dx_i dy_i}{\sum_{\tau \in \mathcal{S}_N} \prod_{i=1}^N \frac{1}{\Lambda_\epsilon(x_i)} \exp\left(-\frac{1}{\epsilon} c(x_i, y_{\tau_i})\right) dx_i dy_i} = \frac{\exp\left(-\frac{1}{\epsilon} c(x, y_\sigma)\right)}{\sum_{\tau \in \mathcal{S}_N} \exp\left(-\frac{1}{\epsilon} c(x, y_\tau)\right)},$$

which is exactly $q^*(\sigma)$ (with $X_i = x_i$ and $Y_i = y_i$).