Schrödinger Bridges: Old and New

Workshop on Stochastic Planning & Control of Dynamical Systems

IEEE CDC 2025

Tryphon T. Georgiou

University of California, Irvine

Rio de Janeiro

Acknowledgements

Thanks to:

Yongxin Chen
Michele Pavon
Asmaa Eldesoukey
Olga Movilla Miangolarra
Ralph Sabbagh

Air Force Office of Scientific Research
Army Research Office
National Science Foundation

Schrödinger Bridges - background & classical concepts

- **Entropy & Relative entropy**
  manifestations
- **Schrödinger's Bridge problem**
  static & dynamic
  Markov chains, diffusion processes
- **Fortet-Sinkhorn algorithm**
  Hilbert metric
- **Stochastic control & steering**
- **A bit on quantum**

Relative Entropy

Kullback-Leibler divergence

P, Q probability laws on any measurable space $\mathcal{X}$ ($dQ \gg dP$),

$$\mathbb{D}(P\|Q) := \int_{\mathcal{X}} dP \log\left(\frac{dP}{dQ}\right)$$
$$= \mathbb{E}_Q\left\{\Lambda \log\left(\Lambda\right)\right\}, \text{ where } \Lambda = \frac{dP}{dQ}$$

If $dQ \not\gg dP$, then $\mathbb{D}(P\|Q) := \infty$

$\mathbb{D}(P\|Q)$ jointly convex, and ≥ 0 always

Relative Entropy

origins

- **degradation of coding efficiency:**
  word-length increase on average when using the wrong code

Average word-length (optimal code) $= -\sum_k p_k \log(p_k)$, i.e., entropy rate

Average word-length using code designed for $\sim q_k$, $-\sum_k p_k \log(q_k)$

Degradation:

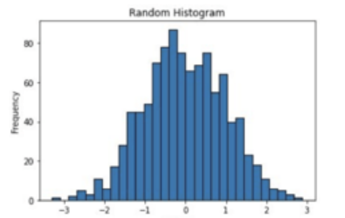
$$\underbrace{-\sum_k p_k \log(q_k)}_{\text{suboptimal}} - \underbrace{(-\sum_k p_k \log(p_k))}_{\text{optimal}} = \mathbb{D}(P \| Q)$$

Relative Entropy

- **quantifying likelihood of rare events:**
  the probability that an empirical average is far away from its mean

Sanov's theorem:

Independent samples X_t $(t \in \{1, \ldots, N\})$, distributed $X_t \sim Q$



Empirical distribution P_N (random histogram)
$P_N(A) = \frac{1}{N} \sum_{t=1}^{N} \mathbb{1}_{X_t \in A}$

Suppose $\mathcal{P}$ is a convex set of distributions,
and $P^\star = \arg\min_{P \in \mathcal{P}} \mathbb{D}(P \| Q)$

$$\mathbb{P}\{P_N \in \mathcal{P}\} \simeq e^{-N \cdot \mathbb{D}(P^\star \| Q)}$$

$P^\star$ representative of P_N in "neighborhood" $\mathcal{P}$

Relative Entropy

- **likelihood estimation:**
  most likely law consistent with statistics/moments

  example: Assuming, e.g., $X \in \{0, \ldots, n\}$ is distributed $X \sim Q$ (prior)

  and given estimated statistics/moments, e.g., $\bar{x} = \frac{1}{N} \sum_{k=1}^{N} X_t$
  what can we say about the distribution of the N-samples?

The most likely (posterior) is:

$$P^* = \arg \min_{\sum_{k=0}^{n} k P_k = \bar{x}} \mathbb{D}(P \| Q)$$

i.e., the closest to the prior that is consistent with the data

Relative Entropy

- **Reconcile statistical data**

  origin in statistics, contigency tables

Example:
X, Y jointly distributed on $\{0, 1, \ldots, n\}$, with prior $Q(x, y)$,
and given (empirical) marginals $p_X(x), p_Y(y)$,
find a most likely posterior $P^\star(x, y)$ in agreement with p_X, p_Y.

$$P^\star = \arg\min_P \left\{ \mathbb{D}(P \| Q) \mid \sum_x P(x, y) = p_Y(y),\ \sum_y P(x, y) = p_X(x) \right\}$$

Form of solution - diagonal scaling[1]

$P^\star = \arg\min_P \left\{ \mathbb{D}(P \| Q) \mid \sum_x P(x,y) = p_Y(y),\ \sum_y P(x,y) = p_X(x) \right\}$

$$\begin{aligned}\mathcal{L}(P,a,b) := &\sum_x \sum_y P(x,y) \log\left(\frac{P(x,y)}{Q(x,y)}\right) \\ &+ \sum_x a(x)(\sum_y P(x,y) - p_X(x)) \\ &+ \sum_y b(y)(\sum_x P(x,y) - p_Y(y))\end{aligned}$$

$$\frac{\partial}{\partial P(x,y)}\mathcal{L} = 0 \Rightarrow \log\left(\frac{P(x,y)}{Q(x,y)}\right) = -1 + a(x) + b(y)$$

$$P^\star(x,y) = e^{-1+a(x)} Q(x,y) e^{b(y)}$$

[1] Sinkhorn-Knopp, Marshall & Olkin, and earlier Schrödinger, Fortet

Fortet-Sinkhorn's algorithm

$$P^\star(x,y) = e^{-1+a(x)} Q(x,y) e^{b(y)} = D_{\text{left}}(x) Q(x,y) D_{\text{right}}(y)$$

Algorithm: Given matrix Q, and vectors p_X, p_Y

Start with $P = Q = [Q(x,y)]_{x,y=1}^n$

$$P \to D_\ell P \text{ where } D_\ell \text{ diagonal, } D_\ell(x) = \frac{p_X(x)}{\sum_y P(x,y)} \quad \text{s.t. } \sum_y D_\ell(y) P(x,y) = p_X(x)$$

$$P \to P D_r \text{ where } D_r \text{ diagonal, } D_r(y) = \frac{p_Y(y)}{\sum_x P(x,y)} \quad \text{s.t. } \sum_x P(x,y) D_r(y) = p_Y(y)$$

repeat until convergence

If $Q(x,y) > 0$ for all x, y convergence is guaranteed.

Applies to multi-marginals and higher-dimensional arrays $Q(x,y,z,\ldots)$, etc.

Schrödinger's bridge problem

for Markov chains

Markov chain $X_t \in \{0, \dots, n\}$
Prior law: $X_0 \sim q_0$, transition probabilities $\Pi_0(x_0, x_1), \Pi_1(x_1, x_2), \dots, \Pi_{T-1}(x_{T-1}, x_T)$.
Data: empirical marginals $X_0 \sim \mathrm{p}_0$, $X_T \sim \mathrm{p}_T$ when $q_0 \neq p_0$ and/or $q_T \neq p_T$
Find the most likely evolution

Path probability/measure:

Prior path probability $Q(x_0, \dots, x_T) = q_0(x_0)\Pi_0(x_0, x_1) \cdots \Pi_{T-1}(x_{T-1}, x_T)$
Posterior path probability $P(x_0, \dots, x_T) = p_0(x_0)\hat{\Pi}_0(x_0, x_1) \cdots \hat{\Pi}_{T-1}(x_{T-1}, x_T)$

Find: transition probabilities

$$P^\star = \arg\min \left\{ \mathbb{D}(P \| Q) \mid \sum_{x_1, \dots, x_T} P(x_1, \dots, x_T) = \mathrm{p}_0(x_0), \quad \sum_{x_0 \dots, x_{T-1}} P(x_0, \dots, x_{T-1}) = \mathrm{p}_T(x_T). \right\}$$

$P^\star = \operatorname{argmin}\{\mathbb{D}(P\|Q) \mid P \in \mathcal{P}(\mathsf{p}_0, \mathsf{p}_T)\}$

Disintegration: Q with respect to the initial and final positions,

$$Q(x_0, x_1, \ldots, x_T) = \underbrace{Q_{x_0,x_T}(x_1, \ldots, x_{T-1})}_{\text{pinned bridge}}\, q_{0T}(x_0, x_T)$$

where $Q_{x_0,x_T}(\cdot) = Q\{\cdot \,|X(0) = x_0, X(T) = x_T\}$; similarly for P

$$\mathbb{D}(P\|Q) = \underbrace{\sum_{x_0 x_T} p_{0T}(x_0, x_T) \log \frac{p_{0T}(x_0, x_T)}{q_{0T}(x_0, x_T)}}_{\geq 0} + \underbrace{\sum_{x} P_{x_0,x_T}(x_{\ldots}) \log \frac{P_{x_0,x_T}(x_{\ldots})}{Q_{x_0,x_T}(x_{\ldots})} q_{0T}(x_0, x_T)}_{\geq 0}$$

$\Rightarrow$ 2nd term $= 0$ when P, Q share pinned bridges

$\Rightarrow$ need to minimize the coupling p_{0T} subject to marginals

For $T = 1$

$$\hat{\Pi}^* = \text{argmin}\{\sum_{x_0 x_1} p_{01}(x_0, x_1) \log \frac{p_{01}(x_0, x_1)}{q_{01}(x_0, x_1)}\}$$

$p_{01}(x_0, x_1) = p(x_0)\hat{\Pi}(x_0, x_1)$, $q_{01}(x_0, x_1) = q(x_0)\Pi(x_0, x_1)$,

$$\mathbb{D}(p_0(\cdot)\hat{\Pi}(\cdot,\cdot)\|q_0(\cdot)\Pi(\cdot,\cdot)) = \sum_{x_0,x_1} p(x_0)\hat{\Pi}(x_0, x_1)\left(\log(\frac{p(x_0)}{q(x_0)}) + \log(\frac{\hat{\Pi}(x_0, x_1)}{\Pi(x_0, x_1)})\right)$$

transition probability: $\sum_{x_1} \hat{\Pi}(x_0, x_1) = 1$

$P^\star = \mathrm{argmin}\{\mathbb{D}(P\|Q) \mid P \in \mathcal{P}(\mathsf{p}_0, \mathsf{p}_T)\}$

$$\hat{\Pi}^* = \operatorname{argmin}\left\{ \sum_{x_0,x_1} p(x_0)\hat{\Pi}(x_0,x_1)\log\left(\frac{\hat{\Pi}(x_0,x_1)}{\Pi(x_0,x_1)}\right) \mid \sum_{x_0} p_0(x_0)\hat{\Pi}(x_0,x_1) = p_1(x_1) \right.$$
$$\left. \sum_{x_1} \hat{\Pi}(x_0,x_1) = 1 \right\}$$

$\Rightarrow$

$$\begin{aligned} \hat{\Pi}^\star(x_0,x_1) &= \mathrm{left}(x_0)\Pi(x_0,x_1)\mathrm{right}(x_1) \\ &= \phi_0(x_0)^{-1}\Pi(x_0,x_1)\phi_1(x_1) \end{aligned}$$

A brief interlude on the Hilbert metric

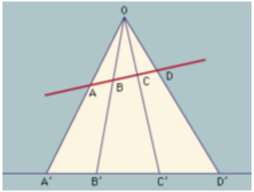
The Hilbert projective metric

Pappus of Alexandria - cross ratio

Convex bounded $\Omega \subset \mathbb{R}^n$

For $B, C \in \Omega$ and A, D points of intersect of AB line with boundary of Ω

$$d_H(A, B) := \log \left(\frac{|BA| \cdot |CD|}{|BD| \cdot |CA|} \right).$$



The Hilbert projective metric

Convex cone $K \subset$ Banach space

- Pointed: $K \cap (-K) = \{0\}$
- Partial order $p \geq q \Leftrightarrow p - q \in K$

$$\begin{aligned} \bar{\lambda}(p,q) &:= \inf\{\lambda \mid p \leq \lambda q\} \\ \underline{\lambda}(p,q) &:= \sup\{\lambda \mid \lambda q \leq p\} \end{aligned}$$

$$d_H(p,q) := \log \frac{\bar{\lambda}(p,q)}{\underline{\lambda}(p,q)}$$

Examples:
positive cone in $\mathbb{R}$
positive definite Hermitian matrices

Hilbert 1895
Birkhoff 1957
Bushell 1973

The Hilbert projective metric

Projective diameter: $\mathrm{diam}(\mathrm{range}(\Pi)) := \sup\{d_H(\Pi(x), \Pi(y)) \mid x, y \in K \backslash \{0\}\}$

Contraction ratio: $\|\Pi\|_H = \inf\{\lambda \mid d_H(\Pi(x), \Pi(y)) \leq \lambda d_H(x, y),\ x, y \in K \backslash \{0\}\}$

Birkhoff-Bushell theorem

Π positive, monotone, homogeneous of degree m, i.e., $\Pi : K \to K$, cone in $\mathbb{R}^n$

$x \leq y \Rightarrow \Pi(x) \leq \Pi(y)$

$\Pi(\alpha x) = \alpha^m \Pi(x)$

Then $\|\Pi\|_H \leq m$, and if, in addition, Π is linear:

$$\|\Pi\|_H = \tanh\left(\frac{1}{4}\mathrm{diam}(\Pi)\right)$$

Corollary: If linear $\Pi : K \to \mathrm{interior}(K)$, then $\|\Pi\|_H < 1$

Bridge for one-step Markov Chain

Start with a stochastic matrix (row sum = 1):

$$\Pi = [\Pi_{x_0,x_T}]_{x_0,x_T=1}^{N}, \text{ with positive entries}$$

& two probability vectors p_0, p_N with strictly positive entries

Schrödinger system

There exist $\phi(0,x_0)$, $\phi(T,x_T)$, $\hat{\phi}(0,x_0)$, $\hat{\phi}(T,x_T)$, $x_0, x_T \in \{1,\ldots,N\}$ such that:

$$\begin{aligned}
\phi(0,x_0) &= \sum_{x_T} \Pi_{x_0,x_T}\phi(T,x_T) \\
\hat{\phi}(T,x_T) &= \sum_{x_0} \Pi_{x_0,x_T}\hat{\phi}(0,x_0) \\
\phi(0,x_0)\hat{\phi}(0,x_0) &= p_0(x_0) \\
\phi(T,x_T)\hat{\phi}(T,x_T) &= p_T(x_T)
\end{aligned}$$

Bridge for one-step Markov Chain

Circular composition of maps:

$$
\begin{array}{ccccccc}
 & & \hat{\phi}(0,x_0) & \xrightarrow{\Pi^T} & \hat{\phi}(T,x_T) & = \sum_{x_0} \Pi_{x_0,x_T} \hat{\phi}(0,x_0) \\
\hat{\phi}(0,x_0) = \frac{p_0(x_0)}{\phi(0,x_0)} & & \uparrow & & \downarrow & \phi(T,x_T) = \frac{p_T(x_T)}{\hat{\phi}(T,x_T)} \\
\sum_{x_N} \Pi_{x_0,x_N} \phi(T,x_T) = & & \phi(0,x_0) & \xleftarrow{\Pi} & \phi(T,x_T) &
\end{array}
$$

The composition

$$
\hat{\phi}(0,x_0) \xrightarrow{\Pi^T} \hat{\phi}(T,x_T) \xrightarrow{\mathcal{D}_T} \phi(T,x_T) \xrightarrow{\Pi} \phi(0,x_0) \xrightarrow{\mathcal{D}_0} \left(\hat{\phi}(0,x_0)\right)_{\text{next}}
$$

is contractive in the Hilbert metric

$$
\mathcal{D}_0 \;:\; \phi(0,x_0) \mapsto \hat{\phi}(0,x_0) = \frac{p_0(x_0)}{\phi(0,x_0)} \text{ and } \mathcal{D}_T \;:\; \hat{\phi}(T,x_T) \mapsto \phi(T,x_T) = \frac{p_T(x_N)}{\hat{\phi}(T,x_T)}
$$

Bridge for one-step Markov Chain

- the ranges of Π^T, Π are strictly in the interior of the cone,

$$\|\Pi\|_H, \|\Pi^T\|_H < 1.$$

- $\mathcal{D}_0$ and $\mathcal{D}_T$ inversion/element-wise scaling are isometries in the Hilbert metric

... a bit more, since Hilbert is a projective metric

The Schrödinger system has a solution (unique up to scaling)

inversion/element-wise scaling = isometries

$$
\begin{aligned}
d_H([x_i],[y_i]) &= \log\left((\max_i(x_i/y_i))\frac{1}{\min_i(x_i/y_i)}\right) \\
&= \log\left(\frac{1}{\min_i((x_i)^{-1}/(y_i)^{-1})}\max_i((x_i)^{-1}/(y_i)^{-1})\right) \\
&= d_H([(x_i)^{-1}],[(y_i)^{-1}])
\end{aligned}
$$

$$
\begin{aligned}
d_H([\mathrm{p}_i x_i],[\mathrm{p}_i y_i]) &= \log\frac{\max_i((\mathrm{p}_i x_i)/(\mathrm{p}_i y_i))}{\min_i((\mathrm{p}_i x_i)/(\mathrm{p}_i y_i))} \\
&= \log\frac{\max_i(x_i/y_i)}{\min_i(x_i/y_i)} = d_H([x_i],[y_i]).
\end{aligned}
$$

$P^\star = \mathrm{argmin}\{\mathbb{D}(P\|Q) \mid P \in \mathcal{P}(\mathsf{p}_0, \mathsf{p}_T)\}$

$$\hat{\Pi}^* = \operatorname{argmin}\left\{\sum_{x_0,x_1} p(x_0)\hat{\Pi}(x_0,x_T)\log\left(\frac{\hat{\Pi}(x_0,x_T)}{\Pi(x_0,x_T)}\right) \mid \sum_{x_0} p_0(x_0)\hat{\Pi}(x_0,x_T) = p_T(x_T) \quad \sum_{x_1}\hat{\Pi}(x_0,x_T) = 1\right\}$$

$\Rightarrow$

$$\begin{aligned}\hat{\Pi}^\star(x_0,x_T) &= \mathrm{left}(x_0)\Pi(x_0,x_T)\mathrm{right}(x_T)\\ &= \phi_0(x_0)^{-1}\Pi(x_0,x_T)\phi_T(x_T)\end{aligned}$$

Schrödinger's bridge

for Markov chains

Markov chain $X_t \in \{0, \dots, n\}$
Prior law: $X_0 \sim q_0$, transition probabilities $\Pi_0(x_0, x_1), \Pi_1(x_1, x_2), \dots, \Pi_{T-1}(x_{T-1}, x_T)$.
Data: empirical marginals $X_0 \sim \mathrm{p}_0$, $X_T \sim \mathrm{p}_T$ when $q_0 \neq p_0$ and/or $q_T \neq p_T$
Find the most likely evolution

Prior path probability $Q(x_0, \dots, x_T) = q_0(x_0)\Pi_0(x_0, x_1) \cdots \Pi_{T-1}(x_{T-1}, x_T)$
Posterior path probability $P^\star(x_0, \dots, x_T) = p_0(x_0)\hat{\Pi}_0(x_0, x_1) \cdots \hat{\Pi}_{T-1}(x_{T-1}, x_T)$

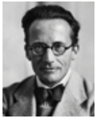
$$P^\star(x_0, \dots, x_T) = p_0(x_0) \overbrace{\left(\phi(0, x_0)^{-1}\Pi_0(x_0, x_1)\phi(1, x_1)\right)}^{\hat{\Pi}_0(x_0, x_1)} \left(\phi(1, x_1)^{-1}\Pi_1(x_1, x_2)\phi(2, x_2)\right) \cdots$$
$$\cdots \left(\phi(T-1, x_{T-1})^{-1}\Pi_{T-1}(x_{T-1}, x_T)\phi(T, x_T)\right)$$

Schrödinger Bridges in earnest

“On the reversal of the laws of nature”

Erwin Schrödinger, 1931

Schrödinger's bridge problem



- Consider a cloud of N independent Brownian particles (N large)
- empirical distributions $\rho_0(x)$ and $\rho_1(y)$ at $t = 0$ and $t = 1$
- ρ_0 and ρ_1 not compatible with transition mechanism

$$\rho_1(y) \neq \int_0^1 \pi(t_0, x, t_1, y)\rho_0(x)dx,$$

where

$$\pi(t_0, y, t_1, x) = \frac{1}{\sqrt{(2\pi)^n(t_1 - t_0)}} e^{-\frac{1}{2}\frac{\|x-y\|^2}{t_1 - t_0}}, \quad s < t$$

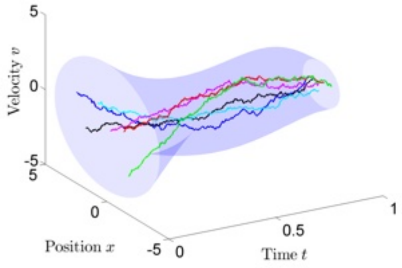
$\Rightarrow$ Particles have been transported in an unlikely way

Schrödinger (1931)

Of the many possible (unlikely) ways, which one is the most likely?

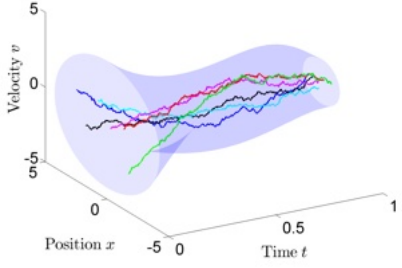
Bridge

Probability law on paths linking two end-point marginals



Bridge

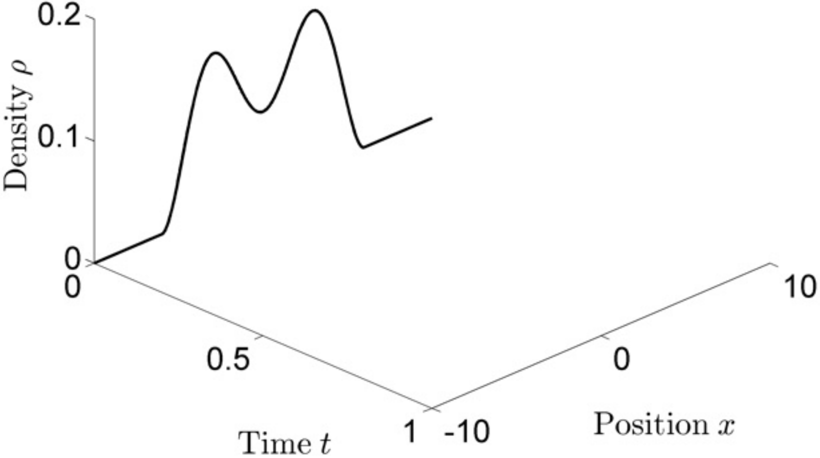
Probability law on paths linking two end-point marginals



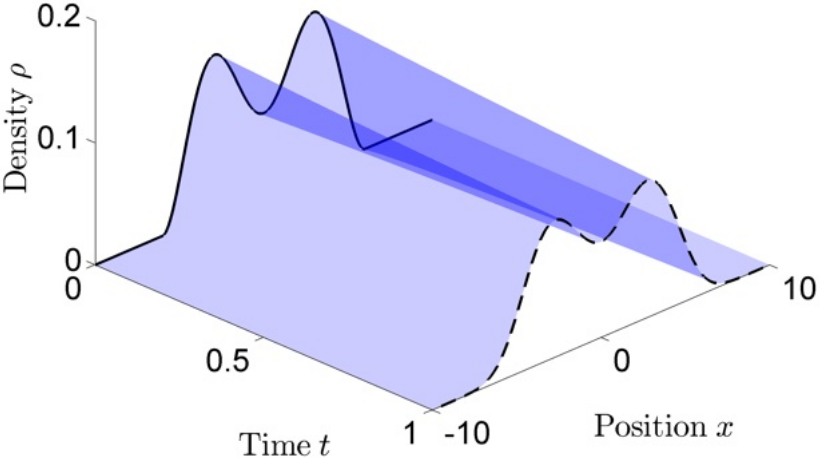
Schrödinger's problem:

- Interpolate in a way that reconciles the two marginals with the prior law
- The new law being the most likely

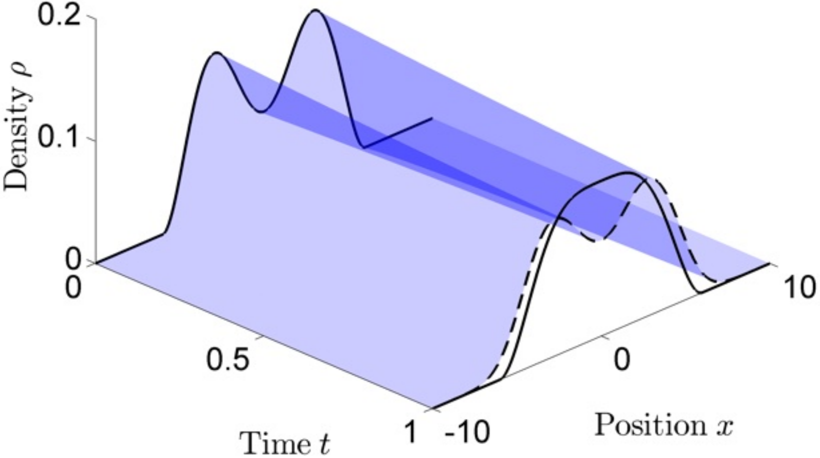
marginal distribution at $t = 0$



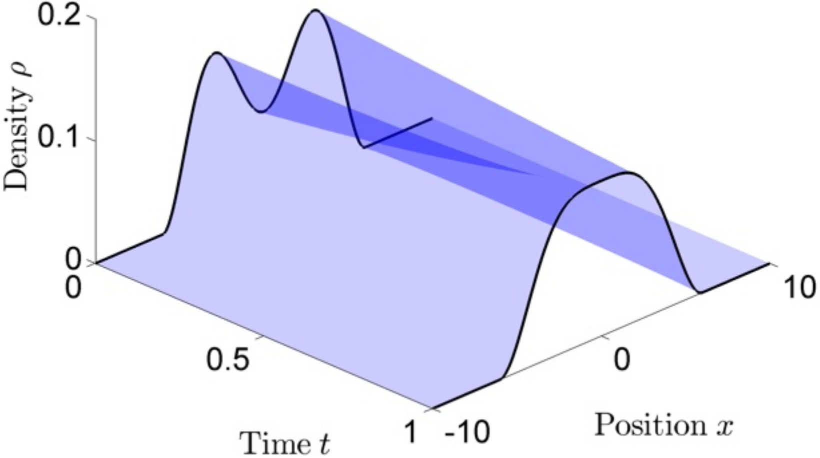
marginal and prior law (flow of one-time densities)



initial marginal, prior law, and end-point marginal



Schrödinger bridge



Föllmer (1988):

Schrödinger's problem concerns **large deviation** of the empirical measure on paths via **Sanov's theorem**

$$\text{Prob}(\mathrm{empirical}\ \mathbb{P}|_{t=0} = \rho_0,\ \mathbb{P}_{t=1} = \rho_1) \simeq e^{-N \int \log\left(\frac{d\mathbb{P}}{d\mathbb{W}}\right) d\mathbb{P}}$$

sampled from the Wiener $\mathbb{W}$: "prior"

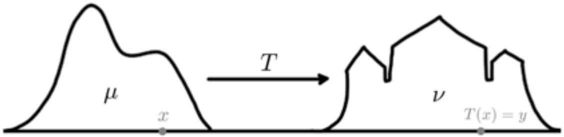
Schrödinger 'sproblem

$$\mathbb{P}^{\star} = \mathrm{argmin} \left\{ \int \log\left(\frac{d\mathbb{P}}{d\mathbb{W}}\right) d\mathbb{P} \ \middle|\ \mathbb{P}|_{t=0} = \rho_0,\ \mathbb{P}_{t=1} = \rho_1 \right\}$$

An brief interlude on Optimal Mass Transport

Optimal Mass Transport

Le mémoire sur les déblais et les remblais
Gaspard Monge 1781



Wasserstein metric

$$W_2(\mu,\nu)^2 := \inf_T \int \|x - \underbrace{T(x)}_{y}\|^2 d\mu(x)$$

where $T\#\mu = \nu$ $\qquad\qquad$ $\mu(dx) = \rho_0 dx, \nu(dx) = \rho_1 dx$

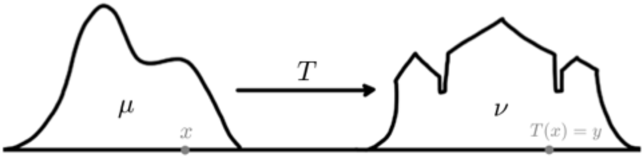
$$\rho_1(x) = \frac{1}{|\det(\nabla T)|}\rho_0(T^{-1}(x))$$

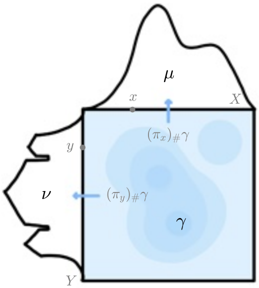
Optimal Mass Transport

$$W_2(\mu,\nu)^2 = \inf_{\pi \in \Pi(\rho_0,\rho_1)} \iint \|x-y\|^2 \quad d\pi(x,y)$$

$\Pi(\mu,\nu)$: "couplings"

$$\int_y \pi(dx,dy) = \rho_0(x)dx = d\mu(x)$$
$$\int_x \pi(dx,dy) = \rho_1(y)dy = d\nu(y)$$



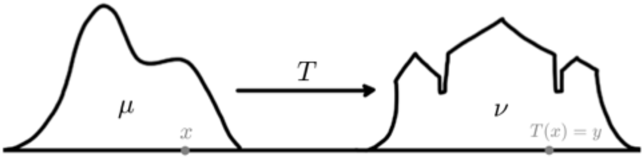


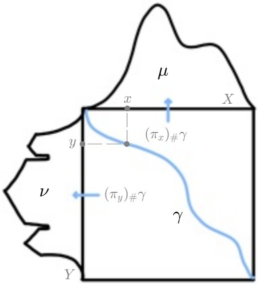
Optimal Mass Transport

$$W_2(\mu,\nu)^2 = \inf_{\pi\in\Pi(\rho_0,\rho_1)} \iint \|x-y\|^2 \quad d\pi(x,y)$$

$\Pi(\mu,\nu)$: "couplings"

$\int_y \pi(dx,dy) = \rho_0(x)dx = d\mu(x)$
$\int_x \pi(dx,dy) = \rho_1(y)dy = d\nu(y)$





Optimal Mass Transport

$$\|x - y\|^2 = \inf\{\int_0^1 \|\dot{x}(t)\|^2 dt \ | \ x(0) = x, \ x(1) = y\}$$

$$W_2(\rho_0, \rho_1)^2 := \inf_{(\rho, v)} t_f \int_{t_0}^{t_f} \int_{\mathbb{R}^n} \rho \|v\|^2 dx dt$$

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (v\rho) = 0$$

$$\rho(x, t_0) = \rho_0(x), \quad \rho(y, t_f) = \rho_1(y)$$

$$W_2(\rho_0, \rho_1)^2 = \inf \underbrace{\int_{\text{time}} \text{average kinetic energy}}_{\text{action integral}}$$

subject to boundary conditions

Riemannian geometry of OMT

ensemble states	$\{\rho \geq 0 : \int \rho = 1\}$
tangent space at ρ are perturbations	$\{\delta : \int \delta = 0\}$

Key insight: $\delta \equiv \frac{\partial \rho}{\partial t} \longleftrightarrow v = \nabla \phi$ (irrotational) via solving

$$\delta = -\nabla \cdot (\rho \nabla \phi)$$

Riemannian geometry of OMT

ensemble states	$\{\rho \geq 0 : \int \rho = 1\}$
tangent space at ρ are perturbations	$\{\delta : \int \delta = 0\}$

Key insight: $\delta \equiv \frac{\partial \rho}{\partial t} \longleftrightarrow v = \nabla \phi$ (irrotational) via solving

$$\delta = -\nabla \cdot (\rho \nabla \phi)$$

Riemannian structure

$$\langle \delta_1, \delta_2 \rangle_\rho := \int \rho \langle v_1, v_2 \rangle dx$$

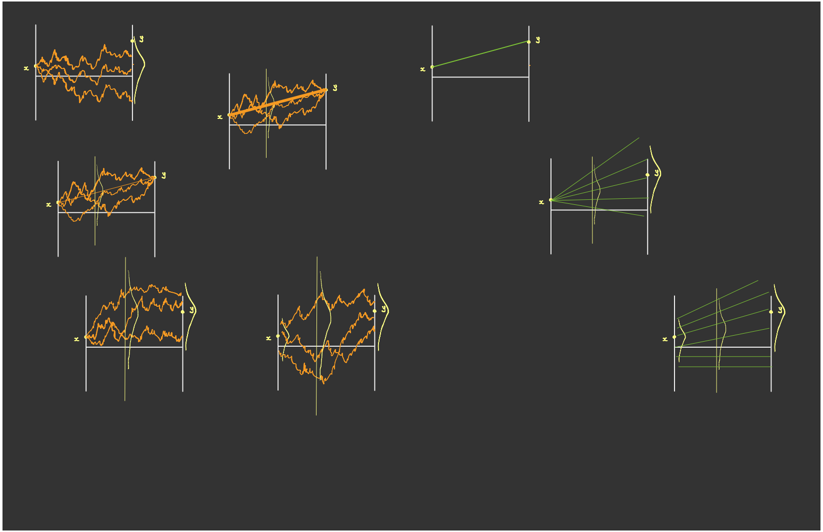
geodesic distance

$$W_2(\rho_0, \rho_1) = \inf_\rho \int_0^1 \sqrt{\left\langle \frac{\partial \rho}{\partial t}, \frac{\partial \rho}{\partial t} \right\rangle_{\rho(t)}} dt$$

Schrödinger Bridges vs. OMT Bridges

Bridges vs. Transport

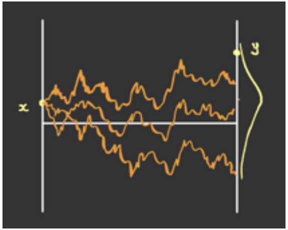
bird's eye view: stochastic bridges vs. Monge-Kantorovich transport (min distance^2)



Stochastic bridges

probability laws on paths linking marginals

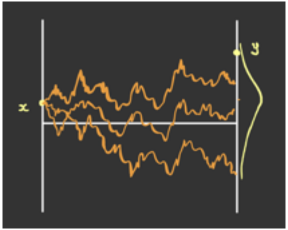
Brownian diffusion - prior law



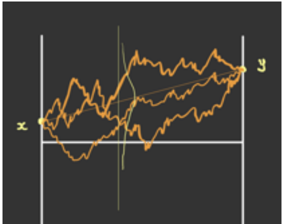
Stochastic bridges

probability laws on paths linking marginals

Brownian diffusion - prior law



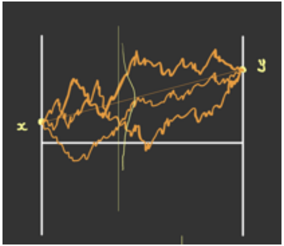
Brownian bridge - conditioned at both end-points (pinned bridge)



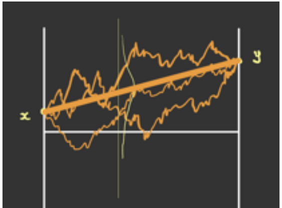
Stochastic bridges

probability laws on paths linking marginals

Brownian bridge - conditioned at both end-points (pinned bridge)



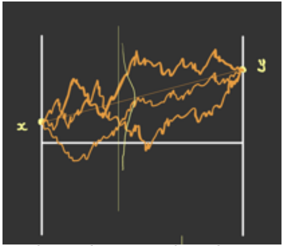
"most-likely" path (most prob. mass in neighborhood)



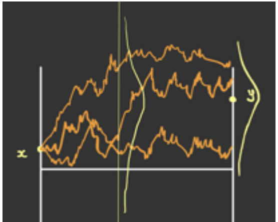
Stochastic bridges

probability laws on paths linking marginals

Brownian bridge - conditioned at both end-points (pinned bridge)



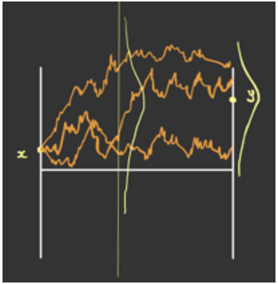
Schrödinger bridge - soft conditioning on one end



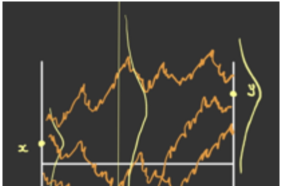
Stochastic bridges

probability laws on paths linking marginals

Schrödinger bridge - soft conditioning on one end

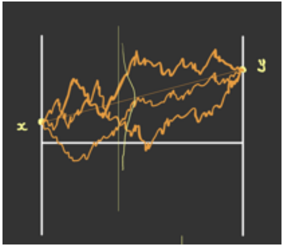


Schrödinger bridge - soft conditioning on both ends

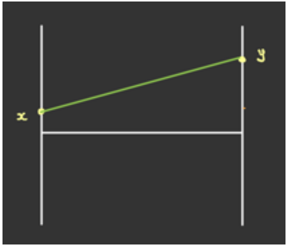


Stochastic bridges vs. optimal transport (deterministic)

Brownian bridge - Conditioned at end-points (Dirac marginals)

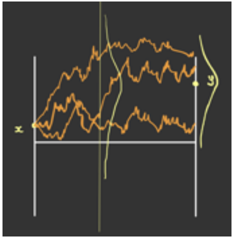


Optimal transport - Conditioned at end-points (Dirac marginals)

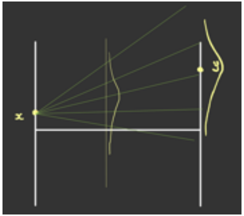


Stochastic bridges vs. optimal transport (deterministic)

Schrödinger bridge - soft conditioning at one end-point

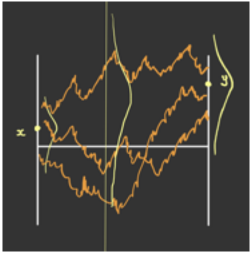


Optimal transport - soft conditioned at one end-point

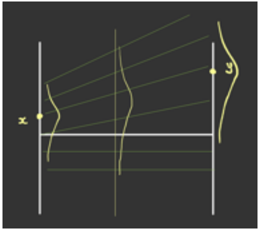


Stochastic bridges vs. optimal transport (deterministic)

Schrödinger bridge - soft conditioning at two ends



Optimal transport - soft conditioned at two ends



Some theory on Schrödinger bridges[2][3]

[2] Léonard, C., 2013. A survey of the schrödinger problem and some of its connections with optimal transport. arXiv preprint arXiv:1308.0215

[3] Chen, Yongxin, Tryphon T. Georgiou, and Michele Pavon. "On the relation between optimal transport and Schrödinger bridges: A stochastic control viewpoint." Journal of Optimization Theory and Applications 169 (2016): 671-691

Schrödinger bridges - first approach

$$\mathbb{P}^\star = \operatorname{argmin}\left\{\int_{\mathrm{paths}} \log\left(\frac{d\mathbb{P}}{d\mathbb{W}}\right) d\mathbb{P} \;\middle|\; \mathbb{P}|_{t=0} = \rho_0,\ \mathbb{P}_{t=1} = \rho_1\right\}$$

i) Disintegration of measures

$$\mathbb{P}(\mathrm{path}) = \underbrace{\mathbb{P}(\mathrm{path}\ |x(0) = x, x(t_f) = y)}_{\mathrm{conditioned}\ =\ \mathrm{pined\ bridge}} \cdot \mathbb{P}_{0,t_f}(x, y)$$

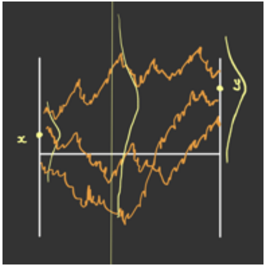
$\Rightarrow$

$$\int \log\left(\frac{d\mathbb{P}}{d\mathbb{W}}\right) d\mathbb{P} = \int \log\left(\frac{d\mathbb{P}_{0,t_f}(x,y)}{d\mathbb{W}_{0,t_f}(x,y)}\right) d\mathbb{P}_{0,t_f}(x,y) + \underbrace{\int \log\left(\frac{d\mathbb{P}(\mathrm{path}\ |x(0), x(t_f))}{d\mathbb{W}(\mathrm{path}\ |x(0), x(t_f))}\right) d\mathbb{P}(\mathrm{path}\ |x(0), x(t_f))}_{=0 \text{ for } \mathbb{P}(\mathrm{path}\ |x(0),x(t_f))\ =\ \mathbb{W}(\mathrm{path}\ |x(0),x(t_f))}$$

Structure of the law

via disintegration of measure

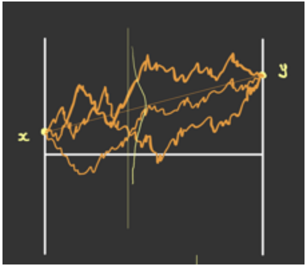
$\mathbb{P}^*(\text{path})$



Schrödinger bridge

$$= \quad \mathbb{P}^*_{0,t_f}(x,y) \times$$

$\mathbb{P}(\text{path}|x,y)$



Pinned bridges

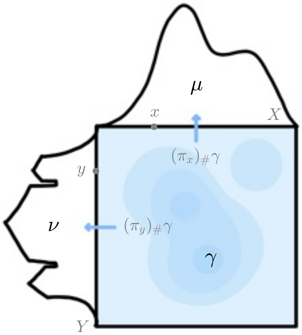
$\mathbb{P}^*_{0,t_f}(x,y)$: optimal end-point coupling

Optimal coupling of two end points

$$\min_{\mathbb{P}_{0,t_f}(x,y)} \int \log\left(\frac{d\mathbb{P}_{0,t_f}(x,y)}{d\mathbb{W}_{0,t_f}(x,y)}\right) d\mathbb{P}_{0,t_f}(x,y)$$

$\mathbb{P}_{0,t_f}(x,y)$: "couplings"

$\int_y \mathbb{P}_{0,t_f}(x,y) = \rho_0(x)dx = d\mu(x)$
$\int_x \mathbb{P}_{0,t_f}(x,y) = \rho_1(y)dy = d\nu(y)$



$\mathbb{P}^*_{0,t_f}(x,y) = \mathbb{W}_{0,t_f}(x,y)a(x)b(y)$

where $a(x) = e^{\lambda^{\mathrm{left}}(x)}$, $b(y) = e^{\lambda^{\mathrm{right}}(y)}$ with λ's Lagrange multipliers

Schrödinger system

Schrödinger (1931/32)

the density factors into

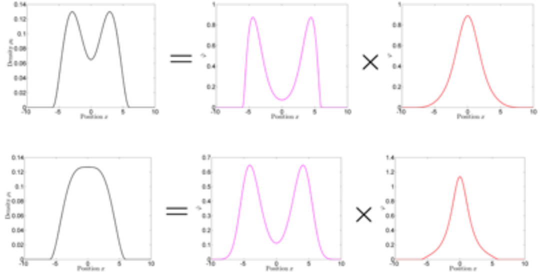
$$\rho(x,t) = \varphi(x,t)\hat{\varphi}(x,t)$$

where φ and $\hat{\varphi}$ solve (Schrödinger's system):

$$\varphi(x,t) = \int p(t,x,1,y)\varphi(y,1)dy, \quad \varphi(x,0)\hat{\varphi}(x,0) = \rho_0(x)$$

$$\hat{\varphi}(x,t) = \int p(0,y,t,x)\hat{\varphi}(y,0)dy, \quad \varphi(x,1)\hat{\varphi}(x,1) = \rho_1(x).$$

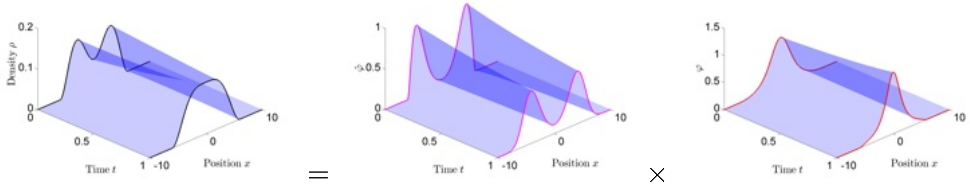
Schrödinger system



$$-\frac{\partial \varphi}{\partial t}(t,x) = \frac{1}{2}\Delta \varphi(t,x)$$

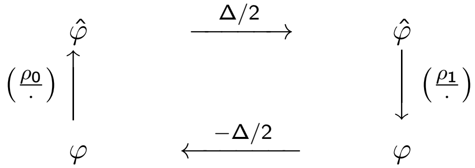
$$\frac{\partial \hat{\varphi}}{\partial t}(t,x) = \frac{1}{2}\Delta \hat{\varphi}(t,x)$$

$$\varphi(0,x)\hat{\varphi}(0,x) = \rho_0(x)$$
$$\varphi(1,x)\hat{\varphi}(1,x) = \rho_1(x)$$



Schrödinger system

Sinkhorn algorithm redux[4]



$$\frac{\partial \varphi}{\partial t}(t,x) = -\frac{1}{2}\Delta\varphi(t,x)$$
$$\frac{\partial \hat{\varphi}}{\partial t}(t,x) = \frac{1}{2}\Delta\hat{\varphi}(t,x)$$

$$\varphi(0,x)\hat{\varphi}(0,x) = \rho_0(x)$$
$$\varphi(1,x)\hat{\varphi}(1,x) = \rho_1(x)$$

$\Rightarrow$ strictly contractive with respect to d_H.

Hilbert metric

$$d_H(p,q) := \log \frac{\bar{\lambda}(p,q)}{\underline{\lambda}(p,q)}$$

$$\begin{aligned} \bar{\lambda}(p,q) &:= \inf\{\lambda \mid p \le \lambda q\} \\ \underline{\lambda}(p,q) &:= \sup\{\lambda \mid \lambda q \le p\} \end{aligned}$$

[4] Chen, Yongxin, Tryphon Georgiou, and Michele Pavon. "Entropic and displacement interpolation: a computational approach using the Hilbert metric." SIAM Journal on Applied Mathematics 76.6 (2016): 2375-2396.

Schrödinger bridges - second approach

$\mathbb{P}^\star = \text{argmin} \left\{ \int_{\text{paths}} \log \left(\frac{d\mathbb{P}}{d\mathbb{W}} \right) d\mathbb{P} \ \middle| \ \mathbb{P}|_{t=0} = \rho_0, \ \mathbb{P}_{t=1} = \rho_1 \right\}$

ii) Girsanov-Cameron-Martin theorem

The law $\mathbb{P}$ of

$$dX_t = v(t, X_t)dt + dB_t$$

and the law of B_t, $\mathbb{W}$, are such that

$$\int_{\text{paths}} \log \left(\frac{d\mathbb{P}}{d\mathbb{W}} \right) d\mathbb{P} = \frac{1}{2} \int \| v(t, X_t) \|^2 d\mathbb{P}$$

$\Rightarrow$ minimum kinetic energy paths matching marginals

Schrödinger bridges - second approach

$\mathbb{P}^\star = \operatorname{argmin} \left\{ \int_{\text{paths}} \log\left(\frac{d\mathbb{P}}{d\mathbb{W}}\right) d\mathbb{P} \;\middle|\; \mathbb{P}|_{t=0} = \rho_0,\ \mathbb{P}_{t=1} = \rho_1 \right\}$

Stochastic control formulation

$$\inf_{(\rho,v)} \int_{\mathbb{R}^n} \int_{t_0}^{t_f} \|v(x,t)\|^2 \rho(x,t) dt dx,$$

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (v\rho) = \frac{1}{2}\Delta\rho$$

$$\rho(x,t_0) = \rho_0(x), \quad \rho(y,t_f) = \rho_1(y).$$

Shift the probability on paths of $dX_t = dB_t$, from ρ_0 to ρ_1, so that it is "concentrated" on paths that correspond to minimum effort of a controlled diffusion $dX_t = vdt + dB_t$.

Schrödinger bridges - second approach

Fisher-information regularization - time-symmetric/fluid dynamic

$$\inf_{(\rho,u)} \int_{\mathbb{R}^n} \int_{t_0}^{t_f} \left(\|u(x,t)\|^2 + \frac{1}{4} \|\nabla \log \rho(x,t)\|^2 \right) \rho(x,t) dt dx,$$

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (u\rho) = 0$$

$$\rho(x, t_0) = \rho_0(x), \quad \rho(y, t_f) = \rho_1(y).$$

$u = v - \frac{1}{2} \nabla \log \rho$

Chen, Georgiou, Pavon, 2016, On the relation between optimal transport and Schrödinger bridges: A stochastic control viewpoint. J. of Opt. Theory and Appl., 169:671-91

Li, Yin, Osher, 2018. Computations of optimal transport distance with Fisher information regularization. J. of Scientific Comp., 75:1581-95

Implications - Geometry & Physics

JKO (Jordan-Kinderlehrer-Otto)

gradient flow of entropy

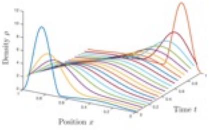
$$\partial_t \rho = -\nabla^{W_2} S(\rho) \ = \ \Delta\rho$$

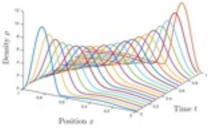
OMT quantifies dissipation in over-damped systems

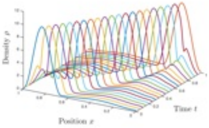
- gradient flows
- convergence to equilibrium
- thermalization
- thermodynamics

OMT as 0-noise limit to SBP & numerics

$\rho_t + \nabla \cdot \rho v = \epsilon \Delta \rho$







$$\inf_{(\rho,v)} \int_{\mathbb{R}^n} \int_{t_0}^{t_f} \|v(x,t)\|^2 \rho(x,t)\,dt\,dx,$$
$$\frac{\partial \rho}{\partial t} + \nabla \cdot (v\rho) = \frac{\epsilon}{2}\Delta\rho$$
$$\rho(x,t_0) = \rho_0(x), \quad \rho(y,t_f) = \rho_1(y).$$

or

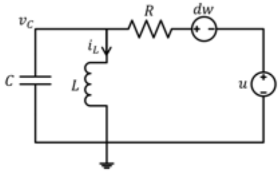
$$\inf_{(\rho,v)} \int_{\mathbb{R}^n} \int_{t_0}^{t_f} \left[\|v(x,t)\|^2 + \|\frac{\epsilon}{2}\nabla \log \rho(x,t)\|^2\right] \rho(x,t)\,dt\,dx,$$
$$\frac{\partial \rho}{\partial t} + \nabla \cdot (v\rho) = 0,$$
$$\rho(x,t_0) = \rho_0(x), \quad \rho(y,t_f) = \rho_1(y).$$

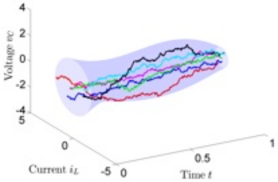
Control applications: active cooling

– thermodynamic systems, controlling collective response
– magnetization distribution in NMR spectroscopy,..

– Nyquist-Johnson noise driven oscillator

$$
\begin{aligned}
L di_L(t) &= v_C(t)dt \\
RC dv_C(t) &= -v_C(t)dt - Ri_L(t)dt + u(t)dt + dw(t)
\end{aligned}
$$



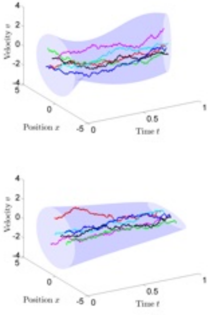


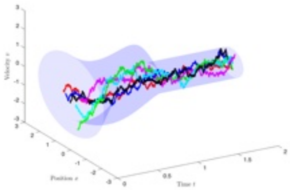
Chen-Georgiou-Pavon, *J. Math. Phys.* 2015.

controlling uncertainty, ensemble control

Inertial particles with stochastic excitation steered between marginals

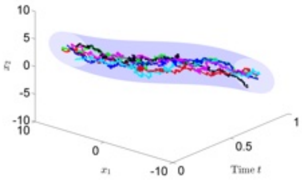
$$
\begin{aligned}
dx(t) &= v(t)dt \\
dv(t) &= -u(t)dt + dw(t)
\end{aligned}
$$



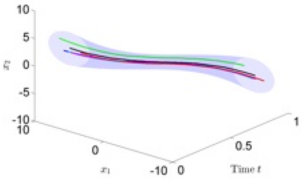


trajectories in phase space
transparent tube: “3σ region”

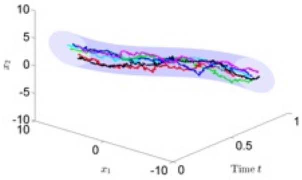
Over prior dynamics



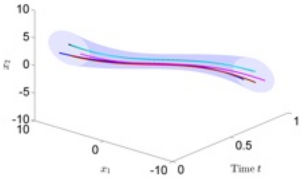
Schrödinger bridge with $\epsilon = 9$



Schrödinger bridge with $\epsilon = 0.01$



Schrödinger bridge with $\epsilon = 4$



Optimal transport with prior

Smooth Bridges/Splines - minimize acceleration

– Mass transports along x in C^2 with $\int \|\dot{x}\|_2 dt < \infty$

Distributional-Spline-Problem:

Find

$$\inf_{x_{t_i}\sharp P=\rho_i} \mathbb{E}_{\mathbb{Q}}\{\int_0^1 \|\ddot{x}(t)\|^2 dt\}$$

with $\mathbb{Q}$ a **probability measure** on path space.

when $\rho_i \sim \mathcal{N}(m_i, \sigma_i) \Rightarrow$ **Semidefinite program**

recap

Schrödinger bridges vs. OMT flows

- minimize cost of transporting between end-point marginals
- minimize cost of traversing paths from beginning to end

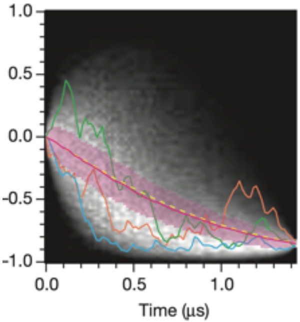
SB efficient iterative computation

OMT provides a Riemannian geometry
$W_2(\cdot, \cdot)$ is a geodesic distance
gradient flow of the entropy & the heat equation (JKO)

topics to keep an eye:

- holonomy of transport (internal DoF) - Abdelgalil & G (TAC Nov25, Jan26)
- transport of tracers (Eldesoukey-Abdelgalil-G CDC25)
- transport spatio-temporal constraints (Dong, Eldesoukey)
- minimal attention (Sabbagh etal arXiv just now)

outlook in the quantum



quantum trajectories

between pre- and post-selected states

Weber, Chantasri, Dressel, Jordan, Murch, Siddiqi
Mapping the optimal route between two quantum states
Nature, 2014, doi:10.1038/nature13559

Quantum Schrödinger Bridges (QSB)

"The (unauthorized) extrapolation ... into the quantum mechanical domain,"

Otto Bergmann, 1988

Interpolation:

Consider initial/final states $|i\rangle$, $|f\rangle$ of an observable A (with discrete spectrum)

$$\exp\left(\frac{1}{2}\pi\frac{t-t_0}{t_1-t_0}\cdot\underbrace{(|f\rangle\langle i| + |i\rangle\langle f|)}_{S}\right)|i\rangle = \cos(\frac{\pi}{2}\tau)|i\rangle + \sin(\frac{\pi}{2}\tau)|f\rangle$$

$\tau = \frac{t-t_0}{t_1-t_0}$

For $S = |f\rangle\langle i| + |i\rangle\langle f|$, then $\exp(\alpha S) = \cos(\alpha)I + \sin(\alpha)S$

Bridge interpolates $|i\rangle\langle i|$ and $|f\rangle\langle f|$

QSB - Bergmann

Restores resemblance with the Brownian bridge, i.e.,

$$\frac{g(t - t_0, x - x_0)g(t_1 - t, x_1 - x)}{Z}$$

Measuring $|k\rangle$ at t, for evolution $U(t) = \exp(-\frac{i}{\hbar}Ht)$,

$$\frac{\sum_k |k\rangle\langle k| \cdot \overbrace{\left(|\langle f|U(t_1 - t)|k\rangle|^2 \cdot |\langle k|U(t - t_0)|i\rangle|^2\right)}^{\text{probability}}}{Z_t}$$

- $|i\rangle$ evolves to a mixed state and back to $|f\rangle$, a decrease in entropy at some point
- the normalization depends on the time t when a projective measurement takes place normalize by $|\langle f|U(t_1 - t_0)|i\rangle|^2$ to restore analogy classical at end-points
- discusses non-selective measurements at t, writes the law in product form

QSB - Bergmann

Bergmann:

• "Schrödinger's main interests... statistical mechanics and the interpretation of quantum mechanics"

• "inspired by the old and almost unknown paper by Schrödinger and no attempt was made to draw any conclusions about its impact, if any, on the theory of measurements. It was written primarily as a historical study"

• "a referee .. informed the author [Bergmann] of"
– Y. Aharonov, P.G. Bergmann and J.L. Lebowitz
– F.J. Belinfante
"contributions to the same problem .. written without Schrödinger's inspiration"

Two-state vector formalism[5]

- pre- and post- selected quantum systems
- a time-symmetric description of QM in which the present is caused by states

$$\langle \psi_1 | \cdot | \psi_0 \rangle,$$

  evolving backwards from the future ($\langle \psi_1 |$) and forward from the past ($| \psi_0 \rangle$)
- time-symmetry by construction

Figure: Watanabe and his son, 1949

[5]Watanabe, S. (1955). Symmetry of physical laws. Part III. Prediction and retrodiction. Reviews of Modern Physics, 27(2), 179.

Two-state vector formalism[6]

- "indubitable asymmetry in time direction" is not due to the principles of QM but from the intrusion of the macroscopic world

Y. Aharonov

P. Bergmann

(1915–2002)

J. Lebowitz

[6] Aharonov, Y., Bergmann, P. G., and Lebowitz, J. L. (1964). Time symmetry in the quantum process of measurement. Physical Review, 134(6B), B1410.

Measurements and time-reversal[7]

- prediction vs postdiction: apparent asymmetry
- Belifante extends to non-ideal measurement processes

Frederik Jozel Belinfante (1913–1991)

[7]Belinfante, F.J. (1975). Measurements and Time Reversal in Objective Quantum Theory: International Series in Natural Philosophy (Vol. 75)

pre/post-selection - link to large deviations[8]

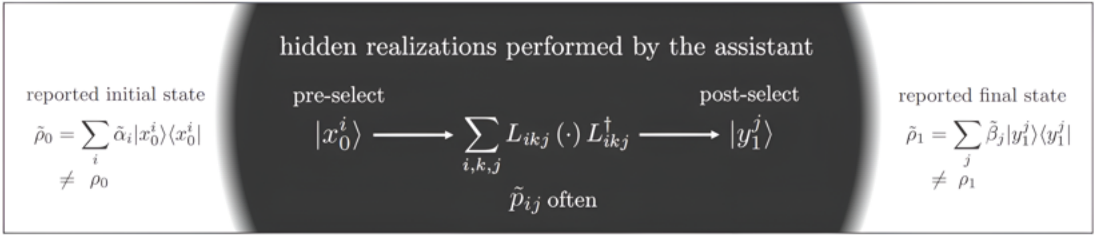
- fix two bases $|\psi_0^i\rangle$ and $|\psi_1^j\rangle$
- assistant prepares ensembles and reports pre/post-selected marginals

Olga Movilla Miangolarra
Universidad de La Laguna, Spain

Bridge problem: What is the most likely coupling of initial and final outcomes?

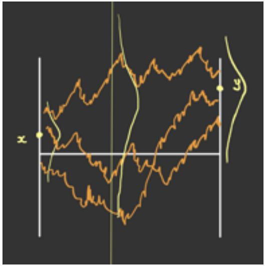
[8] Quantum Schrödinger bridges: large deviations and time-symmetric ensembles, Olga et al., PRR

pre/post-selection

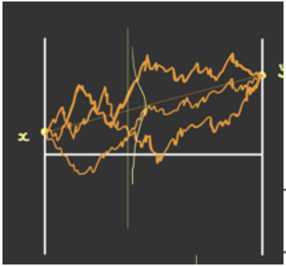


intervening measurements

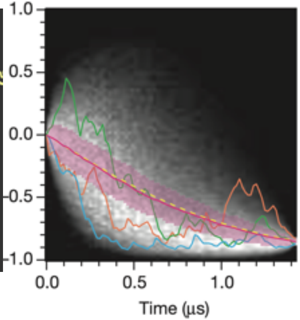
Statistics, path probabilities



Schrödinger bridge



Pinned bridges



quantum trajectories of Weber etal.

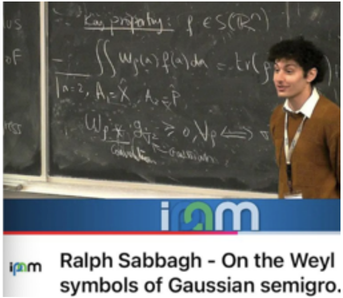
recap & discussion

Schrödinger Bridges: interpolation of probability laws

uncertainty control, stochastic control

diffusion models in ML

quantum SBs & the time arrow in physics



thank you for your attention