



El Problema del Puente de Schrödinger como Generalización de los Modelos de Difusión

Fernando Fêtis Riquelme

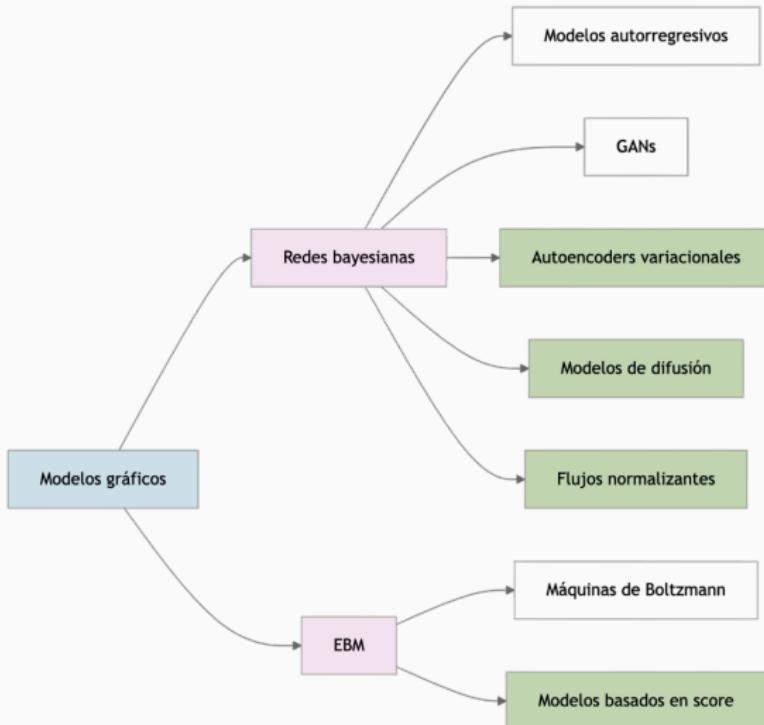
Otoño, 2025

Facultad de Ciencias Físicas y Matemáticas
Universidad de Chile

Tabla de contenidos

1. Modelos de difusión
2. Problema de Schrödinger
3. Transporte óptimo

Taxonomía de los modelos gráficos



Modelos de difusión

Modelos de difusión | Diagrama general

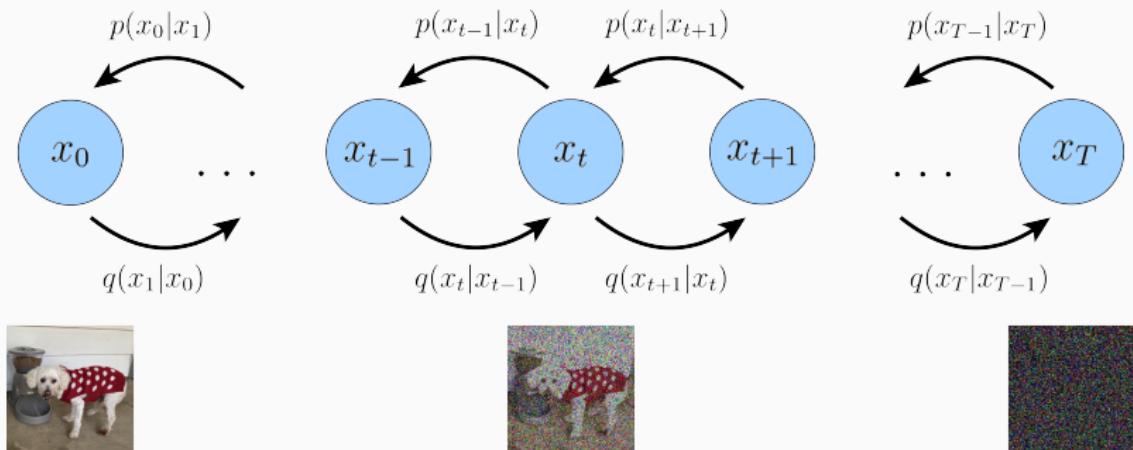


Figura 1: Imagen obtenida desde Luo, 2022.

Modelos de difusión | Proceso forward

El proceso de inyección de ruido es una cadena de Markov en \mathbb{R}^d factorizada de forma causal, cuyos hiperparámetros corresponden a una secuencia finita y decreciente $(\alpha_t)_{t=1}^T \subset (0, 1)$:

Proceso forward

$$q(x_{0:T}) = q(x_0) \prod_{t=1}^T q(x_t|x_{t-1}),$$

con $q(x_0) = p_{\text{data}}(x_0)$ y transiciones gaussianas isotrópicas:

$$q(x_t|x_{t-1}) \sim \mathcal{N}(\sqrt{\alpha_t}x_{t-1}, (1 - \alpha_t)\mathbf{I}_d).$$

La secuencia $(\alpha_t)_{t=1}^T$ debe ser tal que

$$q(x_T) = \int_{(\mathbb{R}^d)^T} q(x_{0:T}) \, dx_{0:(T-1)} \approx p_{\text{prior}}(x_T) \sim \mathcal{N}(0, \mathbf{I}_d).$$

Modelos de difusión | Proceso forward

El proceso de inyección de ruido es una cadena de Markov en \mathbb{R}^d factorizada de forma causal, cuyos hiperparámetros corresponden a una secuencia finita y decreciente $(\alpha_t)_{t=1}^T \subset (0, 1)$:

Proceso forward

$$q(x_{0:T}) = q(x_0) \prod_{t=1}^T q(x_t|x_{t-1}),$$

con $q(x_0) = p_{\text{data}}(x_0)$ y transiciones gaussianas isotrópicas:

$$q(x_t|x_{t-1}) \sim \mathcal{N}(\sqrt{\alpha_t}x_{t-1}, (1 - \alpha_t)\mathbf{I}_d).$$

La secuencia $(\alpha_t)_{t=1}^T$ debe ser tal que

$$q(x_T) = \int_{(\mathbb{R}^d)^T} q(x_{0:T}) \, dx_{0:(T-1)} \approx p_{\text{prior}}(x_T) \sim \mathcal{N}(0, \mathbf{I}_d).$$

Modelos de difusión | Proceso backward

El proceso de reconstrucción es otra cadena de Markov factorizada de forma anticausal:

Proceso backward

$$p_{\theta}(x_{0:T}) = p(x_T) \prod_{t=1}^T p_{\theta}(x_{t-1}|x_t),$$

con $p(x_T) = p_{\text{prior}}(x_T)$ y transiciones gaussianas:

$$p_{\theta}(x_{t-1}|x_t) \sim \mathcal{N}(\mu_{\theta}(x_t, t), \Sigma_{\theta}(x_t, t)).$$

- La función de costo buscará que $p_{\theta}(x_{t-1}|x_t) \approx q(x_{t-1}|x_t, x_0)$.
- Dado que $q(x_{t-1}|x_t, x_0) \sim \mathcal{N}(\mu_q(x_0, x_t, t), \sigma_q^2(t) I_d)$, se puede fijar $\Sigma_{\theta}(x_t, t) = \sigma_q^2(t) I_d$.
- Con esto, solo se necesitará aprender el vector de medias de $p_{\theta}(x_{t-1}|x_t)$ mediante una red neuronal $\mu_{\theta} : \mathbb{R}^d \times \{0, \dots, T\} \rightarrow \mathbb{R}^d$.

Modelos de difusión | Proceso backward

El proceso de reconstrucción es otra cadena de Markov factorizada de forma anticausal:

Proceso backward

$$p_{\theta}(x_{0:T}) = p(x_T) \prod_{t=1}^T p_{\theta}(x_{t-1}|x_t),$$

con $p(x_T) = p_{\text{prior}}(x_T)$ y transiciones gaussianas:

$$p_{\theta}(x_{t-1}|x_t) \sim \mathcal{N}(\mu_{\theta}(x_t, t), \Sigma_{\theta}(x_t, t)).$$

- La función de costo buscará que $p_{\theta}(x_{t-1}|x_t) \approx q(x_{t-1}|x_t, x_0)$.
- Dado que $q(x_{t-1}|x_t, x_0) \sim \mathcal{N}(\mu_q(x_0, x_t, t), \sigma_q^2(t) I_d)$, se puede fijar $\Sigma_{\theta}(x_t, t) = \sigma_q^2(t) I_d$.
- Con esto, solo se necesitará aprender el vector de medias de $p_{\theta}(x_{t-1}|x_t)$ mediante una red neuronal $\mu_{\theta} : \mathbb{R}^d \times \{0, \dots, T\} \rightarrow \mathbb{R}^d$.

Modelos de difusión | Proceso backward

El proceso de reconstrucción es otra cadena de Markov factorizada de forma anticausal:

Proceso backward

$$p_{\theta}(x_{0:T}) = p(x_T) \prod_{t=1}^T p_{\theta}(x_{t-1}|x_t),$$

con $p(x_T) = p_{\text{prior}}(x_T)$ y transiciones gaussianas:

$$p_{\theta}(x_{t-1}|x_t) \sim \mathcal{N}(\mu_{\theta}(x_t, t), \Sigma_{\theta}(x_t, t)).$$

- La función de costo buscará que $p_{\theta}(x_{t-1}|x_t) \approx q(x_{t-1}|x_t, x_0)$.
- Dado que $q(x_{t-1}|x_t, x_0) \sim \mathcal{N}(\mu_q(x_0, x_t, t), \sigma_q^2(t) I_d)$, se puede fijar $\Sigma_{\theta}(x_t, t) = \sigma_q^2(t) I_d$.
- Con esto, solo se necesitará aprender el vector de medias de $p_{\theta}(x_{t-1}|x_t)$ mediante una red neuronal $\mu_{\theta} : \mathbb{R}^d \times \{0, \dots, T\} \rightarrow \mathbb{R}^d$.

Modelos de difusión | Proceso backward

El proceso de reconstrucción es otra cadena de Markov factorizada de forma anticausal:

Proceso backward

$$p_{\theta}(x_{0:T}) = p(x_T) \prod_{t=1}^T p_{\theta}(x_{t-1}|x_t),$$

con $p(x_T) = p_{\text{prior}}(x_T)$ y transiciones gaussianas:

$$p_{\theta}(x_{t-1}|x_t) \sim \mathcal{N}(\mu_{\theta}(x_t, t), \Sigma_{\theta}(x_t, t)).$$

- La función de costo buscará que $p_{\theta}(x_{t-1}|x_t) \approx q(x_{t-1}|x_t, x_0)$.
- Dado que $q(x_{t-1}|x_t, x_0) \sim \mathcal{N}(\mu_q(x_0, x_t, t), \sigma_q^2(t) I_d)$, se puede fijar $\Sigma_{\theta}(x_t, t) = \sigma_q^2(t) I_d$.
- Con esto, solo se necesitará aprender el vector de medias de $p_{\theta}(x_{t-1}|x_t)$ mediante una red neuronal $\mu_{\theta} : \mathbb{R}^d \times \{0, \dots, T\} \rightarrow \mathbb{R}^d$.

Modelos de difusión | Entrenamiento e inferencia

Para entrenar μ_θ , la verosimilitud $p_\theta(x_0) = \int_{(\mathbb{R}^d)^T} p_\theta(x_{0:T}) \, dx_{1:T}$ no es tratable. En cambio, se maximiza $\mathbb{E}_{x_0 \sim p_{\text{data}}(x_0)} [\text{ELBO}(x_0)]$, donde

$$\text{ELBO}(x_0) := \log p_\theta(x_0) - D_{\text{KL}}(q(x_{1:T}|x_0) \parallel p_\theta(x_{1:T}|x_0)).$$

La ELBO se puede evaluar eficientemente:

ELBO para DDPM

Dada una muestra $x_0 \sim p_{\text{data}}(x_0)$, entonces:

$$\text{ELBO}(x_0) = - \sum_{t=1}^T \frac{1}{2\sigma_q^2(t)} \|\mu_q(x_0, x_t, t) - \mu_\theta(x_t, t)\|^2 + \text{constante}.$$

Para la generación de nuevas muestras desde $p_\theta(x_0)$, se simula el proceso backward comenzando con una muestra $x_T \sim p_{\text{prior}}(x_T)$.

Modelos de difusión | Entrenamiento e inferencia

Para entrenar μ_θ , la verosimilitud $p_\theta(x_0) = \int_{(\mathbb{R}^d)^T} p_\theta(x_{0:T}) \, dx_{1:T}$ no es tratable. En cambio, se maximiza $\mathbb{E}_{x_0 \sim p_{\text{data}}(x_0)} [\text{ELBO}(x_0)]$, donde

$$\text{ELBO}(x_0) := \log p_\theta(x_0) - D_{\text{KL}}(q(x_{1:T}|x_0) \parallel p_\theta(x_{1:T}|x_0)).$$

La ELBO se puede evaluar eficientemente:

ELBO para DDPM

Dada una muestra $x_0 \sim p_{\text{data}}(x_0)$, entonces:

$$\text{ELBO}(x_0) = - \sum_{t=1}^T \frac{1}{2\sigma_q^2(t)} \|\mu_q(x_0, x_t, t) - \mu_\theta(x_t, t)\|^2 + \text{constante}.$$

Para la generación de nuevas muestras desde $p_\theta(x_0)$, se simula el proceso backward comenzando con una muestra $x_T \sim p_{\text{prior}}(x_T)$.

Modelos de difusión | Entrenamiento e inferencia

Para entrenar μ_θ , la verosimilitud $p_\theta(x_0) = \int_{(\mathbb{R}^d)^T} p_\theta(x_{0:T}) \, dx_{1:T}$ no es tratable. En cambio, se maximiza $\mathbb{E}_{x_0 \sim p_{\text{data}}(x_0)} [\text{ELBO}(x_0)]$, donde

$$\text{ELBO}(x_0) := \log p_\theta(x_0) - D_{\text{KL}}(q(x_{1:T}|x_0) \parallel p_\theta(x_{1:T}|x_0)).$$

La ELBO se puede evaluar eficientemente:

ELBO para DDPM

Dada una muestra $x_0 \sim p_{\text{data}}(x_0)$, entonces:

$$\text{ELBO}(x_0) = - \sum_{t=1}^T \frac{1}{2\sigma_q^2(t)} \|\mu_q(x_0, x_t, t) - \mu_\theta(x_t, t)\|^2 + \text{constante}.$$

Para la generación de nuevas muestras desde $p_\theta(x_0)$, se simula el proceso backward comenzando con una muestra $x_T \sim p_{\text{prior}}(x_T)$.

Modelos de difusión | Formulación basada en score

Considerando que

$$\mu_q(x_0, x_t, t) = \frac{1}{\sqrt{\alpha_t}} x_t - \frac{1 - \alpha_t}{\sqrt{\alpha_t}} \nabla_{x_t} \log q(x_t),$$

la red neuronal $\mu_\theta(x_t, t)$ puede ser reparametrizada por otra red neuronal $s_\theta(x_t, t)$ que aprenda directamente $\nabla_{x_t} \log q(x_t)$.

- Esto conecta los modelos de difusión con SM y con EBM.
- Entrega un método de generación condicional (guidance):

$$\underbrace{\nabla_{x_t} \log p_\theta(x_t|y)}_{\text{score condicional}} = \underbrace{\nabla_{x_t} \log p_\theta(y|x_t)}_{\text{modelo discriminativo}} - \underbrace{\nabla_{x_t} \log p_\theta(x_t)}_{\text{score incondicional}},$$

con $p_\theta(y|x_t)$ un clasificador o un modelo tipo CLIP.

Modelos de difusión | Formulación basada en score

Considerando que

$$\mu_q(x_0, x_t, t) = \frac{1}{\sqrt{\alpha_t}} x_t - \frac{1 - \alpha_t}{\sqrt{\alpha_t}} \nabla_{x_t} \log q(x_t),$$

la red neuronal $\mu_\theta(x_t, t)$ puede ser reparametrizada por otra red neuronal $s_\theta(x_t, t)$ que aprenda directamente $\nabla_{x_t} \log q(x_t)$.

- Esto conecta los modelos de difusión con SM y con EBM.
- Entrega un método de generación condicional (guidance):

$$\underbrace{\nabla_{x_t} \log p_\theta(x_t|y)}_{\text{score condicional}} = \underbrace{\nabla_{x_t} \log p_\theta(y|x_t)}_{\text{modelo discriminativo}} - \underbrace{\nabla_{x_t} \log p_\theta(x_t)}_{\text{score incondicional}},$$

con $p_\theta(y|x_t)$ un clasificador o un modelo tipo CLIP.

Modelos de difusión | Formulación basada en score

Considerando que

$$\mu_q(x_0, x_t, t) = \frac{1}{\sqrt{\alpha_t}} x_t - \frac{1 - \alpha_t}{\sqrt{\alpha_t}} \nabla_{x_t} \log q(x_t),$$

la red neuronal $\mu_\theta(x_t, t)$ puede ser reparametrizada por otra red neuronal $s_\theta(x_t, t)$ que aprenda directamente $\nabla_{x_t} \log q(x_t)$.

- Esto conecta los modelos de difusión con SM y con EBM.
- Entrega un método de generación condicional (guidance):

$$\underbrace{\nabla_{x_t} \log p_\theta(x_t|y)}_{\text{score condicional}} = \underbrace{\nabla_{x_t} \log p_\theta(y|x_t)}_{\text{modelo discriminativo}} - \underbrace{\nabla_{x_t} \log p_\theta(x_t)}_{\text{score incondicional}},$$

con $p_\theta(y|x_t)$ un clasificador o un modelo tipo CLIP.

Modelos de difusión | Formulación continua

La formulación basada en score permite extender los modelos de difusión a tiempo continuo usando SDEs:

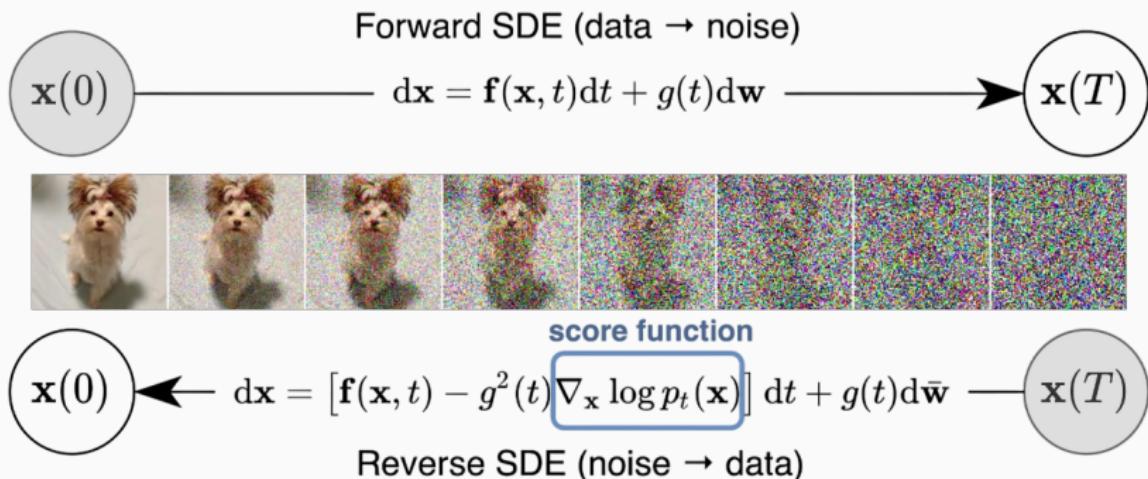


Figura 2: Imagen obtenida desde Song et al., 2021.

La función de costo se puede extender de forma *natural*, obteniendo una expresión análoga a la ELBO en tiempo discreto.

Modelos de difusión | Formulación continua

- DDPM y DSM son discretizaciones de SDEs específicas.

$$\begin{cases} dx_t = \frac{1}{2}(1 - \alpha(t))x_t dt + \sqrt{1 - \alpha(t)} dw_t & (\text{DDPM}) \\ dx_t = \sqrt{\frac{d}{dt}\sigma^2(t)} dw_t & (\text{DSM}) \end{cases}$$

- Se pueden usar diferentes solvers para el proceso backward durante la generación.
- Existe un proceso determinista con las mismas distribuciones marginales que los procesos de difusión y denoising.
- En particular, se puede calcular la verosimilitud de forma exacta.

Modelos de difusión | Formulación continua

- DDPM y DSM son discretizaciones de SDEs específicas.
- Se pueden usar diferentes solvers para el proceso backward durante la generación.
- Existe un proceso determinista con las mismas distribuciones marginales que los procesos de difusión y denoising.
- En particular, se puede calcular la verosimilitud de forma exacta.

Modelos de difusión | Formulación continua

- DDPM y DSM son discretizaciones de SDEs específicas.
- Se pueden usar diferentes solvers para el proceso backward durante la generación.
- Existe un proceso determinista con las mismas distribuciones marginales que los procesos de difusión y denoising.

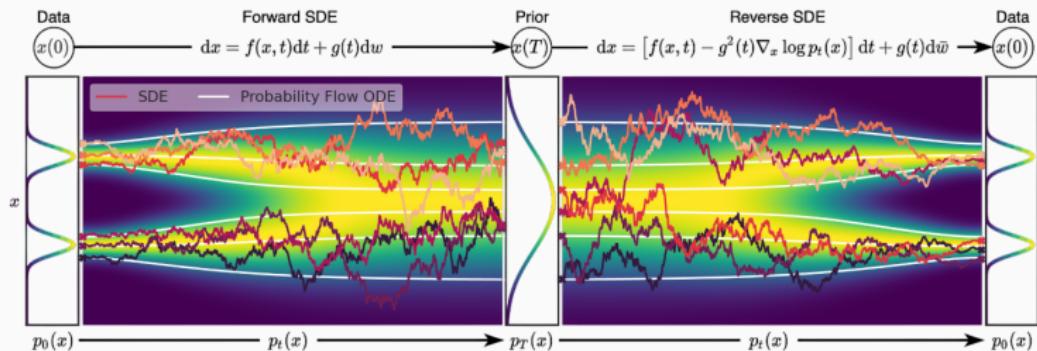


Figura 3: Imagen obtenida desde Song et al., 2021.

- En particular, se puede calcular la verosimilitud de forma exacta.

Modelos de difusión | Formulación continua

- DDPM y DSM son discretizaciones de SDEs específicas.
- Se pueden usar diferentes solvers para el proceso backward durante la generación.
- Existe un proceso determinista con las mismas distribuciones marginales que los procesos de difusión y denoising.
- En particular, se puede calcular la verosimilitud de forma exacta.

Modelos de difusión | Limitaciones

- Sensibilidad a la elección del proceso de difusión.
- No permite transformación entre distribuciones.
- Convergencia asintótica a p_{prior} y generación lenta.



Figura 3: Imágenes obtenidas desde Nichol y Dhariwal, 2021 y Chen, 2023.

Modelos de difusión | Limitaciones

- Sensibilidad a la elección del proceso de difusión.
- No permite transformación entre distribuciones.
- Convergencia asintótica a p_{prior} y generación lenta.



Figura 3: Imagen obtenida desde Zhu et al., 2020.

Modelos de difusión | Limitaciones

- Sensibilidad a la elección del proceso de difusión.
- No permite transformación entre distribuciones.
- Convergencia asintótica a p_{prior} y generación lenta.

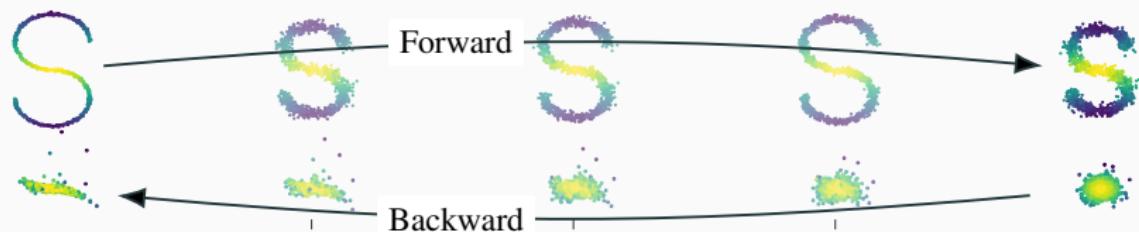


Figura 3: Imagen obtenida desde Bortoli et al., 2023.

Problema de Schrödinger

Problema de Schrödinger | Formulación dinámica

Considerando $\mathcal{X}, \mathcal{Y} \subset \mathbb{R}^d$, el siguiente problema generativo corrige las limitaciones mencionadas:

SBP (formulación dinámica)

El puente de Schrödinger entre dos medidas $\mu \in \mathcal{M}_+^1(\mathcal{X})$ y $\nu \in \mathcal{M}_+^1(\mathcal{Y})$ es el (único) proceso estocástico que resuelve el problema

$$\min_{P: (P_0 \sim \mu) \wedge (P_1 \sim \nu)} D_{\text{KL}}(P \| R),$$

donde $\begin{cases} R \text{ es la ley de } x_t = f(x_t, t) dt + g(t) dw_t \\ P \text{ es la ley de } x_t = [f(x_t, t) + v(x_t, t)] dt + g(t) dw_t \end{cases}$

Por simplicidad, se considerará siempre que R es el movimiento browiano con difusividad $\sigma > 0$ (i.e., $f \equiv 0$ y $g(t) = \sigma$).

Problema de Schrödinger | Formulación dinámica

Considerando $\mathcal{X}, \mathcal{Y} \subset \mathbb{R}^d$, el siguiente problema generativo corrige las limitaciones mencionadas:

SBP (formulación dinámica)

El puente de Schrödinger entre dos medidas $\mu \in \mathcal{M}_+^1(\mathcal{X})$ y $\nu \in \mathcal{M}_+^1(\mathcal{Y})$ es el (único) proceso estocástico que resuelve el problema

$$\min_{P: (P_0 \sim \mu) \wedge (P_1 \sim \nu)} D_{\text{KL}}(P \| R),$$

donde $\begin{cases} R \text{ es la ley de } x_t = f(x_t, t) dt + g(t) dw_t \\ P \text{ es la ley de } x_t = [f(x_t, t) + v(x_t, t)] dt + g(t) dw_t \end{cases}$

Por simplicidad, se considerará siempre que R es el movimiento browiano con difusividad $\sigma > 0$ (i.e., $f \equiv 0$ y $g(t) = \sigma$).

Problema de Schrödinger | Formulación dinámica

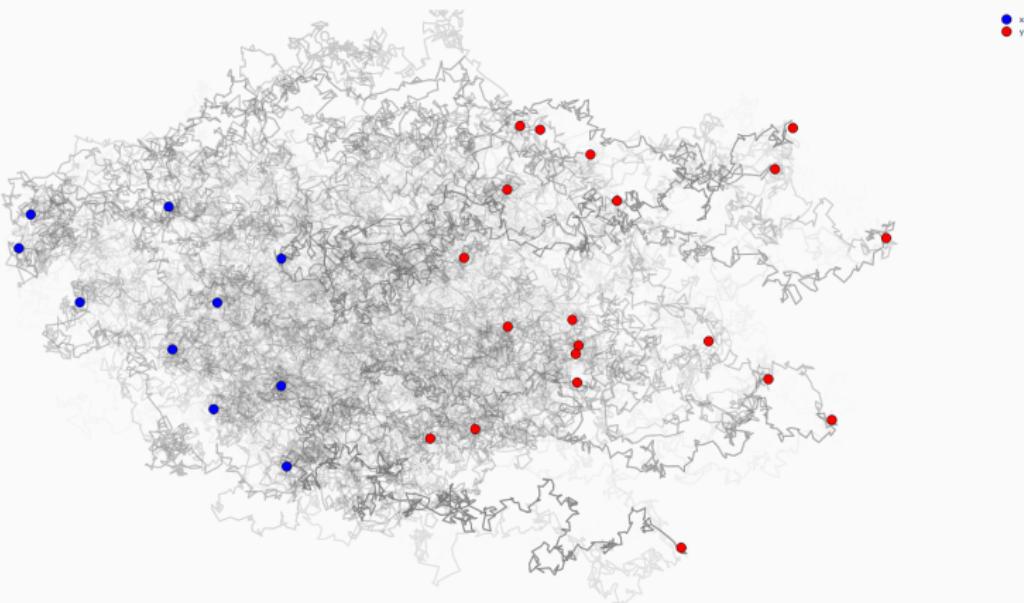


Figura 4: Puentes de Schrödinger para $\sigma = 1$.

Problema de Schrödinger | Formulación dinámica

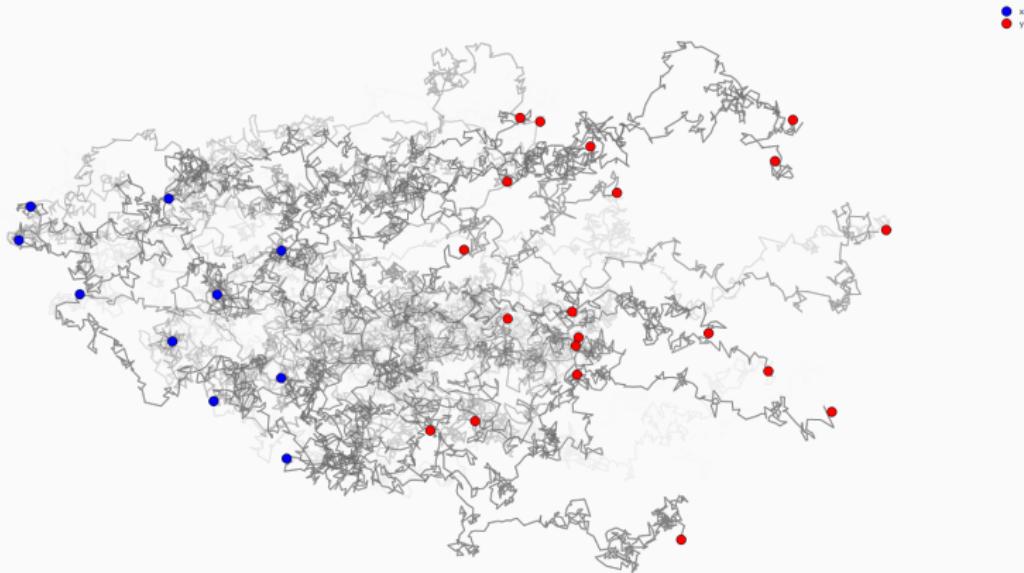


Figura 4: Puentes de Schrödinger para $\sigma = 0,1$.

Problema de Schrödinger | Formulación dinámica

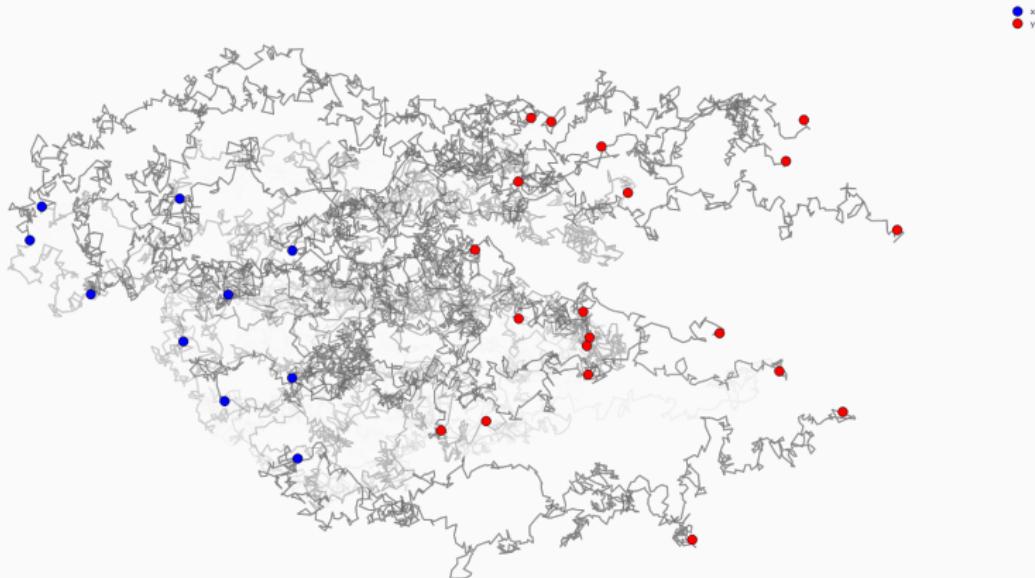


Figura 4: Puentes de Schrödinger para $\sigma = 0,05$.

Problema de Schrödinger | Formulación estática

Se puede demostrar la siguiente descomposición:

$$D_{KL}(P \parallel R) = D_{KL}(P_{01} \parallel R_{01}) + \mathbb{E}_{(x,y) \sim P_{01}} [D_{KL}(P_{|xy} \parallel R_{|xy})].$$

Por lo tanto, el SBP dinámico se puede reducir a un problema estático enfocado únicamente en los extremos del proceso:

$$\underbrace{\min_{P: (P_0 \sim \mu) \wedge (P_1 \sim \nu)} D_{KL}(P \parallel R)}_{\text{problema dinámico}} = \underbrace{\min_{P_{01} \in \Pi(\mu, \nu)} D_{KL}(P_{01} \parallel R_{01})}_{\text{problema estático}},$$

donde $\Pi(\mu, \nu) := \{\pi \in \mathcal{M}_+^1(\mathcal{X} \times \mathcal{Y}) : (\pi_0 \sim \mu) \wedge (\pi_1 \sim \nu)\}$.

Si P_{01}^* es la (única) solución del SBP estático, la (única) solución del SBP dinámico es

$$P^*(\cdot) = \int_{\mathcal{X} \times \mathcal{Y}} R_{|xy}(\cdot) dP_{01}^*(x, y).$$

Problema de Schrödinger | Formulación estática

Se puede demostrar la siguiente descomposición:

$$D_{KL}(P \parallel R) = D_{KL}(P_{01} \parallel R_{01}) + \mathbb{E}_{(x,y) \sim P_{01}} [D_{KL}(P_{|xy} \parallel R_{|xy})].$$

Por lo tanto, el SBP dinámico se puede reducir a un problema estático enfocado únicamente en los extremos del proceso:

$$\underbrace{\min_{P: (P_0 \sim \mu) \wedge (P_1 \sim \nu)} D_{KL}(P \parallel R)}_{\text{problema dinámico}} = \underbrace{\min_{P_{01} \in \Pi(\mu, \nu)} D_{KL}(P_{01} \parallel R_{01})}_{\text{problema estático}},$$

donde $\Pi(\mu, \nu) := \{\pi \in \mathcal{M}_+^1(\mathcal{X} \times \mathcal{Y}) : (\pi_0 \sim \mu) \wedge (\pi_1 \sim \nu)\}$.

Si P_{01}^* es la (única) solución del SBP estático, la (única) solución del SBP dinámico es

$$P^*(\cdot) = \int_{\mathcal{X} \times \mathcal{Y}} R_{|xy}(\cdot) dP_{01}^*(x, y).$$

Problema de Schrödinger | Formulación estática

Se puede demostrar la siguiente descomposición:

$$D_{KL}(P \parallel R) = D_{KL}(P_{01} \parallel R_{01}) + \mathbb{E}_{(x,y) \sim P_{01}} [D_{KL}(P_{|xy} \parallel R_{|xy})].$$

Por lo tanto, el SBP dinámico se puede reducir a un problema estático enfocado únicamente en los extremos del proceso:

$$\underbrace{\min_{P: (P_0 \sim \mu) \wedge (P_1 \sim \nu)} D_{KL}(P \parallel R)}_{\text{problema dinámico}} = \underbrace{\min_{P_{01} \in \Pi(\mu, \nu)} D_{KL}(P_{01} \parallel R_{01})}_{\text{problema estático}},$$

donde $\Pi(\mu, \nu) := \{\pi \in \mathcal{M}_+^1(\mathcal{X} \times \mathcal{Y}) : (\pi_0 \sim \mu) \wedge (\pi_1 \sim \nu)\}$.

Si P_{01}^* es la (única) solución del SBP estático, la (única) solución del SBP dinámico es

$$P^*(\cdot) = \int_{\mathcal{X} \times \mathcal{Y}} R_{|xy}(\cdot) dP_{01}^*(x, y).$$

Problema de Schrödinger | Formulación estática

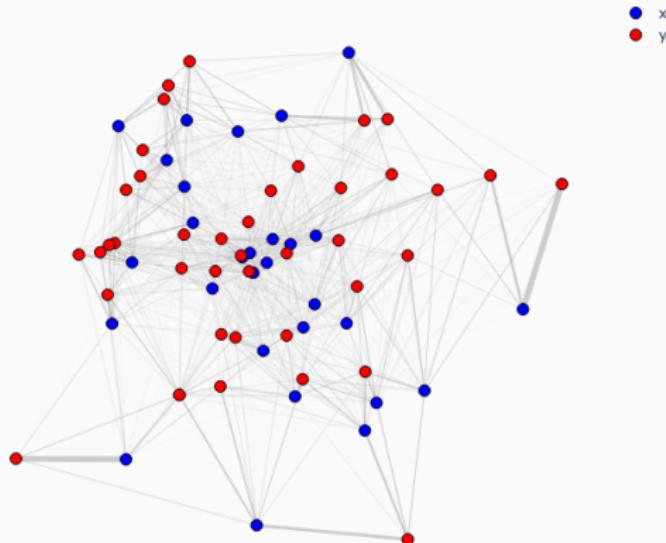


Figura 5: Plan de transporte para $\sigma = 1$.

Problema de Schrödinger | Formulación estática

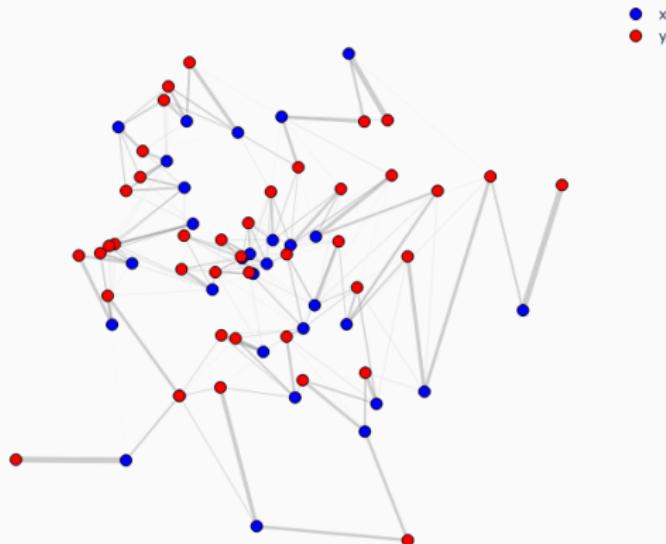


Figura 5: Plan de transporte para $\sigma = 0,1$.

Problema de Schrödinger | Formulación estática

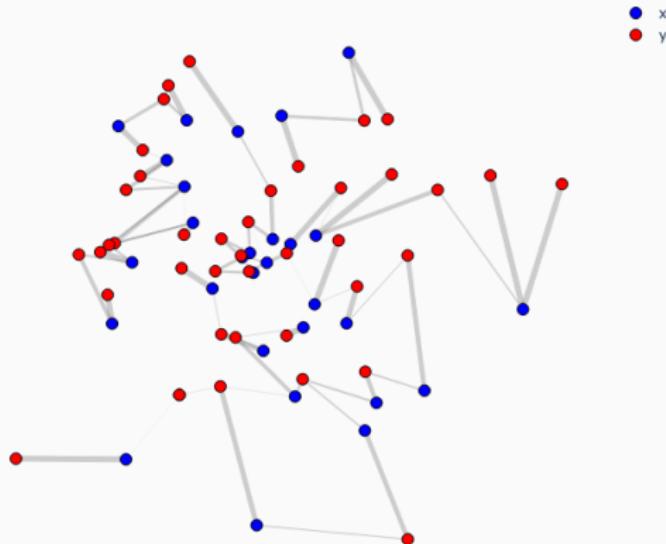


Figura 5: Plan de transporte para $\sigma = 0,01$.

Problema de Schrödinger | Formulación estática

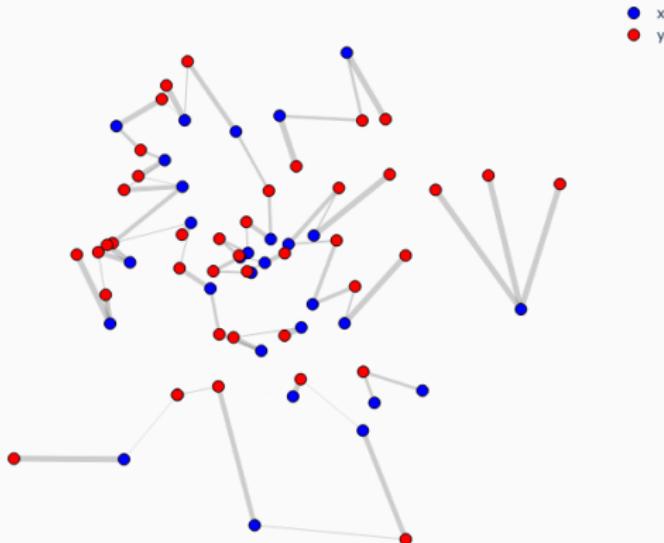


Figura 5: Plan de transporte para $\sigma = 0,005$.

Problema de Schrödinger | Formulación estática

El SBP estático es equivalente al problema de transporte óptimo con regularización entrópica (EOT).

$$D_{KL}(P_{01} \parallel R_{01})$$

$$= \frac{1}{\sigma} \left[\underbrace{\int_{\mathcal{X} \times \mathcal{Y}} \frac{1}{2} \|x - y\|^2 dP_{01}(x, y)}_{\text{costo de transporte}} + \underbrace{-\sigma \cdot \mathcal{H}(P_{01})}_{\text{regularizador}} \right] + \text{constante},$$

donde $\mathcal{H}(P_{01})$ es la entropía (diferencial) de P_{01} .

En general, todo SBP (con una cierta medida de referencia) puede ser transformado a un problema de EOT (con una cierta función de costo), y viceversa.

Problema de Schrödinger | Formulación estática

El SBP estático es equivalente al problema de transporte óptimo con regularización entrópica (EOT).

$$D_{KL}(P_{01} \| R_{01})$$

$$= \frac{1}{\sigma} \left[\underbrace{\int_{\mathcal{X} \times \mathcal{Y}} \frac{1}{2} \|x - y\|^2 dP_{01}(x, y)}_{\text{costo de transporte}} + \underbrace{-\sigma \cdot \mathcal{H}(P_{01})}_{\text{regularizador}} \right] + \text{constante},$$

donde $\mathcal{H}(P_{01})$ es la entropía (diferencial) de P_{01} .

En general, todo SBP (con una cierta medida de referencia) puede ser transformado a un problema de EOT (con una cierta función de costo), y viceversa.

Problema de Schrödinger | Formulación estática

El SBP estático es equivalente al problema de transporte óptimo con regularización entrópica (EOT).

$$D_{KL}(P_{01} \parallel R_{01})$$

$$= \frac{1}{\sigma} \left[\underbrace{\int_{\mathcal{X} \times \mathcal{Y}} \frac{1}{2} \|x - y\|^2 dP_{01}(x, y)}_{\text{costo de transporte}} + \underbrace{-\sigma \cdot \mathcal{H}(P_{01})}_{\text{regularizador}} \right] + \text{constante},$$

donde $\mathcal{H}(P_{01})$ es la entropía (diferencial) de P_{01} .

En general, todo SBP (con una cierta medida de referencia) puede ser transformado a un problema de EOT (con una cierta función de costo), y viceversa.

Transporte óptimo

Transporte óptimo | Problema de Kantorovich

Problema de Kantorovich

Si $c : \mathcal{X} \times \mathcal{Y} \rightarrow \mathbb{R}$ es una medida de discrepancia, el problema de Kantorovich entre $\mu \in \mathcal{M}_+^1(\mathcal{X})$ y $\nu \in \mathcal{M}_+^1(\mathcal{Y})$ es:

$$\min_{\pi \in \Pi(\mu, \nu)} \int_{\mathcal{X} \times \mathcal{Y}} c(x, y) \, d\pi(x, y).$$

- Bajo ciertas condiciones, el problema de Kantorovich tiene solución.
- Este funcional de costo, más un término regularizador, equivale al SBP estático.
- Por simplicidad, se considerará siempre que $c(x, y) = \frac{1}{2} \|x - y\|^2$.

Problema de Kantorovich

Si $c : \mathcal{X} \times \mathcal{Y} \rightarrow \mathbb{R}$ es una medida de discrepancia, el problema de Kantorovich entre $\mu \in \mathcal{M}_+^1(\mathcal{X})$ y $\nu \in \mathcal{M}_+^1(\mathcal{Y})$ es:

$$\min_{\pi \in \Pi(\mu, \nu)} \int_{\mathcal{X} \times \mathcal{Y}} c(x, y) \, d\pi(x, y).$$

- Bajo ciertas condiciones, el problema de Kantorovich tiene solución.
- Este funcional de costo, más un término regularizador, equivale al SBP estático.
- Por simplicidad, se considerará siempre que $c(x, y) = \frac{1}{2} \|x - y\|^2$.

Problema de Kantorovich

Si $c : \mathcal{X} \times \mathcal{Y} \rightarrow \mathbb{R}$ es una medida de discrepancia, el problema de Kantorovich entre $\mu \in \mathcal{M}_+^1(\mathcal{X})$ y $\nu \in \mathcal{M}_+^1(\mathcal{Y})$ es:

$$\min_{\pi \in \Pi(\mu, \nu)} \int_{\mathcal{X} \times \mathcal{Y}} c(x, y) d\pi(x, y).$$

- Bajo ciertas condiciones, el problema de Kantorovich tiene solución.
- Este funcional de costo, más un término regularizador, equivale al SBP estático.
- Por simplicidad, se considerará siempre que $c(x, y) = \frac{1}{2} \|x - y\|^2$.

Problema de Kantorovich

Si $c : \mathcal{X} \times \mathcal{Y} \rightarrow \mathbb{R}$ es una medida de discrepancia, el problema de Kantorovich entre $\mu \in \mathcal{M}_+^1(\mathcal{X})$ y $\nu \in \mathcal{M}_+^1(\mathcal{Y})$ es:

$$\min_{\pi \in \Pi(\mu, \nu)} \int_{\mathcal{X} \times \mathcal{Y}} c(x, y) \, d\pi(x, y).$$

- Bajo ciertas condiciones, el problema de Kantorovich tiene solución.
- Este funcional de costo, más un término regularizador, equivale al SBP estático.
- Por simplicidad, se considerará siempre que $c(x, y) = \frac{1}{2} \|x - y\|^2$.

Transporte óptimo | Problema de Kantorovich

El problema de Kantorovich discreto es un problema lineal, por lo que se puede resolver eficientemente.

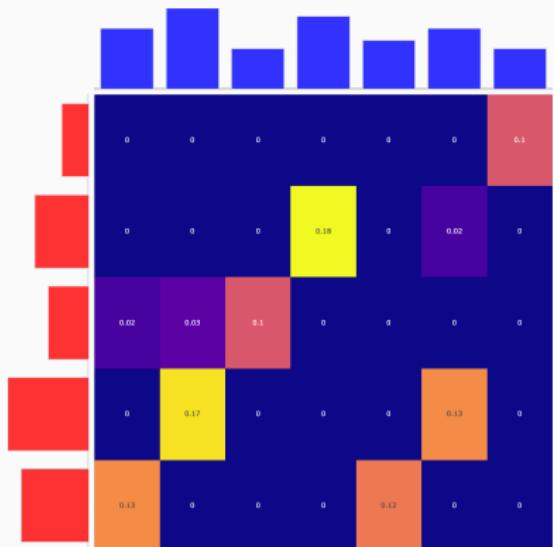


Figura 6: Caso discreto.

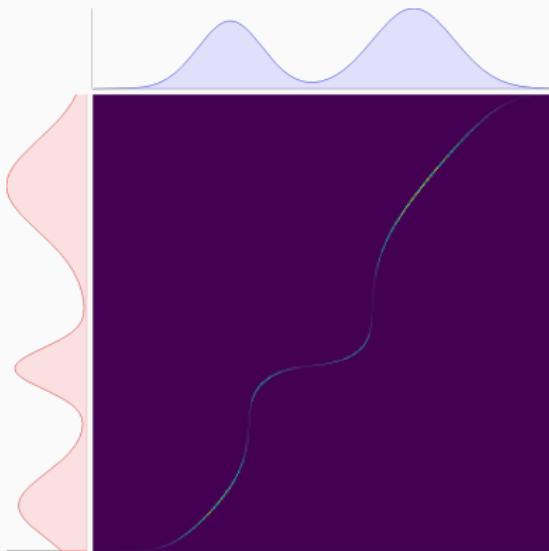


Figura 7: Caso continuo discretizado.

Transporte óptimo | Distancia de Wasserstein

El problema de Kantorovich induce una métrica en (un subconjunto de) $\mathcal{M}_+^1(\mathcal{X})$, llamada distancia de Wasserstein. Para $p \in [1, \infty]$:

$$\mathcal{W}_p(\mu, \nu)^p := \min_{\pi \in \Pi(\mu, \nu)} \int_{\mathcal{X} \times \mathcal{Y}} \|x - y\|^p d\pi(x, y)$$

Transporte óptimo | Distancia de Wasserstein

El problema de Kantorovich induce una métrica en (un subconjunto de) $\mathcal{M}_+^1(\mathcal{X})$, llamada distancia de Wasserstein. Para $p \in [1, \infty]$:

$$\mathcal{W}_p(\mu, \nu)^p := \min_{\pi \in \Pi(\mu, \nu)} \int_{\mathcal{X} \times \mathcal{Y}} \|x - y\|^p d\pi(x, y)$$

Es posible interpolar entre dos distribuciones de probabilidad a través de una curva de distribuciones en $\mathcal{M}_+^1(\mathcal{X})$.

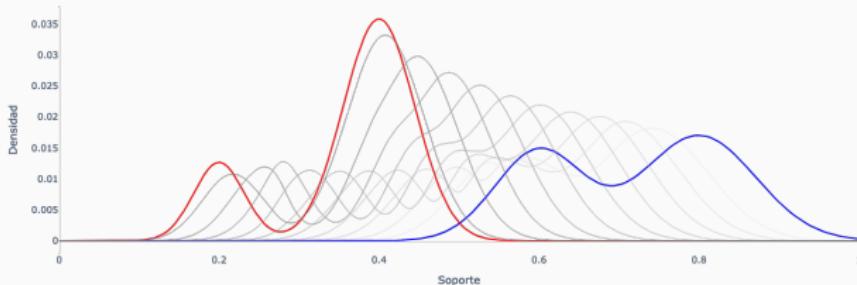


Figura 8: Interpolación entre dos distribuciones gaussianas.

Transporte óptimo | Distancia de Wasserstein

La distancia de Wasserstein tiene buenas propiedades.

- Es una distancia completa.
- Es más débil que la distancia en variación total.
- Más aún, metriza la convergencia débil de medidas.
- Puede ser optimizada por redes neuronales: si $g_\theta(z) \sim \mu_\theta$ es un modelo generativo neuronal de variable latente $z \sim \mathcal{N}(0, I_l)$, entonces $\theta \mapsto \mathcal{W}_1(\mu_\theta, \mu_{\text{true}})$ es diferenciable.

Todas estas propiedades motivan a usar la distancia de Wasserstein en modelos generativos entrenados con redes neuronales.

Transporte óptimo | Distancia de Wasserstein

La distancia de Wasserstein tiene buenas propiedades.

- Es una distancia completa.
- Es más débil que la distancia en variación total.
- Más aún, metriza la convergencia débil de medidas.
- Puede ser optimizada por redes neuronales: si $g_\theta(z) \sim \mu_\theta$ es un modelo generativo neuronal de variable latente $z \sim \mathcal{N}(0, I_l)$, entonces $\theta \mapsto \mathcal{W}_1(\mu_\theta, \mu_{\text{true}})$ es diferenciable.

Todas estas propiedades motivan a usar la distancia de Wasserstein en modelos generativos entrenados con redes neuronales.

Transporte óptimo | Distancia de Wasserstein

La distancia de Wasserstein tiene buenas propiedades.

- Es una distancia completa.
- Es más débil que la distancia en variación total.
- Más aún, metriza la convergencia débil de medidas.
- Puede ser optimizada por redes neuronales: si $g_\theta(z) \sim \mu_\theta$ es un modelo generativo neuronal de variable latente $z \sim \mathcal{N}(0, I_l)$, entonces $\theta \mapsto \mathcal{W}_1(\mu_\theta, \mu_{\text{true}})$ es diferenciable.

Todas estas propiedades motivan a usar la distancia de Wasserstein en modelos generativos entrenados con redes neuronales.

Transporte óptimo | Distancia de Wasserstein

La distancia de Wasserstein tiene buenas propiedades.

- Es una distancia completa.
- Es más débil que la distancia en variación total.
- Más aún, metriza la convergencia débil de medidas.
- Puede ser optimizada por redes neuronales: si $g_\theta(z) \sim \mu_\theta$ es un modelo generativo neuronal de variable latente $z \sim \mathcal{N}(0, I_l)$, entonces $\theta \mapsto \mathcal{W}_1(\mu_\theta, \mu_{\text{true}})$ es diferenciable.

Todas estas propiedades motivan a usar la distancia de Wasserstein en modelos generativos entrenados con redes neuronales.

Transporte óptimo | Distancia de Wasserstein

La distancia de Wasserstein tiene buenas propiedades.

- Es una distancia completa.
- Es más débil que la distancia en variación total.
- Más aún, metriza la convergencia débil de medidas.
- Puede ser optimizada por redes neuronales: si $g_\theta(z) \sim \mu_\theta$ es un modelo generativo neuronal de variable latente $z \sim \mathcal{N}(0, I_l)$, entonces $\theta \mapsto \mathcal{W}_1(\mu_\theta, \mu_{\text{true}})$ es diferenciable.

Todas estas propiedades motivan a usar la distancia de Wasserstein en modelos generativos entrenados con redes neuronales.

Transporte óptimo | Distancia de Wasserstein

La distancia de Wasserstein tiene buenas propiedades.

- Es una distancia completa.
- Es más débil que la distancia en variación total.
- Más aún, metriza la convergencia débil de medidas.
- Puede ser optimizada por redes neuronales: si $g_\theta(z) \sim \mu_\theta$ es un modelo generativo neuronal de variable latente $z \sim \mathcal{N}(0, I_l)$, entonces $\theta \mapsto \mathcal{W}_1(\mu_\theta, \mu_{\text{true}})$ es diferenciable.

Todas estas propiedades motivan a usar la distancia de Wasserstein en modelos generativos entrenados con redes neuronales.

Transporte óptimo | Regularización

El problema de Kantorovich es costoso de resolver, por lo que se preferirá una versión regularizada:

EOT (a.k.a. SBP estático)

El problema de Kantorovich regularizado es

$$\min_{\pi \in \Pi(\mu, \nu)} \int_{\mathcal{X} \times \mathcal{Y}} \|x - y\|^2 d\pi(x, y) - \sigma \mathcal{H}(\pi),$$

donde $\mathcal{H}(\pi) = - \int_{\mathcal{X} \times \mathcal{Y}} \log \left(\frac{d\pi}{dxdy}(x, y) \right) d\pi(x, y)$ es la entropía de π y $\sigma > 0$ es un ponderador de regularización.

Esta versión regularizada es más fácil de resolver y tiene buenas propiedades.

Transporte óptimo | Regularización

El problema de Kantorovich es costoso de resolver, por lo que se preferirá una versión regularizada:

EOT (a.k.a. SBP estático)

El problema de Kantorovich regularizado es

$$\min_{\pi \in \Pi(\mu, \nu)} \int_{\mathcal{X} \times \mathcal{Y}} \|x - y\|^2 d\pi(x, y) - \sigma \mathcal{H}(\pi),$$

donde $\mathcal{H}(\pi) = - \int_{\mathcal{X} \times \mathcal{Y}} \log \left(\frac{d\pi}{dxdy}(x, y) \right) d\pi(x, y)$ es la entropía de π y $\sigma > 0$ es un ponderador de regularización.

Esta versión regularizada es más fácil de resolver y tiene buenas propiedades.

Transporte óptimo | Regularización

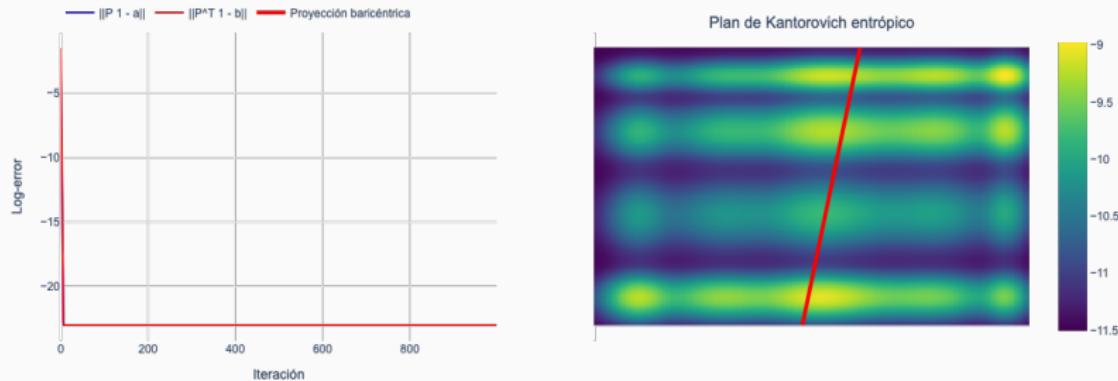


Figura 9: Plan de transporte entrópico para $\sigma = 1$.

Transporte óptimo | Regularización

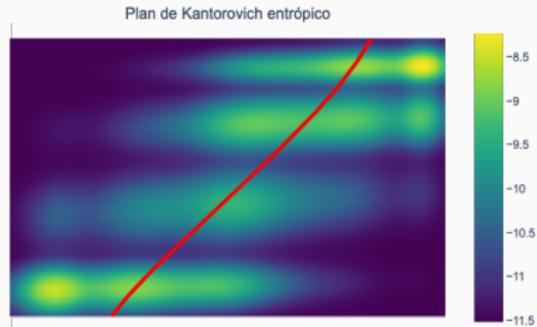
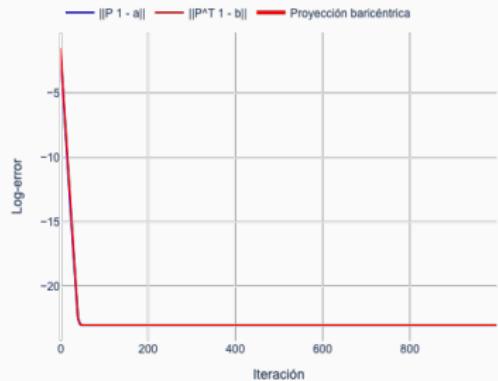


Figura 9: Plan de transporte entrópico para $\sigma = 0,1$.

Transporte óptimo | Regularización

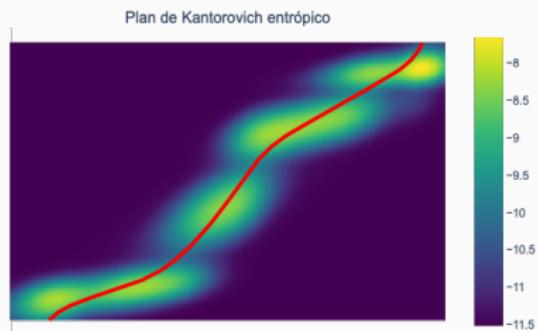
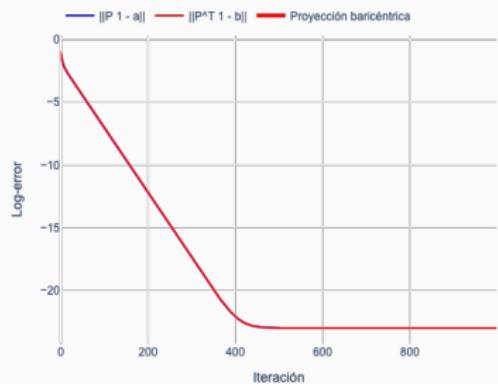


Figura 9: Plan de transporte entrópico para $\sigma = 0,01$.

Transporte óptimo | Regularización

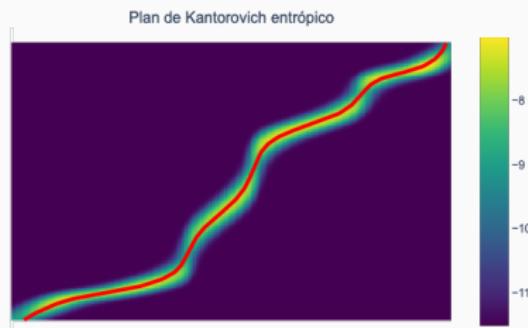
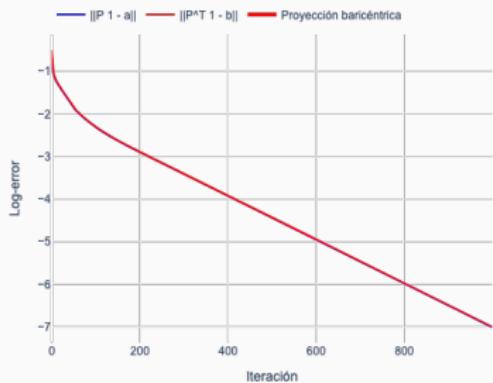


Figura 9: Plan de transporte entrópico para $\sigma = 0,001$.

Transporte óptimo | Regularización

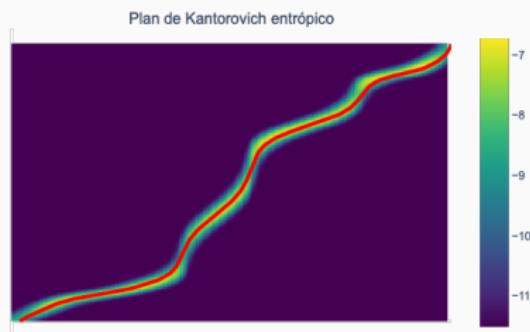
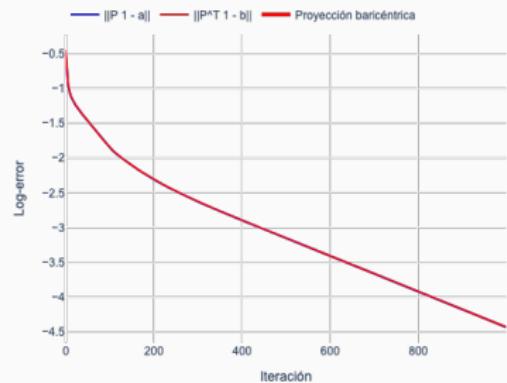


Figura 9: Plan de transporte entrópico para $\sigma = 0,0005$.

OT y SBP | Formulación dual

OT

El problema dual del problema de Kantorovich es

$$\sup_{\phi, \psi} \int_{\mathcal{X}} \phi(x) \, d\mu(x) + \int_{\mathcal{Y}} \psi(y) \, d\nu(y) \quad \text{s.a } \phi(x) + \psi(y) \leq \|x - y\|^2.$$

SBP estático

El problema dual del SBP estático es

$$\sup_{\phi, \psi} \int_{\mathcal{X}} \phi(x) \, d\mu(x) + \int_{\mathcal{Y}} \psi(y) \, d\nu(y) - \sigma \int e^{\frac{\phi(x) + \psi(y) - \|x - y\|^2}{\sigma} - 1} \, d\pi(x, y).$$

Además, la única solución del SBP estático es

$$d\pi(x, y) = e^{\frac{\phi(x) + \psi(y) - \|x - y\|^2}{\sigma} - 1} \, d\mu(x) \, d\nu(y)$$

OT

El problema dual del problema de Kantorovich es

$$\sup_{\phi, \psi} \int_{\mathcal{X}} \phi(x) \, d\mu(x) + \int_{\mathcal{Y}} \psi(y) \, d\nu(y) \quad \text{s.a } \phi(x) + \psi(y) \leq \|x - y\|^2.$$

SBP estático

El problema dual del SBP estático es

$$\sup_{\phi, \psi} \int_{\mathcal{X}} \phi(x) \, d\mu(x) + \int_{\mathcal{Y}} \psi(y) \, d\nu(y) - \sigma \int e^{\frac{\phi(x) + \psi(y) - \|x - y\|^2}{\sigma} - 1} \, d\pi(x, y).$$

Además, la única solución del SBP estático es

$$d\pi(x, y) = e^{\frac{\phi(x) + \psi(y) - \|x - y\|^2}{\sigma} - 1} \, d\mu(x) \, d\nu(y)$$

OT dinámico (Benamou-Brenier)

$$\min_v \mathbb{E}_x \left[\int_0^1 \frac{1}{2} \|v(x_t, t)\|^2 dt \right] \quad \text{sujeto a} \quad \begin{cases} dx_t = v(x_t, t) dt \\ (x_0 \sim \mu) \wedge (x_1 \sim \nu), \end{cases}$$

Se puede sustituir la ODE por su ecuación de continuidad.

SBP dinámico

$$\min_v \mathbb{E}_x \left[\int_0^1 \frac{1}{2} \|v(x_t, t)\|^2 dt \right] \quad \text{sujeto a} \quad \begin{cases} dx_t = v(x_t, t) dt + \sigma dw_t \\ (x_0 \sim \mu) \wedge (x_1 \sim \nu), \end{cases}$$

Se puede sustituir la SDE por su ecuación de Fokker-Planck.

OT dinámico (Benamou-Brenier)

$$\min_v \mathbb{E}_x \left[\int_0^1 \frac{1}{2} \|v(x_t, t)\|^2 dt \right] \quad \text{sujeto a} \quad \begin{cases} dx_t = v(x_t, t) dt \\ (x_0 \sim \mu) \wedge (x_1 \sim \nu), \end{cases}$$

Se puede sustituir la ODE por su ecuación de continuidad.

SBP dinámico

$$\min_v \mathbb{E}_x \left[\int_0^1 \frac{1}{2} \|v(x_t, t)\|^2 dt \right] \quad \text{sujeto a} \quad \begin{cases} dx_t = v(x_t, t) dt + \sigma dw_t \\ (x_0 \sim \mu) \wedge (x_1 \sim \nu), \end{cases}$$

Se puede sustituir la SDE por su ecuación de Fokker-Planck.

OT y SBP | Como problemas de control óptimo

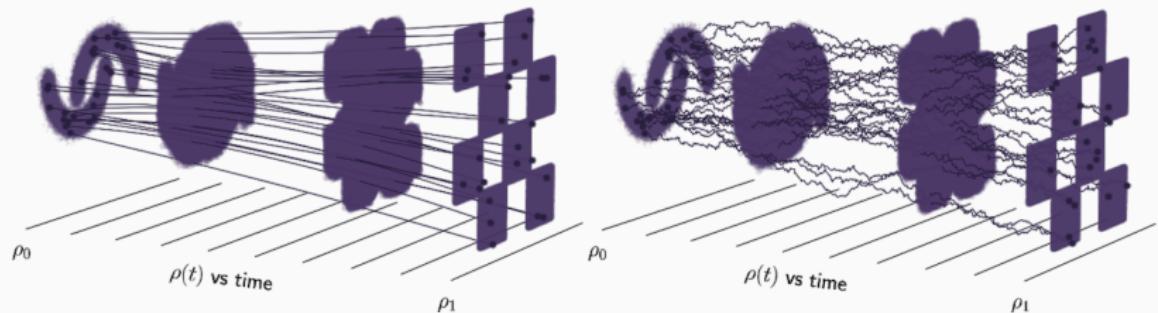


Figura 10: Imagen obtenida desde Albergo et al., 2023.

OT y SBP | Como sistema de PDEs

- Se pueden sustituir las dinámicas por PDEs que las caractericen.
- La optimalidad de los problemas de control también se puede caracterizar por PDEs.

Caracterización de optimalidad para OT

$$\begin{cases} \partial_t \rho + \nabla \cdot (\rho \nabla \lambda) = 0 & \text{(eq. continuidad)} \\ \partial_t \lambda + \frac{1}{2} \|\nabla \lambda\|^2 = 0 & \text{(Hamilton-Jacobi)} \end{cases}$$

Caracterización de optimalidad para SBP (sistema de Schrödinger)

$$\begin{cases} \partial_t \rho + \nabla \cdot (\rho \nabla \lambda) - \frac{\sigma^2}{2} \Delta \rho = 0 & \text{(Fokker-Planck)} \\ \partial_t \lambda + \frac{1}{2} \|\nabla \lambda\|^2 + \frac{\sigma^2}{2} \Delta \lambda = 0 & \text{(Hamilton-Jacobi-Bellman)} \end{cases}$$

OT y SBP | Como sistema de PDEs

- Se pueden sustituir las dinámicas por PDEs que las caractericen.
- La optimalidad de los problemas de control también se puede caracterizar por PDEs.

Caracterización de optimalidad para OT

$$\begin{cases} \partial_t \rho + \nabla \cdot (\rho \nabla \lambda) = 0 & \text{(eq. continuidad)} \\ \partial_t \lambda + \frac{1}{2} \|\nabla \lambda\|^2 = 0 & \text{(Hamilton-Jacobi)} \end{cases}$$

Caracterización de optimalidad para SBP (sistema de Schrödinger)

$$\begin{cases} \partial_t \rho + \nabla \cdot (\rho \nabla \lambda) - \frac{\sigma^2}{2} \Delta \rho = 0 & \text{(Fokker-Planck)} \\ \partial_t \lambda + \frac{1}{2} \|\nabla \lambda\|^2 + \frac{\sigma^2}{2} \Delta \lambda = 0 & \text{(Hamilton-Jacobi-Bellman)} \end{cases}$$

- Se pueden sustituir las dinámicas por PDEs que las caractericen.
- La optimalidad de los problemas de control también se puede caracterizar por PDEs.

Caracterización de optimalidad para OT

$$\begin{cases} \partial_t \rho + \nabla \cdot (\rho \nabla \lambda) = 0 & \text{(eq. continuidad)} \\ \partial_t \lambda + \frac{1}{2} \|\nabla \lambda\|^2 = 0 & \text{(Hamilton-Jacobi)} \end{cases}$$

Caracterización de optimalidad para SBP (sistema de Schrödinger)

$$\begin{cases} \partial_t \rho + \nabla \cdot (\rho \nabla \lambda) - \frac{\sigma^2}{2} \Delta \rho = 0 & \text{(Fokker-Planck)} \\ \partial_t \lambda + \frac{1}{2} \|\nabla \lambda\|^2 + \frac{\sigma^2}{2} \Delta \lambda = 0 & \text{(Hamilton-Jacobi-Bellman)} \end{cases}$$

OT y SBP | Como sistema de PDEs

- Se pueden sustituir las dinámicas por PDEs que las caractericen.
- La optimalidad de los problemas de control también se puede caracterizar por PDEs.

Caracterización de optimalidad para OT

$$\begin{cases} \partial_t \rho + \nabla \cdot (\rho \nabla \lambda) = 0 & \text{(eq. continuidad)} \\ \partial_t \lambda + \frac{1}{2} \|\nabla \lambda\|^2 = 0 & \text{(Hamilton-Jacobi)} \end{cases}$$

Caracterización de optimalidad para SBP (sistema de Schrödinger)

$$\begin{cases} \partial_t \rho + \nabla \cdot (\rho \nabla \lambda) - \frac{\sigma^2}{2} \Delta \rho = 0 & \text{(Fokker-Planck)} \\ \partial_t \lambda + \frac{1}{2} \|\nabla \lambda\|^2 + \frac{\sigma^2}{2} \Delta \lambda = 0 & \text{(Hamilton-Jacobi-Bellman)} \end{cases}$$

OT y SBP | Distintos modelos

- Todos estas formulaciones son equivalentes entre sí.
- Esta diversidad de formulaciones permite atacar el problema de Schrödinger desde distintos ángulos.
- La mayoría de técnicas usadas son variantes del algoritmo de Sinkhorn.
- Un modelo para el SBP puede, en principio, resolver los mismos problemas que un modelo de difusión.

OT y SBP | Distintos modelos

- Todos estas formulaciones son equivalentes entre sí.
- Esta diversidad de formulaciones permite atacar el problema de Schrödinger desde distintos ángulos.
- La mayoría de técnicas usadas son variantes del algoritmo de Sinkhorn.
- Un modelo para el SBP puede, en principio, resolver los mismos problemas que un modelo de difusión.

OT y SBP | Distintos modelos

- Todos estas formulaciones son equivalentes entre sí.
- Esta diversidad de formulaciones permite atacar el problema de Schrödinger desde distintos ángulos.
- La mayoría de técnicas usadas son variantes del algoritmo de Sinkhorn.
- Un modelo para el SBP puede, en principio, resolver los mismos problemas que un modelo de difusión.

OT y SBP | Distintos modelos

- Todos estas formulaciones son equivalentes entre sí.
- Esta diversidad de formulaciones permite atacar el problema de Schrödinger desde distintos ángulos.
- La mayoría de técnicas usadas son variantes del algoritmo de Sinkhorn.
- Un modelo para el SBP puede, en principio, resolver los mismos problemas que un modelo de difusión.

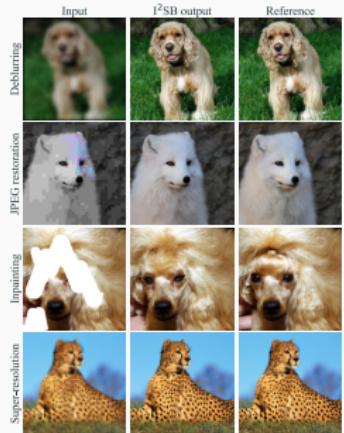


Figura 11: Imagen obtenida desde Liu et al., 2023.

OT y SBP | Distintos modelos

- Todos estas formulaciones son equivalentes entre sí.
- Esta diversidad de formulaciones permite atacar el problema de Schrödinger desde distintos ángulos.
- La mayoría de técnicas usadas son variantes del algoritmo de Sinkhorn.
- Un modelo para el SBP puede, en principio, resolver los mismos problemas que un modelo de difusión.



Figura 11: Imagen obtenida desde Korotin et al., 2024.

Conclusión

- El problema de Schrödinger puede verse como un paradigma generativo que generaliza los modelos de difusión.
- Este modelo es complejo de formular dado que ha sido estudiado principalmente desde una perspectiva matemática.
- En consecuencia, en este trabajo se buscó una presentación sencilla de las cosas más importantes, con demostraciones simples e implementaciones minimales.
- No obstante, se intentó mantener siempre la coherencia matemática, la cual muchas veces es, o bien escasa, o bien engorrosa en la formulación.
- Se espera que este trabajo pueda servir como punto de introducción a cualquier persona interesada en el problema de Schrödinger desde una perspectiva del aprendizaje automático.

Conclusión

- El problema de Schrödinger puede verse como un paradigma generativo que generaliza los modelos de difusión.
- Este modelo es complejo de formular dado que ha sido estudiado principalmente desde una perspectiva matemática.
- En consecuencia, en este trabajo se buscó una presentación sencilla de las cosas más importantes, con demostraciones simples e implementaciones minimales.
- No obstante, se intentó mantener siempre la coherencia matemática, la cual muchas veces es, o bien escasa, o bien engorrosa en la formulación.
- Se espera que este trabajo pueda servir como punto de introducción a cualquier persona interesada en el problema de Schrödinger desde una perspectiva del aprendizaje automático.

Conclusión

- El problema de Schrödinger puede verse como un paradigma generativo que generaliza los modelos de difusión.
- Este modelo es complejo de formular dado que ha sido estudiado principalmente desde una perspectiva matemática.
- En consecuencia, en este trabajo se buscó una presentación sencilla de las cosas más importantes, con demostraciones simples e implementaciones minimales.
- No obstante, se intentó mantener siempre la coherencia matemática, la cual muchas veces es, o bien escasa, o bien engorrosa en la formulación.
- Se espera que este trabajo pueda servir como punto de introducción a cualquier persona interesada en el problema de Schrödinger desde una perspectiva del aprendizaje automático.

Conclusión

- El problema de Schrödinger puede verse como un paradigma generativo que generaliza los modelos de difusión.
- Este modelo es complejo de formular dado que ha sido estudiado principalmente desde una perspectiva matemática.
- En consecuencia, en este trabajo se buscó una presentación sencilla de las cosas más importantes, con demostraciones simples e implementaciones minimales.
- No obstante, se intentó mantener siempre la coherencia matemática, la cual muchas veces es, o bien escasa, o bien engorrosa en la formulación.
- Se espera que este trabajo pueda servir como punto de introducción a cualquier persona interesada en el problema de Schrödinger desde una perspectiva del aprendizaje automático.

Conclusión

- El problema de Schrödinger puede verse como un paradigma generativo que generaliza los modelos de difusión.
- Este modelo es complejo de formular dado que ha sido estudiado principalmente desde una perspectiva matemática.
- En consecuencia, en este trabajo se buscó una presentación sencilla de las cosas más importantes, con demostraciones simples e implementaciones minimales.
- No obstante, se intentó mantener siempre la coherencia matemática, la cual muchas veces es, o bien escasa, o bien engorrosa en la formulación.
- Se espera que este trabajo pueda servir como punto de introducción a cualquier persona interesada en el problema de Schrödinger desde una perspectiva del aprendizaje automático.

El Problema del Puente de Schrödinger como Generalización de los Modelos de Difusión