# Disentangling entropy and suboptimality in Entropic optimal transport

arXiv:2306.06940

Hugo Malamut and Maxime Sylvestre
(Université Paris Dauphine PSL, CEREMADE)

March 24, 2025

# Optimal transport

# Transport map

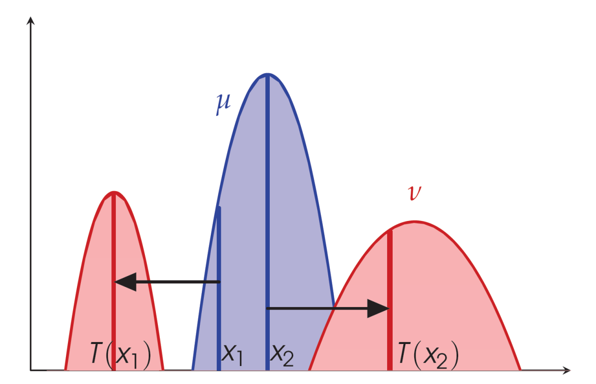


**Figure 1:** the map T transports $\mu$ onto $\nu$ : $T_{\#}\mu = \nu$

# Transport plan

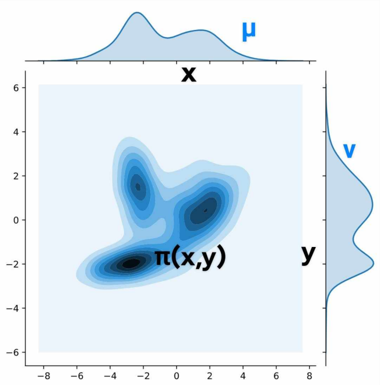


**Figure 2:** The plan $\pi$ transports $\mu$ onto $\nu$ : $\pi \in \Pi(\mu, \nu)$ (image from Wikipedia)

# Optimal transport

Given a cost function $c : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ and two measures $\mu_0, \mu_1 \in \mathcal{P}(\mathbb{R}^d)$,

# Optimal transport

Given a cost function $c : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ and two measures $\mu_0, \mu_1 \in \mathcal{P}(\mathbb{R}^d)$,

- Monge problem (1781) is

$$\inf_{T\#\mu_0=\mu_1} \int c(x, T(x)) \mathrm{d}\mu_0(x) \tag{1}$$

- Kantorovich formulation:

$$OT_0(\mu, \nu) := \inf_{\gamma \in \Pi(\mu_0,\mu_1)} \int c(x, y) \mathrm{d}\gamma(x, y) \tag{P}$$

- Dual problem:

$$OT_0(\mu, \nu) := \sup_{\phi \oplus \psi \leq c} \int \phi \mathrm{d}\mu_0 + \int \psi \mathrm{d}\mu_1 \tag{D}$$

# Optimal transport

Given a cost function $c : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ and two measures $\mu_0, \mu_1 \in \mathcal{P}(\mathbb{R}^d)$,

- Monge problem (1781) is

$$\inf_{T\#\mu_0=\mu_1} \int c(x, T(x))\mathrm{d}\mu_0(x) \tag{1}$$

- Kantorovich formulation:

$$OT_0(\mu,\nu) := \inf_{\gamma\in\Pi(\mu_0,\mu_1)} \int c(x,y)\mathrm{d}\gamma(x,y) \tag{P}$$

- Dual problem:

$$OT_0(\mu,\nu) := \sup_{\phi\oplus\psi\leq c} \int \phi\mathrm{d}\mu_0 + \int \psi\mathrm{d}\mu_1 \tag{D}$$

In the following, the cost function is $c(x,y) = \frac{1}{2}|x-y|^2$. It defines the Wasserstein distance $W_2^2(\mu,\nu) := \inf_{\gamma\in\Pi(\mu_0,\mu_1)} \int |x-y|^2\mathrm{d}\gamma(x,y)$

# Legendre transform

**Definition**
The Legendre transform of a function $f$ is

$$f^*(y) = \sup_x x \cdot y - f(x) \tag{2}$$

# Legendre transform

**Definition**
The Legendre transform of a function $f$ is

$$f^*(y) = \sup_x x \cdot y - f(x) \tag{2}$$

- $f^*$ is always convex
- if $f$ is convex and if both $f$ and $f^*$ are differentiable, then

$$(\nabla f)^{-1} = \nabla f^* \tag{3}$$

# Legendre transform

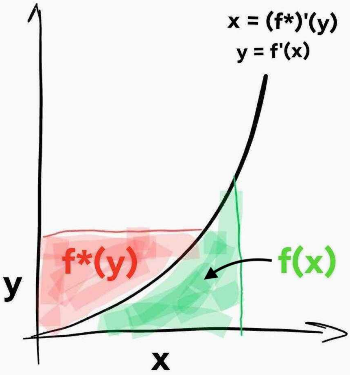


**Figure 3:** The Legendre transform in 1d

# Structure of the optimizers

$$\inf_{T\#\mu_0=\mu_1} \int c(x, T(x))\mathrm{d}\mu_0(x) \geq \min_{\gamma\in\Pi(\mu_0,\mu_1)} \int c(x,y)\mathrm{d}\gamma(x,y) = \sup_{\phi\oplus\psi\leq c} \int \phi\mathrm{d}\mu_0 + \int \psi\mathrm{d}\mu_1 \tag{4}$$

# Structure of the optimizers

$$\inf_{T\#\mu_0=\mu_1} \int c(x, T(x))\mathrm{d}\mu_0(x) \geq \min_{\gamma\in\Pi(\mu_0,\mu_1)} \int c(x,y)\mathrm{d}\gamma(x,y) = \sup_{\phi\oplus\psi\leq c} \int \phi\mathrm{d}\mu_0 + \int \psi\mathrm{d}\mu_1 \tag{4}$$

**Theorem (Brenier)**

*For the quadratic cost $c(x,y) = \frac{1}{2}|x-y|^2$:*

*If $\mu_0 \in \mathcal{P}_{2,ac}(\mathbb{R}^d)$ then the solutions $T, \gamma_0, \phi, \psi$ can be expressed thanks to a unique convex function $f$:*

## Structure of the optimizers

$$\inf_{T\#\mu_0=\mu_1} \int c(x, T(x))\mathrm{d}\mu_0(x) \geq \min_{\gamma\in\Pi(\mu_0,\mu_1)} \int c(x,y)\mathrm{d}\gamma(x,y) = \sup_{\phi\oplus\psi\leq c} \int \phi \mathrm{d}\mu_0 + \int \psi \mathrm{d}\mu_1 \tag{4}$$

**Theorem (Brenier)**

*For the quadratic cost $c(x,y) = \frac{1}{2}|x-y|^2$:*

*If $\mu_0 \in \mathcal{P}_{2,ac}(\mathbb{R}^d)$ then the solutions $T, \gamma_0, \phi, \psi$ can be expressed thanks to a unique convex function $f$:*

- $T = \nabla f$,
- $\gamma_0 = (id, \nabla f)_{\#}\mu_0$,
- $\phi(x) = \frac{1}{2}|x|^2 - f(x)$,
- $\psi(y) = \frac{1}{2}|y|^2 - f^*(y)$.

# Energy gap

Let define the Energy gap function

$$E(x, y) := c(x, y) - \phi(x) - \psi(y) \tag{5}$$

# Energy gap

Let define the Energy gap function

$$E(x,y) := c(x,y) - \phi(x) - \psi(y) \tag{5}$$

- $\forall (x,y) \in \mathbb{R}^{2d} \quad E(x,y) \geq 0$,
- $(x,y) \in Supp(\gamma_0) \Rightarrow E(x,y) = 0$,
- If $\gamma \in \Pi(\mu_0, \mu_1)$ then
$$\int c \mathrm{d}\gamma - \int c \mathrm{d}\gamma_0 = \int E \mathrm{d}\pi,$$
- For the quadratic cost,

$$E(x,y) = f(x) + f^*(y) - x \cdot y,$$

# Legendre transform

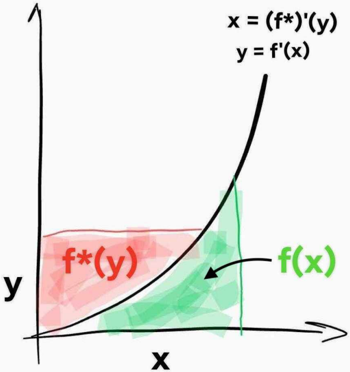


**Figure 4:** We can identify $E(x, y) = f(x) + f^*(y) - x \cdot y$ on this drawing

# Energy gap

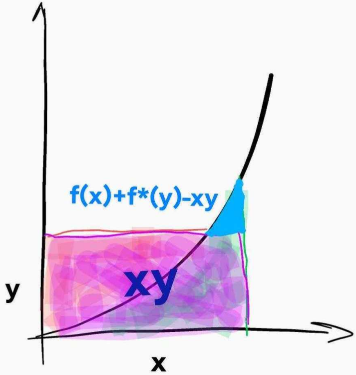


**Figure 5:** The energy gap in 1d

# Benamou-Brenier formulation

Let $\mu_0, \mu_1 \in \mathcal{P}_2(\mathbb{R}^d)$, for the quadratic cost, the Benamou-Brenier formulation is:

$$OT_0(\mu_0, \mu_1) = \inf_{\substack{\partial_t \rho + \nabla \cdot (\rho v) = 0 \\ \rho_0 = \mu_0, \rho_1 = \mu_1}} \iint \frac{1}{2} |v_t|^2 \rho_t \mathrm{d}t \tag{BB}$$

# Benamou-Brenier formulation

Let $\mu_0, \mu_1 \in \mathcal{P}_2(\mathbb{R}^d)$, for the quadratic cost, the Benamou-Brenier formulation is:

$$OT_0(\mu_0, \mu_1) = \inf_{\substack{\partial_t \rho + \nabla \cdot (\rho v) = 0 \\ \rho_0 = \mu_0, \rho_1 = \mu_1}} \iint \frac{1}{2} |v_t|^2 \rho_t \mathrm{d}t \tag{BB}$$

(The advection equation $\partial_t \rho + \nabla \cdot (\rho v) = 0$ means that the density $\rho$ is transported by the velocity field $v$)

# Entropic optimal transport

Let $\rho \in \mathcal{P}_{2,ac}(\mathbb{R}^k)$ be a continuous measure with finite variance, define

$$\underbrace{H(\rho) := \int_{\mathbb{R}^k} \rho(x) \ln \rho(x) \mathrm{d}x}_{\textit{differential entropy}} \quad \text{and} \quad \underbrace{I(\rho) := \int_{\mathbb{R}^k} \rho(x) |\nabla \ln \rho(x)|^2 \mathrm{d}x}_{\textit{Fisher information}} \tag{6}$$

# Entropic optimal transport

Let $\rho \in \mathcal{P}_{2,ac}(\mathbb{R}^k)$ be a continuous measure with finite variance, define

$$\underbrace{H(\rho) := \int_{\mathbb{R}^k} \rho(x) \ln \rho(x) \mathrm{d}x}_{\textit{differential entropy}} \quad \text{and} \quad \underbrace{I(\rho) := \int_{\mathbb{R}^k} \rho(x) |\nabla \ln \rho(x)|^2 \mathrm{d}x}_{\textit{Fisher information}} \tag{6}$$

$I(\rho) < +\infty$ implies $H(\rho) < +\infty$ Let $c$ be a $C^2$ cost function.

Let $\mu_0, \mu_1 \in \mathcal{P}_{ac}(\mathbb{R}^d)$ be such that $H(\mu_i) < +\infty$.

# Entropic optimal transport

Let $\rho \in \mathcal{P}_{2,ac}(\mathbb{R}^k)$ be a continuous measure with finite variance, define

$$\underbrace{H(\rho) := \int_{\mathbb{R}^k} \rho(x) \ln \rho(x) \mathrm{d}x}_{\textit{differential entropy}} \quad \text{and} \quad \underbrace{I(\rho) := \int_{\mathbb{R}^k} \rho(x) |\nabla \ln \rho(x)|^2 \mathrm{d}x}_{\textit{Fisher information}} \tag{6}$$

$I(\rho) < +\infty$ implies $H(\rho) < +\infty$ Let $c$ be a $C^2$ cost function.
Let $\mu_0, \mu_1 \in \mathcal{P}_{ac}(\mathbb{R}^d)$ be such that $H(\mu_i) < +\infty$.
For $\varepsilon \geq 0$

$$OT_\varepsilon(\mu_0, \mu_1) := \inf_{\gamma \in \Pi(\mu_0, \mu_1)} \int c\gamma + \varepsilon H(\gamma) \tag{$\varepsilon$EOT}$$

# Entropic optimal transport

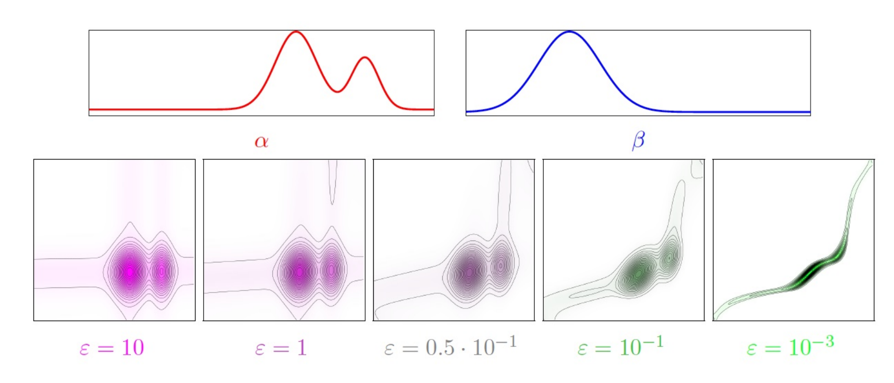


**Figure 6:** Entropic optimal transport in 1d (image from [PC18]

# Sinkhorn algorithm

The dual problem of ($\varepsilon$EOT) is

$$OT_\varepsilon(\mu_0, \mu_1) = \sup_{\phi,\psi} \int \phi \mathrm{d}\mu_0 + \int \psi \mathrm{d}\mu_1 - \varepsilon \int e^{\frac{\phi \oplus \psi - c}{\varepsilon}} \mathrm{d}\mu_0 \mathrm{d}\mu_1 + \varepsilon \quad (7)$$

# Sinkhorn algorithm

The dual problem of ($\varepsilon$EOT) is

$$OT_\varepsilon(\mu_0,\mu_1) = \sup_{\phi,\psi} \int \phi \mathrm{d}\mu_0 + \int \psi \mathrm{d}\mu_1 - \varepsilon \int e^{\frac{\phi\oplus\psi - c}{\varepsilon}} \mathrm{d}\mu_0 \mathrm{d}\mu_1 + \varepsilon \tag{7}$$

The optimality conditions are

$$\phi(x) = -\varepsilon \ln \int e^{\frac{\psi(y)-c(x,y)}{\varepsilon}} \mathrm{d}\mu_1(y), \quad \psi(y) = -\varepsilon \ln \int e^{\frac{\phi(x)-c(x,y)}{\varepsilon}} \mathrm{d}\mu_0(x). \tag{8}$$

# Sinkhorn algorithm

The dual problem of ($\varepsilon$EOT) is

$$OT_\varepsilon(\mu_0, \mu_1) = \sup_{\phi,\psi} \int \phi \mathrm{d}\mu_0 + \int \psi \mathrm{d}\mu_1 - \varepsilon \int e^{\frac{\phi \oplus \psi - c}{\varepsilon}} \mathrm{d}\mu_0 \mathrm{d}\mu_1 + \varepsilon \tag{7}$$

The optimality conditions are

$$\phi(x) = -\varepsilon \ln \int e^{\frac{\psi(y) - c(x,y)}{\varepsilon}} \mathrm{d}\mu_1(y), \quad \psi(y) = -\varepsilon \ln \int e^{\frac{\phi(x) - c(x,y)}{\varepsilon}} \mathrm{d}\mu_0(x). \tag{8}$$

## Sinkhorn algorithm

$$\phi_{n+1}(x) = -\varepsilon \ln \int e^{\frac{\psi_n(y) - c(x,y)}{\varepsilon}} \mathrm{d}\mu_1(y), \quad \psi_{n+1}(y) = -\varepsilon \ln \int e^{\frac{\phi_{n+1}(x) - c(x,y)}{\varepsilon}} \mathrm{d}\mu_0(x). \tag{9}$$

# Entropic optimal transport

For $\varepsilon \geq 0$

$$OT_\varepsilon(\mu_0, \mu_1) := \inf_{\gamma \in \Pi(\mu_0, \mu_1)} \int c\gamma + \varepsilon H(\gamma) \tag{$\varepsilon$EOT}$$

Question: What happens when $\varepsilon \to 0$ ?

# Prior Works

Qualitative convergence results.

- $\Gamma$-convergence : [Mik04],[MT08],[Lé13],[CDPS15]

Quantitative convergence results.

- Discrete optimal transport: [CM94]
- Semi-discrete optimal transport: [ANWS21],[Del21]
- Finite Fisher information: [ADPZ11],[EMR15],[Con19]
- Finite entropy: [Pal19],[EN22],[CPT22]
- Multimarginal: [NP23]
- Sinkhorn divergence: [FSV+18, CRL+20]

# Convergence of the value

**Proposition [ADPZ11][EMR15]**

Assume $c(x, y) = \frac{1}{2}\|x - y\|^2$, and that $Supp(\mu_i)$ are compact with $I(\mu_i) < +\infty$ then

$$OT_\varepsilon - OT_0 = -\frac{d}{2}\varepsilon \ln(2\pi\varepsilon) + \varepsilon\frac{H(\mu_0) + H(\mu_1)}{2} + o(\varepsilon) \qquad \text{(TE-OT}_\varepsilon\text{)}$$

# Convergence of the value

## Proposition [ADPZ11][EMR15]

Assume $c(x, y) = \frac{1}{2}\|x - y\|^2$, and that $Supp(\mu_i)$ are compact with $I(\mu_i) < +\infty$ then

$$OT_\varepsilon - OT_0 = -\frac{d}{2}\varepsilon \ln(2\pi\varepsilon) + \varepsilon \frac{H(\mu_0) + H(\mu_1)}{2} + o(\varepsilon) \qquad (\text{TE-OT}_\varepsilon)$$

## Proposition [EN22, CPT22]

Assume $c$ is infinitesimally twisted and $Supp(\mu_i)$ compact then

$$\left(-\frac{d}{2}\varepsilon \ln(\varepsilon) + C'\varepsilon \leq\right) OT_\varepsilon - OT_0 \leq -\frac{d}{2}\varepsilon \ln(\varepsilon) + C\varepsilon \qquad (10)$$

# Questions

Question 1:

How did they get these results?

# Questions

Question 1:

How did they get these results?

Question 2:

$$OT_\varepsilon - OT_0 = \underbrace{\int c\gamma_\varepsilon - \int c\gamma_0}_{suboptimality} + \varepsilon \underbrace{H(\gamma_\varepsilon)}_{entropy}$$

Can we disentangle ?

# Questions

Question 1:

How did they get these results?

Question 2:

$$OT_\varepsilon - OT_0 = \underbrace{\int c\gamma_\varepsilon - \int c\gamma_0}_{\textit{suboptimality}} + \varepsilon \underbrace{H(\gamma_\varepsilon)}_{\textit{entropy}}$$

Can we disentangle ?

Question 3:

Is there a rate of convergence for $W_2(\gamma_\varepsilon, \gamma_0)$ ?

# Block approximation

**Definition**

Let $\mu, \nu \in \mathcal{P}_{2,ac}(\mathbb{R}^d)$ and $\gamma \in \Pi(\mu, \nu)$.

Let $\delta > 0$ and $(Q_i^\delta)_i$ be a partition of $\mathbb{R}^d$ into cubes of side length $\delta$.

The block approximation of $\gamma$ is defined by

$$\forall x, y \in Q_j^\delta \times Q_k^\delta, \quad \gamma^\delta(x, y) = \gamma(Q_j^\delta \times Q_k^\delta) \frac{\mu(x)\nu(y)}{\mu(Q_j^\delta)\nu(Q_k^\delta)} \tag{11}$$

# Block approximation

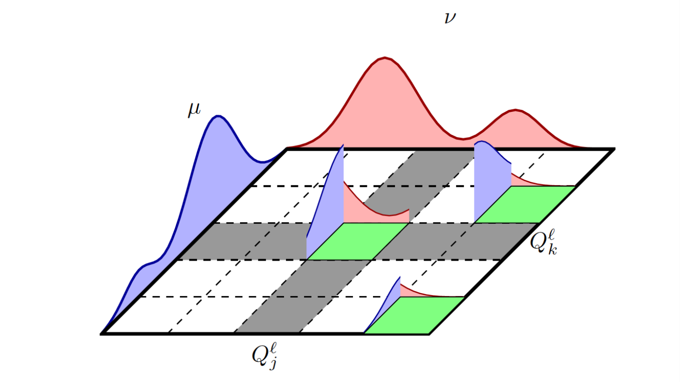


**Figure 7:** The block approximation in 1d (image from [CDPS17])

# Block approximation

[[CPT22]] Let $\mu, \nu \in \mathcal{P}_{2,ac}(\mathbb{R}^d)$ and $\gamma \in \Pi(\mu, \nu)$ be the optimal transport plan.
Let $\delta > 0$ and $\gamma^\delta$ be the block approximation of $\gamma$.
Then

$$\int c\gamma^\delta - \int c\gamma = \int E\gamma^\delta \leq C\delta^2 \quad \text{and} \quad H(\gamma^\delta) \leq -d\ln(\delta) + C \tag{12}$$

# Block approximation

[[CPT22]] Let $\mu, \nu \in \mathcal{P}_{2,ac}(\mathbb{R}^d)$ and $\gamma \in \Pi(\mu, \nu)$ be the optimal transport plan.
Let $\delta > 0$ and $\gamma^\delta$ be the block approximation of $\gamma$.
Then

$$\int c\gamma^\delta - \int c\gamma = \int E\gamma^\delta \leq C\delta^2 \quad \text{and} \quad H(\gamma^\delta) \leq -d\ln(\delta) + C \tag{12}$$

So for $\delta = \sqrt{\varepsilon}$, we have

$$OT_\varepsilon(\mu, \nu) \leq \int c\gamma^\delta + \varepsilon H(\gamma^\delta) \leq -\frac{d}{2}\ln(\varepsilon) + C\varepsilon \tag{13}$$

# Convergence of the value

**Proposition [ADPZ11][EMR15]**

Assume $c(x, y) = \frac{1}{2}\|x - y\|^2$, and that $Supp(\mu_i)$ are compact with $I(\mu_i) < +\infty$ then

$$OT_\varepsilon - OT_0 = -\frac{d}{2}\varepsilon \ln(2\pi\varepsilon) + \varepsilon\frac{H(\mu_0) + H(\mu_1)}{2} + o(\varepsilon) \qquad \text{(TE-OT}_\varepsilon\text{)}$$

**Proposition [EN22, CPT22]**

Assume $c$ is infinitesimally twisted and $Supp(\mu_i)$ compact then

$$\left(-\frac{d}{2}\varepsilon \ln(\varepsilon) + C'\varepsilon \leq\right) OT_\varepsilon - OT_0 \leq -\frac{d}{2}\varepsilon \ln(\varepsilon) + C\varepsilon \qquad (14)$$

# Questions

Question 1:

How did they get these results? by Block approximation

# Questions

Question 1:

How did they get these results? by Block approximation

Question 2:

$$OT_\varepsilon - OT_0 = \underbrace{\int c\gamma_\varepsilon - \int c\gamma_0}_{suboptimality} + \varepsilon \underbrace{H(\gamma_\varepsilon)}_{entropy}$$

Can we disentangle ?

Question 3:

Is there a rate of convergence for $W_2(\gamma_\varepsilon, \gamma_0)$ ?

# Fisher information and quadratic cost

# Main result

**Theorem [MS23]**

Suppose that the cost is quadratic, that is $c(x,y) = \frac{1}{2}|x-y|^2$. Further assume that $I(\mu_i) < \infty$ and $Supp(\mu_i)$ compact. Then

# Main result

**Theorem [MS23]**

Suppose that the cost is quadratic, that is $c(x,y) = \frac{1}{2}|x-y|^2$. Further assume that $I(\mu_i) < \infty$ and $Supp(\mu_i)$ compact. Then

$$H(\gamma_\varepsilon) = -\frac{d}{2}\ln(2\pi\varepsilon) + H_m - \frac{d}{2} + o(1) \tag{15}$$

where $H_m = \frac{H(\mu_0)+H(\mu_1)}{2}$.

# Main result

**Theorem [MS23]**

Suppose that the cost is quadratic, that is $c(x,y) = \frac{1}{2}|x-y|^2$. Further assume that $I(\mu_i) < \infty$ and $Supp(\mu_i)$ compact. Then

$$H(\gamma_\varepsilon) = -\frac{d}{2}\ln(2\pi\varepsilon) + H_m - \frac{d}{2} + o(1) \tag{15}$$

where $H_m = \frac{H(\mu_0)+H(\mu_1)}{2}$. Moreover

$$\int c\gamma_\varepsilon - \int c\gamma_0 = \frac{d}{2}\varepsilon + o(\varepsilon) \tag{16}$$

# Sketch of proof

Recall $H_m = \frac{H(\mu_0)+H(\mu_1)}{2}$. The dynamic formulation [Lé13] is

$$OT_\varepsilon = \varepsilon H_m - \frac{d}{2}\varepsilon \ln(2\pi\varepsilon) + \min_{\substack{\partial\rho+\nabla\cdot(\rho v)=0 \\ \rho_0=\mu_0,\rho_1=\mu_1}} \iint \frac{1}{2}|v_t|^2 \rho_t t + \frac{\varepsilon^2}{8}\int_0^1 I(\rho_t) t \qquad (\varepsilon\text{BB})$$

# Sketch of proof

Recall $H_m = \frac{H(\mu_0)+H(\mu_1)}{2}$. The dynamic formulation [Lé13] is

$$OT_\varepsilon = \varepsilon H_m - \frac{d}{2}\varepsilon \ln(2\pi\varepsilon) + \min_{\substack{\partial\rho+\nabla\cdot(\rho v)=0\\ \rho_0=\mu_0,\rho_1=\mu_1}} \iint \frac{1}{2}|v_t|^2\rho_t t + \frac{\varepsilon^2}{8}\int_0^1 I(\rho_t)t \qquad (\varepsilon\text{BB})$$

Known asymptotics (TE-OT$_\varepsilon$) is

$$OT_\varepsilon - OT_0 = -\frac{d}{2}\varepsilon \ln(2\pi\varepsilon) + \varepsilon H_m + o(\varepsilon) \qquad (\text{TE-OT}_\varepsilon)$$

# Sketch of proof

Recall $H_m = \frac{H(\mu_0)+H(\mu_1)}{2}$. The dynamic formulation [Lé13] is

$$OT_\varepsilon = \varepsilon H_m - \frac{d}{2}\varepsilon \ln(2\pi\varepsilon) + \iint \frac{1}{2}|v_t^\varepsilon|^2 \rho_t^\varepsilon t + \frac{\varepsilon^2}{8}\int_0^1 I(\rho_t^\varepsilon) t \tag{$\varepsilon$BB}$$

Known asymptotics (TE-OT$_\varepsilon$) is

$$OT_\varepsilon - OT_0 + \frac{d}{2}\varepsilon \ln(2\pi\varepsilon) - \varepsilon H_m = o(\varepsilon) \tag{17}$$

# Sketch of proof

Recall $H_m = \frac{H(\mu_0)+H(\mu_1)}{2}$. The dynamic formulation [Lé13] is

$$OT_\varepsilon = \varepsilon H_m - \frac{d}{2}\varepsilon \ln(2\pi\varepsilon) + \iint \frac{1}{2}|v_t^\varepsilon|^2 \rho_t^\varepsilon t + \frac{\varepsilon^2}{8}\int_0^1 I(\rho_t^\varepsilon) t \tag{$\varepsilon$BB}$$

Known asymptotics (TE-OT$_\varepsilon$) is

$$OT_\varepsilon - OT_0 + \frac{d}{2}\varepsilon \ln(2\pi\varepsilon) - \varepsilon H_m = o(\varepsilon) \tag{17}$$

Thus thanks to ($\varepsilon$BB)

$$\frac{1}{\varepsilon}\underbrace{\left(\iint \frac{1}{2}|v_t^\varepsilon|^2 \rho_t^\varepsilon t - OT_0\right)}_{\textit{suboptimality}} + \frac{\varepsilon}{8}\underbrace{\int_0^1 I(\rho_t^\varepsilon) t}_{\textit{regularity term}} = o(1) \tag{18}$$

# Sketch of proof

Recall $H_m = \frac{H(\mu_0)+H(\mu_1)}{2}$. The dynamic formulation [Lé13] is

$$OT_\varepsilon = \varepsilon H_m - \frac{d}{2}\varepsilon \ln(2\pi\varepsilon) + \iint \frac{1}{2}|v_t^\varepsilon|^2 \rho_t^\varepsilon t + \frac{\varepsilon^2}{8}\int_0^1 I(\rho_t^\varepsilon) t \qquad (\varepsilon\text{BB})$$

Known asymptotics (TE-OT$_\varepsilon$) is

$$OT_\varepsilon - OT_0 + \frac{d}{2}\varepsilon \ln(2\pi\varepsilon) - \varepsilon H_m = o(\varepsilon) \qquad (17)$$

Thus thanks to ($\varepsilon$BB)

$$\frac{1}{\varepsilon}\underbrace{\left(\iint \frac{1}{2}|v_t^\varepsilon|^2 \rho_t^\varepsilon t - OT_0\right)}_{suboptimality} + \frac{\varepsilon}{8}\underbrace{\int_0^1 I(\rho_t^\varepsilon) t}_{regularity\ term} = o(1) \qquad (18)$$

Since both terms are positive they both tend to 0.

# Envelop theorem

**Theorem**
*Let $X$ be a set and $f : \mathbb{R} \times X \to \mathbb{R}$ be a function.*
*Let $F : \mathbb{R} \to \mathbb{R}$ be defined by*

$$F(\varepsilon) = \inf_{x \in X} f(\varepsilon, x) \tag{19}$$

*Let $x_\varepsilon$ be a minimizer of $f(\varepsilon, \cdot)$. If $F$ and $f(., x_\varepsilon)$ are differentiable at $\varepsilon$ then there derivative are the same:*

$$F'(\varepsilon) = \frac{\partial f}{\partial \varepsilon}(\varepsilon, x_\varepsilon) \tag{20}$$

# From dynamic to static and back

$$\underbrace{\int c\gamma_\varepsilon + \varepsilon H(\gamma_\varepsilon)}_{(a)\ static} = \underbrace{\varepsilon H_m - \frac{d}{2}\varepsilon \ln(2\pi\varepsilon)}_{(b)} + \underbrace{\iint \frac{1}{2}|v_t^\varepsilon|^2 \rho_t^\varepsilon t + \frac{\varepsilon^2}{8}\int_0^1 I(\rho_t^\varepsilon) t}_{(c)\ dynamic} \qquad (\varepsilon\mathrm{BB})$$

# From dynamic to static and back

$$\underbrace{\int c\gamma_\varepsilon + \varepsilon H(\gamma_\varepsilon)}_{(a)\ static} = \underbrace{\varepsilon H_m - \frac{d}{2}\varepsilon \ln(2\pi\varepsilon)}_{(b)} + \underbrace{\iint \frac{1}{2}|v_t^\varepsilon|^2 \rho_t^\varepsilon t + \frac{\varepsilon^2}{8}\int_0^1 I(\rho_t^\varepsilon) t}_{(c)\ dynamic} \qquad (\varepsilon\text{BB})$$

Envelop theorem

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}\varepsilon}(a) &= \frac{\mathrm{d}}{\mathrm{d}\varepsilon}(b) + \frac{\mathrm{d}}{\mathrm{d}\varepsilon}(c) \\ H(\gamma_\varepsilon) &= H_m - \frac{d}{2}\ln(2\pi\varepsilon) - \frac{d}{2} + \frac{\varepsilon}{4}\int I(\rho_t^\varepsilon) t \end{aligned}$$

# From dynamic to static and back

$$\underbrace{\int c\gamma_\varepsilon + \varepsilon H(\gamma_\varepsilon)}_{(a)\ static} = \underbrace{\varepsilon H_m - \frac{d}{2}\varepsilon \ln(2\pi\varepsilon)}_{(b)} + \underbrace{\iint \frac{1}{2}|v_t^\varepsilon|^2 \rho_t^\varepsilon t + \frac{\varepsilon^2}{8}\int_0^1 I(\rho_t^\varepsilon) t}_{(c)\ dynamic} \tag{$\varepsilon$BB}$$

Envelop theorem

$$\frac{\mathrm{d}}{\mathrm{d}\varepsilon}(a) = \frac{\mathrm{d}}{\mathrm{d}\varepsilon}(b) + \frac{\mathrm{d}}{\mathrm{d}\varepsilon}(c)$$

$$H(\gamma_\varepsilon) = H_m - \frac{d}{2}\ln(2\pi\varepsilon) - \frac{d}{2} + \frac{\varepsilon}{4}\int I(\rho_t^\varepsilon) t$$

$$\begin{cases} \int c\gamma_\varepsilon - OT_0 = \iint \frac{1}{2}|v_t^\varepsilon|^2 \rho_t^\varepsilon t - OT_0 - \frac{\varepsilon^2}{8}\int I(\rho_t^\varepsilon) t + \frac{d}{2}\varepsilon \\ H(\gamma_\varepsilon) = \frac{\varepsilon}{4}\int_0^1 I(\rho_t^\varepsilon) t - \frac{d}{2}\ln(2\pi\varepsilon) + H_m - \frac{d}{2} \end{cases} \tag{21}$$

# Quadratic cost without Fisher information

# Main result

**Theorem [MS23]**

Suppose that the cost is quadratic, that is $c(x, y) = \frac{1}{2}\|x - y\|^2$. Further assume that $\mu_i \in \mathcal{P}_{2+\delta,ac}(\mathbb{R}^d)$ for some $\delta > 0$.

# Main result

**Theorem [MS23]**

Suppose that the cost is quadratic, that is $c(x,y) = \frac{1}{2}\|x-y\|^2$. Further assume that $\mu_i \in \mathcal{P}_{2+\delta,ac}(\mathbb{R}^d)$ for some $\delta > 0$. Then

$$\int c\gamma_\varepsilon - \int c\gamma_0 = \Theta(\varepsilon), \quad H(\gamma_\varepsilon) = -\frac{d}{2}\ln(\varepsilon) + O(1), \tag{22}$$

$$W_2(\gamma_\varepsilon, \gamma_0) \geq C\sqrt{\varepsilon}. \tag{23}$$

# Main result

**Theorem [MS23]**

Suppose that the cost is quadratic, that is $c(x,y) = \frac{1}{2}\|x-y\|^2$. Further assume that $\mu_i \in \mathcal{P}_{2+\delta,ac}(\mathbb{R}^d)$ for some $\delta > 0$ and that the Monge map $\nabla f$ is Lipschitz. Then

$$\int c\gamma_\varepsilon - \int c\gamma_0 = \Theta(\varepsilon), \quad H(\gamma_\varepsilon) = -\frac{d}{2}\ln(\varepsilon) + O(1) \tag{24}$$

$$C_1\sqrt{\varepsilon} \geq W_2(\gamma_\varepsilon, \gamma_0) \geq C_2\sqrt{\varepsilon} \tag{25}$$

# Main result

**Theorem [MS23]**

Suppose that the cost is quadratic, that is $c(x, y) = \frac{1}{2}\|x - y\|^2$. Further assume that $\mu_i \in \mathcal{P}_{2+\delta,ac}(\mathbb{R}^d)$ for some $\delta > 0$ and that the Monge map $\nabla f$ is Lipschitz.
Then

$$\int c\gamma_\varepsilon - \int c\gamma_0 = \Theta(\varepsilon), \quad H(\gamma_\varepsilon) = -\frac{d}{2}\ln(\varepsilon) + O(1) \tag{24}$$

$$C_1\sqrt{\varepsilon} \geq W_2(\gamma_\varepsilon, \gamma_0) \geq C_2\sqrt{\varepsilon} \tag{25}$$

Key idea :
"If $W_2(\gamma, \gamma_0)$ is small, then $H(\gamma)$ explodes"

# Main result

**Theorem [MS23]**

Suppose that the cost is quadratic, that is $c(x,y) = \frac{1}{2}\|x-y\|^2$. Further assume that $\mu_i \in \mathcal{P}_{2+\delta,ac}(\mathbb{R}^d)$ for some $\delta > 0$ and that the Monge map $\nabla f$ is Lipschitz.

Then

$$\int c\gamma_\varepsilon - \int c\gamma_0 = \Theta(\varepsilon), \quad H(\gamma_\varepsilon) = -\frac{d}{2}\ln(\varepsilon) + O(1) \tag{24}$$

$$C_1\sqrt{\varepsilon} \geq W_2(\gamma_\varepsilon, \gamma_0) \geq C_2\sqrt{\varepsilon} \tag{25}$$

Key idea :

"If $W_2(\gamma, \gamma_0)$ is small, then $H(\gamma)$ explodes"

"If $\int c\gamma - \int c\gamma_0$ is small, then $H(\gamma)$ explodes"

# Key idea

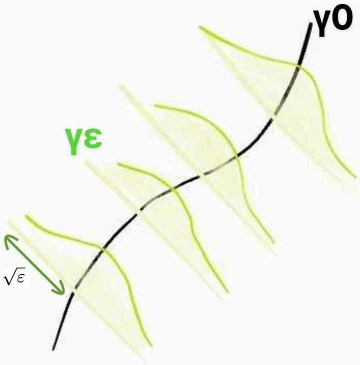


**Figure 8:** The best to do is to convolve $\gamma_0$ with a Gaussian in the transverse direction

# Minty's coordinates

Minty's coordinates $(u, v)$ are defined by

$$u = \frac{x + y}{\sqrt{2}}, \quad v = \frac{y - x}{\sqrt{2}} \tag{26}$$

# Minty's coordinates

Minty's coordinates $(u, v)$ are defined by

$$u = \frac{x + y}{\sqrt{2}}, \quad v = \frac{y - x}{\sqrt{2}} \tag{26}$$

Let $G \subset \mathbb{R}^{2d}$ be the graph of a function $y = T(x)$ in the classical coordinates and of a function $v = S(u)$ in the Minty's coordinates.

# Minty's coordinates

Minty's coordinates $(u, v)$ are defined by

$$u = \frac{x+y}{\sqrt{2}}, \quad v = \frac{y-x}{\sqrt{2}} \tag{26}$$

Let $G \subset \mathbb{R}^{2d}$ be the graph of a function $y = T(x)$ in the classical coordinates and of a function $v = S(u)$ in the Minty's coordinates.
Then

$$T = \nabla f \text{with f convex} \Leftrightarrow S \text{ 1-Lipschitz}$$

# Minty's coordinates

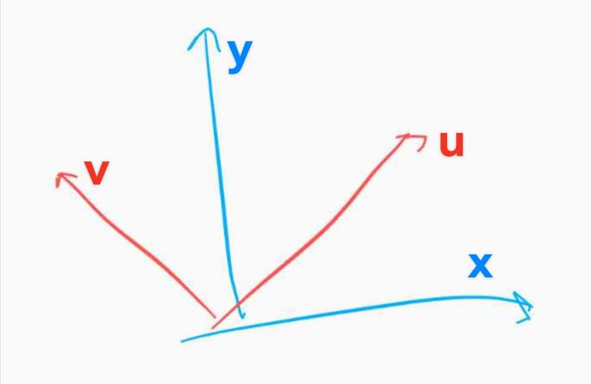


**Figure 9:** Minty's coordinates in 1d

# Minty's coordinates

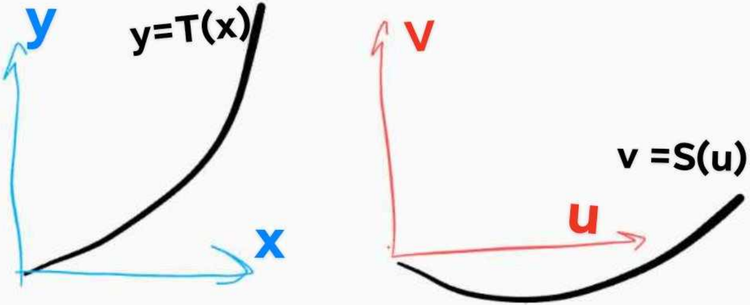


**Figure 10:** Minty's coordinates in 1d

# Linearity of the entropy

Let $\gamma \in \mathcal{P}_{2,ac}(\mathbb{R}^{2d})$ be disentegrated in Minty's coordinates as

$$\gamma(u, v) = \gamma^u(v)\hat{\gamma}(u). \tag{27}$$

# Linearity of the entropy

Let $\gamma \in \mathcal{P}_{2,ac}(\mathbb{R}^{2d})$ be disentegrated in Minty's coordinates as

$$\gamma(u, v) = \gamma^u(v)\hat{\gamma}(u). \tag{27}$$

Then, by linearity of the entropy,

$$H(\gamma) = H(\hat{\gamma}) + \int H(\gamma^u)\mathrm{d}\hat{\gamma}(u) \tag{28}$$

# Linearity of the entropy

Let $\gamma \in \mathcal{P}_{2,ac}(\mathbb{R}^{2d})$ be disentegrated in Minty's coordinates as

$$\gamma(u, v) = \gamma^u(v)\hat{\gamma}(u). \tag{27}$$

Then, by linearity of the entropy,

$$H(\gamma) = H(\hat{\gamma}) + \int H(\gamma^u)\mathrm{d}\hat{\gamma}(u) \tag{28}$$

Question: how to choose $\gamma^u$ in order to minimize $H(\gamma)$ if $\hat{\gamma}$ and $\int |v - S(u)|^2\gamma(u, v)\mathrm{d}u\mathrm{d}v$ are fixed?

# Linearity of the entropy

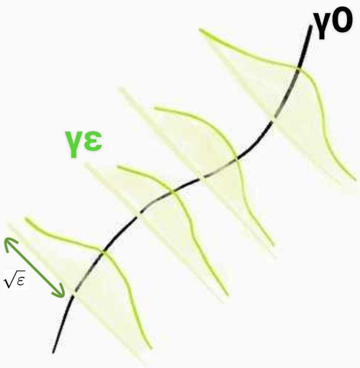


**Figure 11:** The best to do is to is to choose $\gamma^u$ to be a Gaussian

# Linearity of the entropy

By linearity of the entropy,

$$H(\gamma) = H(\hat{\gamma}) + \int H(\gamma^u)\mathrm{d}\hat{\gamma}(u) \tag{29}$$

Question: how to choose $\gamma^u$ in order to minimize $H(\gamma)$ if $\hat{\gamma}$ and $\int |v - S(u)|^2\gamma(u,v)\mathrm{d}u\mathrm{d}v$ are fixed?

## Linearity of the entropy

By linearity of the entropy,

$$H(\gamma) = H(\hat{\gamma}) + \int H(\gamma^u) \mathrm{d}\hat{\gamma}(u) \tag{29}$$

Question: how to choose $\gamma^u$ in order to minimize $H(\gamma)$ if $\hat{\gamma}$ and $\int |v - S(u)|^2 \gamma(u, v) \mathrm{d}u \mathrm{d}v$ are fixed?

Answer: the best to do is to choose $\gamma^u$ to be a Gaussian, with center $S(u)$.

# Linearity of the entropy

By linearity of the entropy,

$$H(\gamma) = H(\hat{\gamma}) + \int H(\gamma^u) d\hat{\gamma}(u) \tag{29}$$

Question: how to choose $\gamma^u$ in order to minimize $H(\gamma)$ if $\hat{\gamma}$ and $\int |v - S(u)|^2 \gamma(u, v) du dv$ are fixed?

Answer: the best to do is to choose $\gamma^u$ to be a Gaussian, with center $S(u)$.

**Proposition**

Let $C_d = -d \ln(\frac{2\pi e}{d})$. Then

$$\int H(\gamma^u) d\hat{\gamma}(u) \geq -\frac{d}{2} \ln(\int |v - S(u)|^2 d\gamma(u, v)) + C_d \tag{30}$$

# Linearity of the entropy

By linearity of the entropy,

$$H(\gamma) = H(\hat{\gamma}) + \int H(\gamma^u) \mathrm{d}\hat{\gamma}(u) \tag{29}$$

Question: how to choose $\gamma^u$ in order to minimize $H(\gamma)$ if $\hat{\gamma}$ and $\int |v - S(u)|^2 \gamma(u, v) \mathrm{d}u \mathrm{d}v$ are fixed?

Answer: the best to do is to choose $\gamma^u$ to be a Gaussian, with center $S(u)$.

**Proposition**

Let $C_d = -d \ln(\frac{2\pi e}{d})$. Then

$$\int H(\gamma^u) \mathrm{d}\hat{\gamma}(u) \geq -\frac{d}{2} \ln(\int |v - S(u)|^2 \mathrm{d}\gamma(u, v)) + C_d \tag{30}$$

Question: What do we do with this term $\int |v - S(u)|^2 \mathrm{d}\gamma(u, v)$?

# Wasserstein distance and Lipschitz function

Since $\gamma_0$ is supported by the graph of S, 1-Lipschitz, $\int |v - S(u)|^2 \mathrm{d}\gamma(u, v)$ is bounded by the Wassertein distance to $\gamma_0$.

# Wasserstein distance and Lipschitz function

Since $\gamma_0$ is supported by the graph of S, 1-Lipschitz, $\int |v - S(u)|^2 \mathrm{d}\gamma(u,v)$ is bounded by the Wassertein distance to $\gamma_0$.

**Proposition**

Let $\gamma, \gamma_0 \in \mathcal{P}_2(\mathbb{R}^{2d})$ be two plans.
If $\gamma_0$ is supported by the graph of a 1-Lipschitz function $S$, then

$$W_2^2(\gamma, \gamma_0) \geq \int \frac{|v - S(u)|^2}{2} \mathrm{d}\gamma(u,v) \tag{31}$$

# Wasserstein distance and Lipschitz function

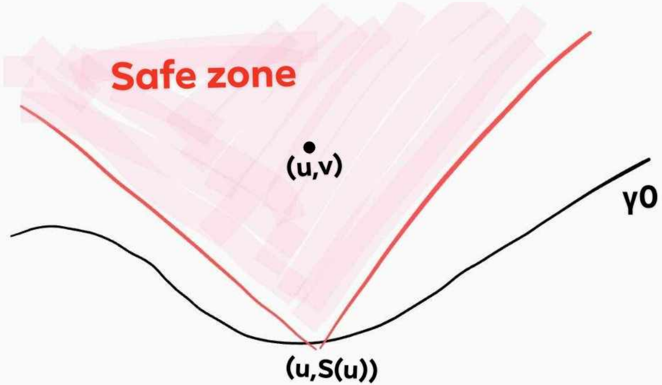


**Figure 12:** There is nothing much better to do than going down

# Quadratic detachment

## Minty's trick

For the quadratic cost, the energy gap has a quadratic detachment in Minty's coordinates:

$$\forall u, v, \quad E(u,v) \geq \frac{|v - S(u)|^2}{2} \tag{32}$$

Consequence: if $\gamma \in \mathcal{P}_{2,ac}(\mathbb{R}^{2d})$, then

$$\int E\gamma \geq \int \frac{|v - S(u)|^2}{2} \gamma(u,v) \mathrm{d}u\mathrm{d}v \tag{33}$$

# Quadratic detachment

## Minty's trick

For the quadratic cost, the energy gap has a quadratic detachment in Minty's coordinates:

$$\forall u, v, \quad E(u, v) \geq \frac{|v - S(u)|^2}{2} \tag{32}$$

Consequence: if $\gamma \in \mathcal{P}_{2,ac}(\mathbb{R}^{2d})$, then

$$\int E\gamma \geq \int \frac{|v - S(u)|^2}{2} \gamma(u, v) \mathrm{d}u\mathrm{d}v \tag{33}$$

# Quadratic detachment

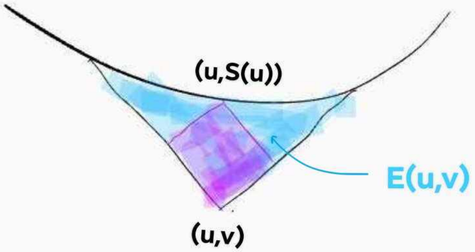


**Figure 13:** Quadratic detachment in 1d

# Bound the transverse distance

Question: What do we do with this term $\int |v - S(u)|^2 \mathrm{d}\gamma(u, v)$?

Answer: we can bound it with meaningful quantities

# Bound the transverse distance

Question: What do we do with this term $\int |v - S(u)|^2 \mathrm{d}\gamma(u, v)$?

Answer: we can bound it with meaningful quantities

**Proposition**

If $\gamma \in \mathcal{P}_2(\mathbb{R}^{2d})$, then

$$\int |v - S(u)|^2 \mathrm{d}\gamma(u, v) \leq 2W_2^2(\gamma, \gamma_0)$$

$$\int |v - S(u)|^2 \mathrm{d}\gamma(u, v) \leq 2 \int E\gamma$$

Remark that if $\gamma \in \Pi(\mu, \nu)$,

$$\int E\gamma = \int c\gamma - \int c\gamma_0.$$

## idea of proof

We had

$$H(\gamma) \geq -\frac{d}{2} \ln \left( \int |v - S(u)|^2 \mathrm{d}\gamma(u, v) \right) + C, \tag{34}$$

## idea of proof

We had

$$H(\gamma) \geq -\frac{d}{2} \ln \left( \int |v - S(u)|^2 \mathrm{d}\gamma(u, v) \right) + C, \tag{34}$$

so we get

$$H(\gamma_\varepsilon) \gtrsim -\frac{d}{2} \ln W_2^2(\gamma_\varepsilon, \gamma_0) \quad \text{and} \quad H(\gamma_\varepsilon) \gtrsim -\frac{d}{2} \ln \left( \int c \mathrm{d}\gamma_\varepsilon - \int c \mathrm{d}\gamma_0 \right)$$

# idea of proof

We had

$$H(\gamma) \geq -\frac{d}{2} \ln \left( \int |v - S(u)|^2 \mathrm{d}\gamma(u, v) \right) + C, \tag{34}$$

so we get

$$H(\gamma_\varepsilon) \gtrsim -\frac{d}{2} \ln W_2^2(\gamma_\varepsilon, \gamma_0) \quad \text{and} \quad H(\gamma_\varepsilon) \gtrsim -\frac{d}{2} \ln \left( \int c \mathrm{d}\gamma_\varepsilon - \int c \mathrm{d}\gamma_0 \right)$$

We are going to combine these inequalities with the existing litterature,

$$-\frac{d}{2} \varepsilon \ln(\varepsilon) \gtrsim OT_\varepsilon - OT_0$$

## idea of proof

We had

$$H(\gamma) \geq -\frac{d}{2} \ln\left(\int |v - S(u)|^2 \mathrm{d}\gamma(u,v)\right) + C, \tag{35}$$

so we get

$$H(\gamma_\varepsilon) \gtrsim -\frac{d}{2} \ln W_2^2(\gamma_\varepsilon, \gamma_0) \quad \text{and} \quad H(\gamma_\varepsilon) \gtrsim -\frac{d}{2} \ln\left(\int c \mathrm{d}\gamma_\varepsilon - \int c \mathrm{d}\gamma_0\right)$$

We are going to combine these inequalities with the existing litterature,

$$-\frac{d}{2} \varepsilon \ln(\varepsilon) \gtrsim \int c \mathrm{d}\gamma_\varepsilon - \int c \mathrm{d}\gamma_0 + \varepsilon H(\gamma_\varepsilon)$$

# Wasserstein distance

Our result:

$$H(\gamma_\varepsilon) \gtrsim -\frac{d}{2} \ln W_2^2(\gamma_\varepsilon, \gamma_0)$$

Existing litterature:

$$-\frac{d}{2}\varepsilon \ln(\varepsilon) \gtrsim \varepsilon H(\gamma_\varepsilon)$$

# Wasserstein distance

Our result:

$$H(\gamma_\varepsilon) \gtrsim -\frac{d}{2} \ln W_2^2(\gamma_\varepsilon, \gamma_0)$$

Existing litterature:

$$-\frac{d}{2}\varepsilon \ln(\varepsilon) \gtrsim \varepsilon H(\gamma_\varepsilon)$$

Combining both,

$$W_2^2(\gamma_\varepsilon, \gamma_0) \gtrsim \varepsilon.$$

# Suboptimality

Our result:

$$H(\gamma_\varepsilon) \gtrsim -\frac{d}{2}\ln\left(\int c\gamma_\varepsilon - \int c\gamma_0\right)$$

Existing litterature:

$$-\frac{d}{2}\varepsilon\ln(\varepsilon) + C\varepsilon \geq \int c\gamma_\varepsilon - \int c\gamma_0 + \varepsilon H(\gamma_\varepsilon)$$

## Suboptimality

Our result:

$$H(\gamma_\varepsilon) \gtrsim -\frac{d}{2}\ln(\int c\gamma_\varepsilon - \int c\gamma_0)$$

Existing litterature:

$$-\frac{d}{2}\varepsilon\ln(\varepsilon) + C\varepsilon \geq \int c\gamma_\varepsilon - \int c\gamma_0 + \varepsilon H(\gamma_\varepsilon)$$

Combining both,

$$C \geq \frac{\int c\gamma_\varepsilon - \int c\gamma_0}{\varepsilon} - \frac{d}{2}\ln\left(\frac{\int c\gamma_\varepsilon - \int c\gamma_0}{\varepsilon}\right) \tag{36}$$

the map $x \mapsto x - \frac{d}{2}\ln(x)$ is coercive, so

$$C_1\varepsilon \leq \int c\gamma_\varepsilon - \int c\gamma_0 \leq C_2\varepsilon$$

# Main result

**Theorem [MS23]**

Suppose that the cost is quadratic, that is $c(x, y) = \frac{1}{2}\|x - y\|^2$. Further assume that $\mu_i \in \mathcal{P}_{2+\delta,ac}(\mathbb{R}^d)$ for some $\delta > 0$.

# Main result

**Theorem [MS23]**

Suppose that the cost is quadratic, that is $c(x, y) = \frac{1}{2}\|x - y\|^2$. Further assume that $\mu_i \in \mathcal{P}_{2+\delta,ac}(\mathbb{R}^d)$ for some $\delta > 0$. Then

$$\int c\gamma_\varepsilon - \int c\gamma_0 = \Theta(\varepsilon), \quad H(\gamma_\varepsilon) = -\frac{d}{2}\ln(\varepsilon) + O(1), \tag{37}$$

$$W_2(\gamma_\varepsilon, \gamma_0) \geq C\sqrt{\varepsilon}. \tag{38}$$

# Maps

**Theorem [MS23]**

Suppose that the cost is quadratic, that is $c(x, y) = \frac{1}{2}\|x - y\|^2$. Further assume that $\mu_i \in \mathcal{P}_{2+\delta,ac}$ for some $\delta > 0$ and that the Monge map $\nabla f$ is Lipschitz. Then

$$C_1\sqrt{\varepsilon} \geq W_2(\gamma_\varepsilon, \gamma_0) \geq C_2\sqrt{\varepsilon}, \tag{39}$$

**Theorem [MS23]**

Suppose that the cost is quadratic, that is $c(x, y) = \frac{1}{2}\|x - y\|^2$. Further assume that $\mu_i \in \mathcal{P}_{2+\delta,ac}$ for some $\delta > 0$ and that the Monge map $\nabla f$ is Lipschitz. Then

$$C_1\sqrt{\varepsilon} \geq W_2(\gamma_\varepsilon, \gamma_0) \geq C_2\sqrt{\varepsilon}, \tag{39}$$

and

$$\|\nabla f_\varepsilon - \nabla f\|^2_{L^2(\mu_0)} \leq C\varepsilon \tag{40}$$

Where $f_\varepsilon$ is the Schrödinger potential and $\nabla f_\varepsilon$ is the barycentric map.

**Theorem [MS23]**

Suppose that the cost is quadratic, that is $c(x,y) = \frac{1}{2}\|x-y\|^2$. Further assume that $\mu_i \in \mathcal{P}_{2+\delta,ac}$ for some $\delta > 0$ and that the Monge map $\nabla f$ is Lipschitz. Then

$$C_1\sqrt{\varepsilon} \geq W_2(\gamma_\varepsilon, \gamma_0) \geq C_2\sqrt{\varepsilon}, \tag{39}$$

and

$$\|\nabla f_\varepsilon - \nabla f\|^2_{L^2(\mu_0)} \leq C\varepsilon \tag{40}$$

Where $f_\varepsilon$ is the Schrödinger potential and $\nabla f_\varepsilon$ is the barycentric map. $\nabla f_\varepsilon$is the barycentric map means

$$\nabla f_\varepsilon(x) = \int y \mathrm{d}\gamma_\varepsilon^x(y)$$

# Idea

We got interested in $\int |v - S(u)|^2 \mathrm{d}\gamma_\varepsilon(u, v)$, we can also be interested in $\int |y - T(x)|^2 \mathrm{d}\gamma_\varepsilon(x, y)$. Two good reasons:

# Idea

We got interested in $\int |v - S(u)|^2 \mathrm{d}\gamma_\varepsilon(u, v)$, we can also be interested in $\int |y - T(x)|^2 \mathrm{d}\gamma_\varepsilon(x, y)$. Two good reasons:

- by Jensen inequality

$$\int \left| \int y \mathrm{d}\gamma_\varepsilon^x(y) - T(x) \right|^2 \mathrm{d}\hat{\gamma}_\varepsilon(x) \leq \int |y - T(x)|^2 \mathrm{d}\gamma_\varepsilon(x, y)$$

- $(x, y) \mapsto x, T(x)$ is a transport map between $\gamma_\varepsilon$ and $\gamma_0$, so

$$W_2^2(\gamma_\varepsilon, \gamma_0) \leq \int |y - T(x)|^2 \mathrm{d}\gamma_\varepsilon(x, y)$$

# Idea

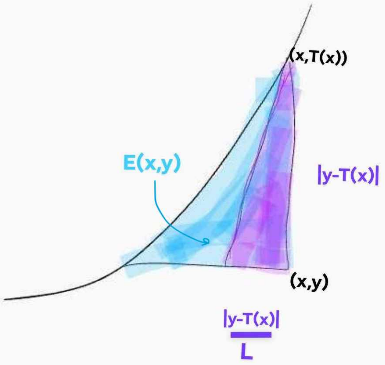


**Figure 14:** The energy gap $E(x, y)$ is bounded by $\frac{|y - T(x)|^2}{2L}$

# Idea

$$E(x, y) \leq \frac{|y - T(x)|^2}{2L}$$

## Idea

$$E(x, y) \leq \frac{|y - T(x)|^2}{2L}$$

So

$$\int |y - T(x)|^2 \mathrm{d}\gamma_\varepsilon(x, y) \leq 2L \int E \mathrm{d}\gamma_\varepsilon$$

# Idea

$$E(x,y) \le \frac{|y - T(x)|^2}{2L}$$

So

$$\int |y - T(x)|^2 \mathrm{d}\gamma_\varepsilon(x,y) \le 2L \int E \mathrm{d}\gamma_\varepsilon$$

**Proposition**

If $\nabla f$ is L-lipschitz, then

$$\int |y - T(x)|^2 \mathrm{d}\gamma_\varepsilon(x,y) \le 2L \left( \int c\gamma_\varepsilon - \int c\gamma_0 \right).$$

**Theorem [MS23]**

Suppose that the cost is quadratic, that is $c(x, y) = \frac{1}{2}\|x - y\|^2$. Further assume that $\mu_i \in \mathcal{P}_{2+\delta,ac}$ for some $\delta > 0$ and that the Monge map $\nabla f$ is Lipschitz. Then

$$C_1\sqrt{\varepsilon} \geq W_2(\gamma_\varepsilon, \gamma_0) \geq C_2\sqrt{\varepsilon}, \tag{41}$$

and

$$\|\nabla f_\varepsilon - \nabla f\|^2_{L^2(\mu_0)} \leq C\varepsilon \tag{42}$$

Where $f_\varepsilon$ is the Schrödinger potential and $\nabla f_\varepsilon$ is the barycentric map.

# Infinitesimally twisted costs and compact supports

# Main result

**Definition**

$c \in \mathcal{C}^2(\Omega^2)$ is said to be infinitesimally twisted if $\nabla^2_{xy} c(x,y) = (\partial^2_{x_i y_j} c(x,y))_{i,j} \in M_d(\mathbb{R})$ is invertible for every $(x,y) \in \Omega^2$.

**Theorem**

Suppose that the cost is $\mathcal{C}^2$ and infinitesimally twisted . Further assume that $\mu_i$ is compactly supported then

$$(c, \gamma_\varepsilon) = OT_0 + \Theta(\varepsilon), \quad H(\gamma_\varepsilon \mid \mathcal{H}^{2d}) = -\frac{d}{2}\ln(\varepsilon) + O(1), \quad \sqrt{\varepsilon} = O(W_2(\gamma_\varepsilon, \gamma_0)) \tag{43}$$

Note that here $\gamma_0$ is any optimal transport plan.

# Thank you !

# References i


Stefan Adams, Nicolas Dirr, Mark A. Peletier, and Johannes Zimmer.
**From a large-deviations principle to the wasserstein gradient flow: A new micro-macro passage.**
Communications in Mathematical Physics, 307(3):791–815, September 2011.

Jason M. Altschuler, Jonathan Niles-Weed, and Austin J. Stromme.
**Asymptotics for semidiscrete entropic optimal transport.**
SIAM Journal on Mathematical Analysis, 54(2):1718–1741, mar 2021.

Guillaume Carlier, Vincent Duval, Gabriel Peyré, and Bernhard Schmitzer.
**Convergence of entropic schemes for optimal transport and gradient flows, 2015.**

- Guillaume Carlier, Vincent Duval, Gabriel Peyré, and Bernhard Schmitzer.
  **Convergence of entropic schemes for optimal transport and gradient flows, 2017.**
- R. Cominetti and J. San Martin.
  **Asymptotic analysis of the exponential penalty trajectory in linear programming.**
  Mathematical Programming, 67(1-3):169–187, October 1994.
- Giovanni Conforti.
  **A second order equation for schrödinger bridges with applications to the hot gas experiment and entropic transportation cost.**
  Probability Theory and Related Fields, 174(1-2):1–47, 2019.

# References iii


Guillaume Carlier, Paul Pegon, and Luca Tamanini.
**Convergence rate of general entropic optimal transport costs.**
arXiv preprint arXiv:2206.03347, 2022.

Lenaic Chizat, Pierre Roussillon, Flavien Léger, François-Xavier Vialard, and Gabriel Peyré.
**Faster wasserstein distance estimation with the sinkhorn divergence.**
Advances in Neural Information Processing Systems, 33:2257–2269, 2020.

Alex Delalande.
**Nearly tight convergence bounds for semi-discrete entropic optimal transport, 2021.**

- Matthias Erbar, Jan Maas, and Michiel Renger.
  **From large deviations to wasserstein gradient flows in multiple dimensions.**
  Electronic Communications in Probability, 20:1–12, 2015.
- Stephan Eckstein and Marcel Nutz.
  **Convergence rates for regularized optimal transport via quantization, 2022.**
- Jean Feydy, Thibault Séjourné, François-Xavier Vialard, Shun ichi Amari, Alain Trouvé, and Gabriel Peyré.
  **Interpolating between optimal transport and mmd using sinkhorn divergences, 2018.**

# References v


Christian Léonard.
**A survey of the schrödinger problem and some of its connections with optimal transport, 2013.**

Toshio Mikami.
**Monge's problem with a quadratic cost by the zero-noise limit of h-path processes.**
Probability Theory and Related Fields, 129(2):245–260, March 2004.

Hugo Malamut and Maxime Sylvestre.
**Convergence rates of the regularized optimal transport: Disentangling suboptimality and entropy.**
arXiv preprint arXiv:2306.06940, 2023.

Toshio Mikami and Michèle Thieullen.
**Optimal transportation problem by stochastic optimal control.**
SIAM Journal on Control and Optimization, 47(3):1127–1139, January 2008.

Luca Nenna and Paul Pegon.
**Convergence rate of entropy-regularized multi-marginal optimal transport costs, 2023.**

Soumik Pal.
**On the difference between entropic cost and the optimal transport cost, 2019.**

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