Disentangling entropy and suboptimality in Entropic optimal transport

arXiv:2306.06940

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

May 27, 2024

The regularized optimal transport

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

$$\underbrace{H(\rho) := \int_{\mathbb{R}^k} \rho(x) \ln \rho(x) \mathrm{d}x}_{\textit{differential entropy}} \quad \text{and} \quad \underbrace{I(\rho) := \int_{\mathbb{R}^k} \rho(x) \|\nabla \ln \rho(x)\|^2 \mathrm{d}x}_{\textit{Fisher information}} \tag{1}$$

The regularized optimal transport

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

$$\underbrace{H(\rho) := \int_{\mathbb{R}^k} \rho(x) \ln \rho(x) \mathrm{d}x}_{\textit{differential entropy}} \quad \text{and} \quad \underbrace{I(\rho) := \int_{\mathbb{R}^k} \rho(x) \|\nabla \ln \rho(x)\|^2 \mathrm{d}x}_{\textit{Fisher information}} \tag{1}$$

The regularized optimal transport

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

$$\underbrace{H(\rho) := \int_{\mathbb{R}^k} \rho(x) \ln \rho(x) \mathrm{d}x}_{\textit{differential entropy}} \quad \text{and} \quad \underbrace{I(\rho) := \int_{\mathbb{R}^k} \rho(x) \|\nabla \ln \rho(x)\|^2 \mathrm{d}x}_{\textit{Fisher information}} \tag{1}$$

Let c be a C^2 cost function.

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

The regularized optimal transport

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

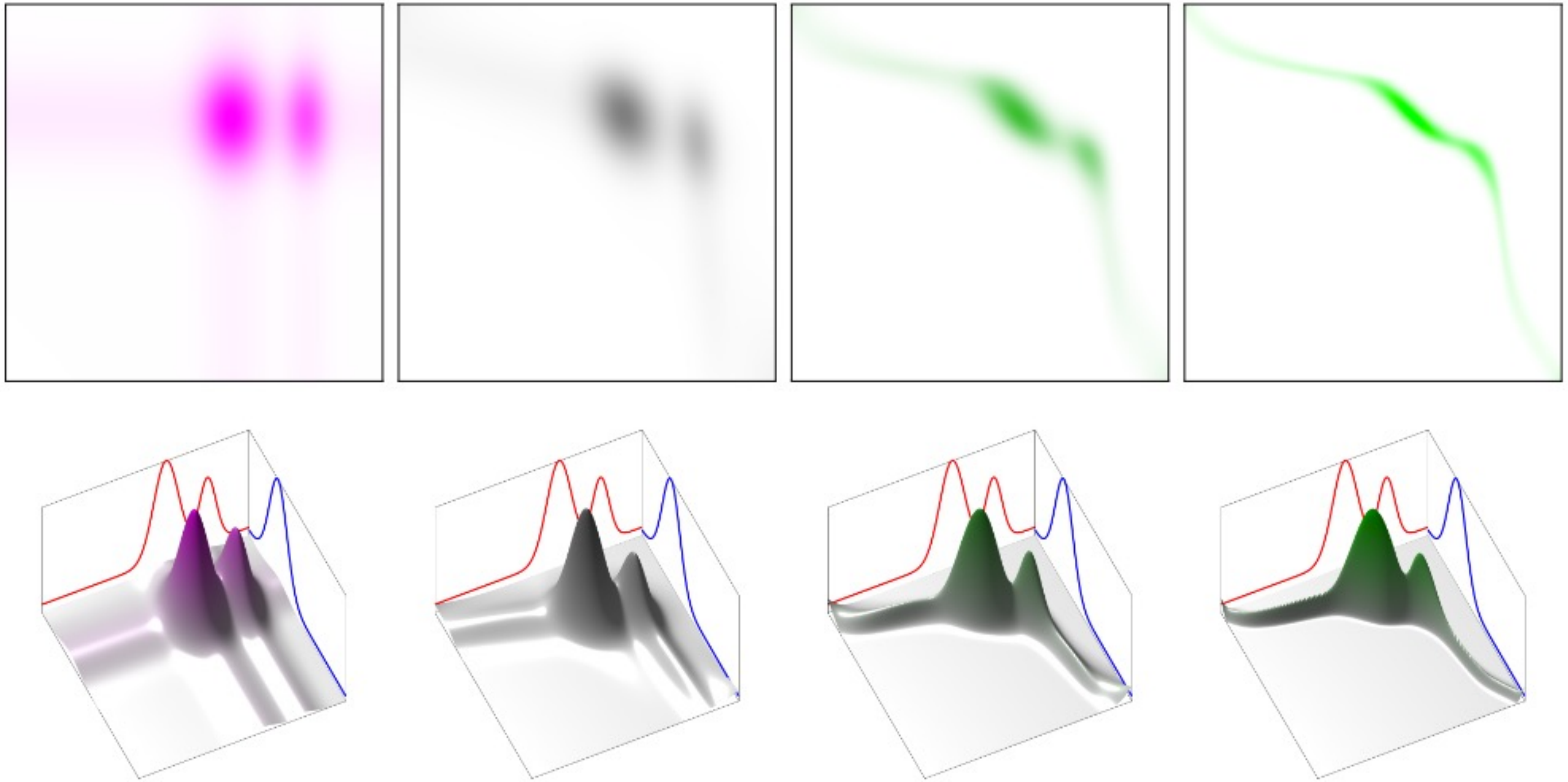
$$\underbrace{H(\rho) := \int_{\mathbb{R}^k} \rho(x) \ln \rho(x) \mathrm{d}x}_{\textit{differential entropy}} \quad \text{and} \quad \underbrace{I(\rho) := \int_{\mathbb{R}^k} \rho(x) \|\nabla \ln \rho(x)\|^2 \mathrm{d}x}_{\textit{Fisher information}} \tag{1}$$

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 $\varepsilon \geq 0$

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



$\varepsilon = 10$ $\varepsilon = 1$ $\varepsilon = 10^{-1}$ $\varepsilon = 10^{-2}$

Peyré and Cuturi (2018)

The regularized optimal transport

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

$$\underbrace{H(\rho) := \int_{\mathbb{R}^k} \rho(x) \ln \rho(x) \mathrm{d}x}_{\textit{differential entropy}} \quad \text{and} \quad \underbrace{I(\rho) := \int_{\mathbb{R}^k} \rho(x) \|\nabla \ln \rho(x)\|^2 \mathrm{d}x}_{\textit{Fisher information}} \tag{1}$$

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 $\varepsilon \geq 0$

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

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

Prior Works

Qualitative convergence results.

- Γ-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

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

Proposition [EN22, CPT22]

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

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

Questions

Question 1 :

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

Can we disentangle ?

Questions

Question 1 :

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

Can we disentangle ?

Question 2 :

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

$$H(\gamma_\varepsilon) = -\frac{d}{2}\ln(2\pi\varepsilon) + H_m - \frac{d}{2} + o(1) \qquad (3)$$

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

$$H(\gamma_\varepsilon) = -\frac{d}{2}\ln(2\pi\varepsilon) + H_m - \frac{d}{2} + o(1) \tag{3}$$

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

$$\int c d\gamma_\varepsilon - \int c d\gamma_0 = \frac{d}{2}\varepsilon + o(\varepsilon) \tag{4}$$

Sketch of proof

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

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

Sketch of proof

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

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

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

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

Known asymptotics (TE-OT_ε) is

$$OT_\varepsilon - OT_0 + \frac{d}{2}\varepsilon \ln(2\pi\varepsilon) - \varepsilon H_m = o(\varepsilon) \qquad (5)$$

Sketch of proof

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

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

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

$$OT_\varepsilon - OT_0 + \frac{d}{2}\varepsilon \ln(2\pi\varepsilon) - \varepsilon H_m = o(\varepsilon) \qquad (5)$$

Thus thanks to (εBB)

$$\underbrace{\frac{1}{\varepsilon}\left(\iint \frac{1}{2}|v_t^\varepsilon|^2 \mathrm{d}\rho_t^\varepsilon \mathrm{d}t - OT_0\right)}_{\textit{suboptimality}} + \frac{\varepsilon}{8}\underbrace{\int_0^1 I(\rho_t^\varepsilon)\mathrm{d}t}_{\textit{regularity term}} = o(1) \qquad (6)$$

Sketch of proof

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

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

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

$$OT_\varepsilon - OT_0 + \frac{d}{2}\varepsilon \ln(2\pi\varepsilon) - \varepsilon H_m = o(\varepsilon) \tag{5}$$

Thus thanks to (εBB)

$$\frac{1}{\varepsilon}\underbrace{\left(\iint \frac{1}{2}|v_t^\varepsilon|^2 \mathrm{d}\rho_t^\varepsilon \mathrm{d}t - OT_0\right)}_{\textit{suboptimality}} + \frac{\varepsilon}{8}\underbrace{\int_0^1 I(\rho_t^\varepsilon)\mathrm{d}t}_{\textit{regularity term}} = o(1) \tag{6}$$

Since both terms are positive they both tend to 0.

From dynamic to static and back

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

From dynamic to static and back

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

Envelop theorem

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

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

From dynamic to static and back

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

Envelop theorem

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

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

$$\begin{cases} \int c\mathrm{d}\gamma_\varepsilon - OT_0 = \iint \frac{1}{2}|v_t^\varepsilon|^2 \mathrm{d}\rho_t^\varepsilon \mathrm{d}t - OT_0 - \frac{\varepsilon^2}{8}\int I(\rho_t^\varepsilon)\mathrm{d}t + \frac{d}{2}\varepsilon \\ H(\gamma_\varepsilon) = \frac{\varepsilon}{4}\int_0^1 I(\rho_t^\varepsilon)\mathrm{d}t - \frac{d}{2}\ln(2\pi\varepsilon) + H_m - \frac{d}{2} \end{cases} \qquad (7)$$

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 $Supp(\mu_i)$ compact.

Main result

Theorem [MS23]

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

$$\int c \mathrm{d}\gamma_\varepsilon - \int c \mathrm{d}\gamma_0 = \Theta(\varepsilon), \quad H(\gamma_\varepsilon) = -\frac{d}{2}\ln(\varepsilon) + O(1), \tag{8}$$

$$W_2(\gamma_\varepsilon, \gamma_0) \geq C\sqrt{\varepsilon}. \tag{9}$$

Main result

Theorem [MS23]

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

$$\int c \mathrm{d}\gamma_\varepsilon - \int c \mathrm{d}\gamma_0 = \Theta(\varepsilon), \quad H(\gamma_\varepsilon) = -\frac{d}{2}\ln(\varepsilon) + O(1) \tag{10}$$

$$C_1\sqrt{\varepsilon} \geq W_2(\gamma_\varepsilon, \gamma_0) \geq C_2\sqrt{\varepsilon} \tag{11}$$

Main result

Theorem [MS23]

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

$$\int c d\gamma_\varepsilon - \int c d\gamma_0 = \Theta(\varepsilon), \quad H(\gamma_\varepsilon) = -\frac{d}{2}\ln(\varepsilon) + O(1) \tag{10}$$

$$C_1\sqrt{\varepsilon} \geq W_2(\gamma_\varepsilon, \gamma_0) \geq C_2\sqrt{\varepsilon} \tag{11}$$

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 $\mu_i \in \mathcal{P}_{2+\delta,ac}$ for some $\delta > 0$ and that the Monge map ∇f is Lipschitz. Then

$$\int c d\gamma_\varepsilon - \int c d\gamma_0 = \Theta(\varepsilon), \quad H(\gamma_\varepsilon) = -\frac{d}{2}\ln(\varepsilon) + O(1) \tag{10}$$

$$C_1\sqrt{\varepsilon} \geq W_2(\gamma_\varepsilon, \gamma_0) \geq C_2\sqrt{\varepsilon} \tag{11}$$

Key idea :

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

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

idea of proof

"Normal distributions minimize the entropy at fixed variance"

idea of proof

"Normal distributions minimize the entropy at fixed variance"

For $\gamma \in \mathcal{P}_{2,ac}(\mathbb{R}^{2d})$,

$$H(\gamma) \geq H(\mathcal{N}(\mathbb{E}(\gamma), Var(\gamma))),$$

idea of proof

"Normal distributions minimize the entropy at fixed variance"

For $\gamma \in \mathcal{P}_{2,ac}(\mathbb{R}^{2d})$,

$$H(\gamma) \geq H(\mathcal{N}(\mathbb{E}(\gamma), Var(\gamma))),$$

$$H(\gamma) \geq -d \ln(Var(\gamma)) - d \ln(\frac{2\pi e}{d}).$$

idea of proof

"Normal distributions minimize the entropy at fixed variance"

For $\gamma \in \mathcal{P}_{2,ac}(\mathbb{R}^{2d})$,

$$H(\gamma) \geq H(\mathcal{N}(\mathbb{E}(\gamma), Var(\gamma))),$$

$$H(\gamma) \geq -d \ln(Var(\gamma)) + C_d.$$

idea of proof

"Normal distributions minimize the entropy at fixed variance"

For $\gamma \in \mathcal{P}_{2,ac}(\mathbb{R}^{2d})$,

$$H(\gamma) \geq H(\mathcal{N}(\mathbb{E}(\gamma), Var(\gamma))),$$

$$H(\gamma) \geq -d \ln W_2^2(\gamma, \delta_{\mathbb{E}(\gamma)}) + C_d.$$

idea of proof

"Normal distributions minimize the entropy at fixed variance"

For $\gamma \in \mathcal{P}_{2,ac}(\mathbb{R}^{2d})$,

$$H(\gamma) \geq H(\mathcal{N}(\mathbb{E}(\gamma), Var(\gamma))),$$

$$H(\gamma) \geq -2d \ln W_2(\gamma, \delta_{\mathbb{E}(\gamma)}) + C_d.$$

idea of proof

Choose $a \in \mathbb{R}^{2d}$.

For $\gamma \in \mathcal{P}_{2,ac}(\mathbb{R}^{2d})$,

$$H(\gamma) \geq -2d \ln W_2(\gamma, \delta_a) + C_d$$

idea of proof

Choose $a \in \mathbb{R}^{2d}$.

For $\gamma \in \mathcal{P}_{2,ac}(\mathbb{R}^{2d})$,

$$H(\gamma) \geq -2d \ln W_2(\gamma, \delta_a) + C_d$$

Idea : This is not true only for diracs:

idea of proof

Choose $a \in \mathbb{R}^{2d}$.

For $\gamma \in \mathcal{P}_{2,ac}(\mathbb{R}^{2d})$,

$$H(\gamma) \geq -2d \ln W_2(\gamma, \delta_a) + C_d$$

Idea : This is not true only for diracs:

If $\gamma_0 \in \mathcal{P}(\mathbb{R}^{2d})$ "is of dimension k,"

$$\forall \gamma \in \mathcal{P}_{2,ac}(\mathbb{R}^{2d}) \quad H(\gamma) \geq -(2d - k) \ln W_2(\gamma, \gamma_0) + C(\gamma_0)$$

idea of proof

Choose $a \in \mathbb{R}^{2d}$.

For $\gamma \in \mathcal{P}_{2,ac}(\mathbb{R}^{2d})$,

$$H(\gamma) \geq -2d \ln W_2(\gamma, \delta_a) + C_d$$

Idea : This is not true only for diracs:

If $\gamma_0 \in \mathcal{P}(\mathbb{R}^{2d})$ "is of dimension k,"

$$\forall \gamma \in \mathcal{P}_{2,ac}(\mathbb{R}^{2d}) \quad H(\gamma) \geq -(2d-k) \ln W_2(\gamma, \gamma_0) + C(\gamma_0)$$

In our case γ_0 is of dimension d (Brenier theorem):

$$H(\gamma_\varepsilon) \geq -d \ln W_2(\gamma_\varepsilon, \gamma_0) + C(\gamma_0)$$

idea of proof

Choose $a \in \mathbb{R}^{2d}$.

For $\gamma \in \mathcal{P}_{2,ac}(\mathbb{R}^{2d})$,

$$H(\gamma) \geq -2d \ln W_2(\gamma, \delta_a) + C_d$$

Idea : This is not true only for diracs:

If $\gamma_0 \in \mathcal{P}(\mathbb{R}^{2d})$ "is of dimension k,"

$$\forall \gamma \in \mathcal{P}_{2,ac}(\mathbb{R}^{2d}) \quad H(\gamma) \geq -(2d - k) \ln W_2(\gamma, \gamma_0) + C(\gamma_0)$$

In our case γ_0 is of dimension d (Brenier theorem):

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

idea of proof

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

idea of proof

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

From the existing litterature,

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

idea of proof

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

From the existing litterature,

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

Combining both,

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

idea of proof

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

idea of proof

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

True because $W_2^2(\gamma_\varepsilon, \gamma_0)$ measures the transverse distance between γ_ε and γ_0

idea of proof

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

True because $W_2^2(\gamma_\varepsilon, \gamma_0)$ measures the transverse distance between γ_ε and γ_0

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

idea of proof

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

True because $W_2^2(\gamma_\varepsilon, \gamma_0)$ measures the transverse distance between γ_ε and γ_0

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

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

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

From the existing litterature

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

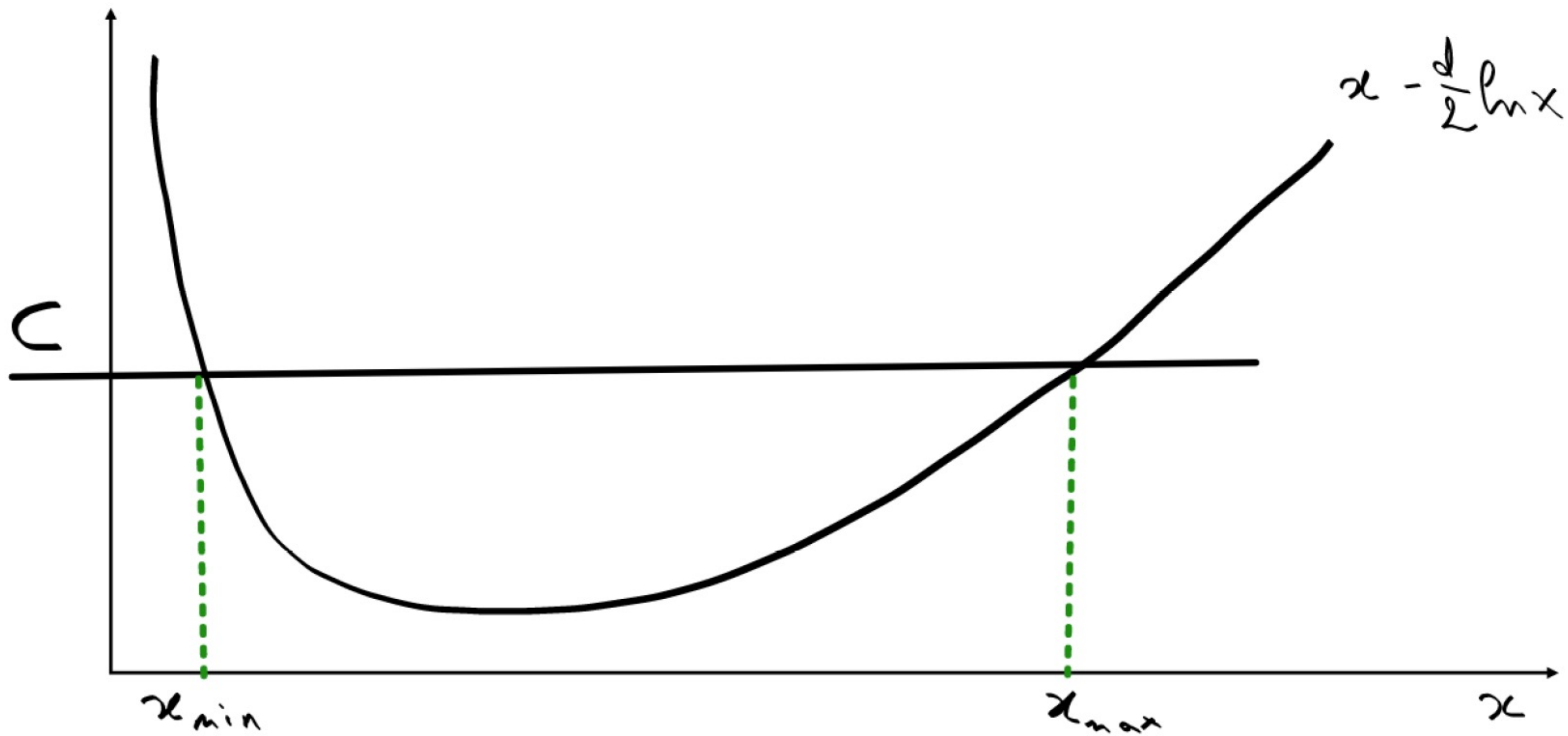
$$H(\gamma_\varepsilon) \gtrsim -\frac{d}{2}\ln(\int c\mathrm{d}\gamma_\varepsilon - \int c\mathrm{d}\gamma_0)$$

From the existing litterature

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

Combining both,

$$C \geq \frac{\int c\mathrm{d}\gamma_\varepsilon - \int c\mathrm{d}\gamma_0}{\varepsilon} - \frac{d}{2}\ln\left(\frac{\int c\mathrm{d}\gamma_\varepsilon - \int c\mathrm{d}\gamma_0}{\varepsilon}\right) \quad (15)$$



x - d/2 ln x
C
x_min
x_max
x

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

From the existing litterature

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

Combining both,

$$C \geq \frac{\int c d\gamma_\varepsilon - \int c d\gamma_0}{\varepsilon} - \frac{d}{2}\ln\left(\frac{\int c d\gamma_\varepsilon - \int c d\gamma_0}{\varepsilon}\right) \tag{15}$$

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

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

Key point

For the problem

$$\min_\gamma \int c \mathrm{d}\gamma + \varepsilon H(\gamma),$$

We got an estimate of the type :

Key point

For the problem

$$\min_{\gamma} \int c \mathrm{d}\gamma + \varepsilon H(\gamma),$$

We got an estimate of the type :

"If the suboptimality $\int c d\gamma - \int c d\gamma_0$ is small enough, then the entropy explodes"

Key point

For the problem

$$\min_\gamma \int c d\gamma + \varepsilon H(\gamma),$$

We got an estimate of the type :

"If the suboptimality $\int c d\gamma - \int c d\gamma_0$ is small enough, then the entropy explodes"

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

Infinitesimally twisted costs and compact supports

Main result

Definition

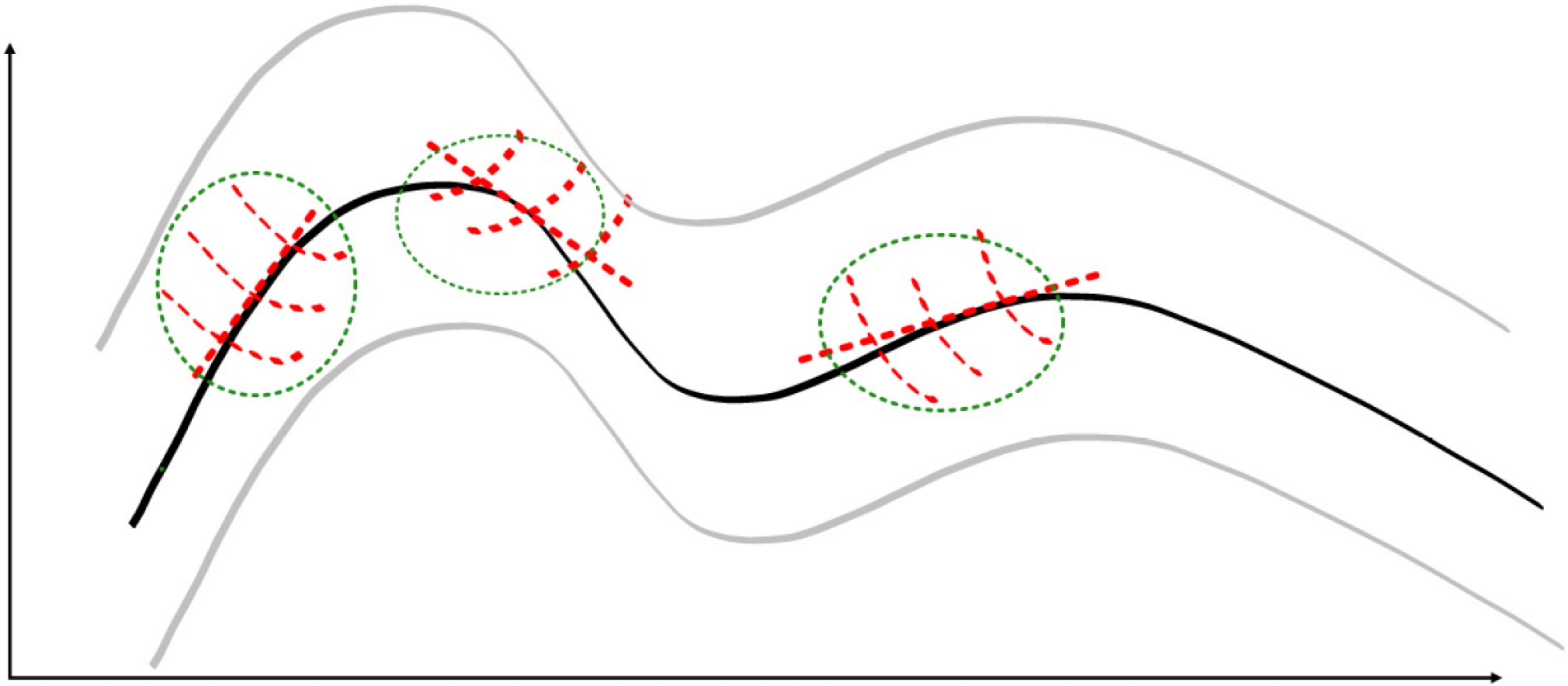
$c \in \mathcal{C}^2(\Omega^2)$ is said to be infinitesimally twisted if $\nabla^2_{xy} c(x,y) = (\partial^2_{x_i y_j} c(x,y))_{i,j} \in M_d(\mathbb{R})$ is invertible for every $(x,y) \in \Omega^2$.

Theorem

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

$$(c, \gamma_\varepsilon) = OT_0 + \Theta(\varepsilon), \quad H(\gamma_\varepsilon \mid \mathcal{H}^{2d}) = -\frac{d}{2}\ln(\varepsilon) + O(1), \quad \sqrt{\varepsilon} = O(W_2(\gamma_\varepsilon, \gamma_0)) \tag{16}$$

Note that here γ_0 is any optimal transport plan.



Thank you !

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.

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.

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

References v

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.