Displacement convexity of the relative entropy in the discrete hypercube

Nathaël Gozlan*

*LAMA Université Paris Est – Marne-la-Vallée

Phenomena in high dimensions in geometric analysis, random matrices, and computational geometry
Roscoff, June 2012

General problem

$(\mathcal{X}, d)$ a complete separable metric space.

General problem

$(\mathcal{X}, d)$ a complete separable metric space.

Definition [Relative entropy]

Let μ, ν be two Borel probability measures on $\mathcal{X}$

$$H(\nu|\mu) := \int \log \frac{d\nu}{d\mu}\, d\nu$$

if ν is absolutely continuous with respect to μ (otherwise we set $H(\nu|\mu) = \infty$).

General problem

$(\mathcal{X}, d)$ a complete separable metric space.

Definition [Relative entropy]

Let μ, ν be two Borel probability measures on $\mathcal{X}$

$$H(\nu|\mu) := \int \log \frac{d\nu}{d\mu}\, d\nu$$

if ν is absolutely continuous with respect to μ (otherwise we set $H(\nu|\mu) = \infty$).

The map $\nu \mapsto H(\nu|\mu)$ is always convex in the usual sense:

$$H((1-t)\nu_0 + t\nu_1|\mu) \leq (1-t)H(\nu_0|\mu) + tH(\nu_1|\mu), \qquad \forall t \in [0,1].$$

General problem

$(\mathcal{X}, d)$ a complete separable metric space.

Definition [Relative entropy]

Let μ, ν be two Borel probability measures on $\mathcal{X}$

$$H(\nu|\mu) := \int \log \frac{d\nu}{d\mu}\, d\nu$$

if ν is absolutely continuous with respect to μ (otherwise we set $H(\nu|\mu) = \infty$).

The map $\nu \mapsto H(\nu|\mu)$ is always convex in the usual sense:

$$H((1-t)\nu_0 + t\nu_1|\mu) \leq (1-t)H(\nu_0|\mu) + tH(\nu_1|\mu), \qquad \forall t \in [0,1].$$

Problem: Examine the convexity of H along other types of paths $(\nu_t)_{t\in[0,1]}$ interpolating between ν_0 and ν_1.

General problem

$(\mathcal{X}, d)$ a complete separable metric space.

Definition [Relative entropy]

Let μ, ν be two Borel probability measures on $\mathcal{X}$

$$H(\nu|\mu) := \int \log \frac{d\nu}{d\mu}\, d\nu$$

if ν is absolutely continuous with respect to μ (otherwise we set $H(\nu|\mu) = \infty$).

The map $\nu \mapsto H(\nu|\mu)$ is always convex in the usual sense:

$$H((1-t)\nu_0 + t\nu_1|\mu) \leq (1-t)H(\nu_0|\mu) + tH(\nu_1|\mu), \qquad \forall t \in [0,1].$$

Problem: Examine the convexity of H along other types of paths $(\nu_t)_{t\in[0,1]}$ interpolating between ν_0 and ν_1.

$\rightsquigarrow$ Convexity along geodesics of the Wasserstein W_2 distance has connections with curvature of the underlying space and functional inequalities (Log-Sobolev, Talagrand) or Brunn-Minkowski type inequalities.

Outline

Based on a joint work in progress with C. Roberto, P-M Samson and P. Tetali

I. Displacement convexity of the entropy in a continuous setting.
$\rightsquigarrow$ Link with curvature and functional inequalities.

II. Displacement convexity in discrete setting.
$\rightsquigarrow$ General framework and the example of the discrete hypercube.

I. Displacement convexity of the entropy in a continuous setting.

Geodesic spaces

A metric space (E, d) is said *geodesic* if for all $x_0, x_1 \in E$ there is at least one path $\gamma : [0,1] \to E$ such that $\gamma(0) = x_0$ and $\gamma(1) = x_1$ and

$$d(\gamma(s), \gamma(t)) = |t - s| d(x_0, x_1), \qquad \forall t, s \in [0,1].$$

Such a path is called a *constant speed geodesic* between x_0 and x_1.

Let $\mathcal{P}_p(\mathcal{X})$, $p \geq 1$, be the set of Borel probability measures having a finite p-th moment.

Definition [L_p-Wasserstein distance]

Let $\nu_0, \nu_1 \in \mathcal{P}_p(\mathcal{X})$;

$$W_p^p(\nu_0, \nu_1) = \inf_{\pi \in P(\nu_0, \nu_1)} \iint d^p(x_0, x_1)\, \pi(dx_0 dx_1),$$

where $P(\nu_0, \nu_1)$ is the set of couplings of ν_0 and ν_1.

Geodesics in the Wasserstein space

Let $\mathcal{P}_p(\mathcal{X})$, $p \geq 1$, be the set of Borel probability measures having a finite p-th moment.

Definition [L_p-Wasserstein distance]

Let $\nu_0, \nu_1 \in \mathcal{P}_p(\mathcal{X})$;

$$W_p^p(\nu_0, \nu_1) = \inf_{\pi \in P(\nu_0, \nu_1)} \iint d^p(x_0, x_1)\, \pi(dx_0 dx_1),$$

where $P(\nu_0, \nu_1)$ is the set of couplings of ν_0 and ν_1.

Theorem

If $p > 1$, the space $(\mathcal{P}_p(\mathcal{X}), W_p)$ is geodesic if and only if $(\mathcal{X}, d)$ is geodesic.

Displacement convexity of the entropy and the Bakry-Emery criterion

Theorem

Let (M,g) be a complete connected Riemannian manifold and suppose that $\mu \in \mathcal{P}(M)$ is absolutely continuous with $\mu(dx) = e^{-V(x)}\,dx$. The following are equivalent:

1. μ verifies the $\mathrm{CD}(K,\infty)$ condition, for some $K \in \mathbf{R}$:
$$\mathrm{Ric} + \mathrm{Hess}\, V \geq Kg$$
2. The relative entropy functional is K-displacement convex with respect to the W_2 metric: for all $\nu_0, \nu_1 \in \mathcal{P}_2(M)$, absolutely continuous with respect to μ, there is a W_2-geodesic $(\nu_t)_{t\in[0,1]}$ connecting ν_0 to ν_1 such that
$$H(\nu_t|\mu) \leq (1-t)H(\nu_0|\mu) + tH(\nu_1|\mu) - K\frac{t(1-t)}{2}W_2^2(\nu_0,\nu_1), \quad \forall t \in [0,1]$$

$\rightsquigarrow$ Mc Cann, Cordero-McCann-Schmuckenschläger, Sturm-Von Renesse, Lott-Villani, Sturm.

Consequences

1. This equivalence sets the ground of a possible definition of the condition $\mathrm{CD}(K, \infty)$ on geodesic spaces (Lott-Sturm-Villani).

Consequences

1. This equivalence sets the ground of a possible definition of the condition $\mathrm{CD}(K,\infty)$ on geodesic spaces (Lott-Sturm-Villani).
2. **Brunn-Minkowski type inequalities.** Suppose $K \geq 0$, then

$$\mu([A,B]_t) \geq \mu(A)^{1-t}\mu(B)^t e^{\frac{Kt(1-t)}{2} d^2(A,B)}, \qquad \forall t \in [0,1],$$

where $[A,B]_t = \{x \in M; \exists (a,b) \in A \times B, d(x,a) = (1-t)d(a,b) \text{ and } d(x,b) = td(a,b)\}$.

Consequences

1. This equivalence sets the ground of a possible definition of the condition $\mathrm{CD}(K, \infty)$ on geodesic spaces (Lott-Sturm-Villani).
2. **Brunn-Minkowski type inequalities.** Suppose $K \geq 0$, then

$$\mu([A,B]_t) \geq \mu(A)^{1-t}\mu(B)^t e^{\frac{Kt(1-t)}{2}d^2(A,B)}, \qquad \forall t \in [0,1],$$

   where $[A,B]_t = \{x \in M; \exists (a,b) \in A \times B, d(x,a) = (1-t)d(a,b) \text{ and } d(x,b) = td(a,b)\}$.
3. **Talagrand's inequality.** Suppose $K > 0$, then

$$W_2^2(\nu_0, \mu) \leq \frac{2}{K} H(\nu_0|\mu), \qquad \forall \nu_0 \in \mathcal{P}_2(M).$$

Consequences

1. This equivalence sets the ground of a possible definition of the condition $\mathrm{CD}(K,\infty)$ on geodesic spaces (Lott-Sturm-Villani).
2. **Brunn-Minkowski type inequalities.** Suppose $K \geq 0$, then

$$\mu([A,B]_t) \geq \mu(A)^{1-t}\mu(B)^t e^{\frac{Kt(1-t)}{2}d^2(A,B)}, \qquad \forall t \in [0,1],$$

   where $[A,B]_t = \{x \in M; \exists (a,b) \in A \times B, d(x,a) = (1-t)d(a,b) \text{ and } d(x,b) = td(a,b)\}$.
3. **Talagrand's inequality.** Suppose $K > 0$, then

$$W_2^2(\nu_0,\mu) \leq \frac{2}{K} H(\nu_0|\mu), \qquad \forall \nu_0 \in \mathcal{P}_2(M).$$

4. **HWI inequality.** Suppose $K \in \mathbf{R}$, then

$$H(\nu_0|\mu) \leq \sqrt{I(\nu_0|\mu)}\,W_2(\nu_0,\mu) - \frac{K}{2}W_2^2(\nu_0,\mu), \qquad \forall \nu_0 \in \mathcal{P}_2(M),$$

   where the Fisher information is defined by

$$I(\nu|\mu) = \int \frac{|\nabla h|^2}{h}\,d\mu, \qquad \nu = h\mu.$$

5. **Log-Sobolev inequality.** Suppose $K > 0$, then

$$H(\nu_0|\mu) \leq \frac{2}{K} I(\nu_0|\mu), \qquad \forall \nu_0.$$

5. **Log-Sobolev inequality.** Suppose $K > 0$, then

$$H(\nu_0|\mu) \leq \frac{2}{K} I(\nu_0|\mu), \qquad \forall \nu_0.$$

6. **Prekopa-Leindler type inequalities.** Suppose $K \in \mathbf{R}$, and fix $t \in (0,1)$. If $f, g, h : M \to \mathbf{R}$ are such that

$$h(z) \geq (1-t)f(x) + tg(y) - \frac{Kt(1-t)}{2} d^2(x,y), \qquad \forall x, y \in M, z \in [x,y]_t$$

then

$$\int e^{h(z)} \mu(dz) \geq \left(\int e^{f(x)} \mu(dx) \right)^{1-t} \left(\int e^{g(y)} \mu(dy) \right)^t$$

II. Displacement convexity of the entropy in a discrete setting.

Extension to discrete setting

Question: Is it possible to extend this theory to discrete settings, for example finite graphs ?

Extension to discrete setting

Question: Is it possible to extend this theory to discrete settings, for example finite graphs ?

Two obstructions:

- Talagrand's inequality

$$W_2^2(\nu_0, \mu) \leq \mathrm{Cste} H(\nu_0|\mu), \qquad \forall \nu_0$$

does not hold in discrete setting unless μ is a Dirac.

Extension to discrete setting

Question: Is it possible to extend this theory to discrete settings, for example finite graphs ?

Two obstructions:

- Talagrand's inequality

$$W_2^2(\nu_0, \mu) \leq \mathrm{Cste} H(\nu_0|\mu), \qquad \forall \nu_0$$

  does not hold in discrete setting unless μ is a Dirac.

- W_2-geodesics do not exist in discrete setting. Namely, if ν_t is a W_2 geodesic between δ_x and δ_y, then it is easy to see that ν_t is supported in $[x, y]_t$. But in discrete, $[x, y]_t$ can be empty. For example, if x, y are neighbors in a graph, then $[x, y]_t = \emptyset$, for all $t \in (0, 1)$.

Extension to discrete setting

Question: Is it possible to extend this theory to discrete settings, for example finite graphs ?

Two obstructions:

- Talagrand's inequality

$$W_2^2(\nu_0, \mu) \leq \mathrm{Cste} H(\nu_0|\mu), \qquad \forall \nu_0$$

  does not hold in discrete setting unless μ is a Dirac.

- W_2-geodesics do not exist in discrete setting. Namely, if ν_t is a W_2 geodesic between δ_x and δ_y, then it is easy to see that ν_t is supported in $[x, y]_t$. But in discrete, $[x, y]_t$ can be empty. For example, if x, y are neighbors in a graph, then $[x, y]_t = \emptyset$, for all $t \in (0, 1)$.

$\rightsquigarrow$ W_2 is not adapted neither for defining the path ν_t nor for measuring the convexity defect/excess of the entropy.

Recent results

$\rightsquigarrow$ Ollivier, Ollivier-Villani, Bonciocat-Sturm, Maas, Erbar-Maas, Mielke, Hillion, Léonard, Lehec, Joulin, Veysseire . . .

Recent results

$\rightsquigarrow$ Ollivier, Ollivier-Villani, Bonciocat-Sturm, Maas, Erbar-Maas, Mielke, Hillion, Léonard, Lehec, Joulin, Veysseire ...

Ollivier-Villani: Brunn-Minkowski type inequality on the discrete hypercube $\Omega_n = \{0,1\}^n$ equipped with the Hamming distance $d(x,y) = \sum_{i=1}^n \mathbb{1}_{x_i \neq y_i}$.

$$|[A,B]_{1/2}| \geq |A|^{1/2}|B|^{1/2} e^{\frac{1}{16n} d^2(A,B)}, \qquad \forall A, B \subset \Omega_n$$

Recent results

$\rightsquigarrow$ Ollivier, Ollivier-Villani, Bonciocat-Sturm, Maas, Erbar-Maas, Mielke, Hillion, Léonard, Lehec, Joulin, Veysseire . . .

Ollivier-Villani: Brunn-Minkowski type inequality on the discrete hypercube $\Omega_n = \{0,1\}^n$ equipped with the Hamming distance $d(x,y) = \sum_{i=1}^n \mathbb{1}_{x_i \neq y_i}$.

$$|[A,B]_{1/2}| \geq |A|^{1/2}|B|^{1/2} e^{\frac{1}{16n} d^2(A,B)}, \qquad \forall A, B \subset \Omega_n$$

Consequence of the following property: $\forall \nu_0, \nu_1$, $\exists \nu_{1/2}$ such that

$$H(\nu_{1/2}|\mu) \leq \frac{1}{2} H(\nu_0|\mu) + \frac{1}{2} H(\nu_1|\mu) - \frac{1}{16n} W_1^2(\nu_0, \nu_1),$$

where μ is the uniform measure on Ω_n.

Recent results

Maas-Erbar: Displacement convexity of the entropy on the discrete hypercube $\Omega_n = \{0,1\}^n$.

$$H(\nu_t|\mu) \leq (1-t)H(\nu_0|\mu) + tH(\nu_1|\mu) - \frac{t(1-t)}{n}\mathcal{W}_2^2(\nu_0,\nu_1)$$

Recent results

Maas-Erbar: Displacement convexity of the entropy on the discrete hypercube $\Omega_n = \{0,1\}^n$.

$$H(\nu_t|\mu) \leq (1-t)H(\nu_0|\mu) + tH(\nu_1|\mu) - \frac{t(1-t)}{n}\mathcal{W}_2^2(\nu_0,\nu_1)$$

The (finite dimensional) space $\mathcal{P}(\Omega_n)$ is equipped with a Riemannian metric.

Maas-Erbar: Displacement convexity of the entropy on the discrete hypercube $\Omega_n = \{0,1\}^n$.

$$H(\nu_t|\mu) \leq (1-t)H(\nu_0|\mu) + tH(\nu_1|\mu) - \frac{t(1-t)}{n}\mathcal{W}_2^2(\nu_0,\nu_1)$$

The (finite dimensional) space $\mathcal{P}(\Omega_n)$ is equipped with a Riemannian metric.

This metric is made in such a way that the continuous time simple random walk becomes a gradient flow of the function $H(\,\cdot\,|\mu)$ (with respect to the Riemannian structure). [This construction is very general and holds for every reversible Markov kernel on a finite graph].

Maas-Erbar: Displacement convexity of the entropy on the discrete hypercube $\Omega_n = \{0,1\}^n$.

$$H(\nu_t|\mu) \leq (1-t)H(\nu_0|\mu) + tH(\nu_1|\mu) - \frac{t(1-t)}{n}\mathcal{W}_2^2(\nu_0,\nu_1)$$

The (finite dimensional) space $\mathcal{P}(\Omega_n)$ is equipped with a Riemannian metric.

This metric is made in such a way that the continuous time simple random walk becomes a gradient flow of the function $H(\,\cdot\,|\mu)$ (with respect to the Riemannian structure). [This construction is very general and holds for every reversible Markov kernel on a finite graph].

The pseudo Wasserstein distance $\mathcal{W}_2$ corresponds to the geodesic distance on $\mathcal{P}(\Omega_n)$ and ν_t is a geodesic connecting ν_0 to ν_1.

Maas-Erbar: Displacement convexity of the entropy on the discrete hypercube $\Omega_n = \{0,1\}^n$.

$$H(\nu_t|\mu) \leq (1-t)H(\nu_0|\mu) + tH(\nu_1|\mu) - \frac{t(1-t)}{n}\mathcal{W}_2^2(\nu_0,\nu_1)$$

The (finite dimensional) space $\mathcal{P}(\Omega_n)$ is equipped with a Riemannian metric.

This metric is made in such a way that the continuous time simple random walk becomes a gradient flow of the function $H(\,\cdot\,|\mu)$ (with respect to the Riemannian structure). [This construction is very general and holds for every reversible Markov kernel on a finite graph].

The pseudo Wasserstein distance $\mathcal{W}_2$ corresponds to the geodesic distance on $\mathcal{P}(\Omega_n)$ and ν_t is a geodesic connecting ν_0 to ν_1.

The function $\mathcal{W}_2^2$ is not a transport cost.

Recent results

Maas-Erbar: Displacement convexity of the entropy on the discrete hypercube $\Omega_n = \{0,1\}^n$.

$$H(\nu_t|\mu) \leq (1-t)H(\nu_0|\mu) + tH(\nu_1|\mu) - \frac{t(1-t)}{n}\mathcal{W}_2^2(\nu_0,\nu_1)$$

The (finite dimensional) space $\mathcal{P}(\Omega_n)$ is equipped with a Riemannian metric.

This metric is made in such a way that the continuous time simple random walk becomes a gradient flow of the function $H(\,\cdot\,|\mu)$ (with respect to the Riemannian structure). [This construction is very general and holds for every reversible Markov kernel on a finite graph].

The pseudo Wasserstein distance $\mathcal{W}_2$ corresponds to the geodesic distance on $\mathcal{P}(\Omega_n)$ and ν_t is a geodesic connecting ν_0 to ν_1.

The function $\mathcal{W}_2^2$ is not a transport cost.

$\rightsquigarrow$ This implies some versions of LSI and HWI, and a Talagrand type inequality for the $\mathcal{W}_2$ and W_1 distances with sharp constants.

Our approach

similar approach developed independently by E Hillion

Our approach

similar approach developed independently by E Hillion

We seek to prove convexity of the entropy along (pseudo)-$\mathbf{W_1}$**-geodesics** ν_t connecting ν_0 to ν_1 :

Our approach

similar approach developed independently by E Hillion

We seek to prove convexity of the entropy along (pseudo)-$\mathbf{W_1}$**-geodesics** ν_t connecting ν_0 to ν_1 :

There is some $K \geq 0$ such that for all ν_0, ν_1 there is some pseudo W_1-geodesic ν_t connecting ν_0 to ν_1 such that for all $t \in [0, 1]$

$$H(\nu_t|\mu) \leq (1-t)H(\nu_0|\mu) + tH(\nu_1|\mu) - \frac{Kt(1-t)}{2} \times \cdots .$$

Our approach

similar approach developed independently by E Hillion

We seek to prove convexity of the entropy along (pseudo)-$\mathbf{W_1}$**-geodesics** ν_t connecting ν_0 to ν_1 :

There is some $K \geq 0$ such that for all ν_0, ν_1 there is some pseudo W_1-geodesic ν_t connecting ν_0 to ν_1 such that for all $t \in [0, 1]$

$$H(\nu_t|\mu) \leq (1-t)H(\nu_0|\mu) + tH(\nu_1|\mu) - \frac{Kt(1-t)}{2} \times \cdots .$$

A good candidate for the transport term in the right hand side is

$$W_1^2(\nu_0, \nu_1).$$

Our approach

similar approach developed independently by E Hillion

We seek to prove convexity of the entropy along (pseudo)-$\mathbf{W_1}$**-geodesics** ν_t connecting ν_0 to ν_1 :

There is some $K \geq 0$ such that for all ν_0, ν_1 there is some pseudo W_1-geodesic ν_t connecting ν_0 to ν_1 such that for all $t \in [0,1]$

$$H(\nu_t|\mu) \leq (1-t)H(\nu_0|\mu) + tH(\nu_1|\mu) - \frac{Kt(1-t)}{2} \times \cdots .$$

A good candidate for the transport term in the right hand side is

$$W_1^2(\nu_0, \nu_1).$$

Another better candidate is the transport cost

$$\widetilde{W}_2^2(\nu_0, \nu_1),$$

introduced by Marton, which is stronger than W_1 and weaker than W_2.

Our main example - the discrete hypercube Ω_n

Theorem [GRST]

Let μ be a probability on $\{0,1\}$ and denote by μ^n its n-fold tensor product. The following displacement convexity properties of the entropy hold true.

For all $\nu_0, \nu_1 \in \mathcal{P}(\Omega_n)$, there is a coupling $\pi \in P(\nu_0, \nu_1)$ such that for all $t \in [0,1]$, it holds

Our main example - the discrete hypercube Ω_n

Theorem [GRST]

Let μ be a probability on $\{0,1\}$ and denote by μ^n its n-fold tensor product. The following displacement convexity properties of the entropy hold true.

For all $\nu_0, \nu_1 \in \mathcal{P}(\Omega_n)$, there is a coupling $\pi \in P(\nu_0, \nu_1)$ such that for all $t \in [0,1]$, it holds

$$H(\nu_t^\pi | \mu^n) \leq (1-t) H(\nu_0|\mu^n) + tH(\nu_1|\mu^n) - \frac{2t(1-t)}{n} W_1^2(\nu_0, \nu_1) \qquad (1)$$

Our main example - the discrete hypercube Ω_n

Theorem [GRST]

Let μ be a probability on $\{0,1\}$ and denote by μ^n its n-fold tensor product. The following displacement convexity properties of the entropy hold true.

For all $\nu_0, \nu_1 \in \mathcal{P}(\Omega_n)$, there is a coupling $\pi \in P(\nu_0, \nu_1)$ such that for all $t \in [0,1]$, it holds

$$(1) \qquad H(\nu_t^\pi|\mu^n) \leq (1-t)H(\nu_0|\mu^n) + tH(\nu_1|\mu^n) - \frac{2t(1-t)}{n} W_1^2(\nu_0, \nu_1)$$

and

$$(2) \qquad H(\nu_t^\pi|\mu^n) \leq (1-t)H(\nu_0|\mu^n) + tH(\nu_1|\mu^n) - \frac{t(1-t)}{2} \widetilde{W}_2^2(\nu_0, \nu_1),$$

Our main example - the discrete hypercube Ω_n

Theorem [GRST]

Let μ be a probability on $\{0,1\}$ and denote by μ^n its n-fold tensor product. The following displacement convexity properties of the entropy hold true.

For all $\nu_0, \nu_1 \in \mathcal{P}(\Omega_n)$, there is a coupling $\pi \in P(\nu_0, \nu_1)$ such that for all $t \in [0,1]$, it holds

$$(1) \qquad H(\nu_t^\pi|\mu^n) \leq (1-t)H(\nu_0|\mu^n) + tH(\nu_1|\mu^n) - \frac{2t(1-t)}{n} W_1^2(\nu_0, \nu_1)$$

and

$$(2) \qquad H(\nu_t^\pi|\mu^n) \leq (1-t)H(\nu_0|\mu^n) + tH(\nu_1|\mu^n) - \frac{t(1-t)}{2} \widetilde{W}_2^2(\nu_0, \nu_1),$$

where

$$\nu_t^\pi = \sum_{x_0, x_1 \in \Omega_n} \pi(x_0, x_1) \sum_{z \in [\![x_0, x_1]\!]} t^{d(x_0,z)}(1-t)^{d(z,x_1)} \delta_z,$$

where $z \in [\![x_0, x_1]\!]$ means that z belongs to some geodesic joining x_0 to x_1.

Our main example - the discrete hypercube Ω_n

Theorem [GRST]

Let μ be a probability on $\{0,1\}$ and denote by μ^n its n-fold tensor product. The following displacement convexity properties of the entropy hold true.

For all $\nu_0, \nu_1 \in \mathcal{P}(\Omega_n)$, there is a coupling $\pi \in P(\nu_0, \nu_1)$ such that for all $t \in [0,1]$, it holds

$$(1) \qquad H(\nu_t^\pi | \mu^n) \leq (1-t)H(\nu_0|\mu^n) + tH(\nu_1|\mu^n) - \frac{2t(1-t)}{n} W_1^2(\nu_0, \nu_1)$$

and

$$(2) \qquad H(\nu_t^\pi | \mu^n) \leq (1-t)H(\nu_0|\mu^n) + tH(\nu_1|\mu^n) - \frac{t(1-t)}{2} \widetilde{W}_2^2(\nu_0, \nu_1),$$

where

$$\nu_t^\pi = \sum_{x_0, x_1 \in \Omega_n} \pi(x_0, x_1) \sum_{z \in [\![x_0, x_1]\!]} t^{d(x_0,z)} (1-t)^{d(z,x_1)} \delta_z,$$

where $z \in [\![x_0, x_1]\!]$ means that z belongs to some geodesic joining x_0 to x_1.

We will define Marton's distance $\widetilde{W}_2$ later on. It verifies $\widetilde{W}_2^2(\nu_0, \nu_1) \geq \frac{2}{n} W_1^2(\nu_0, \nu_1)$. So (2) is stronger than (1) (up to a factor 2).

A first consequence - Transport inequalities and concentration

μ^n verifies the following transport-entropy inequalities

$$W_1^2(\nu_0, \mu^n) \leq \frac{n}{2} H(\nu_0|\mu^n), \qquad \forall \nu_0 \in \mathcal{P}(\Omega_n)$$

and

$$\widetilde{W}_2^2(\nu_0, \mu^n) \leq 2\,\mathrm{H}(\nu_0|\mu^n), \qquad \forall \nu_0 \in \mathcal{P}(\Omega_n).$$

A first consequence - Transport inequalities and concentration

μ^n verifies the following transport-entropy inequalities

$$W_1^2(\nu_0, \mu^n) \leq \frac{n}{2} H(\nu_0|\mu^n), \qquad \forall \nu_0 \in \mathcal{P}(\Omega_n)$$

and

$$\widetilde{W}_2^2(\nu_0, \mu^n) \leq 2\,\mathrm{H}(\nu_0|\mu^n), \qquad \forall \nu_0 \in \mathcal{P}(\Omega_n).$$

They imply respectively, the following well known concentration inequalities

$$\mu^n(f \geq \mu^n(f) + t) \leq e^{-\frac{2}{n}t^2}, \qquad \forall t \geq 0,$$

for all function $f : \Omega_n \to \mathbf{R}$, 1-Lipschitz with respect to Hamming distance.

A first consequence - Transport inequalities and concentration

μ^n verifies the following transport-entropy inequalities

$$W_1^2(\nu_0, \mu^n) \leq \frac{n}{2} H(\nu_0|\mu^n), \qquad \forall \nu_0 \in \mathcal{P}(\Omega_n)$$

and

$$\widetilde{W}_2^2(\nu_0, \mu^n) \leq 2\,\mathrm{H}(\nu_0|\mu^n), \qquad \forall \nu_0 \in \mathcal{P}(\Omega_n).$$

They imply respectively, the following well known concentration inequalities

$$\mu^n(f \geq \mu^n(f) + t) \leq e^{-\frac{2}{n}t^2}, \qquad \forall t \geq 0,$$

for all function $f : \Omega_n \to \mathbf{R}$, 1-Lipschitz with respect to Hamming distance.

$$\mu^n(f \geq \mu^n(f) + t) \leq e^{-\frac{t^2}{4}}, \qquad \forall t \geq 0,$$

for all $f : [0,1]^n \to \mathbf{R}$ convex and 1-Lipschitz for the Euclidean distance on $\mathbf{R}^n$.

A first consequence - Transport inequalities and concentration

μ^n verifies the following transport-entropy inequalities

$$W_1^2(\nu_0, \mu^n) \leq \frac{n}{2} H(\nu_0|\mu^n), \qquad \forall \nu_0 \in \mathcal{P}(\Omega_n)$$

and

$$\widetilde{W}_2^2(\nu_0, \mu^n) \leq 2\,\mathrm{H}(\nu_0|\mu^n), \qquad \forall \nu_0 \in \mathcal{P}(\Omega_n).$$

They imply respectively, the following well known concentration inequalities

$$\mu^n(f \geq \mu^n(f) + t) \leq e^{-\frac{2}{n}t^2}, \qquad \forall t \geq 0,$$

for all function $f : \Omega_n \to \mathbf{R}$, 1-Lipschitz with respect to Hamming distance.

$$\mu^n(f \geq \mu^n(f) + t) \leq e^{-\frac{t^2}{4}}, \qquad \forall t \geq 0,$$

for all $f : [0,1]^n \to \mathbf{R}$ convex and 1-Lipschitz for the Euclidean distance on $\mathbf{R}^n$.

We will see other consequences of (2) (Prekopa-Leindler - HWI) later on.

A general framework.

In all the sequel, $G = (V, E)$ will be a finite connected graph equipped with the classical graph distance d.

If $\gamma = (x_0, x_1, \ldots, x_n)$ $(n = d(x, y))$ is a geodesic between x and y, we write $\gamma(k) = x_k$.

Definition [L_1-Wasserstein distance]

For all $\nu_0, \nu_1 \in \mathcal{P}(V)$, $\quad W_1(\nu_0, \nu_1) = \inf_{\pi \in P(\nu_0, \nu_1)} \iint d(x_0, x_1)\, \pi(dx_0 dx_1)$.

Important example:

Let $V = \{0, 1, \ldots, n\}$; consider the Binomial distribution $\mathcal{B}(n, t)$ defined by

$$\mathcal{B}(n, t) = \sum_{k=0}^{n} \binom{n}{k} t^k (1-t)^{n-k} \delta_k \in \mathcal{P}(V).$$

Then $(\mathcal{B}(n, t))_{t \in [0,1]}$ is a W_1 constant speed geodesic connecting δ_0 to δ_n.

Construction of (pseudo)-W_1 geodesic.

Let π be a coupling between ν_0 and ν_1 and (X_0, X_1) a random variable with law π.

Construction of (pseudo)-W_1 geodesic.

Let π be a coupling between ν_0 and ν_1 and (X_0, X_1) a random variable with law π.

For all $x_0, x_1 \in V$ let $\mathcal{G}(x_0, x_1)$ be the set of geodesics connecting x_0 to x_1 and consider the following random variables:

Construction of (pseudo)-W_1 geodesic.

Let π be a coupling between ν_0 and ν_1 and (X_0, X_1) a random variable with law π.

For all $x_0, x_1 \in V$ let $\mathcal{G}(x_0, x_1)$ be the set of geodesics connecting x_0 to x_1 and consider the following random variables:

- Γ^{x_0,x_1} is a random variable uniformly distributed over $\mathcal{G}(x_0, x_1)$ and independent of (X_0, X_1).

Construction of (pseudo)-W_1 geodesic.

Let π be a coupling between ν_0 and ν_1 and (X_0, X_1) a random variable with law π.

For all $x_0, x_1 \in V$ let $\mathcal{G}(x_0, x_1)$ be the set of geodesics connecting x_0 to x_1 and consider the following random variables:

- Γ^{x_0,x_1} is a random variable uniformly distributed over $\mathcal{G}(x_0, x_1)$ and independent of (X_0, X_1).

- for all $t \in [0, 1]$, $N_t^{x_0,x_1}$ is a random variable following a $\mathcal{B}(n, t)$ distribution with $n = d(x_0, x_1)$ independent of Γ^{x_0,x_1} and of (X_0, X_1).

Construction of (pseudo)-W_1 geodesic.

Let π be a coupling between ν_0 and ν_1 and (X_0, X_1) a random variable with law π.

For all $x_0, x_1 \in V$ let $\mathcal{G}(x_0, x_1)$ be the set of geodesics connecting x_0 to x_1 and consider the following random variables:

- Γ^{x_0,x_1} is a random variable uniformly distributed over $\mathcal{G}(x_0, x_1)$ and independent of (X_0, X_1).

- for all $t \in [0, 1]$, $N_t^{x_0,x_1}$ is a random variable following a $\mathcal{B}(n, t)$ distribution with $n = d(x_0, x_1)$ independent of Γ^{x_0,x_1} and of (X_0, X_1).

Finally, set

$$X_t = \Gamma^{X_0,X_1}(N_t^{X_0,X_1}) \qquad \text{and} \qquad \nu_t^\pi = \mathrm{Law}(X_t).$$

Construction of (pseudo)-W_1 geodesic.

Let π be a coupling between ν_0 and ν_1 and (X_0, X_1) a random variable with law π.

For all $x_0, x_1 \in V$ let $\mathcal{G}(x_0, x_1)$ be the set of geodesics connecting x_0 to x_1 and consider the following random variables:

- Γ^{x_0,x_1} is a random variable uniformly distributed over $\mathcal{G}(x_0, x_1)$ and independent of (X_0, X_1).

- for all $t \in [0,1]$, $N_t^{x_0,x_1}$ is a random variable following a $\mathcal{B}(n,t)$ distribution with $n = d(x_0, x_1)$ independent of Γ^{x_0,x_1} and of (X_0, X_1).

Finally, set

$$X_t = \Gamma^{X_0,X_1}(N_t^{X_0,X_1}) \qquad \text{and} \qquad \nu_t^\pi = \mathrm{Law}(X_t).$$

Proposition [GRST]

If π is an optimal coupling between ν_0 and ν_1 then ν_t^π is a W_1-geodesic between ν_0 and ν_1.

Construction of (pseudo)-W_1 geodesic.

Let π be a coupling between ν_0 and ν_1 and (X_0, X_1) a random variable with law π.

For all $x_0, x_1 \in V$ let $\mathcal{G}(x_0, x_1)$ be the set of geodesics connecting x_0 to x_1 and consider the following random variables:

- Γ^{x_0,x_1} is a random variable uniformly distributed over $\mathcal{G}(x_0, x_1)$ and independent of (X_0, X_1).

- for all $t \in [0,1]$, $N_t^{x_0,x_1}$ is a random variable following a $\mathcal{B}(n,t)$ distribution with $n = d(x_0, x_1)$ independent of Γ^{x_0,x_1} and of (X_0, X_1).

Finally, set

$$X_t = \Gamma^{X_0,X_1}(N_t^{X_0,X_1}) \qquad \text{and} \qquad \nu_t^\pi = \mathrm{Law}(X_t).$$

Proposition [GRST]

If π is an optimal coupling between ν_0 and ν_1 then ν_t^π is a W_1-geodesic between ν_0 and ν_1.

When $\nu_0 = \delta_x$ and $\nu_1 = \delta_y$, we write $\nu_t^{x,y}$.
When π is not optimal we say that ν_t^π is a pseudo W_1-geodesic.

Explicit formula for ν_t^π

Proposition

For all $\nu_0, \nu_1 \in \mathcal{P}(V)$, and $t \in [0,1]$,

$$\nu_t^\pi = \sum_{x_0, x_1} \pi(x_0, x_1) \sum_{z \in V} \binom{d(x_0, x_1)}{d(x_0, z)} t^{d(x_0,z)} (1-t)^{d(z,x_1)} \frac{|\mathcal{G}(x_0, z, x_1)|}{|\mathcal{G}(x_0, x_1)|} \delta_z,$$

where

- $\mathcal{G}(x_0, x_1)$ is the set of geodesic connecting x_0 to x_1.
- $\mathcal{G}(x_0, z, x_1)$ is the set of geodesics connecting x_0 to x_1 and crossing the vertex z.

Explicit formula for ν_t^π

Proposition

For all $\nu_0, \nu_1 \in \mathcal{P}(V)$, and $t \in [0,1]$,

$$\nu_t^\pi = \sum_{x_0,x_1} \pi(x_0,x_1) \sum_{z \in V} \binom{d(x_0,x_1)}{d(x_0,z)} t^{d(x_0,z)}(1-t)^{d(z,x_1)} \frac{|\mathcal{G}(x_0,z,x_1)|}{|\mathcal{G}(x_0,x_1)|}\delta_z,$$

where

- $\mathcal{G}(x_0,x_1)$ is the set of geodesic connecting x_0 to x_1.
- $\mathcal{G}(x_0,z,x_1)$ is the set of geodesics connecting x_0 to x_1 and crossing the vertex z.

In particular, on the cube, if $z \in [\![x_0,x_1]\!]$

$$|\mathcal{G}(x_0,x_1)| = d(x_0,x_1)! \qquad |\mathcal{G}(x_0,z,x_1)| = d(x_0,z)!d(z,x_1)!$$

and so we recover

$$\nu_t^\pi = \sum_{x_0,x_1 \in \Omega_n} \pi(x_0,x_1) \sum_{z \in [\![x_0,x_1]\!]} t^{d(x_0,z)}(1-t)^{d(z,x_1)}\delta_z$$

Marton's weak W_2 distance

Let $\nu_0, \nu_1 \in \mathcal{P}(V)$.

Marton's weak W_2 distance

Let $\nu_0, \nu_1 \in \mathcal{P}(V)$.
If $\pi \in P(\nu_0, \nu_1)$, consider its conditional disintegration w.r.t the first coordinate:

$$\pi(dx_0 dx_1) = \nu_0(dx_0)p(x_o, dx_1),$$

where $p : V \to \mathcal{P}(V)$ is a probability kernel.

Marton's weak W_2 distance

Let $\nu_0, \nu_1 \in \mathcal{P}(V)$.
If $\pi \in P(\nu_0, \nu_1)$, consider its conditional disintegration w.r.t the first coordinate:

$$\pi(dx_0 dx_1) = \nu_0(dx_0) p(x_o, dx_1),$$

where $p : V \to \mathcal{P}(V)$ is a probability kernel.

Definition [Marton's transport cost]

$$\mathcal{T}_2(\nu_1|\nu_0) = \inf_{\pi \in P(\nu_0,\nu_1)} \int \left(\int d(x_0, x_1)\, p(x_0, dx_1) \right)^2 \nu_0(dx_0).$$

Marton's weak W_2 distance

Let $\nu_0, \nu_1 \in \mathcal{P}(V)$.
If $\pi \in P(\nu_0, \nu_1)$, consider its conditional disintegration w.r.t the first coordinate:

$$\pi(dx_0 dx_1) = \nu_0(dx_0) p(x_o, dx_1),$$

where $p : V \to \mathcal{P}(V)$ is a probability kernel.

Definition [Marton's transport cost]

$$\mathcal{T}_2(\nu_1|\nu_0) = \inf_{\pi \in P(\nu_0,\nu_1)} \int \left(\int d(x_0, x_1)\, p(x_0, dx_1) \right)^2 \nu_0(dx_0).$$

By Jensen,

$$W_2^2(\nu_0, \nu_1) \geq \mathcal{T}_2(\nu_1|\nu_0).$$

Marton's weak W_2 distance

Let $\nu_0, \nu_1 \in \mathcal{P}(V)$.
If $\pi \in P(\nu_0, \nu_1)$, consider its conditional disintegration w.r.t the first coordinate:

$$\pi(dx_0 dx_1) = \nu_0(dx_0) p(x_o, dx_1),$$

where $p : V \to \mathcal{P}(V)$ is a probability kernel.

Definition [Marton's transport cost]

$$\mathcal{T}_2(\nu_1|\nu_0) = \inf_{\pi \in P(\nu_0,\nu_1)} \int \left(\int d(x_0, x_1)\, p(x_0, dx_1) \right)^2 \nu_0(dx_0).$$

By Jensen,

$$W_2^2(\nu_0, \nu_1) \geq \mathcal{T}_2(\nu_1|\nu_0).$$

Define

$$\widetilde{W}_2^2(\nu_0, \nu_1) = \mathcal{T}_2(\nu_0|\nu_1) + \mathcal{T}_2(\nu_1|\nu_0).$$

Marton's weak W_2 distance

Proposition

On any space $\mathcal{X}$ equipped with the trivial distance $\mathbb{1}_{x\neq y}$, the following holds

$$\mathcal{T}_2(\nu_1|\nu_0) = \int \left[1 - \frac{f_1}{f_0}\right]_+^2 f_0 \, d\mu,$$

where $\nu_0 = f_0 \, \mu$, $\nu_1 = f_1 \, \mu$.

This formula will be useful on the two point space $\{0, 1\}$ or more generally on the complete graph K_n.

The two point space

On the two point space

$$\nu_t^\pi = (1-t)\nu_0 + t\nu_1$$

whatever the coupling π is.

The two point space

On the two point space

$$\nu_t^\pi = (1-t)\nu_0 + t\nu_1$$

whatever the coupling π is.

Differentiating twice the function $H(t) := H((1-t)\nu_0 + t\nu_1|\mu)$ it is easy to check that

$$H''(t) \geq \int \left[1 - \frac{f_1}{f_0}\right]_+^2 f_0\, d\mu + \int \left[1 - \frac{f_0}{f_1}\right]_+^2 f_1\, d\mu = \widetilde{W}_2^2(\nu_0, \nu_1).$$

The two point space

On the two point space

$$\nu_t^\pi = (1-t)\nu_0 + t\nu_1$$

whatever the coupling π is.

Differentiating twice the function $H(t) := H((1-t)\nu_0 + t\nu_1|\mu)$ it is easy to check that

$$H''(t) \geq \int \left[1 - \frac{f_1}{f_0}\right]_+^2 f_0 \, d\mu + \int \left[1 - \frac{f_0}{f_1}\right]_+^2 f_1 \, d\mu = \widetilde{W}_2^2(\nu_0, \nu_1).$$

So

Proposition

For all probability measure μ on $\Omega_1 = \{0,1\}$, it holds

$$H(\nu_t^\pi|\mu) \leq (1-t)H(\nu_0|\mu) + tH(\nu_1|\mu) - \frac{t(1-t)}{2}\widetilde{W}_2^2(\nu_0, \nu_1).$$

The two point space

On the two point space

$$\nu_t^\pi = (1-t)\nu_0 + t\nu_1$$

whatever the coupling π is.

Differentiating twice the function $H(t) := H((1-t)\nu_0 + t\nu_1|\mu)$ it is easy to check that

$$H''(t) \geq \int \left[1 - \frac{f_1}{f_0}\right]_+^2 f_0 \, d\mu + \int \left[1 - \frac{f_0}{f_1}\right]_+^2 f_1 \, d\mu = \widetilde{W}_2^2(\nu_0, \nu_1).$$

So

Proposition

For all probability measure μ on $\Omega_1 = \{0,1\}$, it holds

$$H(\nu_t^\pi|\mu) \leq (1-t)H(\nu_0|\mu) + tH(\nu_1|\mu) - \frac{t(1-t)}{2}\widetilde{W}_2^2(\nu_0, \nu_1).$$

To get the result for the hypercube Ω_n, we use a general tensorisation technique.

Tensorisation

Theorem [GRST]

Assume that a probability μ on V verifies the following property: There is $K \geq 0$ such that for all $\nu_0, \nu_1 \in \mathcal{P}(V)$, there is some $\pi \in P(\nu_0, \nu_1)$ such that for all $t \in [0,1]$, it holds

$$H(\nu_t^\pi | \mu) \leq (1-t) H(\nu_0|\mu) + tH(\nu_1|\mu) - K \frac{t(1-t)}{2} \widetilde{W}_2^2(\nu_0, \nu_1).$$

Tensorisation

Theorem [GRST]

Assume that a probability μ on V verifies the following property: There is $K \geq 0$ such that for all $\nu_0, \nu_1 \in \mathcal{P}(V)$, there is some $\pi \in P(\nu_0, \nu_1)$ such that for all $t \in [0,1]$, it holds

$$H(\nu_t^\pi|\mu) \leq (1-t)H(\nu_0|\mu) + tH(\nu_1|\mu) - K\frac{t(1-t)}{2}\widetilde{W}_2^2(\nu_0, \nu_1).$$

Then the product measure μ^n, verifies the following property: for all $\nu_0, \nu_1 \in \mathcal{P}(V^n)$, there is some $\hat{\pi} \in P(\nu_0, \nu_1)$ such that for all $t \in [0,1]$, it holds

$$H(\nu_t^{\hat{\pi}}|\mu^n) \leq (1-t)H(\nu_0|\mu^n) + tH(\nu_1|\mu^n) - K\frac{t(1-t)}{2}\widetilde{W}_{2,n}^2(\nu_0, \nu_1),$$

with the same constant K as above, where $\widetilde{W}_{2,n}$ is the tensorised Marton's distance.

Tensorisation

Ingredients of the proof:

Tensorisation

Ingredients of the proof:

1. **Tensorisation of the pseudo W_1 geodesic.**
   Denoting by $\nu_t^{x,y}$ the W_1-geodesic connecting δ_x to δ_y, $x, y \in V^n$. Then

$$\nu_t^{x,y} = \bigotimes_{i=1}^{n} \nu_t^{x_i,y_i}.$$

   Strongly connected to the choice of the Binomial distributions.

Tensorisation

Ingredients of the proof:

1. **Tensorisation of the pseudo W_1 geodesic.**
   Denoting by $\nu_t^{x,y}$ the W_1-geodesic connecting δ_x to δ_y, $x, y \in V^n$. Then

$$\nu_t^{x,y} = \bigotimes_{i=1}^{n} \nu_t^{x_i,y_i}.$$

   Strongly connected to the choice of the Binomial distributions.

2. **Knothe-Rosenblatt construction.**
   $(n = 2)$ Write

$$\nu_0(dx_1 dx_2) = \nu_0^1(dx_1)\nu_0^2(dx_2|x_1) \quad \text{and} \quad \nu_1(dy_1 dy_2) = \nu_1^1(dy_1)\nu_1^2(dy_2|y_1).$$

Tensorisation

Ingredients of the proof:

1. **Tensorisation of the pseudo W_1 geodesic.**
   Denoting by $\nu_t^{x,y}$ the W_1-geodesic connecting δ_x to δ_y, $x, y \in V^n$. Then
   $$\nu_t^{x,y} = \bigotimes_{i=1}^{n} \nu_t^{x_i,y_i}.$$
   Strongly connected to the choice of the Binomial distributions.
2. **Knothe-Rosenblatt construction.**
   $(n = 2)$ Write
   $$\nu_0(dx_1 dx_2) = \nu_0^1(dx_1)\nu_0^2(dx_2|x_1) \quad \text{and} \quad \nu_1(dy_1 dy_2) = \nu_1^1(dy_1)\nu_1^2(dy_2|y_1).$$
   Suppose that
   $$\pi_1 \in P(\nu_0^1, \nu_1^1) \quad \text{and} \quad \pi_2(\,\cdot\,|x_1, y_1) \in P(\nu_0^2(\,\cdot\,|x_1), \nu_1^2(\,\cdot\,|y_1)).$$

Ingredients of the proof:

1. **Tensorisation of the pseudo W_1 geodesic.**
   Denoting by $\nu_t^{x,y}$ the W_1-geodesic connecting δ_x to δ_y, $x, y \in V^n$. Then

$$\nu_t^{x,y} = \bigotimes_{i=1}^{n} \nu_t^{x_i,y_i}.$$

   Strongly connected to the choice of the Binomial distributions.

2. **Knothe-Rosenblatt construction.**
   $(n = 2)$ Write

$$\nu_0(dx_1 dx_2) = \nu_0^1(dx_1)\nu_0^2(dx_2|x_1) \quad \text{and} \quad \nu_1(dy_1 dy_2) = \nu_1^1(dy_1)\nu_1^2(dy_2|y_1).$$

   Suppose that

$$\pi_1 \in P(\nu_0^1, \nu_1^1) \quad \text{and} \quad \pi_2(\,\cdot\,|x_1, y_1) \in P(\nu_0^2(\,\cdot\,|x_1), \nu_1^2(\,\cdot\,|y_1)).$$

   Then $\hat{\pi}$ is obtained as follows

$$\hat{\pi}(dx_1 dx_2 dy_1 dy_2) = \pi_1(dx_1 dy_1)\pi_2(dx_2 dy_2|x_1, y_1) \in P(\nu_0, \nu_1).$$

Consequences in terms of functional inequalities

Theorem [GRST]

Suppose that a probability μ on V verifies the following property: $\exists K \geq 0$ such that $\forall \nu_0, \nu_1 \in \mathcal{P}(V)$, $\exists \pi \in P(\nu_0, \nu_1)$ such that $\forall t \in [0,1]$, it holds

$$H(\nu_t^\pi|\mu) \leq (1-t)H(\nu_0|\mu) + tH(\nu_1|\mu) - K\frac{t(1-t)}{2}(I_2(\pi) + \bar{I}_2(\pi)).$$

Then, for all $n \in \mathbb{N}^*$, the following *HWI* inequality holds: for all $\nu_0 \in \mathcal{P}(V^n)$, $\exists \pi \in P(\nu_0, \mu)$,

$$H(\nu_0|\mu) \leq \sqrt{\sum_{x \in V^n} \sum_{i=1}^{n} \left(\sum_{z \in N_i(x)} \left[\log \frac{\nu_0(x)}{\mu(x)} - \log \frac{\nu_0(z)}{\mu(z)}\right]_+ \right)^2 \nu_0(x)} \sqrt{I_2^{(n)}(\pi)} - \frac{K}{2}(I_2^{(n)}(\pi) + \bar{I}_2^{(n)}(\pi)),$$

where $N_i(x) = \{y \in V^n ; x_i \neq y_i \text{ and } d(x,y) = 1\}$.

A weak Log-Sobolev

Theorem [GRST]

Suppose that a probability μ on V verifies the following property: $\exists K \geq 0$ such that $\forall \nu_0, \nu_1 \in \mathcal{P}(V)$, $\exists \pi \in P(\nu_0, \nu_1)$ such that $\forall t \in [0,1]$, it holds

$$H(\nu_t^\pi|\mu) \leq (1-t)H(\nu_0|\mu) + tH(\nu_1|\mu) - K\frac{t(1-t)}{2}\widetilde{W}_2^2(\nu_0,\nu_1).$$

Then, for all $n \in \mathbb{N}^*$, the following *HWI* inequality holds: for all $\nu_0 \in \mathcal{P}(V^n)$,

$$H(\nu_0|\mu) \leq \frac{2}{K}\sum_{x\in V^n}\sum_{i=1}^{n}\left(\sum_{z\in N_i(x)}\left[\log\frac{\nu_0(x)}{\mu(x)} - \log\frac{\nu_0(z)}{\mu(z)}\right]_+\right)^2 \nu_0(x)$$

$\rightsquigarrow$ This LSI applied to the hypercube gives back Log-Sobolev for the Gaussian measure using the central limit theorem.

Discrete forms of the Prekopa-Leindler inequality

Theorem [GRST]

Suppose that a probability μ on V verifies the following property: $\exists K \geq 0$ such that $\forall \nu_0, \nu_1 \in \mathcal{P}(V)$, $\exists \pi \in P(\nu_0, \nu_1)$ such that $\forall t \in [0,1]$, it holds

$$H(\nu_t^\pi|\mu) \leq (1-t)H(\nu_0|\mu) + tH(\nu_1|\mu) - K\frac{t(1-t)}{2}\widetilde{W}_2^2(\nu_0, \nu_1).$$

Discrete forms of the Prekopa-Leindler inequality

Theorem [GRST]

Suppose that a probability μ on V verifies the following property: $\exists K \geq 0$ such that $\forall \nu_0, \nu_1 \in \mathcal{P}(V)$, $\exists \pi \in P(\nu_0, \nu_1)$ such that $\forall t \in [0,1]$, it holds

$$H(\nu_t^\pi|\mu) \leq (1-t)H(\nu_0|\mu) + tH(\nu_1|\mu) - K\frac{t(1-t)}{2}\widetilde{W}_2^2(\nu_0,\nu_1).$$

Then, for all $n \in \mathbb{N}^*$ and for all triple of functions $f, g, h : V^n \to \mathbf{R}$ such that for some $t \in (0,1)$ and for all $m \in \mathcal{P}(V^n)$,

$$\iint h(z)\,\nu_t^{x,y}(dz)\,m(dy) \geq (1-t)f(x) + t\int g(y)\,m(dy) - \frac{Kt(1-t)}{2}\sum_{i=1}^{n}\left(\int d(x_i,y_i)\,m(dy)\right)^2, \qquad \forall x \in V^n,$$

Discrete forms of the Prekopa-Leindler inequality

Theorem [GRST]

Suppose that a probability μ on V verifies the following property: $\exists K \geq 0$ such that $\forall \nu_0, \nu_1 \in \mathcal{P}(V)$, $\exists \pi \in P(\nu_0, \nu_1)$ such that $\forall t \in [0,1]$, it holds

$$H(\nu_t^\pi | \mu) \leq (1-t) H(\nu_0|\mu) + t H(\nu_1|\mu) - K \frac{t(1-t)}{2} \widetilde{W}_2^2(\nu_0, \nu_1).$$

Then, for all $n \in \mathbb{N}^*$ and for all triple of functions $f, g, h : V^n \to \mathbf{R}$ such that for some $t \in (0,1)$ and for all $m \in \mathcal{P}(V^n)$,

$$\iint h(z)\, \nu_t^{x,y}(dz)\, m(dy) \geq (1-t) f(x) + t \int g(y)\, m(dy) - \frac{Kt(1-t)}{2} \sum_{i=1}^n \left(\int d(x_i, y_i)\, m(dy) \right)^2, \qquad \forall x \in V^n,$$

it holds

$$\int e^{h(z)}\, \mu^n(dz) \geq \left(\int e^{f(x)}\, \mu^n(dx) \right)^{1-t} \left(\int e^{g(y)}\, \mu^n(dy) \right)^t.$$

The proof follows easily from the following general duality formula

$$\log\left(\int e^h\, d\mu\right) = \sup_{\nu\in\mathcal{P}(\mathcal{X})}\left\{\int h\, d\nu - H(\nu|\mu)\right\}.$$

Consequences: PL implies a form of the log-Sobolev inequality and a transport-entropy inequality involving Marton's transport costs.

Thank you for your attention !

Congratulations to Alain !