# A CURVED BRUNN–MINKOWSKI INEQUALITY ON THE DISCRETE HYPERCUBE
## OR: WHAT IS THE RICCI CURVATURE OF THE DISCRETE HYPERCUBE?

Y. OLLIVIER AND C. VILLANI

Abstract. We compare two approaches to Ricci curvature on non-smooth spaces, in the case of the discrete hypercube $\{0,1\}^N$. While the coarse Ricci curvature of the first author readily yields a positive value for curvature, the displacement convexity property of Lott, Sturm and the second author could not be fully implemented. Yet along the way we get new results of a combinatorial and probabilistic nature, including a curved Brunn–Minkowski inequality on the discrete hypercube.

## Introduction

Let $A_0, A_1$ be two compact, nonempty subsets of $\mathbb{R}^n$. In one of its guises, the remarkable Brunn–Minkowski inequality states that

$$\ln \operatorname{vol} A_t \geqslant (1-t) \ln \operatorname{vol} A_0 + t \ln \operatorname{vol} A_1$$

where $0 \leqslant t \leqslant 1$ and $A_t = \{(1-t)a_0 + ta_1,\ a_0 \in A_0,\ a_1 \in A_1\}$ is the set of $t$-midpoints between $A_0$ and $A_1$. In other words, the logarithm of the volume of $A_t$ is concave. We refer to [Gar02] for a nice survey. This is the “infinite-dimensional” version of the Brunn–Minkowski inequality, from which the more common version using $1/n$-th powers instead of logarithms can be derived (see Eq. (22) in [Gar02]).

If $\mathbb{R}^n$ is replaced with a Riemannian manifold, the presence of positive curvature *improves* this inequality. Indeed, in [CMS06] (elaborating on [CMS01]) it is proved that if $X$ is a smooth and complete Riemannian manifold with Ricci curvature at least $K$ for some $K \in \mathbb{R}$, then for any two compact, nonempty subsets $A_0, A_1 \subset X$, we have

$$\ln \operatorname{vol} A_t \geqslant (1-t) \ln \operatorname{vol} A_0 + t \ln \operatorname{vol} A_1 + \frac{K}{2} t(1-t)\, d(A_0, A_1)^2.$$

Here the set of $t$-midpoints $A_t$ is defined as the set of all $\gamma(t)$ where $\gamma$ is any minimizing geodesic such that $\gamma(0) \in A_0$ and $\gamma(1) \in A_1$. The distance $d(A_0, A_1)$ is $\inf_{a_0 \in A_0,\, a_1 \in A_1} d(a_0, a_1)$.

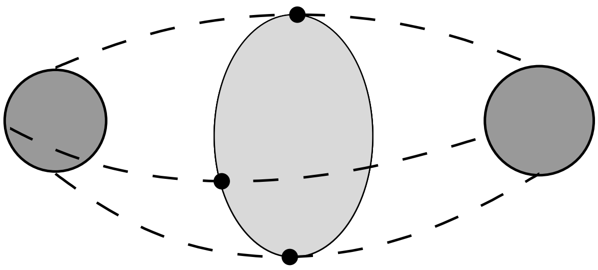

FIGURE 1. In positive curvature, midpoints spread out.

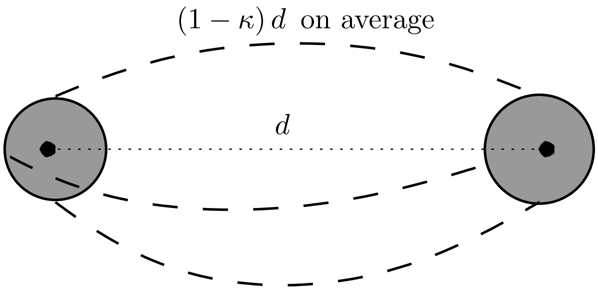


FIGURE 2. In positive curvature, balls are closer than their centers.

Actually this kind of inequality has been used as a tentative *definition* of positive Ricci curvature on more general, non-smooth spaces. The idea is that, in positive curvature, "midpoints spread out" so that the set of midpoints of two given sets is larger than in the reference Euclidean case (Fig. 1). This led to the notion of *displacement convexity of entropy* for Riemannian manifolds [RS05, CMS01, OV00], later developed by Sturm [Stu06] and Lott and the second author [LV09]. However, it is not clear how this fares for discrete spaces [BS09].

Another approach to define the Ricci curvature of discrete spaces is *coarse Ricci curvature*, developed by the first author [Oll07, Oll09]. The motto is that, in positive curvature, "balls are closer than their centers are" in transportation distance (Fig. 2).

We compare both approaches applied to the discrete hypercube $X = \{0,1\}^N$. This is the most simple discrete space expected to have positive Ricci curvature in some sense, for a variety of reasons (see, e.g., paragraph $3\frac{1}{2}.21$ "Spheres, cubes, and the law of large numbers" in [Gro99]). The subtitle question "What is the Ricci curvature of the discrete hypercube?" was asked verbatim by Stroock in a seminar as early as 1998, in a context of logarithmic Sobolev inequalities.

The formalism of coarse Ricci curvature is readily available for the hypercube and yields a value of $\frac{2}{N+1}$ for the Ricci curvature of $\{0,1\}^N$ (section 2.1). On the other hand, we could not fully implement the displacement convexity of entropy (properly discretized) in the hypercube. Yet, along the way, we still get a combinatorial *Brunn–Minkowski inequality on the hypercube*, including a positive curvature term. The resulting value of curvature is $\approx 1/N$, compatible with coarse Ricci curvature.

**Acknowledgements:** The authors would like to thank Prasad Tetali for helpful comments on concentration in the symmetric group, which led to improved constants.

## 1. Statement of results

### 1.1. **Brunn–Minkowski inequality in the hypercube.**

We consider the discrete hypercube $X := \{0,1\}^N$, $N \in \mathbb{N}$, equipped with the Hamming (or $\ell^1$) metric

$$d((x_i),(y_i)) := \#\{i,\ x_i \neq y_i\}.$$

For $A$ and $B$ nonempty subsets of $X$, we define $d(A,B) := \inf_{a\in A, b\in B} d(a,b)$.

Let $a$ and $b$ be two points in $X$. A *midpoint* of $a$ and $b$ is any point $m$ such that $d(m,a) + d(m,b) = d(a,b)$ and $|d(m,a) - d(a,b)/2| < 1$. More explicitly: if $d(a,b)$ is even, a midpoint is the middle point on any shortest path from $a$ to $b$ in $X$, and if $d(a,b)$ is odd, a midpoint is one the two middlemost points on such a shortest path. In the hypercube, midpoints are by no means unique: the number of midpoints of $a$ and $b$ is the binomial coefficient $\binom{d(a,b)}{d(a,b)/2}$ if $d(a,b)$ is even, and $2\binom{d(a,b)}{(d(a,b)-1)/2}$ if $d(a,b)$ is odd.

If $A$ and $B$ are two subsets of $X$, the set of midpoints of $A$ and $B$ is the set of midpoints of all pairs $(a,b) \in A \times B$.

**Theorem 1.** *Let $A$ and $B$ be two nonempty subsets of $\{0,1\}^N$. Let $M$ be the set of midpoints of $A$ and $B$. Then*

$$\ln \#M \geqslant \frac{1}{2} \ln \#A + \frac{1}{2} \ln \#B + \frac{K}{8}\, d(A,B)^2$$

*with $K = \frac{1}{2N}$.*

This is analogous to the curved Brunn–Minkowski inequality above in Riemannian manifolds (for $t = 1/2$), with $K$ playing the role of a curvature lower bound.

The order of magnitude $\frac{1}{N}$ for $K$ is optimal: indeed, when $A$ and $B$ are singletons lying at distance $N$, then $d(A,B)^2 = N^2$, while the number of midpoints is $\binom{N}{N/2} \sim 2^N\sqrt{\frac{2}{\pi N}}$, so that $\ln \#M$ grows linearly in $N$.

We will now see that this theorem can be improved by replacing $d(A, B)$ with a transportation distance.

1.2. **Entropy of midpoints in the hypercube.** Theorem 1 appears as a particular case of a refined statement using probability measures instead of sets.

Let $\mu$ be a probability measure on a discrete set $X$. Its *Shannon entropy* is

$$S(\mu) := -\sum_{x\in X} \mu(x) \ln \mu(x).$$

In particular, if $\mu$ is the uniform distribution on a finite subset $A \subset X$, then $S(\mu) = \ln \#A$.

In this paper, we shall also use the *relative entropy* (or Kullback–Leibler divergence) of a measure $\mu$ with respect to a reference probability measure $\nu$, defined as

$$H(\mu|\nu) := \sum_{x\in X} \mu(x) \ln \frac{\mu(x)}{\nu(x)} \geqslant 0.$$

If $X$ is finite and the reference measure $\nu$ is uniform on $X$, then we have $H(\mu|\nu) = \ln \#X - S(\mu)$.

To state an entropic version of Theorem 1 we define the midpoints of two measures as follows. Loosely speaking, we first pick a random point $a$ under $\mu_0$, then an *independent* random point $b$ under $\mu_1$, and finally we pick a random midpoint of $a$ and $b$ uniformly over all such midpoints.

More precisely, let $a$ and $b$ be two points of the hypercube $X$. The *midpoint measure* $mid(a, b)$ is defined as the uniform probability measure on all midpoints of $a$ and $b$. Let now $\mu_0, \mu_1$ be two probability measures on $X$. The *midpoint measure* of $\mu_0$ and $\mu_1$ is defined as

$$mid(\mu_0, \mu_1) := \iint mid(a, b)\, \mathrm{d}\mu_0(a)\mathrm{d}\mu_1(b).$$

**Theorem 2.** *Let $\mu_0$ and $\mu_1$ be two probability measures on the discrete hypercube $X = \{0, 1\}^N$. Let $\mu_{1/2} = mid(\mu_0, \mu_1)$ be their midpoint measure. Then*

$$S(\mu_{1/2}) \geqslant \frac{1}{2} \left(S(\mu_0) + S(\mu_1)\right) + \frac{K}{8}\, W_1(\mu_0, \mu_1)^2$$

*with $K = \frac{1}{2N}$. Equivalently,*

$$H(\mu_{1/2}|\nu) \leqslant \frac{1}{2} \left(S(\mu_0|\nu) + S(\mu_1|\nu)\right) - \frac{K}{8}\, W_1(\mu_0, \mu_1)^2$$

*with $\nu$ the uniform probability measure on $\{0, 1\}^N$.*

Here we use the $L^1$ *Wasserstein distance*

$$W_1(\mu, \mu') := \inf_{\xi} \iint d(a,b) \, \mathrm{d}\xi(a,b)$$

where the infimum is taken over all measures $\xi$ on $X \times X$ such that $\int_b \mathrm{d}\xi(a,b) = \mathrm{d}\mu(a)$ and $\int_a \mathrm{d}\xi(a,b) = \mathrm{d}\mu'(b)$, i.e., all couplings of $\mu$ and $\mu'$. We refer to [Vil03] for more background on this topic.

Note that $W_1(\mu_0, \mu_1)$ is always at least $d(A, B)$ for $\mu_0$ and $\mu_1$ supported in sets $A$ and $B$; in particular, if $\mu_0$ and $\mu_1$ are taken uniform in $A$ and $B$, Theorem 2 is really a refinement of Theorem 1.

1.3. **Limitations and open questions.** A first limitation of these results is the necessity to take $t = 1/2$. This comes from the combinatorial nature of our proof, which, for the most basic situation $K = 0$, consists in building an injection from $A \times B$ into $M \times M$.

This can probably be circumvented if we assume that the sets $A$ and $B$ are convex (i.e. the midpoint of two points in $A$ lies in $A$, and likewise for $B$): then, we can describe $t$-midpoints of $A$ and $B$ as iterated $1/2$-midpoints. (If $A$ or $B$ are not convex, iterating only yields midpoints of several points in $A$ and several points in $B$, which is not what we want.)

The injection from $A \times B$ into $M \times M$ used in our proof very naturally extends to an injection from $A \times B$ into $M_t \times M_{(1-t)}$, with $M_t$ the set of $t$-midpoints. This leads to a lower bound for $\ln \#M_t + \ln \#M_{(1-t)}$ in terms of $\ln \#A + \ln \#B$ plus a curvature term. This also holds in the Riemannian case (by adding the Brunn–Minkowski inequality for $t$ and for $(1-t)$). We do not know if there is a particular interpretation of this inequality.

Our initial goal was to prove that the discrete hypercube has positive Ricci curvature in the sense of Lott, Sturm and the second author, i.e., that the hypercube satisfies displacement convexity of entropy (see below). The main difference with our result is that, in the Brunn–Minkowski inequality, we consider all midpoints of all pairs of points $(a, b)$ with law $\mu_0 \otimes \mu_1$; whereas for displacement convexity, one should first choose an optimal coupling between $\mu_0$ and $\mu_1$ and then only consider the midpoints of those pairs $(a, b)$ that make up the optimal coupling. The two properties coincide only when $\mu_0$ is a Dirac measure, in which case our result is related to Sturm's *measure contraction property* [Stu06].

So as far as we know, the problem of computing the Ricci curvature of the hypercube using the displacement convexity approach is still open.

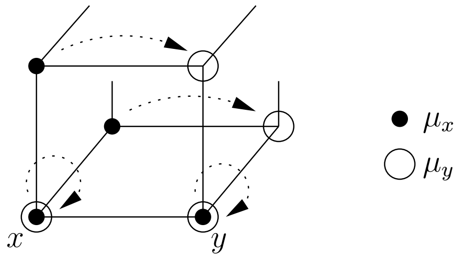


FIGURE 3. Coarse Ricci curvature in the hypercube.

## 2. Two approaches to discrete Ricci curvature

We now present in more detail the two known approaches for Ricci curvature on discrete spaces. This is not necessary to understand our results and proofs, but provides the original motivation.

2.1. **Coarse Ricci curvature (after the first author).** The basic idea of coarse Ricci curvature is to take two small balls and compute the transportation distance between them. If this distance is smaller than the distance between the centers of the balls, then coarse Ricci curvature is positive.

This is formalized as follows [Oll07, Oll09]. Let $(X, d)$ be a metric space equipped with a measure $\mu$. Let $\varepsilon$ be a discretization parameter (we take $\varepsilon = 1$ for a graph) and assume that all $\varepsilon$-balls in $X$ have finite and non-zero measure. For $x \in X$ define the measure $\mu_x$ by restricting $\mu$ to the closed $\varepsilon$-ball around $x$:

$$\mu_x := \frac{\mu_{|B(x,\varepsilon)}}{\mu(B(x,\varepsilon))}$$

with $B(x, \varepsilon) = \{y \in X, d(x, y) \leqslant \varepsilon\}$.

If $x$ and $y$ are two points in $X$, then the *coarse Ricci curvature along* $(x, y)$ is the number $\kappa(x, y)$ defined by

$$W_1(\mu_x, \mu_y) =: (1 - \kappa(x, y))\, d(x, y)$$

where $W_1$ is the $L^1$ Wasserstein distance as defined earlier. If this is applied to a Riemannian manifold, this gives back the ordinary Ricci curvature when $\varepsilon \to 0$, up to scaling by $\varepsilon^2$.

Let us apply this to the discrete hypercube $X = \{0, 1\}^N$ equipped with the uniform measure. The measure $\mu_x$ is uniform on the $N + 1$ neighbors of $x$ (counting $x$ itself). When $x$ and $y$ are neighbors, it is very easy to compute the curvature $\kappa(x, y)$, as illustrated on Figure 3. Indeed, we have to move the $N + 1$ neighbors of $x$ to the

$N+1$ neighbors of $y$; out of these $N+1$ points, two are already in place ($x$ and $y$ themselves) and do not need to move, and the others have to move by a distance 1. So $W_1(\mu_x, \mu_y) = 1 - 2/(N+1)$ and $\kappa(x,y) = 2/(N+1)$.

If $x$ and $y$ are not neighbors, we use a locality property of coarse Ricci curvature. Namely, if the space $X$ is $\delta$-geodesic (i.e. if the distance between two points is realized by a sequence of points with jumps at most $\delta$), then it is enough to compute $\kappa(x,y)$ for $d(x,y) \leqslant \delta$ (Exercise 2 in [Oll07]). A graph is 1-geodesic by definition of the graph metric, so it is enough to work with neighbors.

A lower bound on coarse Ricci curvature comes with a number of consequences [Oll09]. For the discrete hypercube equipped with the uniform measure these properties were already known (but not on the hypercube with e.g. Bernoulli($\theta/N$) measures [JO10]).

In general, one may directly choose an arbitrary Markov kernel $\mu_x$ (without using a global measure $\mu$); this leads to interesting applications [JO10].

2.2. **Displacement convexity (after Lott, Sturm and the second author).** In [RS05] (following ideas from [OV00]), Renesse and Sturm present a characterization of Ricci curvature on Riemannian manifolds, based on the idea that in positive curvature, "midpoints spread out".

Let $X$ be a smooth, complete Riemannian manifold. Let $\mathrm{d}x$ be the Riemannian volume measure on $X$. Given a probability measure $\mu$ on $X$, define its relative entropy as $H(\mu|\mathrm{d}x) := \int \ln \frac{\mathrm{d}\mu}{\mathrm{d}x}\,\mathrm{d}\mu$ if the integral makes sense, or $+\infty$ otherwise.

Let $\mathcal{P}^2(X)$ be the set of probability measures on $X$ with finite second moment, i.e. those probability measures $\mu$ such that $\int d(\mathrm{pt}, x)^2\,\mathrm{d}\mu(x) < \infty$ for some (hence any) point $\mathrm{pt} \in X$. On $\mathcal{P}^2(X)$, the Wasserstein distance $W_2$ is well-defined. Moreover, $\mathcal{P}^2(X)$ equipped with the metric $W_2$ is a geodesic space: given any two probability measures $\mu_0, \mu_1 \in \mathcal{P}^2(X)$, there exists a curve $(\mu_t)_{t\in(0;1)}$ in $\mathcal{P}^2(X)$ with $W_2(\mu_t, \mu_{t'}) = |t-t'|\, W_2(\mu_0, \mu_1)$ for $t, t' \in [0;1]$. Such a curve is called a *displacement interpolation* between $\mu_0$ and $\mu_1$. We refer to Chapter 7 of [Vil08] for more details.

Theorem 1.1 in [RS05] asserts that the Riemannian manifold $X$ has Ricci curvature at least $K \in \mathbb{R}$ if and only if the following inequality is satisfied: for any two measures $\mu_0, \mu_1 \in \mathcal{P}^2(X)$, for any $W_2$-geodesic $(\mu_t)_{t\in(0;1)}$ joining them, we have

$$H(\mu_t|\mathrm{d}x) \leqslant (1-t)H(\mu_0|\mathrm{d}x) + tH(\mu_1|\mathrm{d}x) - \frac{K}{2}t(1-t)W_2(\mu_0,\mu_1)^2,$$

a property called *displacement convexity* of the entropy function.

For any probability measure $\mu$ we have $H(\mu|\mathrm{d}x) \geqslant -\ln \operatorname{vol} \operatorname{Supp}(\mu)$, with equality when $\mu$ is uniform on its support. Taking $\mu_0$ and $\mu_1$ to be uniform probability distributions on sets $A_0$ and $A_1$ respectively, we see that displacement convexity of entropy implies an inequality between the logarithms of the volumes of the support of $\mu_t$, $\mu_0$ and $\mu_1$. This inequality is very similar to the Brunn–Minkowski inequality mentioned earlier. Actually, an important property of displacement interpolation is that the measure $\mu_t$ will charge only $t$-midpoints between the supports of $\mu_0$ and $\mu_1$ (Corollary 7.22 in [Vil08], basically due to Brenier and McCann), and so the Brunn–Minkowski inequality in a Riemannian manifold really follows from convexity of entropy.

Displacement convexity of entropy makes sense in an arbitrary geodesic space. In [Stu06, LV09], it is taken as the basis for a notion of Ricci curvature in such spaces. The definition depends on two parameters $K$ (the curvature) and $N$ (a "dimension"). Displacement convexity of entropy as written here corresponds to $N = \infty$, the simplest and weakest case.

Interestingly, this approach applies to spaces with positive curvature in the sense of Alexandrov [Pet].

Application to discrete spaces requires some changes: for instance, in the case of the hypercube considered in this article, clearly if two points are at odd distance they do not have an exact midpoint, but they have an approximate midpoint up to an error term $\pm 1/2$. Such an approach is used in [Bon09] to define the Brunn–Minkowski inequality on discrete spaces. In [BS09], Bonciocat and Sturm use approximate midpoints in the space of probability measures to extend the definition of displacement convexity of entropy to discrete spaces, and provide examples of planar graphs satisfying this property. To our knowledge, these planar graphs are the only discrete examples so far.

## 3. Brunn–Minkowski inequality without curvature

To make the idea clearer and introduce necessary concepts, we begin with a simplified version of Theorem 1, namely the same statement with $K = 0$. So let $A, B$ be two nonempty subsets of the hypercube $X = \{0,1\}^N$. Let $M$ be the set of midpoints of $A$ and $B$. We want to prove that

$$\ln \#M \geqslant \frac{1}{2} (\ln \#A + \ln \#B)$$

or equivalently

$$\#M \geqslant \sqrt{\#A \, \#B}.$$

Let $a = (a_i)_{1 \leqslant i \leqslant N} \in A$ and $b = (b_i)_{1 \leqslant i \leqslant N} \in B$. A midpoint $m = (m_i)$ of $a$ and $b$ is a sequence of bits such that $m_i = a_i$ whenever $a_i = b_i$ and such that half the remaining bits coincide with those of $a$ and the other half with those of $b$. Let $r = d(a, b)$ be the number of distinct bits between $a$ and $b$. For fixed $a$ and $b$, there is a one-to-one correspondence between the midpoints $m$ of $a$ and $b$ and the subsets $c \subset \{1, \dots, r\}$ with cardinality $r/2$ (if $r$ is even) or $r/2 \pm 1/2$ ($r$ odd): among the $r$ distinct bits between $a$ and $b$, the set $c$ describes those picked from $a$ in the construction of $m$.

We shall call *$r$-crossover* such a $c \subset \{1, \dots, r\}$ with $|\#c - r/2| \leqslant 1/2$. We shall denote $m = \varphi_c(a, b)$ the midpoint of $a$ and $b$ defined by crossover $c$. If $c$ is a crossover, we shall denote by $\bar{c}$ its complement, which is also a crossover.

Note that, given a fixed $d(a, b)$-crossover $c$, the pair $\Phi_c(a, b) := (\varphi_c(a, b), \varphi_{\bar{c}}(a, b)) = (m, m')$ allows to recover $a$ and $b$. Indeed, the identical bits in $m$ and $m'$ are the same as in $a$ and $b$; the bits that differ between $m$ and $m'$ also differ between $a$ and $b$, and knowledge of the crossover $c$ tells us exactly which of those come from $a$ or $b$.

In particular, for each $r \in \{0, \dots, N\}$, let us define the $r$-crossover $c_r := \{1, 2, \dots, \lfloor r/2 \rfloor\}$. Then the map $(a, b) \to \Phi_{c_{d(a,b)}}(a, b)$ is an injection from $A \times B$ to $M \times M$ where $M$ is the set of midpoints of $A$ and $B$. This proves that $\#(A \times B) \leqslant \#(M \times M)$ as needed.

For later use, let us state a property of the coding maps $\varphi_c$ and $\Phi_c$. If $\Phi_c(a, b) = (m, m')$, we denote $a = \varphi_c^{-1}(m, m')$ and $b = \varphi_{\bar{c}}^{-1}(m, m') = \varphi_c^{-1}(m', m)$.

Let us equip the set of crossovers $C_r$ with the distance

$$d(c, c') := \#(c \backslash c') + \#(c' \backslash c).$$

**Proposition 3** (Decoding is isometric)**.** *Let $m, m' \in \{0, 1\}^N$. Let $c_1, c_2 \in C_{d(m,m')}$. Let $a_1 = \varphi_{c_1}^{-1}(m, m')$ and $a_2 = \varphi_{c_2}^{-1}(m, m')$. Then $d(a_1, a_2) = d(c_1, c_2)$.*

*Proof.* Given $m$ and $m'$, modifying the crossover $c$ changes the preimage $\varphi_c^{-1}(m, m')$ by the same amount. □

## 4. Concentration in the set of crossovers

To get an improved inequality with positive curvature $K$, we will need to study geometric properties of the set of crossovers; more precisely we show that this set exhibits concentration of measure. This is obtained from the well-known concentration of measure in the permutation group by a quotienting argument. (We refer to [Led01] for more background about concentration of measure.) We first state concentration in the permutation group under the form we need.

**Lemma 4** (Concentration in $S_n$)**.** *Let $S_n$ be the permutation group on $\{1, \dots, n\}$. Equip $S_n$ with the distance $d(\sigma, \sigma') = \#\{i, \sigma(i) \neq \sigma'(i)\}$ for $\sigma, \sigma' \in S_n$. Let $\nu$ be the uniform probability measure on $S_n$.*

*Let $f : S_n \to \mathbb{R}$ be a 1-Lipschitz function. Then $f$ satisfies the concentration inequality*

$$\nu(\{f \geqslant \textstyle\int f \mathrm{d}\nu + t\}) \leqslant e^{-t^2/2(n-1)} \qquad \forall t \geqslant 0$$

*and the Laplace transform estimate*

$$\int e^{\lambda f} \mathrm{d}\nu \leqslant e^{\lambda \int f \mathrm{d}\nu + (n-1)\lambda^2/2} \qquad \forall \lambda \in \mathbb{R}.$$

*Proof.* The second statement is Proposition 6.1 in [BHT06]. The first statement follows by the exponential Markov inequality. □

**Proposition 5** (The set of crossovers is concentrated)**.** *Let $n \geqslant 1$ and let $C_n$ be the set of parts $c \subset \{1, \dots, n\}$ with $|\#c - n/2| < 1$. Equip $C_n$ with the distance $d(c, c') := \#(c \backslash c') + \#(c' \backslash c)$ as above and with the uniform probability measure $\mu$.*

*Let $f : C_n \to \mathbb{R}$ be a 1-Lipschitz function. Then $f$ satisfies the concentration inequality*

$$\mu(\{f \geqslant \textstyle\int f \mathrm{d}\mu + t\}) \leqslant e^{-t^2/2n} \qquad \forall t \geqslant 0$$

*and the Laplace transform estimate*

$$\int e^{\lambda f} \mathrm{d}\mu \leqslant e^{\lambda \int f \mathrm{d}\mu + n\lambda^2/2} \qquad \forall \lambda \in \mathbb{R}.$$

*Proof.* Let us begin with even $n$. Then the natural action of $S_n$ on $\{1, \dots, n\}$ preserves $C_n$. Let us fix an origin $c_0 := \{1, \dots, n/2\} \in C_n$ and define the projection map $\pi : S_n \to C_n$ by $\sigma \mapsto \sigma(c_0)$. Each fiber of $\pi$ has the same cardinality $((n/2)!)^2$. Moreover, if we equip $S_n$ and $C_n$ with the distances as above, then the map $\pi$ is 1-Lipschitz.

Thus, if $f : C_n \to \mathbb{R}$ is a 1-Lipschitz function, the function $\widetilde{f} := f \circ \pi$ is 1-Lipschitz on $S_n$. So $\widetilde{f}$ satisfies the concentration property $\nu(\{\widetilde{f} \geqslant \int \widetilde{f}\, d\nu + t\}) \leqslant e^{-t^2/2(r-1)}$ where $\nu$ is the uniform probability measure on $S_n$. Since all fibers of $\pi$ have the same cardinality, $\pi$ sends $\nu$ to the uniform measure $\mu$ and so the same estimate holds for $f$ in $C_n$ under $\mu$. The argument is identical for the Laplace transform estimate.

For odd $n$ we proceed as follows. Let us fix $c_0 = \{1, \dots, \lfloor n/2 \rfloor\} \in C_n$ and $c_1 = \{1, \dots, \lceil n/2 \rceil\} \in C_n$. Let us define the set $S_n^* := S_n \times \{0\} \sqcup S_n \times \{1\}$. Define the map $\pi : S_n^* \to C_n$ by $(\sigma, i) \mapsto \sigma(c_i)$ for $i = 0, 1$. Then each fiber of $\pi$ has the same cardinality $\lfloor n/2 \rfloor! \lceil n/2 \rceil!$. Let us equip $S_n^*$ with the metric $d((\sigma, i), (\sigma', i')) = |i - i'| + d(\sigma, \sigma')$. Then one checks that $\pi$ is 1-Lipschitz from $S_n^*$ to $C_n$. (A more elegant construction would have used $c \mapsto \bar{c}$ to get a group structure on $S_n^*$, but this has bad metric properties.)

Given a 1-Lipschitz function $f : C_n \to \mathbb{R}$, consider as above the function $\widetilde{f} := f \circ \pi$ on $S_n^*$. Applying, for instance, the technique of Theorem 4.2 in [Led01] to get concentration of measure in $S_n^*$ instead of $S_n$, we get that $\widetilde{f}$ satisfies the Laplace transform estimate

$$\int e^{\lambda \widetilde{f}}\, \mathrm{d}\nu \leqslant e^{\lambda \int \widetilde{f} \mathrm{d}\nu + (r-1)\lambda^2/2 + \lambda^2/8} \leqslant e^{\lambda \int \widetilde{f} \mathrm{d}\nu + r\lambda^2/2}$$

with $\nu$ the uniform probability measure on $S_n^*$. This implies that $\nu(\{\widetilde{f} \geqslant \int \widetilde{f}\, d\nu + t\}) \leqslant e^{-t^2/2r}$. Just as above, this estimate then holds for $f$ on $C_n$. □

**Corollary 6.** *Let $A$ be a subset of the set of crossovers $C_n$ and let $\bar{A} := \{\bar{c},\, c \in A\}$. Suppose that $d(A, \bar{A}) \geqslant k$. Then*

$$\#A \leqslant e^{-k^2/8n}\, \#C_n.$$

*Proof.* Consider the function $f : C_n \to \mathbb{R}$ given by $f(c) := \frac{1}{2}\left(d(c, A) - d(c, \bar{A})\right)$. This function is 1-Lipschitz, and takes values at least $k/2$ on $A$. By symmetry the average of $f$ is 0. So applying the above, we get that the (relative) measure of $A$ in $C_n$ is at most $e^{-k^2/8n}$. □

The following is a refined version of Corollary[1]6, in which the set $A$ is replaced with a measure $\xi$, cardinals are replaced with entropies, and the distance $d(A, \bar{A})$ is replaced with $W_1(\xi, \bar{\xi})$.

**Corollary 7.** *Let $\xi$ be a probability measure on the set of crossovers $C_n$. Let $\bar{\xi}$ be the complement of $\xi$ i.e. $\bar{\xi}(c) := \xi(\bar{c})$ for $c \in C_n$. Then*

$$S(\xi) \leqslant \ln \#C_n - \frac{1}{8n} W_1(\xi, \bar{\xi})^2$$

*with $S$ the Shannon entropy.*

*Proof.* The proof uses the following consequence of Proposition 5.

**Lemma 8** ($W_1 H$ inequality for crossovers)**.** *Let $\xi$ be a probability measure on $C_n$. Then*

$$W_1(\xi, \mu)^2 \leqslant 2nH(\xi|\mu)$$

*where $\mu$ is the uniform probability measure on $C_n$ and $H$ the relative entropy.*

Indeed, by a result of Bobkov and Götze (Theorem 3.1 in [BG99]), the inequality $W_1(\xi, \mu)^2 \leqslant 2\gamma H(\xi|\nu)$ for all measures $\xi$, is equivalent to the Laplace transform estimate $\int e^{\lambda f} \mathrm{d}\mu \leqslant e^{\lambda \int f \mathrm{d}\mu + \gamma\lambda^2/2}$ for all $\lambda \in \mathbb{R}$ and all 1-Lipschitz functions $f$. So the lemma is actually equivalent to Proposition 5.

Now, since $W_1(\xi,\bar\xi) \leqslant W_1(\xi,\mu) + W_1(\mu,\bar\xi) = 2W_1(\xi,\mu)$ by symmetry, we get

$$H(\xi|\mu) \geqslant \frac{1}{8n} W_1(\xi,\bar\xi)^2.$$

Finally, using $H(\xi|\mu) = \ln \#C_n - S(\xi)$, this rewrites in terms of the Shannon entropy as

$$S(\xi) \leqslant \ln \#C_n - \frac{1}{8n} W_1(\xi,\bar\xi)^2.$$

□

## 5. Positively curved Brunn–Minkowski inequality

Let us now prove Theorem 1. So let again $A, B$ be two nonempty subsets of the hypercube $X = \{0,1\}^N$, and let $M$ be the set of midpoints of $A$ and $B$. We have to prove that

$$\ln \#M \geqslant \frac{1}{2}(\ln \#A + \ln \#B) + \frac{K\, d(A,B)^2}{8}, \qquad K = \frac{1}{2N}.$$

The difference with the case $K = 0$ is that we now consider all crossovers at once. Let $C_r$ be the set of $r$-crossovers. Let $Y := \{(a,b,c),\ a \in A,\ b \in B,\ c \in C_{d(a,b)}\}$. Consider the map $f : (a,b,c) \mapsto \Phi_c(a,b)$ from $Y$ to $M \times M$. This map $f$ may not be one-to-one; but we will show that it is not too-many-to-one. The idea is that, given a pair of midpoints $(m,m')$, the geometry of $A$ and $B$ allows to guess, to some extent, which crossover was used, so that the cardinality of $f^{-1}(m,m')$ is bounded. (This is most clear when $A$ is a singleton $\{00\dots00\}$, in which case there is no ambiguity on the crossover: every '1' in $m$ or $m'$ was taken from $B$.)

Let $Y_r := \{(a,b,c) \in Y,\ d(a,b) = r\}$ and let likewise $(M \times M)_r := \{(m,m') \in M \times M,\ d(m,m') = r\}$. Now fix $(m,m') \in (M \times M)_r$. The fiber $f^{-1}(m,m')$ is in bijection with the set $E$ of crossovers $c \in C_r$ such that $\Phi_c^{-1}(m,m') \in A \times B$. Consider, symmetrically, the set $E' = \{c \in C_r,\ \Phi_c^{-1}(m,m') \in B \times A\}$. By definition $\Phi_c = (\varphi_c, \varphi_{\bar c})$, so the elements of $E'$ are the complements of the elements of $E$.

We claim that $d(E,E') \geqslant d(A,B)$. Indeed, if $c \in E$, $c' \in E'$ we have $\varphi_{c_1}^{-1}(m,m') \in A$ and $\varphi_{c'}^{-1}(m,m') \in B$. Since decoding is isometric (Proposition 3) we have $d(c,c') \geqslant d(A,B)$.

Corollary 6 then states that the cardinality of $E$ is at most $\#C_r e^{-d(A,B)^2/8r}$. Since the cardinality of $E$ is also the cardinality of the fiber $f^{-1}(m,m')$, this shows that the map $f : Y_r \to (M \times M)_r$ is at most $(\#C_r e^{-d(A,B)^2/8r})$-to-one. Consequently, $\#Y_r \leqslant \#C_r e^{-d(A,B)^2/8r}\, \#(M \times M)_r$.

Setting $(A\times B)_r := \{(a,b) \in A\times B,\ d(a,b) = r\}$, we have $\#Y_r = \#(A \times B)_r \times \#C_r$ so that

$$\#(M \times M)_r \geqslant e^{d(A,B)^2/8r}\,\#(A \times B)_r.$$

Finally, summing over $r$ from 1 to $N$ we find

$$\#(M \times M) \geqslant e^{d(A,B)^2/8N}\#(A \times B)$$

which proves Theorem 1.

## 6. Entropy of the set of midpoints

We now turn to the proof of Theorem 2.

Remember that, given $a$ and $b$ in the hypercube $X$, the midpoint measure $mid(a,b)$ is the uniform probability measure on all midpoints of $a$ and $b$. The midpoint measure of two probability measures $\mu_A$ and $\mu_B$ is defined as

$$mid(\mu_A, \mu_B) := \iint mid(a,b)\,\mathrm{d}\mu_A(a)\mathrm{d}\mu_B(b)$$

that is, the average of $mid(a,b)$ where $a$ and $b$ are taken *independently* at random under $\mu_A$ and $\mu_B$.

The proof follows the same lines as in the deterministic case, using probability measures instead of sets. The reader should think of the probability measures below as being nothing but weighted sets, and their Shannon entropy as being the logarithm of their cardinality. The main differences are as follows:

- In the set-theoretic version, a key point was an estimation of the cardinality of the fibers of the map $(a,b,c) \mapsto (m,m') = \Phi_c(a,b)$. The lower bound on the cardinality of the set $\{(m,m')\}$ followed. Here, we will use the associativity of Shannon entropy to express the same relationship, yielding a lower bound on the entropy of $(m,m')$ if the entropy of the fibers is known.
- The final result involves $W_1(\mu_A,\mu_B)$ instead of $d(A,B)$. In the set-theoretic version, we used the map $c \mapsto \bar{c}$ and the fact that $\Phi_c(a,b) = \Phi_{\bar{c}}(b,a)$ to conclude that, if $\Phi_c(a,b) = \Phi_{c'}(a',b')$ then $d(\bar{c},c') = d(b,a') \geqslant d(A,B)$. Then Corollary 6 was used to bound the cardinality of the set $E$ of such crossovers $c$ in a fiber. The refined version uses the relation $d(\bar{c},c') = d(b,a')$ to turn any coupling between $E$ and $\bar{E}$, into a coupling between $A$ and $B$ with the same transportation distance. Then, Corollary 7 is used as a refined version of Corollary 6 and yields a bound on the entropy of the crossovers $c$ in a fiber.

So let $a$ and $b$ be independent random variables with law $\mu_A$ and $\mu_B$. Let as above $C_r$ be the set of $r$-crossovers. Let $c$ be a random variable uniformly distributed on

$C_{d(a,b)}$, independent of $a$ and $b$ conditionally to $d(a,b)$. Let us define the random variables $m := \varphi_c(a,b)$ and $m' := \varphi_{\bar{c}}(a,b)$. Thus the law of $m$ is $mid(\mu_A, \mu_B)$, as is the law of $m'$.

Let us slightly abuse notation and denote by $S((y))$ the Shannon entropy of the law of a random variable $y$. We have $S((m,m')) \leqslant S((m)) + S((m'))$ but since $m$ and $m'$ have the same law $mid(\mu_A, \mu_B)$, we get

$$S(mid(\mu_A, \mu_B)) \geqslant \frac{1}{2} S((m,m')).$$

Consider as above the map $\Phi$ sending $(a,b,c)$ to $\Phi_c(a,b) = (m,m')$. Let $Y_{(m,m')}$ be the law of $(a,b,c)$ knowing $(m,m')$. By the associativity of entropy, the Shannon entropy of the law of $(m,m')$ is the entropy of the law of $(a,b,c)$ minus the average entropy of fibers of $\Phi$, namely:

$$S((m,m')) = S((a,b,c)) - \mathbb{E} S(Y_{(m,m')}).$$

The first term is computed as follows. The random variables $a$ and $b$ are independent, and, conditionally to $d(a,b)$, the variable $c$ is independent of $a$ and $b$ with law the uniform distribution $U_{d(a,b)}$ on $C_{d(a,b)}$. So

$$S((a,b,c)) = S((a)) + S((b)) + \mathbb{E} S(U_{d(a,b)}) = S(\mu_A) + S(\mu_B) + \mathbb{E} \ln \# C_{d(a,b)}.$$

Let us turn to the second term $\mathbb{E} S(Y_{(m,m')})$. This means we have to evaluate the entropy of the fibers of $\Phi$, as in the non-random case.

Let $E_{(m,m')}$ be the law of $c$ knowing $(m,m')$ (i.e., the third marginal of $Y_{(m,m')}$). Given $(m,m')$, the value of $c$ determines $a$ and $b$, and so, $S((a,b,c)|(m,m')) = S((c)|(m,m'))$ i.e.

$$S(Y_{(m,m')}) = S(E_{(m,m')})$$

so that

$$S((m,m')) = S(\mu_A) + S(\mu_B) + \mathbb{E} \ln \# C_{d(a,b)} - \mathbb{E} S(E_{(m,m')}).$$

If, at this point, we apply the crude estimate $S(E_{(m,m')}) \leqslant \ln \# C_{d(m,m')}$, we get $S((m,m')) \geqslant S(\mu_A) + S(\mu_B) + \mathbb{E} \ln \# C_{d(a,b)} - \mathbb{E} \ln \# C_{d(m,m')} = S(\mu_A) + S(\mu_B)$ since $d(a,b) = d(m,m')$. This implies $S((m)) \geqslant \frac{1}{2}(S(\mu_A) + S(\mu_B))$ i.e. the case $K = 0$ in the theorem.

As in the set-theoretic case, we will show that $E_{(m,m')}$ has small Shannon entropy by using concentration properties in the set of crossovers. Corollary 7 tells us that

$$S(E_{(m,m')}) \leqslant \ln \# C_{d(m,m')} - \frac{1}{8d(m,m')} W_1(E_{(m,m')}, \bar{E}_{(m,m')})^2$$

where $\bar{E}_{(m,m')}$ is the image of $E_{(m,m')}$ by $c \mapsto \bar{c}$. Thus, we need to evaluate the distance between $E_{(m,m')}$ and $\bar{E}_{(m,m')}$, as in the deterministic case.

Actually we only need an estimate on average over $(m, m')$. We claim that

$$\mathbb{E}W_1(E_{(m,m')}, \bar{E}_{(m,m')})^2 \geqslant W_1(\mu_A, \mu_B)^2.$$

Indeed, let us fix $(m, m')$ for now, and let $A_{(m,m')}$ and $B_{(m,m')}$ be the laws of $a$ and $b$ knowing $(m, m')$, respectively. Since $a = \varphi_c^{-1}(m, m')$ and $b = \varphi_{\bar{c}}^{-1}(m, m')$, any coupling between $E_{(m,m')}$ and $\bar{E}_{(m,m')}$ determines a coupling between $A_{(m,m')}$ and $B_{(m,m')}$. Moreover, since decoding is isometric by Proposition 3, these couplings will define the same transportation distance. So we get $W_1(A_{(m,m')}, B_{(m,m')}) \leqslant W_1(E_{(m,m')}, \bar{E}_{(m,m')})$.

If for each $(m, m')$ we are given a coupling between $A_{(m,m')}$ and $B_{(m,m')}$, by summation this defines a coupling between $\mu_A$ and $\mu_B$ and so $W_1(\mu_A, \mu_B) \leqslant \mathbb{E}W_1(A_{(m,m')}, B_{(m,m')})$. Thus $W_1(\mu_A, \mu_B) \leqslant \mathbb{E}W_1(E_{(m,m')}, \bar{E}_{(m,m')})$. Then, by convexity we get

$$W_1(\mu_A, \mu_B)^2 \leqslant \mathbb{E}W_1(E_{(m,m')}, \bar{E}_{(m,m')})^2$$

as announced.

Putting everything together and using that $d(m, m') = d(a, b)$, we get

$$\begin{aligned}
S((m, m')) &= S((a, b, c)) - \mathbb{E}S(Y_{(m,m')}) \\
&= S(\mu_A) + S(\mu_B) + \mathbb{E}\ln \#C_{d(a,b)} - \mathbb{E}S(E_{(m,m')}) \\
&\geqslant S(\mu_A) + S(\mu_B) + \mathbb{E}\ln \#C_{d(a,b)} - \mathbb{E}\ln \#C_{d(m,m')} + \mathbb{E}\left[\frac{W_1(E_{(m,m')}, \bar{E}_{(m,m')})^2}{8d(m, m')}\right] \\
&\geqslant S(\mu_A) + S(\mu_B) + \frac{1}{8N}\,\mathbb{E}W_1(E_{(m,m')}, \bar{E}_{(m,m')})^2 \\
&\geqslant S(\mu_A) + S(\mu_B) + \frac{1}{8N} W_1(\mu_A, \mu_B)^2
\end{aligned}$$

and so

$$S((m)) \geqslant \frac{1}{2}\left(S(\mu_A) + S(\mu_B)\right) + \frac{1}{16N} W_1(\mu_A, \mu_B)^2$$

which ends the proof.

## References

[BG99] S. Bobkov, F. Götze, *Exponential integrability and transportation cost related to logarithmic Sobolev inequalities*, J. Funct. Anal. **163** (1999), 1–28.

[BHT06] S. Bobkov, C. Houdré, P. Tetali, *The subgaussian constant and concentration inequalities*, Israel J. Math. **156** (2006), 255–283.

[Bon09] M. Bonnefont, *A discrete version of the Brunn–Minkowski inequality and its stability*, Ann. Math. Blaise Pascal **16** (2009), n° 2, 245–257.

[BS09] A.-I. Bonciocat, K.-T. Sturm, *Mass transportation and rough curvature bounds for discrete spaces*, J. Funct. Anal. **256** (2009), n° 9, 2944–2966.

[CMS01] D. Cordero-Erausquin, R.J. McCann, M. Schmuckenschläger, *A Riemannian interpolation inequality à la Borell, Brascamp and Lieb*, Invent. Math. **146** (2001), 219–257.

[CMS06] D. Cordero-Erausquin, R.J. McCann, M. Schmuckenschläger, *Prékopa-Leindler type inequalities on Riemannian manifolds, Jacobi fields and optimal transport*, Ann. Fac. Sci. Toulouse Math. (6) **15** (2006), n° 4, 613–635.

[Gar02] R. J. Gardner, *The Brunn–Minkowski inequality*, Bull. Amer. Math. Soc. (N.S.) **39** (2002), n° 3, 355–405.

[Gro99] M. Gromov, *Metric Structures for Riemannian and Non-Riemannian Spaces*, Progress in Math. **152**, Birkhäuser (1999).

[JO10] A. Joulin, Y. Ollivier, *Curvature, concentration, and error estimates for Markov chain Monte Carlo*, Ann. Probab. **38** (2010), n° 6, 2418–2442.

[Led01] M. Ledoux, *The concentration of measure phenomenon*, Mathematical Surveys and Monographs **89**, AMS (2001).

[LV09] J. Lott, C. Villani, *Ricci curvature for metric-measure spaces via optimal transport*, Ann. of Math. (2) **169** (2009), n° 3, 903–991.

[Oll07] Y. Ollivier, *Ricci curvature of metric spaces*, C. R. Math. Acad. Sci. Paris **345** (2007), n° 11, 643–646.

[Oll09] Y. Ollivier, *Ricci curvature of Markov chains on metric spaces*, J. Funct. Anal. **256** (2009), n° 3, 810–864.

[OV00] F. Otto, C. Villani, *Generalization of an inequality by Talagrand and links with the logarithmic Sobolev inequality*, J. Funct. Anal. **173** (2000), 361–400.

[Pet] A. Petrunin, *Alexandrov meets Lott–Villani–Sturm*, preprint (2010), `arxiv.org/abs/1003.5948`

[RS05] M.-K. von Renesse, K.-T. Sturm, *Transport inequalities, gradient estimates, and Ricci curvature*, Comm. Pure Appl. Math. **68** (2005), 923–940.

[Stu06] K.-T. Sturm, *On the geometry of metric measure spaces*, Acta Math. **196** (2006), n° 1, 65–177.

[Vil03] C. Villani, *Topics in optimal transportation*, Graduate Studies in Mathematics **58**, American Mathematical Society, Providence (2003).

[Vil08] C. Villani, *Optimal transport, old and new*, Grundlehren der mathematischen Wissenschaften **338**, Springer (2008).