Transport-entropy inequalities on locally acting groups of permutations

Paul-Marie Samson

TRANSPORT-ENTROPY INEQUALITIES ON LOCALLY ACTING GROUPS OF PERMUTATIONS

PAUL-MARIE SAMSON

Abstract. Following Talagrand's concentration results for permutations picked uniformly at random from a symmetric group [Tal95], Luczak and McDiarmid have generalized it to more general groups G of permutations which act suitably 'locally'. Here we extend their results by setting transport-entropy inequalities on these permutations groups. Talagrand and Luczak-Mc-Diarmid concentration properties are consequences of these inequalities. The results are also generalised to a larger class of measures including Ewens distributions of arbitrary parameter θ on the symmetric group. By projection, we derive transport-entropy inequalities for the uniform law on the slice of the discrete hypercube and more generally for the multinomial law. These results are new examples, in discrete setting, of weak transport-entropy inequalities introduced in [GRST15], that contribute to a better understanding of the concentration properties of measures on permutations groups. One typical application is deviation bounds for the so-called configuration functions, such as the number of cycles of given lenght in the cycle decomposition of a random permutation.

1. Introduction

Let S_n denote the symmetric group of permutations acting on a set Ω of cardinality n, and μ_o denote the uniform law on S_n, $\mu_o(\sigma) := \frac{1}{n!}, \sigma \in S_n$. A seminal concentration result on S_n obtained by Maurey is the following.

Theorem 1.1. [Mau79] *Let d_H be the Hamming distance on the symmetric group, for all $\sigma, \tau \in S_n$,*

$$d_H(\sigma,\tau) := \sum_{i\in\Omega} \mathbb{1}_{\sigma(i)\neq\tau(i)}.$$

Then for any subset $A \subset S_n$ such that $\mu_o(A) \geq 1/2$, and for all $t \geq 0$, one has

$$\mu_o(A_t) \geq 1 - 2e^{-\frac{t^2}{64n}},$$

where $A_t := \{y \in S_n, d_H(x, A) \leq t\}$.

Milman and Schechtman [MS86] generalized this result to some groups whose distance is invariant by translation. For example, in the above result we may replace

---

Date: June 23, 2017.

1991 *Mathematics Subject Classification.* 60E15, 32F32, 39B62, 26D10.

Key words and phrases. Concentration of measure, Random permutation, Symmetric group, Slices of the discrete cube, Transport inequalities, Ewens distribution, Chinese restaurant process, Deviation's inequalities for configuration functions.

Supported by the grants ANR 2011 BS01 007 01, ANR 10 LABX-58.

(up to constants) the Hamming distance by the *transposition distance* $d_T(\sigma,\tau)$ that corresponds to the minimal number of transpositions $t_1,...,t_k$ such that $\sigma t_1\cdots t_k = \tau$. The distances d_T and d_H are comparable,

$$\frac{1}{2}d_H(\sigma,\tau)\le d_T(\sigma,\tau)\le d_H(\sigma,\tau)-1,\qquad \forall\sigma\neq\tau.$$

(We refer to [BHT06] for comments about these comparison inequalities).

Let us also observe that Theorem 1.1 can be also recover from the transportation cost inequality approach of Theorem 1 of [Mar03].

A few years later, a stronger concentration property in terms of dependence in the parameter n, has been settled by Talagrand using the so-called "convex-hull" method [Tal95] (see also [Led01]). This property implies Maurey's result with a slightly worse constant. Let us recall some notations from [Tal95]. For each $A\subset S_n$ and $\sigma\in S_n$, let $V(\sigma,A)\subset\mathbb{R}^\Omega$ be the set of vectors $z=(z_j)_{j\in\Omega}\in\mathbb{R}^\Omega$ with $z_j := \mathbb{1}_{\sigma(j)\neq y(j)}$ for $y\in A$. Let $\mathrm{conv}(V(\sigma,A))$ denote the convex hull of $V(\sigma,A)$ in $\mathbb{R}^\Omega$,

$$V(\sigma,A):=\left\{x=(x_j)_{j\in\Omega},\exists p\in\mathcal{P}(A),\forall j\in\Omega,x_j=\int\mathbb{1}_{\sigma(j)\neq y(j)}dp(y)\right\},$$

where $\mathcal{P}(A)$ denotes the set of probability measures on A. Talagrand introduced the quantity

$$f(\sigma,A):=\inf\{\|x\|_2^2;x\in\mathrm{conv}(V(\sigma,A))\}.$$

with $\|x\|_2^2:=\sum_{i\in\Omega}x_i^2$, that measures the distance from σ to the subset A.

Theorem 1.2. [Tal95] *For any subset $A\subset S_n$,*

$$\int_{S_n}e^{f(\sigma,A)/16}d\mu_o(\sigma)\le\frac{1}{\mu_o(A)}.$$

Maurey's concentration result easily follows by observing that

$$f(\sigma,A)\ge\frac{1}{n}\left(\inf\left\{\sum_{i\in\Omega}x_i;x\in\mathrm{conv}(V(\sigma,A))\right\}\right)^2=\frac{1}{n}d_H^2(\sigma,A)$$

and applying Tchebychev inequality with usual optimization arguments.

Talagrand's result has been first extended to the uniform probability measure on product of symmetric groups by McDiarmid [McD02], and then further by Luczak and McDiarmid to cover more general permutation groups which act suitably "locally" [LM03].

For any finite subset A, let $\#A$ denote the *cardinality* of A. For any $\sigma\in S_n$, the *support* of σ, denoted by $\mathrm{supp}(\sigma)$, is the set $\{i\in\Omega,\sigma(i)\neq i\}$ and the *degree* of σ, denoted by $\deg(\sigma)$, is the cardinality of $\mathrm{supp}(\sigma)$, $\deg(\sigma):=\#\ \mathrm{supp}(\sigma)$.

By definition, according to [LM03], a group of permutations G is *ℓ-local*, $\ell\in\{2,\dots,n\}$, if for any $\sigma\in G$ and any $i,j\in\Omega$ with $\sigma(i)=j$, there exists $\tau\in G$ such that $\mathrm{supp}(\tau)\subset\mathrm{supp}(\sigma)$, $\deg(\tau)\le\ell$ and $\tau(i)=j$.

The *orbit* of an element $j\in\Omega$, denoted by $\mathrm{orb}(j)$, is the set of elements in Ω connected to j by a permutation of G,

$$\mathrm{orb}(j):=\{\sigma(j),\sigma\in G\}.$$

The set of orbits provides a partition of G.

As explained in [LM03], any 2-local group is a direct product of symmetric groups on its orbits, the alternating group (consisting of even permutations) is 3-local, and any 3-local group is a direct product of symmetric or alternating groups on its orbits.

In the present paper, the concentration result by Luczak-McDiarmid and Talagrand is a consequence of a weak transport-entropy inequality satisfied by the uniform law on G, μ_o. We also prove weaker types of transport entropy inequalities. Moreover we extend the results to a larger class of probability measures on G, denoted by $\mathcal{M}$.

For a better comprehension of the class of measures $\mathcal{M}$, let us first consider the case of the symmetric group S_n on $[n] := \{1, \ldots, n\}$. Let (i, j) denote the transposition in S_n that exchanges the elements i and j in $[n]$. It follows by induction that the map

$$U : \begin{array}{ccc} \{1,2\} \times \{1,2,3\} \times \cdots \times \{1,\ldots,n\} & \to & S_n \\ i_2, i_3, \ldots, i_n & \mapsto & (i_2, 2)(i_3, 3)\cdots(i_n, n), \end{array}$$

is one to one.

The set of measures $\mathcal{M}$ consists of probability measures on S_n which are pushed forward by the map U of product probability measures on $\{1,2\} \times \{1,2,3\} \times \cdots \times \{1, \ldots, n\}$,

$$\mathcal{M} := \left\{ U\#\hat{\nu}, \hat{\nu} = \hat{\nu}_2 \otimes \cdots \otimes \hat{\nu}_n \text{ with } \hat{\nu}_j \in \mathcal{P}([j]),\ \forall j \in \{2, \ldots, n\} \right\}, \tag{1}$$

where by definition $U\#\hat{\nu}(C) = \hat{\nu}(U^{-1}(C))$ for any subset C in S_n.

The uniform measure μ_o on S_n belongs to the set $\mathcal{M}$ since $\mu_o = U\#\hat{\mu}$ with $\hat{\mu} = \hat{\mu}_2 \otimes \cdots \otimes \hat{\mu}_n$, where for each i, $\hat{\mu}_i$ denotes the uniform law on $[i]$.

The Ewens distribution of parameter $\theta > 0$, denoted by μ^θ, is also an example of measure of $\mathcal{M}$. Indeed, it is well known (see [ABT03, Chapter 5], [JKB97]) that $\mu^\theta = U\#\hat{\mu}^\theta$ with $\hat{\mu}^\theta = \hat{\mu}_2^\theta \otimes \cdots \otimes \hat{\mu}_n^\theta$, where for any $j \in \{2, \ldots, n\}$, the measure $\hat{\mu}_j^\theta \in \mathcal{P}([j])$ is given by

$$\hat{\mu}_j^\theta(j) = \frac{\theta}{\theta + j - 1}, \quad ,\hat{\mu}_j^\theta(1) = \cdots = \hat{\mu}_j^\theta(j-1) = \frac{1}{\theta + j - 1}.$$

This definition provides an easy algorithm for simulating a random permutation with law μ^θ. This procedure is known as *a Chinese restaurant process* (see [ABT03, Chapter 2], [Pit06]).

Let us observe that the uniform distribution μ_o corresponds to the Ewens distribution with parameter 1, μ^1.

The Ewens distribution is also given by the following expression (see [ABT03, Chapter 5]),

$$\mu^\theta(\sigma) := \frac{\theta^{|\sigma|}}{\theta^{(n)}}, \quad \sigma \in S_n, \tag{2}$$

where $|\sigma|$ denotes the number of cycles in the cycle decomposition of σ and $\theta^{(n)}$ is the Pochhammer symbol defined by

$$\theta^{(n)} := \frac{\Gamma(\theta+n)}{\Gamma(\theta)}, \qquad \text{with} \quad \Gamma(\theta) := \int_0^{+\infty} s^{\theta-1}e^{-s}ds.$$

Let us now construct the class of measures $\mathcal{M}$ for any group G of permutations. To clarify the notations, the elements of Ω are labelled with integers, $\Omega = [n]$. Let $G_n := G$ and for any $j \in [n-1]$, let G_j denotes the subgroup of G defined by

$$G_j := \{\sigma \in G, \sigma(j+1) = j+1, \ldots, \sigma(n) = n\},$$

We denote by O_j the orbit of j in G_j,

$$O_j := \left\{\sigma(j), \sigma \in G_j\right\}.$$

Let us observe that $\{j\} \subset O_j \subset [j]$.

Definition 1.1. *Let G be a group of permutations. A family $\mathcal{T} = (t_{i_j,j})$ of permutations of G, indexed by $j \in \{2,\ldots,n\}$ and $i_j \in O_j$, is called "ℓ-local base of G" if for every $j \in \{2,\ldots,n\}$, $t_{j,j} := id$, for every $i_j \neq j$, $t_{i_j,j} \in G_j$ and*

$$t_{i_j,j}(i_j) = j, \quad \text{and} \quad deg(t_{i_j j}) \leq \ell.$$

Lemma 1.1. *Let $\mathcal{T} = (t_{i_j,j})$ be a ℓ-local base of a group of permutations G. Then the map*

$$\begin{array}{rccc} & O_2 \times O_3 \times \cdots \times O_n & \to & G \\ U_{\mathcal{T}} : & i_2, i_3, \ldots, i_n & \mapsto & t_{i_2,2}t_{i_3,3}\cdots t_{i_n,n}, \end{array} \tag{3}$$

is one to one.

Lemma 1.2. *Any ℓ-local group of permutations admits a "ℓ-local base".*

For completeness, a proof of these two lemmas is given in the Appendix.

As a consequence of these lemmas, if G is a ℓ-local group, then there exists a ℓ-local base $\mathcal{T}$, such that the uniform probability measure μ_o satisfies $\mu_o = U_{\mathcal{T}}\#\hat{\mu}$, with $\hat{\mu} = \hat{\mu}_2 \otimes \cdots \otimes \hat{\mu}_n$, where for each j, $\hat{\mu}_j$ is the uniform law on O_j.

As for the symmetric group, given a ℓ-local base $\mathcal{T}$ of a group G, the class of measures $\mathcal{M} = \mathcal{M}_{\mathcal{T}}$ on G is made up of all probability measures on G which are pushed forward of product probability measures on $O_2 \times O_3 \times \cdots \times O_n$ by the map $U_{\mathcal{T}}$ defined by (37),

$$\mathcal{M}_{\mathcal{T}} := \left\{U_{\mathcal{T}}\#\hat{\nu}, \hat{\nu} = \hat{\nu}_2 \otimes \cdots \otimes \hat{\nu}_n \text{ with } \hat{\nu}_j \in \mathcal{P}(O_j),\ \forall j \in \{2,\ldots,n\}\right\}. \tag{4}$$

As explained above, if G is a ℓ-local group, the class $\mathcal{M}_{\mathcal{T}}$ contains the uniform law μ_o on G for a well choosen ℓ-local base $\mathcal{T}$.

In this paper, the concentration results are derived from weak transport-entropy inequalities, involving the relative entropy $H(\nu|\mu)$ between two probability measures μ, ν on G given by

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

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

The terminology "weak transport-entropy" introduced in [GRST15], encompass many kinds of transport-entropy inequalities from the well-known Talagrand's transport inequality satisfied by the standard Gaussian measure on $\mathbb{R}^n$ [Tal96], to the usual Csizár-Kullback-Pinsker inequality [Pin64, Csi67, Kul67] that holds for any (reference) probability measure μ on a Polish metric space $\mathcal{X}$, namely

$$\|\mu - \nu\|_{TV}^2 \leq 2\,H(\nu|\mu), \qquad \forall \nu \in \mathcal{P}(\mathcal{X}). \tag{5}$$

where $\|\mu - \nu\|_{TV}$ denotes the total variation distance between μ and ν,

$$\|\mu - \nu\|_{TV} := 2\sup_A |\mu(A) - \nu(A)|.$$

Above, the supremum runs over all measurable subset A of $\mathcal{X}$. We refer to the survey [Sam16, Sam17] for other examples of weak transport-entropy inequalities and their connections with the concentration of measure principle.

The next theorem is one of the main result of this paper. It presents new weak transport inequalities for the uniform measure on G or any measure in the class $\mathcal{M}_{\mathcal{T}}$, that recover the concentration results of Theorems 1.1 and 1.2.

We also denote by d_H the Hamming distance on G: for any $\sigma, \tau \in G$,

$$d_H(\sigma, \tau) := \deg(\sigma\tau^{-1}) = \sum_{i=1}^{n} \mathbb{1}_{\sigma(i)\neq\tau(i)},$$

and the distance $d_T(\sigma, \tau)$ is defined as the minimal number of elements of G, $t_1, ..., t_k$, with degree less than ℓ, such that $\sigma t_1 \cdots t_k = \tau$.

For any measures $\nu_1, \nu_2 \in \mathcal{P}(G)$, the set $\Pi(\nu_1, \nu_2)$ denotes the set of all probability measures on $G \times G$ with first marginal ν_1 and second marginal ν_2. The Wasserstein distance between ν_1 and ν_2, according to the distance $d = d_H$ or $d = d_T$, is given by

$$W_1(\nu_1, \nu_2) := \inf_{\pi\in\Pi(\nu_1,\nu_2)} \iint d(\sigma, \tau)\, d\pi(\sigma, \tau).$$

We also consider two other optimal weak transport costs, $\widetilde{T}_2(\nu_2|\nu_1)$ and $\widehat{T}_2(\nu_2|\nu_1)$ defined by

$$\widetilde{T}_2(\nu_2|\nu_1) := \inf_{\pi\in\Pi(\nu_1,\nu_2)} \int \left(\int d(\sigma, \tau)\, dp_\sigma(\tau) \right)^2 d\nu_1(\sigma), \tag{6}$$

and

$$\widehat{T}_2(\nu_2|\nu_1) := \inf_{\pi\in\Pi(\nu_1,\nu_2)} \int \sum_{i=1}^{n} \left(\int \mathbb{1}_{\sigma(i)\neq\tau(i)}\, dp_\sigma(\tau) \right)^2 d\nu_1(\sigma),$$

where p_σ represents any probability measure such that $\pi(\sigma, \tau) = \nu_1(\sigma)p_\sigma(\tau)$ for all $\sigma, \tau \in G$. By Jensen's inequality, these weak transport costs are comparable, namely

$$W_1^2(\nu_1, \nu_2) \leq \widetilde{T}_2(\nu_2|\nu_1) \leq n\widehat{T}_2(\nu_2|\nu_1),$$

where the last inequality only holds for $d = d_H$.

By definition a subgroup G of S_n is *normal* if for any $t \in S_n$,$t^{-1}Gt = G$.

In the next theorem the constant K_n is the cardinality of the set $\{ j \in \{2, \ldots, n\}, O_j \neq \{j\}\}$. It follows that $0 \leq K_n \leq (n-1)$ and $K_n = 0$ if and only if $G = \{id\}$.

Theorem 1.3. *Let G be a group of permutations with ℓ-local base $\mathcal{T}$. Let $\mu \in \mathcal{P}(G)$ be a measure of the set $\mathcal{M}_{\mathcal{T}}$ defined by* (4).

(a) For all probability measures ν_1 and ν_2 on G, one has

$$\frac{2}{c(\ell)^2} W_1^2(\nu_1, \nu_2) \leq K_n \left(\sqrt{H(\nu_1|\mu)} + \sqrt{H(\nu_2|\mu)} \right)^2, \tag{7}$$

and

$$\frac{1}{2c(\ell)^2} \widetilde{T}_2(\nu_2|\nu_1) \leq K_n \left(\sqrt{H(\nu_1|\mu)} + \sqrt{H(\nu_2|\mu)} \right)^2, \tag{8}$$

where

$$c(\ell) := \begin{cases} \min(2\ell - 1, n) & \text{if } d = d_H, \\ 2 & \text{if } d = d_T. \end{cases}$$

When $\mu = \mu_o$ is the uniform law of a ℓ-local group G, inequalities (7) *and* (8) *hold with*

$$c(\ell) := \begin{cases} \ell & \text{if } d = d_H, \\ 1 & \text{if } d = d_T. \end{cases}$$

(b)
- *Assume that $\mu = \mu_o$ is the uniform law of a ℓ-local group G. Then, for all probability measures ν_1 and ν_2 on G,*

$$\frac{1}{2c(\ell)^2} \widehat{T}_2(\nu_2|\nu_1) \leq \left(\sqrt{H(\nu_1|\mu)} + \sqrt{H(\nu_2|\mu)} \right)^2, \tag{9}$$

with $c(\ell)^2 = 2(\ell - 1)^2 + 2$.

- *Assume that G is a normal subgroup of S_n, and that μ satisfies for all $\sigma \in G, t \in S_n$*

$$\mu(\sigma) = \mu(\sigma^{-1}) \quad \text{and} \quad \mu(\sigma) = \mu(t^{-1}\sigma t). \tag{10}$$

Then, the inequality (9) *holds with $c(\ell)^2 = 8(\ell - 1)^2 + 2$.*

The proofs of these results, given in the next section, are inspired by Talagrand seminal work on S_n [Tal95], and Luczak-McDiarmid extension to ℓ-local groups [LM03].

Comments :

- If $G = S_n$ and the class of measure $\mathcal{M}$ is given by (1), the Ewens distribution μ^θ introduced before, is an interesting example of measure in $\mathcal{M}$, satisfying condition (10). This simply follows from its expression given by (2), since for any $\sigma, t \in S_n$, $|\sigma^{-1}| = |\sigma|$ and $|t^{-1}\sigma t| = |\sigma|$.

  An open question is to generalize the above transport-entropy inequalities to the *generalized Ewens distribution* (see the definition in [MNZ12, HNNZ13]). This measure no longer belongs to the class of measure $\mathcal{M}$. In other words, no Chinese restaurant process are known for simulating the generalized Ewens distribution.

- From the triangular inequality satisfied by the Wasserstein distance W_1, the transport-entropy inequality (7) is clearly equivalent to the following transport-entropy inequality, for all probability measure ν on G,

$$\frac{2}{c(\ell)^2} W_1^2(\nu,\mu) \leq K_n\, H(\nu|\mu).$$

Here is a popular dual formulation of this transport-entropy inequality: for all 1-Lipschitz functions $\varphi : G \to \mathbb{R}$ (with respect to the distance d),

$$\int e^{\varphi} d\mu \leq e^{\int \varphi\, d\mu + K_n c(\ell)^2 t^2/8}, \qquad \forall t \geq 0. \tag{11}$$

For the uniform measure on S_n, $K_n = n-1$ and this property is widely commented in [BHT06]; it is also a consequence of Hoeffding inequalities for bounded martingales (see page 18 of [Hoe63]). The concentration result derived from item (a) are of the same nature as the one obtained by the "bounded differences approach" in [Mau79, McD89, McD02, LM03, BDR15].

- Similarly, by Proposition 4.5 and Theorem 2.7 of [GRST15] and using the identity

$$\left(\sqrt{u} + \sqrt{v}\right)^2 = \inf_{\alpha\in(0,1)} \left\{\frac{u}{\alpha} + \frac{v}{1-\alpha}\right\},$$

we may easily show that the weak transport-entropy inequality (8) is equivalent to the following dual property: for any real function φ on G and for any $0 < \alpha < 1$,

$$\left(\int e^{\alpha \widetilde{Q}_{K_n}\varphi} d\mu\right)^{1/\alpha} \left(\int e^{-(1-\alpha)\varphi} d\mu\right)^{1/(1-\alpha)} \leq 1, \tag{12}$$

where the infimum-convolution operator $\widetilde{Q}_t\varphi$, $t \geq 0$, is defined by

$$\widetilde{Q}_t\varphi(\sigma) := \inf_{p\in\mathcal{P}(G)} \left\{\int \varphi\, dp + \frac{1}{2c^2(\ell)t}\left(\int d(\sigma,y)\, dp(y)\right)^2\right\}, \quad \sigma \in G.$$

Moreover, let us observe that following our proof of (12) in the next section, for each $\alpha \in (0,1)$ the inequality (12) can be improved by replacing the square cost function by the convex cost $c_\alpha(u) \geq u^2/2, u \geq 0$ given in Lemma 2.2. More precisely, (12) holds replacing $\widetilde{Q}_{K_n}\varphi$ by $\widetilde{Q}^{\alpha}_{K_n}\varphi$ defined by

$$\widetilde{Q}^{\alpha}_t\varphi(\sigma) := \inf_{p\in\mathcal{P}(S_n)} \left\{\int \varphi\, dp + t c_\alpha\left(\frac{1}{c(\ell)t}\int d(\sigma,y)\, dp(y)\right)^2\right\},$$

for any $\sigma \in G, t > 0$.

- Proposition 4.5 and Theorem 9.5 of [GRST15] also provide a dual formulation of the weak transport-entropy inequality (9): for any real function φ on G and for any $0 < \alpha < 1$,

$$\left(\int e^{\alpha \widehat{Q}\varphi} d\mu\right)^{1/\alpha} \left(\int e^{-(1-\alpha)\varphi} d\mu\right)^{1/(1-\alpha)} \leq 1, \tag{13}$$

where the infimum convolution operator $\widehat{Q}\varphi$ is defined by

$$\widehat{Q}\varphi(\sigma) = \inf_{p\in\mathcal{P}(G)} \left\{ \int \varphi\, dp + \frac{1}{2c(\ell)^2} \sum_{k=1}^{n} \left(\int \mathbb{1}_{\sigma(k)\neq y(k)}\, dp(y) \right)^2 \right\}, \quad \sigma \in G.$$

As explained at the end of this section, the property (13) directly provides the following version of the Talagrand's concentration result for any measure on G of the set $\mathcal{M}_{\mathcal{T}}$.

Corollary 1.1. *Let G be a group of permutations with ℓ-local base $\mathcal{T}$. Let $\mu \in \mathcal{P}(G)$ be a measure of the set $\mathcal{M}_{\mathcal{T}}$ defined by* (4)*. Assume that μ and G satisfy the conditions of* (*b*) *in Theorem 1.3. Then, for all $A \subset G$ and all $\alpha \in (0,1)$, one has*

$$\int e^{\frac{\alpha}{2c(\ell)^2} f(\sigma,A)}\, d\mu(\sigma) \leq \frac{1}{\mu(A)^{\alpha/(1-\alpha)}},$$

with the same definition for $c(\ell)^2$ as in part (*b*) *of Theorem 1.3. As a consequence, by Tchebychev inequality, for any $\alpha \in (0,1)$ and all $t \geq 0$,*

$$\mu\big(\{\sigma \in G, f(\sigma, A) \geq t\}\big) \leq \frac{e^{-\frac{\alpha t}{2c(\ell)^2}}}{\mu(A)^{\alpha/(1-\alpha)}}.$$

For $\alpha = 1/2$ and $\mu = \mu_o$ the uniform law on a ℓ-local group of G, this result is exactly Theorem 2.1 by Luczak-McDiarmid [LM03], that generalizes Theorem 1.2 on S_n (since S_n is a 2-local group).

By projection arguments, Theorem 1.3 applied with the uniform law μ_o on the symmetric group S_n, also provides transport-entropy inequalities for the uniform law on the slices of the discrete cube $\{0,1\}^n$. Namely, for $n \geq 1$, let us denote by $\mathcal{X}_{k,n-k}$, $k \in \{0, \ldots, n\}$, *the slices of discrete cube* defined by

$$\mathcal{X}_{k,n-k} := \left\{ x = (x_1, \ldots, x_n) \in \{0,1\}^n, \sum_{i=1}^{n} x_i = k \right\}.$$

The uniform law on $\mathcal{X}_{k,n-k}$, denoted by $\mu_{k,n-k}$, is the pushed forward of μ_o by the projection map

$$\begin{array}{rccc} & S_n & \to & \mathcal{X}_{k,n-k} \\ P: & \sigma & \mapsto & \mathbb{1}_{\sigma([k])}, \end{array}$$

where $\sigma([k]) := \{\sigma(1), \ldots, \sigma(k)\}$ and for any subset A of $[n]$, $\mathbb{1}_A$ is the vector with coordinates $\mathbb{1}_A(i), i \in [n]$. In other terms, $\mu_{k,n-k} = P\#\mu_o$ and $\mu_{k,n-k}(x) = \binom{n}{k}^{-1}$ for all $x \in \mathcal{X}_{k,n-k}$. Let d_h denotes the *Hamming distance on $\mathcal{X}_{k,n-k}$* defined by

$$d_h(x,y) := \frac{1}{2} \sum_{i=1}^{n} \mathbb{1}_{x_i \neq y_i}, \qquad x, y \in \mathcal{X}_{k,n-k}.$$

Theorem 1.4. *Let $\mu_{k,n-k}$ be the uniform law on $\mathcal{X}_{k,n-k}$, a slice of the discrete cube.*

(*a*) *For all probability measures ν_1 and ν_2 on $\mathcal{X}_{k,n-k}$,*

$$\frac{2}{C_{k,n-k}} W_1^2(\nu_1, \nu_2) \leq \left(\sqrt{H(\nu_1|\mu_{k,n-k})} + \sqrt{H(\nu_2|\mu_{k,n-k})} \right)^2,$$

and

$$\frac{1}{2C_{k,n-k}}\widetilde{T}_2(\nu_2|\nu_1) \leq \left(\sqrt{H(\nu_1|\mu_{k,n-k})} + \sqrt{H(\nu_2|\mu_{k,n-k})}\right)^2,$$

where W_1 is the Wasserstein distance associated to d_h, $\widetilde{T}_2$ is the weak optimal transport cost defined by (6) *with $d = d_h$, and $C_{k,n-k} = \min(k, n-k)$.*

(b) For all probability measures ν_1 and ν_2 on $\mathcal{X}_{k,n-k}$,

$$\frac{1}{8}\widehat{T}_2(\nu_2|\nu_1) \leq \left(\sqrt{H(\nu_1|\mu_{k,n-k})} + \sqrt{H(\nu_2|\mu_{k,n-k})}\right)^2, \tag{14}$$

where

$$\widehat{T}_2(\nu_2|\nu_1) := \inf_{\pi\in\Pi(\nu_1,\nu_2)} \int \sum_{i=1}^n \left(\int \mathbb{1}_{x_i\neq y_i} dp_x(y)\right)^2 d\nu_1(x),$$

with $\pi(x,y) = \nu_1(x)p_x(y)$ for all $x, y \in \mathcal{X}_{k,n-k}$.

Up to constants, the weak transport inequality (14) is the stronger one since for all $\nu_1, \nu_2 \in \mathcal{P}(\mathcal{X}_{k,n-k})$,

$$W_1^2(\nu_1,\nu_2) \leq \widetilde{T}_2(\nu_2|\nu_1) \leq \frac{n}{4}\widehat{T}_2(\nu_2|\nu_1).$$

The proof of Theorem 1.4 is given in section 3. The transport-entropy inequality (14) is derived by projection from the transport-entropy inequality (9) for the uniform measure μ_o on S_n. The same projection argument could be used to reach the results of (a) from the transport-entropy inequality of (a) in Theorem 1.3, but it provides worse constants. The constant $C_{k,n-k}$ is obtained by working directly on $\mathcal{X}_{k,n-k}$ and following similar arguments as in the proof of Theorem 1.3.

Remark : The results of Theorem 1.4 also extend to the multinomial law. Let $E = \{e_1,\ldots,e_m\}$ be a set of cardinality m and let $k_1,\ldots,k_m$ be a collection of non-zero integers satisfying $k_1 + \cdots + k_m = n$. *The multinomial law* $\mu_{k_1,\ldots,k_m}$ is by definition the uniform law on the set

$$\mathcal{X}_{k_1,\ldots,k_m} := \left\{x \in E^n, \text{ such that for all } l \in [m], \#\{i \in [n], x_i = e_l\} = k_l\right\}.$$

For any $x \in \mathcal{X}_{k_1,\ldots,k_m}$, one has $\mu_{k_1,\ldots,k_m}(x) = \frac{k_1!\cdots k_m!}{n!}$. As a result, the weak transport-entropy inequality (14) holds on $\mathcal{X}_{k_1,\ldots,k_m}$ replacing the measure $\mu_{k,n-k}$ by the measure $\mu_{k_1,\ldots,k_m}$. The proof of this result is a simple generalization of the one on $\mathcal{X}_{k,n-k}$, by using the projection map $P : S_n \to \mathcal{X}_{k_1,\ldots,k_m}$ defined by: $P(\sigma) = x$ if and only if

$$x_i = e_l, \quad \forall l \in [m],\ \forall i \in J_l,$$

where $J_l := \{i \in [n], k_0 + \cdots + k_{l-1} < i \leq k_0 + \cdots + k_l\}$, with $k_0 = 0$. The details of this proof are left to the reader.

A straightforward application of transport-entropy inequalities is deviation's bounds for different classes of functions. For more comprehension, we present below deviations bounds that can be reached from Theorem 1.3 for any measure in $\mathcal{M}_{\mathcal{T}}$. A similar corollary can be derived from Theorem 1.4 on the slices of the discrete cube.

For any $h : G \to \mathbb{R}$, the mean of h is denoted by $\mu(h) := \int h\, d\mu$.

Corollary 1.2. *Let G be a group of permutations with ℓ-local base $\mathcal{T}$, $G \neq \{id\}$. Let $\mu \in \mathcal{P}(G)$ be a measure of the set $\mathcal{M}_{\mathcal{T}}$ defined by* (4)*. Let g be a real function on G.*

(a) *Assume that there exists a function $\beta : G \to \mathbb{R}^+$ such that for all $\tau, \sigma \in G$,*

$$g(\tau) - g(\sigma) \leq \beta(\tau) d(\tau, \sigma),$$

where $d = d_T$ or $d = d_H$. Then for all $u \geq 0$, one has

$$\mu\left(g \geq \mu(g) + u\right) \leq \exp\left(-\frac{2u^2}{K_n c(\ell)^2 \, \sup_{\sigma \in G} \beta(\sigma)^2}\right).$$

and

$$\mu\left(g \leq \mu(g) - u\right) \leq \exp\left(-\frac{2u^2}{K_n c(\ell)^2 \, \min(\sup_{\sigma \in G} \beta(\sigma)^2, 4\mu(\beta^2))}\right),$$

where the constants $c(\ell)$ and K_n are defined as in part (a) *of Theorem 1.3.*

(b) *Assume that μ and G satisfy the conditions of* (b) *in Theorem 1.3. Let g be a so-called configuration function. This means that there exist functions $\alpha_k : G \to \mathbb{R}^+$, $k \in \{1, \ldots, n\}$ such that for all $\tau, \sigma \in G$,*

$$g(\tau) - g(\sigma) \leq \sum_{k=1}^{n} \alpha_k(\tau) \mathbb{1}_{\tau(k) \neq \sigma(k)}.$$

Then, for all $v \geq 0, \lambda \geq 0$, one has

$$\mu\left(g \geq \mu(g) + v + \frac{\lambda c(\ell)^2 |\alpha|_2}{2}\right) \leq e^{-\lambda v},$$

and for all $u \geq 0$,

$$\mu\left(g \leq \mu(g) - u\right) \leq \exp\left(-\frac{u^2}{2c(\ell)^2 \mu\left(|\alpha|_2^2\right)}\right),$$

where $|\alpha(\sigma)|_2^2 := \sum_{k=1}^{n} \alpha_k^2(\sigma)$ and $c(\ell)$ is defined as in part (b) *of Theorem 1.3. We also have, for all $u \geq 0$*

$$\mu\left(g \geq \mu(g) + u\right) \leq \exp\left(-\frac{u^2}{2c(\ell)^2 \, \sup_{\sigma \in G} |\alpha(\sigma)|_2^2}\right),$$

and if there exists $M \geq 0$ such that $|\alpha|_2^2 \leq Mg$, then for all $u \geq 0$

$$\mu\left(g \geq \mu(g) + u\right) \leq \exp\left(-\frac{u^2}{2c(\ell)^2 M(\mu(g) + u)}\right),$$

Comments and examples:

- The above deviation's bounds of g around its mean $\mu(g)$ are directly derived from the dual representations (11),(12),(13) of the transport-entropy inequalities of Theorem 1.3, when α goes to 0 or α goes to 1. By classical arguments (see [Led01]), Corollary 1.2 also implies deviation's bounds around a median $M(g)$ of g, but we loose in the constants with this procedure. However, starting directly from Corollary 1.1, we get the following bound under the assumption of (b): for all $u \geq 0$,

$$\mu(g \geq M(g) + u) \leq \frac{1}{2} \exp\left(-w\left(\frac{u}{\sqrt{2}c(\ell)\sup_{\sigma\in G}|\alpha(\sigma)|_2}\right)\right), \tag{15}$$

where $w(u) = u(u - 2\sqrt{\log 2})$, $u \geq 0$.

The idea of the proof is to choose the set $A = \{\sigma \in G, g(\sigma) \leq M(g)\}$ of measure $\mu(A) \geq 1/2$ and to show that the asumption of (b) implies

$$\{\sigma \in G, f(\sigma, A) < t\} \subset \left\{\sigma \in G, g(\sigma) < M(g) + t \sup_{\sigma\in G}|\alpha(\sigma)|_2\right\}, \quad t \geq 0.$$

Then, the deviation bound above the median directly follows from Corollary 1.1 by optimizing over all $\alpha \in (0, 1)$. With identical arguments, the same bound can be reached for $\mu(g \leq M(g) - u)$.

- In (a), the bound above the mean is a simple consequence of (11). As settled in (a), this bound also holds for the deviations under the mean, and it can be slightly improved by replacing $\sup_{\sigma\in G}\beta(\sigma)^2$ by $4\mu(\beta^2)$. This small improvement is a consequence of the weak transport inequality with stronger cost $\widetilde{T}_2$. The same kind of improvement could be reached for the deviations above the mean under additional Lipschitz regularity conditions on the function β.
- Let $\varphi : [0, 1]^n \to \mathbb{R}$ be a 1-Lipschitz convex function and let $x = (x_1, \ldots, x_n)$ be a fixed vector of $[0, 1]^n$. For any $\sigma \in G$, let $x_\sigma := (x_{\sigma(1)}, \ldots, x_{\sigma(n)})$. By applying the results of (b) (or even (15)) to the particular function $g_x(\sigma) = \varphi(x_\sigma)$, $\sigma \in G$, we recover and extend to any group G with ℓ-local base $\mathcal{T}$ and to any measure in $\mathcal{M}_\mathcal{T}$ satisfying (10), the deviation inequality by Adamczak, Chafaï and Wolff [ACW14] (Theorem 3.1) obtained from Theorem 1.2 by Talagrand. Namely, since for any $\sigma, \tau \in G$,

$$\varphi(x_\tau) - \varphi(x_\sigma) \leq \sum_{k=1}^n \partial_k\varphi(x_\tau)(x_{\tau(k)} - x_{\sigma(k)}) \leq \sum_{k=1}^n |\partial_k\varphi(x_\tau)|\mathbb{1}_{\tau(k)\neq\sigma(k)},$$

with $\sum_{k=1}^n |\partial_k\varphi(x_\tau)|^2 = |\nabla\varphi(x_\tau)|^2 \leq 1$, Corollary 1.2 implies, for any choice of vector $x = (x_1, \ldots, x_n) \in [0, 1]^n$,

$$\mu(|g_x - \mu(g_x)| \geq u) \leq 2\exp\left(-\frac{u^2}{2c(\ell)^2}\right), \quad u \geq 0.$$

This concentration property on S_n (with $\ell = 2$) plays a key role in the approach by Adamczak and al. [ACW14], to study the convergence of the empirical spectral measure of random matrices with exchangeable entries, when the size of the matrices is increasing.
- As a second example, for any t in a finite set $\mathcal{F}$, let $(a^t_{i,j})_{1\leq i,j\leq n}$ be a collection of non negative real numbers and consider the function

$$g(\sigma) = \sup_{t\in\mathcal{F}}\left(\sum_{k=1}^n a^t_{k,\sigma(k)}\right), \qquad \sigma \in G.$$

This function satisfies, for any $\sigma, \tau \in G$,

$$g(\tau) - g(\sigma) \le \sum_{k=1}^{n} \left(a_{k,\tau(k)}^{t(\tau)} - a_{k,\sigma(k)}^{t(\tau)} \right) \mathbb{1}_{\tau(k)\neq\sigma(k)} \le \sum_{k=1}^{n} a_{k,\tau(k)}^{t(\tau)} \mathbb{1}_{\tau(k)\neq\sigma(k)},$$

where $t(\tau) \in \mathcal{F}$ is chosen so that

$$g(\tau) = \sum_{k=1}^{n} a_{k,\tau(k)}^{t(\tau)}.$$

Let us consider the function

$$h(\sigma) = \sup_{t\in\mathcal{F}} \left(\sum_{k=1}^{n} (a_{k,\sigma(k)}^{t})^2 \right), \qquad \sigma \in G.$$

The mean of h, $\mu(h)$, can be interpreted as a variance term as regards to g. Observing that g satisfies the condition of (b) with

$$\alpha_k(\tau) := a_{k,\tau(k)}^{t(\tau)},$$

and $|\alpha|_2^2 \le h$, Corollary 1.2 provides the following Bernstein deviation's bounds, for all $u \ge 0$,

$$\mu\left(g \le \mu(g) - u\right) \le \exp\left(- \frac{u^2}{2c(\ell)^2 \mu\left(h\right)} \right),$$

and for all $\lambda, v \ge 0$,

$$\mu\left(g \ge \mu(g) + v + \frac{\lambda c(\ell)^2 h}{2} \right) \le e^{-\lambda v}.$$

If the real numbers $a_{i,j}$ are bounded by M, then $|\alpha|_2^2 \le Mg$ and therefore Corollary 1.2 also provides for all $u \ge 0$,

$$\mu\left(g \ge \mu(g) + u\right) \le \exp\left(- \frac{u^2}{2c(\ell)^2 M(\mu(g) + u)} \right).$$

If we want to bound the deviation above the mean in terms of the variance term $\mu(h)$, it suffises to observe that the last inequality provides deviations bounds for the function h, replacing g by h and M by M^2. Then, as a consequence of all the above deviation's results, it follows that for all $\lambda, v, \gamma \ge 0$,

$$\begin{aligned} &\mu\left(g \ge \mu(g) + v + \frac{\lambda c(\ell)^2(\mu(h) + \gamma)}{2} \right) \\ &\le \mu\left(g \ge \mu(g) + v + \frac{\lambda c(\ell)^2 h}{2} \right) + \mu(h \ge \mu(h) + \gamma) \\ &\le e^{-\lambda v} + \exp\left(- \frac{\gamma^2}{2c(\ell)^2 M^2(\mu(h) + \gamma)} \right). \end{aligned}$$

By choosing $\gamma = Mu$, $\lambda = \frac{u}{c(\ell)^2 M^2(\mu(h)+Mu)}$, and $v = u/2$, we get the following Bernstein deviation inequality for the deviation of g above its mean, for all $u \geq 0$

$$\mu(g \geq \mu(g) + u) \leq 2 \exp\left(-\frac{u^2}{2c(\ell)^2(\mu(h) + Mu)}\right).$$

All the previous deviation's inequalities extend to countable sets $\mathcal{F}$ by monotone convergence.

When $\mathcal{F}$ is reduced to a singleton, these deviation's results simply implies Bernstein deviation's results for $g(\sigma) = \sum_{k=1}^n a_{k,\sigma(k)}$ when $-M \leq a_{i,j} \geq M$ for all $1 \leq i, j \leq n$, by following for example the procedure presented in [BDR15, Section 4.2]. Thus, we extend the deviation's results of [BDR15] to probability measures in $\mathcal{M}_{\mathcal{T}}$.

- As a last example, let $g(\sigma) = |\sigma|_l$ denotes the number of cycles of lenght l in the cycle decomposition of a permutation σ. Let us show that g is a configuration function. Let $C_l(\tau)$ denotes the set of cycles of lenght l in the cycle decomposition of a permutation τ. One has

$$\begin{aligned}|\tau|_l &= \#\{C_l(\tau) \cap C_l(\sigma)\} + \#\{c \in C_l(\tau),\ \text{such that } c \notin C_l(\sigma)\} \\ &\leq |\sigma|_l + \#\{c \in C_l(\tau),\ \text{such that } c \notin C_l(\sigma)\}.\end{aligned}$$

If $c \in C_l(\tau)$ and $c \notin C_l(\sigma)$ then there exists k in the support of c such that $\tau(k) \neq \sigma(k)$. As a consequence, one has

$$\#\{c \in C_l(\tau),\ \text{such that } c \notin C_l(\sigma)\} \leq \sum_{k=1}^n \alpha_k(\tau) \mathbb{1}_{\sigma(k) \neq \tau(k)},$$

where $\alpha_k(\tau) = 1$ if k is in the support of a cycle of lenght l of the cycle decomposition of τ, and $\alpha_l(\tau) = 0$ otherwise. Thus, we get that the function g satisfies the condition of (b), g is a configuration function. Finally, observing that $|\alpha|_2^2 = lg$, Corollary 1.2 provides for any measure $\mu \in \mathcal{M}_{\mathcal{T}}$ satisfying (10), for all $u \geq 0$,

$$\mu\left(g \leq \mu(g) - u\right) \leq \exp\left(-\frac{u^2}{2c(\ell)^2 l \mu\left(h\right)}\right),$$

and

$$\mu\left(g \geq \mu(g) + u\right) \leq \exp\left(-\frac{u^2}{2c(\ell)^2 l(\mu(g) + u)}\right).$$

- The aim of this paper is to clarify the links between Talagrand's type of concentration results on the symmetric group and functional inequalities derived from the transport-entropy inequalities. For brevity's sake, applications of these functional inequalities are not fully developped in the present paper. However, let us briefly mention some other applications using concentration results on the symmetric group: the stochastic travelling salesman problem for sampling without replacement (see Appendix [Pau14]), graph coloring problems (see [McD02]). We also refer to the surveys and books [DP09, MR02] for other numerous examples of application of the concentration of measure principle in randomized algorithms.

Proof of Corollary 1.2. We start with the proof of (b). From the assumption on the function g, we get that for any $p \in \mathcal{P}(G)$

$$\int g\,dp \geq g(\sigma) - \sum_{k=1}^n \left(\alpha_k(\sigma) \int \mathbb{1}_{\sigma(k)\neq\tau(k)}\,dp(\tau)\right)$$
$$\geq g(\sigma) - |\alpha(\sigma)|_2 \left(\sum_{k=1}^n \left(\int \mathbb{1}_{\sigma(k)\neq\tau(k)} dp(\tau)\right)^2\right)^{1/2}.$$

Let $\lambda \geq 0$. Plugging this estimate into the definition of $\widehat{Q}(\lambda g)$, it follows that for any $\sigma \in G$

$$\widehat{Q}(\lambda g)(\sigma) \geq \lambda g(\sigma) - \sup_{u\geq 0}\left\{\lambda|\alpha(\sigma)|_2 u - \frac{u^2}{2c(\ell)^2}\right\} = \lambda g(\sigma) - \frac{\lambda^2|\alpha(\sigma)|_2^2 c(\ell)^2}{2}.$$

As α goes to 1, (13) applied to the function λg yields

$$\int e^{\widehat{Q}(\lambda g)} d\mu \leq e^{\lambda\mu(g)},$$

and therefore

$$\int \exp\left(\lambda g - \frac{\lambda^2 c(\ell)^2 |\alpha|_2^2}{2}\right) d\mu \leq e^{\lambda\mu(g)}, \tag{16}$$

$$\int e^{\lambda g} d\mu \leq \exp\left(\lambda\mu(g) + \frac{\lambda^2 c(\ell)^2 \sup_{\sigma\in G}|\alpha(\sigma)|_2^2}{2}\right), \tag{17}$$

and if $|\alpha|_2^2 \leq Mg$,

$$\int \exp\left(\lambda\left(1 - \frac{\lambda c(\ell)^2 M}{2}\right) g\right) d\mu \leq e^{\lambda\mu(g)}. \tag{18}$$

As α goes to 0, (13) yields

$$\int e^{-\lambda g} d\mu \leq e^{\lambda\mu(\widehat{Q}(\lambda g))},$$

and therefore

$$\int e^{-\lambda g} d\mu \leq \exp\left(-\lambda\mu(g) + \frac{\lambda^2 c(\ell)^2 \mu(|\alpha|_2^2)}{2}\right). \tag{19}$$

The deviation bounds of (b) follows from (16), (19), (17), (18) by Tchebychev inequality, and by optimizing over all $\lambda \geq 0$.

The deviation bounds of (a) are similarly obtained from (12) by Tchebychev inequality. As above, the improvement for the deviation under the mean is a consequence of (12) applied to λg, as α goes to 0, and using the estimate

$$\widetilde{Q}_{K_n}(\lambda g)(\sigma) \geq \lambda g(\sigma) - \frac{\lambda^2 \beta(\sigma)^2 c(\ell)^2 K_n}{2}.$$

□

Proof of Corollary 1.1. Take a subset $A \subset G$ and consider the function φ_λ which takes the values 0 on A and $\lambda > 0$ on $G \setminus A$. It holds

$$\begin{aligned}\widehat{Q}\varphi_\lambda(\sigma) &= \inf_{p\in\mathcal{P}(G)} \left\{ \lambda(1-p(A)) + \frac{1}{2c(\ell)^2}\sum_{j=1}^n \left(\int \mathbb{1}_{\sigma(j)\neq y(j)}\, dp(y)\right)^2 \right\}\\ &= \inf_{\beta\in[0,1]} \{\lambda(1-\beta)+\psi(\beta,\sigma)\},\end{aligned}$$

denoting by

$$\psi(\beta,\sigma) = \inf\left\{ \frac{1}{2c(\ell)^2}\sum_{j=1}^n \left(\int \mathbb{1}_{\sigma(j)\neq y(j)}\, dp(y)\right)^2 ; p(A)=\beta \right\}.$$

So it holds

$$\begin{aligned}\widehat{Q}\varphi_\lambda(\sigma) &= \min\left(\inf_{\beta\in[0,1-\varepsilon]}\{\lambda(1-\beta)+\psi(\beta,\sigma)\}; \inf_{\beta\in[1-\varepsilon,1]}\{\lambda(1-\beta)+\psi(\beta,\sigma)\}, \right)\\ &\geq \min\left(\lambda\varepsilon; \inf_{\beta\geq 1-\varepsilon}\psi(\beta,\sigma)\right) \to \inf_{\beta\geq 1-\varepsilon}\psi(\beta,\sigma),\end{aligned}$$

as $\lambda\to\infty$. It is easy to check that for any fixed σ, the function $\psi(\,\cdot\,,\sigma)$ is continuous on $[0,1]$, so letting ε go to 0, we get $\liminf_{\lambda\to\infty}\widehat{Q}\varphi_\lambda(\sigma) \geq \psi(1,\sigma)$. On the other hand, $\widehat{Q}\varphi_\lambda(\sigma)\leq\psi(1,\sigma)$ for all $\lambda>0$. This proves that $\lim_{\lambda\to\infty}\widehat{Q}\varphi_\lambda(\sigma)=\psi(1,\sigma)$. Applying (13) to φ_λ and letting λ go to infinity yields to

$$\int e^{\alpha\psi(1,\sigma)}\, d\mu\cdot\mu(A)^{\alpha/(1-\alpha)} \leq 1.$$

It remains to observe that $\psi(1,\sigma) = \frac{f(\sigma,A)}{2c(\ell)^2}$. □

2. Proof of Theorem 1.3

Let $\mathcal{T}_n = (t_{i_j,j}, j\in\{2,\dots,n\}, i_j\in O_j)$ be a ℓ-local base of G. Let μ be a probability measure of the set $\mathcal{M}_{\mathcal{T}_n}$ given by (4). Then, there exists a product probability measure $\hat{\nu}=\hat{\nu}_1\otimes\cdots\otimes\hat{\nu}_n$ such that $\mu = U_{\mathcal{T}_n}\#\hat{\nu}$ where the map $U_{\mathcal{T}_n}$ is given by (37).

Each transport-entropy inequality of Theorem 1.3 is obtained by induction over n and using the partition $(H_i)_{i\in\mathrm{orb}(n)}$ of the group G defined by: for any $i\in orb(n)=O_n$,

$$H_i := \{\sigma\in G, \sigma(i)=n\}. \tag{20}$$

According to our notations, $H_n = G_{n-1}$ is a subgroup of G, and we may easily check that $\mathcal{T}_{n-1}$ is a ℓ-local base of this subgroup. We also observe that if G is a normal subgroup of S_n then G_{n-1} is a normal subgroup of S_{n-1}.

Moreover, for any $i\in O_n$, H_i is the coset defined by $H_i = H_n t_{in}$. From the definition of μ, if $\sigma\in H_i$, then there exist $i_2,\dots,i_{n-1}$ such that $\sigma = t_{i_2,2}\cdots t_{i_{n-1},n-1}t_{i,n}$ and therefore

$$\mu(\sigma) = \hat{\nu}_2(i_2)\cdots\hat{\nu}_{n-1}(i_{n-1})\hat{\nu}_n(i).$$

As a consequence, one has $\mu(H_i) = \hat{\nu}_n(i)$. Let μ_i denote the restriction of μ to H_i defined by

$$\mu_i(\sigma) = \frac{\mu(\sigma)}{\mu(H_i)} \mathbb{1}_{\sigma\in H_i}.$$

From the construction of μ, $\mu_n = U_{\mathcal{T}_{n-1}}\#(\hat{\nu}_1 \otimes \cdots \otimes \hat{\nu}_{n-1})$. Moreover, for all $\sigma \in H_n$, one has $\sigma t_{i,n} \in H_i$ and

$$\mu_n(\sigma) = \frac{\mu(\sigma)}{\mu(H_n)} = \frac{\mu(\sigma t_{i,n})}{\mu(H_i)} = \mu_i(\sigma t_{i,n}). \tag{21}$$

Moreover if μ satisfies the condition (10), then $\mu_n \in \mathcal{P}(G_{n-1})$ satisfies the same condition at rank $n-1$: namely, for any $\sigma \in G_{n-1}$, $t \in S_{n-1}$,

$$\mu_n(\sigma) = \mu_n(\sigma^{-1}) \quad \text{ and } \quad \mu_n(\sigma) = \mu_n(t^{-1}\sigma t).$$

These properties are needed in the induction step of the proofs.

When G is a ℓ-local group, let us note that if i and l are elements of $O_n = \mathrm{orb}(n)$, then from the ℓ-local property, there exists $t_{i,l} \in G$ such that $t_{i,l}(i) = l$ and $\deg(t_{i,l}) \leq \ell$. We also have $H_l = H_i t_{i,l}$. If moreover $\mu = \mu_o$ is the uniform law on G, then for any $i, l \in O_n$, $\mu_i(H_i) = \mu_l(H_l) = \frac{1}{\#O_n}$. In that case we will use in the proofs the following property: for any $\sigma \in H_n$, one has $\sigma t_{i,n} \in H_i$, $\sigma t_{i,n} t_{i,l}^{-1} \in H_l$, and

$$\mu_n(\sigma) = \frac{\#O_n}{\#G} = \mu_i(\sigma t_{i,n}) = \mu_l(\sigma t_{i,n} t_{i,l}^{-1}). \tag{22}$$

The measure μ_n is the uniform measure on the ℓ-local subgroup $H_n = G_{n-1}$.

Proof of (a) in Theorem 1.3. As already mentioned, since W_1 satisfies a triangular inequality, the transport-entropy inequality (7) is equivalent to the following one: for all $\nu \in \mathcal{P}(G)$,

$$\frac{2}{c(\ell)^2} W_1^2(\nu, \mu) \leq K_n\, H(\nu|\mu).$$

A dual formulation of this property given by Theorem 2.7 in [GRST15] and Proposition 3.1 in [Sam17] is the following: for all functions φ on G and all $\lambda \geq 0$,

$$\int e^{\lambda Q\varphi} d\mu \leq e^{\int \lambda\varphi\, d\mu + K_n c(\ell)^2 \lambda^2/8}, \tag{23}$$

with

$$Q\varphi(\sigma) = \inf_{p\in\mathcal{P}(S_n)} \left\{ \int \varphi dp + \int d(\sigma,\tau)\, dp(\tau) \right\}$$

We will prove the inequality (23) by induction on n.

Assume that $n = 2$. If $G = \{id\}$ then $K_n = 0$ and the inequality (23) is obvious. If $G \neq \{id\}$, then G is the two points space, $G = S_2$, $\ell = 2$ and one has

$$Q\varphi(\sigma) = \inf_{p\in\mathcal{P}(S_2)} \left\{ \int \varphi dp + c(2) \int \mathbb{1}_{\sigma\neq\tau}\, dp(\tau) \right\}.$$

In that case, (23) exactly corresponds to the following dual form of the Csiszar-Kullback-Pinsker inequality (5) (see Proposition 3.1 in [Sam17]): for any probability measure ν on a Polish space $\mathcal{X}$, for any measurable function $f : \mathcal{X} \to \mathbb{R}$,

$$\int e^{\lambda R^c f} d\nu \leq e^{\lambda \int f\, d\nu + \lambda^2 c^2/8}, \qquad \forall \lambda, c \geq 0, \tag{24}$$

with $R^c f(x) = \inf_{p\in\mathcal{P}(\mathcal{X})} \left\{ \int f dp + c \int \mathbb{1}_{x\neq y} dp(y) \right\}$, $x \in \mathcal{X}$.

The induction step will be also a consequence of (24). Let $(H_i)_{i\in O_n}$ be the partition of G defined by (20). Any $p \in \mathcal{P}(G)$ admits a unique decomposition defined by

$$p = \sum_{i\in O_n} \hat{p}(i) p_i, \qquad \text{with} \quad p_i \in \mathcal{P}(H_i) \qquad \text{and} \quad \hat{p}(i) = p(H_i). \tag{25}$$

This decomposition defines a probability measure $\hat{p}$ on O_n. In particular, according to the definition of the measure $\mu \in \mathcal{M}_{\mathcal{T}_n}$ and since $\hat{\nu}_n(i) = \mu(H_i)$, one has

$$\mu = \sum_{i\in O_n} \hat{\nu}_n(i)\, \mu_i.$$

It follows that

$$\int e^{\lambda Q\varphi} d\mu = \sum_{i\in O_n} \hat{\nu}_n(i) \int e^{\lambda Q\varphi(\sigma)} d\mu_i(\sigma) = \sum_{i\in O_n} \hat{\nu}_n(i) \int e^{\lambda Q\varphi(\sigma t_{i,n})} d\mu_n(\sigma),$$

where the last equality is a consequence of property (21). Now, we will bound the right-hand side of this equality by using the induction hypotheses.

For any function $g : G \to \mathbb{R}$ and any $t \in G$, let $g^t : G \to \mathbb{R}$ denote the function defined by $g^t(\sigma) := g(\sigma t)$.

For any function $f : H_n \to \mathbb{R}$ and any $\sigma \in H_n$, let us note

$$Q^{H_n} f(\sigma) := \inf_{p\in\mathcal{P}(H_n)} \left\{ \int f\, dp + \int d(\sigma,\tau)\, dp(\tau) \right\}.$$

The next step of the proof relies on the following Lemma.

Lemma 2.1. *Let $i \in O_n$, for any function $\varphi : H_i \to \mathbb{R}$ and any $\sigma \in H_n$, one has*

(1) $Q\varphi(\sigma t_{i,n}) \leq \inf_{\hat{p}\in\mathcal{P}(O_n)} \left\{ \sum_{l\in O_n} Q^{H_n} \varphi^{t_{n,l}}(\sigma)\hat{p}(l) + c(\ell) \sum_{l\in O_n} \mathbb{1}_{l\neq i} \hat{p}(l) \right\},$
where $c(\ell) = \min(2\ell - 1, n)$ if $d = d_H$ and $c(\ell) = 2$ if $d = d_T$.

(2) $Q\varphi(\sigma t_{i,n}) \leq \inf_{\hat{p}\in\mathcal{P}(O_n)} \left\{ \sum_{l\in O_n} Q^{H_n} \varphi^{t_{i,n}t_{i,l}^{-1}}(\sigma)\hat{p}(l) + c(\ell) \sum_{l\in O_n} \mathbb{1}_{l\neq i} \hat{p}(l) \right\},$
where $c(\ell) = \ell$ if $d = d_H$ and $c(\ell) = 1$ if $d = d_T$, and $t_{i,l}$ denotes an element of G with $\deg(t_{i,l}) \leq \ell$ and such that $t_{i,l}(i) = l$.

This lemma is obtained using the decomposition (25) of the measures $p \in \mathcal{P}(G)$ on the H_j's. Let $\sigma \in H_n$. By the triangular inequality and using the invariance by translation of the distance d, one has

$$\begin{aligned}
\int d(\sigma t_{i,n}, \tau)\, dp(\tau) &= \sum_{l\in O_n} \int_{H_l} d(\sigma t_{i,n}, \tau) dp_l(\tau)\hat{p}(l) \\
&\leq \sum_{l\in O_n} d(\sigma t_{i,n}, \sigma t_{l,n})\hat{p}(l) + \sum_{l\in O_n} \int_{H_l} d(\sigma t_{l,n}, \tau) dp_l(\tau)\hat{p}(l) \\
&= \sum_{l\in O_n} d(t_{i,n}, t_{l,n})\hat{p}(l) + \sum_{l\in O_n} \int_{H_l} d(\sigma, \tau t_{l,n}^{-1}) dp_l(\tau)\hat{p}(l)
\end{aligned}$$

and therefore, since $d(t_{i,n}, t_{l,n}) \leq c(\ell)$ with $c(\ell) = \min(2\ell - 1, n)$ if $d = d_H$ and $c(\ell) = 2$ if $d = d_T$,

$$\int d(\sigma t_{i,n}, \tau)\, dp(\tau) \leq \sum_{l\in O_n} \int_{H_l} d(\sigma, \tau t_{l,n}^{-1}) dp_l(\tau)\hat{p}(l) + c(\ell) \sum_{l\in O_n} \mathbb{1}_{l\neq i}\hat{p}(l). \tag{26}$$

It follows that

$$\begin{aligned}
Q\varphi(\sigma t_{i,n}) \leq & \inf_{\hat{p}\in\mathcal{P}(O_n)} \inf_{p_l\in\mathcal{P}(H_l), l\in O_n} \\
& \left\{ \sum_{l\in O_n} \left[\int \varphi\, dp_l + \int_{H_l} d(\sigma, \tau t_{l,n}^{-1}) dp_l(\tau) \right] \hat{p}(l) + c(\ell) \sum_{l\in O_n} \mathbb{1}_{l\neq i}\hat{p}(l) \right\} \\
= & \inf_{\hat{p}\in\mathcal{P}(O_n)} \inf_{q_l\in\mathcal{P}(H_n), l\in O_n} \\
& \left\{ \sum_{l\in O_n} \left[\int \varphi^{t_{l,n}}\, dq_l + \int_{H_n} d(\sigma, \tau) dq_l(\tau) \right] \hat{p}(l) + c(\ell) \sum_{l\in O_n} \mathbb{1}_{l\neq i}\hat{p}(l) \right\} \\
= & \inf_{\hat{p}\in\mathcal{P}(O_n)} \left\{ \sum_{l\in O_n} Q^{H_n}\varphi^{t_{l,n}}(\sigma)\hat{p}(l) + c(\ell) \sum_{l\in O_n} \mathbb{1}_{l\neq i}\hat{p}(l) \right\}.
\end{aligned}$$

The proof of the second inequality of Lemma 2.1 is similar, starting from the following triangular inequality

$$\begin{aligned}
\int d(\sigma t_{i,n}, \tau)\, dp(\tau) &= \sum_{l\in O_n} \int_{H_l} d(\sigma t_{i,n}, \tau) dp_l(\tau)\hat{p}(l) \\
&\leq \sum_{l\in O_n} \int d(\sigma t_{i,n}, \tau t_{i,l}) dp_l(\tau)\hat{p}(l) + \sum_{l\in O_n} \int_{H_l} d(\tau t_{i,l}, \tau) dp_l(\tau)\hat{p}(l) \\
&= \sum_{l\in O_n} \int d(\sigma, \tau t_{i,l} t_{i,n}^{-1}) dp_l(\tau)\hat{p}(l) + \sum_{l\in O_n} d(t_{i,l}, id)\hat{p}(l) \\
&\leq \sum_{l\in O_n} \int_{H_l} d(\sigma, \tau t_{i,l} t_{i,n}^{-1}) dp_l(\tau)\hat{p}(l) + c(\ell) \sum_{l\in O_n} \mathbb{1}_{l\neq i}\hat{p}(l), \qquad (27)
\end{aligned}$$

with $c(\ell) = \ell$ if $d = d_H$ and $c(\ell) = 1$ if $d = d_T$. The end of the proof of the second inequality of Lemma 2.1 is left to the reader.

The induction step of the proof of (23) continues by applying consecutively Lemma 2.1 (1), the Hölder inequality, and the induction hypotheses to the measure μ_n on the subgroup $H_n = G_{n-1}$ with ℓ-local base $\mathcal{T}_{n-1}$.

If $O_n = \{n\}$ then $K_n = K_{n-1}$ and

$$\int e^{\lambda Q\varphi} d\mu = \int e^{\lambda Q\varphi(\sigma)} d\mu_n(\sigma) \leq e^{\int \lambda\varphi d\mu_n + K_{n-1}c(\ell)^2/8} = e^{\int \lambda\varphi d\mu + K_n c(\ell)^2/8}$$

If $O_n \neq \{n\}$ then $K_n = K_{n-1} + 1$ and for any $i \in O_n$,

$$\int e^{\lambda Q\varphi(\sigma t_{i,n})} d\mu_n(\sigma) \leq \inf_{\hat{p}\in\mathcal{P}(O_n)} \left\{ \prod_{l\in O_n} \left(\int e^{\lambda Q^{H_n}\varphi^{t_{l,n}}} d\mu_n \right)^{\hat{p}(l)} e^{c(\ell)\lambda \sum_{l=1}^{n} \mathbb{1}_{l\neq i}\hat{p}(l)} \right\}$$

$$\leq \exp\left[\inf_{\hat{p}\in\mathcal{P}(O_n)} \left\{ \lambda \sum_{l\in O_n} \left(\int \varphi^{t_{l,n}} d\mu_n \right) \hat{p}(l) + K_{n-1}c(\ell)^2 \frac{\lambda^2}{8} + c(\ell)\lambda \sum_{l\in O_n} \mathbb{1}_{l\neq i}\hat{p}(l) \right\} \right]$$

$$= \exp\left[\lambda \inf_{\hat{p}\in\mathcal{P}(O_n)} \left\{ \sum_{l\in O_n} \hat{\varphi}(l)\hat{p}(l) + c(\ell) \sum_{l\in O_n} \mathbb{1}_{l\neq i}\hat{p}(l) \right\} + K_{n-1}c(\ell)^2 \frac{\lambda^2}{8} \right],$$

where, by using property (21), $\hat{\varphi}(l) := \int \varphi d\mu_l = \int \varphi^{t_{l,n}} d\mu_n$. Let us consider again the above infimum-convolution $R^c\hat{\varphi}$ defined on the space $\mathcal{X} = O_n$, with $c = c(\ell)$, one has

$$R^c\hat{\varphi}(i) = \inf_{\hat{p}\in\mathcal{P}(O_n)} \left\{ \sum_{l\in O_n} \hat{\varphi}(l)\hat{p}(l) + c \sum_{l\in O_n} \mathbb{1}_{l\neq i}\hat{p}(l) \right\}.$$

By applying (24) with the probability measure $\nu = \hat{\nu}_n$ on O_n, the previous inequality gives

$$\int e^{\lambda Q\varphi} d\mu = \sum_{i\in O_n} \hat{\nu}_n(i) \int e^{\lambda Q\varphi(\sigma t_{i,n})} d\mu_n(\sigma) \leq \left(\sum_{i\in O_n} e^{\lambda R^{c(\ell)}\hat{\varphi}(i)} \hat{\nu}_n(i) \right) e^{K_{n-1}\lambda^2/8}$$

$$\leq \exp\left[\sum_{i=1}^{n} \hat{\varphi}(i)\hat{\nu}_n(i) + \frac{\lambda^2 c(\ell)^2}{8} + K_{n-1}c(\ell)^2 \frac{\lambda^2}{8} \right] = \exp\left[\lambda \int \varphi\, d\mu + K_n c(\ell)^2 \frac{\lambda^2}{8} \right].$$

This ends the proof of (23) for any $\mu \in \mathcal{M}_{\mathcal{T}_n}$.

The scheme of the induction proof of (23), with a better constant $c(\ell)$ when $\mu = \mu_o$ is the uniform measure on a ℓ-local group G, is identical, starting from the second result of Lemma 2.1 and using the property (22). This is left to the reader.

We now turn to the induction proof of the dual formulation (12) of the weak transport-entropy inequality (8). The sketch of the proof is identical to the one of (23).

For the initial step $n = 2$, one has $G = S_2$ and $\ell = 2$, and one may easily check that

$$\widetilde{Q}_1\varphi(\sigma) = \inf_{p\in\mathcal{P}(S_2)} \left\{ \int \varphi dp + \frac{1}{2} \left(\int \mathbb{1}_{\sigma\neq\tau}\, dp(\tau) \right)^2 \right\}.$$

In that case, the result follows from the following infimum-convolution property.

Lemma 2.2. *For any probability measure ν on a Polish metric space $\mathcal{X}$, for all $\alpha \in (0,1)$ and all measurable functions $f : \mathcal{X} \to \mathbb{R}$, bounded from below*

$$\left(\int e^{\alpha \widetilde{R}^\alpha f} d\nu \right)^{1/\alpha} \left(\int e^{-(1-\alpha)f} d\nu \right)^{1/(1-\alpha)} \leq 1,$$

where for all $x \in \mathcal{X}$,

$$\widetilde{R}^\alpha f(x) = \inf_{p\in\mathcal{P}(\mathcal{X})} \left\{ \int f(y)dp(y) + c_\alpha \left(\int \mathbb{1}_{x\neq y} dp(y) \right)^2 \right\},$$

and c_α is the convex function defined by

$$c_\alpha(u) = \frac{\alpha(1-u)\log(1-u) - (1-\alpha u)\log(1-\alpha u)}{\alpha(1-\alpha)}, \quad u \in [0,1].$$

Observing that $c_\alpha(u) \geq u^2/2$ for all $u \in [0,1]$, the above inequality also holds replacing $\widetilde{R}^\alpha f$ by

$$\widetilde{R}f(x) = \inf_{p\in\mathcal{P}(\mathcal{X})} \left\{ \int f(y)dp(y) + \frac{1}{2} \left(\int \mathbb{1}_{x\neq y} dp(y) \right)^2 \right\}, \qquad x \in \mathcal{X}. \tag{28}$$

The proof of this Lemma can be found in [Sam07] (inequality (4)). For a sake of completeness, we give in the Appendix a new proof of this result on finite spaces $\mathcal{X}$ by using a localization argument (Lemma 4.1).

Let us now present the key lemma for the induction step of the proof. For any function $f : H_n \to \mathbb{R}$ and any $\sigma \in H_n$, we define

$$\widetilde{Q}_t^{H_n} f(\sigma) := \inf_{p\in\mathcal{P}(H_n)} \left\{ \int f\, dp + \frac{1}{2c(\ell)^2 t} \left(\int d(\sigma,\tau)\, dp(\tau) \right)^2 \right\}.$$

Here, writing $Q_t^{H_n} f$, we omit the dependence in $c(\ell)$ to simplify the notations. The proof relies on the following Lemma.

Lemma 2.3. *Let $i \in O_n$. For any function $\varphi : H_i \to \mathbb{R}$ and any $\sigma \in H_n$, one has*

(1) $\widetilde{Q}_{K_n}\varphi(\sigma t_{i,n}) \leq \inf_{\hat{p}\in\mathcal{P}(O_n)} \left\{ \sum_{l\in O_n} \widetilde{Q}_{K_{n-1}}^{H_n} \varphi^{t_{l,n}}(\sigma)\hat{p}(l) + \frac{1}{2}\Big(\sum_{l\in O_n} \mathbb{1}_{l\neq i}\hat{p}(l) \Big)^2 \right\},$
with $c(\ell) = \min(2\ell - 1, n)$ if $d = d_H$ and $c(\ell) = 2$ if $d = d_T$.

(2) $\widetilde{Q}_{K_n}\varphi(\sigma t_{i,n}) \leq \inf_{\hat{p}\in\mathcal{P}(O_n)} \left\{ \sum_{l\in O_n} \widetilde{Q}_{K_{n-1}}^{H_n} \varphi^{t_{i,n}t_{i,l}^{-1}}(\sigma)\hat{p}(l) + \frac{1}{2}\Big(\sum_{l\in O_n} \mathbb{1}_{l\neq i}\hat{p}(l) \Big)^2 \right\},$
where $c(\ell) = \ell$ if $d = d_H$ and $c(\ell) = 1$ if $d = d_T$, and $t_{i,l}$ denotes an element of G with $deg(t_{i,l}) \leq \ell$ and such that $t_{i,l}(i) = l$.

The proof of this lemma is similar to the one of Lemma 2.1. By (26) and the inequality

$$(u+v)^2 \leq \frac{u^2}{s} + \frac{v^2}{1-s}, \qquad u, v \in \mathbb{R}, \quad s \in (0,1),$$

we get for any $s \in (0,1)$,

$$
\begin{aligned}
\left(\int d(\sigma t_{l,n}, \tau)\, dp(\tau)\right)^2 &\leq \left(\sum_{l\in O_n} \int_{H_l} d(\sigma, \tau t_{l,n}^{-1}) dp_l(\tau)\hat{p}(l) + c(\ell) \sum_{l\in O_n} \mathbb{1}_{l\neq i}\hat{p}(l)\right)^2 \\
&\leq \frac{1}{s}\left(\sum_{l\in O_n} \int_{H_l} d(\sigma, \tau t_{l,n}^{-1}) dp_l(\tau)\hat{p}(l)\right)^2 + \frac{c(\ell)^2}{1-s}\Big(\sum_{l\in O_n} \mathbb{1}_{l\neq i}\hat{p}(l)\Big)^2 \\
&\leq \frac{1}{s}\sum_{l\in O_n}\left(\int_{H_l} d(\sigma, \tau t_{l,n}^{-1}) dp_l(\tau)\right)^2 \hat{p}(l) + \frac{c(\ell)^2}{1-s}\Big(\sum_{l\in O_n} \mathbb{1}_{l\neq i}\hat{p}(l)\Big)^2.
\end{aligned}
$$

It follows that for any $\sigma \in H_n$,

$$
\begin{aligned}
&\widetilde{Q}_{K_n}\varphi(\sigma t_{l,n}) \\
&\leq \inf_{\hat{p}\in\mathcal{P}(O_n)} \inf_{p_l\in\mathcal{P}(H_l), l\in O_n} \left\{\sum_{l\in O_n}\left[\int \varphi\, dp_l + \frac{1}{2c(\ell)^2 s K_n}\left(\int_{H_l} d(\sigma, \tau t_{l,n}^{-1}) dp_l(\tau)\right)^2\right]\hat{p}(l)\right. \\
&\qquad\qquad\left. + \frac{1}{2(1-s)K_n}\Big(\sum_{l\in O_n} \mathbb{1}_{l\neq i}\hat{p}(l)\Big)^2\right\} \\
&= \inf_{\hat{p}\in\mathcal{P}(O_n)} \inf_{q_l\in\mathcal{P}(H_n), l\in O_n} \left\{\sum_{l\in O_n}\left[\int \varphi^{t_{l,n}}\, dq_l + \frac{1}{2c(\ell)^2 s K_n}\left(\int_{H_n} d(\sigma, \tau) dq_l(\tau)\right)^2\right]\hat{p}(l)\right. \\
&\qquad\qquad\left. + \frac{1}{2(1-s)K_n}\Big(\sum_{l\in O_n} \mathbb{1}_{l\neq i}\hat{p}(l)\Big)^2\right\} \\
&= \inf_{\hat{p}\in\mathcal{P}(O_n)} \left\{\sum_{l\in O_n} \widetilde{Q}^{H_n}_{K_{n-1}}\varphi^{t_{l,n}}(\sigma)\hat{p}(l) + \frac{1}{2}\Big(\sum_{l\in O_n} \mathbb{1}_{l\neq i}\hat{p}(l)\Big)^2\right\},
\end{aligned}
$$

where the last equality follows by choosing $s = K_{n-1}/K_n$, which ends the proof of the first inequality of Lemma 2.3. The second inequality of Lemma 2.3 is obtained identically starting from (27).

We now turn to the induction step of the proof. By the decomposition of the measure μ on the H_i's, we want to bound

$$
\int e^{\alpha\widetilde{Q}_{K_n}\varphi} d\mu = \sum_{i\in O_n} \hat{\nu}_n(i) \int e^{\alpha\widetilde{Q}_{K_n}\varphi(\sigma)} d\mu_i(\sigma) = \sum_{i\in O_n} \hat{\nu}_n(i) \int e^{\alpha\widetilde{Q}_{K_n}\varphi(\sigma t_{i,n})} d\mu_n(\sigma),
$$

where the last equality is a consequence of property (21).

If $O_n = \{n\}$, then the result simply follows from the induction hypotheses applied to the measure μ_n.

If $O_n \neq \{n\}$, then applying successively Lemma 2.3 (1), the Hölder inequality, and the induction hypotheses, we get

$$\int e^{\alpha \widetilde{Q}_{K_n}\varphi(\sigma t_{i,n})} d\mu_n(\sigma) \leq \inf_{\hat{p}\in\mathcal{P}(O_n)} \left\{ \prod_{l\in O_n} \left(\int e^{\alpha \widetilde{Q}^{H_n}_{K_{n-1}} \varphi^{t_{l,n}}} d\mu_n \right)^{\hat{p}(l)} \exp\left[\frac{1}{2} \Big(\sum_{l\in O_n} \mathbb{1}_{l\neq i} \hat{p}(l) \Big)^2 \right] \right\}$$

$$\leq \inf_{\hat{p}\in\mathcal{P}(O_n)} \left\{ \prod_{l\in O_n} \left(\int e^{-(1-\alpha)\varphi^{t_{l,n}}} d\mu_n \right)^{-\frac{\hat{p}(l)\alpha}{1-\alpha}} \exp\left[\frac{1}{2} \Big(\sum_{l\in O_n} \mathbb{1}_{l\neq i} \hat{p}(l) \Big)^2 \right] \right\}$$

$$= \exp\left[\alpha \inf_{\hat{p}\in\mathcal{P}(O_n)} \left\{ \sum_{l\in O_n} \hat{\varphi}(l)\hat{p}(l) + \frac{1}{2} \Big(\sum_{l\in O_n} \mathbb{1}_{l\neq i} \hat{p}(l) \Big)^2 \right\} \right],$$

where by property (21), we set

$$\hat{\varphi}(l) := \log \left(\int e^{-(1-\alpha)\varphi} d\mu_l \right)^{-\frac{1}{1-\alpha}} = \log \left(\int e^{-(1-\alpha)\varphi^{t_{l,n}}} d\mu_n \right)^{-\frac{1}{1-\alpha}}.$$

According to the definition of the infimum convolution $\widetilde{R}\hat{\varphi}$ on the space $\mathcal{X} = O_n$ given in Lemma 2.2, the last inequality is

$$\int e^{\alpha \widetilde{Q}_{K_n}\varphi(\sigma t_{i,n})} d\mu_n(\sigma) \leq e^{\alpha \widetilde{R}\hat{\varphi}(i)},$$

and therefore Lemma 2.2, applied with the measure $\nu = \hat{\nu}_n$, provides

$$\int e^{\alpha \widetilde{Q}_{K_n}\varphi} d\mu = \sum_{i\in O_n} e^{\alpha \widetilde{R}\hat{\varphi}(i)} \hat{\nu}_n(i) \leq \left(\sum_{i\in O_n} e^{-(1-\alpha)\hat{\varphi}(i)} \hat{\nu}_n(i) \right)^{-\frac{\alpha}{1-\alpha}}$$

$$= \left(\sum_{i\in O_n} \hat{\nu}_n(i) \int e^{-(1-\alpha)\varphi} d\mu_i \right)^{-\frac{\alpha}{1-\alpha}} = \left(\int e^{-(1-\alpha)\varphi} d\mu \right)^{-\frac{\alpha}{1-\alpha}}.$$

The proof of (12) is completed for any measure $\mu \in \mathcal{M}$. To improve the constant when $\mu = \mu_o$ is the uniform law on a ℓ-local group G, the proof is similar using the second inequality of Lemma 2.3 together with property (22). □

Proof of (b) in Theorem 1.3. We prove the dual equivalent property (13) as a consequence of the stronger following result: for any real function φ on G, for any $j \in \{1, \ldots, n\}$

$$\left(\int e^{\alpha Q^j \varphi} d\mu \right)^{1/\alpha} \left(\int e^{-(1-\alpha)\varphi} d\mu \right)^{1/(1-\alpha)} \leq 1, \tag{29}$$

where the infimum convolution operator $Q^j\varphi$ is defined as follows, for $\sigma \in G$

$$Q^j\varphi(\sigma) = \inf_{p\in\mathcal{P}(G)} \left\{ \int \varphi dp + \frac{1}{c(\ell)^2} \left(\int \mathbb{1}_{\sigma(j)\neq y(j)} dp(y) \right)^2 + \frac{1}{2c(\ell)^2} \sum_{k\in[n]\setminus\{j\}} \left(\int \mathbb{1}_{\sigma(k)\neq y(k)} dp(y) \right)^2 \right\}. \tag{30}$$

The proof of (29) relies on Lemma 2.2 and the following ones. For any $\sigma \in G$, we define

$$Q^{H_n}\varphi(\sigma) = \inf_{p\in\mathcal{P}(H_n)} \left\{ \int \varphi dp + \frac{1}{2c(\ell)^2} \sum_{k=1}^{n-1} \left(\int \mathbb{1}_{\sigma(k)\neq y(k)} dp(y) \right)^2 \right\},$$

and for $j \in [n-1]$,

$$Q^{H_n,j}\varphi(\sigma) = \inf_{p\in\mathcal{P}(H_n)} \left\{ \int \varphi dp + \frac{1}{c(\ell)^2} \left(\int \mathbb{1}_{\sigma(j)\neq y(j)} dp(y) \right)^2 \right.$$
$$\left. + \frac{1}{2c(\ell)^2} \sum_{k\in[n-1]\setminus\{j\}} \left(\int \mathbb{1}_{\sigma(k)\neq y(k)} dp(y) \right)^2 \right\}.$$

Lemma 2.4. *Let $j \in [n]$. For any $\sigma \in G$, one has*

$$Q^j\varphi(\sigma) = Q^{\sigma(j)}\varphi^{\{-1\}}(\sigma^{-1}),$$

where $\varphi^{\{-1\}}(z) = \varphi(z^{-1}), z \in G$.

This result follows from the change of variables $\sigma(k) = l$ in the definition (30) of $Q^j\varphi(\sigma)$, one has

$$Q^j\varphi(\sigma) = \inf_{p\in\mathcal{P}(G)} \left\{ \int \varphi dp + \frac{1}{c(\ell)^2} \left(\int \mathbb{1}_{y^{-1}(\sigma(j))\neq\sigma^{-1}(\sigma(j))} dp(y) \right)^2 \right.$$
$$\left. + \frac{1}{2c(\ell)^2} \sum_{l,l\neq\sigma(j)} \left(\int \mathbb{1}_{l\neq y(\sigma^{-1}(l))} dp(y) \right)^2 \right\}$$
$$= \inf_{q\in\mathcal{P}(G)} \left\{ \int \varphi(z^{-1})\, dq(z) + \frac{1}{c(\ell)^2} \left(\int \mathbb{1}_{z(\sigma(j))\neq\sigma^{-1}(\sigma(j))} dq(z) \right)^2 \right.$$
$$\left. + \frac{1}{2c(\ell)^2} \sum_{l,l\neq\sigma(j)} \left(\int \mathbb{1}_{z(l)\neq\sigma^{-1}(l)} dq(z) \right)^2 \right\},$$

where for the last equality, we use the fact that the map that associates to any measure $p \in \mathcal{P}(G)$ the image measure $q := R\#p$ with $R : \sigma \in G \mapsto \sigma^{-1} \in G$, is one to one from $\mathcal{P}(G)$ to $\mathcal{P}(G)$.

Here is the key lemma for the induction step of the proof of (29).

Lemma 2.5. *(1) Let $j \in O_n$. For any $\sigma \in H_n$, one has*

$$Q^j\varphi(\sigma t_{j,n}) \leq Q^{H_n}\varphi^{t_{j,n}}(\sigma).$$

(2) For any $\ell \geq 2$, let $c^2(\ell) := 8(\ell-1)^2 + 2$. Assume that $O_n \neq \{n\}$ and let $i, j \in O_n$, $i \neq j$. We note $D_i = supp(t_{j,n}^{-1} t_{i,n}) \setminus \{i\}$ and $d = |D_i|$. For any $\sigma \in H_n$, for any $\theta \in [0,1]$ one has

$$Q^i\varphi(\sigma t_{i,n}) \leq \frac{1}{d} \sum_{l\in t_{i,n}(D_i)} \left[\theta Q^{H_n,l}\varphi^{t_{i,n}}(\sigma) + (1-\theta) Q^{H_n}\varphi^{t_{j,n}}(\sigma) \right] + \frac{1}{2}(1-\theta)^2.$$

(3) *For any $\ell \geq 2$, let $c^2(\ell) := 2(\ell-1)^2+2$. Assume that $O_n \neq \{n\}$ and let $i, j \in O_n$, $i \neq j$. Let $t_{i,j} \in G$ such that $t_{i,j}(i) = j$ and $deg(t_{i,j}) \leq \ell$. We note $D_i = supp(t_{i,j}) \setminus \{i\}$ and $d = |D_i|$. For any $\sigma \in H_n$, for any $\theta \in [0,1]$ one has*

$$Q^i\varphi(\sigma t_{i,n}) \leq \frac{1}{d}\sum_{l\in t_{i,n}(D_i)}\left[\theta Q^{H_n,l}\varphi^{t_{i,n}}(\sigma) + (1-\theta)Q^{H_n}\varphi^{t_{i,n}t_{i,j}^{-1}}(\sigma)\right] + \frac{1}{2}(1-\theta)^2.$$

Proof. The first part of this Lemma follows from the fact that $\mathcal{P}(H_j) \subset \mathcal{P}(G)$ and the fact that $\int \mathbb{1}_{\sigma t_{j,n}(j)\neq y(j)}dp(y) = 0$ for $\sigma \in H_n$ and $p \in \mathcal{P}(H_j)$. Therefore, according to the definition of $Q^j\varphi$, one has for $\sigma \in H_j$,

$$Q^j\varphi(\sigma t_{j,n}) \leq \inf_{p\in\mathcal{P}(H_j)}\left\{\int \varphi dp + \frac{1}{2c(\ell)^2}\sum_{k\in[n]\setminus\{j\}}\left(\int \mathbb{1}_{\sigma t_{j,n}(k)\neq y(k)}dp(y)\right)^2\right\}$$

$$= \inf_{q\in\mathcal{P}(H_n)}\left\{\int \varphi^{t_{j,n}}dq + \frac{1}{2c(\ell)^2}\sum_{k\in[n]\setminus\{j\}}\left(\int \mathbb{1}_{\sigma t_{j,n}(k)\neq yt_{j,n}(k)}dq(y)\right)^2\right\} = Q^{H_n}\varphi^{t_{j,n}}(\sigma).$$

For the proof of the second part of Lemma 2.5, we set

$$\tilde{t}_{i,j} := t_{j,n}^{-1}t_{i,n}.$$

Let us consider $p_i^l, l \in D_i$, a collection of measures in $\mathcal{P}(H_i)$, and $p_j \in \mathcal{P}(H_j)$ ($j \neq i$). For $\theta \in [0,1]$,

$$p := \frac{1}{d}\sum_{l\in D_i}[\theta p_i^l + (1-\theta)p_j],$$

is a probability measure on G. Therefore, according to the definition of $Q^i\varphi$, for any $\sigma \in H_n$,

$$Q^i\varphi(\sigma t_{i,n}) \leq \frac{1}{d}\sum_{l\in D_i}\left[\theta\int f dp_i^l + (1-\theta)\int f dp_j\right] + \frac{1}{2c(\ell)^2}(A+B+C),$$

with

$$A = \sum_{k\in[n]\setminus\,\mathrm{supp}(\tilde{t}_{i,j})}\left(\int \mathbb{1}_{\sigma t_{i,n}(k)\neq y(k)}dp(y)\right)^2, \quad B = \sum_{k\in D_i}\left(\int \mathbb{1}_{\sigma t_{i,n}(k)\neq y(k)}dp(y)\right)^2,$$

and $C = 2\left(\int \mathbb{1}_{\sigma t_{i,n}(i)\neq y(i)}dp(y)\right)^2$.

Since $\sigma \in H_n$ and $p_i^l \in \mathcal{P}(H_i)$, one has $\int \mathbb{1}_{\sigma t_{i,n}(i)\neq y(i)}dp_i^l(y) = 0$ and $\int \mathbb{1}_{\sigma t_{i,n}(i)\neq y(i)}dp_j(y) = 1$. It follows that

$$C = 2(1-\theta)^2.$$

For any $k \in [n]$ and $l \in D_i$, let us note

$$U_i(k,l) := \int \mathbb{1}_{\sigma t_{i,n}(k)\neq y(k)}dp_i^l(y), \quad \text{and} \quad U_j(k) := \int \mathbb{1}_{\sigma t_{i,n}(k)\neq y(k)}dp_j(y).$$

By the Cauchy-Schwarz inequality, one has

$$A \leq \frac{1}{d}\sum_{l\in D_i}\left[\theta \sum_{k\in[n]\setminus \operatorname{supp}(\tilde{t}_{i,j})} U_i^2(k,l) + (1-\theta)\sum_{k\in[n]\setminus \operatorname{supp}(\tilde{t}_{i,j})} U_j^2(k)\right].$$

We also have

$$\begin{aligned}
B &= \sum_{k\in D_i}\left(\frac{\theta}{d}U_i(k,k) + (1-\theta)U_j(k) + \frac{\theta}{d}\sum_{l\in D_i\setminus\{k\}} U_i(k,l)\right)^2 \\
&\leq \sum_{k\in D_i}\left[d\left(\frac{\theta}{d}U_i(k,k) + (1-\theta)U_j(k)\right)^2 + \frac{\theta^2}{d}\sum_{l\in D_i\setminus\{k\}} U_i^2(k,l)\right] \\
&\leq \sum_{k\in D_i}\left[\frac{2\theta^2}{d}U_i^2(k,k) + 2d(1-\theta)^2 + \frac{\theta^2}{d}\sum_{l\in D_i\setminus\{k\}} U_i^2(k,l)\right] \\
&\leq 2d^2(1-\theta)^2 + \frac{\theta}{d}\sum_{l\in D_i}\left[2U_i^2(l,l) + \sum_{k\in D_i\setminus\{l\}} U_i^2(k,l)\right]
\end{aligned}$$

All the above estimates together provide

$$\begin{aligned}
A + B + C &\leq (2d^2+2)(1-\theta)^2 \\
&\quad + \frac{1}{d}\sum_{l\in D_i}\left[\theta\left(2U_i^2(l,l) + \sum_{k\in[n]\setminus\{i,l\}} U_i^2(k,l)\right) + (1-\theta)\sum_{k\in[n]\setminus \operatorname{supp}(\tilde{t}_{i,j})} U_j^2(k)\right].
\end{aligned}$$

Observe that

$$d = \deg(\tilde{t}_{i,j}) - 1 = \deg(t_{j,n}^{-1}t_{i,n}) - 1 \leq 2\ell - 2.$$

Therefore, according to the definition of $c(\ell)$, one has $2d^2 + 2 \leq c(\ell)^2$. As a consequence we get from all estimates above, by optimizing over all $p_i^l \in \mathcal{P}(H_i)$ and all $p_j \in \mathcal{P}(H_j)$,

$$Q^i\varphi(\sigma t_{i,n}) \leq \frac{1}{d}\sum_{l\in D_i}\left[\theta V_l + (1-\theta)W_j\right] + \frac{1}{2}(1-\theta)^2,$$

with

$$\begin{aligned}
V_l &:= \inf_{p_i\in\mathcal{P}(H_i)} \left\{ \int \varphi dp_i + \frac{1}{c(\ell)^2}\left(\int \mathbb{1}_{\sigma t_{i,n}(l)\neq y(l)} dp_i(y)\right)^2 \right.\\
&\qquad\qquad \left. + \frac{1}{2c(\ell)^2}\sum_{k\in[n]\setminus\{i,l\}}\left(\int \mathbb{1}_{\sigma t_{i,n}(k)\neq y(k)} dp_i(y)\right)^2 \right\}\\
&= \inf_{q_i\in\mathcal{P}(H_n)} \left\{ \int \varphi^{t_{i,n}} dq_i + \frac{1}{c(\ell)^2}\left(\int \mathbb{1}_{\sigma(t_{i,n}(l))\neq y(t_{i,n}(l))} dq_i(y)\right)^2 \right.\\
&\qquad\qquad \left. + \frac{1}{2c(\ell)^2}\sum_{k\in[n-1]\setminus\{t_{i,n}(l)\}}\left(\int \mathbb{1}_{\sigma(k)\neq y(k)} dq_i(y)\right)^2 \right\}\\
&= Q^{H_n,t_{i,n}(l)}\varphi^{t_{i,n}}(\sigma)
\end{aligned}$$

and

$$\begin{aligned}
W_j &:= \inf_{p_j\in\mathcal{P}(H_j)} \left\{ \int \varphi dp_j + \frac{1}{2c(\ell)^2}\sum_{k\in[n]\setminus\ \mathrm{supp}(\tilde{t}_{i,j})}\left(\int \mathbb{1}_{\sigma t_{i,n}(k)\neq y(k)} dp_j(y)\right)^2 \right\}\\
&= \inf_{q_j\in\mathcal{P}(H_n)} \left\{ \int \varphi^{t_{j,n}} dq_j + \frac{1}{2c(\ell)^2}\sum_{k\in[n]\setminus\ \mathrm{supp}(t_{j,n}^{-1}t_{i,n})}\left(\int \mathbb{1}_{\sigma t_{i,n}(k)\neq y t_{j,n}(k)} dq_j(y)\right)^2 \right\}\\
&\leq \inf_{q_j\in\mathcal{P}(H_n)} \left\{ \int \varphi^{t_{j,n}} dq_j + \frac{1}{2c(\ell)^2}\sum_{k\in[n]\setminus\{i\}}\left(\int \mathbb{1}_{\sigma t_{i,n}(k)\neq y t_{i,n}(k)} dq_j(y)\right)^2 \right\}\\
&= \inf_{q_j\in\mathcal{P}(H_n)} \left\{ \int \varphi^{t_{j,n}} dq_j + \frac{1}{2c(\ell)^2}\sum_{k\in[n-1]}\left(\int \mathbb{1}_{\sigma(k)\neq y(k)} dq_j(y)\right)^2 \right\}\\
&= Q^{H_n}\varphi^{t_{j,n}}(\sigma)
\end{aligned}$$

where we used successively the following arguments: $H_n t_{j,n} = H_j$; if $k \in [n] \setminus \mathrm{supp}(t_{j,n}^{-1}t_{i,n})$ then $t_{i,n}(k) = t_{j,n}(k)$; $[n] \setminus \mathrm{supp}(t_{j,n}^{-1}t_{i,n}) \subset [n] \setminus \{i\}$. This ends the proof of part (2) of Lemma 2.5.

The proof of part (3) Lemma 2.5 is identical replacing $\tilde{t}_{i,j}$ by $t_{i,j}$. In that case

$$2d^2 + 2 \leq 2(\ell-1)^2 + 2 = c^2(\ell).$$

Then, the only minor change is for the last step

$$\begin{aligned}
W_j &:= \inf_{p_j\in\mathcal{P}(H_j)} \left\{ \int \varphi dp_j + \frac{1}{2c(\ell)^2} \sum_{k\in[n]\setminus \operatorname{supp}(t_{i,j})} \left(\int \mathbb{1}_{\sigma t_{i,n}(k)\neq y(k)} dp_j(y) \right)^2 \right\} \\
&= \inf_{q_j\in\mathcal{P}(H_n)} \left\{ \int \varphi^{t_{i,n}t_{i,j}^{-1}} dq_j + \frac{1}{2c(\ell)^2} \sum_{k\in[n]\setminus \operatorname{supp}(t_{i,j})} \left(\int \mathbb{1}_{\sigma t_{i,n}(k)\neq y t_{i,n} t_{i,j}^{-1}(k)} dq_j(y) \right)^2 \right\} \\
&\leq \inf_{q_j\in\mathcal{P}(H_n)} \left\{ \int \varphi^{t_{i,n}t_{i,j}^{-1}} dq_j + \frac{1}{2c(\ell)^2} \sum_{k\in[n]\setminus\{i\}} \left(\int \mathbb{1}_{\sigma t_{i,n}(k)\neq y t_{i,n}(k)} dq_j(y) \right)^2 \right\} \\
&= \inf_{q_j\in\mathcal{P}(H_n)} \left\{ \int \varphi^{t_{i,n}t_{i,j}^{-1}} dq_j + \frac{1}{2c(\ell)^2} \sum_{k\in[n-1]} \left(\int \mathbb{1}_{\sigma(k)\neq y(k)} dq_j(y) \right)^2 \right\} \\
&= Q^{H_n} \varphi^{t_{i,n}t_{i,j}^{-1}}(\sigma)
\end{aligned}$$

where we used successively the following arguments: $H_n t_{i,n} t_{i,j}^{-1} = H_j$; if $k \in [n] \setminus \operatorname{supp}(t_{i,j})$ then $t_{i,j}(k) = k$; $[n] \setminus \operatorname{supp}(t_{i,j}) \subset [n] \setminus \{i\}$. The proof of Lemma 2.5 is completed. □

We will now prove (29) by induction over n. For $n = 2$, G is the two points space S_2 which is 2-local. For $i \in \{1, 2\}$, and for any $p \in \mathcal{P}(G)$,

$$\begin{aligned}
&\frac{1}{c(2)^2} \left(\int \mathbb{1}_{\sigma(i)\neq y(i)} dp(y) \right)^2 + \frac{1}{2c(2)^2} \sum_{k,k\neq i} \left(\int \mathbb{1}_{\sigma(k)\neq y(k)} dp(y) \right)^2 \\
&\qquad = \frac{3}{8} \left(\int \mathbb{1}_{\sigma\neq y} dp(y) \right)^2 \leq \frac{1}{2} \left(\int \mathbb{1}_{\sigma\neq y} dp(y) \right)^2.
\end{aligned}$$

As a consequence, we get the expected result from Lemma 2.2 applied with $\mathcal{X} = G$.

We will now present the induction step. We assume that (29) holds at the rank $n-1$ for all $j \in \{1, \ldots, n-1\}$.

Let us first explain that it suffices to prove (29) for $j = n$. For any $t \in S_n$, let $G^{(t)} = t^{-1}Gt$. The isomorphism $c_t : G \to G^{(t)}$, $\sigma \mapsto t^{-1}\sigma t$ pushes forward the measure μ on the measure $\mu^{(t)} := c_t\#\mu \in \mathcal{P}(G^{(t)})$, and conversely $\mu = c_{t^{-1}}\#\mu^{(t)}$. Let

$j \in [n]$. For any $\sigma \in G^{(t)}$ and any real function φ on G, one has

$$\begin{aligned}(Q^j\varphi)\circ c_{t^{-1}}(\sigma) &= \inf_{p\in\mathcal{P}(G)}\left\{\int \varphi\, dp + \frac{1}{c(\ell)^2}\left(\int \mathbb{1}_{t\sigma t^{-1}(j)\neq y(j)}\, dp(y)\right)^2\right.\\ &\qquad\left. + \frac{1}{2c(\ell)^2}\sum_{k\in[n]}\left(\int \mathbb{1}_{t\sigma t^{-1}(k)\neq y(k)}\, dp(y)\right)^2\right\}\\ &= \inf_{q\in\mathcal{P}(G^{(t)})}\left\{\int \varphi\circ c_{t^{-1}}\, dq + \frac{1}{c(\ell)^2}\left(\int \mathbb{1}_{t\sigma t^{-1}(j)\neq tyt^{-1}(j)}\, dq(y)\right)^2\right.\\ &\qquad\left. + \frac{1}{2c(\ell)^2}\sum_{k\in[n]}\left(\int \mathbb{1}_{t\sigma t^{-1}(k)\neq tyt^{-1}(k)}\, dq(y)\right)^2\right\}\\ &= \inf_{q\in\mathcal{P}(G^{(t)})}\left\{\int \varphi\circ c_{t^{-1}}\, dq + \frac{1}{c(\ell)^2}\left(\int \mathbb{1}_{\sigma t^{-1}(j)\neq yt^{-1}(j)}\, dq(y)\right)^2\right.\\ &\qquad\left. + \frac{1}{2c(\ell)^2}\sum_{k\in[n]}\left(\int \mathbb{1}_{\sigma(k)\neq y(k)}\, dq(y)\right)^2\right\}\\ &= Q^{t^{-1}(j)}(\varphi\circ c_{t^{-1}})(\sigma).\end{aligned}$$

From this observation, by choosing $t^{-1} = t_{jn}$, and setting $\psi = \varphi\circ c_{t^{-1}}$, one has

$$\begin{aligned}&\left(\int_G e^{\alpha Q^j\varphi} d\mu\right)^{1/\alpha}\left(\int_G e^{-(1-\alpha)\varphi} d\mu\right)^{1/(1-\alpha)}\\ &= \left(\int_{G^{(t)}} e^{\alpha(Q^j\varphi)\circ c_{t^{-1}}} d\mu^{(t)}\right)^{1/\alpha}\left(\int_{G^{(t)}} e^{-(1-\alpha)\varphi\circ c_{t^{-1}}} d\mu^{(t)}\right)^{1/(1-\alpha)}\\ &= \left(\int_{G^{(t)}} e^{\alpha Q^n\psi} d\mu^{(t)}\right)^{1/\alpha}\left(\int_{G^{(t)}} e^{-(1-\alpha)\psi} d\mu^{(t)}\right)^{1/(1-\alpha)}\end{aligned}$$

If we assume that G is a normal subgroup of S_n and that μ satisfies the second property of (10), then $G^{(t)} = G$ and $\mu^{(t)} = \mu$. Therefore the above expression is bounded by 1 as soon as (29) holds for $j = n$. If we assume G is a ℓ-local group and $\mu = \mu_o$ is the uniform law on G, then $G^{(t)}$ is also a ℓ-local group and $\mu^{(t)}$ is exactly the uniform law on $G^{(t)}$. Therefore the last expression is bounded by 1 as soon as (29) holds with $j = n$ for any uniform law on a ℓ-local group. As a conclusion, it remains to prove inequality (29) for $j = n$.

We may assume that $O_n \neq \{n\}$, otherwise the induction step is obvious. We first apply Lemma 2.4, by the first property of (10) satisfied by μ,

$$\int e^{\alpha Q^n\varphi} d\mu = \int e^{\alpha Q^{\sigma(n)}\varphi^{\{-1\}}(\sigma^{-1})} d\mu(\sigma) = \int e^{\alpha Q^{\sigma^{-1}(n)}\varphi^{\{-1\}}(\sigma)} d\mu(\sigma).$$

Let $g = \varphi^{\{-1\}}$. According to the decomposition of the measure μ on the sets $H_i, i \in O_n$,

$$\int e^{\alpha Q^n\varphi} d\mu = \sum_{i\in O_n} \hat{\nu}_n(i) \int e^{\alpha Q^i g} d\mu_i. \tag{31}$$

For $k \in O_n$, let us note

$$\hat{g}(k) := \log\left(\int e^{-(1-\alpha)g} d\mu_k\right)^{-1/(1-\alpha)}.$$

We choose $j \in O_n$ such that

$$\min_{k\in O_n} \hat{g}(k) = \hat{g}(j).$$

By property (21) and then applying Lemma 2.5 (1), we get

$$\int e^{\alpha Q^j g} d\mu_j = \int e^{\alpha Q^j g(\sigma t_{i,n})} d\mu_n(\sigma) \le \int e^{\alpha Q^{H_n} g^{t_{j,n}}} d\mu_n.$$

By the induction hypotheses applied to the measure μ_n on the subgroup $H_n = G_{n-1}$, it follows that

$$\begin{aligned}\int e^{\alpha Q^j g} d\mu_j &\le \left(\int e^{-(1-\alpha) g^{t_{jn}}}\, d\mu_n\right)^{-\alpha/(1-\alpha)}\\ &= \left(\int e^{-(1-\alpha)g}\, d\mu_j\right)^{-\alpha/(1-\alpha)} = e^{\alpha \hat{g}(j)}. \qquad (32)\end{aligned}$$

Let us now consider $i \neq j$, $i \in O_n$. When G is a normal subgroup of S_n, property (21), the second part of Lemma 2.5 and Jensen's inequality yield: for any $\theta \in [0, 1]$,

$$\int e^{\alpha Q^i g} d\mu_i = \int e^{\alpha Q^i g(\sigma t_{i,n})} d\mu_n(\sigma)$$
$$\le \exp\left\{\frac{1}{d} \sum_{l \in t_{i,n}(D_i)} \left[\theta \log \int e^{\alpha Q^{H_n,l} g^{t_{i,n}}} d\mu_n + (1-\theta) \log \int e^{\alpha Q^{H_n} g^{t_{j,n}}} d\mu_n\right] + \frac{\alpha}{2}(1-\theta)^2\right\}$$

By the induction hypotheses applied with the measure μ_n on the normal subgroup $G_{n-1} = H_n$ of S_{n-1}, and from property (21), it follows that

$$\int e^{\alpha Q^i g} d\mu_i \le \exp\left\{\theta\alpha\hat{g}(i) + (1-\theta)\alpha\hat{g}(j) + \frac{\alpha}{2}(1-\theta)^2\right\}. \qquad (33)$$

We get the same inequality when G is a ℓ-local group and $\mu = \mu_o$ is the uniform law on G, by using property (22), the third part of Lemma 2.5 and the induction hypotheses applied to the uniform measure μ_n on the ℓ-local subgroup $G_{n-1} = H_n$.

According to the definition (28) of the infimum-convolution operator $\widetilde{R}\hat{g}$ defined on the space $\mathcal{X} = O_n$, we may easily check that for every $i \in O_n$,

$$\widetilde{R}\hat{g}(i) = \inf_{\theta\in[0,1]} \left\{\theta\hat{g}(i) + (1-\theta)\min_{k\in O_n} \hat{g}(k) + \frac{1}{2}(1-\theta)^2\right\}.$$

Therefore optimizing over all $\theta \in [0, 1]$, we get from (32) and (33): for all $i \in O_n$,

$$\int e^{\alpha Q^i g} d\mu_i \le e^{\alpha \widetilde{R}\hat{g}(i)}.$$

Finally, from Lemma 2.2 applied with the measure $\nu = \hat\nu_n$ on O_n, the equality (31) gives

$$\int e^{\alpha Q^n\varphi} d\mu \le \int e^{\alpha \widetilde{R}\hat g}\, d\hat\nu_n \le \left(\int e^{-(1-\alpha)\hat g}\, d\hat\nu_n\right)^{-\alpha/(1-\alpha)} = \left(\sum_{i\in O_n} \hat\nu_n(i)\int e^{-(1-\alpha)g}\, d\mu_i\right)^{-\alpha/(1-\alpha)}$$
$$= \left(\int e^{-(1-\alpha)g}\, d\mu\right)^{-\alpha/(1-\alpha)} = \left(\int e^{-(1-\alpha)\varphi}\, d\mu\right)^{-\alpha/(1-\alpha)}.$$

The proof of (29) is completed. □

3. Transport-entropy inequalities on the slice of the cube.

Proof of (a) in Theorem 1.4. We adapt to the space $\mathcal{X}_{k,n-k}$ the proof of (a) in Theorem 1.3. In order to avoid redundancy, we only present the main steps of the proof.

By duality, it suffices to prove that for all functions φ on $\mathcal{X}_{k,n-k}$ and all $\lambda \ge 0$,

$$\int e^{\lambda Q\varphi} d\mu_{k,n-k} \le e^{\int \lambda\varphi\, d\mu_{k,n-k} + C_{k,n-k}\lambda^2/2}, \tag{34}$$

where

$$Q\varphi(x) = \inf_{p\in\mathcal{P}(\mathcal{X}_{k,n-k})}\left\{\int \varphi dp + \int d_h(x,y)\, dp(x)\right\}, \qquad x\in\mathcal{X}_{k,n-k},$$

and for any $0<\alpha<1$,

$$\left(\int e^{\alpha \widetilde{Q}_{C_{k,n-k}}\varphi} d\mu\right)^{1/\alpha}\left(\int e^{-(1-\alpha)\varphi} d\mu\right)^{1/(1-\alpha)} \le 1, \tag{35}$$

where for $t>0$,

$$\widetilde{Q}_t\varphi(x) = \inf_{p\in\mathcal{P}(\mathcal{X}_{k,n-k})}\left\{\int \varphi\, dp + \frac{1}{2t}\left(\int d_h(x,y)\, dp(y)\right)^2\right\}, \quad x\in\mathcal{X}_{k,n-k}.$$

The proof is by induction over n and $0\le k\le n$.

For any $n\ge 1$, if $k=n$ or $k=0$, the set $\mathcal{X}_{k,n-k}$ is reduced to a singleton and the inequalities (34) or (35) are obvious.

For $n=2$ and $k=1$, $\mathcal{X}_{k,n-k}$ is a two points set, (34) and (35) directly follows from property (24) and Lemma 2.2 on $\mathcal{X}=\mathcal{X}_{1,1}$.

For the induction step, we consider the collection of subset $\Omega_{i,j}$, with $i,j\in\{1,\dots,n\}, i\ne j$, defined by

$$\Omega^{i,j} := \left\{x\in\mathcal{X}_{k,n-k}, x_i = 0, x_j = 1\right\}.$$

Since for any $x\in\mathcal{X}_{k,n-k}$,

$$\sum_{(i,j),i\ne j} \mathbb{1}_{\Omega^{i,j}}(x) = k(n-k),$$

any probability measure p on $\mathcal{X}_{k,n-k}$ admits a unique decomposition defined by

$$p = \sum_{(i,j),i\ne j} \hat p(i,j) p^{i,j}, \qquad \text{with} \quad p^{i,j} = \frac{\mathbb{1}_{\Omega^{i,j}} p}{p(\Omega^{i,j})} \qquad \text{and} \quad \hat p(i,j) = \frac{p(\Omega^{i,j})}{k(n-k)}.$$

Thus, we define probability measures $p^{i,j} \in \mathcal{P}(\Omega^{i,j})$ and a probability measure $\hat{p}$ on the set $I(n) = \{(i,j) \in \{1,\ldots,n\}^2, i \neq j\}$. For the uniform law μ on $\mathcal{X}_{k,n-k}$, one has

$$\mu = \frac{1}{n(n-1)} \sum_{(i,j)\in I(n)} \mu^{i,j},$$

where $\mu^{i,j}$ is the uniform law on $\Omega^{i,j}$, $\mu_{i,j}(x) = \binom{n-2}{k-1}$, for any $x \in \Omega^{i,j}$.

For any $(i,j),(l,m) \in I(n)$, let $s_{(i,j),(l,m)} : \mathcal{X}_{k,n-k} \to \mathcal{X}_{k,n-k}$ denote the map that exchanges the coordinates x_i by x_l and x_j by x_m for any point $x \in \mathcal{X}_{k,n-k}$. This map is one to one from $\Omega^{i,j}$ to $\Omega^{l,m}$. For any $(i,j) \in I(n)$, the set $\Omega^{i,j}$ can be identify to $\mathcal{X}_{k-1,n-k-1}$ and therefore the induction hypotheses apply for the uniform law $\mu^{i,j}$ on $\Omega^{i,j}$ with Hamming distance

$$d_h^{i,j}(x,y) = \frac{1}{2} \sum_{k\in[n]\setminus\{i,j\}} \mathbb{1}_{x_k\neq y_k}, \qquad x,y \in \Omega^{i,j}.$$

For any function $f : \Omega^{i,j} \to \mathbb{R}$ and any $x \in \Omega^{i,j}$, we define

$$Q^{\Omega^{i,j}} f(x) := \inf_{p\in\mathcal{P}(\Omega^{i,j})} \left\{ \int f\,dp + \int d_h^{i,j}(x,y)\,dp(y) \right\},$$

and

$$\widetilde{Q}_t^{\Omega^{i,j}} f(x) := \inf_{p\in\mathcal{P}(H_n)} \left\{ \int f\,dp + \frac{1}{2t} \left(\int d_h^{i,j}(x,y)\,dp(x) \right)^2 \right\}.$$

The key lemma of the proof that replaces Lemma 2.1 and 2.3 is the following.

Lemma 3.1. *For any function $\varphi : \Omega^{i,j} \to \mathbb{R}$ and any $x \in \Omega^{i,j}$, one has*

$$Q\varphi(x) \leq \inf_{\hat{p}\in\mathcal{P}(I(n))} \left\{ \sum_{(l,m)\in I(n)} Q^{\Omega^{i,j}}(\varphi \circ s_{(i,j),(l,m)})(x)\hat{p}(l,m) + \sum_{(l,m)\in I(n)} \mathbb{1}_{(l,m)\neq(i,j)}\hat{p}(l,m) \right\},$$

and

$$\widetilde{Q}_{C_{k,n-k}}\varphi(x) \leq \inf_{\hat{p}\in\mathcal{P}(I(n))} \left\{ \sum_{(l,m)\in I(n)} \widetilde{Q}^{\Omega^{i,j}}_{C_{k-1,n-k-1}}(\varphi \circ s_{(i,j),(l,m)})(x)\hat{p}(l,m) + \frac{1}{2}\Big(\sum_{(l,m)\in I(n)} \mathbb{1}_{(l,m)\neq(i,j)}\hat{p}(l,m) \Big)^2 \right\}.$$

The proof of this lemma is obtained by decomposition of the measures $p \in \mathcal{P}(\mathcal{X}_{k,n-k})$ on the sets $\Omega^{i,j}$, and using the following inequality

$$d_h(x,y) \leq d_h^{i,j}(x, s_{(i,j),(l,m)}(y)) + d_h(s_{(i,j),(l,m)}(y), y) \leq d_h^{i,j}(x, s_{(i,j),(l,m)}(y)) + 2,$$

for any $x \in \Omega^{i,j}$, $y \in \Omega^{l,m}$.

Finally, the proof of the induction step based on Lemma 3.1 and the identity $C_{k,n-k} = C_{k-1,n-k-1} + 1$, is left to the reader. □

Proof of (b) in Theorem 1.4. We will explain the projection argument on the dual formulations of the transport-entropy inequalities. According to Proposition 4.5

and Theorem 9.5 of [GRST15], the weak transport-entropy inequality (14) is equivalent to the following property that we want to establish: for any real function f on $\mathcal{X}_{k,n-k}$ and for any $0 < \alpha < 1$,

$$\left(\int e^{\alpha \widehat{Q}f} d\mu_{k,n-k}\right)^{1/\alpha} \left(\int e^{-(1-\alpha)f} d\mu_{k,n-k}\right)^{1/(1-\alpha)} \leq 1, \tag{36}$$

where

$$\widehat{Q}f(x) := \inf_{p\in\mathcal{P}(\mathcal{X}_{k,n-k}}\left\{\int \varphi\, dp + \frac{1}{8}\sum_{k=1}^{n}\left(\int \mathbb{1}_{x_k\neq y_k}\, dp(y)\right)^2\right\}, \qquad x\in \mathcal{X}_{k,n-k}.$$

Let us apply property (13) to the function $f\circ P : S_n \to \mathbb{R}$. Since $\mu_{k,n-k} = P\#\mu$, we get

$$\left(\int e^{\alpha \widehat{Q}(f\circ P)} d\mu\right)^{1/\alpha} \left(\int e^{-(1-\alpha)f} d\mu_{k,n-k}\right)^{1/(1-\alpha)} \leq 1.$$

The inequality (36) is an easy consequence of the following result.

Lemma 3.2. *For any $\sigma\in S_n$, $\widehat{Q}(f\circ P)(\sigma) \geq \widehat{Q}f(P(\sigma))$.*

It remains to prove this lemma. By definition, one has

$$\begin{aligned}
\widehat{Q}(f\circ P)(\sigma) &= \inf_{p\in\mathcal{P}(S_n)}\left\{\int f\circ P\, dp + \sum_{j=1}^{n}\left(\int \mathbb{1}_{\sigma(j)\neq\tau(j)} dp(\tau)\right)^2\right\}\\
&= \inf_{q\in\mathcal{P}(\mathcal{X}_{k,n-k})}\inf_{p\in S_n, P\#p=q}\left\{\int f\circ P\, dp + \sum_{j=1}^{n}\left(\int \mathbb{1}_{\sigma(j)\neq\tau(j)} dp(\tau)\right)^2\right\}\\
&= \inf_{q\in\mathcal{P}(\mathcal{X}_{k,n-k})}\left\{\int f\, dq + \inf_{p\in S_n, P\#p=q}\left[\sum_{j=1}^{n}\left(\int \mathbb{1}_{\sigma(j)\neq\tau(j)} dp(\tau)\right)^2\right]\right\}.
\end{aligned}$$

Let $p\in S_n$ such that $P\#p = q$.

$$\int \mathbb{1}_{\sigma(j)\neq\tau(j)} dp(\tau) = \sum_{y\in\mathcal{X}_{k,n-k}}\sum_{\tau\in S_n} \mathbb{1}_{P(\tau)=y,\sigma(j)\neq\tau(j)}\, p(\tau).$$

For $y\in\mathcal{X}_{k,n-k}$, let us note $Y = \{i\in[n], y_i = 1\}$. Then $P(\tau) = y$ if and only if $\tau([k]) = Y$.

Assume that $j\in[k]$, if $\tau([k]) = Y$ and $\sigma(j)\notin Y$ then $\tau(j)\neq\sigma(j)$. Therefore one has

$$\{\tau, \tau([k]) = Y, \sigma(j)\notin Y\} \subset \{\tau, P(\tau) = y, \sigma(j)\neq\tau(j)\}.$$

Assume now that $j\notin[k]$, if $\tau([k]) = Y$ and $\sigma(j)\in Y$ then we also have $\tau(j)\neq\sigma(j)$. It follows that

$$\{\tau, \tau([k]) = Y, \sigma(j)\in Y\} \subset \{\tau, P(\tau) = y, \sigma(j)\neq\tau(j)\}.$$

From these observations, we get

$$\begin{aligned}
\sum_{j=1}^{n}\left(\int \mathbb{1}_{\sigma(j)\neq\tau(j)}dp(\tau)\right)^2 &\geq \sum_{j\in[k]}\left(\int \mathbb{1}_{P(\tau)=y,\sigma(j)\notin Y}dp(\tau)\right)^2+\sum_{j\in[n]\setminus[k]}\left(\int \mathbb{1}_{P(\tau)=y,\sigma(j)\in Y}dp(\tau)\right)^2\\
&=\sum_{j\in[k]}\left(\int \mathbb{1}_{\sigma(j)\notin Y}dq(y)\right)^2+\sum_{j\in[n]\setminus[k]}\left(\int \mathbb{1}_{\sigma(j)\in Y}dq(y)\right)^2\\
&=\sum_{i\in\sigma([k])}\left(\int \mathbb{1}_{i\notin Y}dq(y)\right)^2+\sum_{i\notin\sigma([k])}\left(\int \mathbb{1}_{i\in Y}dq(y)\right)^2\\
&=\sum_{i\in\sigma([k])}\left(\int \mathbb{1}_{y_i=0}dq(y)\right)^2+\sum_{i\notin\sigma([k])}\left(\int \mathbb{1}_{y_i=1}dq(y)\right)^2
\end{aligned}$$

Setting $x=P(\sigma)$, it follows that

$$\begin{aligned}
\sum_{j=1}^{n}\left(\int \mathbb{1}_{\sigma(j)\neq\tau(j)}dp(\tau)\right)^2 &\geq \sum_{i=1}^{n}\left[\mathbb{1}_{x_i=1}\left(\int \mathbb{1}_{y_i=0}dq(y)\right)^2+\mathbb{1}_{x_i=0}\left(\int \mathbb{1}_{y_i=1}dq(y)\right)^2\right]\\
&=\sum_{i=1}^{n}\left(\int \mathbb{1}_{y_i\neq x_i}dq(y)\right)^2.
\end{aligned}$$

This inequality provides

$$\widehat{Q}(f\circ P)(\sigma)\geq \widehat{Q}f(x)=\widehat{Q}f(P(\sigma)).$$

The proof of Lemma 3.2 and (b) in Theorem 1.4 is completed. □

4. Appendix

Proof of Lemma 1.1. Let $\mathcal{T}=(t_{i_j,j})$ be a ℓ-local base of a group of permutations $G=G_n$. In order to prove that the map

$$U_{\mathcal{T}}:\begin{array}{ccc} O_2\times O_3\times\cdots\times O_n & \to & G\\ i_2,i_3,\dots,i_n & \mapsto & t_{i_2,2}t_{i_3,3}\cdots t_{i_n,n},\end{array}\tag{37}$$

is one to one, it suffises to construct its inverse.

For any $j\in\{2,\dots,n\}$, let U_j denotes the map defined by

$$U_j(i_2,i_3,\dots,i_j)=t_{i_2,2}t_{i_3,3}\cdots t_{i_j,j}.$$

Let $\sigma=\sigma^{(n)}\in G$. We want to find the unique vector $(i_1,\dots,i_n)\in O_1\times\cdots\times O_n$ such that

$$U_n(i_1,\dots,i_n)=U_{\mathcal{T}}(i_1,\dots,i_n)=\sigma.$$

Since $U_n(i_1,\dots,i_n)(i_n)=n$, necessarily, one has to fixe $i_n=(\sigma^{(n)})^{-1}(n)$. i_n belongs to O_n. Let $\sigma^{(n-1)}=\sigma^{(n)}t^{-1}_{(\sigma^{(n)})^{-1}(n),n}$. On has $\sigma^{(n-1)}\in G_{n-1}$. Then, since

$$n-1=U_{n-1}(i_1,\dots,i_{n-1})(i_{n-1})=\sigma^{(n-1)}(i_{n-1}),$$

we necessarily have $i_{n-1}=(\sigma^{(n-1)})^{-1}(n-1)\in O_{n-1}$. We set

$$\sigma^{(n-2)}=\sigma^{(n-1)}t^{-1}_{(\sigma^{(n-1)})^{-1}(n-1),n-1}\in G_{n-2}.$$

Following this induction procedure, we construct a family of permutations $\sigma^{(j)} \in G_j$ for $j \in [n]$, such that $i_j = (\sigma^{(j)})^{-1}(j) \in O_j$ for all $j \in \{2, \ldots, n\}$. Observing that $G_1 = \{Id\}$, it follows that $\sigma^{(1)} = id$ and therefore

$$\sigma = \sigma^{(n)} = t_{i_2,2} t_{i_3,3} \cdots t_{i_n,n}.$$

This ends the proof of Lemma 1.1. □

Proof of Lemma 1.2. Let $G = G_n$ be a ℓ-local group. From the definition of the ℓ-local property, it is clear that any of the subgroup G_j, $j \in \{2, \ldots, n\}$ is ℓ-local. As a consequence, for any $i_j \in O_j$, $i_j \neq j$, there exists $t_{i_j,j} \in G_j$ such that

$$t_{i_j,j}(i_j) = j, \quad \text{and} \quad \deg(t_{i_j j}) \leq \ell.$$

This completes the proof of Lemma 1.2. □

Proof of Lemma 2.2. Let $\alpha \in (0,1)$ and f be a real function on the finite set $\mathcal{X}$. We want to show that for any probability measure ν on $\mathcal{X}$,

$$\left(\int e^{\alpha \widetilde{R}^\alpha f} d\nu\right)^{1/\alpha} \left(\int e^{-(1-\alpha)h} d\nu\right)^{1/(1-\alpha)} \leq 1.$$

We will apply the following lemma whose proof is given at the end of this section.

Lemma 4.1. *Let F be a real function on $\mathcal{X}$ and $K \in \mathbb{R}$. Let us consider the set*

$$C := \left\{\nu \in \mathcal{P}(\mathcal{X}), \int F\, d\nu = K\right\}.$$

If C is not empty, then the extremal points of this convex set are Dirac measures or convex combinations of two Dirac measures on $\mathcal{X}$.

Given a real function f on $\mathcal{X}$, for any $K \in \mathbb{R}$, let

$$C_K = \left\{\nu \in \mathcal{P}(\mathcal{X}), \int e^{-(1-\alpha)f}\, d\nu = K\right\}.$$

One has

$$\sup_{\nu \in \mathcal{P}(\mathcal{X})} \left(\int e^{\alpha \widetilde{R}^\alpha f} d\nu\right)^{1/\alpha} \left(\int e^{-(1-\alpha)f} d\nu\right)^{1/(1-\alpha)} = \sup_{K, C_K \neq \emptyset} \left(\sup_{\nu \in C_K} \int e^{\alpha \widetilde{R}^\alpha f} d\nu\right)^{1/\alpha} K^{1/(1-\alpha)}$$

The supremum of the linear function $\nu \mapsto \int e^{\alpha \widetilde{R}^\alpha f} d\nu$ on the non empty convex set C_K is reached at an extremal point of C_K. Therefore, by Lemma 4.1, we get

$$\sup_{\nu \in \mathcal{P}(\mathcal{X})} \left(\int e^{\alpha \widetilde{R}^\alpha f} d\nu\right)^{1/\alpha} \left(\int e^{-(1-\alpha)h} d\nu\right)^{1/(1-\alpha)}$$

$$= \sup_{x,y \in \mathcal{X}} \sup_{\lambda \in [0,1]} \left((1-\lambda) e^{\alpha \widetilde{R}^\alpha f(x)} + \lambda e^{\alpha \widetilde{R}^\alpha f(y)}\right)^{1/\alpha} \left((1-\lambda) e^{-(1-\alpha)f(x)} + \lambda e^{-(1-\alpha)f(y)}\right)^{1/(1-\alpha)}$$

Now, let x and y be some fixed points of $\mathcal{X}$. It remains to show that for any real function f on E and for any $x, y \in \mathcal{X}$,

$$\left((1-\lambda) e^{\alpha \widetilde{R}^\alpha f(x)} + \lambda e^{\alpha \widetilde{R}^\alpha f(y)}\right)^{1/\alpha} \left((1-\lambda) e^{-(1-\alpha)f(x)} + \lambda e^{-(1-\alpha)f(y)}\right)^{1/(1-\alpha)} \leq 1.$$

The left-hand side of this inequality is invariant by translation of the function f by a constant. Therefore, by symmetry, we may assume that $0 = f(y) \le f(x)$. It follows that $\widetilde{R}^\alpha f(y) = 0$. Therefore we want to check that for any non-negative function f on $\{x, y\}$, for any $\lambda \in [0, 1]$,

$$\left((1-\lambda)e^{\alpha \widetilde{R}^\alpha f(x)} + \lambda\right)^{1/\alpha} \left((1-\lambda)e^{-(1-\alpha)f(x)} + \lambda\right)^{1/(1-\alpha)} \le 1,$$

or equivalently, setting $\psi(\lambda) = \left((1-\lambda)e^{-(1-\alpha)f(x)} + \lambda\right)^{-\alpha/(1-\alpha)} - \lambda$,

$$e^{\alpha \widetilde{R}^\alpha f(x)} \le \inf_{\lambda \in [0,1)} \frac{\psi(\lambda) - \psi(1)}{1-\lambda} = -\psi'(1) = \frac{\alpha}{1-\alpha}\left(1 - e^{-(1-\alpha)f(x)}\right) + 1,$$

since ψ is a convex function on $[0, 1]$.

So, it suffices to check that $\widetilde{R}^\alpha f(x) \le \phi(f(x))$, where

$$\phi(h) = \frac{1}{\alpha}\log\left(\frac{\alpha}{1-\alpha}\left(1 - e^{-(1-\alpha)h}\right) + 1\right), \qquad h \ge 0.$$

The function ϕ is concave and $\phi(0) = 0$. For all $h \ge 0$, one has

$$\phi'(h) = \frac{1-\alpha}{e^{(1-\alpha)h} - \alpha}.$$

The function ϕ' is a bijection from $[0, +\infty)$ to $(0, 1]$. It follows that

$$\phi(h) = \inf_{\theta \in (0,1]} \{\theta h + c_\alpha(1-\theta)\}, \quad h \ge 0,$$

where c_α is the convex function defined by

$$c_\alpha(1-\theta) = \sup_{h \in [0,+\infty)} \{-\theta h + \phi(h)\}, \quad \theta \in (0, 1].$$

After computations, we get

$$c_\alpha(u) := \frac{\alpha(1-u)\log(1-u) - (1-\alpha u)\log(1-\alpha u)}{\alpha(1-\alpha)},$$

and therefore we exactly have for any $x \in \mathcal{X}$,

$$\phi(f(x)) = \inf_{\theta \in [0,1]} \{\theta f(x) + c_\alpha(1-\theta)\} = \widetilde{R}^\alpha f(x).$$

The proof of Lemma 2.2 is completed. □

Proof of Lemma 4.1. We will show that, if $\nu \in C$ is a convex combination of three probability measures ν_1, ν_2, ν_3,

$$\nu = \alpha_1 \nu_1 + \alpha_2 \nu_2 + \alpha_3 \nu_3,$$

with $\alpha_1 \ne 0$, $\alpha_2 \ne 0$, $\alpha_3 \ne 0$, and $\alpha_1 + \alpha_2 + \alpha_3 = 1$, and $\nu_1(\mathcal{X}) > 0$, $\nu_2(\mathcal{X}) > 0$, $\nu_3(\mathcal{X}) > 0$, then there exists two measures $\hat{\nu}_1, \hat{\nu}_2$ in C and $\lambda \in [0, 1]$ such that

$$\nu = \lambda \hat{\nu}_1 + (1-\lambda)\hat{\nu}_2.$$

Setting $F_i = \int F d\nu_i$, for $i = 1, 2, 3$, we may assume, without loss of generality, that $F_1 \le F_2 \le F_3$. Then one has either $F_1 \le K \le F_2$, either $F_2 \le K \le F_3$.

We will assume that $F_1 \leq K \leq F_2$. The case $F_2 \leq K \leq F_3$ can be treated identically and the proof in that case is let to the reader. Since $F_1 \leq K \leq F_2$ and $F_1 \leq K \leq F_3$, there exists $\beta, \gamma \in [0, 1]$ such that

$$K = \beta F_1 + (1 - \beta)F_2 \quad \text{and} \quad K = \gamma F_1 + (1 - \gamma)F_3. \tag{38}$$

If $F_1 = F_3$ then $F_1 = F_2 = F_3 = K$ and therefore $\nu_1, \nu_2, \nu_3 \in C$. We may choose $\lambda = \alpha_1$, $\hat{\nu}_1 = \nu_1$ and $\hat{\nu}_2 = \frac{\alpha_2\nu_2+\alpha_3\nu_3}{\alpha_2+\alpha_3}$.

If $F_1 = F_2$ then necessarily $F_1 = F_2 = F_3 = K$ and we are reduced to the previous case.

So, we may now assume that $F_1 \neq F_3$ and $F_1 \neq F_2$ and therefore $F_1 < K \leq F_2 \leq F_3$. In that case, we exactly have

$$\beta = \frac{F_2 - K}{F_2 - F_1} \quad \text{and} \quad \gamma = \frac{F_3 - K}{F_3 - F_1}.$$

Let us choose

$$\lambda = \frac{\alpha_2}{1 - \beta} = \alpha_2 \frac{F_2 - F_1}{K - F_1}, \quad \hat{\nu}_1 = \beta\nu_1 + (1 - \beta)\nu_2, \quad \hat{\nu}_2 = \gamma\nu_1 + (1 - \gamma)\nu_2.$$

The equalities (38) ensure that $\hat{\nu}_1 \in C$ and $\hat{\nu}_2 \in C$. The proof of Lemma 4.1 ends by checking that $\lambda\hat{\nu}_1 + (1 - \lambda)\hat{\nu}_2 = \hat{\mu}$. One has

$$\lambda\hat{\nu}_1 + (1 - \lambda)\hat{\nu}_2 = (\lambda\beta + (1 - \lambda)\gamma)\nu_1 + \lambda(1 - \beta)\nu_2 + (1 - \lambda)(1 - \gamma)\nu_3. \tag{39}$$

According to the definitions of λ, β, γ, we may easily check that $\lambda(1 - \beta) = \alpha_2$, and

$$(1 - \lambda)(1 - \gamma) = \frac{K - F_1}{F_3 - F_1} - \alpha_2 \frac{F_2 - F_1}{F_3 - F_1}.$$

Since $\hat{\mu} \in C$, one has $(1 - (\alpha_2 + \alpha_3))F_1 + \alpha_2 F_2 + \alpha_3 F_3$ and therefore

$$(1 - \lambda)(1 - \gamma) = \alpha_3.$$

As a consequence $\lambda\beta + (1 - \lambda)\gamma = 1 - \alpha_2 - \alpha_3 = \alpha_1$ and according to (39), we get

$$\lambda\hat{\nu}_1 + (1 - \lambda)\hat{\nu}_2 = \alpha_1\nu_1 + \alpha_2\nu_2 + \alpha_3\nu_3 = \nu.$$

$\square$

Acknowledgements. The author thanks R. Adamczak and D. Chafaï for useful discussions, that lead to consider the Ewens distribution, and the number of cycles of given lenght, as examples to illustrate the results of this paper.

References

[ABT03] R. Arratia, A. D. Barbour, and S. Tavaré. *Logarithmic combinatorial structures: a probabilistic approach*. EMS Monographs in Mathematics. European Mathematical Society (EMS), Zürich, 2003.

[ACW14] R. Adamczak, D. Chafaï, and P. Wolff. Circular law for random matrices with exchangeable entries. *ArXiv e-prints*, February 2014.

[BDR15] B. Bercu, B. Delyon, and E. Rio. *Concentration inequalities for sums and martingales*. SpringerBriefs in Mathematics. Springer, Cham, 2015.

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

[Csi67] I. Csiszár. Information-type measures of difference of probability distributions and indirect observations. *Studia Sci. Math. Hungar.*, 2:299–318, 1967.

[DP09] D. Dubhashi and A. Panconesi. *Concentration of Measure for the Analysis of Randomized Algorithms*. Cambridge University Press, New York, NY, USA, 1st edition, 2009.

[GRST15] N. Gozlan, C. Roberto, P.-M. Samson, and P. Tetali. Kantorovich duality for general transport costs and applications. *ArXiv e-prints*, December 2015.

[HNNZ13] C. Hughes, J. Najnudel, A. Nikeghbali, and D. Zeindler. Random permutation matrices under the generalized Ewens measure. *Ann. Appl. Probab.*, 23(3):987–1024, 2013.

[Hoe63] W. Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58(301):13–30, March 1963.

[JKB97] N. L. Johnson, S. Kotz, and N. Balakrishnan. *Discrete multivariate distributions*. Wiley Series in Probability and Statistics: Applied Probability and Statistics. John Wiley & Sons, Inc., New York, 1997. A Wiley-Interscience Publication.

[Kul67] S. Kullback. Lower bound for discrimination information in terms of variation. *IEEE Trans. Information Theory*, 4:126127, 1967.

[Led01] M. Ledoux. *The concentration of measure phenomenon*, volume 89 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 2001.

[LM03] M. J. Luczak and C. McDiarmid. Concentration for locally acting permutations. *Discrete Math.*, 265(1-3):159–171, 2003.

[Mar03] K. Marton. Measure concentration and strong mixing. *Studia Sci. Math. Hungar.*, 40(1-2):95–113, 2003.

[Mau79] B. Maurey. Construction de suites symétriques. *C. R. Acad. Sci. Paris Sér. A-B*, 288(14):A679–A681, 1979.

[McD89] C. McDiarmid. On the method of bounded differences. In *Surveys in combinatorics, 1989 (Norwich, 1989)*, volume 141 of *London Math. Soc. Lecture Note Ser.*, pages 148–188. Cambridge Univ. Press, Cambridge, 1989.

[McD02] C. McDiarmid. Concentration for independent permutations. *Combin. Probab. Comput.*, 11(2):163–178, 2002.

[MNZ12] K. Maples, A. Nikeghbali, and D. Zeindler. On the number of cycles in a random permutation. *Electron. Commun. Probab.*, 17:no. 20, 13, 2012.

[MR02] M. Molloy and B. Reed. *Graph Colouring and the Probabilistic Method*. Algorithms and Combinatorics. Springer, 2002.

[MS86] V. D. Milman and G. Schechtman. *Asymptotic theory of finite-dimensional normed spaces*, volume 1200 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 1986. With an appendix by M. Gromov.

[Pau14] D. Paulin. The convex distance inequality for dependent random variables, with applications to the stochastic travelling salesman and other problems. *Electron. J. Probab.*, 19:no. 68, 34, 2014.

[Pin64] M. S. Pinsker. *Information and information stability of random variables and processes*. Translated and edited by Amiel Feinstein. Holden-Day Inc., San Francisco, Calif., 1964.

[Pit06] J. Pitman. *Combinatorial stochastic processes*, volume 1875 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 2006. Lectures from the 32nd Summer School on Probability Theory held in Saint-Flour, July 7–24, 2002, With a foreword by Jean Picard.

[Sam07] P.-M. Samson. Infimum-convolution description of concentration properties of product probability measures, with applications. *Ann. Inst. H. Poincaré Probab. Statist.*, 43(3):321–338, 2007.

[Sam16] P.-M. Samson. *High dimensionnal concentration phenomena, transport of measure and functional inequalities.* Accreditation to supervise research, Université Paris-Est, June 2016.

[Sam17] P.-M. Samson. Concentration of measure principle and entropy-inequalities. In E. Carlen, M. Madiman, and E. M Werner, editors, *Convexity and Concentration*, pages 55–10. Springer New Yor, New York, N, 2017.

[Tal95] M. Talagrand. Concentration of measure and isoperimetric inequalities in product spaces. *Inst. Hautes Études Sci. Publ. Math.*, (81):73–205, 1995.

[Tal96] M. Talagrand. Transportation cost for Gaussian and other product measures. *Geom. Funct. Anal.*, 6(3):587–600, 1996.

Samson P.-M., Université Paris-Est, Laboratoire d'Analyse et de Mathématiques Appliquées (UMR 8050), UPEM, UPEC, CNRS, F-77454, Marne-la-Vallée, France

E-mail address: paul-marie.samson@u-pem.fr