



**HAL**  
open science

# Transport-entropy inequalities on locally acting groups of permutations

Paul-Marie Samson

► **To cite this version:**

Paul-Marie Samson. Transport-entropy inequalities on locally acting groups of permutations. 2016.  
hal-01370491v1

**HAL Id: hal-01370491**

**<https://hal.science/hal-01370491v1>**

Preprint submitted on 22 Sep 2016 (v1), last revised 23 Jun 2017 (v3)

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# 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-McDiarmid concentration properties are consequences of these transport-entropy inequalities. We also consider transport-entropy inequalities on  $G$  for a larger class of measures. 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. The transport-entropy inequalities settled in this paper are new examples, in discrete setting, of weak transport-entropy inequalities introduced in [GRST14].

## 1. INTRODUCTION

Let  $S_n$  denote the symmetric group of permutations acting on a set  $\Omega$  of cardinality  $n$ , and  $\mu$  denote the uniform law on  $S_n$ ,  $\mu(\sigma) := \frac{1}{n!}$ ,  $\sigma \in S_n$ . 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  $\mu(A) \geq 1/2$ , and for all  $t \geq 0$ , one has*

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

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

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

---

*Date:* September 22, 2016.

*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 .

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

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

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  $\text{conv}(V(\sigma, A))$  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

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

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

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

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

Maurey’s concentration result easily follows by observing that

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

and applying Tchebychev inequality with usual optimization arguments.

Talagrand’s result has been first extended to 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  $\sigma \in S_n$ , the *support* of  $\sigma$ , denoted by  $\text{supp}(\sigma)$ , is the set  $\{i \in \Omega, \sigma(i) \neq i\}$  and the *degree* of  $\sigma$ , denoted by  $\text{deg}(\sigma)$ , is the cardinality of  $\text{supp}(\sigma)$ ,  $\text{deg}(\sigma) = |\text{supp}(\sigma)|$ . By definition, according to [LM03], a subgroup  $G$  of  $S_n$  is  $\ell$ -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  $\text{supp}(\tau) \subset \text{supp}(\sigma)$ ,  $\text{deg}(\tau) \leq \ell$  and  $\tau(i) = j$ .

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  $G$ . We also prove weaker type of transport entropy inequalities for 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  $[n] := \{1, \dots, n\}$ . Let  $(i, j)$  denote the transposition in  $S_n$  that exchanges the elements  $i$  and  $j$  in  $[n]$ . One may easily

check that the map

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

is one to one.

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

$$\mathcal{M} := \{T\#\hat{\nu}, \hat{\nu} = \hat{\nu}_2 \otimes \cdots \otimes \hat{\nu}_n \text{ with } \hat{\nu}_i \in \mathcal{P}([i]), \forall i \in [n]\},$$

where by definition  $T\#\hat{\nu}(C) = \nu(T^{-1}(C))$  for any subset  $C$  in  $S_n$ . The uniform measure  $\mu$  on  $S_n$  belongs to  $\mathcal{M}$  since  $\mu = T\#\hat{\mu}$  with  $\hat{\mu} = \hat{\mu}_2 \otimes \cdots \otimes \hat{\mu}_n$ , where for each  $i$ ,  $\hat{\mu}_i$  denotes the uniform probability measure on  $[i]$ .

Let us now construct the class of measures  $\mathcal{M}$  for any  $\ell$ -local group  $G$ . To clarify the notations, the elements of  $\Omega$  are labelled with integers,  $\Omega = [n]$ .

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

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

Let us observe that for any  $j \in [n]$ ,  $j \in \text{orb}(j)$  since the identity  $id$  belongs to  $G$ .

**Lemma 1.1.** *Let  $G$  be a  $\ell$ -local subgroup of  $S_n$ . For any  $j \in \{2, \dots, n\}$  let  $I_j = \text{orb}(j) \cap [j]$ . There exists a fixed family of permutations  $\mathcal{T} = (t_{i,j})$  with  $j \in \{2, \dots, n\}$  and  $i_j \in I_j$  with  $t_{j,j} = id$  and for every  $i_j \neq j$*

$$t_{i,j}(i_j) = j, \quad \text{supp}(t_{i,j}) \subset [j], \quad \text{and} \quad \deg(t_{i,j}) \leq \ell,$$

such that the map

$$(1) \quad \begin{array}{ccc} I_2 \times I_3 \times \cdots \times I_n & \rightarrow & G \\ T_{\mathcal{T}} : (i_2, i_3, \dots, i_n) & \mapsto & t_{i_2,2} t_{i_3,3} \cdots t_{i_n,n}, \end{array}$$

is one to one. Such a family  $\mathcal{T}$  is called “ $\ell$ -local base of  $G$ ” in the present paper.

As for the symmetric group, the class  $\mathcal{M}$  is made up of all probability measures on  $G$  which are pushed forward of product probability measures on  $I_2 \times I_3 \times \cdots \times I_n$  by a map  $T_{\mathcal{T}}$  defined by (1),

$$(2) \quad \mathcal{M} := \{T_{\mathcal{T}}\#\hat{\nu}, \hat{\nu} = \hat{\nu}_2 \otimes \cdots \otimes \hat{\nu}_n \text{ with } \hat{\nu}_i \in \mathcal{P}(I_i), \forall i \in \{2, \dots, n\}\},$$

where  $\mathcal{T}$  is a  $\ell$ -local base of  $G$ . The class  $\mathcal{M}$  contains the uniform law on  $G$ , obtained by choosing for each  $\hat{\nu}_i$  the uniform law on  $I_i$ .

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

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

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

The terminology “weak transport-entropy” introduced in [GRST14], 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  $\mu$  on a Polish metric space  $\mathcal{X}$ , namely

$$(3) \quad \|\mu - \nu\|_{TV}^2 \leq 2H(\nu|\mu), \quad \forall \nu \in \mathcal{P}(\mathcal{X}).$$

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

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

where the supremum runs over all measurable subset  $A$  of  $\mathcal{X}$ . We refer to the survey [Sam16b, Sam16a] 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 measures in the class  $\mathcal{M}$ , that recover the concentration results of Theorems 1.1 and 1.2.

We 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  $G$ ,  $t_1, \dots, t_k$ , with degree less than  $\ell$ , 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  $\nu_1$  and second marginal  $\nu_2$ . We denote by  $W_1(\nu_1, \nu_2)$  the Wasserstein distance between  $\nu_1$  and  $\nu_2$  according to the distance  $d = d_H$  or  $d = d_T$ .

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

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

$$(4) \quad \tilde{T}_2(\nu_2|\nu_1) := \inf_{\pi \in \Pi(\nu_1, \nu_2)} \int \left( \int d(\sigma, \tau) dp_\sigma(\tau) \right)^2 d\nu_1(\sigma),$$

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  $p_\sigma$  represents a probability measure such that  $\pi(\sigma, \tau) = \nu_1(\sigma)p_\sigma(\tau)$ . By Jensen's inequality, these weak transport costs are comparable, namely

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

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

**Theorem 1.3.** *Let  $G$  be a  $\ell$ -local subgroup of  $S_n$ .*

(a) *Let  $\mu$  be a probability measure of the set  $\mathcal{M}$  defined by (2). Then, for all probability measures  $\nu_1$  and  $\nu_2$  on  $G$ ,*

$$(5) \quad \frac{2}{K_n c(\ell)^2} W_1^2(\nu_1, \nu_2) \leq \left( \sqrt{H(\nu_1|\mu)} + \sqrt{H(\nu_2|\mu)} \right)^2,$$

and

$$(6) \quad \frac{1}{2K_n c(\ell)^2} \tilde{T}_2(\nu_2|\nu_1) \leq \left( \sqrt{H(\nu_1|\mu)} + \sqrt{H(\nu_2|\mu)} \right)^2,$$

where

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

When  $\mu$  is the uniform law on  $G$ , the inequalities (5) and (6) hold with

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

(b) Let  $\mu$  denotes the uniform law on  $G$ . Then, for all probability measures  $\nu_1$  and  $\nu_2$  on  $G$ ,

$$(7) \quad \frac{1}{2c(\ell)^2} \widehat{\mathcal{T}}_2(\nu_2|\nu_1) \leq \left( \sqrt{H(\nu_1|\mu)} + \sqrt{H(\nu_2|\mu)} \right)^2,$$

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

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  $\ell$ -local groups [LM03].

**Comments :**

- Here is a more popular dual formulation of the transport-entropy inequality (5): for all 1-Lipschitz functions  $\varphi : G \rightarrow \mathbb{R}$  (with respect to the distance  $d$ ),

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

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].

- Similarly, by Proposition 4.5 and Theorem 2.7 of [GRST14] 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 (6) is equivalent to the following dual property: for any real function  $\varphi$  on  $G$  and for any  $0 < \alpha < 1$ ,

$$(9) \quad \left( \int e^{\alpha \tilde{Q}_{K_n} \varphi} d\mu \right)^{1/\alpha} \left( \int e^{-(1-\alpha)\varphi} d\mu \right)^{1/(1-\alpha)} \leq 1,$$

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

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

Moreover, let us observe that following our proof of (9) in the next section, for each  $\alpha \in (0, 1)$  the inequality (9) 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, (9) holds replacing  $\tilde{Q}_{K_n}\varphi$  by  $\tilde{Q}_{K_n}^\alpha\varphi$  defined by

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

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

- Proposition 4.5 and Theorem 9.5 of [GRST14] also provide a dual formulation of the weak transport-entropy inequality (7): for any real function  $\varphi$  on  $G$  and for any  $0 < \alpha < 1$ ,

$$(10) \quad \left( \int e^{\alpha \widehat{Q}\varphi} d\mu \right)^{1/\alpha} \left( \int e^{-(1-\alpha)\varphi} d\mu \right)^{1/(1-\alpha)} \leq 1,$$

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 (10) directly provides the following version of the Talagrand's concentration result for the uniform law on  $G$ .

**Corollary 1.1.** *Let  $\mu$  denotes the uniform law on a  $\ell$ -local subgroup  $G$  of  $S_n$ . 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  $c(\ell)^2 = 2(\ell - 1)^2 + 2$ . As a consequence, by Tchebychev inequality, for any  $\alpha \in (0, 1)$  and all  $t \geq 0$ ,

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

For  $\alpha = 1/2$ , 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 to the uniform law  $\mu$  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  $\mathcal{X}_{k, n-k}$ ,  $k \in \{0, \dots, n\}$ , the slice of discrete cube defined by

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

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

$$\begin{aligned} S_n &\rightarrow \mathcal{X}_{k, n-k} \\ U : \sigma &\mapsto \mathbb{1}_{\sigma([k])}, \end{aligned}$$

where  $\sigma([k]) := \{\sigma(1), \dots, \sigma(k)\}$  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,  $\mu_{k,n-k} = U\#\mu$  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

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

**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  $\nu_1$  and  $\nu_2$  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

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

where  $W_1$  is the Wasserstein distance associated to  $d_h$ ,  $\tilde{T}_2$  is the weak optimal transport cost defined by (4) with  $d = d_h$ , and  $C_{k,n-k} = \min(k, n-k)$ .

(b) *For all probability measures  $\nu_1$  and  $\nu_2$  on  $\mathcal{X}_{k,n-k}$ ,*

$$(11) \quad \frac{1}{8} \widehat{T}_2(\nu_2|\nu_1) \leq \left( \sqrt{H(\nu_1|\mu_{k,n-k})} + \sqrt{H(\nu_2|\mu_{k,n-k})} \right)^2,$$

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 (11) is the stronger one since for all  $\nu_1, \nu_2 \in \mathcal{P}(\mathcal{X}_{k,n-k})$ ,

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

The proof of Theorem 1.4 is given in section 3. The transport-entropy inequality (11) is derived by projection from the transport-entropy inequality (7) for the uniform measure  $\mu$  on  $S_n$ . The same projection argument could be used to reach the results of (a) from the transport-entropy inequality of item (a) of Theorem 1.3, but it provides worse constants. The constant  $C_{k,n-k}$  in (a) 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  $E = \{e_1, \dots, e_m\}$  be a set of cardinality  $m$  and let  $k_1, \dots, k_m$  be a collection of non-zero integers satisfying  $k_1 + \dots + k_m = n$ . The multinomial law  $\mu_{k_1, \dots, k_m}$  is by definition the uniform law on the set

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



For any  $x \in \mathcal{X}_{k_1, \dots, k_m}$ , one has  $\mu_{k_1, \dots, k_m}(x) = \frac{k_1! \cdots k_m!}{n!}$ . As a result, the weak transport-entropy inequality (11) holds on  $\mathcal{X}_{k_1, \dots, k_m}$  replacing the measure  $\mu_{k, n-k}$  by the measure  $\mu_{k_1, \dots, k_m}$ . The proof of this result is a simple generalization of the one on  $\mathcal{X}_{k, n-k}$ , by using the projection map  $V : S_n \rightarrow \mathcal{X}_{k_1, \dots, k_m}$  defined by:  $V(\sigma) = x$  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 let to the reader.

A straightforward application of transport-entropy inequalities of Theorem 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 the uniform law  $\mu$  on a  $\ell$ -local subgroup of  $S_n$ . Obviously some of these results extend to the class of measures  $\mathcal{M}$  and a similar corollary can be derived from Theorem 1.4 on the slices of the discrete cube. For any  $h : G \rightarrow \mathbb{R}$ , the mean of  $h$  is denoted by  $\mu(h) := \int h d\mu$ .

**Corollary 1.2.** *Let  $\mu$  be the uniform law on a  $\ell$ -local subgroup  $G$  of  $S_n$ . Let  $g$  be a real function on  $G$ .*

(a) *Assume that there exists a function  $\beta : G \rightarrow \mathbb{R}^+$  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*

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

*and*

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

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

(b) *Assume that there exist functions  $\alpha_k : G \rightarrow \mathbb{R}^+$ ,  $k \in \{1, \dots, n\}$  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  $u \geq 0$ , one has*

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

*and*

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

*where  $|\alpha(\sigma)|_2^2 := \sum_{k=1}^n \alpha_k^2(\sigma)$  and  $c(\ell)^2 = 2(\ell - 1)^2 + 2$ .*

**Comments :**

- The above deviation's bounds of  $g$  around its mean  $\mu(g)$  are directly derived from the dual representations (8),(9),(10) of the transport-entropy inequalities of Theorem 1.3, when  $\alpha$  goes to 0 or  $\alpha$  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 properties of type Theorem 1.1, one simply get the following bound under the condition of (b): for all  $u \geq 0$ ,

$$(12) \quad \mu(g \geq M(g) + u) \leq \frac{1}{2} \exp \left( -w \left( \frac{u}{\sqrt{2}c(\ell) \sup_{\sigma \in G} |\alpha(\sigma)|_2} \right) \right),$$

where  $w(h) = h(h - 2\sqrt{\log 2})$ ,  $h \geq 0$ , and  $c(\ell)^2 = 2(\ell - 1)^2 + 2$ .

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 under the conditions of (b),

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

Then, the deviation bound above the median directly follows from Theorem 1.1 by optimizing over all  $\alpha \in (0, 1)$ . With similar 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 (8). 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  $\tilde{T}_2$ . The same kind of improvement could be reached for the deviations above the mean under additional Lipschitz regularity conditions on the function  $\beta$ .
- Let  $\varphi : [0, 1]^n \rightarrow \mathbb{R}$  be a 1-Lipschitz convex function and let  $x = (x_1, \dots, x_n) \in [0, 1]^n$  be a fixed vector. For any  $\sigma \in G$ , let  $x_\sigma := (x_{\sigma(1)}, \dots, x_{\sigma(n)})$ . By applying the results of item (b) (or even (12)), to the particular function  $g_x(\sigma) = \varphi(x_\sigma)$ ,  $\sigma \in G$ , we recover and extend to any  $\ell$  subgroup  $G$ , 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  $x = (x_1, \dots, x_n) \in [0, 1]^n$ ,

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

This concentration property on  $S_n$  ( $\ell = 2$ ) 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.

- 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 will not be examined in the present paper. However, let us briefly mentioned 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 [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). Under the conditions on the function  $g$ , for any  $p \in \mathcal{P}(G)$

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

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

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

As  $\alpha$  goes to 1, (10) applied to the function  $\lambda g$  yields

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

and therefore

$$(13) \quad \int e^{\lambda g} d\mu \leq \exp \left( \lambda \mu(g) + \frac{\lambda^2 c(\ell)^2 \sup_{\sigma \in G} |\alpha(\sigma)|_2^2}{2} \right).$$

As  $\alpha$  goes to 0, (10) yields

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

and therefore

$$(14) \quad \int e^{-\lambda g} d\mu \leq \exp \left( -\lambda \mu(g) + \frac{\lambda^2 c(\ell)^2 \mu(|\alpha|_2^2)}{2} \right).$$

The deviation bounds of (b) then follows from (10) and (14) by Tchebychev inequality by optimizing over all  $\lambda \geq 0$ .

The deviation bounds of (a) are also obtained from (9) by Tchebychev inequality. As above, the improvement for the deviation under the mean is a consequence of (9) applied to  $\lambda g$ , as  $\alpha$  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  $\varphi_\lambda$  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

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

as  $\lambda \rightarrow \infty$ . It is easy to check that for any fixed  $\sigma$ , the function  $\psi(\cdot, \sigma)$  is continuous on  $[0, 1]$ , so letting  $\varepsilon$  go to 0, we get  $\liminf_{\lambda \rightarrow \infty} \widehat{Q}\varphi_\lambda(\sigma) \geq \psi(1, \sigma)$ . On the other hand,  $\widehat{Q}\varphi_\lambda(\sigma) \leq \psi(1, \sigma)$  for all  $\lambda > 0$ . This proves that  $\lim_{\lambda \rightarrow \infty} \widehat{Q}\varphi_\lambda(\sigma) = \psi(1, \sigma)$ . Applying (10) to  $\varphi_\lambda$  and letting  $\lambda$  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  $G$  be a  $\ell$ -local subgroup of  $S_n$  and  $\mathcal{T}_n = (t_{i,j}, j \in \{2, \dots, n\}, i_j \in I_j)$  be a  $\ell$ -local base of  $G$ . Let  $\mu$  be a probability measure in the set  $\mathcal{M}$ . By definition of  $\mathcal{M}$ , there exists a product probability measure  $\hat{\nu} = \hat{\nu}_1 \otimes \dots \otimes \hat{\nu}_n$  such that  $\mu = T_{\mathcal{T}_n} \# \hat{\nu}$  where the map  $T_{\mathcal{T}_n}$  is defined by (1).

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

$$(15) \quad H_i = \{ \sigma \in G, \sigma(i) = n \}.$$

If  $H_n = \{Id\}$ , then  $\text{orb}(n) = \{n\}$  and if  $H_n \neq \{Id\}$ , then  $\text{orb}(n) \supsetneq \{n\}$ . In any case, from the definition of the  $\ell$ -local property, we may easily check that  $H_n$  is a  $\ell$ -local group that realizes a natural embedding of a  $\ell$ -local subgroup of  $S_{n-1}$  and  $\mathcal{T}_{n-1}$  is a  $\ell$ -local base of this subgroup.

Moreover, it is rather easy to get that for any  $i \in \text{orb}(n)$ ,  $H_i$  is the coset defined by  $H_i = H_n t_{in}$ . From the definition of  $\mu$ , 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  $\mu_i$  denote the restriction of  $\mu$  to  $H_i$  defined by

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

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

$$(16) \quad \mu_n(\sigma) = \frac{\mu(\sigma)}{\mu(H_n)} = \frac{\mu(\sigma t_{i,n})}{\mu(H_i)} = \mu_i(\sigma t_{i,n}).$$

This property is needed in the induction step of the proofs.

Let us note that if  $i$  and  $l$  are elements of  $\text{orb}(n)$ , then from the  $\ell$ -local property of  $G$ , 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 the measure  $\mu$  is the uniform law on  $G$ , then for any  $i, l \in \text{orb}(n)$ ,  $\mu_i(H_i) = \mu_l(H_l) = \frac{|G|}{|\text{orb}(n)|}$ . In that case we will use in the proofs the following property: for any  $\sigma \in H_n$ ,  $\sigma t_{i,n} t_{i,l}^{-1} \in H_l$  and

$$(17) \quad \mu_n(\sigma) = \mu_l(\sigma t_{i,n} t_{i,l}^{-1}).$$

*Proof of item (a) of Theorem 1.3.* Since the distance  $W_1$  satisfies a triangular inequality, the transport-entropy inequality (5) follows from the following one: for all probability measures  $\nu_1$  on  $G$ ,

$$\frac{2}{K_n c(\ell)^2} W_1^2(\nu_1, \mu) \leq H(\nu_1 | \mu).$$

A dual formulation of this property given by Theorem 2.7 in [GRST14] and Proposition 3.1 in [Sam16a] is the following: for all functions  $\varphi$  on  $G$  and all  $\lambda \geq 0$ ,

$$(18) \quad \int e^{\lambda Q\varphi} d\mu \leq e^{\int \lambda \varphi d\mu + K_n c(\ell)^2 \lambda^2 / 8},$$

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 (18) by induction on  $n$ .

When  $n = 2$ , either  $G = \{id\}$  and the inequality (18) is obvious, either  $G = S_2$  is the two points space,  $\ell = 2$  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, (18) exactly corresponds to the following dual form of the Csiszar-Kullback-Pinsker inequality (3) (see Proposition 3.1 in [Sam16a]): for any probability measure  $\nu$  on a Polish space  $\mathcal{X}$ , for any measurable function  $f : \mathcal{X} \rightarrow \mathbb{R}$ ,

$$(19) \quad \int e^{\lambda R^c f} d\nu \leq e^{\lambda \int f d\nu + \lambda^2 c^2 / 8}, \quad \forall \lambda, c \geq 0,$$

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 recurrence step will be also a consequence of (19). Let  $(H_i)_{i \in \text{orb}(n)}$  be the partition of  $G$  defined by (15). Any  $p \in \mathcal{P}(G)$  admits a unique decomposition defined by

$$(20) \quad p = \sum_{i \in \text{orb}(n)} \hat{p}(i) p_i, \quad \text{with } p_i \in \mathcal{P}(H_i) \quad \text{and} \quad \hat{p}(i) = p(H_i).$$

Thus, we define a probability measure  $\hat{p}$  on  $\text{orb}(n)$ . In particular, according to the definition of the measure  $\mu \in \mathcal{M}$ , one has

$$\mu = \sum_{i \in \text{orb}(n)} \hat{\nu}_n(i) \mu_i.$$

It follows that

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

where the last equality is a consequence of property (16). Now, we will bound the right-hand side of this equality by using the induction hypotheses.

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

For any function  $f : H_n \rightarrow \mathbb{R}$  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  $i \in \text{orb}(n)$ , for any function  $\varphi : H_i \rightarrow \mathbb{R}$  and any  $\sigma \in H_n$ , one has*

$$(1) \quad Q\varphi(\sigma t_{i,n}) \leq \inf_{\hat{p} \in \mathcal{P}(\text{orb}(n))} \left\{ \sum_{l \in \text{orb}(n)} Q^{H_n} \varphi^{t_{i,n} l}(\sigma) \hat{p}(l) + c(\ell) \sum_{l \in \text{orb}(n)} \mathbb{1}_{l \neq i} \hat{p}(l) \right\},$$

where  $c(\ell) = \min(2\ell - 1, n)$  if  $d = d_H$  and  $c(\ell) = 2$  if  $d = d_T$ .

$$(2) \quad Q\varphi(\sigma t_{i,n}) \leq \inf_{\hat{p} \in \mathcal{P}(\text{orb}(n))} \left\{ \sum_{l \in \text{orb}(n)} Q^{H_n} \varphi^{t_{i,n} t_{i,l}^{-1}}(\sigma) \hat{p}(l) + c(\ell) \sum_{l \in \text{orb}(n)} \mathbb{1}_{l \neq i} \hat{p}(l) \right\},$$

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 (20) 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

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

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  $d = d_T$ ,

$$(21) \quad \int d(\sigma t_{i,n}, \tau) dp(\tau) \leq \sum_{l \in \text{orb}(n)} \int_{H_l} d(\sigma, \tau t_{l,n}^{-1}) dp_l(\tau) \hat{p}(l) + c(\ell) \sum_{l \in \text{orb}(n)} \mathbb{1}_{l \neq i} \hat{p}(l).$$

It follows that

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

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

$$\begin{aligned} \int d(\sigma t_{i,n}, \tau) dp(\tau) &= \sum_{l \in \text{orb}(n)} \int_{H_l} d(\sigma t_{i,n}, \tau) dp_l(\tau) \hat{p}(l) \\ &\leq \sum_{l \in \text{orb}(n)} \int d(\sigma t_{i,n}, \tau t_{i,l}) dp_l(\tau) \hat{p}(l) + \sum_{l \in \text{orb}(n)} \int_{H_l} d(\tau t_{i,l}, \tau) dp_l(\tau) \hat{p}(l) \\ &= \sum_{l \in \text{orb}(n)} \int d(\sigma, \tau t_{i,l} t_{i,n}^{-1}) dp_l(\tau) \hat{p}(l) + \sum_{l \in \text{orb}(n)} d(t_{i,l}, id) \hat{p}(l) \\ (22) \quad &\leq \sum_{l \in \text{orb}(n)} \int_{H_l} d(\sigma, \tau t_{i,l} t_{i,n}^{-1}) dp_l(\tau) \hat{p}(l) + c(\ell) \sum_{l \in \text{orb}(n)} \mathbb{1}_{l \neq i} \hat{p}(l), \end{aligned}$$

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 (18) continues by applying consecutively Lemma 2.1, the Hölder inequality, and the induction hypotheses to the measure  $\mu_n$ .

If  $I_n = \text{orb}(n) = \{n\}$  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  $I_n = \text{orb}(n) \neq \{n\}$  then  $K_n = K_{n-1} + 1$  and for any  $i \in \text{orb}(n)$ ,

$$\begin{aligned} \int e^{\lambda Q\varphi(\sigma_{t_i,n})} d\mu_n(\sigma) &\leq \inf_{\hat{p} \in \mathcal{P}(\text{orb}(n))} \left\{ \prod_{l \in \text{orb}(n)} \left( \int e^{\lambda Q^{H_n} \varphi^{t_l,n}} d\mu_n \right)^{\hat{p}(l)} e^{c(\ell)\lambda \sum_{l=1}^n \mathbb{1}_{l \neq i} \hat{p}(l)} \right\} \\ &\leq \exp \left[ \inf_{\hat{p} \in \mathcal{P}(\text{orb}(n))} \left\{ \lambda \sum_{l \in \text{orb}(n)} \left( \int \varphi^{t_l,n} d\mu_n \right) \hat{p}(l) + K_{n-1} c(\ell)^2 \frac{\lambda^2}{8} + c(\ell)\lambda \sum_{l \in \text{orb}(n)} \mathbb{1}_{l \neq i} \hat{p}(l) \right\} \right] \\ &= \exp \left[ \lambda \inf_{\hat{p} \in \mathcal{P}(\text{orb}(n))} \left\{ \sum_{l \in \text{orb}(n)} \hat{\varphi}(l) \hat{p}(l) + c(\ell) \sum_{l \in \text{orb}(n)} \mathbb{1}_{l \neq i} \hat{p}(l) \right\} + K_{n-1} c(\ell)^2 \frac{\lambda^2}{8} \right], \end{aligned}$$

where, by using property (16),  $\hat{\varphi}(l) = \int \varphi^{t_l,n} d\mu_n = \int \varphi d\mu_l$ . Let us consider again the above infimum-convolution  $R^c \hat{\varphi}$  defined on the space  $\mathcal{X} = \text{orb}(n)$ , with  $c = c(\ell)$ , one has

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

By applying (19) with the probability measure  $\hat{\nu}_n$  on  $\text{orb}(n)$ , the previous inequality gives

$$\begin{aligned} \int e^{\lambda Q\varphi} d\mu &= \sum_{i \in \text{orb}(n)} \hat{\nu}_n(i) \int e^{\lambda Q\varphi(\sigma_{t_i,n})} d\mu_n(\sigma) \leq \left( \sum_{i \in \text{orb}(n)} e^{\lambda R^c \hat{\varphi}(i)} \hat{\nu}_n(i) \right) e^{K_{n-1} \lambda^2 / 8} \\ &\leq \exp \left[ \sum_{i=1}^n \hat{\varphi}(i) \hat{\nu}_n(i) + \frac{\lambda^2 c(\ell)^2}{8} + K_{n-1} c(\ell)^2 \frac{\lambda^2}{8} \right] = \exp \left[ \lambda \int \varphi d\mu + K_n c(\ell)^2 \frac{\lambda^2}{8} \right]. \end{aligned}$$

The scheme of the induction proof of (18), with a better constant  $c(\ell)$  when  $\mu$  is the uniform measure on  $G$ , is identical, starting from the second result of Lemma 2.1 and using the property (17). This is let to the reader.

We now turn to the induction proof of the dual formulation (9) of the weak transport-entropy inequality (6). The sketch of the proof is identical to the one of (18).

For the initial step  $n = 2$ , in the non-trivial case  $G = S_2$  and  $\ell = 2$ , and one may easily check that

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

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

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

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

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

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



and  $c_\alpha$  is the convex function defined by

$$c_\alpha(1 - \theta) = \frac{\alpha(1 - u) \log(1 - u) - (1 - \alpha u) \log(1 - \alpha u)}{\alpha(1 - \alpha)}.$$

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

$$(23) \quad \tilde{R}f(x) = \inf_{p \in \mathcal{P}(\mathcal{X})} \left\{ \int f(y) dp(y) + \frac{1}{2} \left( \int \mathbb{1}_{x \neq y} dp(y) \right)^2 \right\}, \quad x \in \mathcal{X}.$$

The proof of this Lemma can be founded in [Sam07] (inequality (4)). For a sake of completeness, we give in 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 in the proof. For any function  $f : H_n \rightarrow \mathbb{R}$  and any  $\sigma \in H_n$ , we define

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

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

**Lemma 2.3.** *Let  $i \in \text{orb}(n)$ . For any function  $\varphi : H_i \rightarrow \mathbb{R}$  and any  $\sigma \in H_n$ , one has*

$$(1) \quad \tilde{Q}_{K_n} \varphi(\sigma t_{i,n}) \leq \inf_{\hat{p} \in \mathcal{P}(\text{orb}(n))} \left\{ \sum_{l \in \text{orb}(n)} \tilde{Q}_{K_{n-1}}^{H_n} \varphi^{t_{i,n}}(\sigma) \hat{p}(l) + \frac{1}{2} \left( \sum_{l \in \text{orb}(n)} \mathbb{1}_{l \neq i} \hat{p}(l) \right)^2 \right\},$$

with  $c(\ell) = \min(2\ell - 1, n)$  if  $d = d_H$  and  $c(\ell) = 2$  if  $d = d_T$ .

$$(2) \quad \tilde{Q}_{K_n} \varphi(\sigma t_{i,n}) \leq \inf_{\hat{p} \in \mathcal{P}(\text{orb}(n))} \left\{ \sum_{l \in \text{orb}(n)} \tilde{Q}_{K_{n-1}}^{H_n} \varphi^{t_{i,n} t_{i,l}^{-1}}(\sigma) \hat{p}(l) + \frac{1}{2} \left( \sum_{l \in \text{orb}(n)} \mathbb{1}_{l \neq i} \hat{p}(l) \right)^2 \right\},$$

where  $c(\ell) = \ell$  if  $d = d_H$  and  $c(\ell) = 1$  if  $d = d_T$ .

The proof of this lemma is similar to the one of Lemma 2.1. By (21) and the inequality

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

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

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

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

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

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 (22).

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

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

where the last equality is a consequence of property (16).

If  $\text{orb}(n) = \{n\}$ , then the result simply follows for the induction hypotheses apply to the measure  $\mu_n$ .

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

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

where by property (16),

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

According to the definition of the infimum convolution  $\tilde{R}\hat{\varphi}$  on the space  $\mathcal{X} = \text{orb}(n)$  given in Lemma 2.2, the last inequality is

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

and therefore Lemma 2.2 provides

$$\begin{aligned} \int e^{\alpha \tilde{Q}_{k_n} \varphi} d\mu &= \sum_{i \in \text{orb}(n)} e^{\alpha \tilde{R}\hat{\varphi}(i)} \hat{\nu}_n(i) \leq \left( \sum_{i \in \text{orb}(n)} e^{-(1-\alpha)\hat{\varphi}(i)} \hat{\nu}_n(i) \right)^{-\frac{\alpha}{1-\alpha}} \\ &= \left( \sum_{i \in \text{orb}(n)} \hat{\nu}_n(i) \int e^{-(1-\alpha)\varphi} d\mu_i \right)^{-\frac{\alpha}{1-\alpha}} = \left( \int e^{-(1-\alpha)\varphi} d\mu \right)^{-\frac{\alpha}{1-\alpha}}. \end{aligned}$$

The proof of (9) is completed for any measure  $\mu \in \mathcal{M}$ . To improve the constant when  $\mu$  is the uniform probability measure on  $G$ , the proof is similar using the second inequality of Lemma 2.3 together with property (17).  $\square$

*Proof of item (b) of Theorem 1.3.* We prove the dual equivalent property (10) as a consequence of the stronger following result, for any real function  $\varphi$  on  $G$ , for any  $j \in \{1, \dots, n\}$

$$(24) \quad \left( \int e^{\alpha Q^j \varphi} d\mu \right)^{1/\alpha} \left( \int e^{-(1-\alpha)\varphi} d\mu \right)^{1/(1-\alpha)} \leq 1,$$

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

$$(25) \quad Q^j \varphi(\sigma) = \inf_{p \in \mathcal{P}(G)} \left\{ \int \varphi dp + \frac{1}{c(\ell)^2} \left( \int \mathbb{1}_{\sigma(j) \neq y(j)} dp(y) \right)^2 + \frac{1}{2c(\ell)^2} \sum_{k \in [n] \setminus \{j\}} \left( \int \mathbb{1}_{\sigma(k) \neq y(k)} dp(y) \right)^2 \right\}.$$

The proof of (24) relies on Lemma 2.2 and the following. For this proof, 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  $j \in [n-1]$ ,

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

**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 (25) of  $Q^j\varphi(\sigma)$ , one has

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

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  $q := R\#p$  under the application  $R : \sigma \in S_n \mapsto \sigma^{-1}$ , 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 (24).

**Lemma 2.5.** (1) *Let  $j \in \text{orb}(n)$ . 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) *If  $\text{orb}(n) \neq \{n\}$ , let  $i, j \in \text{orb}(n)$  with  $i \neq j$ , and let  $t_{i,j} \in G$  such that  $t_{i,j}(i) = j$  and  $\deg(t_{i,j}) \leq \ell$ . We note  $D_i = \text{supp}(t_{i,j}) \setminus \{i\}$  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  $\sigma \in H_j$ ,

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

For the proof of the second part of Lemma 2.5, let us consider  $p'_i, l \in D_i$ , a collection of measures in  $\mathcal{P}(H_i)$ , and  $p_j \in \mathcal{P}(H_j)$  ( $j \neq i$ ). For  $\theta \in [0, 1]$ ,

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

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

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

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

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

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

We also have

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

All these estimates together provide

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

Since  $2d^2 + 2 \leq 2(\ell - 1)^2 + 2 = c(\ell)^2$ , we get by optimizing over all  $p_i^l \in \mathcal{P}(H_i)$  and all  $p_j \in \mathcal{P}(H_j)$ ,

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

with

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

and

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

where we use successively the following arguments:  $t_{i,j}(k) = k$  for any  $k \in [n] \setminus \text{supp}(t_{i,j})$ , the set  $[n] \setminus \text{supp}(t_{i,j})$  is a subset of  $[n] \setminus \{i\}$ , for any  $y \in H_j$  one has  $y t_{i,j} t_{i,n}^{-1} \in H_n$ . This ends the proof of Lemma 2.5.  $\square$

We will now prove (24) by induction over  $n$ . For  $n = 2$ , in the non-trivial case, the set  $G$  is the 2 points space  $S_2 = \{Id, t_{12}\}$  which is 2-local. For  $i \in \{1, 2\}$ , and for any  $p \in \mathcal{P}(G)$ ,

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

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

Let us now present the induction step. We assume that (24) holds at the rank  $n - 1$  for all  $j \in \{1, \dots, n - 1\}$ . By symmetry, we may assume without loss of generality that  $j = n$  in (24) since  $\mu$  is the uniform probability measure on  $G$ .

We assume that  $\text{orb}(n) \neq \{n\}$ , otherwise the induction step is obvious. We first apply Lemma 2.4,

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

Let  $g = \varphi^{(-1)}$ . According to the decomposition of the measure  $\mu$  on the sets  $H_i$ ,  $i \in \text{orb}(n)$ ,

$$(26) \quad \int e^{\alpha Q^n \varphi} d\mu = \frac{1}{|\text{orb}(n)|} \sum_{i \in \text{orb}(n)} \int e^{\alpha Q^i g} d\mu_i,$$

where  $\mu_i$  is the uniform probability measure on  $H_i$ . For  $k \in \text{orb}(n)$ , let us note

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

We choose  $j \in \text{orb}(n)$  such that

$$\min_{k \in \text{orb}(n)} \hat{g}(k) = \hat{g}(j).$$

By property (16) and then applying Lemma 2.5, we get

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

By the induction hypotheses applied to the measure  $\mu_n$ , it follows that

$$(27) \quad \begin{aligned} \int e^{\alpha Q^j g} d\mu_j &\leq \left( \int e^{-(1-\alpha)g^{i,j,n}} d\mu_n \right)^{-\alpha/(1-\alpha)} \\ &= \left( \int e^{-(1-\alpha)g} d\mu_j \right)^{-\alpha/(1-\alpha)} = e^{\alpha \hat{g}(j)}. \end{aligned}$$

Let us now consider  $i \neq j$ ,  $i \in \text{orb}(n)$ , property (16), the second part of Lemma 2.5 and Jensen's inequality yield: for any  $\theta \in [0, 1]$ ,

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

By the induction hypotheses applied with the uniform measure  $\mu_n$  on the  $\ell$ -local subgroup  $H_n$ , and from property (17), it follows that

$$(28) \quad \int e^{\alpha Q^i g} d\mu_i \leq \exp \left\{ \theta \alpha \hat{g}(i) + (1 - \theta) \alpha \hat{g}(j) + \frac{\alpha}{2} (1 - \theta)^2 \right\}.$$

According to the definition (23) of the infimum-convolution operator  $\tilde{R}\hat{g}$  defined on the space  $\mathcal{X} = \text{orb}(n)$ , we may easily check that for every  $i \in \text{orb}(n)$ ,

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

Therefore optimizing over all  $\theta \in [0, 1]$ , we get from (27) and (28): for all  $i \in \text{orb}(n)$ ,

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

Finally, from Lemma 2.2 applied with the uniform probability measure  $\hat{\mu}$  on  $\text{orb}(n)$ , the equality (26) gives

$$\begin{aligned} \int e^{\alpha Q^n \varphi} d\mu &\leq \int e^{\alpha \tilde{R}\hat{g}} d\hat{\mu} \leq \left( \int e^{-(1-\alpha)\hat{g}} d\hat{\mu} \right)^{-\alpha/(1-\alpha)} \\ &= \left( \frac{1}{|\text{orb}(n)|} \sum_{i \in \text{orb}(n)} \int e^{-(1-\alpha)g} d\mu_i \right)^{-\alpha/(1-\alpha)} = \left( \int e^{-(1-\alpha)g} d\mu \right)^{-\alpha/(1-\alpha)}. \end{aligned}$$

The proof of (24) is completed.  $\square$

### 3. TRANSPORT-ENTROPY INEQUALITIES ON THE SLICE OF THE CUBE.

*Proof of item (a) of Theorem 1.4.* We adapt to the space  $\mathcal{X}_{k,n-k}$  the proof of item (a) of 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  $\varphi$  on  $\mathcal{X}_{k,n-k}$  and all  $\lambda \geq 0$ ,

$$(29) \quad \int e^{\lambda Q \varphi} d\mu_{k,n-k} \leq e^{\int \lambda \varphi d\mu_{k,n-k} + C_{k,n-k} \lambda^2 / 2},$$

where

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

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

$$(30) \quad \left( \int e^{\alpha \tilde{Q} C_{k,n-k} \varphi} d\mu \right)^{1/\alpha} \left( \int e^{-(1-\alpha)\varphi} d\mu \right)^{1/(1-\alpha)} \leq 1,$$

where for  $t > 0$ ,

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

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

For any  $n \geq 1$ , if  $k = n$  or  $k = 0$ , the set  $\mathcal{X}_{k,n-k}$  is reduced to a singleton and the inequalities (29) or (30) are obvious.

For  $n = 2$  and  $k = 1$ ,  $\mathcal{X}_{k,n-k}$  is a two points set, (29) and (30) directly follows from property (19) 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  $i, j \in \{1, \dots, n\}, i \neq j$ , defined by

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



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

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

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

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

Thus, we define probability measures  $p^{i,j} \in \mathcal{P}(\Omega^{i,j})$  and a probability measure  $\hat{p}$  on the set  $I(n) = \{(i,j) \in \{1, \dots, n\}^2, i \neq j\}$ . For the uniform law  $\mu$  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  $s_{(i,j),(l,m)} : \mathcal{X}_{k,n-k} \rightarrow \mathcal{X}_{k,n-k}$  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

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

For any function  $f : \Omega^{i,j} \rightarrow \mathbb{R}$  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

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

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

**Lemma 3.1.** *For any function  $\varphi : \Omega^{i,j} \rightarrow \mathbb{R}$  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

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

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 by using Lemma 3.1 and the identity  $C_{k,n-k} = C_{k-1,n-k-1} + 1$  is left to the reader.  $\square$

*Proof of item (b) of 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 [GRST14], the weak transport-entropy inequality (11) 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  $0 < \alpha < 1$ ,

$$(31) \quad \left( \int e^{\alpha \widehat{Q}f} d\mu_{k,n-k} \right)^{1/\alpha} \left( \int e^{-(1-\alpha)f} d\mu_{k,n-k} \right)^{1/(1-\alpha)} \leq 1,$$

where

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

Let us apply property (10) to the function  $f \circ U : S_n \rightarrow \mathbb{R}$ . Since  $\mu_{k,n-k} = U\#\mu$ , we get

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

The inequality (31) is an easy consequence of the following result.

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

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

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

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

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

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

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

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

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

From these observations, we get

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

Setting  $x = U(\sigma)$ , 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

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

The proof of Lemma 3.2 and item (b) of Theorem 1.4 is completed.  $\square$

#### 4. APPENDIX

*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  $\nu$  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

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

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

$$\begin{aligned} & \sup_{\nu \in \mathcal{P}(\mathcal{X})} \left( \int e^{\alpha \tilde{R}^\alpha f} d\nu \right)^{1/\alpha} \left( \int e^{-(1-\alpha)f} d\nu \right)^{1/(1-\alpha)} \\ &= \sup_{x, y \in \mathcal{X}} \sup_{\lambda \in [0, 1]} \left( (1-\lambda)e^{\alpha \tilde{R}^\alpha f(x)} + \lambda e^{\alpha \tilde{R}^\alpha f(y)} \right)^{1/\alpha} \left( (1-\lambda)e^{-(1-\alpha)f(x)} + \lambda e^{-(1-\alpha)f(y)} \right)^{1/(1-\alpha)} \end{aligned}$$

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  $x, y \in \mathcal{X}$ ,

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

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  $0 = f(y) \leq f(x)$ . It follows that  $\tilde{R}^\alpha f(y) = 0$ . Therefore we want to check that for any non-negative function  $f$  on  $\{x, y\}$ , for any  $\lambda \in [0, 1]$ ,

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

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

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

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

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

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

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

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

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

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

where  $c_\alpha$  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  $x \in \mathcal{X}$ ,

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

The proof of Lemma 2.2 is completed.  $\square$

*Proof of Lemma 4.1.* We will show that, if  $\nu \in \mathcal{C}$  is a convex combination of three probability measures  $\nu_1, \nu_2, \nu_3$ ,

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

with  $\alpha_1 \neq 0, \alpha_2 \neq 0, \alpha_3 \neq 0$ , 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  $\mathcal{C}$  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  $F_1 \leq F_2 \leq F_3$ . Then one has either  $F_1 \leq K \leq F_2$ , either  $F_2 \leq K \leq F_3$ .

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

$$(32) \quad K = \beta F_1 + (1-\beta)F_2 \quad \text{and} \quad K = \gamma F_1 + (1-\gamma)F_3.$$

If  $F_1 = F_3$  then  $F_1 = F_2 = F_3 = K$  and therefore  $\nu_1, \nu_2, \nu_3 \in \mathcal{C}$ . 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 (32) ensure that  $\hat{\nu}_1 \in \mathcal{C}$  and  $\hat{\nu}_2 \in \mathcal{C}$ . The proof of Lemma 4.1 ends by checking that  $\lambda \hat{\nu}_1 + (1-\lambda) \hat{\nu}_2 = \hat{\mu}$ . One has

$$(33) \quad \lambda \hat{\nu}_1 + (1-\lambda) \hat{\nu}_2 = (\lambda\beta + (1-\lambda)\gamma) \nu_1 + \lambda(1-\beta) \nu_2 + (1-\lambda)(1-\gamma) \nu_3.$$

According to the definitions of  $\lambda, \beta, \gamma$ , 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  $\hat{\mu} \in \mathcal{C}$ , 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 (33), we get

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

□

## REFERENCES

- [ACW14] R. Adamczak, D. Chafaï, and P. Wolff. Circular law for random matrices with exchangeable entries. *ArXiv e-prints*, February 2014.
- [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.
- [GRST14] N. Gozlan, C. Roberto, P.-M. Samson, and P. Tetali. Kantorovich duality for general transport costs and applications. *ArXiv e-prints*, December 2014.
- [Hoe63] W. Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58(301):13–30, March 1963.
- [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.
- [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.
- [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.
- [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.
- [Sam16a] P.-M. Samson. Concentration of measure principle and entropy inequalities. working paper or preprint, April 2016.
- [Sam16b] P.-M. Samson. *High dimensionnal concentration phenomena, transport of measure and functional inequalities*. Accreditation to supervise research, Université Paris-Est, June 2016.
- [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