CATALYTIC MAJORIZATION AND ℓ_p NORMS

GUILLAUME AUBRUN AND ION NECHITA

Abstract. An important problem in quantum information theory is the mathematical characterization of the phenomenon of quantum catalysis: when can the surrounding entanglement be used to perform transformations of a jointly held quantum state under LOCC (local operations and classical communication) ? Mathematically, the question amounts to describe, for a fixed vector y, the set $T(y)$ of vectors x such that we have $x \otimes z \prec y \otimes z$ for some z, where $\prec$ denotes the standard majorization relation.

Our main result is that the closure of $T(y)$ in the ℓ_1 norm can be fully described by inequalities on the ℓ_p norms: $\|x\|_p \leqslant \|y\|_p$ for all $p \geqslant 1$. This is a first step towards a complete description of $T(y)$ itself. It can also be seen as a ℓ_p-norm analogue of Ky Fan dominance theorem about unitarily invariant norms. The proofs exploits links with another quantum phenomenon: the possibiliy of multiple-copy transformations ($x^{\otimes n} \prec y^{\otimes n}$ for given n). The main new tool is a variant of Cramér's theorem on large deviations for sums of i.i.d. random variables.

1. Introduction

The increasing interest that quantum entanglement has received in the past decade is due, in part, to its use as a *resource* in quantum information processing. We investigate the problem of entanglement transformation: under which conditions can an entangled state $|\phi\rangle$ be transformed into another entangled state $|\psi\rangle$? We restrict ourselves to LOCC protocols: Alice and Bob share $|\phi\rangle$ and have at their disposal only *local operations* (such as unitaries $U_A \otimes I_B$ for Alice) and *classical communication*. Nielsen showed in [13] that such a transformation is possible if and only if $\lambda_\phi \prec \lambda_\psi$, where "$\prec$" is the *majorization* relation and λ_ϕ, λ_ψ are the Schmidt coefficients vectors of $|\phi\rangle$ and $|\psi\rangle$ respectively. Practically in the same time, Jonathan and Plenio [8] discovered a striking phenomenon: entanglement can help LOCC communication, without even being consumed. Precisely, they have found states $|\phi\rangle$ and $|\psi\rangle$ such that $|\phi\rangle$ cannot be transformed into $|\psi\rangle$, but, with the help of a *catalyst* state $|\chi\rangle$, the transformation $|\phi\rangle \otimes |\chi\rangle \to |\psi\rangle \otimes |\chi\rangle$ is possible. When such a catalyst exists, we say that the state $|\phi\rangle$ is *trumped* by $|\psi\rangle$ and we write $\lambda_\phi \prec_T \lambda_\psi$. We say then that $|\phi\rangle$ can be transformed into $|\psi\rangle$ by entanglement-assisted LOCC or ELOCC. It turns out that the trumping relation is much more complicated that the majorization relation; one can easily check on two given states $|\phi\rangle$ and $|\psi\rangle$ whether $\lambda_\phi \prec \lambda_\psi$ is satisfifted or not, but there is no direct way to determine if $\lambda_\phi \prec_T \lambda_\psi$. Later, Bandyopadhyay et al. [1] discovered that a similar situation occurs when trying to transform by LOCC *multiple copies* of $|\phi\rangle$ into $|\psi\rangle$. It may happen that the transformation $|\phi\rangle \to |\psi\rangle$ is not possible, but when considering n copies, one can transform $|\phi\rangle^{\otimes n}$ into $|\psi\rangle^{\otimes n}$. The phenomenon of multiple simultaneous LOCC transformations, or MLOCC, has been intensively studied in the last years and many similarities with ELOCC have been found [6, 7].

In this note, we make some progress towards a complete characterization of both ELOCC and MLOCC. We show that a set of inequalities involving ℓ_p norms (see the remark on Conjecture 1 at the end of the paper) is equivalent to the fact that $|\phi\rangle$ can be approached by a sequence of states $|\phi_n\rangle$ which are MLOCC/ELOCC-dominated by $|\psi\rangle$. An important point is that we allow the dimension of $|\phi_n\rangle$ to exceed the dimension of $|\phi\rangle$. Our proof uses probabilistic tools; we introduce probability measures associated to $|\phi\rangle$ and $|\psi\rangle$ and we use large deviation techniques to show the desired result.

Interestingly, the result can be reversed to give a characterization of ℓ_p norms that is similar to the Ky Fan characterization of unitarily invariant norms. We refer the interested reader to Section 3. The rest of the paper is organized as follows: in Section 2 we introduce the notation and the general framework of entanglement transformation of bipartite states. We also state our main result, Theorem 1. The theorem is proved in Section 4. Conclusions and some directions for further study are sketched in Section 5. The appendix at the end of the paper contains basic results from large deviation theory needed in the proof of the main theorem.

Acknowledgement: we thank the referees for several helpful remarks that improved the presentation of the paper.

2. Notation and statement of the results

For $d \in \mathbf{N}^*$, let P_d be the set of d-dimensional probability vectors : $P_d = \{x \in \mathbf{R}^d \text{ s.t. } x_i \geqslant 0, \sum x_i = 1\}$. If $x \in P_d$, we write $x^\downarrow$ for the decreasing rearrangement of x, i.e. the vector $x^\downarrow \in P_d$ such that x and $x^\downarrow$ have the same coordinates up to permutation, and $x_i^\downarrow \geqslant x_{i+1}^\downarrow$. We shall also write $x_{\max}$ for $x_1^\downarrow$ and $x_{\min}$ for the smallest nonzero coordinate of x.

There is an operation on probability vectors that is fundamental in what follows: the tensor product $\otimes$. If $x = (x_1, \ldots, x_d) \in P_d$ and $x' = (x'_1, \ldots, x'_{d'}) \in P_{d'}$, the tensor product $x \otimes x'$ is the vector $(x_i x'_j)_{ij} \in P_{dd'}$; the way we order the coordinates of $x \otimes x'$ is immaterial for our purposes. We also define the direct sum $x \oplus x'$ as the concatenated vector $(x_1, \ldots, x_d, x'_1, \ldots, x'_{d'}) \in \mathbf{R}^{d+d'}$.

It $x \in P_d$ satisfies $x_d = 0$, it will be useful to identify x with the truncated vector $(x_1, \ldots, x_{d-1}) \in P_{d-1}$. This identification induces a canonical inclusion $P_{d-1} \subset P_d$. Thus, every vector $x \in P_d$ can be thought of as a vector of $P_{d'}$ for all $d' \geqslant d$ by appending $d' - d$ null elements to x. We consider thus the set of all probability vectors $P_{<\infty} = \bigcup_{d>0} P_d$. In other words, $P_{<\infty}$ is the set of finitely supported probability vectors.

Let us now introduce the classical majorization relation [11, 2]. If $x, y \in \mathbf{R}^d$ we define the submajorization relation $\prec_w$ as follows

$$x \prec_w y \quad \text{iff.} \quad \forall k \in \{1, \ldots d\}, \sum_{i=1}^k x_i^\downarrow \leqslant \sum_{i=1}^k y_i^\downarrow,$$

and the majorization relation $\prec$ as

$$x \prec y \quad \text{iff.} \quad \sum_{i=1}^d x_i = \sum_{i=1}^d y_i \quad \text{and} \quad \forall k \in \{1, \ldots d-1\}, \sum_{i=1}^k x_i^\downarrow \leqslant \sum_{i=1}^k y_i^\downarrow.$$

We usually work with probability vectors, for which both relations coincide. However, it will be useful in the proof to work with deficient vectors (of total mass less than 1) and to use submajorization. We write $S_d(y)$ for the set of vectors x in P_d which are majorized by y. It is well-known that $S_d(y)$ is a compact convex set whose extreme points are the vectors obtained by permuting the coordinates of y; many other characterizations of $S_d(y)$ are known [14, 3]. This relation behaves well with respect to direct sums and tensor products: $x \prec y$ implies $x \oplus z \prec y \oplus z$ and $x \otimes z \prec y \otimes z$ for any $z \in P_{<\infty}$. The majorization relation has been shown to have a very important role in quantum information. Nielsen has proved [13] that a state $|\phi\rangle$ belonging to Alice and Bob can be transformed into the state $|\psi\rangle$ by using local operations and classical communication (LOCC) if and only if

$$\lambda_\phi \prec \lambda_\psi,$$

where λ_ϕ (respectively λ_ψ) is the vector of eigenvalues of the density matrix for Alice's system when the joint system is in the state $|\phi\rangle$ (respectively $|\psi\rangle$). Not long after Nielsen's theorem, Jonathan and Plenio have discovered a very intriguing phenomenon: there exist states $|\phi\rangle$ and $|\psi\rangle$ such that the transformation $|\phi\rangle \to |\psi\rangle$ is impossible by LOCC, but, with the aid of a *catalyst state* $|\chi\rangle$, the transformation $|\phi\rangle \otimes |\chi\rangle \to |\psi\rangle \otimes |\chi\rangle$ becomes possible; we say that $|\phi\rangle$ can be transformed into $|\psi\rangle$ by *Entanglement-assisted LOCC* or ELOCC. This result has motivated a more complex relation between probability vectors: if $x, y \in P_d$, we say that y trumps x and write $x \prec_T y$ if there exists $z \in P_{<\infty}$ such that $x \otimes z \prec y \otimes z$. It is important to require that the auxiliary vector z (called the catalyst) is finitely supported (see Remark 2). Given $y \in P_d$, we write $T_d(y)$ for the set of d-dimensional vectors trumped by y, that is

$$T_d(y) = \{x \in P_d \text{ s.t. } x \prec_T y\}.$$

The set $T_d(y)$ is in general larger than $S_d(y)$ [4] and much more complicated to describe. Up to now, there is no known simple procedure to decide whether $x \in T_d(y)$ or not. Hence, finding a tractable characterization of the relation $\prec_T$ (or, equivalently, of the set $T_d(y)$) has become an important open problem in quantum information theory [16]. The geometry of $T_d(y)$ has been studied in [3, 4]: it is a bounded convex set that it is neither closed nor open (provided y is not too simple). We shall introduce now another important extension of LOCC transformations. Bandyopadhyay et al [1] found an example of entangled states $|\phi\rangle$ and $|\psi\rangle$ with the property that the LOCC transformation $|\phi\rangle \to |\psi\rangle$ is impossible but, when one tries to transform multiple copies of the states, the transformation $|\phi\rangle^{\otimes n} \to |\psi\rangle^{\otimes n}$ becomes possible. We say that $|\psi\rangle$ MLOCC-dominates $|\phi\rangle$. We introduce the analogue of the trumping relation for probability vectors:

$$x \prec_M y \quad \text{iff} \quad \exists n \geqslant 1 \, \text{s.t.} \, x^{\otimes n} \prec y^{\otimes n},$$

and the set of probability vectors MLOCC-dominated by a given vector y:

$$M_d(y) = \{x \in P_d \text{ s.t. } x \prec_M y\}.$$

Not much is known about the set $M_d(y)$. It has been studied in [6] and shown to have many similarities with the set $T_d(y)$: for example it is neither closed nor open in general. One important point is that, for all y, we have $M_d(y) \subseteq T_d(y)$ (see [6]).

We report progress towards a description of the sets of $M_d(y)$ and $T_d(y)$. The main ingredient of our approach is the following observation. Consider two vectors $x, y \in P_d$. Whether $x \prec y$, $x \prec_M y$, $x \prec_T y$ or not depends only on the non-zero coordinates of x and y. Thus, it is possible to $\prec/\prec_M/\prec_T$-compare vectors of different sizes by appending the necessary amount of zero coordinates to the end of one of them. Hence, it seems more natural (at least from a mathematical point of view) to consider the sets

$$T_{<\infty}(y) = \{x \in P_{<\infty} \text{ s.t. } x \prec_T y\} = \{x \in P_{<\infty} \text{ s.t. } \exists z \in P_{<\infty} \text{ s.t. } x \otimes z \prec y \otimes z\} = \bigcup_{d' \geqslant d} T_{d'}(y)$$

and

$$M_{<\infty}(y) = \{x \in P_{<\infty} \text{ s.t. } x \prec_M y\} = \{x \in P_{<\infty} \text{ s.t. } \exists n \geqslant 1 \text{ s.t. } x^{\otimes n} \prec y^{\otimes n}\} = \bigcup_{d' \geqslant d} M_{d'}(y).$$

The important point here is that both $T_{<\infty}(y)$ and $M_{<\infty}(y)$ do not depend anymore on the size of y, but only on the non-null coordinates of y. Of course, if $y \in P_d$, $T_d(y) = T_{<\infty}(y) \cap P_d$ and $M_d(y) = M_{<\infty}(y) \cap P_d$; this shows that the sets $T_{<\infty}(y)$ and $M_{<\infty}(y)$ are not closed either in general (otherwise $T_d(y)$ and $M_d(y)$ would be also closed). We then write $\overline{T_{<\infty}(y)}$ and $\overline{M_{<\infty}(y)}$ to denote the closure taken with respect to the ℓ_1-norm, the natural topology in this setting (see Remark 3). Recall that for $p \geqslant 1$, the ℓ_p norm of a vector $x \in P_d$ is defined as

$$\|x\|_p = \left(\sum_{i=1}^{d} x_i^p \right)^{1/p} \tag{1}$$

and $\|x\|_\infty = \max x_i$. We now come to our main result:

Theorem 1. *Consider two vectors $x, y \in P_{<\infty}$. The following assertions are equivalent:*
(a) $x \in \overline{M_{<\infty}(y)}$,
(b) $x \in \overline{T_{<\infty}(y)}$,
(c) $\|x\|_p \leqslant \|y\|_p \ \forall p \geqslant 1$.

Remark 1. *Note that instead of demanding that $\|x\|_p \leqslant \|y\|_p$ for all $p \geqslant 1$, it suffices to ask for $x, y \in P_d$ that the inequality holds for all $p \in [1, p_{max}(x, y)]$, where $p_{max}(x, y) = \log d/(\log y_{max} - \log x_{max})$. The inequalities for $p > p_{max}$ follow by simple computation. For such results in a more general setting, see* [12].

Remark 2. *It is important to see at this point how the set $\overline{T_{<\infty}(y)}$ is related to the set $T_d(y)$. First of all, note that if we drop the closure, we have equality: $T_{<\infty}(y) \cap P_d = T_d(y)$ for $y \in P_d$. However, when taking the ℓ_1 closure of the left hand side, we obtain a strict inclusion: $\overline{T_d(y)} \subsetneq \overline{T_{<\infty}(y)} \cap P_d$. An example for such a vector is provided by the phenomenon of infinite-dimensional catalysis, discovered by Daftuar* [3]. *Take $y = (0.5, 0.25, 0.25)$ and $x = (0.4, 0.4, 0.2)$. It is obvious that $x \notin \overline{T_d(y)}$ because $x_3 < y_3$ and the condition $x_d \geqslant y_d$ is necessary for $x \in \overline{T_d(y)}$. However, there exist an infinite-dimensional catalyst $z = (1-\alpha)(1, \alpha, \alpha^2, \ldots, \alpha^k, \ldots)$, where $\alpha = 2^{-\frac{1}{8}}$, such that $x \otimes z \prec y \otimes z$ and $||x \otimes z||_p \leqslant ||y \otimes z||_p$ for all $p \geqslant 1$. Note that z is ℓ_p-bounded and thus $\|x\|_p \leqslant \|y\|_p$ for all $p \geqslant 1$. By the preceding theorem, we have that $x \in \overline{T_{<\infty}(y)} \cap P_3$. For further remarks on this topic, see Section 5.*

Remark 3. *The use of the ℓ_1 norm is natural in this context from a mathematical point of view since $P_{<\infty}$ is a subset of the norm-closed hyperplane of ℓ_1 defined by $\sum x_i = 1$. Let us explain also how it relates to other physically motivated distances between the approaching states $|\phi_n\rangle$ and the original state $|\phi\rangle$. Recall that x is the eigenvalue vector of the reduced density matrix corresponding to Alice's (or, equivalently to Bob's) part of the system. From the details of the proof (see also Section 5), one sees that the size of the approaching vectors x_n increases with n. So, in order to compare ρ and ρ_n, we have to realize them as density matrices on the same Hilbert space $\mathcal{H}$. Moreover, we can suppose that the two states are diagonalizable in the same basis (Alice can achieve this by applying a local unitary basis change). As usually, we append the necessary number of zero eigenvalues to x in order to have the same size as x_n. We obtain the following equality:*

$$\|x - x_n\|_1 = \|\rho - \rho_n\|_{tr}.$$

So, for Alice's part of the system, we obtain a convergence in the trace norm sense. It is well known that the trace norm distance is related to the probability that the two states can be distinguished by some measurement. Moreover, by using some classical inequalities (see [15]*, Chapter 9), the fidelity $F(\rho, \rho_n)$ can be shown to converge to 1.*

3. A ℓ_p version of Ky Fan theorem

In this section, we explain how Theorem 1 can be seen as an analogue of Ky Fan dominance theorem. We refer to [2] for background. We denote by $\mathscr{M}_d$ the space of complex $d \times d$ matrices. A norm $|||\cdot|||$

on $\mathscr{M}_d$ is said to be unitarily invariant if $|||UAV||| = |||A|||$ for all unitary matrices U, V. A norm $||\cdot||$ on $\mathbf{R}^d$ is said to be symmetric if

$$||(x_1, \ldots, x_d)|| = ||(\pm x_{\sigma(1)}, \ldots, \pm x_{\sigma(d)})||$$

for all choices of signs in $\{\pm 1\}^d$ and all permutations $\sigma \in \mathfrak{S}_d$. It is well-known ([2], Theorem IV.2.1) that unitarily invariant norms on $\mathscr{M}_d$ are in 1-to-1 correspondance with symmetric norms on $\mathbf{R}^d$ (consider the restriction of $|||\cdot|||$ to diagonal matrices).

Examples of unitarily invariant norms are given by Ky Fan norms, defined for $k = 1, 2, \ldots, d$ by

$$|||A|||_{(k)} = \sum_{j=1}^{k} s_j(A),$$

where $s_1(A) \geqslant \cdots \geqslant s_d(A)$ denote the ordered singular values of a matrix A. The Ky Fan dominance theorem asserts that these norms are extremal among unitarily invariant norms in the following sense: if A, B satisfy $|||A|||_{(k)} \leqslant |||B|||_{(k)}$ for any $k = 1, \cdots, d$, then $|||A||| \leqslant |||B|||$ for any unitarily invariant norm ; this condition can also be formulated as $s(A) \prec_w s(B)$, where $s(\cdot)$ denotes the vector of singular values of a matrix.

This gives a way to derive an infinite family of inequalities from a finite one. However this may be a too strong requirement and one can wonder what happens for an important special class of unitarily invariant norms: the Schatten p-norms (or noncommutative ℓ_p norms), defined for $p \geqslant 1$ by

$$|||A|||_p = \left(\sum_{j=1}^{d} s_j(A)^p \right)^{1/p}.$$

To state our result, we need to compare matrices of different sizes. If $d < d'$ we identify $\mathscr{M}_d$ with the top-left corner of $\mathscr{M}_{d'}$; this gives a natural inclusion $\mathscr{M}_d \subset \mathscr{M}_{d'}$ and we write $\mathscr{M}_{<\infty} = \bigcup_d \mathscr{M}_d$. Note that the tensor product of matrices is a well-defined operation on $\mathscr{M}_{<\infty}$.

Theorem 2. *Let $A, B \in \mathscr{M}_d$. The following are equivalent*

(1) $|||A|||_p \leqslant |||B|||_p$ *for all* $p \geqslant 1$.

(2) *There exists in $\mathscr{M}_{<\infty}$ a sequence (A_n) so that $\lim_{n\to\infty} |||A_n - A|||_1 = 0$ and $|||A_n^{\otimes n}||| \leqslant |||B_n^{\otimes n}|||$ for all unitarily invariant norms $|||.|||$ (or, equivalently, so that $s(A_n^{\otimes n}) \prec_w s(B^{\otimes n})$).*

Of course, a main difference between this result and Ky Fan dominance theorem is that condition (ii) here is hard to check and involves infinitely many inequalities.

Proof (sketch). Because of the bijective correspondance between unitarily invariant norms on matrices and symmetric norms on vectors, it is enough to prove the theorem for positive diagonal matrices. This is almost the content of the equivalence (a) $\iff$ (c) of Theorem 1. The only slight remark that we need in order to get condition (2) as stated here is the following: in Lemma 2 below, it follows from the proof that we can actually choose the integer n so that $x^{\otimes N} \prec_w y^{\otimes N}$ for any $N \geqslant n$. $\square$

4. The proof of the theorem

We shall prove the sequence of implications (a) $\Rightarrow$ (b) $\Rightarrow$ (c) $\Rightarrow$ (a). The first two are well known; we sketch their proof for completeness. The third is the most difficult one and represents our contribution to the theorem.

(a) $\Rightarrow$ (b) Because the closure is taken with respect to the same topology (ℓ_1) for both $\overline{M_{<\infty}(y)}$ and $\overline{T_{<\infty}(y)}$, it is enough to show $M_{<\infty}(y) \subset T_{<\infty}(y)$. Let $x \in M_{<\infty}(y)$ and consider n such that

$x^{\otimes n} \prec y^{\otimes n}$. The trick here (see [6]) is to use the following z as a catalyst

$$z = x^{\otimes(n-1)} \oplus x^{\otimes(n-2)} \otimes y \oplus \cdots \oplus x \otimes y^{\otimes(n-2)} \oplus y^{\otimes(n-1)}.$$

For simplicity we do not normalize z, but this is irrelevant. The vector z has been constructed such that

$$x \otimes z = x^{\otimes n} \oplus w \text{ and } y \otimes z = y^{\otimes n} \oplus w,$$

where w is the same in both expressions. This implies that $x \otimes z \prec y \otimes z$, i.e. $x \in T_{<\infty}(y)$.

(b) $\Rightarrow$ **(c)** Let $z \in P_{<\infty}$ be the catalyst for $x \prec_T y$: $x \otimes z \prec y \otimes z$. A function $\varphi : \mathbf{R}^d \to \mathbf{R}$ is said to be Schur-convex if $a \prec b$ implies $\varphi(a) \leqslant \varphi(b)$. It is well-known (see [11, 14]) that if $h : \mathbf{R} \to \mathbf{R}$ is a convex function, then $\varphi : x \mapsto \sum_{i=1}^d h(x_i)$ is Schur-convex. Consequently, the functions $x \mapsto \|x\|_p^p$ are Schur-convex for $p \geqslant 1$. Moreover, they satisfy the identity $\|x \otimes z\|_p = \|x\|_p \|z\|_p$, and similarly for y. Since $\|z\|_p$ is finite, we get that $\|x\|_p \leqslant \|y\|_p$. To show that the same is true for $x \in \overline{T_{<\infty}(y)}$, it suffices to check that the set of $x \in \ell_1$ such that $\|x\|_p \leqslant \|y\|_p$ is norm-closed; this follows from the inequality $\| \cdot \|_p \leqslant \| \cdot \|_1$.

(c) $\Rightarrow$ **(a)** We will adapt some techniques used by G. Kuperberg in a slightly different context [10]. In our proof, we allow deficient vectors, i.e. vectors with total mass smaller than 1, and we use submajorization.

As in [10], we associate to a positive vector $x \in \mathbf{R}^d$ the measure $\mu_x = \sum_{i=1}^d x_i \delta_{\log x_i}$, where δ_z is the Dirac measure at point z. The basic property is that the tensor product operation of vectors corresponds to the convolution of associated measures:

$$\mu_{x \otimes y} = \mu_x * \mu_y.$$

The convolution of two measures μ and ν is defined by the relation

$$\mu * \nu(A) = (\mu \times \nu)\left(\{(x, y) \in \mathbf{R}^2 : x + y \in A\}\right).$$

Moreover, if μ and ν are probability measures and X_μ and X_ν denote independent random variables with laws respectively μ and ν, then $\mu * \nu$ is the law of $X_\mu + X_\nu$.

The following lemma gives a way to prove majorization using comparison of the tails of the associated measures

Lemma 1. *Let x and y be two vectors of $\mathbf{R}^d$ with non-negative components. Consider the measures μ_x and μ_y associated with x and y. Assume that, for all $t \in \mathbf{R}$, $\mu_x[t, \infty) \leqslant \mu_y[t, \infty)$. Then $x \prec_w y$.*

Proof. Note that

$$\mu_x[t, \infty) = \sum_{i : \log x_i \geqslant t} x_i = \sum_{i : x_i \geqslant \exp(t)} x_i.$$

Thus, for all $u > 0$, $\sum_{i : x_i \geqslant u} x_i \leqslant \sum_{i : y_i \geqslant u} y_i$. For simplicity, we assume first that all coordinates of y are distinct. We will show by induction on $k \in \{1, \ldots, d\}$ that $\sum_{i=1}^k x_i^\downarrow \leqslant \sum_{i=1}^k y_i^\downarrow$. For the first step; use $u = y_1^\downarrow$ to conclude that $x_1^\downarrow \leqslant y_1^\downarrow$. Now, fix $k \in \{1, \ldots, d-1\}$ and suppose that $\sum_{i=1}^k x_i^\downarrow \leqslant \sum_{i=1}^k y_i^\downarrow$. If $x_{k+1}^\downarrow \leqslant y_{k+1}^\downarrow$, the induction step is obvious. If $x_{k+1}^\downarrow > y_{k+1}^\downarrow$, we use $u = x_{k+1}^\downarrow$ to get

$$\sum_{i=1}^{k+1} x_i^\downarrow \leqslant \sum_{i : x_i \geqslant x_{k+1}^\downarrow} x_i \leqslant \sum_{i : y_i \geqslant x_{k+1}^\downarrow} y_i \leqslant \sum_{i : y_i \geqslant y_{k+1}^\downarrow} y_i = \sum_{i=1}^{k+1} y_i^\downarrow.$$

This completes the induction when y has distinct coordinates. The general case follows by approximating y by $y + \varepsilon_n$, where (ε_n) is a suitable sequence of positive vectors tending to 0. The approximation is possible since the set of vectors y majorizing a fixed x is closed. $\square$

We now get to the key lemma in our argument. We shall use a slightly modified version of Cramér large deviations theorem — see Appendix.

Lemma 2. *Let x, y in $\mathbf{R}^d$, with nonnegative coordinates. Assume that for any $1 \leqslant p \leqslant \infty$, we have the* **strict** *inequality $\|x\|_p < \|y\|_p$. Then there exists an integer n such that $x^{\otimes n} \prec_w y^{\otimes n}$.*

Proof. Consider x and y satisfying the hypotheses of the lemma. We can assume by multiplying both vectors by a positive constant K that $\|y\|_1 = 1$. Let $p = 1 - \|x\|_1 > 0$. We introduce the measures μ_x and μ_y associated to x and y; μ_y is a probability measure but μ_x is not, so we add a mass at $-\infty$ by setting $\overline{\mu}_x = \mu_x + p\delta_{-\infty}$. Let X and Y be random variables distributed according to $\overline{\mu}_x$ and μ_y respectively. We denote by (X_n) (resp. (Y_n)) a sequence of i.i.d. copies of X (resp. Y). We are going to show that for n large enough

$$\forall t \in \mathbf{R},\ \mathbf{P}(X_1 + \cdots + X_n \geqslant nt) \leqslant \mathbf{P}(Y_1 + \cdots + Y_n \geqslant nt). \tag{2}$$

This is equivalent to showing that

$$\int_{nt}^{\infty} d\mu_x^{*n} = \int_{nt}^{\infty} d\overline{\mu}_x^{*n} \leqslant \int_{nt}^{\infty} d\mu_y^{*n},$$

which, by the previous lemma implies $x^{\otimes n} \prec_w y^{\otimes n}$. Note that the asymptotic behavior of the quantities appearing in (2) is governed by Cramér's theorem. Let $f_n(t) = \mathbf{P}(X_1 + \cdots + X_n \geqslant nt)^{1/n}$ and $g_n(t) = \mathbf{P}(Y_1 + \cdots + Y_n \geqslant nt)^{1/n}$. Applying Cramér's theorem (see Appendix), we obtain

$$f(t) := \lim_{n\to\infty} f_n(t) = \begin{cases} 1 - p \text{ if } t \leqslant \mathbf{E}(X | X \neq -\infty) \\ e^{-\Lambda_X^*(t)} \text{ otherwise.} \end{cases}$$

$$g(t) := \lim_{n\to\infty} g_n(t) = \begin{cases} 1 \text{ if } t \leqslant \mathbf{E}(Y) \\ e^{-\Lambda_Y^*(t)} \text{ otherwise.} \end{cases}$$

Note also that the log-Laplace of X, defined for $\lambda \in \mathbf{R}$ by $\Lambda_X(\lambda) = \log \mathbf{E} e^{\lambda X}$, is related to the ℓ_p norms of x:

$$\forall \lambda \geqslant 0, \quad \Lambda_X(\lambda) = \log \|x\|_{\lambda+1}^{\lambda+1}.$$

The same holds for Y: $\Lambda_Y(\lambda) = \log \|y\|_{\lambda+1}^{\lambda+1}$ and thus we have $\Lambda_X(\lambda) < \Lambda_Y(\lambda)$ for $\lambda \geqslant 0$.

Let $M_X = \operatorname{esssup} X = \log \|x\|_\infty$ and $M_Y = \operatorname{esssup} Y = \log \|y\|_\infty$; by hypothesis $M_X < M_Y$. First of all, note that $f_n(t) = 0$ for $t \geqslant M_X$, so it suffices to show that $f_n \leqslant g_n$ on $(-\infty, M_X]$, for n large enough. We claim that $f < g$ on $(-\infty, M_Y)$, and thus on $(-\infty, M_X]$. Indeed, for $\mathbf{E}(Y) \leqslant t < M_Y$, the supremum in the definition of $\Lambda_Y^*(t)$ is attained at a point $\lambda_0 \geqslant 0$ (cf Appendix), so we have that

$$f(t) \leqslant e^{-(\lambda_0 t - \Lambda_X(\lambda_0))} < e^{-(\lambda_0 t - \Lambda_Y(\lambda_0))} = g(t),$$

where the **strict** inequality follows from the fact that $\Lambda_X(\lambda) < \Lambda_Y(\lambda)$, for all $\lambda \geqslant 0$. For $t < \mathbf{E}(Y)$, $g(t) = 1$ and $f(t) \leqslant 1 - p < 1$. Moreover, the functions f and g admit finite limits in $-\infty$: $\lim_{t\to-\infty} f(t) = 1 - p$ and $\lim_{t\to-\infty} g(t) = 1$. Thus, on the compact set $[-\infty, M_X]$, the functions f and g are well-defined, non-increasing, continuous and satisfy $f < g$.

We now use the following elementary fact: if a sequence of non-increasing functions defined on a compact interval I converges pointwise towards a continuous limit, then the convergence is actually uniform on I (for a proof see [17] Part 2, Problem 127; this statement is attributed to Pólya or to Dini depending on authors). We apply this result to (f_n) and (g_n) on the interval $I = [-\infty, M_X]$ to conclude that the convergence is uniform for both sequences. As $f < g$, we can therefore find n large enough such that $f_n \leqslant g_n$ on I, and thus on $\mathbf{R}$. This is equivalent to (2) and completes the proof of the lemma. $\square$

Remark 4. *It is possible to avoid the use of Cramér's theorem by using low-technology estimates on large deviations probability instead, as done in* [10]. *This requires additional care to get the required uniform bounds and slightly obfuscates the argument. The only advantage is to give explicit bounds for the value of n in Lemma 2, which our compactness argument does not. These bounds are quite bad anyway, and for example do not allow to replace the ℓ_1-closure in the main theorem by a ℓ_p-closure for some $p < 1$.*

Proof of (c) $\Rightarrow$ (a) (continued) Recall that x and y are such that $\|x\|_p \leqslant \|y\|_p$ for any $p \geqslant 1$ and that we want to find, for any $\varepsilon > 0$ small enough, a vector $x_\varepsilon \in M_{<\infty}(y)$ such that $\|x - x_\varepsilon\|_1 \leqslant \varepsilon$. Let d_x (resp. d_y) be the number of nonzero coordinates of x (resp. y). We proceed as follows : let $0 < \varepsilon < 2d_x x_{min}$ and consider the (deficient) vector x'_ε obtained from x by subtracting $\varepsilon/2d_x$ to each of its nonzero coordinates. This implies that x'_ε is a positive vector, $\|x - x'_\varepsilon\|_1 = \varepsilon/2$ and that x'_ε satisfies the hypotheses of Lemma 2. Applying the lemma, we obtain the existence of an integer n such that $(x'_\varepsilon)^{\otimes n} \prec_w y^{\otimes n}$.

Remember that x'_ε is deficient; we now enlarge it into a vector $x_\varepsilon \in P_{<\infty}$ by adding mass $\varepsilon/2$. But since we want to keep the property $x_\varepsilon^{\otimes n} \prec_w y^{\otimes n}$ (which is identical to $x_\varepsilon^{\otimes n} \prec_w y^{\otimes n}$), a safe way to do this is to add a large number of coordinates, each of them being very small. More precisely, let $x_\varepsilon = x'_\varepsilon \oplus \delta^{\oplus D}$, where $\delta D = \varepsilon/2$ and δ is a positive number such that $\delta (x'_\varepsilon)^{n-1}_{\max} \leqslant \min((x'_\varepsilon)^n_{\min}, y^n_{\min})$. We claim that $x_\varepsilon^{\otimes n} \prec y^{\otimes n}$, that is, for any $k \geqslant 1$,

$$\sum_{i=1}^{k} (x_\varepsilon^{\otimes n})^{\downarrow}_i \leqslant \sum_{i=1}^{k} (y^{\otimes n})^{\downarrow}_i . \tag{3}$$

Indeed, δ has been chosen so that the d_x^n largest coordinates of $x_\varepsilon^{\otimes n}$ are exactly the coordinates of $(x'_\varepsilon)^{\otimes n}$, so when $1 \leqslant k \leqslant d_x^n$, (3) follows from the relation $(x'_\varepsilon)^{\otimes n} \prec_w y^{\otimes n}$. If $d_x < k \leqslant d_y^n$, the inequality also holds since the choice of δ guarantees $(x_\varepsilon^{\otimes n})^{\downarrow}_k \leqslant (y^{\otimes n})^{\downarrow}_k$. Finally if $k \geqslant d_y^n$, (3) holds trivially since the right-hand side equals 1.

In conclusion, $x_\varepsilon^{\otimes n} \prec y^{\otimes n}$, and thus $x_\varepsilon \in M_{<\infty}(y)$. But x_ε has been constructed such that $\|x - x_\varepsilon\|_1 \leqslant \varepsilon$ and thus $x \in \overline{M_{<\infty}(y)}$ which completes the proof of the theorem.

5. Conclusion and further remarks

In conclusion, we are able to give a nice description of the ℓ_1-closure of the set $T_{<\infty}(y)$. However, this closure may be substantially larger than the usual closure $\overline{T_d(y)}$ in P_d, and requires approximation by vectors with growing support. Our result can be seen as a contribution to a conjecture attributed to Nielsen [3]:

Conjecture 1. *Fix a vector $y \in P_d$. Then a vector $x \in P_d$ belongs to $\overline{T_d(y)}$ if and only if the following conditions are verified.*

(1) *For $p \geqslant 1$, $\|x\|_p \leqslant \|y\|_p$.*
(2) *For $0 < p \leqslant 1$, $\|x\|_p \geqslant \|y\|_p$.*
(3) *For $p < 0$, $\|x\|_p \geqslant \|y\|_p$.*

M. Klimesh announced a proof of this conjecture in a short communication [9], but the solution has not appeared in print yet. However, his methods are different from our approach (private communication). Note that the definition of $\|\cdot\|_p$ given in (1) is extended to any $p \in \mathbf{R}^*$. For $p < 1$, $\|\cdot\|_p$ is not a norm in the usual sense. We have shown that the condition (1) above is equivalent to $x \in \overline{T_{<\infty}(y)}$. Notice however that $\overline{T_{<\infty}(y)}$ is in general larger than $\overline{T_d(y)}$; note also that the set of $x \in P_d$ that satisfy conditions (1–3) is closed. The "only if" part of the conjecture follows from standard convexity/concavity properties of functionals $\|\cdot\|_p$, see [14, 3].

This question also appears in [6] where it is formulated using the Rényi entropies. For any real $p \neq 1$, the p-Rényi entropy is defined for $x \in P_d$ as

$$H_p(x) = \frac{\mathrm{sgn}(p)}{p-1} \log_2 \left(\sum_{i=1}^{d} x_i^p \right).$$

The limit case $p = 1$ corresponds to the usual entropy. The conditions (1–3) of the conjecture can be concisely reformulated as "$H_p(x) \leqslant H_p(y)$ for all p".

An intermediate notion is the following: for $y \in P_d$, let $\overline{T_{<\infty}(y)}^b$ be the set of vectors $x \in P_d$ such that there is a sequence (x_n) in $T_{<\infty}(y)$ tending to x, with a **uniform** bound on the size of the support of x_n. We think that a description of $\overline{T_{<\infty}(y)}^b$ could be related to the set of vectors which satisfy conditions (1) and (2) — but not necessarily (3) — in Conjecture 1.

There is one more consequence of our main theorem we would like to discuss. Recall that when defining catalysis, we insisted on the fact that the catalyst should be finitely-supported. Let $P_\infty \subset \ell_1$ be the set of infinite-dimensional probability vectors, and for y in $P_{<\infty}$, define the set $T'(y)$ of (finitely supported) vectors trumped by y using infinite catalysts:

$$T'(y) = \{x \in P_{<\infty} \text{ s.t. } \exists z \in P_\infty \text{ s.t. } x \otimes z \prec y \otimes z\}.$$

As shown in [3] (Section 4.3), in general $T_{<\infty}(y) \neq T'(y)$. However, since $x \in T'(y)$ implies $\|x\|_p \leqslant \|y\|_p$ for all $p \geqslant 1$, it follows from our main theorem that $\overline{T_{<\infty}(y)} = \overline{T'(y)}$.

6. Appendix: On Cramér's theorem

We review here some facts from large deviations theory. A complete reference for all the material contained here is [5]. Let X be a random variable taking values in $[-\infty, \infty)$. We allow X to equal $-\infty$ with positive probability; this is a nonstandard hypothesis. We however exclude the trivial case $\mathbf{P}(X = -\infty) = 1$. We write $\mathbf{E}$ for the expectation. We assume also that the conditional expectation $\mathbf{E}(X | X \neq -\infty)$ is finite. The cumulant generating function Λ_X of the random variable X is defined for any $\lambda \in \mathbf{R}$ by

$$\Lambda_X(\lambda) = \log \mathbf{E} e^{\lambda X}.$$

It is a convex function taking values in $(-\infty, +\infty]$. Its convex conjugate Λ_X^*, sometimes called the Cramér transform, is defined as

$$\Lambda_X^*(x) = \sup_{\lambda \in \mathbf{R}} \lambda x - \Lambda_X(\lambda). \tag{4}$$

Note that Λ_X is a smooth and strictly convex function on $[0, +\infty]$. Moreover, $\Lambda_X'(0) = \mathbf{E}(X | X \neq -\infty)$ and $\lim_{\lambda \to +\infty} \Lambda_X'(\lambda) = \mathrm{esssup}(X)$. Consequently, for any x such that $\mathbf{E}(X | X \neq -\infty) < x < \mathrm{esssup}(X)$, the supremum in (4) is attained at a unique point $\lambda \geqslant 0$. We now state Cramér's theorem in a suitable formulation

Proposition 1. *Let X be a $[-\infty, +\infty)$-valued random variable such that $\Lambda_X(\lambda) < +\infty$ for any $\lambda \geqslant 0$. Let (X_i) be a sequence of i.i.d. copies of X. Then for any $t \in \mathbf{R}$*

$$\lim_{n \to \infty} \frac{1}{n} \log \mathbf{P}(X_1 + \cdots + X_n \geqslant tn) = \begin{cases} \log \mathbf{P}(X \neq -\infty) \text{ if } t \leqslant \mathbf{E}(X | X \neq -\infty) \\ -\Lambda_X^*(t) \text{ otherwise.} \end{cases}$$

Proof. Let $\hat{X}$ denote the random variable X conditioned to be finite, that is for any Borel set $B \subset \mathbf{R}$

$$\mathbf{P}(\hat{X} \in B) = \frac{1}{1-p} \mathbf{P}(X \in B),$$

where $p = \mathbf{P}(X = -\infty)$. A consequence of the classical Cramér theorem ([5], Corollary 2.2.19) states that

$$\forall t \in \mathbf{R}, \quad \lim_{n\to\infty} \frac{1}{n} \log \mathbf{P}(\hat{X}_1 + \cdots + \hat{X}_n \geqslant tn) = -\inf_{s \geqslant t} \Lambda^*_{\hat{X}}(s). \tag{5}$$

One checks that $\Lambda_{\hat{X}} = \Lambda_X - \log(1-p)$, and consequently

$$\Lambda^*_{\hat{X}} = \Lambda^*_X + \log(1-p). \tag{6}$$

Note also that

$$\mathbf{P}(X_1 + \cdots + X_n \geqslant tn) = (1-p)^n \mathbf{P}(\hat{X}_1 + \cdots + \hat{X}_n \geqslant tn). \tag{7}$$

Finally, note that the infimum on the right hand side of (5) is null for $t \leqslant \mathbf{E}(\hat{X})$ and equals $\Lambda^*_{\hat{X}}(t)$ for $t > \mathbf{E}(\hat{X})$. This follows from the fact that the convex function $t \mapsto \Lambda^*_{\hat{X}}(t)$ attains its zero minimum at $t = \mathbf{E}(\hat{X})$ and is increasing for $t \geqslant \mathbf{E}(\hat{X})$. Thus, we can rewrite equation (5) as:

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbf{P}(\hat{X}_1 + \cdots + \hat{X}_n \geqslant tn) = \begin{cases} 0 \text{ if } t \leqslant \mathbf{E}(\hat{X}) \\ -\Lambda^*_{\hat{X}}(t) \text{ otherwise.} \end{cases} \tag{8}$$

The proposition follows from the equations (6), (7) and (8). □

References

[1] S. Bandyopadhyay, V. Roychowdhury and U. Sen *Classification of nonasymptotic bipartite pure-state entanglement transformations*, Phys. Rev. A, 65 (2002), 052315.

[2] R. Bhatia, *Matrix Analysis.* Graduate Texts in Mathematics, 169. Springer-Verlag, New York, 1997.

[3] S. K. Daftuar, *Eigenvalues Inequalities in Quantum Information Processing*, Ph. D. Thesis, California Institute of technology, 2004. Available at `http://resolver.caltech.edu/CaltechETD:etd-03312004-100014`

[4] S. K. Daftuar and M. Klimesh, *Mathematical structure of entanglement catalysis*, Phys. Rev. A (3) **64** (2001), no. 4, 042314

[5] A. Dembo and O. Zeitouni, *Large deviations Techniques and Applications*, Second edition. Applications of Mathematics (New York), 38. Springer-Verlag, New York, 1998.

[6] R. Duan, Y. Feng, X. Li and M. Ying, *Multiple-copy entanglement transformation and entanglement catalysis*, Phys. Rev. A **71** (2005), 042319

[7] R. Duan, Z. Ji, Y. Feng, X. Li and M. Ying, *Some issues in quantum information theory*, J. Comput. Sci. & Technol. **21** (2006), no. 5, 776–789.

[8] D. Jonathan and M. B. Plenio, *Entanglement-assisted local manipulation of pure quantum states*, Phys. Rev. Lett. **83** (1999), no. 17, 3566–3569.

[9] M. Klimesh, *Entropy measures and catalysis of bipartite quantum state transformations*, extended abstract, ISIT 2004, Chicago, USA.

[10] G. Kuperberg, *The capacity of hybrid quantum memory*, IEEE Trans. Inform. Theory **49** (2003), 1465–1473.

[11] A. Marshall and I. Olkin, *Inequalities: theory of majorization and its applications*, Mathematics in Science and Engineering, 143. Academic Press Inc., New York-London, 1979

[12] T. Mitra and E. Ok, *Majorization by L^p-Norms*, preprint, available at `http://homepages.nyu.edu/~eo1/Papers-PDF/Major.pdf`

[13] M. Nielsen, *Conditions for a class of entanglement transformations*, Phys. Rev. Lett. **83**, 436 (1999).

[14] M. Nielsen, *An introduction to majorization and its applications to quantum mechanics*, preprint, available at `www.qinfo.org/talks/2002/maj/book.ps`

[15] M. Nielsen and I. Chuang, *Quantum computation and quantum information.* Cambridge University Press, Cambridge, 2000.

[16] Open problems in Quantum Information Theory, available at `http://www.imaph.tu-bs.de/qi/problems/` or `quant-ph/0504166`

[17] G. Pólya and G. Szegö, *Problems and Theorems in Analysis*, Springer-Verlag, Berlin-New York, 1978.

Address :
Université de Lyon,
Université Lyon 1,
CNRS, UMR 5208 Institut Camille Jordan,
Batiment du Doyen Jean Braconnier,
43, boulevard du 11 novembre 1918,
F - 69622 Villeurbanne Cedex,
France

Email: aubrun@math.univ-lyon1.fr, nechita@math.univ-lyon1.fr