



**HAL**  
open science

# Stochastic domination for iterated convolutions and catalytic majorization

Guillaume Aubrun, Ion Nechita

► **To cite this version:**

Guillaume Aubrun, Ion Nechita. Stochastic domination for iterated convolutions and catalytic majorization. 2007. hal-00159197v1

**HAL Id: hal-00159197**

**<https://hal.science/hal-00159197v1>**

Preprint submitted on 2 Jul 2007 (v1), last revised 11 Apr 2008 (v2)

**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.

# STOCHASTIC DOMINATION FOR ITERATED CONVOLUTIONS AND CATALYTIC MAJORIZATION

GUILLAUME AUBRUN AND ION NECHITA

ABSTRACT. We study how iterated convolutions of probability measures compare under stochastic domination. We give necessary and sufficient conditions for the existence of an integer  $n$  such that  $\mu^{*n}$  is stochastically dominated by  $\nu^{*n}$  for two given probability measures  $\mu$  and  $\nu$ . As a consequence we obtain a similar theorem on the majorization order for vectors in  $\mathbf{R}^d$ . In particular we prove results about catalysis in quantum information theory.

## NOTATIONS

Let us start by introducing some notation and recalling basic facts about probability measures. We write  $\mathbf{P}(\mathbf{R})$  for the set of probability measures on  $\mathbf{R}$ . We denote by  $\delta_x$  the Dirac mass at point  $x$ . If  $\mu \in \mathbf{P}(\mathbf{R})$ , we write  $\text{supp } \mu$  for the support of  $\mu$ . We write respectively  $\min \mu$  and  $\max \mu$  for  $\min \text{supp } \mu$  and  $\max \text{supp } \mu$ . We also write  $\mu(a, b)$  and  $\mu[a, b]$  as a shortcut for  $\mu((a, b))$  and  $\mu([a, b])$ . The convolution of two measures  $\mu$  and  $\nu$  is denoted  $\mu * \nu$ . Recall that if  $X$  and  $Y$  are independent random variables of respective laws  $\mu$  and  $\nu$ , the law of  $X + Y$  is given by  $\mu * \nu$ . The results of this paper are stated for convolutions of measures, they admit immediate translations in the language of sums of independent random variables. For  $\lambda \in \mathbf{R}$ , the function  $e_\lambda$  is defined by  $e_\lambda(x) = \exp(\lambda x)$ .

## 1. STOCHASTIC DOMINATION

A natural way of comparing two probability measures is given by the following relation

**Definition 1.** *Let  $\mu$  and  $\nu$  be two probability measures on the real line. We say that  $\mu$  is stochastically dominated by  $\nu$  and we write  $\mu \leq_{\text{st}} \nu$  if*

$$(1) \quad \forall t \in \mathbf{R}, \mu[t, \infty) \leq \nu[t, \infty).$$

It is immediately checked that (1) is equivalent to

$$\forall t \in \mathbf{R}, \mu(-\infty, t] \leq \nu(-\infty, t].$$

Stochastic domination is an order relation (in particular,  $\mu \leq_{\text{st}} \nu$  and  $\nu \leq_{\text{st}} \mu$  imply  $\mu = \nu$ ). The following result [9] provides a useful characterization of stochastic domination

**Theorem (Strassen).** *Let  $\mu$  and  $\nu$  be two probability measures on the real line. Then  $\mu \leq_{\text{st}} \nu$  if and only if there is a probability space  $(\Omega, \mathcal{F}, \mathbf{P})$  with two random variables  $X$  and  $Y$  such that*

- $X$  has law  $\mu$ ,
- $Y$  has law  $\nu$ ,
- $X(\omega) \leq Y(\omega)$ , for all  $\omega \in \Omega$ .

The following lemmas can be proved either using Strassen's theorem or by direct calculation

---

Research was supported in part by the European Network Phenomena in High Dimensions, FP6 Marie Curie Actions, MCRN-511953.

**Lemma 1.** *Let  $\mu, \nu$  be two probability measures on the real line such that  $\mu \leq_{\text{st}} \nu$ . Then for any increasing function  $f : \mathbf{R} \rightarrow \mathbf{R}^+$  and any decreasing function  $g : \mathbf{R} \rightarrow \mathbf{R}^+$ , we have*

$$\int f d\mu \leq \int f d\nu \text{ and } \int g d\mu \geq \int g d\nu.$$

**Lemma 2.** *Let  $\mu_1, \mu_2, \nu_1, \nu_2$  be probability measures on the real line. If  $\mu_1 \leq_{\text{st}} \nu_1$  and  $\mu_2 \leq_{\text{st}} \nu_2$ , then  $\mu_1 * \mu_2 \leq_{\text{st}} \nu_1 * \nu_2$ .*

**Corollary 1.** *Let  $\mu$  and  $\nu$  be two probability measures on the real line such that  $\mu \leq_{\text{st}} \nu$ . Then, for all  $n \geq 2$ ,  $\mu^{*n} \leq_{\text{st}} \nu^{*n}$ .*

The converse to the corollary is false, as shown by the following proposition

**Proposition 1.** *For all  $n \in 2\mathbf{N}^*$ , there exist probability measures  $\mu$  and  $\nu$  such that  $\mu^{*(n+1)} \leq_{\text{st}} \nu^{*(n+1)}$  and, for all  $1 \leq k \leq n$ ,  $\mu^{*k} \not\leq_{\text{st}} \nu^{*k}$ . Moreover,  $\mu$  and  $\nu$  can be chosen to be a convex combination of two Dirac masses.*

*Proof.* Let  $\mu = (1-p)\delta_0 + p\delta_a$  and  $\nu = \frac{1}{2}\delta_{a-(1+\varepsilon)} + \frac{1}{2}\delta_{a+1}$ , where  $a \in \mathbf{N}^*$ ,  $p \in (0, 1)$  and  $\varepsilon > 0$  will be defined later. For  $k \geq 1$ ,

$$\mu^{*k} = \sum_{i=0}^k \binom{k}{i} p^i (1-p)^{k-i} \delta_{ia},$$

and

$$\nu^{*k} = \frac{1}{2^k} \sum_{i=0}^k \binom{k}{i} \delta_{ka+k-2i-i\varepsilon}.$$

Hence,  $\text{supp}(\mu^{*k}) \subset [0, ka]$  and  $\text{supp}(\nu^{*k}) \subset [ka - k - k\varepsilon, ka + k]$ . Choose  $\varepsilon > 0$  small enough such that  $k - 2i - i\varepsilon \geq 0 \Leftrightarrow i < \frac{k}{2}$  for all  $1 \leq k \leq n+1$ ,  $0 \leq i \leq k$ ; also, choose  $a > (n+1)(1+\varepsilon)$ . Then, for  $1 \leq k \leq n+1$ , the only atom of  $\mu^{*k}$  inside  $[\min \nu^{*k}, \max \nu^{*k}]$  is  $\delta_{ka}$ . It can be easily showed that, for all  $1 \leq k \leq n+1$ ,  $\mu^{*k} \leq_{\text{st}} \nu^{*k}$  iff.  $\mu^{*k}[ka, \infty) \leq \nu^{*k}[ka, \infty)$ . We have:

$$\mu^{*k}[ka, \infty) = p^k,$$

$$\nu^{*k}[ka, \infty) = \frac{1}{2^k} \sum_{0 \leq i < k/2} \binom{k}{i} = \begin{cases} \frac{1}{2} & \text{if } k \text{ is odd,} \\ \frac{1}{2} - \frac{1}{2^{k+1}} \binom{k}{k/2} & \text{if } k \text{ is even.} \end{cases}$$

Consider now  $p \in (0, 1)$  such that  $p^n > 1/2$  and  $p^{n+1} \leq 1/2$ . For all  $1 \leq k \leq n$ , we have

$$\nu^{*k}[ka, \infty) \leq \frac{1}{2} < p^k = \mu^{*k}[ka, \infty),$$

and thus  $\mu^{*k} \not\leq_{\text{st}} \nu^{*k}$ . However, if  $n$  is even,  $\mu^{*(n+1)}[(n+1)a, \infty) = p^{n+1} \leq 1/2 = \nu^{*(n+1)}[(n+1)a, \infty)$  and so  $\mu^{*(n+1)} \leq_{\text{st}} \nu^{*(n+1)}$ .  $\square$

## 2. STOCHASTIC DOMINATION FOR ITERATED CONVOLUTIONS AND CRAMÉR'S THEOREM

In light of Proposition 1, we are going to study the following relation which generalizes stochastic domination

**Definition 2.** *We define a relation  $\leq_{\text{st}}^*$  on  $\mathbf{P}(\mathbf{R})$  as follows*

$$\mu \leq_{\text{st}}^* \nu \iff \exists n \geq 1 \text{ s.t. } \mu^{*n} \leq_{\text{st}} \nu^{*n}.$$

It turns that when defined on  $\mathbf{P}(\mathbf{R})$ , this relation is not an order relation due to pathological poorly integrable measures. Indeed, there exist two probability measures  $\mu$  and  $\nu$  so that  $\mu \neq \nu$  and  $\mu * \mu = \nu * \nu$  (see [7], p. 479). For this reason, we restrict ourselves to sufficiently integrable measures (however, most of what follows generalizes to wider classes of measures).

**Definition 3.** A measure  $\mu$  on  $\mathbf{R}$  is said to be exponentially integrable if  $\int e_\lambda d\mu < +\infty$  for all  $\lambda \in \mathbf{R}$  (recall that  $e_\lambda(x) = \exp(\lambda x)$ ). We write  $\mathbf{P}_{\text{exp}}(\mathbf{R})$  for the set of exponentially integrable probability measures.

Notice that the space of exponentially integrable measures is stable under convolution. On restriction to  $\mathbf{P}_{\text{exp}}(\mathbf{R})$ , the relation  $\leq_{\text{st}}^*$  is a partial order. Indeed, if  $\mu^{*k} \leq_{\text{st}} \nu^{*k}$  and  $\nu^{*l} \leq_{\text{st}} \mu^{*l}$ , then  $\mu^{*kl} \leq_{\text{st}} \nu^{*kl} \leq_{\text{st}} \mu^{*kl}$  and therefore  $\mu^{*kl} = \nu^{*kl}$ . But if  $\mu$  and  $\nu$  are exponentially integrable, this implies that  $\mu = \nu$ . One can see this in the following way: if we denote the moments of  $\mu$  by  $m_p(\mu) = \int x^p d\mu(x)$ , one checks by induction on  $p$  that  $m_p(\mu) = m_p(\nu)$  for all  $p \in \mathbf{N}$ . On the other hand, exponential integrability implies that  $m_{2p}(\mu)^{1/2p} \leq Cp$  for some constant  $C$ , so that Carleman's condition is satisfied (see [7], p. 224). Therefore  $\mu$  is determined by its moments and  $\mu = \nu$ .

We would like to give a description of the relation  $\leq_{\text{st}}^*$ . We start with the following lemma

**Lemma 3.** Let  $\mu, \nu \in \mathbf{P}_{\text{exp}}(\mathbf{R})$  such that  $\mu \leq_{\text{st}}^* \nu$ . Then the following inequalities hold

- (a)  $\forall \lambda > 0, \int e_\lambda d\mu \leq \int e_\lambda d\nu,$
- (b)  $\forall \lambda < 0, \int e_\lambda d\mu \geq \int e_\lambda d\nu,$
- (c)  $\int x d\mu(x) \leq \int x d\nu(x),$
- (d)  $\min \mu \leq \min \nu,$
- (e)  $\max \mu \leq \max \nu,$

*Proof.* Let  $\mu \leq_{\text{st}}^* \nu$  and  $\lambda > 0$ . Since  $\mu^{*n} \leq \nu^{*n}$  for some  $n$ , we get from Lemma 1 that

$$\int e_\lambda d\mu^{*n} \leq \int e_\lambda d\nu^{*n}.$$

It remains to notice that

$$\int e_\lambda d\mu^{*n} = \left( \int e_\lambda d\mu \right)^n$$

and we get (a). The proof of (b) is completely symmetric, while (c) follows also from Lemma 1. Conditions (d) and (e) are obvious since  $\min(\mu^{*n}) = n \min(\mu)$  and  $\max(\mu^{*n}) = n \max(\mu)$ .  $\square$

The following lemma shows that the necessary conditions of Lemma 3 are “almost sufficient”.

**Lemma 4.** Let  $\mu, \nu \in \mathbf{P}_{\text{exp}}(\mathbf{R})$ . Assume that the following inequalities hold:

- (a)  $\forall \lambda > 0, \int e_\lambda d\mu < \int e_\lambda d\nu.$
- (b)  $\forall \lambda < 0, \int e_\lambda d\nu < \int e_\lambda d\mu.$
- (c)  $\int x d\mu(x) < \int x d\nu(x).$
- (d)  $\max \mu < \max \nu \leq +\infty.$
- (e)  $-\infty \leq \min \mu < \min \nu.$

Then  $\mu \leq_{\text{st}}^* \nu$ , and more precisely there exists an integer  $N \in \mathbf{N}$  such that for any  $n \geq N$ ,  $\mu^{*n} \leq_{\text{st}} \nu^{*n}$ .

We give in Proposition 2 a counter-example showing that Lemma 4 is not true when stated with large inequalities.

We are going to use Cramér's theorem on large deviations. The cumulant generating function  $\Lambda_\mu$  of the probability measure  $\mu$  is defined for any  $\lambda \in \mathbf{R}$  by

$$\Lambda_\mu(\lambda) = \log \int e_\lambda d\mu.$$

It is a convex function taking values in  $\mathbf{R}$ . Its convex conjugate  $\Lambda_\mu^*$ , sometimes called the Cramér transform, is defined as

$$\Lambda_\mu^*(t) = \sup_{\lambda \in \mathbf{R}} \lambda t - \Lambda_\mu(\lambda).$$

Note that  $\Lambda_\mu^* : \mathbf{R} \rightarrow [0, +\infty]$  is a smooth convex function, which takes the value  $+\infty$  on  $\mathbf{R} \setminus [\min \mu, \max \mu]$ . Moreover, for  $t \in (\min \mu, \max \mu)$ , the supremum in the definition of  $\Lambda_\mu^*(t)$  is attained at a unique point  $\lambda_t$ . Moreover,  $\lambda_t > 0$  if  $t > \int x d\mu(x)$  and  $\lambda_t < 0$  if  $t < \int x d\mu(x)$ . Also,  $\Lambda_\mu^*(\int x d\mu(x)) = 0$  since  $\Lambda_\mu'(0) = \int x d\mu(x)$ . We now state Cramér's theorem. The theorem can be equivalently stated in the language of sums of i.i.d. random variables [5, 9].

**Theorem** (Cramér's theorem). *Let  $\mu \in \mathbf{P}_{\text{exp}}(\mathbf{R})$ . Then for any  $t \in (\min \mu, \max \mu)$ ,*

$$(2) \quad \lim_{n \rightarrow \infty} \frac{1}{n} \log \mu^{*n}[tn, +\infty) = \begin{cases} 0 & \text{if } t \leq \int x d\mu(x) \\ -\Lambda_X^*(t) & \text{otherwise.} \end{cases}$$

$$(3) \quad \lim_{n \rightarrow \infty} \frac{1}{n} \log (1 - \mu^{*n}[tn, +\infty)) = \begin{cases} 0 & \text{if } t \geq \int x d\mu(x) \\ -\Lambda_X^*(t) & \text{otherwise.} \end{cases}$$

*Proof of Lemma 4.* Note that the hypotheses imply that the quantities  $\max \mu$  and  $\min \nu$  are finite. We write also  $M_\mu = \int x d\mu(x)$  and  $M_\nu = \int x d\nu(x)$ .

For  $n \geq 1$ , define  $(f_n)$  and  $(g_n)$  by

$$\begin{aligned} f_n(t) &= \mu^{*n}[tn, +\infty), \\ g_n(t) &= \nu^{*n}[tn, +\infty). \end{aligned}$$

We need to prove that  $f_n \leq g_n$  on  $\mathbf{R}$  for  $n$  large enough. If  $t > \max \mu$ , the inequality is trivial since  $f_n(t) = 1$ . Similarly, if  $t < \min \nu$  we have  $g_n(t) = 1$  and there is nothing to prove.

Fix a real number  $t_0$  such that  $M_\mu < t_0 < M_\nu$ . We first work on the interval  $I = [t_0, \max \mu]$ . By Cramér's theorem, the sequences  $(f_n^{1/n})$  and  $(g_n^{1/n})$  converge respectively on  $I$  toward  $f$  and  $g$  defined by

$$\begin{aligned} f(t) &= \exp(-\Lambda_\mu^*(t)), \\ g(t) &= \begin{cases} 1 & \text{if } t_0 \leq t \leq M_\nu \\ \exp(-\Lambda_\nu^*(t)) & \text{if } M_\nu \leq t \leq \max \mu \end{cases}. \end{aligned}$$

Note that  $f$  and  $g$  are continuous on  $I$ . We claim also that  $f < g$  on  $I$ . The inequality is clear on  $[t_0, M_\nu]$  since  $f < 1$ . If  $t \in (M_\nu, \max \mu]$ , note that the supremum in the definition of  $\Lambda_\nu^*(t)$  is attained for some  $\lambda > 0$  — to show this we used hypothesis (d). Using (a) and the definition of the convex conjugate, it implies that  $\Lambda_\nu^*(t) > \Lambda_\mu^*(t)$ . We now use the following elementary fact: if a sequence of non-increasing functions defined on a compact interval  $I$  converges pointwise toward a continuous limit, then the convergence is actually uniform on  $I$  (for a proof see [16] Part 2, Problem 127; this statement is attributed to Pólya or to Dini depending on authors). We apply this result to both  $(f_n^{1/n})$  and  $(g_n^{1/n})$ ; and since  $f < g$ , uniform convergence implies that for  $n$  large enough,  $f_n^{1/n} < g_n^{1/n}$  on  $I$ , and thus  $f_n \leq g_n$ .

Finally, we apply a similar argument on the interval  $J = [\min \nu, t_0]$ , except that we consider the sequences  $(1 - f_n)^{1/n}$  and  $(1 - g_n)^{1/n}$ , and we use (3) to compute the limit. We omit the details since the argument is totally symmetric.

We eventually showed that for  $n$  large enough,  $f_n \leq g_n$  on  $I \cup J$ , and thus on  $\mathbf{R}$ . This is exactly the conclusion of the lemma.  $\square$

3. GEOMETRY AND TOPOLOGY OF  $\leq_{\text{st}}^*$ 

We investigate here the topology of the relation  $\leq_{\text{st}}^*$ . We first need to define an adequate topology on  $\text{P}_{\text{exp}}(\mathbf{R})$ . This space can be topologized in several ways, an important point for us is that the map  $\mu \mapsto \int e_\lambda d\mu$  should be continuous.

**Definition 4.** A function  $f : \mathbf{R} \rightarrow \mathbf{R}$  is said to be subexponential if there exist constants  $c, C$  so that for every  $x \in \mathbf{R}$

$$|f(x)| \leq C \exp(c|x|).$$

**Definition 5.** Let  $\tau$  be the topology defined on the space of exponentially integrable measures, generated by the family of seminorms  $(N_f)$

$$N_f(\mu) = \left| \int f d\mu \right|,$$

where  $f$  belongs to the class of continuous subexponential functions.

The topology  $\tau$  is a locally convex vector space topology. It can be shown that the relation  $\leq_{\text{st}}^*$  is not  $\tau$ -closed (see Proposition 2). However, its closure can be described by the following theorem.

**Theorem 1.** Let  $R \subset \text{P}_{\text{exp}}(\mathbf{R})^2$  be the set of couples  $(\mu, \nu)$  of exponentially integrable probability measures so that  $\mu \leq_{\text{st}}^* \nu$ . Then

$$(4) \quad \overline{R} = \left\{ (\mu, \nu) \in \text{P}_{\text{exp}}(\mathbf{R})^2 \text{ s.t. } \forall \lambda \geq 0, \int e_\lambda d\mu \leq \int e_\lambda d\nu \text{ and } \forall \lambda \leq 0, \int e_\lambda d\mu \geq \int e_\lambda d\nu \right\},$$

the closure being taken with respect to the topology  $\tau$ .

*Proof.* Let us write  $X$  for the set on the right-hand side of (4). We get from Lemma 3 that  $R \subset X$ . Moreover, it is easily checked that  $X$  is  $\tau$ -closed, therefore  $\overline{R} \subset X$ . Conversely, we are going to show that the set of couples  $(\mu, \nu)$  satisfying the hypotheses of Lemma 4 is  $\tau$ -dense in  $X$ . Let  $(\mu, \nu) \in X$ . We get from the inequalities satisfied by  $\mu$  and  $\nu$  that

- $\int x d\mu(x) \leq \int x d\nu(x)$  (taking derivatives at  $\lambda = 0$ ),
- $\min \mu \leq \min \nu$  (taking  $\lambda \rightarrow -\infty$ ),
- $\max \mu \leq \max \nu$  (taking  $\lambda \rightarrow +\infty$ ).

We want to define two sequences  $(\mu_n, \nu_n)$  which  $\tau$ -converge toward  $(\mu, \nu)$ , with  $\mu_n \leq_{\text{st}} \mu$  and  $\nu \leq_{\text{st}} \nu_n$  and for which the above inequalities become strict. Assume for example that  $\max \mu = \max \nu = +\infty$  and  $\min \mu = \min \nu = -\infty$ . Then we can define  $\mu_n$  and  $\nu_n$  as follows: let  $\varepsilon_n = \mu[n, +\infty)$  and  $\eta_n = \nu(-\infty, -n]$ , and set

$$\begin{aligned} \mu_n &= \mu|_{(-\infty, n)} + \varepsilon_n \delta_n, \\ \nu_n &= \nu|_{(-n, +\infty)} + \eta_n \delta_{-n}. \end{aligned}$$

We check using dominated convergence that  $\lim \mu_n = \mu$  and  $\lim \nu_n = \nu$  with respect to  $\tau$ , while by Lemma 4 we have  $\mu_n \leq_{\text{st}}^* \nu_n$ . The other cases are treated in a similar way: we can always play with small Dirac masses to make all inequalities strict (for example, if  $\max \mu = \max \nu = M < +\infty$ , replace  $\nu$  by  $(1 - \varepsilon)\nu + \varepsilon\delta_{M+1}$ , and so on).  $\square$

A more comfortable way of describing the relation  $\leq_{\text{st}}^*$  is given by the following sets

**Definition 6.** Let  $\nu \in \text{P}_{\text{exp}}(\mathbf{R})$ . We define  $D(\nu)$  to be the following set

$$D(\nu) = \{ \mu \in \text{P}_{\text{exp}}(\mathbf{R}) \text{ s.t. } \mu \leq_{\text{st}}^* \nu \}.$$

Using the ideas in the proof of Theorem 1, it can easily be showed that for  $\nu \in \mathbf{P}_{\text{exp}}(\mathbf{R})$  such that  $\min \nu > -\infty$ , one has

$$(5) \quad \overline{D(\nu)} = \left\{ \mu \in \mathbf{P}_{\text{exp}}(\mathbf{R}) \text{ s.t. } \forall \lambda \geq 0, \int e_{\lambda} d\mu \leq \int e_{\lambda} d\nu \text{ and } \forall \lambda \leq 0, \int e_{\lambda} d\mu \geq \int e_{\lambda} d\nu \right\},$$

where the closure is taken in the topology  $\tau$ . However, for measures  $\nu$  with  $\min \nu = -\infty$ , the condition (e) of Lemma 4 is violated and we do not know if the relation (5) holds.

Another consequence of equation (5) is that the  $\tau$ -closure of  $D(\nu)$  is a convex set. It is not clear that the set  $D(\nu)$  itself is convex. We shall see in Proposition 3 that this is not the case in general for measures  $\nu \notin \mathbf{P}_{\text{exp}}(\mathbf{R})$ . Not also that for fixed  $\nu \in \mathbf{P}(\mathbf{R})$  the set  $\{\mu \in \mathbf{P}(\mathbf{R}) \text{ s.t. } \mu \leq_{\text{st}} \nu\}$  is easily checked to be convex.

**Remark 1.** *One can analogously define for  $\mu \in \mathbf{P}_{\text{exp}}(\mathbf{R})$  the “dual” set  $E(\mu) = \{\nu \in \mathbf{P}_{\text{exp}}(\mathbf{R}) \text{ s.t. } \mu \leq_{\text{st}}^* \nu\}$ . Results about  $D(\nu)$  or  $E(\mu)$  are equivalent. Indeed, let  $\mu^{\leftrightarrow}$  be the measure defined for a Borel set  $B$  by  $\mu^{\leftrightarrow}(B) = \mu(-B)$ . We have  $\mu \leq_{\text{st}}^* \nu \iff \nu^{\leftrightarrow} \leq_{\text{st}}^* \mu^{\leftrightarrow}$  and therefore  $E(\mu) = D(\mu^{\leftrightarrow})^{\leftrightarrow}$ .*

We now give an example showing that the relation  $\leq_{\text{st}}^*$  is not  $\tau$ -closed.

**Proposition 2.** *There exists a probability measure  $\nu \in \mathbf{P}_{\text{exp}}(\mathbf{R})$  so that the set  $D(\nu)$  is not  $\tau$ -closed. Consequently, the set  $R$  appearing in (4) is not closed either.*

*Proof.* For  $k \geq 1$ , let  $a_k = (k+2)!$ ,  $b_k = (k+2)! + 1$  and  $\gamma_k = c \exp(-k^k)$ , where the constant  $c$  is chosen so that  $\sum \gamma_k = 1$ . We check that  $(a_k)$  and  $(b_k)$  satisfy the following inequalities

$$(6) \quad (k-1)b_k + b_{k-1} < ka_k,$$

$$(7) \quad kb_k < a_{k+1}.$$

Start with the following fact which follows from Proposition 1: for each  $k \in \mathbf{N}$  there exist  $\mu_k$  and  $\nu_k$ , probability measures with finite support such that  $\mu_k \in D(\nu_k)$  while  $\mu_k^{*k} \not\leq_{\text{st}} \nu_k^{*k}$ . Moreover, we can assume that  $\text{supp}(\mu_k) \subset (a_k, b_k)$  and  $\text{supp}(\nu_k) \subset (a_k, b_k)$ . Indeed, we can apply to both measures a suitable affine transformation (increasing affine transformations preserve stochastic domination and are compatible with convolution). We now define  $\mu$  and  $\nu$  as

$$\mu = \sum_{k=1}^{\infty} \gamma_k \mu_k \quad \text{and} \quad \nu = \sum_{k=1}^{\infty} \gamma_k \nu_k.$$

Note that the sequence  $(\gamma_k)$  has been chosen to tend very quickly to 0 to ensure that  $\mu$  and  $\nu$  are exponentially integrable. We also introduce the following sequences of measures

$$\begin{aligned} \tilde{\mu}_n &= \sum_{k=1}^n \gamma_k \mu_k + \left( \sum_{k=n+1}^{\infty} \gamma_k \right) \delta_0, \\ \tilde{\nu}_n &= \sum_{k=1}^n \gamma_k \nu_k + \left( \sum_{k=n+1}^{\infty} \gamma_k \right) \delta_0. \end{aligned}$$

One checks using Lebesgue’s dominated convergence theorem that the sequences  $(\tilde{\mu}_n)$  and  $(\tilde{\nu}_n)$  converge respectively toward  $\mu$  and  $\nu$  for the topology  $\tau$ . Note also that this sequences are increasing with respect to stochastic domination, so that  $\tilde{\nu}_n \leq_{\text{st}} \nu$ . For fixed  $k$ ,  $\mu_k$  and  $\nu_k$  satisfy the hypotheses of Lemma 4 and thus the same holds for  $\tilde{\mu}_n$  and  $\tilde{\nu}_n$ . Therefore  $\tilde{\mu}_n \in D(\tilde{\nu}_n) \subset D(\nu)$ . This proves that  $\mu \in \overline{D(\nu)}$ .

We now prove by contradiction that  $\mu \notin D(\nu)$ . Assume that  $\mu \in D(\nu)$ , i.e.  $\mu^{*k} \leq_{\text{st}} \nu^{*k}$  for some  $k \geq 1$ . Let  $s_k = ka_k$  and  $t_k = kb_k$ . Fix a sequence  $i_1, \dots, i_k$  of nonzero integers. Set  $m = \mu_{i_1} * \dots * \mu_{i_k}$

or  $m = \nu_{i_1} * \cdots * \nu_{i_k}$ . We know that  $\text{supp}(m) \subset (a, b)$ , with  $a = \sum_{j=1}^k a_{i_j}$  and  $b = \sum_{j=1}^k b_{i_j}$ . It is possible to locate precisely  $\text{supp}(m)$  using the inequalities (6) and (7).

- (a) If  $i_j > k$  for some  $j$ , then  $a \geq a_{k+1} > t_k$  and therefore  $\text{supp}(m) \subset (t_k, +\infty)$ .
- (b) If  $i_j = k$  for all  $j$ , then  $a = s_k$  and  $b = t_k$  and therefore  $\text{supp}(m) \subset (s_k, t_k)$ .
- (c) If  $i_j \leq k$  for all  $j$  and  $i_{j_0} < k$  for some  $j_0$ , then  $b \leq b_{k-1} + (k-1)b_k < s_k$  and therefore  $\text{supp}(m) \subset [0, s_k)$ .

Consequently,

$$\mu^{*k}[t_k, +\infty) = \sum_{i_1, \dots, i_k} \gamma_{i_1} \cdots \gamma_{i_k} \mu_{i_1} * \cdots * \mu_{i_k}[t_k, +\infty) = \sum_{i_1, \dots, i_k \text{ satisfying (a)}} \gamma_{i_1} \cdots \gamma_{i_k} = \nu^{*k}[t_k, +\infty).$$

Moreover, because of (b) and (c), we get that for  $s_k \leq t \leq t_k$ ,

$$\mu^{*k}[t, t_k) = \gamma_k^k \mu_k^{*k}[t, t_k) = \gamma_k^k \mu_k^{*k}[t, +\infty).$$

and similarly

$$\nu^{*k}[t, t_k) = \gamma_k^k \nu_k^{*k}[t, +\infty).$$

We assumed that  $\mu^{*k} \leq_{\text{st}} \nu^{*k}$ , i.e.  $\mu^{*k}[t, +\infty) \leq \nu^{*k}[t, +\infty)$  for all  $t$ . If  $t \leq t_k$ , since  $\mu^{*k}(t_k, +\infty) = \nu^{*k}(t_k, +\infty)$ , we get that  $\mu^{*k}[t, t_k) \leq \nu^{*k}[t, t_k)$ . Since  $\gamma_k > 0$ , this implies that for all  $t \geq s_k$ ,  $\mu_k^{*k}[t, +\infty) \leq \nu_k^{*k}[t, +\infty)$ . This contradicts the fact that  $\mu_k^{*k} \not\leq_{\text{st}} \nu_k^{*k}$ . Therefore  $\mu \in \overline{D(\nu)} \setminus D(\nu)$ , and so  $D(\nu)$  is not closed.  $\square$

We now give an example of pathologies that can happen if we consider measures with poor integrability properties.

**Proposition 3.** *There exists a probability measure  $\nu \in \mathbf{P}(\mathbf{R})$  such that the set*

$$(8) \quad \{\mu \in \mathbf{P}(\mathbf{R}) \text{ s.t. } \mu \leq_{\text{st}}^* \nu\}$$

*is not convex.*

The difference between equation (8) and our definition of  $D(\nu)$  is that here we do not suppose the measures to be exponentially integrable.

*Proof.* We rely on the following fact which we already alluded to (see [7], p. 479): there exist two distinct real characteristic functions  $\phi_1$  and  $\phi_2$  such that  $\phi_1^2 = \phi_2^2$  identically. Consider now the measures  $\mu$  and  $\nu$  with respective characteristic functions  $\phi_1$  and  $\phi_2$ , i.e.  $\phi_1(t) = \int e^{it} d\mu(t)$  and  $\phi_2(t) = \int e^{it} d\nu(t)$ . Obviously, we have  $\nu \leq_{\text{st}}^* \nu$  and  $\mu \leq_{\text{st}}^* \nu$  since  $\mu^{*2} = \nu^{*2}$ . Let  $\chi = \frac{1}{2}\mu + \frac{1}{2}\nu$  and let us show that  $\chi \not\leq_{\text{st}}^* \nu$ . We have

$$\begin{aligned} \chi^{*2n} &= \frac{1}{2^{2n}} \sum_{i=0}^{2n} \binom{2n}{i} \mu^{*i} * \nu^{*2n-i} = \\ &= \frac{1}{2^{2n}} \left[ \sum_{i \text{ even}} \binom{2n}{i} \nu^{*2n} + \sum_{i \text{ odd}} \binom{2n}{i} \nu^{*2n-1} * \mu \right]. \end{aligned}$$

Thus  $\chi^{*2n} \leq_{\text{st}} \nu^{*2n}$ , is equivalent to  $\nu^{*2n-1} * \mu \leq_{\text{st}} \nu^{*2n}$ . Let us show that this is impossible. Indeed, the measures  $\nu^{*2n-1} * \mu$  and  $\nu^{*2n}$  have real characteristic functions and thus they are symmetric probability measures. Note however that two symmetric probability distributions cannot be compared with  $\leq_{\text{st}}$  unless they are equal. But it cannot be that  $\nu^{*2n-1} * \mu = \nu^{*2n}$  because their characteristic functions are different ( $\phi_1(\xi) = \phi_2(\xi)$  iff.  $\phi_1(\xi) = 0$ ). A similar argument holds for  $\chi^{*2n+1} \not\leq_{\text{st}} \nu^{*2n+1}$ .  $\square$

We conclude this section with few remarks on a relation which is very similar to  $\leq_{\text{st}}^*$ . It is the analogue of catalytic majorization in quantum information theory (see Section 4).



**Definition 7.** Let  $\mu, \nu \in \mathbf{P}_{\text{exp}}(\mathbf{R})$ . We say that  $\mu$  is *catalytically stochastically dominated* by  $\nu$  and write  $\mu \leq_{\text{st}}^{\text{C}} \nu$  if there exists a probability measure  $\pi \in \mathbf{P}_{\text{exp}}(\mathbf{R})$  such that  $\mu * \pi \leq_{\text{st}} \nu * \pi$ .

The following lemma shows a connection between the two relations.

**Lemma 5.** Let  $\mu, \nu \in \mathbf{P}_{\text{exp}}(\mathbf{R})$ . Assume  $\mu \leq_{\text{st}}^* \nu$ . Then  $\mu \leq_{\text{st}}^{\text{C}} \nu$ .

*Proof.* Assume that  $\mu^{*n} \leq_{\text{st}} \nu^{*n}$  for some  $n$ . Let  $\pi$  the probability measure defined by

$$\pi = \frac{1}{n} \sum_{k=0}^{n-1} \mu^{*k} * \nu^{*(n-k)}.$$

Let also  $\rho$  be the measure defined by

$$\rho = \frac{1}{n} \sum_{k=1}^{n-1} \mu^{*k} * \nu^{*(n-k)},$$

then one has  $\mu * \pi = \frac{1}{n} \mu^{*n} + \rho$  and  $\nu * \pi = \frac{1}{n} \nu^{*n} + \rho$ , and since  $\mu^{*n} \leq_{\text{st}} \nu^{*n}$  this implies  $\mu * \pi \leq_{\text{st}} \nu * \pi$ . Since  $\pi \in \mathbf{P}_{\text{exp}}(\mathbf{R})$ , we get  $\mu \leq_{\text{st}}^{\text{C}} \nu$ .  $\square$

From Theorem 1 and Lemma 5 one can easily derive the

**Corollary 2.** The analogue of Theorem 1 is true if we substitute  $\leq_{\text{st}}^*$  with  $\leq_{\text{st}}^{\text{C}}$ .

#### 4. CATALYTIC MAJORIZATION

This section is dedicated to the study of the majorization relation, the notion which was the initial motivation of this work. The majorization relation provides, much as the stochastic domination for probability measures, a partial order on the set of probability vectors. Originally introduced in linear algebra [12], [3], it has found many application in quantum information theory with the work of Nielsen [13], [14]. We shall not focus on quantum-theoretical aspects of majorization; we refer the interested reader to [1] and references therein. Here, we study majorization by adapting previously obtained results for stochastic domination.

The majorization relation is defined for *probability vectors*, i.e. vectors  $x \in \mathbf{R}^{\mathbf{N}}$  with non-negative components ( $x_i \geq 0$ ) which sum up to one ( $\sum_i x_i = 1$ ). Before defining precisely majorization, let us introduce some notation. For  $d \in \mathbf{N}^*$ , let  $P_d$  be the set of  $d$ -dimensional probability vectors :  $P_d = \{x \in \mathbf{R}^d \text{ s.t. } x_i \geq 0, \sum x_i = 1\}$ . Consider also the set of finitely supported probability vectors  $P_{<\infty} = \bigcup_{d>0} P_d$ . For a vector  $x \in P_{<\infty}$ , we write  $x_{\max}$  for the largest component of  $x$  and  $x_{\min}$  for its smallest non-zero component. In this section we shall consider only finitely supported vectors. For the general case, see Section 6. We shall identify an element  $x \in P_d$  with the corresponding element in  $P_{d'}$  ( $d' > d$ ) or  $P_{<\infty}$  obtained by appending null components at the end of  $x$ .

Next, we define  $x^\downarrow$ , the decreasing rearrangement of a vector  $x \in P_d$  as the vector which has the same coordinates as  $x$  up to permutation and such that  $x_i^\downarrow \geq x_{i+1}^\downarrow$  for all  $1 \leq i < d$ . We can now define majorization in terms of the ordered vectors:

**Definition 8.** For  $x, y \in P_d$  we say that  $x$  is *majorized* by  $y$  and we write  $x \prec y$  if for all  $k \in \{1, \dots, d\}$

$$(9) \quad \sum_{i=1}^k x_i^\downarrow \leq \sum_{i=1}^k y_i^\downarrow.$$

Note however that there are several equivalent definitions of majorization which do not use the ordering of the vectors  $x$  and  $y$  (see [3] for further details):

**Proposition 4.** *The following assertions are equivalent:*

- (1)  $x \prec y$ ,
- (2)  $\forall t \in \mathbf{R}, \sum_{i=1}^d |x_i - t| \leq \sum_{i=1}^d |y_i - t|$ ,
- (3)  $\forall t \in \mathbf{R}, \sum_{i=1}^d (x_i - t)^+ \leq \sum_{i=1}^d (y_i - t)^+$ , where  $z^+ = \max(z, 0)$ ,
- (4) *There is a bistochastic matrix  $B$  such that  $x = By$ .*

There are two operations on probability vectors which are of particular interest to us: the tensor product and the direct sum. For  $x = (x_1, \dots, x_d) \in P_d$  and  $x' = (x'_1, \dots, x'_{d'}) \in P_{d'}$ , we define the tensor product  $x \otimes x'$  as the vector  $(x_i x'_j)_{ij} \in P_{dd'}$ . We also define the direct sum  $x \oplus x'$  as the concatenated vector  $(x_1, \dots, x_d, x'_1, \dots, x'_{d'}) \in \mathbf{R}^{d+d'}$ . Note that if we take  $\oplus$ -convex combinations, we get probability vectors:  $\lambda x \oplus (1 - \lambda)x' \in P_{d+d'}$ .

The construction which permits us to use tools from stochastic domination in the framework of majorization is the following (inspired by [11]): to a probability vector  $z \in P_{<\infty}$  we associate a probability measure  $\mu_z$  defined by:

$$\mu_z = \sum z_i \delta_{\log z_i}.$$

These measures behave well with respect to tensor products:

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

The connection between majorization and stochastic domination is provided by the following lemma:

**Lemma 6.** *Let  $x, y \in P_{<\infty}$ . Assume that  $\mu_x \leq_{\text{st}} \mu_y$ . Then  $x \prec y$ .*

*Proof.* We can assume that  $x = x^\downarrow$  and  $y = y^\downarrow$ . Note that

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

Thus, for all  $u > 0$ ,  $\sum_{i: x_i \geq u} x_i \leq \sum_{i: y_i \geq u} y_i$ . To start, use  $u = y_1$  to conclude that  $x_1 \leq y_1$ . Notice that it suffices to show that  $\sum_{i=1}^k x_i \leq \sum_{i=1}^k y_i$  only for those  $k$  such that  $x_k > y_k$  (indeed, if  $x_k \leq y_k$ , the  $(k+1)$ -th inequality in (9) can be deduced from the  $k$ -th inequality). Consider such a  $k$  and let  $x_k > u > y_k$ . We get:

$$\sum_{i=1}^k x_i \leq \sum_{i: x_i \geq u} x_i \leq \sum_{i: y_i \geq u} y_i \leq \sum_{i=1}^k y_i,$$

which completes the proof of the lemma.  $\square$

**Remark 2.** *The converse of this lemma does not hold. Indeed, consider  $x = (0.5, 0.5)$  and  $y = (0.9, 0.1)$ . Obviously,  $x \prec y$  but  $1 = \mu_x[\log 0.5, \infty) > \mu_y[\log 0.5, \infty) = 0.9$  and thus  $\mu_x \not\leq_{\text{st}} \mu_y$ .*

We can describe the majorization relation by the sets:

$$S_d(y) = \{x \in P_d | x \prec y\},$$

where  $y$  is a finitely supported probability vector. Mathematically, such a set is characterized by the following lemma, which is a simple consequence of Birkhoff's theorem on bistochastic matrices:

**Lemma 7.** *For  $y$  a  $d$ -dimensional probability vector, the set  $S(y)$  is a polytope whose extreme points are  $y$  and its permutations.*

The initial motivation for our work was the following phenomena discovered in quantum information theory (see [10] and respectively [2]). It turns out that additional vectors can act as *catalysts* for the majorization relation: there are vectors  $x, y, z \in P_{<\infty}$  such that  $x \not\prec y$  but  $x \otimes z \prec y \otimes z$ ; in such a situation we say that  $x$  is catalytically majorized (or *trumped*) by  $y$  and we write  $x \prec_T y$ . Another form of catalysis is provided by *multiple-copies* of vectors: we can find vectors  $x$  and  $y$  such that  $x \not\prec y$  but still, for some  $n \geq 2$ ,  $x^{\otimes n} \prec y^{\otimes n}$ ; in this case we write  $x \prec_M y$ . We have thus two new order relations on probability vectors, analogues of  $\leq_{\text{st}}^{\text{C}}$  and respectively  $\leq_{\text{st}}^*$ . As before, for  $y \in P_d$ , we introduce the sets

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

and

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

It turns out that the relations  $\prec_T$  and  $\prec_M$  (and thus the sets  $T_d(y)$  and  $M_d(y)$ ) are not as simple as  $\prec$  and  $S_d(y)$ . It is known that the inclusion  $M_d(y) \subset T_d(y)$  holds (this is the analogue of Lemma 5) and that it can be strict [8]. In general, the sets  $T_d(y)$  and  $M_d(y)$  are neither closed nor open, and although  $T_d(y)$  is known to be convex, nothing is known about the convexity of  $M_d(y)$  (such questions have been intensively studied in the physical literature; see [6, 4] and the references therein). As explained in [1] it is natural from a mathematical point of view to introduce the sets  $T_{<\infty}(y) = \bigcup_{d \in \mathbf{N}} T_d(y)$  and  $M_{<\infty}(y) = \bigcup_{d \in \mathbf{N}} M_d(y)$ . A key notion in characterizing them is *Schur-convexity*:

**Definition 9.** A function  $f : P_d \rightarrow \mathbf{R}$  is said to be

- *Schur-convex* if  $f(x) \leq f(y)$  whenever  $x \prec y$ ,
- *Schur-concave* if  $f(x) \geq f(y)$  whenever  $x \prec y$ ,
- *strictly Schur-convex* if  $f(x) < f(y)$  whenever  $x \not\prec y$ ,
- *strictly Schur-concave* if  $f(x) > f(y)$  whenever  $x \not\prec y$ ,

where  $x \not\prec y$  means  $x \prec y$  and  $x^\perp \neq y^\perp$ .

Examples are provided as follows: if  $\Phi : \mathbf{R} \rightarrow \mathbf{R}$  is a (strictly) convex/concave function, then the following function  $h : P_d \rightarrow \mathbf{R}$  defined by  $h(x_1, \dots, x_d) = \Phi(x_1) + \dots + \Phi(x_d)$  is (strictly) Schur-convex/Schur-concave.

For  $x \in P_d$  and  $p \in \mathbf{R}$ , we define  $N_p(x)$  as

$$N_p(x) = \sum_{\substack{1 \leq i \leq d \\ x_i > 0}} x_i^p.$$

We will also use the Shannon entropy  $H$

$$H(x) = - \sum_{i=1}^d x_i \log x_i.$$

Note that  $-H(x)$  is the derivative of  $p \mapsto N_p(x)$  at  $p = 1$  and that  $N_0(x)$  is the number of non-zero components of the vector  $x$ . These functions satisfy the following properties:

- (1) If  $p > 1$ ,  $N_p$  is strictly Schur-convex on  $P_{<\infty}$ .
- (2) If  $0 < p < 1$ ,  $N_p$  is strictly Schur-concave on  $P_{<\infty}$ .
- (3) If  $p < 0$ ,  $N_p$  is strictly Schur-convex on  $P_d$  for any  $d$ . However, for  $p < 0$ , it is not possible to compare vectors with a different number of non-zero components.
- (4)  $H$  is strictly Schur-concave on  $P_{<\infty}$ .

One possible way of describing the relations  $\prec_M$  and  $\prec_T$  is to find a family (the smallest possible) of Schur-convex functions which characterizes them. In this direction, Nielsen conjectured the following result:

**Conjecture 1.** Fix a vector  $y \in P_d$ , with nonzero coordinates. Then  $\overline{T_d(y)} = \overline{M_d(y)}$  and they both are equal to the set of  $x \in P_d$  satisfying

- (C1) For  $p \geq 1$ ,  $N_p(x) \leq N_p(y)$ .
- (C2) For  $0 < p \leq 1$ ,  $N_p(x) \geq N_p(y)$ .
- (C3) For  $p < 0$ ,  $N_p(x) \leq N_p(y)$ .

Here, the closures are taken in  $\mathbf{R}^d$  (recall that neither  $M_d(y)$  nor  $T_d(y)$  is closed). By the previous remarks, any vector in  $T_d(y)$  or  $M_d(y)$  (and by continuity, also in the closures) must satisfy conditions (C1-C3). The other direction is still open.

Note that the conjecture can be reformulated as follows: if  $x, y \in P_d$  and satisfy (C1-C3), then there exists a sequence  $(x_n)$  in  $M_d(y)$  (and therefore in  $T_d(y)$ ) such that  $(x_n)$  converges to  $x$ . If we relax the condition that  $x_n$  and  $y$  have the same dimension, we can prove the following two theorems:

**Theorem 2.** If  $x, y \in P_d$  and satisfy (C1), then there exists a sequence  $(x_n)$  in  $M_{<\infty}(y)$  (and therefore in  $T_{<\infty}(y)$ ) such that  $(x_n)$  converges to  $x$  in  $\ell_1$ -norm.

Recall that the  $\ell_1$  norm is defined by

$$\|x\|_1 = \sum_i |x_i|.$$

**Theorem 3.** If  $x, y \in P_d$  and satisfy (C1-C2), then there exists a sequence  $(x_n)$  in  $M_{d+1}(y)$  (and therefore in  $T_{d+1}(y)$ ) such that  $(x_n)$  converges to  $x$ .

Theorem 2 restates the authors' previous result in [1]; however, the proof presented in the next section is more transparent than the previous one. The second theorem answers a question of [1]. It is an intermediate result between Theorem 2 and Nielsen's Conjecture.

### 5. PROOF OF THE THEOREMS

We show here how to derive Theorems 2 and 3. We first state a lemma which is the translation of Lemma 4 in terms of majorization.

**Lemma 8.** Let  $x, y \in P_{<\infty}$ . Assume that  $x$  and  $y$  have nonzero coordinates, and respective dimensions  $d_x$  and  $d_y$ . Assume that

- (1)  $x_{\min} < y_{\min}$ .
- (2)  $x_{\max} < y_{\max}$ .
- (3)  $H(x) > H(y)$ .
- (4)  $N_p(x) < N_p(y)$  for all  $p \in ]1, +\infty[$ .
- (5)  $N_p(x) > N_p(y)$  for all  $p \in ]-\infty, 1[$ .

Then there exists an integer  $N$  such that for all  $n \geq N$ , we have  $x^{\otimes n} \prec y^{\otimes n}$ .

It is important to notice that since  $N_0(x) = d_x$  and  $N_0(y) = d_y$ , the conditions of the lemma can be satisfied only when  $d_x > d_y$ . This is the main reason why our approach fails to prove Conjecture 1.

*Proof.* One checks that the probability measures  $\mu_x$  and  $\mu_y$  associated to the vectors  $x$  and  $y$  satisfy the hypotheses of Lemma 4. Indeed, for  $p \in \mathbf{R}$ , one has

$$N_p(x) = \int e_\lambda d\mu_x, \quad \text{with } \lambda = p - 1.$$

As  $\mu_x^{*n} = \mu_{x^{\otimes n}}$ , there exists a integer  $N$  such that for  $n \geq N$ , we have  $\mu_{x^{\otimes n}} \leq_{st} \mu_{y^{\otimes n}}$ . It remains to apply the Lemma 6 in order to complete the proof.  $\square$

The main idea used in the following proofs is to slightly modify the vector  $x$  so that the couple  $(x, y)$  satisfies the hypotheses of Lemma 8.

*Proof of Theorem 2.* Let  $x, y \in P_d$  satisfying  $N_p(x) \leq N_p(y)$  for all  $p \geq 1$ . Since  $N_1(x) = N_1(y) = 1$  and  $-H = \frac{dN_p}{dp}|_{p=1}$ , we also have  $-H(x) \leq -H(y)$ . For  $0 < \varepsilon < \frac{d}{d+1}x_{\min}$ , define  $x_\varepsilon \in P_{d+1}$  by

$$x_\varepsilon = (x_1 - \frac{\varepsilon}{d}, \dots, x_d - \frac{\varepsilon}{d}, \varepsilon).$$

One checks that  $x_\varepsilon \not\geq x$  and therefore  $N_p(x_\varepsilon) < N_p(x) \leq N_p(y)$  for any  $p > 1$ , and  $-H(x_\varepsilon) < -H(x) \leq -H(y)$ . Since  $-H = \frac{dN_p}{dp}|_{p=1}$  and the function  $p \mapsto N_p(\cdot)$  is continuous, this means that there exists some  $0 < p_\varepsilon < 1$  such that  $N_p(x_\varepsilon) \geq N_p(y)$  for any  $p \in [p_\varepsilon, 1]$ . Choose an integer  $k \geq 2$ , depending on  $\varepsilon$ , such that

$$k > \max\{d^{1/(1-p_\varepsilon)}\varepsilon^{-p_\varepsilon/(1-p_\varepsilon)}, \frac{\varepsilon}{y_{\min}}, d\}$$

and define  $x_{\varepsilon,k} \in P_{<\infty}$  as

$$x_{\varepsilon,k} = (x_1 - \frac{\varepsilon}{d}, \dots, x_d - \frac{\varepsilon}{d}, \underbrace{\frac{\varepsilon}{k}, \dots, \frac{\varepsilon}{k}}_{k \text{ times}}).$$

For any  $0 \leq p \leq p_\varepsilon$  we have

$$N_p(x_{\varepsilon,k}) \geq k \left(\frac{\varepsilon}{k}\right)^p > d \geq N_p(y),$$

and for any  $p < 0$  we have

$$N_p(x_{\varepsilon,k}) \geq k \left(\frac{\varepsilon}{k}\right)^p > dy_{\min}^p \geq N_p(y).$$

We also have  $x_{\varepsilon,k} \not\geq x_\varepsilon$  and therefore  $N_p(x_{\varepsilon,k}) > N_p(x_\varepsilon) \geq N_p(y)$  for  $p_\varepsilon \leq p < 1$ . Similarly,  $N_p(x_{\varepsilon,k}) < N_p(x_\varepsilon) \leq N_p(y)$  for  $p > 1$ . This means that  $x_{\varepsilon,k}$  and  $y$  satisfy the hypotheses of Lemma 8, and therefore  $x_{\varepsilon,k} \in M_{<\infty}(y)$ . Since  $\|x_{\varepsilon,k} - x\|_1 \leq 2\varepsilon$  and  $\varepsilon$  can be chosen arbitrarily small, this completes the proof of the theorem.  $\square$

*Proof of Theorem 3.* Let  $x, y \in P_d$  satisfying  $N_p(x) \leq N_p(y)$  for  $p \geq 1$  and  $N_p(x) \geq N_p(y)$  for  $0 \leq p \leq 1$ . As in the previous proof, we consider for  $0 < \varepsilon < \frac{d}{d+1}x_{\min}$  the vector  $x_\varepsilon$  defined as

$$x_\varepsilon = (x_1 - \frac{\varepsilon}{d}, \dots, x_d - \frac{\varepsilon}{d}, \varepsilon).$$

We are going to show using Lemma 8 that for  $\varepsilon$  small enough,  $x_\varepsilon$  is in  $M_{d+1}(y)$ . Note that  $x_\varepsilon \not\geq x$ , and therefore  $N_p(x_\varepsilon) < N_p(x) \leq N_p(y)$  for  $p > 1$ , and  $N_p(x_\varepsilon) > N_p(x) \geq N_p(y)$  for  $0 < p < 1$ . Also, since  $N_0(x_\varepsilon) = d+1$  and  $N_0(y) = d$ , there exists by continuity a number  $p_0 < 0$  (not depending on  $\varepsilon$ ) such that  $N_p(y) < d+1$  for all  $p \in [p_0, 0]$ . Thus for  $p \in [p_0, 0]$  we have

$$N_p(x_\varepsilon) \geq N_0(x_\varepsilon) = d+1 > N_p(y).$$

It remains to notice that for  $\varepsilon < d^{1/p_0}y_{\min}$ , we have for any  $p \leq p_0$

$$N_p(x_\varepsilon) \geq \varepsilon^p > dy_{\min}^p \geq N_p(y).$$

We checked that  $x_\varepsilon$  and  $y$  satisfy the hypotheses of Lemma 8, and therefore  $x_\varepsilon \in M_{d+1}(y)$ . Since  $\|x_\varepsilon - y\|_1 \leq 2\varepsilon$  and  $\varepsilon$  can be chosen arbitrarily small, this completes the proof of the theorem.  $\square$

6. INFINITE DIMENSIONAL CATALYSIS

In light of the recent paper [15], we investigate the majorization relation and its generalizations for infinitely-supported probability vectors. Let us start by adapting the key tools used in the previous section to this non-finite setting.

First, note that when defining the decreasing rearrangement  $x^\downarrow$  of a vector  $x$ , we shall ask that only the *non-zero* components of  $x$  and  $x^\downarrow$  should be the same up to permutation. The majorization relation  $\prec$  extends trivially to  $P_\infty$ , the set of (possibly infinite) probability vectors. The same holds for the relations  $\prec_M$  and  $\prec_T$  (note however that for  $\prec_T$ , we allow now infinite-dimensional catalysts).

Note that for a general probability vector, there is no reason that  $N_p$  for  $p \in (0, 1)$  or  $H$  should be finite; moreover, if the support of  $x$  is not finite, then  $N_p(x) = \infty$  for all  $p \leq 0$ . He have thus to replace the hypothesis (C1) by the following one:

(C1') For  $p \geq 1$ ,  $N_p(x) \leq N_p(y)$  and  $H(x) < \infty$ .

Notice however that the inequalities  $N_p(x) \leq N_p(y)$  for  $p \rightarrow 1^+$  imply that  $H(y) \leq H(x) < \infty$  and thus both entropies are finite.

**Theorem 4.** *If  $x, y \in P_\infty$  and satisfy (C1'), then, for all  $\varepsilon > 0$  there exist finitely supported vectors  $x_\varepsilon, y_\varepsilon \in P_{<\infty}$  and  $n \in \mathbf{N}$  such that  $\|x - x_\varepsilon\|_1 \leq \varepsilon$ ,  $\|y - y_\varepsilon\|_1 \leq \varepsilon$  and  $x_\varepsilon^{\otimes n} \prec y_\varepsilon^{\otimes n}$ .*

*Proof.* Fix  $\varepsilon > 0$  small enough. If  $y$  has infinite support, consider the truncated vector  $y_\varepsilon = (y_1 + R(\varepsilon), y_2, \dots, y_{N(\varepsilon)})$ , where  $N(\varepsilon)$  and  $R(\varepsilon)$  are such that  $R(\varepsilon) = \sum_{i=N(\varepsilon)+1}^\infty y_i \leq \varepsilon$ ; otherwise put  $y_\varepsilon = y$ . Clearly, we have  $\|y - y_\varepsilon\|_1 \leq 2\varepsilon$  and  $N_p(y_\varepsilon) \geq N_p(y)$  for all  $p > 1$ . If the vector  $x$  is finite, use Theorem 2 with  $x_\varepsilon = x$  and  $y_\varepsilon$  to conclude. Otherwise, consider  $M(\varepsilon)$  such that  $S(\varepsilon) = \sum_{i=M(\varepsilon)+1}^\infty x_i \leq \varepsilon$  and define the vector

$$x_\varepsilon = (x_1, x_2, \dots, x_{M(\varepsilon)}, \underbrace{\frac{S(\varepsilon)}{k}, \frac{S(\varepsilon)}{k}, \dots, \frac{S(\varepsilon)}{k}}_{k \text{ times}}),$$

where  $k$  is a constant depending on  $\varepsilon$  which will be chosen later. For all  $k \geq 1$ ,  $x_\varepsilon$  is a finite vector of size  $M(\varepsilon) + k$  and we have  $\|x - x_\varepsilon\|_1 \leq 2\varepsilon$ . Let us now show that we can chose  $k$  such that  $N_p(x_\varepsilon) \leq N_p(x)$  for all  $p \geq 1$ . In order to do this, consider the function  $\phi : (1, \infty) \rightarrow \mathbf{R}_+$

$$\phi(p) = \left[ \frac{S(\varepsilon)^p}{\sum_{i=M(\varepsilon)+1}^\infty x_i^p} \right]^{\frac{1}{p-1}}.$$

The function  $\phi$  takes finite values on  $(1, \infty)$  and  $\lim_{p \rightarrow \infty} \phi(p) = \frac{S(\varepsilon)}{x_{M(\varepsilon)+1}} < \infty$ . Moreover, as the Shannon entropy of  $x$  is finite, one can also show that  $\lim_{p \rightarrow 1^+} \phi(p) < \infty$ . Thus, the function  $\phi$  is bounded and we can choose  $k \in \mathbf{N}$  such that  $k \geq \phi(p)$  for all  $p \geq 1$ . This implies that

$$N_p(x_\varepsilon) - N_p(x) = k \left( \frac{S(\varepsilon)}{k} \right)^p - \sum_{i=M(\varepsilon)+1}^\infty x_i^p \leq 0.$$

In conclusion, we have found two finitely supported vectors  $x_\varepsilon$  and  $y_\varepsilon$  such that  $\|x - x_\varepsilon\|_1 \leq 2\varepsilon$ ,  $\|y - y_\varepsilon\|_1 \leq 2\varepsilon$  and  $N_p(x_\varepsilon) \leq N_p(y_\varepsilon)$  for all  $p \geq 1$ . To conclude, it suffices to apply Theorem 2 to  $x_\varepsilon$  and  $y_\varepsilon$ .  $\square$

REFERENCES

[1] G. Aubrun and I. Nechita, *Catalytic majorization and  $\ell_p$  norms*, preprint, available at [quant-ph/0702153](https://arxiv.org/abs/quant-ph/0702153). To appear in Comm. in Math. Physics.

- [2] S. Bandyopadhyay, V. Roychowdhury and U. Sen *Classification of nonasymptotic bipartite pure-state entanglement transformations*, Phys. Rev. A, 65 (2002), 052315.
- [3] R. Bhatia, *Matrix Analysis*. Graduate Texts in Mathematics, 169. Springer-Verlag, New York, 1997.
- [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, 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.
- [7] W. Feller, *An introduction to probability theory and its applications. Vol. II*. John Wiley & Sons, 1966.
- [8] Y. Feng, R. Duan and M. Ying, *Relation Between Catalyst-assisted Entanglement Transformation and Multiple-copy Transformation*. Phys. Rev. A (3) **74** (2006), 042312.
- [9] G. Grimmett and D. Stirzaker, *Probability and random processes*. Third edition. Oxford University Press, New York, 2001.
- [10] D. Jonathan and M. B. Plenio, *Entanglement-assisted local manipulation of pure quantum states*, Phys. Rev. Lett. **83** (1999), no. 17, 3566–3569.
- [11] G. Kuperberg, *The capacity of hybrid quantum memory*, IEEE Trans. Inform. Theory **49** (2003), 1465–1473.
- [12] 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.
- [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](http://www.qinfo.org/talks/2002/maj/book.ps)
- [15] M. Owari et al.,  *$\epsilon$ -convertibility of entangled states and extension of Schmidt rank in infinite-dimensional systems*, preprint, available at [quant-ph/0609167v3](http://quant-ph/0609167v3).
- [16] G. Pólya and G. Szegő, *Problems and Theorems in Analysis*, Springer-Verlag, Berlin-New York, 1978.