Extremal distributions under approximate majorization

Michał Horodecki,[1] Jonathan Oppenheim,[2] and Carlo Sparaciari[3]

[1]*Institute of Theoretical Physics and Astrophysics, National Quantum Information Centre, Faculty of Mathematics, Physics and Informatics, Univeristy of Gdańsk, Wita Stwosza 57, 80-308 Gdańsk, Poland*

[2]*University College of London, Department of Physics & Astronomy, London, WC1E 6BT and London Interdisciplinary Network for Quantum Science*

[3]*University College of London, Department of Physics & Astronomy, London, WC1E 6BT*

(Dated: May 24, 2018)

Although an input distribution may not majorize a target distribution, it may majorize a distribution which is close to the target. Here we consider a notion of approximate majorization. For any distribution, and given a distance δ, we find the approximate distributions which majorize (are majorized by) all other distributions within the distance δ. We call these the steepest and flattest approximation. This enables one to compute how close one can get to a given target distribution under a process governed by majorization. We show that the flattest and steepest approximations preserve ordering under majorization. Furthermore, we give a notion of majorization distance. This has applications ranging from thermodynamics, entanglement theory, and economics.

INTRODUCTION

The theory of majorization [1, 2] has important applications in topics as diverse as matrix theory, geometry, combinatorics, statistics, thermodynamics, coherence [3] and entanglement theory, uncertainty relations [4, 5], and economics. It defines a partial ordering over vectors of real numbers, as follow. For two vectors $a, b \in \mathbb{R}^k$, one define $a^{\downarrow}$, $b^{\downarrow}$ as the same vectors whose elements are non-increasingly ordered. Then, one says that *a weakly majorizes b from below*, $a \succ_w b$, iff

$$\sum_{i=1}^{l} a_i^{\downarrow} \geq \sum_{i=1}^{l} b_i^{\downarrow}, \quad \forall\, l = 1, \ldots, k. \tag{1}$$

When the sums of all k elements of the two vectors are equal, one says that *a majorizes b*, $a \succ b$. Hardy, Littlewood, and Polya [6] showed that $a \succ b$ iff $b = Da$, where D is a doubly-stochastic matrix (alternatively a probabilistic mixture of permutations), i.e. a stochastic matrix that preserves the uniform distribution. Here, we consider discrete probability distributions, represented by k-dimensional vectors with non-negative elements whose sum adds to 1. In many physical situations, a system can be described by a discrete probability distributions p, and the process the system is subjected to can be represented by a doubly-stochastic matrix D. Examples of these situations, especially relevant for the field of quantum information theory, concern entanglement theory, purity theory, quantum thermodynamics, and asymmetry theory.

For instance, in entanglement theory we have that a pure entangled state can be transformed into another, by means of Local Operations and Classical Communication [7], iff the vector of squared Schmidt coefficients of the final state can be mapped into the vector of squared Schmidt coefficients of the initial one by a doubly-stochastic matrix [2]. In purity theory, instead, a quantum state can be transformed into another, using Noisy Operations [8], iff the vector of the eigenvalues of the initial state can be mapped into the vector of eigenvalues of the final one by a doubly-stochastic matrix [9]. Similarly for quantum thermodynamics, a semi-classical state (diagonal in the energy eigenbasis) can be transformed into another semi-classical state, with Thermal Operations [10], iff the vector of the eigenvalues of the initial state can be mapped into the vector of eigenvalues of the final one by a Gibbs-preserving stochastic matrix, i.e. a stochastic matrix that preserves the Gibbs distribution instead of the uniform distribution. This latter result can be extended to any quantum state if a coherence reservoir [11] is used. In all these examples, except for quantum thermodynamics, majorization determines whether there exists a process mapping a quantum state into another. For quantum thermodynamics, the relevant order is given by β-majorization, a generalisation of majorization (see the concept of "d-majorization" in Ref. [1]).

In all the above-mentioned examples, exact transformations between discrete distributions p and q are considered. However, in physics we are very often more interested in the question of whether a process gets us "close" (under some physical distance) to the target distribution. Indeed, in the laboratory an experimentalist can only perform this kind on approximate transformations over a physical system. In the context of single-shot information theory, and of certain entropic functions, finding an approximation to the target distribution which minimises resources has been termed *smoothing* [12]. In this paper, we are interested in a different notion of smoothing which can be applied to finding the optimal approximation of the output or input state for the purposes of majorization. Namely, we consider the set of probability distributions which are δ-close (under a given distance) to a given distribution, and we find the extremal distributions of this set with respect to the majorization order. The distance we use throughout the paper is the one induced by the ℓ_1-norm, due to its clear operational meaning in terms of the degree of distinguishability between two probability distributions. For a real vector of k elements a, the ℓ_1-norm is defined as $\|a\| := \sum_{i=1}^{k} |a_i|$.

The same notion of approximate majorization we use, involving real vectors and probability distributions δ-close to a given one, can be found in the literature, see for instance Refs. [13, 14]. Furthermore, in Ref. [15] a generalisation of majorization called "relative submajorization" is given, and from it a more general notion of approximate majorization is derived. While the notion of approximate majorization we use in this paper is not novel, our contribution consists in the analysis of the properties of the extremal elements, with respect to majorization order, of the set of distributions δ-close to a give probability distribution. Finally, the reader should notice that the term approximate majorization has also been used in other contexts with a slightly different mathematical meaning. For example, in Ref. [16] the authors introduce a different notion of approximate majorization from ours, which is linked to the "fairness" of a resource-allocating protocol.

In the following we first introduce two smoothed versions of a given probability distribution, namely, the steepest and flattest δ-approximation of this distribution. Then, we show that the steepest approximation majorizes any probability distribution whose distance from the original distribution is less or equal than δ, while the flattest approximation is majorized by all these distributions. We also show that smoothing preserves monotonicity under majorization, for both the steepest and flattest approximation. Finally, we apply our findings to the analysis of the smooth version of Schur concave/convex functions. The present work[1] has recently found application in the context of thermodynamics [17, 18], and the extremal elements that we find have been independently rediscovered in the context of convex optimisation [19].

RESULTS

As we anticipated, our main tool consists in two specific approximations of a given probability distribution p, each of them δ-close to the original distribution. These approximations are (i) the *flattest δ-approximation* of p, and (ii) the *steepest δ-approximation* of p. In the following we will assume the elements of the probability distribution p to be non-increasingly ordered.

The steepest δ-approximation of p, which we denote by $\bar{p}^{(\delta)}$, is constructed as follows. If $\|p - \mathrm{e}_1\| \leq \delta$, where e_1 is the distribution whose first element is equal to 1, then we take $\bar{p}^{(\delta)} = \mathrm{e}_1$. Otherwise, we maximally increase the largest element of p, and we cut the tail. More precisely, we first add $\frac{\delta}{2}$ to the largest element of p (which is possible, since $\|p - \mathrm{e}_1\| > \delta$). This procedure returns a non-normalized distribution which we will denote by r, whose elements are defined as

$$r_i = \begin{cases} p_1 + \frac{\delta}{2} & \text{for} \;\; i = 1, \\ p_i & \text{for} \;\; i \neq 1. \end{cases} \tag{2}$$

Then we cut $\frac{\delta}{2}$ from the tail of this distribution. Formally, we take the integer $l^* \in \{1, \dots, k\}$ such that

$$\sum_{i=1}^{l^*} r_i \leq 1 \quad \text{and} \quad \sum_{i=1}^{l^*+1} r_i > 1, \tag{3}$$

and we define the steepest δ-approximation of p as

$$\bar{p}_i^{(\delta)} = \begin{cases} r_i & \text{for} \;\; i < l^* + 1, \\ 1 - x & \text{for} \;\; i = l^* + 1, \\ 0 & \text{for} \;\; i > l^* + 1. \end{cases} \tag{4}$$

where $x = \sum_{i=1}^{l^*} r_i$. In Fig. 1, the process of steepening a probability distribution p is shown, together with the resulting steepest δ-approximation $\bar{p}^{(\delta)}$. It is worth noting that the steepening process always maximally reduces the distance between the initial distribution p and the distribution e_1. Indeed, for $\|p - e_1\| \geq \delta$, it is easy to show that $\|\bar{p}^{(\delta)} - e_1\| = \|p - e_1\| - \delta$.

---

[1] whose initial draft circulated in 2013.

The flattest δ-approximation of p, denoted by $\underline{p}^{(\delta)}$, is constructed in the following way. If $\|p-\eta\| \le \delta$, where η is the uniform distribution, then we define $\underline{p}^{(\delta)} = \eta$. Otherwise, we proceed as follows. For a given $x, y \in [0,1]$, we define the following subsets,

$$I_x = \{i \in \{1,\ldots,k\} \mid p_i \ge x\}\,, \tag{5}$$
$$J_y = \{i \in \{1,\ldots,k\} \mid p_i \le y\}\,. \tag{6}$$

In the following we will drop the subscripts of these two sets, as the context should make them clear. Let us now introduce the functions

$$\epsilon(x) = \sum_{i\in I} (p_i - x), \tag{7}$$
$$\gamma(y) = \sum_{i\in J} (y - p_i). \tag{8}$$

Then, we choose $x^* \in [0,1]$ such that $\epsilon(x^*) = \frac{\delta}{2}$, and $y* \in [0,1]$ such that $\gamma(y^*) = \frac{\delta}{2}$. It is worth noting that, since $\|p-\eta\| > \delta$, both x^* and y^* exist and are unique, and moreover $x^* > y^*$. We can now define the flattest δ-approximation of p as

$$\underline{p}_i^{(\delta)} = \begin{cases} x^* & \text{for} \quad i \in I, \\ y^* & \text{for} \quad i \in J, \\ p_i & \text{else.} \end{cases} \tag{9}$$

In Fig. 1, the process of flattening a probability distribution p is shown, together with the resulting flattest δ-approximation $\underline{p}^{(\delta)}$. In analogous fashion with the steepening process, the flattening one always maximally reduces the distance between the initial distribution p and the uniform distribution η. Indeed, for $\|p-\eta\| \ge \delta$, it is easy to show that $\|\underline{p}^{(\delta)} - \eta\| = \|p-\eta\| - \delta$.

Remark. Let us note that the above constructions preserve the order of the elements, i.e., if the probability distribution p is non-increasingly ordered, then the same applies to both $\bar{p}^{(\delta)}$ and $\underline{p}^{(\delta)}$.

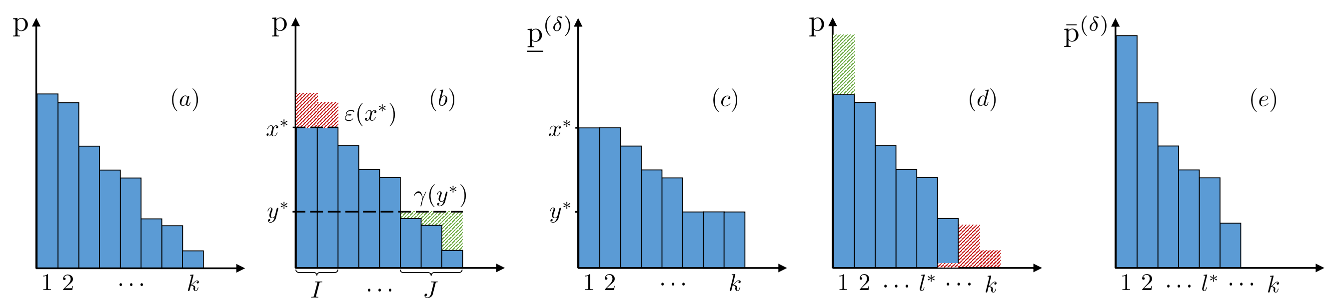

FIG. 1. The procedure of flattening and steepening the probability distribution p. The added portion is green, while the removed one is red. The two portions have the same area equal to $\frac{\delta}{2}$. (a) The original distribution p. (b) The procedure of flattening the probability distribution p. (c) The flattest δ-approximation of p, $\underline{p}^{(\delta)}$. (d) The procedure of steepening the distribution p. (e) The steepest δ-approximation of p, $\bar{p}^{(\delta)}$.

The following lemma, concerning the majorization properties of $\bar{p}^{(\delta)}$ and $\underline{p}^{(\delta)}$, singles out these two distributions among all the other distributions which are δ-close to p.

Lemma 1. *For a given probability distribution p of k elements, the distributions $\bar{p}^{(\delta)}$ and $\underline{p}^{(\delta)}$ are extremal δ-approximations of p in the sense of majorization order,*

(i) The steepest δ-approximation $\bar{p}^{(\delta)}$ of p majorizes any arbitrary distribution p' satisfying $\|p-p'\| \le \delta$,

$$\bar{p}^{(\delta)} \succ p'. \tag{10}$$

(ii) The flattest δ-approximation $\underline{p}^{(\delta)}$ of p is majorized by every arbitrary distribution p' satisfying $\|p-p'\| \le \delta$,

$$p' \succ \underline{p}^{(\delta)}. \tag{11}$$

Thus, the above lemma (whose proof can be found in the appendix) shows that the steepest and flattest δ-approximations of p are extremal points (with respect to the majorization order) of the set of all probability distributions which are δ-close to p.

Let us now consider the Lorenz curves of the steepest and flattest approximations. Lorenz curves are useful tools for studying majorization in a visual way, and they have been utilised in the context of the resource theory of purity and thermodynamics [9, 10]. For this reason, in the following we provide a description of the two extremal distributions we have found in terms of their Lorenz curves, see Fig. 2. Consider a probability distribution p of k elements, where we take the elements to be non-increasingly ordered. We define the Lorenz curve of p as the continuous curve L_p from the real interval $[0, k]$ to the real interval $[0, 1]$, such that $L_p(l) = \sum_{i=1}^{l} p_i$ for $l = 1, \ldots, k$, and $L_p(0) = 0$; for non-integer values of l, the curve is a straight line. In the following, we refer to the points of contact between these straight lines, $\left\{ \left(l, \sum_{i=1}^{l} p_i \right) \right\}_{l=1}^{k-1}$, as the elbows of the curve. Lorenz curves are concave functions, and can be used to study majorization as well as to extend its scope [15, 20]. Indeed, given two probability distributions p and q, such that $p \succ q$, we have that the Lorenz curve of p, L_p, always lays above the Lorenz curve of q, L_q.

The Lorenz curve of $\bar{p}^{(\delta)}$ is obtained from L_p by shifting all its elbows upward by $\frac{\delta}{2}$, until we reach the normalisation threshold equal to 1. Then, the curve is concluded by an horizontal segment. Formally, the Lorenz curve of the steepest δ-approximation is defined as

$$L_{\bar{p}^{(\delta)}}(l) = \sum_{i=1}^{l} \bar{p}_i^{(\delta)} = \begin{cases} \sum_{i=1}^{l} p_i + \frac{\delta}{2} & \text{for } \ l \leq l^*, \\ 1 & \text{for } \ l > l^*. \end{cases} \tag{12}$$

The Lorenz curve of $\underline{p}^{(\delta)}$ begins as a straight segment connecting the origin of the axes with the elbow $\left(l_I, \sum_{i \in I} p_i - \frac{\delta}{2} \right)$, where l_I is the maximum index of the set I. The final part of the curve is also a straight segment, connecting the elbow $\left(l_J - 1, 1 - \left(\sum_{i \in J} p_i + \frac{\delta}{2} \right) \right)$, where l_J is the minimum index of the set J, with the point $(k, 1)$. Finally, the other elbows of the curve are simply shifted downward by $\frac{\delta}{2}$. More formally, the Lorenz curve of the flattest δ-approximation is

$$L_{\underline{p}^{(\delta)}}(l) = \sum_{i=1}^{l} \underline{p}_i^{(\delta)} = \begin{cases} l \, x^* & \text{for } \ l \in I, \\ \sum_{i=1}^{l} p_i - \frac{\delta}{2} & \text{for } \ l \notin I \cup J, \\ 1 - (k - l) \, y^* & \text{for } \ l \in J. \end{cases} \tag{13}$$

Then, from Lemma 1 it follows that the Lorenz curve of any probability distribution p' (such that $\| p - p' \| \leq \delta$) entirely lies above the Lorenz curve of $\underline{p}^{(\delta)}$, and below the Lorenz curve of $\bar{p}^{(\delta)}$.

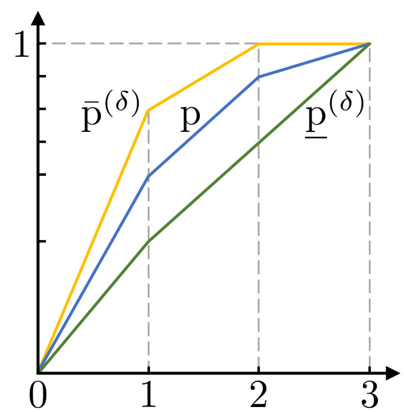

FIG. 2. The Lorenz curve of the probability distribution $p = (0.6, 0.3, 0.1)$ is shown in blue. For $\delta = 0.4$, we find that the steepest δ-approximation of p is $\bar{p}^{(\delta)} = (0.8, 0.2, 0)$, and its Lorenz curve is shown in orange. The flattest δ-approximation of p is $\underline{p}^{(\delta)} = (0.4, 0.3, 0.3)$, and its Lorenz curve is shown in green.

Now, let us come back to the analysis of the steepest and flattest approximations. An additional property of the processes of steepening and flattening a probability distribution consists in the fact that they preserve the majorization order.

Lemma 2. *Given two probability distributions of k elements, p and q, which satisfy $p \succ q$, we have*

$$\underline{p}^{(\delta)} \succ \underline{q}^{(\delta)}, \tag{14}$$

$$\bar{p}^{(\delta)} \succ \bar{q}^{(\delta)}. \tag{15}$$

The above lemma (whose proof can be found in the appendix), together with Lemma 1, has implications for the smooth versions of Schur convex/concave functions, i.e. functions from the space of probability distribution to $\mathbb{R}$ that are, respectively, monotonic increasing/decreasing in the majorization order. For any function f from the space of probability distribution to $\mathbb{R}$, let us define the following two smooth versions,

$$\bar{f}^{(\delta)}(p) = \max_{\|q-p\|\leq\delta} f(q), \tag{16}$$

$$\underline{f}^{(\delta)}(p) = \min_{\|q-p\|\leq\delta} f(q). \tag{17}$$

Then, we have the following proposition, which allows for explicitly computing the smoothed entropies for a given value of δ,

Proposition 3. *Let f be Schur-convex function, and p a probability distribution of k elements. Then*

$$\bar{f}^{(\delta)}(p) = f(\bar{p}^{(\delta)}), \quad \underline{f}^{(\delta)}(p) = f(\underline{p}^{(\delta)}). \tag{18}$$

If f is a Schur-concave function, then

$$\bar{f}^{(\delta)}(p) = f(\underline{p}^{(\delta)}), \quad \underline{f}^{(\delta)}(p) = f(\bar{p}^{(\delta)}). \tag{19}$$

Proof. For f being Schur-convex we have that $p \succ q$ implies

$$f(p) \geq f(q). \tag{20}$$

Thus the function preserves majorization order, hence on the set of δ-approximations of p it is maximal on $\bar{p}^{(\delta)}$ and minimal on $\underline{p}^{(\delta)}$. Thus from definition of $\bar{f}^{(\delta)}$ and $\underline{f}^{(\delta)}$ we obtain Eq. (18). An analogous argument applies when f is Schur-concave. □

The above result has applications in several resource theories, some of which were considered in the introduction, when approximate transformations are applied on a single copy of the system. For example, with Proposition 3 one can explicitly evaluate the smoothed versions of the min- and max-entropies (both of which are Schur-concave functions), used in the theory of purity. These Rényi entropies are associated, respectively, with the single-shot non-uniformity of formation and the single-shot distillable non-uniformity of a given state [9]. Similar quantities exists in thermodynamics, known as the smoothed single-shot work of formation and smoothed single-shot extractable work [10]. These quantities are associated with, respectively, the ∞-order and the 0-order Rényi-divergence from a given thermal state, that is, a Gibbs state with fixed inverse temperature β and given Hamiltonian H. It is in the thermodynamic setting that our result has found a first application, see Refs. [17, 18]. There, the authors use the property of Schur-concave functions shown in Proposition 3 to explicitly evaluate the smooth version of a family of Rényi-divergences associated with the generalised free energies introduced in Ref. [21]. In order to explicitly evaluate these quantities, the authors of Ref. [18] find the equivalent of the steepest and flattest distributions we have derived, but in the context of β-majorization. For this setting, they show that the flattest distribution can always be explicitly defined, while this is not the case for the steepest one.

From Lemma 2 and Proposition 3 it directly follows that

Corollary 4. *The smoothed versions of a Schur-convex function, $\bar{f}^{(\delta)}$ and $\underline{f}^{(\delta)}$, are monotonic under majorization order, i.e., given two probability distributions of k elements, p and q, where $p \succ q$, we have*

$$\bar{f}^{(\delta)}(p) \geq \bar{f}^{(\delta)}(q), \quad \underline{f}^{(\delta)}(p) \geq \underline{f}^{(\delta)}(q). \tag{21}$$

We close the paper with a result about the minimum distance δ which allows the δ-approximation of p to majorize q (and the δ-approximation of q to be majorized by p), when $p \not\succ q$.

Proposition 5. *Consider two probability distributions of k elements, p and q, such that $p \not\succ q$. Let δ_1 be the minimal δ such that $\bar{p}^{(\delta)} \succ q$, and δ_2 the minimal δ such that $p \succ \underline{q}^{(\delta)}$. Then we have*

$$\delta_1 = \delta_2 = \delta^* \equiv 2 \max_{l\in\{1,\ldots,k\}} \sum_{i=1}^{l} (q_i - p_i)\,. \tag{22}$$

The above proposition (whose proof can be found in the appendix) provides a measure of how much a distribution p majorizes a distribution q, in the sense that it tells us how much we have to distort q in order for p to majorize it, or equally how much we have to distort p in order for q to be majorized by it. By proving the statement of Proposition 5 we have shown that this measure is given by δ^*, which is the minimal distance that allows the steepest approximation of p to majorize q, and the flattest approximation of q to be majorized by p. Other measures of majorization distance include the mixing character/distance [22, 23], the information/work distance [21, 24, 25], and the maximum probability of transition [26].

CONCLUSIONS

In this paper, we study a notion of approximate majorization for discrete probability distributions. In particular, we consider the set of probability distributions which are δ-close, in the ℓ_1-distance, to a given distribution p, and we identify the extremal elements of this set under majorization order. These extremal distributions are the steepest and flattest δ-approximations of p, and in Lem. 1 we show that they majorize/are majorized by all the other δ-apporximations, respectively. We then show, in Lem. 2, that taking the steepest (or the flattest) δ-approximation of two probability distributions preserves their majorization order. In Prop. 3, we use the properties of the steepest and flattest approximations to explicitly compute the smoothed versions of Schur-convex (concave) functions. This is a non-trivial task, since to compute these quantities one needs to perform an optimisation over the set of δ-approximations of a given probability distribution. In Cor. 4 we show that the smoothed versions of Schur-convex (concave) functions are monotonic under majorization order. Finally, in Prop. 5 we provide a new measure for quantifying how much a probability distribution majorizes another one, using the notion of extremal distributions here introduced.

Appendix A: Steepest and flattest approximations

Lemma 1 (restatement). *For a given probability distribution p of k elements, the distributions $\bar{p}^{(\delta)}$ and $\underline{p}^{(\delta)}$ are extremal δ-approximations of p in the sense of majorization order,*

(i) The steepest δ-approximation $\bar{p}^{(\delta)}$ of p majorizes any arbitrary distribution p' satisfying $\|p - p'\| \leq \delta$,

$$\bar{p}^{(\delta)} \succ p'. \tag{A1}$$

(ii) The flattest δ-approximation $\underline{p}^{(\delta)}$ of p is majorized by every arbitrary distribution p' satisfying $\|p - p'\| \leq \delta$,

$$p' \succ \underline{p}^{(\delta)}. \tag{A2}$$

Proof. Let p be in non-increasing order. Let $\tilde{p}$ be an arbitrary δ-approximation of p, satisfying $\|\tilde{p} - p\| \leq \delta$. Then we can obtain $\tilde{p}$ as $\tilde{p}_i = p_i + \delta_i$, where $\sum_{i=1}^{k} |\delta_i| \leq \delta$. Notice that the obtained probability distribution might be not ordered, and therefore we define $\tilde{p}^{\downarrow}$ as the non-increasingly ordered probability distribution obtained from $\tilde{p}$. Also, notice that for all $m, l = 1, \ldots, k$, with $m \leq l$, we have that $\sum_{i=m}^{l} \delta_i \in \left[-\frac{\delta}{2}, \frac{\delta}{2}\right]$. We can now prove the Lemma.

Proof of part (i). We will exploit the distribution r, see Eq. (2), used in the definition of $\bar{p}^{(\delta)}$ (which is equal to p with the largest element increased by $\frac{\delta}{2}$, before the tail is cut by the same amount). Clearly, for any l

$$\sum_{i=1}^{l} \tilde{p}_i^{\downarrow} \leq \sum_{i=1}^{l} p_i + \frac{\delta}{2} = \sum_{i=1}^{l} r_i. \tag{A3}$$

Now, the procedure of cutting the tail only affects sums that are larger than 1, and makes them to be equal to 1. Since in $\tilde{p}$ all sums are no greater than 1, this does not affect the majorization conditions. Thus, we find that $\bar{p}^{(\delta)}$ majorizes $\tilde{p}$.

Proof of part (ii). Note that, over the interval I, the elements of $\underline{p}^{(\delta)}$ are all equal, see Eq. (9). The same is true for elements of $\underline{p}^{(\delta)}$ with indices in J. Those that are neither in I nor in J are the same as in the original distribution p. Since p is in non-increasing order, we have that $I = \{1, \ldots, l_I\}$, and $J = \{l_J, \ldots, k\}$. Let us first consider sums up to l elements for $l \leq l_I$. As we noticed, over the interval I the distribution $\underline{p}^{(\delta)}$ is flat, and its norm is equal to $\sum_{i=1}^{l_I} p_i - \frac{\delta}{2}$ due to the definition of $\epsilon(x^*)$, see Eq. (7). Let us now consider $\tilde{p}$, and its non-increasingly ordered version $\tilde{p}^{\downarrow}$. Then it is clear that $\sum_{i=1}^{l_I} \tilde{p}_i^{\downarrow} \geq \sum_{i=1}^{l_I} \tilde{p}_i \geq \sum_{i=1}^{l_I} p_i - \frac{\delta}{2}$, since the most we can diminish the first l_I largest elements of p is by subtracting all $\frac{\delta}{2}$ from them. Therefore, we have two distributions over I, one is flat, and the second has larger total sum. Since all distributions majorize the flat one, we get that for all $l \le l_I$,

$$\sum_{i=1}^{l} \tilde{p}_i^{\downarrow} \ge \sum_{i=1}^{l} \underline{p}_i^{(\delta)}. \tag{A4}$$

For $l \notin I \cup J$ we have that, according to its definition, $\underline{p}_i^{(\delta)} = p_i$, and therefore the conditions of Eq. (A4) are still satisfied. To deal with the set of indices J, we rewrite the related majorization inequalities which we need to prove as

$$\sum_{i=l}^{k} \underline{p}_i^{(\delta)} \ge \sum_{i=l}^{k} \tilde{p}_i^{\downarrow}, \quad \forall\, l > l_J. \tag{A5}$$

As a first step, we consider these sums for $l = l_J$. We have that,

$$\sum_{i=l_J}^{k} \tilde{p}_i^{\downarrow} \le \sum_{i=l_J}^{k} \tilde{p}_i = \sum_{i=l_J}^{k} (p_i + \delta_i) \le \sum_{i=l_J}^{k} p_i + \frac{\delta}{2} = \sum_{i=l_J}^{k} \underline{p}_i^{(\delta)}, \tag{A6}$$

where the last equality follows from the definitions of $\underline{p}^{(\delta)}$ and $\gamma(y^*)$, see Eq. (8). To prove Eq. (A5), note that for $l \ge l_J$ we have $\underline{p}_i^{(\delta)} = y^*$, where y^* is some positive number defined in the course of the construction of $\underline{p}^{(\delta)}$, see Eq. (9). For now, the value of y^* is not important, and what we need is that all $\underline{p}_i^{(\delta)}$ are constant for $i \in J$. Then, we have to prove that

$$\sum_{i=l}^{k} \tilde{p}_i^{\downarrow} \le (k + 1 - l)\, y^*, \tag{A7}$$

for $l > l_J$. But this is a consequence of the following easy-to-prove observation. Consider non-negative numbers $\{a_i\}_{i=1}^n$ put in increasing order. Let $\sum_{i=1}^n a_i \le n\,\lambda$, where $\lambda \ge 0$ is some constant. Then $\sum_{i=1}^l a_i \le l\,\lambda$. This observation ends the proof. □

Lemma 2 (restatement). *Given two probability distributions of k elements, p and q, which satisfy $p \succ q$, we have*

$$\underline{p}^{(\delta)} \succ \underline{q}^{(\delta)}, \tag{A8}$$

$$\bar{p}^{(\delta)} \succ \bar{q}^{(\delta)}. \tag{A9}$$

Proof. Let us first prove that the *steepest δ-approximation preserves the majorization order.* Let l^* and m^* be the indices defined as in Eq. (3) for, respectively, $\bar{p}^{(\delta)}$ and $\bar{q}^{(\delta)}$. Note that $l^* \le m^*$. Indeed, from the construction of $\bar{p}^{(\delta)}$ it follows that l^* is the largest l such that $\sum_{i=1}^l p_i + \frac{\delta}{2} \le 1$. Then, from the fact that $p \succ q$, we get that $l^* \le m^*$. Now, for $l \le l^*$ we have

$$\sum_{i=1}^{l} \bar{p}_i^{(\delta)} = \sum_{i=1}^{l} p_i + \frac{\delta}{2}, \quad \sum_{i=1}^{l} \bar{q}_i^{(\delta)} = \sum_{i=1}^{l} q_i + \frac{\delta}{2}, \tag{A10}$$

hence $p \succ q$ implies

$$\sum_{i=1}^{l} \bar{p}_i^{(\delta)} \ge \sum_{i=1}^{l} \bar{q}_i^{(\delta)}, \quad \forall\, l \le l^*. \tag{A11}$$

For $l > l^*$, instead, we have $\sum_{i=1}^l \bar{p}_i^{(\delta)} = 1$, hence the rest of the majorization conditions is automatically satisfied.

Now we will prove that the *flattest approximation preserves the majorization order.* Following the definition of Eq. (9), let us denote $x = x^*(p)$, $x' = x^*(q)$ and $y = y^*(p)$, $y' = y^*(q)$, where x^*, y^* are defined in the course of constructing the flattest approximation; x^* is the level at which the first largest elements are cut, and y^* is the level to which the smallest elements are enlarged. Similarly, let us denote $l_I \equiv l_I(p)$, $l'_I \equiv l_I(q)$ and $l_J \equiv l_J(p)$, $l'_J \equiv l_J(q)$. Recall that the interval $I = \{1, \ldots, l_I\}$ labels the elements that are cut (and have become equal to x^*), while the interval $J = \{l_J, \ldots, k\}$ labels the elements that are enlarged (and have become equal to y^*). In the following, we will frequently use the result of Lemma 6, that $x \ge x'$, and $y \le y'$. In fact, these inequalities are necessary conditions for majorization (indeed, x and x' are the largest elements, while y and y' are the smallest elements of $\underline{p}^{(\delta)}$ and $\underline{q}^{(\delta)}$, respectively).

We will divide the range $l = 1, \ldots, k$ into five intervals; (i) $[1, l_I]$, (ii) $[l_I + 1, l'_I]$, (iii) $[l_J, k]$, (iv) $[l'_J, l_J - 1]$, and (v) $(\max\{l_I, l'_I\}, \min\{l_J, l'_J\})$. Notice that the intervals (ii) and (iv) may be empty. For each interval we will prove that

$$\sum_{i=1}^{l} \underline{p}_i^{(\delta)} \geq \sum_{i=1}^{l} \underline{q}_i^{(\delta)}, \tag{A12}$$

for l belonging to the specific interval.

(i) $[1, l_I]$: This case is immediate. For all $i \leq l_I$, and independently of whether $l_I > l'_I$ or vice versa, we have

$$\underline{p}_i^{(\delta)} = x \geq x' \geq \underline{q}_i^{(\delta)}, \tag{A13}$$

where the first inequality follows from Lemma 6, and the second one from the fact that x' is largest element of $\underline{q}^{(\delta)}$. Notice that the second inequality is saturated for all $i \leq l'_I$. Summing up we obtain Eq. (A12) for $l \leq l_I$.

(ii) $[l_I + 1, l'_I]$: This case is trivial if the set is empty. When the set is not empty, instead, we start by considering the case of $l = l'_I$. In this situation we have

$$\sum_{i=1}^{l'_I} \underline{q}_i^{(\delta)} = \sum_{i=1}^{l'_I} q_i - \frac{\delta}{2}, \quad \sum_{i=1}^{l'_I} \underline{p}_i^{(\delta)} \geq \sum_{i=1}^{l'_I} p_i - \frac{\delta}{2}, \tag{A14}$$

which follows from the definition of $\epsilon(x)$ and $\epsilon(x')$, see Eq. (7), and the fact that $l_I \leq l'_I$. Thus, due to the fact that $p \succ q$, we obtain

$$\sum_{i=1}^{l'_I} \underline{p}_i^{(\delta)} \geq \sum_{i=1}^{l'_I} \underline{q}_i^{(\delta)}. \tag{A15}$$

Then, since $\underline{p}^{(\delta)}$ has no smaller norm than $\underline{q}^{(\delta)}$ on this interval, and moreover $\underline{q}^{(\delta)}$ is flat on the interval, we have that $\underline{p}^{(\delta)}$ (as well as any other distribution with no smaller norm) majorizes $\underline{q}^{(\delta)}$ on the interval. This proves Eq. (A12) for this interval.

(iii) $[l_J, k]$: In this interval we will prove equivalent relation to the one of Eq. (A12), namely

$$\sum_{i=l}^{k} \underline{p}_i^{(\delta)} \leq \sum_{i=l}^{k} \underline{q}_i^{(\delta)}, \tag{A16}$$

for $l > l_J$. For all $i \geq l_J$, and independently of whether $l_J < l'_J$ or vice versa, we have

$$\underline{p}_i^{(\delta)} = y \leq y' \leq \underline{q}_i^{(\delta)}, \tag{A17}$$

where the first inequality follows from Lemma 6, and the second one from the fact that y' is the smallest element of $\underline{q}^{(\delta)}$. Summing up, we obtain Eq. (A16) for $l \geq l_J$.

(iv) $[l'_J, l_J - 1]$: This case is trivial if the set is empty. When the set is not empty, instead, we start by considering the case of $l = l'_J$. We have that

$$\sum_{i=l'_J}^{k} \underline{q}_i^{(\delta)} = \sum_{i=l'_J}^{k} q_i + \frac{\delta}{2}, \quad \sum_{i=l'_J}^{k} \underline{p}_i^{(\delta)} \leq \sum_{i=l'_J}^{k} p_i + \frac{\delta}{2}, \tag{A18}$$

which follows from the definition of $\gamma(y)$ and $\gamma(y')$, see Eq. (8), and the fact that $l_J \geq l'_J$. Therefore, by $p \succ q$ we obtain that

$$\sum_{i=l'_J}^{k} \underline{p}_i^{(\delta)} \leq \sum_{i=l'_J}^{k} \underline{q}_i^{(\delta)}. \tag{A19}$$

Thus, on this interval $\underline{q}^{(\delta)}$ has no smaller norm than $\underline{p}^{(\delta)}$, and moreover $\underline{q}^{(\delta)}$ is flat. If the norms were equal to each other, $\underline{p}^{(\delta)}$ would majorize $\underline{q}^{(\delta)}$ on the interval, since any distribution majorizes the flat distribution. Therefore the conditions

$$\sum_{i=l}^{k} \underline{p}_i^{(\delta)} \leq \sum_{i=l}^{k} \underline{q}_i^{(\delta)}, \tag{A20}$$

would be satisfied for $l > l'_J$. Since norm of $\underline{q}^{(\delta)}$ may only be larger, the above inequalities still hold.

(v) $\max\{l_I, l'_I\} < l < \min\{l_J, l'_J\}$: Note that for l in such interval we have

$$\sum_{i=1}^{l} \underline{p}_i^{(\delta)} = \sum_{i=1}^{l} p_i - \frac{\delta}{2}, \quad \sum_{i=1}^{l} \underline{q}_i^{(\delta)} = \sum_{i=1}^{l} q_i - \frac{\delta}{2}, \tag{A21}$$

which follows from the definition of $\epsilon(x)$ and $\epsilon(x')$, see Eq. (7). So, by $p \succ q$ we obtain for the considered interval

$$\sum_{i=1}^{l} \underline{p}_i^{(\delta)} \geq \sum_{i=1}^{l} \underline{q}_i^{(\delta)}. \tag{A22}$$

This concludes the proof of the majorization relations for all l.

□

Lemma 6. *Let us consider two probability distributions of k elements, p and q, where $p \succ q$. We denote $x = x^*(p)$, $x' = x^*(q)$ and $y = y^*(p)$, $y' = y^*(q)$, where x^*, y^* are defined in the course of constructing the flattest approximation; x^* is the level at which first largest elements are cut, and y^* is the level to which the smallest elements are enlarged. Then*

$$x \geq x', \quad y \leq y'. \tag{A23}$$

Proof. Let us first denote $l_I \equiv l_I(p)$, $l'_I \equiv l_I(q)$, and $l_J \equiv l_J(p)$, $l'_J \equiv l_J(q)$. Recall that the interval $I = \{1, \ldots, l_I\}$ labels the elements that are cut (and have become equal to x^*), while the interval $J = \{l_J, \ldots, k\}$ labels the elements that are enlarged (and have become equal to y^*).

To prove that $x \geq x'$, notice first that for any $l = 1, \ldots, k$ we have

$$\sum_{i=1}^{l} p_i \leq l\, x + \frac{\delta}{2}, \tag{A24}$$

Indeed, for $l \leq l_I$ we have

$$\sum_{i=1}^{l} p_i = \sum_{i=1}^{l} (x + \delta_i) \leq l\, x + \frac{\delta}{2}, \tag{A25}$$

where $\delta_i \equiv p_i - \underline{p}_i^{(\delta)}$ satisfy $\sum_{i=1}^{l_I} \delta_i = \frac{\delta}{2}$ and $\sum_{i=1}^{l} \delta_i \geq 0$ for any l, which follows from the construction of $\underline{p}^{(\delta)}$. For $l > l_I$, instead, we have

$$\sum_{i=1}^{l} p_i = \sum_{i=1}^{l_I} p_i + \sum_{i=l_I+1}^{l} p_i = l_I\, x + \frac{\delta}{2} + \sum_{i=l_I+1}^{l} p_i \leq l_I\, x + \frac{\delta}{2} + (l - l_I)\, x = l\, x + \frac{\delta}{2}, \tag{A26}$$

where the inequality follows from the definition of the interval I, see Eq. (5). Now we use Eq. (A24) for $l = l'_I$, in conjunction with the majorization condition $p \succ q$, to get $x \geq x'$. We write

$$l'_I\, x + \frac{\delta}{2} \geq \sum_{i=1}^{l'_I} p_i \geq \sum_{i=1}^{l'_I} q_i = l'_I\, x' + \frac{\delta}{2}, \tag{A27}$$

which implies $x \geq x'$, since $l'_I \geq 1$ by definition.

The relation $y \le y'$ is proved in an analogous way. First, for any $l = 1, \dots, k$ we have

$$\sum_{i=l}^{k} p_i \ge (k - l + 1)\, y - \frac{\delta}{2}. \tag{A28}$$

Indeed, for $l \ge l_J$ we have

$$\sum_{i=l}^{k} p_i = \sum_{i=l}^{k} (y - \epsilon_i) \ge (k - l + 1)\, y - \frac{\delta}{2}, \tag{A29}$$

where $\epsilon_i \equiv \underline{p}_i^{(\delta)} - p_i$ satisfy $\sum_{i=l_J}^{k} \epsilon_i = \frac{\delta}{2}$ and $\sum_{i=l}^{k} \epsilon_i \ge 0$ for any l, which follows from the construction of $\underline{p}^{(\delta)}$. For $l < l_J$, instead, we have

$$\sum_{i=l}^{k} p_i = \sum_{i=l}^{l_J - 1} p_i + \sum_{i=l_J}^{k} p_i = \sum_{i=l}^{l_J - 1} p_i + (k - l_J + 1)\, y - \frac{\delta}{2} \ge (k - l + 1)\, y - \frac{\delta}{2}, \tag{A30}$$

where the inequality follows from the definition of the interval J, see Eq. (6). Now we use Eq. (A28) for $l = l'_J$, in conjunction with the majorization condition $p \succ q$, to show that $y \le y'$. We write

$$(k - l'_J + 1)\, y - \frac{\delta}{2} \le \sum_{i=l'_J}^{k} p_i \le \sum_{i=l'_J}^{k} q_i = (k - l'_J + 1)\, y' - \frac{\delta}{2}, \tag{A31}$$

which implies $y \le y'$. □

Proposition 5 (restatement). *Consider two probability distributions of k elements, p and q, such that $p \not\succ q$. Let δ_1 be the minimal δ such that $\bar{p}^{(\delta)} \succ q$, and δ_2 the minimal δ such that $p \succ \underline{q}^{(\delta)}$. Then we have*

$$\delta_1 = \delta_2 = \delta^* \equiv 2 \max_{l \in \{1,\dots,k\}} \sum_{i=1}^{l} (q_i - p_i)\,. \tag{A32}$$

Proof. Let us begin by showing that δ^* is the minimum distance δ such that $\bar{p}^{(\delta)} \succ q$. As a first step, we want to show that $\bar{p}^{(\delta^*)}$ majorizes q. To this aim, consider the non-normalised distribution r obtained from p by adding $\frac{\delta^*}{2}$ to its first element, Eq. (2). Then, for all $l \le l^*$, we have

$$\sum_{i=1}^{l} \bar{p}_i^{(\delta^*)} = \sum_{i=1}^{l} r_i = \sum_{i=1}^{l} p_i + \frac{\delta^*}{2} \ge \sum_{i=1}^{l} p_i + \left(\sum_{i=1}^{l} (q_i - p_i) \right) = \sum_{i=1}^{l} q_i, \tag{A33}$$

where the inequality follows from the definition of δ^*. When $l > l^*$, instead, we have that $\sum_{i=1}^{l} \bar{p}_i^{(\delta^*)} = 1$, and due to the normalisation condition on q we have that $\sum_{i=1}^{l} \bar{p}_i^{(\delta^*)} \ge \sum_{i=1}^{l} q_i$. Then, $\bar{p}^{(\delta^*)} \succ q$.

To show that δ^* is minimum, we consider $\bar{\delta} < \delta^*$, and we show that $\bar{p}^{(\bar{\delta})} \not\succ q$. In this case, it exists an $\bar{l}$ such that

$$\frac{\bar{\delta}}{2} < \sum_{i=1}^{\bar{l}} (q_i - p_i)\,. \tag{A34}$$

Then,

$$\sum_{i=1}^{\bar{l}} \bar{p}_i^{(\bar{\delta})} \le \sum_{i=1}^{\bar{l}} p_i + \frac{\bar{\delta}}{2} < \sum_{i=1}^{\bar{l}} p_i + \sum_{i=1}^{\bar{l}} (q_i - p_i) = \sum_{i=1}^{\bar{l}} q_i, \tag{A35}$$

where the first inequality is saturated for $\bar{l} \le l^*$, and the second inequality follows from Eq. (A34). Thus, we have that $\bar{p}^{(\bar{\delta})} \not\succ q$ for all $\bar{\delta} < \delta^*$.

Now, we show that δ^* is the distance δ such that $p \succ \underline{q}^{(\delta)}$. In particular, we initially want to show that $p \succ \underline{q}^{(\delta^*)}$. As a first step, we consider the interval $I = \{1, \dots, l_I\}$ in which $\underline{q}^{(\delta^*)}$ is flat, and all its elements are equal to x^*. In particular, we have that

$$\sum_{i=1}^{l_I} \underline{q}_i^{(\delta^*)} = \sum_{i=1}^{l_I} x_i^* = \sum_{i=1}^{l_I} q_i - \frac{\delta^*}{2} \le \sum_{i=1}^{l_I} q_i - \left(\sum_{i=1}^{l_I} (q_i - p_i) \right) = \sum_{i=1}^{l_I} p_i, \tag{A36}$$

where the second equality directly follows from Eq. (7) and from the fact that $\epsilon(x^*) = \frac{\delta^*}{2}$, while the inequality follows from the definition of δ^*. The above equation proves that, on the interval I, the norm of $\underline{q}^{(\delta^*)}$ is smaller or equal to the one of p. Then, since $\underline{q}^{(\delta^*)}$ is flat over the interval I, we have that p majorizes it, that is,

$$\sum_{i=1}^{l} p_i \geq \sum_{i=1}^{l} \underline{q}_i^{(\delta^*)}, \quad \forall\, l \leq l_I. \tag{A37}$$

We can now consider the interval in between I and J, where $J = \{l_J, \dots, k\}$. For all l in this interval, $l_I < l < l_J$, we have

$$\sum_{i=1}^{l} \underline{q}_i^{(\delta^*)} = \sum_{i=1}^{l_I} x^* + \sum_{i=l_I+1}^{l} q_i = \sum_{i=1}^{l} q_i - \frac{\delta^*}{2} \leq \sum_{i=1}^{l} q_i - \left(\sum_{i=1}^{l} (q_i - p_i) \right) = \sum_{i=1}^{l} p_i, \tag{A38}$$

which, again, follows from the definition of $\epsilon(x^*)$ and the one of δ^*. Thus, we find that

$$\sum_{i=1}^{l} p_i \geq \sum_{i=1}^{l} \underline{q}_i^{(\delta^*)}, \quad \forall\, l \in \{l_I + 1, \dots, l_J - 1\}\,. \tag{A39}$$

Finally, we consider the interval J. In this case, we will prove that

$$\sum_{i=l}^{k} p_i \leq \sum_{i=l}^{k} \underline{q}_i^{(\delta^*)}, \quad \forall\, l > l_J. \tag{A40}$$

To do so, let us consider the case $l = l_J$, where we have

$$\sum_{i=l_J}^{k} \underline{q}_i^{(\delta^*)} = \sum_{i=l_J}^{k} y^* = \sum_{i=l_J}^{k} q_i + \frac{\delta^*}{2} \geq \sum_{i=l_J}^{k} q_i + \left(\sum_{i=l_J}^{k} (p_i - q_i) \right) = \sum_{i=l_J}^{k} p_i, \tag{A41}$$

which follows from the definition of $\gamma(y^*)$ and the one of δ^*. Thus, we have that, over the interval J, $\underline{q}^{(\delta^*)}$ has bigger norm than p. Then, following the same argument used in the proof of Lemma 2 (iv), we have that since $\underline{q}^{(\delta^*)}$ is flat over J, then it is majorized by p, which proves Eq. (A40). Therefore, we have that $p \succ \underline{q}^{(\delta^*)}$.

To conclude the proof, we need to show that δ^* is minimum, that is, for all $\bar{\delta} < \delta^*$, we have that $p \not\succ \underline{q}^{(\bar{\delta})}$. When $\bar{\delta}$ is considered, we have seen that an $\bar{l}$ exists such that Eq. (A34) is satisfied. Then, for $l = \bar{l}$,

$$\sum_{i=1}^{\bar{l}} \underline{q}_i^{(\bar{\delta})} \geq \sum_{i=1}^{\bar{l}} q_i - \frac{\bar{\delta}}{2} > \sum_{i=1}^{\bar{l}} q_i - \sum_{i=1}^{\bar{l}} (q_i - p_i) = \sum_{i=1}^{\bar{l}} p_i, \tag{A42}$$

where the first inequality is saturated when $l_I \leq \bar{l} < l_J$, and the second one follows from Eq. (A34). Thus, we have that $p \not\succ \underline{q}^{(\bar{\delta})}$ for all $\bar{\delta} < \delta^*$. □

Acknowledgements We thank Fernando Brandão, Nelly Ng and Stephanie Wehner for discussions. MH is partially supported by a grant from the John Templeton Foundation. The opinions expressed in this publication are those of the authors and do not necessarily reflect the views of the John Templeton Foundation. JO thanks the Royal Society and an EPSRC Established Career Fellowship for their support. CS is supported by the EPSRC [grant number EP/L015242/1].

---

[1] Albert W. Marshall, Ingram Olkin, and Barry C. Arnold. *Inequalities: Theory of Majorization and Its Applications.* Springer Series in Statistics. Springer New York, New York, NY, 2011.

[2] Michael A. Nielsen and Guifré Vidal. Majorization and the Interconversion of Bipartite States. *Quantum Information and Computation*, 1(1):76–93, 2001.

[3] Huangjun Zhu, Zhihao Ma, Zhu Cao, Shao-Ming Fei, and Vlatko Vedral. Operational one-to-one mapping between coherence and entanglement measures. *arXiv:1704.01935 [quant-ph]*, 2017.

[4] Zbigniew Puchała, Łukasz Rudnicki, and Karol Życzkowski. Majorization entropic uncertainty relations. *Journal of Physics A: Mathematical and Theoretical*, 46(27):272002, 2013.

[5] Shmuel Friedland, Vlad Gheorghiu, and Gilad Gour. Universal Uncertainty Relations. *Physical Review Letters*, 111(23):230401, 2013.

[6] G. H. Hardy, J. E. Littlewood, and G. Pólya. *Inequalities.* Cambridge University Press, 1952.

[7] Charles H. Bennett, David P. DiVincenzo, John A. Smolin, and William K. Wootters. Mixed-state entanglement and quantum error correction. *Physical Review A*, 54(5):3824, 1996.

[8] Micha Horodecki, Pawe Horodecki, and Jonathan Oppenheim. Reversible transformations from pure to mixed states and the unique measure of information. *Physical Review A*, 67(6), 2003.

[9] Gilad Gour, Markus P. Mller, Varun Narasimhachar, Robert W. Spekkens, and Nicole Yunger Halpern. The resource theory of informational nonequilibrium in thermodynamics. *Physics Reports*, 2015.

[10] Micha Horodecki and Jonathan Oppenheim. Fundamental limitations for quantum and nanoscale thermodynamics. *Nature Communications*, 4, 2013.

[11] Johan Åberg. Catalytic Coherence. *Physical Review Letters*, 113(15):150402, 2014.

[12] Renato Renner. Security of Quantum Key Distribution. *arXiv:quant-ph/0512258*, 2005.

[13] Erik Nikolai Torgersen. Comparison of experiments when the parameter space is finite. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 16(3):219–249, 1970.

[14] Erik Nikolai Torgersen. *Comparison of Statistical Experiments.* Cambridge University Press, Cambridge, 1991.

[15] Joseph M. Renes. Relative submajorization and its use in quantum resource theories. *Journal of Mathematical Physics*, 57(12):122202, 2016.

[16] Rishi Bhargava, Ashish Goel, and Adam Meyerson. Using approximate majorization to characterize protocol fairness. In *SIGMETRICS '01 Proceedings of the 2001 ACM SIGMETRICS international conference on Measurement and modeling of computer systems*, pages 330–331, New York, NY, USA, 2001. ACM.

[17] Remco van der Meer. *The Properties of Thermodynamical Operations.* PhD thesis, Delft University of Technology, 2016.

[18] Remco van der Meer, Nelly Huei Ying Ng, and Stephanie Wehner. Smoothed generalized free energies for thermodynamics. *arXiv:1706.03193 [quant-ph]*, 2017.

[19] Eric P. Hanson and Nilanjana Datta. Maximum and minimum entropy states yielding local continuity bounds. *arXiv:1706.02212 [math-ph, physics:quant-ph]*, 2017.

[20] Francesco Buscemi and Gilad Gour. Quantum relative Lorenz curves. *Physical Review A*, 95(1):012110, 2017.

[21] Fernando Brandão, Michał Horodecki, Nelly Ng, Jonathan Oppenheim, and Stephanie Wehner. The second laws of quantum thermodynamics. *Proceedings of the National Academy of Sciences*, 112(11):3275–3279, 2015.

[22] Ernst Ruch and Alden Mead. The principle of increasing mixing character and some of its consequences. *Theoretical Chemistry Accounts: Theory, Computation, and Modeling (Theoretica Chimica Acta)*, 41(2):95–117, 1976.

[23] Ernst Ruch, Rudolf Schranner, and Thomas H. Seligman. The mixing distance. *Journal of Chemical Physics*, 69(1):386–392, 1978.

[24] Michał Horodecki and Jonathan Oppenheim. Fundamental limitations for quantum and nano thermodynamics. *Nature Communications*, 4(2059), 2013.

[25] Philippe Faist, Frédéric Dupuis, Jonathan Oppenheim, and Renato Renner. The minimal work cost of information processing. *Nature Communications*, 6:7669, 2015.

[26] Álvaro M. Alhambra, Jonathan Oppenheim, and Christopher Perry. Fluctuating States: What is the Probability of a Thermodynamical Transition? *Physical Review X*, 6(4):041016, 2016.