RÉNYI ENTROPY, SIGNED PROBABILITIES, AND THE QUBIT

ADAM BRANDENBURGER, PIERFRANCESCO LA MURA, STUART ZOBLE

JULY 28, 2019

ABSTRACT. The states of the qubit, the basic unit of quantum information, are 2x2 positive semi-definite Hermitian matrices with trace 1. We characterize these states in terms of an entropic uncertainty principle formulated on an eight-point phase space.

1. Introduction and Preliminaries

A basis for the space of 2×2 Hermitian matrices is given by $\{\boldsymbol{\sigma}_0, \boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2, \boldsymbol{\sigma}_3\}$, where $\boldsymbol{\sigma}_0 = \mathbf{I}$ is the 2×2 identity matrix and $\boldsymbol{\sigma}_1$, $\boldsymbol{\sigma}_2$, $\boldsymbol{\sigma}_3$ are the Pauli matrices

$$\boldsymbol{\sigma}_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \ \boldsymbol{\sigma}_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \ \boldsymbol{\sigma}_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

A 2×2 Hermitian matrix $\mathbf{M}$ satisfies $\mathrm{Tr}(\mathbf{M}) = 1$ if and only if

$$\mathbf{M} = \frac{1}{2}(\mathbf{I} + r_1\boldsymbol{\sigma}_1 + r_2\boldsymbol{\sigma}_2 + r_3\boldsymbol{\sigma}_3)$$

for some vector $\boldsymbol{r} = (r_1, r_2, r_3) \in \mathbb{R}^3$.

Definition 1. A 2×2 Hermitian matrix $\mathbf{M}$ with $\mathrm{Tr}(\mathbf{M}) = 1$ is called a *potential quantum state*. If, in addition, $\mathbf{M}$ is positive semi-definite, then $\mathbf{M}$ is a *quantum state*, or a state of the qubit. We also refer to the corresponding vectors $\boldsymbol{r}$ as potential quantum states and quantum states.

This is the model of the simplest quantum system, namely a two-level system such as the spin of a particle. Empirically, the experimenter can observe a property such as spin in three arbitrarily chosen mutually orthogonal directions x_1, x_2, and x_3. In each direction, the outcome of a measurement will be labeled $+1$ or -1. The expectation of the outcome in direction i is

$$\mathrm{Tr}(\mathbf{M}\boldsymbol{\sigma}_i) = r_i.$$

(See, e.g., Sakurai and Napolitano, 2011, p.181.) We want to associate an entropy with an empirical model. This step is not immediate because entropy is a measure of the uncertainty in a single probability distribution, and an empirical model contains three probability distributions (one for each direction). Our solution is to move to phase space, where an empirical model is represented by a single probability distribution. The *phase space* for a two-level system contains eight points,

$$\{+1,-1\}^3 = \{\boldsymbol{e}_n \mid n = 1, .., 8\},$$

where $\boldsymbol{e}_n(i) = (-1)^{n_i}$ for (n_1, n_2, n_3) the base 2 digits of $n-1$. Each point in phase space specifies the outcomes of each the three possible measurements. Non-deterministic responses to measurement are incorporated by specifying probabilities over the points in phase space. Let

$$Q = \{\boldsymbol{q} \in \mathbb{R}^8 \mid \Sigma_{i=1}^8 q_i = 1\}$$

denote the set of all signed probability distributions on phase space. That is, we do not require the probabilities to be positive, only that they sum to 1. We define a map ϕ from Q to the set of potential quantum states by

$$\phi(\boldsymbol{q}) = \frac{1}{2}(\mathbf{I} + r_1\boldsymbol{\sigma}_1 + r_2\boldsymbol{\sigma}_2 + r_3\boldsymbol{\sigma}_3),$$

where

$$r_i = \sum_{\{n \mid \boldsymbol{e}_n(i)=+1\}} q_n \times (+1) \;+ \sum_{\{n \mid \boldsymbol{e}_n(i)=-1\}} q_n \times (-1).$$

The map ϕ gives the correct transformation from phase space to the space of potential quantum states, in the sense of preserving the empirical probabilities. This map is linear and it will be helpful to fix some notation surrounding a matrix representation. Note we have folded the condition that $\boldsymbol{q}$ is a probability distribution in as the last equation in the definition of representation below.

Definition 2. Let $\mathbf{A}$ denote the matrix

$$\begin{pmatrix} 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 \\ 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix}.$$

For $\boldsymbol{r} \in \mathbb{R}^3$ define $\hat{\boldsymbol{r}} = (r_1, r_2, r_3, 1) \in \mathbb{R}^4$. For $\boldsymbol{q} \in \mathbb{R}^8$ and $\boldsymbol{r} \in \mathbb{R}^3$ we say $\boldsymbol{q}$ *represents* $\boldsymbol{r}$ if $\mathbf{A}\boldsymbol{q} = \hat{\boldsymbol{r}}$.

We are going to use phase space to formulate an entropic uncertainty principle as an axiom, and derive the quantum states this way. In particular, we will allow only those potential quantum states $\boldsymbol{r}$ for which there is a phase-space representation $\boldsymbol{q}$ whose entropy exceeds a lower bound. The non-classicality of the qubit becomes apparent because there are quantum states for which the only representations with entropy exceeding the bound are signed probability distributions. The use of negative probabilities on phase space to represent quantum systems goes back to the Wigner quasi-probability probability distribution (Wigner, 1932). The first task then is to choose a suitable definition of entropy for signed probabilities. We extend Rényi entropy (Rényi, 1961) to signed probabilities and then impose a smoothness condition that identifies a particular family of entropy functionals. Fix a finite set $X = \{x_1, ..., x_n\}$ together with an ordinary (unsigned) probability distribution $\boldsymbol{q}$ on X. Rényi entropy is the family of functionals

$$H_\alpha(\boldsymbol{q}) = -\frac{1}{\alpha - 1}\log_2(\sum_{i=1}^{n} q_i^\alpha),$$

where $0 < \alpha < \infty$ is a free parameter. (Shannon entropy is the special case $\alpha = 1$.) We can preserve the real-valuedness of entropy under signed probabilities by taking absolute values

$$H_\alpha(\boldsymbol{q}) = -\frac{1}{\alpha - 1}\log_2(\sum_{i=1}^{n} |q_i|^\alpha).$$

This formula can also be derived axiomatically (See Brandenburger and La Mura, 2019, who modify the original axioms for Rényi entropy in Rényi, 1961 and Dareczy, 1963.) Rényi entropy with ordinary probabilities is smooth in the interior of its domain. We now impose smoothness at 0, since this is no longer a boundary value of $\boldsymbol{q}$, namely, we require that

$$H_\alpha((q, 1-q)) \text{ is } C^\infty \text{ at } q = 0.$$

If α is not an integer let k be the least integer with $k > \alpha$. Then

$$\frac{\partial^k H_\alpha}{\partial q^k}((q, 1-q)) = \frac{f(q)}{g(q)}$$

where $f(0) \neq 0$ and $g(0) = 0$. Thus α must be an integer. If α is an odd integer then $H_\alpha((q, 1-q))$ is eventually not differentiable at 0. Thus Rényi entropy takes the following form under the smoothness assumption.

Definition 3. *Rényi entropy for signed probability distributions* is the family of functionals

$$H_{2k}(\boldsymbol{q}) = -\frac{1}{2k-1}\log_2(\sum_{i=1}^{n} {q_i}^{2k}) = -\frac{2k}{2k-1}\log_2(\|\boldsymbol{q}\|_{2k}),$$

where $k = 1, 2, \ldots$ is a free parameter.

Finally we give an example of a quantum state such that the only representatives with Rényi entropy satisfying the lower bound are signed probabilities. Consider the quantum state $(r_1, r_2, r_3) = (\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}})$. Set $k = 1$. The (unique) maximum 2-entropy representation is

$$\boldsymbol{q} = \frac{1}{8}(1+\sqrt{3}, 1+\frac{1}{\sqrt{3}}, 1+\frac{1}{\sqrt{3}}, 1-\frac{1}{\sqrt{3}}, 1+\frac{1}{\sqrt{3}}, 1-\frac{1}{\sqrt{3}}, 1-\frac{1}{\sqrt{3}}, 1-\sqrt{3}),$$

with negative final component. The 2-entropy of $\boldsymbol{q}$ is 2, which is the lower bound we impose below, so we cannot find a representation with all non-negative components with sufficiently high 2-entropy. In fact any state with $|r_1|+|r_2|+|r_3| > 1$ will have this property.

2. Main Theorem

We can now state an entropic uncertainty principle as an axiom on phase space. The axiom is inspired by the use of entropic uncertainty relations in quantum information. (See Coles et al., 2017 for a survey.)

> **Uncertainty Principle**: A potential quantum state $\boldsymbol{r}$ satisfies the uncertainty principle if for every k, there is a phase-space probability distribution $\boldsymbol{q}$ that represents $\boldsymbol{r}$ and satisfies $H_{2k}(\boldsymbol{q}) \geq 2$.

This says that we allow as potential quantum states only those states $\boldsymbol{r}$ containing a minimum amount of uncertainty, as measured by the entropy of a corresponding probability distribution $\boldsymbol{q}$ on phase space. Note that our uncertainty principle is a sequence of conditions, one for each k. This is because Rényi entropy itself is not a single functional but a sequence of functionals (indexed by k).

Theorem 1. *The potential quantum states satisfying the uncertainty principle are precisely the states of the qubit.*

Proof. We first show that the potential quantum states satisfying the uncertainty principle at $k = 1$ are the states of the qubit. Note that

$$H_2(\boldsymbol{q}) \geq 2 \text{ if and only if } \|\boldsymbol{q}\|_2^2 \leq \frac{1}{4}.$$

For a general $\boldsymbol{r}$, the representation $\boldsymbol{q}^*$ which maximizes 2-entropy is given by

$$\boldsymbol{q}^* = \mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^{-1}\hat{\boldsymbol{r}}.$$

Using the fact that $\mathbf{A}\mathbf{A}^T = 8\mathbf{I}$ we have

$$\|\boldsymbol{q}^*\|_2^2 = \hat{\boldsymbol{r}}^T(\mathbf{A}\mathbf{A}^T)^{-1}\hat{\boldsymbol{r}} = \frac{1}{8}\boldsymbol{r}^T\boldsymbol{r} + \frac{1}{8} \leq \frac{1}{4}$$

if and only if

$$r_1^2 + r_2^2 + r_3^2 \leq 1,$$

and the result follows since the matrix $\frac{1}{2}(\mathbf{I} + r_1\boldsymbol{\sigma}_1 + r_2\boldsymbol{\sigma}_2 + r_3\boldsymbol{\sigma}_3)$ is positive semi-definite if and only if $r_1^2 + r_2^2 + r_3^2 \leq 1$.

We now show that if a potential state $\boldsymbol{r}$ satisfies the uncertainty principle at $k = 1$ then it satisfies the uncertainty principle at all k. This is the main mathematical argument. Fix $k > 1$ and let $\boldsymbol{r} \in \mathbb{R}^3$ be a state of the qubit. Choose a $\boldsymbol{q}$ to maximize the $2k$-entropy of a representative of $\boldsymbol{r}$. We want to show $H_{2k}(\boldsymbol{q}) \geq 2$ which is equivalent to $\|\boldsymbol{q}\|_{2k} \leq (\frac{1}{2})^{\frac{2k-1}{k}}$. Observe that $\boldsymbol{q}$ solves the norm minimization problem

$$\min_{\boldsymbol{q} \in \mathbb{R}^8} \|\boldsymbol{q}\|_{2k}$$
$$\text{subject to } \mathbf{A}\boldsymbol{q} = \hat{\boldsymbol{r}}.$$

The dual problem is

$$\max_{x\in\mathbb{R}^4} \hat{\boldsymbol{r}}^T \boldsymbol{x}$$
$$\text{subject to } \|\mathbf{A}^T\boldsymbol{x}\|_{\frac{2k}{2k-1}} \leq 1.$$

(See Boyd and Vandenberghe, 2004, pp.221-222.) Note that $\|\cdot\|_{\frac{2k}{2k-1}}$ is the dual norm of $\|\cdot\|_{2k}$.) Strong duality holds so the values of the primal and dual problems are equal. Let $\boldsymbol{y}^1, \boldsymbol{y}^k$ be the maximizers of the dual problems for 2-entropy and $2k$-entropy respectively. Let

$$C_1 = \{\boldsymbol{x} \in \mathbb{R}^4 \mid \|\mathbf{A}^T\boldsymbol{x}\|_2 \leq 1\}$$

and

$$C_k = \{\boldsymbol{x} \in \mathbb{R}^4 \mid \|\mathbf{A}^T\boldsymbol{x}\|_{\frac{2k}{2k-1}} \leq 1\}.$$

Note that $C_k \subseteq C_1$ are both convex and, in fact, C_1 is the ball of radius $\frac{1}{\sqrt{8}}$. Let

$$\boldsymbol{z}^k = (\hat{\boldsymbol{r}}^T\boldsymbol{y}^k/\|\hat{\boldsymbol{r}}\|_2)\hat{\boldsymbol{r}}$$

be the projection of $\boldsymbol{y}^k$ onto $\hat{\boldsymbol{r}}$. Since $\hat{\boldsymbol{r}}^T\boldsymbol{y}^1 = \frac{\|\hat{\boldsymbol{r}}\|_2}{\sqrt{8}}\cos\theta$, where θ is the angle between them, we must have $\theta = 0$ and so

$$\boldsymbol{y}^1 = (\hat{\boldsymbol{r}}^T\boldsymbol{y}^1/\|\hat{\boldsymbol{r}}\|_2)\hat{\boldsymbol{r}}.$$

Since the values of the primal and dual problems are equal, these values are positive so $\frac{\|\boldsymbol{z}^k\|_2}{\|\boldsymbol{y}^1\|_2}$ is equal to the ratio of ratio of the value of the general k problem to the value of the $k = 1$ problem. By assumption

$$\hat{\boldsymbol{r}}^T\boldsymbol{y}^1 \leq \frac{1}{2},$$

so it suffices to show

$$\frac{\|\boldsymbol{z}^k\|_2}{\|\boldsymbol{y}^1\|_2} \leq (\frac{1}{2})^{\frac{k-1}{k}}.$$

We will bound this expression by a function that can be explicitly maximized. Note that for every nonzero vector $\boldsymbol{w}$ there are unique $\lambda < \nu$ such that

$$\|\mathbf{A}^T\nu\boldsymbol{w}\|_2 = 1$$

and

$$\|\mathbf{A}^T\lambda\boldsymbol{w}\|_{\frac{2k}{2k-1}} = 1.$$

This follows immediately from linearity, homogeneity, the fact that $\mathbf{A}$ has full rank, and the fact that $\frac{2k}{2k-1} < 2$. Now let

$$f(\boldsymbol{w}) = \frac{\|\mathbf{A}^T\boldsymbol{w}\|_2}{\|\mathbf{A}^T\boldsymbol{w}\|_{\frac{2k}{2k-1}}}.$$

By the previous observation and the fact that $f(\lambda\boldsymbol{w}) = f(\boldsymbol{w})$ for any nonzero scalar λ, we see that $f(\boldsymbol{w})$ is the ratio of the distance to the boundary of C_1 along the ray through $\boldsymbol{w}$ to the distance to the boundary of C_k. Let $\boldsymbol{w}^1 = \nu \boldsymbol{y}^k$ belong to the boundary of C_1. Figure 1 depicts the situation in the plane containing $\hat{\boldsymbol{r}}$ and $\boldsymbol{y}^k$.

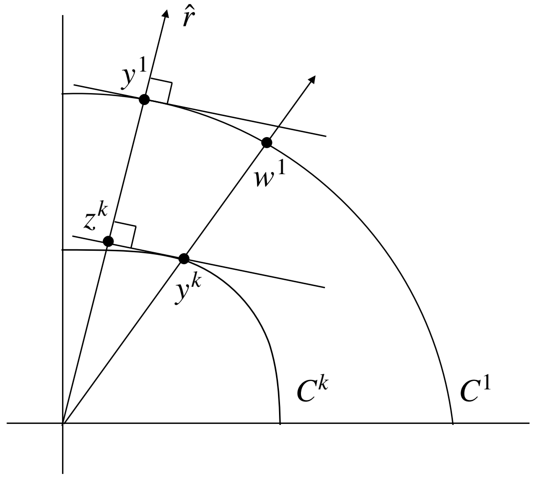

Figure 1

Claim 1. $\frac{\|\boldsymbol{z}^k\|_2}{\|\boldsymbol{y}^1\|_2}$ is bounded by a value of f.

Proof. We claim that

$$\frac{\|\boldsymbol{z}^k\|_2}{\|\boldsymbol{y}^1\|_2} \leq \frac{\|\boldsymbol{y}^k\|_2}{\|\boldsymbol{w}^1\|_2}.$$

Note that

$$\hat{\boldsymbol{r}}^T \boldsymbol{w}^1 \leq \hat{\boldsymbol{r}}^T \boldsymbol{y}^1,$$

so the length of the projection of $\boldsymbol{w}^1$ onto $\hat{\boldsymbol{r}}$ (call this vector $\boldsymbol{v}$) cannot exceed the length of $\boldsymbol{y}^1$. By similar triangles then

$$\frac{\|\boldsymbol{y}^k\|_2}{\|\boldsymbol{z}^k\|_2} = \frac{\|\boldsymbol{w}^1\|_2}{\|\boldsymbol{v}\|_2} \geq \frac{\|\boldsymbol{w}^1\|_2}{\|\boldsymbol{y}^1\|_2},$$

so $\frac{\|\boldsymbol{z}^k\|_2}{\|\boldsymbol{y}^1\|_2} \leq \frac{\|\boldsymbol{y}^k\|_2}{\|\boldsymbol{w}^1\|_2} = f(\boldsymbol{w}^1)$. □

We now prove that

$$\max\{f(\boldsymbol{w}) \mid \boldsymbol{w} \in \mathbb{R}^4\} = (\frac{1}{2})^{\frac{k-1}{k}},$$

which completes the proof of the theorem. Let $\boldsymbol{w} \in \mathbb{R}^4$ and $\boldsymbol{v} = \boldsymbol{w}\mathbf{A}$. Let $\boldsymbol{t} \in \mathbb{R}^8$ be defined by $t_i = v_i^{\frac{1}{2k-1}}$. Note that the critical points of f are the same as the critical points of

$$\frac{f^2(\boldsymbol{w})}{8} = \frac{\|\boldsymbol{v}\|_2^2}{8\|\boldsymbol{v}\|_{2k/2k-1}^2} = \frac{\boldsymbol{w}\mathbf{A}\mathbf{A}^T\boldsymbol{w}^T}{8\|\boldsymbol{v}\|_{2k/2k-1}^2} = \frac{\|\boldsymbol{w}\|_2^2}{\|\boldsymbol{v}\|_{2k/2k-1}^2},$$

which are the solutions of the system of first-order conditions

$$w_i = h(\boldsymbol{w}) r_i \boldsymbol{t}^T \quad i = 1, 2, 3, 4,$$

where

$$h(\boldsymbol{w}) = \frac{\|\boldsymbol{w}\|_2^2}{\|\boldsymbol{v}\|_{2k/2k-1}^{2k/2k-1}} > 0$$

and r_i is the i^{th} row of the matrix $\mathbf{A}$. It is helpful to write out the system with γ denoting $\frac{1}{2k-1}$ for readability:

$$w_1 = h(\boldsymbol{w})[(w_1+w_2+w_3+w_4)^\gamma - (-w_1+w_2+w_3+w_4)^\gamma + (w_1-w_2+w_3+w_4)^\gamma - (-w_1-w_2+w_3+w_4)^\gamma +$$
$$(w_1+w_2-w_3+w_4)^\gamma - (-w_1+w_2-w_3+w_4)^\gamma + (w_1-w_2-w_3+w_4)^\gamma - (-w_1-w_2-w_3+w_4)^\gamma],$$

$$w_2 = h(\boldsymbol{w})[(w_1+w_2+w_3+w_4)^\gamma + (-w_1+w_2+w_3+w_4)^\gamma - (w_1-w_2+w_3+w_4)^\gamma - (-w_1-w_2+w_3+w_4)^\gamma +$$
$$(w_1+w_2-w_3+w_4)^\gamma + (-w_1+w_2-w_3+w_4)^\gamma - (w_1-w_2-w_3+w_4)^\gamma - (-w_1-w_2-w_3+w_4)^\gamma],$$

$$w_3 = h(\boldsymbol{w})[(w_1+w_2+w_3+w_4)^\gamma + (-w_1+w_2+w_3+w_4)^\gamma + (w_1-w_2+w_3+w_4)^\gamma + (-w_1-w_2+w_3+w_4)^\gamma -$$
$$(w_1+w_2-w_3+w_4)^\gamma - (-w_1+w_2-w_3+w_4)^\gamma - (w_1-w_2-w_3+w_4)^\gamma - (-w_1-w_2-w_3+w_4)^\gamma],$$

$$w_4 = h(\boldsymbol{w})[(w_1+w_2+w_3+w_4)^\gamma + (-w_1+w_2+w_3+w_4)^\gamma + (w_1-w_2+w_3+w_4)^\gamma + (-w_1-w_2+w_3+w_4)^\gamma +$$
$$(w_1+w_2-w_3+w_4)^\gamma + (-w_1+w_2-w_3+w_4)^\gamma + (w_1-w_2-w_3+w_4)^\gamma + (-w_1-w_2-w_3+w_4)^\gamma].$$

Claim 2. The system $\boldsymbol{w} = h(\boldsymbol{w})\mathbf{A}\boldsymbol{t}^T$ has the following properties:

(1) If $\boldsymbol{w}$ is a solution then so is $\lambda\boldsymbol{w}$ for any $\lambda \neq 0$.
(2) If $\boldsymbol{w}$ is a solution then $\boldsymbol{v}$ is a solution, where $\boldsymbol{v}$ is obtained from $\boldsymbol{w}$ by permuting coordinates.

Proof. For (1) we have $h(\lambda\boldsymbol{w})\boldsymbol{t}^T(\lambda\boldsymbol{w}) = \frac{\lambda^2\lambda^{1/2k-1}}{\lambda^{2k/2k-1}}h(\boldsymbol{w})\boldsymbol{t}^T = \lambda\boldsymbol{w}$. For (2) we have

$$w_1 = h(w_4, w_2, w_3, w_1)r_4\boldsymbol{t}^T(w_4, w_2, w_3, w_1)$$

and

$$w_4 = h(w_4, w_2, w_3, w_1)r_1\boldsymbol{t}^T(w_4, w_2, w_3, w_1),$$

and similarly for w_2, w_3. □

Claim 3. Assume $w_4 \neq 0$. Let $i, j < 4$. Then

$$|w_i| = |w_j| \text{ or } w_i w_j = 0.$$

Proof. We may assume $w_4 > 0$. For N sufficiently large we have $\|\boldsymbol{w} - \boldsymbol{a}\|_2 < \|\boldsymbol{a}\|_2$ where $\boldsymbol{a} = (0, 0, 0, N)$. Thus the Taylor series expansion of $r_i\boldsymbol{t}^T$ at the point $\boldsymbol{a}$ converges at $\boldsymbol{w}$. We have

$$w_1 = 8h(\boldsymbol{w}) \sum_{\substack{\alpha_1 \in O \\ \alpha_2, \alpha_3 \in E \\ \alpha_4 \in \mathbb{N}}} \frac{(\boldsymbol{w}-\boldsymbol{a})^{\boldsymbol{\alpha}}}{\boldsymbol{\alpha}!} C(\sum_{i=1}^4 \alpha_i),$$

$$w_2 = 8h(\boldsymbol{w}) \sum_{\substack{\alpha_2\in O\\ \alpha_1,\alpha_3\in E\\ \alpha_4\in\mathbb{N}}} \frac{(\boldsymbol{w}-\boldsymbol{a})^{\boldsymbol{\alpha}}}{\boldsymbol{\alpha}!} C(\sum_{i=1}^{4}\alpha_i),$$

$$w_3 = 8h(\boldsymbol{w}) \sum_{\substack{\alpha_3\in O\\ \alpha_1,\alpha_2\in E\\ \alpha_4\in\mathbb{N}}} \frac{(\boldsymbol{w}-\boldsymbol{a})^{\boldsymbol{\alpha}}}{\boldsymbol{\alpha}!} C(\sum_{i=1}^{4}\alpha_i),$$

$$w_4 = 8h(\boldsymbol{w}) \sum_{\substack{\alpha_1,\alpha_2,\alpha_3\in E\\ \alpha_4\in\mathbb{N}}} \frac{(\boldsymbol{w}-\boldsymbol{a})^{\boldsymbol{\alpha}}}{\boldsymbol{\alpha}!} C(\sum_{i=1}^{4}\alpha_i),$$

where $\boldsymbol{\alpha}\in\mathbb{N}^4$ is a multi-index, $\mathbb{N} = E\cup O = \{0,2,4,...\}\cup\{1,3,5,...\}$,

$$\boldsymbol{\alpha}! = \alpha_1!\alpha_2!\alpha_3!\alpha_4!,$$

$$(\boldsymbol{w}-\boldsymbol{a})^{\boldsymbol{\alpha}} = w_1^{\alpha_1}w_2^{\alpha_2}w_3^{\alpha_3}(w_4-N)^{\alpha_4},$$

and C is defined by

$$C(1) = \frac{1}{(2k-1)N^{\frac{2k}{2k-1}}},$$

$$C(n) = \frac{(-1)^{n-1}\prod_{j=1}^{n-1}(j(2k-1)-1)}{(2k-1)^n N^{\frac{n(2k-1)-1}{2k-1}}} \quad \text{for } n>1.$$

Note that $C(\sum_{i=1}^4\alpha_i) > 0$ if and only if $\sum_{i=1}^4\alpha_i\in O$. Assume $w_1, w_2\neq 0$. We have

$$w_1 \sum_{\substack{\alpha_2\in O\\ \alpha_1,\alpha_3\in E\\ \alpha_4\in\mathbb{N}}} \frac{(\boldsymbol{w}-\boldsymbol{a})^{\boldsymbol{\alpha}}}{\boldsymbol{\alpha}!} C(\sum_{i=1}^{4}\alpha_i) = w_2 \sum_{\substack{\alpha_1\in O\\ \alpha_2,\alpha_3\in E\\ \alpha_4\in\mathbb{N}}} \frac{(\boldsymbol{w}-\boldsymbol{a})^{\boldsymbol{\alpha}}}{\boldsymbol{\alpha}!} C(\sum_{i=1}^{4}\alpha_i),$$

equivalently

$$\sum_{\substack{\alpha_2\in O\\ \alpha_1,\alpha_3\in E\\ \alpha_4\in\mathbb{N}}} \frac{(\boldsymbol{w}-\boldsymbol{a})^{\boldsymbol{\alpha}+(1,0,0,0)}}{\boldsymbol{\alpha}!} C(\sum_{i=1}^{4}\alpha_i) = \sum_{\substack{\alpha_1\in O\\ \alpha_2,\alpha_3\in E\\ \alpha_4\in\mathbb{N}}} \frac{(\boldsymbol{w}-\boldsymbol{a})^{\boldsymbol{\alpha}+(0,1,0,0)}}{\boldsymbol{\alpha}!} C(\sum_{i=1}^{4}\alpha_i).$$

Re-indexing we have

$$\sum_{\substack{\alpha_1,\alpha_2\in O\\ \alpha_3\in E\\ \alpha_4\in\mathbb{N}}} \alpha_1\frac{(\boldsymbol{w}-\boldsymbol{a})^{\boldsymbol{\alpha}}}{\boldsymbol{\alpha}!} C(\sum_{i=1}^{4}\alpha_i-1) = \sum_{\substack{\alpha_1,\alpha_2\in O\\ \alpha_3\in E\\ \alpha_4\in\mathbb{N}}} \alpha_2\frac{(\boldsymbol{w}-\boldsymbol{a})^{\boldsymbol{\alpha}}}{\boldsymbol{\alpha}!} C(\sum_{i=1}^{4}\alpha_i-1).$$

Collecting terms we have

$$\sum_{\substack{\alpha_1<\alpha_2\\ \alpha_1,\alpha_2\in O\\ \alpha_3\in E\\ \alpha_4\in\mathbb{N}}} (\alpha_1-\alpha_2)(w_1^{\alpha_1+1}w_2^{\alpha_2+1}-w_1^{\alpha_2+1}w_2^{\alpha_1+1})w_3^{\alpha_3}(w_4-N)^{\alpha_4}\frac{C(\sum_{i=1}^4\alpha_i-1)}{\boldsymbol{\alpha}!}=0,$$

equivalently

$$\sum_{\substack{\alpha_1<\alpha_2\\ \alpha_1,\alpha_2\in O\\ \alpha_3\in E\\ \alpha_4\in\mathbb{N}}} (\alpha_1-\alpha_2)(w_1^{\alpha_1+1}w_2^{\alpha_2+1})(1-(\frac{w_1}{w_2})^{\alpha_2-\alpha_1})w_3^{\alpha_3}(w_4-N)^{\alpha_4}\frac{C(\sum_{i=1}^4\alpha_i-1)}{\boldsymbol{\alpha}!}=0.$$

The key point is that

$$(\alpha_1-\alpha_2)(w_1^{\alpha_1+1}w_2^{\alpha_2+1})w_3^{\alpha_3}(w_4-N)^{\alpha_4}\frac{C(\sum_{i=1}^4\alpha_i-1)}{\boldsymbol{\alpha}!}$$

is always negative. Thus since $\alpha_2-\alpha_1\in E$ we conclude that $|w_1|=|w_2|$. □

Claim 4. If $w_i, w_j \neq 0$ then $|w_i|=|w_j|$.

Proof. By Claim 2 and Claim 3 we may assume that $w_1, w_4\neq 0$ and $w_2, w_3 = 0$. We may further assume that $w_4=1$. Thus the equation for w_1 becomes

$$w_1=\frac{(w_1+1)^{1/2k-1}-(-w_1+1)^{1/2k-1}}{(w_1+1)^{1/2k-1}+(-w_1+1)^{1/2k-1}}$$

so

$$(w_1+1)(-w_1+1)^{1/2k-1}=(-w_1+1)(w_1+1)^{1/2k-1},$$

from which we conclude that $w_1\in\{-1,1\}$ as desired. □

We have thus shown that for every $i,j\leq 4$ either $|w_i|=|w_j|$ or $w_iw_j=0$ so we need only consider critical points with

$$w_i\in\{0,1,-1\}$$

for each $i=1,...,4$. It is easy to check that the maximum of the original function f occurs when exactly two of the weights are 0 and this maximum value is $(\frac{1}{2})^{\frac{k-1}{k}}$, completing the proof of Theorem 1.

Acknowledgements

We are grateful to Samson Abramsky, Časlav Brukner, Matthew Deutsch, Christopher Fuchs, Leslie Greengard, Michael J.W. Hall, Jerry Keisler, Elliot Lipnowski, David Pine, Rui Soares Barbosa, Yuri Tschinkel, and Noson Yanofsky for valuable input. Financial support from the NYU Stern School of Business, J.P. Valles, NYU Abu Dhabi, NYU Shanghai, and the HHL - Leipzig Graduate School of Management is gratefully acknowledged.

Stern School of Business, Tandon School of Engineering, NYU Shanghai, New York University, New York, NY 10012, U.S.A., adam.brandenburger@stern.nyu.edu, http://www.adambrandenburger.com

HHL - Leipzig Graduate School of Management, 04109 Leipzig, Germany, plamura@hhl.de

Signal Fox, Princeton, NJ 08542, U.S.A., szoble@gmail.com, https://signalfox.com

References

Boyd, S., and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, 2004.

Brandenburger, A., and P. La Mura, "Axioms for Rényi Entropy with Signed Measures," 2019, at http://www.adambrandenburger.com.

Coles, P., M. Beta, M. Tomamichel, and S. Wehner, "Entropic Uncertainty Relations and 13 Their Applications," *Reviews of Modern Physics*, 89, 2017, 015002.

Daróczy, Z., "Über die gemeinsame Charakterisierung der zu den nicht vollständigen Verteilungen gehörigen Entropien von Shannon und von Rényi," *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 1, 1963, 381-388.

Rényi, A., "On Measures of Information and Entropy," in Neymann, J., (ed.), *Proceedings of the 4th Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press, 1961, 547-561.

Sakurai, J.J., and J. Napolitano, *Modern Quantum Mechanics*, Addison-Wesley, 2nd edition, 2011.

Shannon, C., "A Mathematical Theory of Communication," *Bell System Technical Journal*, 27, 1948, 379-423 and 623-656.

Wigner, E., "On the Quantum Correction For Thermodynamic Equilibrium," *Physical Review*, 40, 1932, 749-759, at at https://doi:10.1103/PhysRev.40.749.