Quantum birthday problems: Geometrical aspects of quantum random coding

Akio Fujiwara[*]
Department of Mathematics, Osaka University
Toyonaka, Osaka 560-0043, Japan

Abstract

This paper explores asymptotics of randomly generated vectors on extended Hilbert spaces. In particular, we are interested to know how 'orthogonal' these vectors are. We investigate two types of asymptotic orthogonality, the weak orthogonality and the strong orthogonality, that are regarded as quantum analogues of the classical birthday problem and its variant. As regards the weak orthogonality, a new characterization of the von Neumann entropy is derived, and a mechanism behind the noiseless quantum channel coding theorem is clarified. As regards the strong orthogonality, on the other hand, a characterization of the quantum Rényi entropy of degree 2 is derived.

Index Terms: asymptotic orthogonality, birthday problem, quantum channel coding, quantum information theory, random vector, Rényi entropy, von Neumann entropy

1 Introduction

Let $\mathcal{H}$ be a Hilbert space and let p be a probability measure on $\mathcal{H}$ the support of which being a countable set $\mathcal{X}$ of unit vectors. We assume that $|\mathcal{X}| \geq 2$ and that vectors in $\mathcal{X}$ are not parallel with each other. Associated with the probability measure p is the density operator

$$\rho := \sum_{\phi \in \mathcal{X}} p(\phi)\, |\phi\rangle\langle\phi|, \tag{1}$$

where, and in what follows, we use the Dirac notation. Let $\{X_k(i)\}_{ki}$ be $\mathcal{X}$-valued random variables i.i.d. with respect to p, and let $\{L_n\}_n$ be an increasing sequence of natural numbers. For each $n \in \boldsymbol{N}$, we define L_n random vectors $\{\Psi^{(n)}(i)\}_{1 \leq i \leq L_n}$ on $\mathcal{H}^{\otimes n}$ by

$$\Psi^{(n)}(i) := X_1(i) \otimes X_2(i) \otimes \cdots \otimes X_n(i), \qquad (i = 1, ..., L_n).$$

[*] email: fujiwara@math.wani.osaka-u.ac.jp

We denote the inner product of the ith and jth vectors by

$$g_{ij}^{(n)} := \langle \Psi^{(n)}(i) | \Psi^{(n)}(j) \rangle = \prod_{k=1}^{n} \langle X_k(i) | X_k(j) \rangle. \tag{2}$$

Note that $g_{ii}^{(n)} = 1$ for all n and $1 \leq i \leq L_n$, and that, for fixed i and j ($i \neq j$), $g_{ij}^{(n)}$ converges to 0 almost surely as $n \to \infty$. Thus it is natural to inquire how 'orthogonal' those random vectors are. For later convenience, we denote the ordered list of the L_n random vectors by $\mathcal{C}^{(n)}$.

To motivate our problem, let us consider the special case when $\mathcal{X}$ forms an orthonormal system. In this case, $g_{ij}^{(n)} = 0$ if $X_k(i) \neq X_k(j)$ for some k, and $g_{ij}^{(n)} = 1$ otherwise. Put differently, the Gram matrix

$$G^{(n)} := \begin{bmatrix} g_{11}^{(n)} & \cdots & g_{1L_n}^{(n)} \\ \vdots & & \vdots \\ g_{L_n 1}^{(n)} & \cdots & g_{L_n L_n}^{(n)} \end{bmatrix}$$

gives an yes/no table indicating whether the ith n-tuple $(X_1(i), ..., X_n(i))$ and the jth n-tuple $(X_1(j), ..., X_n(j))$ are identical ($g_{ij}^{(n)} = 1$) or not ($g_{ij}^{(n)} = 0$). As a consequence, orthogonality problems for the random vectors are reduced to combinatorial ones when $\mathcal{X}$ is orthonormal.

In his paper [1], Rényi posed several combinatorial problems that can be regarded as asymptotic versions of the classical birthday problem (cf. [2]) and its variants, and characterized classical entropies. From among them, let us recast two problems in terms of orthogonalities of random vectors. Let $\mathcal{X}$ be orthonormal. We say that $\mathcal{C}^{(n)}$ satisfies *weak orthogonality* condition with respect to the ith vector if the event

$$E_i^{(n)} := \{\text{The } i\text{th vector } \Psi^{(n)}(i) \text{ is orthogonal to the other vectors in } \mathcal{C}^{(n)}\}$$

occurs, and that $\mathcal{C}^{(n)}$ satisfies *strong orthogonality* condition if the event

$$F^{(n)} := \{\text{The vectors in } \mathcal{C}^{(n)} \text{ are mutually orthogonal}\}$$

occurs. We are interested to know how fast can L_n be increased under the condition that the probability $P(E_i^{(n)})$ for some (then any) i, or $P(F^{(n)})$, tends to 1 as $n \to \infty$. Let $C_w(p)$ [resp. $C_s(p)$] be the supremum of $\limsup_{n\to\infty} \log L_n / n$ over all sequences $\{L_n\}_n$ that satisfy $P(E_i^{(n)}) \to 1$ [resp. $P(F^{(n)}) \to 1$]. We may call $C_w(p)$ [resp. $C_s(p)$] the *weak* [resp. *strong*] *orthogonality capacity* of the probability measure p. Since we are now dealing with a probability measure p that has an orthonormal support $\mathcal{X}$, the problems are essentially combinatorial, and it is not too difficult to show that $C_w(p) = H(p)$ and $C_s(p) = H_2(p)/2$, where $H(p)$ and $H_2(p)$ are the Shannon entropy and the Rényi entropy [1] of degree 2 with respect to the probability measure p.

Let us now return to the general case when $\mathcal{X}$ is not necessarily orthonormal. Although the vectors in $\mathcal{C}^{(n)}$ may not be strictly orthogonal in this case, it would be quite possible that they are 'almost' orthogonal for sufficiently large n. It is therefore expected that quantum entropies might be characterized via asymptotic properties of the set $\mathcal{C}^{(n)}$ of random vectors (as Rényi did for classical entropies via combinatorics). The purpose of this paper is to extend the notions of weak and strong orthogonality of random vectors to a general probability measure p, and to determine the corresponding capacities. In fact, with proper definitions of 'asymptotic' orthogonalities, it is shown in Theorems 1 and 4 that the weak orthogonality capacity $C_w(p)$ is given by the von Neumann entropy

$$H(\rho) := -\mathrm{Tr}\,\rho\log\rho,$$

and the strong orthogonality capacity $C_s(p)$ is given by half the quantum Rényi entropy of degree 2

$$H_2(\rho) := -\log\mathrm{Tr}\,\rho^2,$$

where the probability measure p and the density operator ρ are connected by Eq. (1). Since these results obviously generalize the above mentioned classical characterizations by Rényi, our problems may be called *quantum birthday problems*. It should be emphasized that each capacity depends only on the density operator ρ, so that, for those probability measures p and q which give the same density operator, $C_w(p) = C_w(q)$ and $C_s(p) = C_s(q)$ hold.

The orthogonality of vectors in $\mathcal{C}^{(n)}$ is closely related to their 'distinguishability' in quantum measurement theory. Let a physical system of interest be represented by the Hilbert space $\mathcal{H}^{\otimes n}$, and let a unit vector in $\mathcal{H}^{\otimes n}$ correspond to a quantum pure state. When $\mathcal{X}$ is orthonormal, any two vectors in $\mathcal{C}^{(n)}$ are either orthogonal or identical, so that there is a quantum mechanical measurement on $\mathcal{H}^{\otimes n}$ that distinguishes distinct vectors in $\mathcal{C}^{(n)}$ with probability one. In this sense, strict orthogonality implies strict distinguishability by a certain measurement. As a matter of fact, this corresponds to the classical situation: unlimited distinguishability for distinct objects is precisely the central dogma of the classical theory, and one can restore Rényi's original problems by replacing the word(s) 'orthogonal (to)' with 'distinct (from)' in the above definitions of the events $E_i^{(n)}$ and $F^{(n)}$. When $\mathcal{X}$ is not orthonormal, on the other hand, distinct vectors in $\mathcal{C}^{(n)}$ are not necessarily orthogonal, so that they are not always strictly distinguishable in the sense of quantum mechanics. We therefore have to deal with, so to say, 'asymptotic distinguishability' of random vectors. In fact, we will clarify a close connection between asymptotic orthogonality and the noiseless quantum channel coding problem.

2 Weak orthogonality

Let us start with a preliminary consideration as to what is the proper extension of the notion of weak orthogonality to the case when $\mathcal{X}$ is not necessarily orthonormal. If the ith vector in $\mathcal{C}^{(n)}$ is 'almost' orthogonal to the other vectors in $\mathcal{C}^{(n)}$, then the inner products $g_{ij}^{(n)}$, $(j = 1, ..., L_n,\ j \neq i)$, must be all sufficiently small. Thus the proper extension of weak orthogonality might be such that the random variables $g_{ij}^{(n)}$ converge to 0 simultaneously for all $j(\neq i)$ as $n \to \infty$ in a certain mode of convergence. For example, the condition that

$$Y_i^{(n)} := \sum_{j(\neq i)}^{L_n} |g_{ij}^{(n)}|^2 \to 0 \quad \text{in probability} \tag{3}$$

might be a candidate. However, in anticiption of a characterization of the von Neumann entropy as well as a relationship with the quantum channel coding problem, we adopt a slightly different approach. (In fact, it is shown in Appendix A that the condition (3) does not characterize the von Neumann entropy.)

For each n, let $\mathcal{L}^{(n)}$ be a subspace of $\mathcal{H}^{\otimes n}$ and let $\Pi_{\mathcal{L}^{(n)}}$ be the projection operator onto the subspace $\mathcal{L}^{(n)}$. Given a pair $(\mathcal{C}^{(n)}, \mathcal{L}^{(n)})$, let us denote the inner product of the projected ith and jth vectors by

$$\hat{g}_{ij}^{(n)} := \langle \Pi_{\mathcal{L}^{(n)}} \Psi^{(n)}(i) | \Pi_{\mathcal{L}^{(n)}} \Psi^{(n)}(j) \rangle,$$

and define random variables

$$\hat{Y}_i^{(n)} := \sum_{j(\neq i)}^{L_n} |\hat{g}_{ij}^{(n)}|^2.$$

We say that a sequence $\{\mathcal{C}^{(n)}\}_n$ satisfies *asymptotic weak orthogonality* condition if there is a sequence $\{\mathcal{L}^{(n)}\}_n$ of subspaces such that the following conditions are satisfied: for all i,

(i) $\hat{g}_{ii}^{(n)} \to 1$ in probability,

(ii) $\hat{Y}_i^{(n)} \to 0$ in probability,

Some remarks are in order. The condition (i) implies that, for sufficiently large n, the ith vector $\Psi^{(n)}(i)$ will be almost parallel to $\mathcal{L}^{(n)}$, so that the projected ith vector $\Pi_{\mathcal{L}^{(n)}} \Psi^{(n)}(i)$ be almost identical to the original vector $\Psi^{(n)}(i)$. The condition (ii) implies that, for sufficiently large n, all the vectors $\Psi^{(n)}(j)$, $(j \neq i)$, will be simultaneously almost orthogonal to the projected ith vector $\Pi_{\mathcal{L}^{(n)}} \Psi^{(n)}(i)$. It should be emphasized that the choice of subspaces $\{\mathcal{L}^{(n)}\}_n$ is independent of the index i.

Theorem 1 *(von Neumann entropy as the weak orthogonality capacity) Given a probability measure p, let $C_w(p)$ be the supremum of $\limsup_{n\to\infty} \log L_n/n$ over all sequences $\{\mathcal{C}^{(n)}\}_n$ that satisfy the asymptotic weak orthogonality condition. Then $C_w(p) = H(\rho)$, where $H(\rho)$ is the von Neumann entropy for the density operator (1).*

Before proceeding to the proof, we mention a close connection between Theorem 1 and the noiseless quantum channel coding theorem [3]. Let us regard $\mathcal{C}^{(n)}$ as a quantum random codebook. Given a vector (a quantum codeword) in $\mathcal{C}^{(n)}$, our task is to estimate, by means of a certain measurement, which vector among $\mathcal{C}^{(n)}$ is the actual one. Associated with a codebook $\mathcal{C}^{(n)}$ and a subspace $\mathcal{L}^{(n)}$ is the Gram operator

$$\mathcal{G} := \sum_{j=1}^{L_n} |\Pi_{\mathcal{L}^{(n)}}\Psi^{(n)}(j)\rangle\langle\Pi_{\mathcal{L}^{(n)}}\Psi^{(n)}(j)|.$$

The operator $\mathcal{G}$ is strictly positive on the subspace

$$\hat{\mathcal{L}}^{(n)} := \mathrm{Span}\{\Pi_{\mathcal{L}^{(n)}}\Psi^{(n)}(j);\ 1 \le j \le L_n\}.$$

Let the operator $\mathcal{G}^{-1}$ be the inverse of $\mathcal{G}$ on $\hat{\mathcal{L}}^{(n)}$ and zero on the orthogonal complement $\hat{\mathcal{L}}^{(n)\perp}$. According to [3], we introduce a measurement $M^{(n)}$ by

$$M^{(n)} := \left\{ |\hat{\mu}(1)\rangle\langle\hat{\mu}(1)|,\ \ldots,\ |\hat{\mu}(L_n)\rangle\langle\hat{\mu}(L_n)|, I - \sum_{j=1}^{L_n} |\hat{\mu}(j)\rangle\langle\hat{\mu}(j)| \right\}, \tag{4}$$

where $\hat{\mu}(j)$ are vectors on $\hat{\mathcal{L}}^{(n)}$ defined by

$$\hat{\mu}(j) := \mathcal{G}^{-1/2}\Pi_{\mathcal{L}^{(n)}}\Psi^{(n)}(j) = \mathcal{G}^{-1/2}\Psi^{(n)}(j).$$

We can regard $M^{(n)}$ as a decoder for the codebook $\mathcal{C}^{(n)}$, in which the ith entry $|\hat{\mu}(i)\rangle\langle\hat{\mu}(i)|$, $(1 \le i \le L_n)$, corresponds to the ith codeword in $\mathcal{C}^{(n)}$, and the $(L_n + 1)$st entry the wild card. (Note that the decoder (4) with the special choice of a subspace $\mathcal{L}^{(n)} = \mathcal{H}^{\otimes n}$ was introduced by Holevo [4]).

The idea for adopting the decoder (4) is this: When the ith vector $\Psi^{(n)}(i)$ is strictly orthogonal to the other vectors in $\mathcal{C}^{(n)}$, then the Gram operator $\mathcal{G}$ with $\mathcal{L}^{(n)} := \mathcal{H}^{\otimes n}$ is decomposed into the orthogonal direct sum

$$\mathcal{G} = |\Psi^{(n)}(i)\rangle\langle\Psi^{(n)}(i)| \oplus \sum_{j\ne i} |\Psi^{(n)}(j)\rangle\langle\Psi^{(n)}(j)|,$$

so that

$$\mathcal{G}^{-1/2} = |\Psi^{(n)}(i)\rangle\langle\Psi^{(n)}(i)| \oplus \left(\sum_{j\ne i} |\Psi^{(n)}(j)\rangle\langle\Psi^{(n)}(j)| \right)^{-1/2},$$

and

$$\hat{\mu}(i) = \mathcal{G}^{-1/2}\Psi^{(n)}(i) = \Psi^{(n)}(i).$$

As a consequence, the decoding error probability $P_e^{(n)}(i)$ for the ith codeword $\Psi^{(n)}(i)$ by the decoder (4) is

$$P_e^{(n)}(i) = 1 - \mathrm{Tr}\, |\Psi^{(n)}(i)\rangle\langle\Psi^{(n)}(i)|\hat{\mu}(i)\rangle\langle\hat{\mu}(i)| = 0.$$

Thus it is expected that, when the ith vector is almost orthogonal to the other vectors in $\mathcal{C}^{(n)}$, the decoding error probability $P_e^{(n)}(i)$ will be small. In fact, this expectation is verified by the following

Lemma 2 *If $|\hat{g}_{ii}^{(n)}|^2 > 1 - \varepsilon$ and $\hat{Y}_i^{(n)} < \varepsilon$ hold for some i and $0 < \varepsilon < 1$, then the decoding error probability $P_e^{(n)}(i)$ for the ith codeword $\Psi^{(n)}(i)$ by the decoder (4) is upper bounded by $\frac{3}{2}\varepsilon$.*

Proof See Appendix B. □

Corollary 3 *If $\{\mathcal{C}^{(n)}\}_n$ satisfies asymptotic weak orthogonality condition, then $E[P_e^{(n)}] \to 0$ as $n \to \infty$, where $P_e^{(n)}$ is the average decoding error probability for the code $(\mathcal{C}^{(n)}, M^{(n)})$, and $E[\,\cdot\,]$ denotes the expectation.*

Proof Let the event $E_i^{(n)}(\varepsilon)$ be defined by

$$E_i^{(n)}(\varepsilon) := \{|\hat{g}_{ii}^{(n)}|^2 > 1 - \varepsilon \;\text{ and }\; \hat{Y}_i^{(n)} < \varepsilon\}.$$

By the assumption of asymptotic weak orthogonality, for all $\varepsilon > 0$ and $\delta > 0$, there is an N such that for all $n \geq N$ and i, $P(E_i^{(n)}(\varepsilon)) > 1 - \delta$ holds. Then

$$\begin{aligned} E[P_e^{(n)}(i)] &= E[P_e^{(n)}(i); E_i^{(n)}(\varepsilon)] + E[P_e^{(n)}(i); E_i^{(n)}(\varepsilon)^c] \\ &\leq E[P_e^{(n)}(i); E_i^{(n)}(\varepsilon)] + P(E_i^{(n)}(\varepsilon)^c) \\ &< \frac{3}{2}\varepsilon + \delta. \end{aligned}$$

Here, $E[X; A] := \int_A X\, dP$, and Lemma 2 is used in the last inequality. Since this upper bound is independent of i, we have

$$E[P_e^{(n)}] = \frac{1}{L_n}\sum_{i=1}^{L_n} E[P_e^{(n)}(i)] < \frac{3}{2}\varepsilon + \delta.$$

This completes the proof. □

Theorem 1 and Corollary 3 clarify why the decoder of the type (4) has fitted to the random coding technique in the proof of the direct part of the noiseless quantum channel coding theorem [3]. The notion of asymptotic weak orthogonality for the random codebook $\mathcal{C}^{(n)}$ thus explicates the physical implication of the probabilistic distinguishability among codewords in the quantum channel coding problem as well as the geometrical mechanism behind the decoder (4).

Proof of Theorem 1 We first prove the direct part $C_w(p) \geq H(\rho)$. Fix an arbitrarily small positive constant δ and, for each n, let L_n be such that $L_n < e^{n(H(\rho)-4\delta)}$. We show that there is a sequence $\{\mathcal{L}^{(n)}\}_n$ of subspaces for which the asymptotic weak orthogonality conditions (i) and (ii) hold for all i. The idea of the proof is similar to [3]: we take $\mathcal{L}^{(n)}$ to be the δ-typical subspace $\Lambda_\delta^{(n)}$ with respect to the density ρ. (For the reader's convenience, the definition and the basic properties of the δ-typical subspace are summarized in Appendix C.) Since

$$\hat{g}_{ii}^{(n)} = \langle \Pi_{\Lambda_\delta^{(n)}} \Psi^{(n)}(i) | \Pi_{\Lambda_\delta^{(n)}} \Psi^{(n)}(i) \rangle = \mathrm{Tr}\, |\Psi^{(n)}(i)\rangle\langle\Psi^{(n)}(i)| \Pi_{\Lambda_\delta^{(n)}},$$

we have

$$E[\hat{g}_{ii}^{(n)}] = \mathrm{Tr}\, \rho^{\otimes n} \Pi_{\Lambda_\delta^{(n)}} > 1 - \delta$$

for all i and all sufficiently large n (see Eq. (9)). This proves (i). On the other hand, for all $j (\neq i)$,

$$\begin{aligned} |\hat{g}_{ij}^{(n)}|^2 &= |\langle \Pi_{\Lambda_\delta^{(n)}} \Psi^{(n)}(i) | \Pi_{\Lambda_\delta^{(n)}} \Psi^{(n)}(j) \rangle|^2 \\ &= \mathrm{Tr}\, \Pi_{\Lambda_\delta^{(n)}} |\Psi^{(n)}(i)\rangle\langle\Psi^{(n)}(i)| \Pi_{\Lambda_\delta^{(n)}} |\Psi^{(n)}(j)\rangle\langle\Psi^{(n)}(j)| \Pi_{\Lambda_\delta^{(n)}}, \end{aligned}$$

so that

$$E|\hat{g}_{ij}^{(n)}|^2 = \mathrm{Tr}\, \Pi_{\Lambda_\delta^{(n)}} \rho^{\otimes n} \Pi_{\Lambda_\delta^{(n)}} \rho^{\otimes n} \Pi_{\Lambda_\delta^{(n)}} = \mathrm{Tr}\, (\rho^{\otimes n})^2 \Pi_{\Lambda_\delta^{(n)}} \leq e^{-n(H(\rho)-3\delta)},$$

(see Eq. (10)), and

$$E[\hat{Y}_i^{(n)}] \leq (L_n - 1)\, e^{-n(H(\rho)-3\delta)} < e^{-n\delta}.$$

Thus $\hat{Y}_i^{(n)} \to 0$ in L^1 for all i, proving (ii). This completes the proof of the direct part $C_w(p) \geq H(\rho)$.

We next prove the converse part $C_w(p) \leq H(\rho)$. Let X be the random variable uniformly distributed over $\mathcal{C}^{(n)}$ and let Y be the random variable representing the outcome of the corresponding decoder $M^{(n)}$ defined by (4). Then by virtue of Fano's inequality,

$$\begin{aligned}
\log 2 + P_e^{(n)} \log L_n &\geq H(X|Y) = H(X) - I(X:Y) \\
&= \log L_n - \frac{1}{L_n}\sum_{j=1}^{L_n} D_{M^{(n)}}\left(|\Psi^{(n)}(j)\rangle\langle\Psi^{(n)}(j)| \middle\| \frac{1}{L_n}\sum_{k=1}^{L_n} |\Psi^{(n)}(k)\rangle\langle\Psi^{(n)}(k)|\right) \\
&\geq \log L_n - \frac{1}{L_n}\sum_{j=1}^{L_n} D\left(|\Psi^{(n)}(j)\rangle\langle\Psi^{(n)}(j)| \middle\| \frac{1}{L_n}\sum_{k=1}^{L_n} |\Psi^{(n)}(k)\rangle\langle\Psi^{(n)}(k)|\right) \\
&= \log L_n - H\left(\frac{1}{L_n}\sum_{k=1}^{L_n} |\Psi^{(n)}(k)\rangle\langle\Psi^{(n)}(k)|\right).
\end{aligned}$$

Here $D(\sigma\|\tau) := \operatorname{Tr}\sigma(\log\sigma - \log\tau)$ is the quantum relative entropy between the quantum states σ and τ (with $\operatorname{supp}\sigma \subset \operatorname{supp}\tau$), and $D_M(\sigma\|\tau)$ denotes the classical Kullback-Leibler divergence between the probability distributions $p(\cdot) := \operatorname{Tr}\sigma M(\cdot)$ and $q(\cdot) := \operatorname{Tr}\tau M(\cdot)$ over the outcomes of the measurement M. (See [5] for notations.) The second inequality is due to the familiar monotonicity relation of the relative entropy [6, Theorem 1.5]. Now taking the expectation for the above inequality, and using the concavity of the von Neumann entropy, we have

$$\begin{aligned}
\log 2 + E[P_e^{(n)}] \log L_n &\geq \log L_n - E\left[H\left(\frac{1}{L_n}\sum_{k=1}^{L_n} |\Psi^{(n)}(k)\rangle\langle\Psi^{(n)}(k)|\right)\right] \\
&\geq \log L_n - H\left(\frac{1}{L_n}\sum_{k=1}^{L_n} E\left[|\Psi^{(n)}(k)\rangle\langle\Psi^{(n)}(k)|\right]\right) \\
&= \log L_n - H(\rho^{\otimes n}) \\
&= \log L_n - nH(\rho).
\end{aligned}$$

Therefore

$$(1 - E[P_e^{(n)}])\frac{\log L_n}{n} \leq H(\rho) + \frac{\log 2}{n}.$$

Thus in order to assure the asymptotic weak orthogonality (so that $E[P_e^{(n)}] \to 0$ as $n \to \infty$ by Corollary 3), $\limsup_n \log L_n/n$ must be less than or equal to $H(\rho)$. This completes the proof of the converse part $C_w(p) \leq H(\rho)$. □

3 Strong orthogonality

If the vectors in $\mathcal{C}^{(n)}$ are mutually strictly orthogonal, then the Gram matrix $G^{(n)} = [g_{ij}^{(n)}]$, where $g_{ij}^{(n)}$ is the inner product (2), is reduced to the identity. Therefore if they are mutually ‘almost’ orthogonal, the Gram matrix is expected to be close to the identity. This observation prompts us to define the strong orthogonality as follows. Given $\mathcal{C}^{(n)}$, let the random variable $Z^{(n)}$ be defined by the squared sum of off-diagonal elements of $G^{(n)}$, i.e.,

$$Z^{(n)} := \sum_{i=1}^{L_n} \sum_{j(\neq i)}^{L_n} |g_{ij}^{(n)}|^2.$$

We say that a sequence $\{\mathcal{C}^{(n)}\}_n$ satisfies *asymptotic strong orthogonality* condition if the following two conditions are satisfied:

(i) $Z^{(n)} \to 0$ in probability,

(ii) the sequence $\{Z^{(n)}\}_n$ is uniformly integrable.

Theorem 4 *(Quantum Rényi entropy as the strong orthogonality capacity) Given a probability measure p, let $C_s(p)$ be the supremum of* $\limsup_{n\to\infty} \log L_n/n$ *over all sequences $\{\mathcal{C}^{(n)}\}_n$ that satisfy asymptotic strong orthogonality condition. Then $C_s(p) = \frac{1}{2}H_2(\rho)$, where $H_2(\rho)$ is the quantum Rényi entropy of degree 2 for the density operator (1).*

Proof For $i \neq j$,

$$|g_{ij}^{(n)}|^2 = |\langle\Psi^{(n)}(i)|\Psi^{(n)}(j)\rangle|^2 = \mathrm{Tr}\,(|\Psi^{(n)}(i)\rangle\langle\Psi^{(n)}(i)|)(|\Psi^{(n)}(j)\rangle\langle\Psi^{(n)}(j)|),$$

so that

$$E|g_{ij}^{(n)}|^2 = \mathrm{Tr}\,(\rho^{\otimes n})^2 = (\mathrm{Tr}\,\rho^2)^n = e^{-nH_2(\rho)},$$

and

$$E[Z^{(n)}] = L_n(L_n-1)\,e^{-nH_2(\rho)}.$$

Let $\delta > 0$ be an arbitrarily small positive consitant. If $L_n < e^{n(H_2(\rho)/2-\delta)}$ then $E[Z^{(n)}] \to 0$, and if $L_n > e^{n(H_2(\rho)/2+\delta)}$ then $E[Z^{(n)}] \to \infty$. This completes the proof. □

Several remarks are in order. If $\{\mathcal{C}^{(n)}\}_n$ satisfies the condition (i) for the asymptotic strong orthogonality, then in a similar way to Lemma 2, it can be shown that the decoding error probabilities $\{P_e^{(n)}(i)\}_{1\leq i\leq L_n}$ by the decoder (4) with $\mathcal{L}^{(n)} := \mathcal{H}^{\otimes n}$ exhibits $\sum_{i=1}^{L_n} P_e^{(n)}(i) \to 0$ in L^1 as $n \to \infty$. (Compare this with Corollary 3.) The notion of asymptotic strong orthogonality thus leads to a new, strong type of probabilistic distinguishability in quantum measurement theory. It is not clear whether there is a coding theorem in which this strong distinguishability playes a pivotal role. A related question whether the converse part of Theorem 4 holds without the uniform integrability condition (ii) is also still open.

4 Conclusions

We have shown that asymptotic orthogonalities of random vectors lead us to new, geometrical, characterizations of the von Neumann entropy and the quantum Rényi entropy of degree 2. These characterizations are closely related to the distinguishability of the vectors by quantum mechanical measurements. In particular, a mechanism behind the random coding technique for the noiseless quantum channel coding theorem was clarified.

Acknowledgments

The author wishes to thank Michael Camarri and Hidetoshi Shimokawa for their helpful comments. He also acknowledges support of the Telecommunications Advancement Foundation.

Appendices

A Remark on the condition (3)

In this appendix, we exemplify that the condition (3) does not characterize the von Neumann entropy. Let ϕ_0, ϕ_1 be unit vectors in $\mathcal{H}$ with $a := |\langle\phi_0|\phi_1\rangle|^2 < 1$, and let p be the probability measure on $\mathcal{H}$ such that $p(\phi_0) = p(\phi_1) = \frac{1}{2}$. The nonzero eigenvalues of the corresponding density operator $\rho = \frac{1}{2}|\phi_0\rangle\langle\phi_0| + \frac{1}{2}|\phi_1\rangle\langle\phi_1|$ are $(1 \pm \sqrt{a})/2$, so the quantum Rényi entropy of degree 2 is

$$h_2 := H_2(\rho) = -\log\frac{1+a}{2}.$$

Let $\{X_k(i)\}_{ki}$ be $\{\phi_0, \phi_1\}$-valued random variables i.i.d. with respect to p, and let $\Psi^{(n)}(i) = X_1(i) \otimes \cdots \otimes X_n(i)$. The squared norm of the inner product $g_{ij}^{(n)}$ then becomes

$$|g_{ij}^{(n)}|^2 = \prod_{k=1}^{n} |\langle X_k(i)|X_k(j)\rangle|^2 = a^{N_{ij}},$$

where N_{ij} is the number of indices k for which $X_k(i) \neq X_k(j)$.

Now fix a number i arbitrarily. Then it is easily shown that for each n, $\{|g_{ij}^{(n)}|^2\,;\, j \in \boldsymbol{N},\ j \neq i\}$ are i.i.d. random variables, each taking the value a^ℓ with probability $\binom{n}{\ell}2^{-n}$, where $\ell = 0, ..., n$. In particular, they have the expectation

$$m^{(n)} := E\left[|g_{ij}^{(n)}|^2\right] = \left(\frac{1+a}{2}\right)^n = e^{-nh_2},$$

and the variance

$$v^{(n)} := V\left[|g_{ij}^{(n)}|^2\right] = \left(\frac{1+a^2}{2}\right)^n - \left(\frac{1+a}{2}\right)^{2n}.$$

We claim the following.

Proposition 5 *Let ε be a positive constant and let*

$$Y_i^{(n)} = \sum_{j(\neq i)}^{L_n} |g_{ij}^{(n)}|^2.$$

If $L_n < e^{n(h_2-\varepsilon)}$, then $Y_i^{(n)}$ converges to 0 in probability as $n \to \infty$, and if $L_n > e^{n(h_2+\varepsilon)}$, then $Y_i^{(n)}$ does not.

Proof Assume first that $L_n < e^{n(h_2-\varepsilon)}$. Then $E[Y_i^{(n)}] = (L_n - 1)\, m^{(n)} < e^{-n\varepsilon} \to 0$ as $n \to \infty$, proving that $Y_i^{(n)} \to 0$ in probability. To prove the second part, we use the following inequality:

$$P\left(Y_i^{(n)} < \frac{1}{2}(L_n - 1)\, m^{(n)}\right) < \frac{4}{(L_n - 1)\, m^{(n)}}, \tag{5}$$

which is verified as follows.

$$\begin{aligned} P\left(Y_i^{(n)} < \frac{1}{2}(L_n - 1)\, m^{(n)}\right) &\leq P\left(\left|Y_i^{(n)} - (L_n - 1)\, m^{(n)}\right| > \frac{1}{2}(L_n - 1)\, m^{(n)}\right) \\ &< \left(\frac{2}{(L_n - 1)\, m^{(n)}}\right)^2 V[Y_i^{(n)}] \\ &= \frac{4}{(L_n - 1)\, m^{(n)}} \cdot \frac{v^{(n)}}{m^{(n)}} < \frac{4}{(L_n - 1)\, m^{(n)}}. \end{aligned}$$

Now assume that $L_n > e^{n(h_2+\varepsilon)}$. Then $(L_n - 1)\, m^{(n)} > e^{n\varepsilon} - m^{(n)} \to \infty$ as $n \to \infty$. This fact and the inequality (5) together prove that $Y_i^{(n)}$ does not converge to 0 in probability. □

According to Proposition 5, it is h_2 (the quantum Rényi entropy of degree 2) that characterizes the asymptotic behavior of $Y_i^{(n)}$ in this example. As a consequence, one cannot characterize the von Neumann entropy as the capacity for the condition (3) in general.

Remark: If the definition of the asymptotic weak orthogonality is such that $Y_i^{(n)} \to 0$ in probability and the sequence $\{Y_i^{(n)}\}_n$ is uniformly integrable for all i, then it can be shown in a similar way to Theorem 4 that $C_w(p) = H_2(\rho)$ for a general probability measure p.

B Proof of Lemma 2

Due to the symmetry, it suffices to consider the case when $i = 1$. Let $\{\hat{e}_k\}_k$ be a complete orthonormal system (CONS) of the finite dimensional subspace $\hat{\mathcal{L}}^{(n)}$ with

$$\hat{e}_1 := \frac{\Pi_{\mathcal{L}^{(n)}}\Psi^{(n)}(1)}{\|\Pi_{\mathcal{L}^{(n)}}\Psi^{(n)}(1)\|}.$$

The $(1,1)$th matrix element of $\mathcal{G}^{-1/2}$ with respect to the CONS $\{\hat{e}_k\}_k$ is

$$\begin{aligned} \left(\mathcal{G}^{-1/2}\right)_{11} &:= \langle \hat{e}_1 | \mathcal{G}^{-1/2} \hat{e}_1 \rangle \\ &= \frac{\langle \Pi_{\mathcal{L}^{(n)}}\Psi^{(n)}(1) | \mathcal{G}^{-1/2} \Pi_{\mathcal{L}^{(n)}}\Psi^{(n)}(1)\rangle}{\|\Pi_{\mathcal{L}^{(n)}}\Psi^{(n)}(1)\|^2} \\ &= \frac{\langle \Psi^{(n)}(1) | \mathcal{G}^{-1/2} \Psi^{(n)}(1)\rangle}{\|\Pi_{\mathcal{L}^{(n)}}\Psi^{(n)}(1)\|^2}. \end{aligned} \tag{6}$$

The error probability $P_e^{(n)}(1)$ for the first codeword $\Psi^{(n)}(1)$ in $\mathcal{C}^{(n)}$ with respect to the decoder (4) is evaluated as

$$\begin{aligned} P_e^{(n)}(1) &= 1 - |\langle \Psi^{(n)}(1) | \hat{\mu}(1) \rangle|^2 \\ &= 1 - |\langle \Psi^{(n)}(1) | \mathcal{G}^{-1/2} \Psi^{(n)}(1) \rangle|^2 \\ &= 1 - \|\Pi_{\mathcal{L}^{(n)}}\Psi^{(n)}(1)\|^4 \left| \left(\mathcal{G}^{-1/2}\right)_{11} \right|^2 \\ &\leq 1 - \|\Pi_{\mathcal{L}^{(n)}}\Psi^{(n)}(1)\|^4 \left(\mathcal{G}_{11}\right)^{-1}. \end{aligned} \tag{7}$$

Here $\mathcal{G}_{11}$ stands for the $(1,1)$th matrix element of $\mathcal{G}$, and we have used (6) and the inequality

$$\left(\mathcal{G}^{-1/2}\right)_{11} \geq \left(\mathcal{G}_{11}\right)^{-1/2},$$

which is verified by Lemma 6 below. On the other hand,

$$\begin{aligned} \mathcal{G}_{11} &= |\langle \hat{e}_1 | \Pi_{\mathcal{L}^{(n)}}\Psi^{(n)}(1) \rangle|^2 + \sum_{j\geq 2} |\langle \hat{e}_1 | \Pi_{\mathcal{L}^{(n)}}\Psi^{(n)}(j) \rangle|^2 \\ &= \|\Pi_{\mathcal{L}^{(n)}}\Psi^{(n)}(1)\|^2 + \frac{1}{\|\Pi_{\mathcal{L}^{(n)}}\Psi^{(n)}(1)\|^2} \sum_{j\geq 2} |\langle \Pi_{\mathcal{L}^{(n)}}\Psi^{(n)}(1) | \Pi_{\mathcal{L}^{(n)}}\Psi^{(n)}(j) \rangle|^2 \\ &= \|\Pi_{\mathcal{L}^{(n)}}\Psi^{(n)}(1)\|^2 \left(1 + \frac{1}{\|\Pi_{\mathcal{L}^{(n)}}\Psi^{(n)}(1)\|^4} \sum_{j\geq 2} |\langle \Pi_{\mathcal{L}^{(n)}}\Psi^{(n)}(1) | \Pi_{\mathcal{L}^{(n)}}\Psi^{(n)}(j) \rangle|^2 \right) \\ &< \|\Pi_{\mathcal{L}^{(n)}}\Psi^{(n)}(1)\|^2 \left(1 + \frac{\varepsilon}{1-\varepsilon} \right) \\ &= \|\Pi_{\mathcal{L}^{(n)}}\Psi^{(n)}(1)\|^2 \frac{1}{1-\varepsilon}. \end{aligned} \tag{8}$$

Substituting (8) into (7), we have

$$P_e^{(n)}(1) < 1 - \|\Pi_{\mathcal{L}^{(n)}}\Psi^{(n)}(1)\|^2(1-\varepsilon) < 1-(1-\varepsilon)^{3/2} < \frac{3}{2}\varepsilon.$$

This completes the proof of Lemma 2. □

Lemma 6 *Let $A = [A_{ij}]$ be a strictly positive Hermitian matrix. Then*[1]

$$(A^{-1/2})_{11} \geq (A_{11})^{-1/2}$$

Proof Let

$$A = \sum_k \lambda_k E_k$$

be the spectral decomposition. Then

$$A^{-1/2} = \sum_k (\lambda_k)^{-1/2} E_k,$$

so that

$$(A^{-1/2})_{11} = \sum_k (\lambda_k)^{-1/2} \langle e_1 | E_k e_1 \rangle \geq \left(\sum_k \lambda_k \langle e_1 | E_k e_1 \rangle \right)^{-1/2} = (A_{11})^{-1/2}.$$

Here we have used Jensen's inequality and the fact that $\langle e_1 | E_k e_1 \rangle \geq 0$ for all k and $\sum_k \langle e_1 | E_k e_1 \rangle = 1$. □

C Typical subspaces

In this appendix, we give a brief account of the so-called typical subspace (cf. [7] [8] [3]). Given a density operator ρ on $\mathcal{H}$, let

$$\rho = \sum_{j \in J} \lambda_j E_j$$

be a Schatten decomposition, where $\lambda_j > 0$ for all $j \in J$ and $\sum_j \lambda_j = 1$. Note that $\lambda := (\lambda_1, \lambda_2, ...)$ is naturally regarded as a probability distribution on the index set J such that $\lambda(j) = \lambda_j$. A Schatten decomposition of $\rho^{\otimes n}$ is given by

$$\rho^{\otimes n} = \sum_{(j_1,\ldots,j_n) \in J^n} (\lambda_{j_1} \cdots \lambda_{j_n})(E_{j_1} \otimes \cdots \otimes E_{j_n}).$$

[1] More generally, we can prove that $(A^m)_{11} \geq (A_{11})^m$ for $m < 0$ or $m > 1$, and $(A^m)_{11} \leq (A_{11})^m$ otherwise.

Obviously, the eigenvalues of $\rho^{\otimes n}$ form a probability distribution λ^n, the i.i.d. extension of λ, on the set J^n.

Given a density operator ρ and a positive constant δ, an eigenvalue $(\lambda_{j_1}\cdots\lambda_{j_n})$ of $\rho^{\otimes n}$ is called *δ-typical* if the sequence $j_1, ..., j_n$ of indices is δ-typical [9, p. 51] with respect to the probability distribution λ^n, that is, if $e^{-n(H(\rho)+\delta)} \leq \lambda^n(j_1, ..., j_n) \leq e^{-n(H(\rho)-\delta)}$. It follows that for all sufficiently large n,

(a) $e^{-n(H(\rho)+\delta)} \leq$ (a δ-typical eigenvalue) $\leq e^{-n(H(\rho)-\delta)}$,

(b) (the sum of δ-typical eigenvalues) $> 1-\delta$,

(c) $(1-\delta)e^{n(H(\rho)-\delta)} \leq$ (the number of δ-typical eigenvalues) $\leq e^{n(H(\rho)+\delta)}$.

Here (a) is a direct consequence of the definition, and (b) and (c) follow from the asymptotic equipartition property [9, Theorem 3.1.2].

Let $\Lambda_\delta^{(n)}(\subset \mathcal{H}^{\otimes n})$ be the linear span of such eigenvectors of $\rho^{\otimes n}$ that correspond to δ-typical eigenvalues. The subspace $\Lambda_\delta^{(n)}$ is called *δ-typical* with respect to the density ρ. Let $\Pi_{\Lambda_\delta^{(n)}}$ be the projection operator onto the δ-typical subspace $\Lambda_\delta^{(n)}$. Clearly the operators ρ and $\Pi_{\Lambda_\delta^{(n)}}$ commute. And it follows immediately from (a)-(c) that, for all sufficiently large n,

$$\mathrm{Tr}\,\rho^{\otimes n}\Pi_{\Lambda_\delta^{(n)}} > 1-\delta, \tag{9}$$

$$\mathrm{Tr}\,(\rho^{\otimes n})^2\Pi_{\Lambda_\delta^{(n)}} \leq \left(e^{-n(H(\rho)-\delta)}\right)^2 e^{n(H(\rho)+\delta)} = e^{-n(H(\rho)-3\delta)}. \tag{10}$$

References

[1] A. Rényi, “On the foundations of information theory,” Review of the International Statistics Institute **33:1**, pp. 1-14 (1965)

[2] M. Camarri and J. Pitman, “Limit distributions and random trees derived from the birthday problem with unequal probabilities,” Electronic Journal of Probability **5**, pp. 1-18 (2000).

[3] P. Hausladen, R. Jozsa, B. Schumacher, M. Westmoreland, and W. K. Wootters, “Classical information capacity of a quantum channel,” Phys. Rev. A **54**, pp. 1869-1876 (1996).

[4] A. S. Holevo, “Capacity of a quantum communication channel,” Probl. Pered. Inform., **15**, pp. 3-11 (1979); translation in Probl. Inform. Transm., **15**, pp.247-253 (1979).

[5] A. Fujiwara and H. Nagaoka, “Operational capacity and pseudoclassicality of a quantum channel,” IEEE Trans. on Inform. Theory **44**, pp. 1071-1086 (1998).

[6] M. Ohya and D. Petz, *Quantum Entropy and Its Use* (Springer, Berlin, 1993).

[7] B. Schumacher, “Quantum coding,” Phys. Rev. A **51**, pp. 2738-2747 (1995).

[8] R. Jozsa and B. Schumacher, “A new proof of the quantum noiseless source coding theorem,” J. Mod. Opt. **41**, pp. 2343-2349 (1994).

[9] T. M. Cover and J. Thomas, *Elements of Information Theory* (Wiley, New York, 1991).