QUANTUM ALGORITHMIC ENTROPY

PETER GÁCS

ABSTRACT. We extend algorithmic information theory to quantum mechanics, taking a universal semicomputable density matrix ("universal probability") as a starting point, and define complexity (an operator) as its negative logarithm.

A number of properties of Kolmogorov complexity extend naturally to the new domain. Approximately, a quantum state is simple if it is within a small distance from a low-dimensional subspace of low Kolmogorov complexity. The von Neumann entropy of a computable density matrix is within an additive constant from the average complexity. Some of the theory of randomness translates to the new domain.

We explore the relations of the new quantity to the quantum Kolmogorov complexity defined by Vitányi (we show that the latter is sometimes as large as $2n - 2\log n$) and the qubit complexity defined by Berthiaume, Dam and Laplante. The "cloning" properties of our complexity measure are similar to those of qubit complexity.

1. INTRODUCTION

Kolmogorov complexity (or by a more neutral name, description complexity) is an attractive concept, helping to shed light onto such subtle concepts as information content, randomness and inductive inference. Quantum information theory, a subject with its own conceptual difficulties, is attracting currently more attention than ever before, due to the excitement around quantum computing, quantum cryptography, and the many connections between these areas. The new interest is also spurring efforts to extend the theory of description complexity to the quantum setting: see [6], [1]. We continue these efforts in the hope that the correct notions will be found at the convergence of approaches from different directions. This has been the case for the theory of classical description complexity and randomness, What we expect from these researches is an eventual deeper understanding of quantum information theory itself.

One of the starting points from wich it is possible to arrive at description complexity is Levin's concept of a universal semicomputable (semi)measure. We follow this approach in the quantum setting, where probability measures are generalized into density matrices.

In contrast to the works [6], [1] we do not find the notion of a quantum computer essential for this theory, even to the notions and results found in these works. The reason is that limitations on computing time do not play a role in the main theory of description complexity, and given enough time, a quantum computer can be simulated by a classical computer to any desired degree of precision.

1.1. **Notation.** It seems that universal probability can also be defined in an infinite-dimensional space (it should be simple to extend the notions to Fock space), but we will confine ourselves to finite-dimensional spaces, in order to avoid issues of convergence and spectral representation for infinite-dimensional operators. Let us fix for each N a finite-dimensional Hilbert space $\mathcal{H}_N$, with a *canonical orthonormal basis* $|\beta_1\rangle, \dots, |\beta_N\rangle$. (We do not use double index here, since we can assume that $\mathcal{H}_N \subset \mathcal{H}_{N+1}$ and the canonical basis of $\mathcal{H}_N$ is also the beginning of that of $\mathcal{H}_{N+1}$.) Let $\mathcal{Q}_n = \bigotimes_{i=1}^n \mathcal{Q}_1$ be the Hilbert space of n qubits. Let $|0\rangle, |1\rangle$ be some fixed orthonormal basis of $\mathcal{Q}_1$.

Date: March 1, 2005.

Key words and phrases. Kolmogorov complexity, quantum information theory.

The paper was written during the author's visit at CWI, Amsterdam, partially supported by a grant of NWO.

Let $\mathbb{Z}_2^n$ be the set of binary sequences of length n. If $x \in \mathbb{Z}_2^n$ then $x = (x(1), x(2), \dots, x(n))$, and we write

$$l(x) = n.$$

We denote, as usual, for $x \in \mathbb{Z}_2^n$:

$$|x\rangle = \bigotimes_{i=1}^{n} |x(i)\rangle.$$

We identify $\mathcal{Q}_n$ with $\mathcal{H}_{2^n}$, with the canonical basis element $|\beta_x\rangle = |x\rangle$.

If we write ψ or $|\psi\rangle$ for a state then the corresponding element of the dual space will be written either as $\psi^\dagger$ or as $\langle\psi|$. Accordingly, the inner product can be written in three ways as

$$\langle\phi|\psi\rangle = \langle\phi, \psi\rangle = \phi^\dagger\psi.$$

As usual, we will sometimes write

$$|\phi\rangle \otimes |\psi\rangle = |\phi\rangle|\psi\rangle = |\phi, \psi\rangle.$$

The operation Tr denotes trace, and over a tensor product space $\mathcal{H}_X \otimes \mathcal{H}_Y$, the operation Tr_Y denotes partial trace.

As usual, for self-adjoint operators ρ, σ, let us write $\rho \leqslant \sigma$ if $\sigma - \rho$ is nonnegative definite.

Let us call a quantum state $|\psi\rangle$, with coefficients $\langle\beta_i|\psi\rangle$ that are *algebraic numbers*, *elementary*. The reason for going to coefficients that are algebraic numbers is that this allows us the usual operations of linear algebra (orthogonalization, finding eigenvalues and eigenvectors) while remaining in the realm of elementary objects.

Whenever we write $U(p) = |\phi\rangle$ for a Turing machine U, we mean that U simply outputs the (algebraic definitions of the) coefficients of the elementary state $|\phi\rangle$. Similarly, let us call a self-adjoint operator T *elementary* if it is given by a matrix with algebraic entries.

We will also write $U(p) = |\phi\rangle$ if $U(p)$ outputs a sequence of tuples $(c_{1k}, \dots, c_{Nk})$ for $k = 1, 2, \dots$, where c_{ik} is an elementary approximation of $\langle\beta_i|\phi\rangle$ to within 2^{-k}. In this case, we say that $|\phi\rangle$ is a *computable* quantum state with program p. We can talk similarly about a program computing a linear operator on the finite-dimensional space, or even computing an infinite sequence $|\phi_1\rangle, |\phi_2\rangle, \dots$ of states, in which case we output progressively better approximations to more and more elements of the sequence.

Let $\overset{+}{<}$ denote inequality to within an additive constant, and $\overset{*}{<}$ inequality to within a multiplicative constant.

We assume that the reader knows the definition and simple properties of Kolmogorov complexity, even the definition of its prefix-free version $K(x)$. For a reference, use [3].

1.2. **Attempts to define a quantum Kolmogorov complexity.** In [6], a notion of the description complexity of a quantum state was introduced. Though that definition uses quantum Turing machines, this does not seem essential. Indeed, a quantum Turing machine can simulate a classical one. And if there is no restriction on computing time then any state output by a quantum Turing machine starting from $|0 \dots 0\rangle$ can also be output with arbitrary approximation by some ordinary Turing machine. We reproduce the definition from [6] as follows. For $|\psi\rangle \in \mathcal{H}_n$, let

$$\mathrm{Kq}(|\psi\rangle \mid N) = \min\{\, l(p) - \log|\langle\phi|\psi\rangle|^2 : U(p, N) = |\phi\rangle \,\}.$$

So, the complexity of $|\psi\rangle$ is made up of the length of a program describing an approximation $|\phi\rangle$ to $|\psi\rangle$ and a term penalizing for bad approximation. It is proved in [6] that for $|\psi\rangle \in \mathcal{Q}_n$,

$$\mathrm{Kq}(|\psi\rangle \mid n) \overset{+}{<} 2n.$$

The lower bounds given in that paper are close to n. The following theorem will be proved in Section 7.

Theorem 1. *For large enough n, there are states $|\psi\rangle \in \mathcal{Q}_n$ with* $\mathrm{Kq}(|\psi\rangle \mid n) > 2n - 2\log n$.

An entirely different approach to quantum Kolmogorov complexity is used in [1], where even the defining programs consist of qubits rather than ordinary bits. I will refer informally to complexity defined in [1] as "qubit complexity". Despite the difference in some of the goals and basic definitions, still a number of results of that paper look somewhat similar to ours.

1.3. **This paper.** The definition of Kq reflects the view that quantum states should not be accorded the status of individual outcomes of experiments, and therefore Kq strives only to approximate specification. We go a little further, and approach quantum complexity using probability distributions to start with. We find a universal semicomputable (semi-) density matrix ("universal probability") and define a "complexity operator" as its negative logarithm. Depending on the order of taking the logarithm and the expectation, two possible complexities are introduced for a quantum state $|\psi\rangle$: $\underline{H}(|\psi\rangle) \stackrel{+}{<} \overline{H}(|\psi\rangle)$.

A number of properties of Kolmogorov complexity extend naturally to the new domain. Approximately, a quantum state is simple if it is within a small distance from a low-dimensional subspace of low Kolmogorov complexity. (Ideally, the three vague terms should play a role in the following decreasing order of significance: dimension, complexity, closeness.) This property can be used to relate our algorithmic entropy to both Vitányi's complexity and qubit complexity. We find that $\underline{H}$ is within constant factor of Vitányi's complexity, that $\overline{H}$ essentially lowerbounds qubit complexity and upperbounds an oracle version of qubit complexity.

Though Vitányi's complexity is typically close to $2n$, while qubit complexity is $\stackrel{+}{<} n$, these are differences only within a constant factor; on the other hand, occasionally $\underline{H}$ can be much smaller than $\overline{H}$ and thus Vitányi's complexity is occasionally much smaller than qubit complexity. This is due to the permissive way in which Vitányi's complexity deals with approximations.

The von Neumann entropy of a computable density matrix is within an additive constant from the average complexity. Some of the theory of randomness translates to the new domain, but new questions arise due to non-commutativity.

The results on the maximal complexity of clones are sharp, and similar to those in [1].

2. Universal probability

Let us call a nonnegative real function $f(x)$ defined on strings a *semimeasure* if $\sum_x f(x) \leqslant 1$, and a *measure* (a probability distribution) if the sum is 1. A function is called *lower semicomputable* if there is a monotonically increasing sequence $g_n(x)$ of functions converging to it such that $(n, x) \mapsto g_n(x)$ is a computable function mapping into rational numbers. It is computable when it is both lower and upper semicomputable. (A lower semicomputable measure can be shown to be also computable.) The reason for introducing semicomputable semimeasures is not that computable measures are not felt general enough; rather, this step is analogous to the introduction of recursively enumerable sets and partial recursive functions. Just as there are "universal" (or, "complete" in terms of, say, many-one reduction) recursively enumerable sets but no universal recursive sets, there is a universal semicomputable semimeasure in the sense of the following proposition, even though there is no universal computable measure.

Let U be an optimal prefix Turing machine used in the definition of $K(x)$, and let $z_1, z_2, \ldots$ be an infinite sequence. Then the quantity $U(z)$ is well-defined: it is the output of U when z is written on the input tape. Let $Z_1, Z_2, \ldots$ be an infinite coin-tossing 0-1 sequence, and let us define

$$\mathbf{m}'(x) = \mathrm{Prob}[U(Z) = x]. \tag{2.1}$$

Proposition 2.1 (Levin)**.** *There is a semicomputable semimeasure μ with the property that for any other semicomputable semimeasure ν there is a constant $c_\nu > 0$ such that for all x we have $c_\nu \nu(x) \leqslant \mu(x)$. Moreover, $\mu \stackrel{*}{=} \mathbf{m}'$.*

Proof sketch. We define a Turing machine T that will output a sequence (p_t, x_t, r_t) where r_t is a positive rational number. At any time t, let $r_t(p, x)$ be defined as follows. If there is no $i \leqslant t$ for

which some (p, x, r_i) has been outputted then $r_t(p, x) = 0$; otherwise, $r_t(p, x)$ is the maximum of those r_i. The machine T will have the following property for all p:

$$\sum_x r_t(p, x) \leqslant 1. \tag{2.2}$$

To define T, take a universal Turing machine $V(p, x, n)$. Let T simulate V simultaneously on all inputs. If at any stage of the simulation, some $V(p, x, n)$ has been found, then T checks whether it can interpret $V(p, x, n)$ as a positive rational number r, and whether it can output the triple (p, x, r) while keeping the condition (2.2). If yes, the triple is outputted, otherwise it is not, and the simulation continues. Define $\nu(p, x) = \lim_t r_t(p, x)$. Then it is easy to check that $\mu(x) = \sum_p 2^{-p-1}\nu(p, x)$ satisfies the conditions of the proposition.

To show $\mu \stackrel{*}{=} \mathbf{m}'$, note that the random variable whose distribution is μ can be represented as a function of the coin-tossing infinite sequence. It is not difficult to check that the function in question now can be implemented by a prefix Turing machine. □

We will call any semicomputable semimeasure μ with the property in the proposition "universal". Any two universal semimeasures dominate each other within a multiplicative constant. We fix one such measure and denote it by

$$\mathbf{m}(x)$$

and call it the *universal probability*. Its significance for complexity theory can be estimated by by the following theorem, deriving the prefix complexity $K(x)$ from the universal probability.

Proposition 2.2 (Levin's Coding Theorem)**.** *We have* $K(x) = -\log \mathbf{m}(x)$.

The lower bound $(-\log \mathbf{m}(x)) \stackrel{+}{<} K(x)$ comes easily from the fact that $K(x)$ is upper semicomputable and satisfies the "Kraft inequality" $\sum_x 2^{-K(x)} \leqslant 1$. For the proof of the upper bound, see [3].

The above concepts and results can be generalized to the case when we have an extra parameter in the condition: we will therefore talk about $\mathbf{m}(x \mid N)$, the universal probability conditional to N, a function maximal within a multiplicative constant among all lower semicomputable functions $f(x, N)$ which also satisfy the condition $\sum_x f(x, N) \leqslant 1$. The coding theorem generalizes to $2^{-K(x|N)} \stackrel{*}{=} \mathbf{m}(x \mid N)$.

Constructive objects other than integers or strings can be encoded into integers in some canonical way. Elementary quantum states $|\psi\rangle \in \mathcal{H}_N$ also correspond to integers, and this is how we understand the expression

$$\mathbf{m}(|\psi\rangle \mid N),$$

which is therefore nonzero only for elementary states $|\psi\rangle$. (This is not our definition of quantum universal probability or complexity, only a tool from classical complexity theory helpful in its discussion.)

The quantum analog of a probability distribution is a density matrix, a self-adjoint positive semidefinite operator with trace 1. Just as with universal probability, let us allow operators with trace less than 1, and call them *semi-density matrices*.

We call a sequence A_N of operators, where A_N is defined over $\mathcal{H}_N$, *lower semicomputable* if there is a double sequence of elementary operators A_{Nk} with the property that for each N, the sequence A_{Nk} is increasing and converges to A_N.

Lemma 2.3.

(1) *A computable sequence of operators is also lower semicomputable.*
(2) *If A_N is nonnegative then the elements of the sequence A_{Nk} can be chosen nonnegative.*

Proof. Both these statements are proved via standard approximations. □

From now on, we suppress the index N whenever it is not necessary to point out its presence for clarity.

Theorem 2. *There is a lower semicomputable semi-density matrix $\boldsymbol{\mu}$ dominating all other such matrices in the sense that for every other such matrix σ there is a constant $c_\sigma > 0$ with $c_\sigma \sigma \leqslant \boldsymbol{\mu}$. We have $\boldsymbol{\mu} \stackrel{*}{=} \boldsymbol{\mu}'$ where*

$$\boldsymbol{\mu}' = \sum_{|\psi\rangle} \mathbf{m}(|\psi\rangle)|\psi\rangle\langle\psi|. \tag{2.3}$$

Also

$$\boldsymbol{\mu} \stackrel{*}{=} \sum_{\nu} \mathbf{m}(\nu)\nu \stackrel{*}{=} \sum_{P} \mathbf{m}(P)P/\dim P$$

where ν runs through all elementary semi-density matrices and P runs through all elementary projections.

Proof. The proof of the existence of $\boldsymbol{\mu}$ is completely analogous to the proof of Proposition 2.1.

To prove $\boldsymbol{\mu} \stackrel{*}{=} \boldsymbol{\mu}'$, note first that the form of its definition guarantees that $\boldsymbol{\mu}'$ is a lower semicomputable semi-density, and therefore $\boldsymbol{\mu}' \stackrel{*}{<} \boldsymbol{\mu}$. It remains to prove $\boldsymbol{\mu} \stackrel{*}{<} \boldsymbol{\mu}'$. Since $\boldsymbol{\mu}$ is lower semicomputable, there is a nondecreasing sequence $\boldsymbol{\mu}_k$ of elementary semi-density matrices such that $\boldsymbol{\mu} = \lim_k \boldsymbol{\mu}_k$, with $\boldsymbol{\mu}_0 = 0$. For $k \geqslant 1$, let $\delta_k = \mu_k - \mu_{k-1}$. Each of the nonnegative self-adjoint operators δ_k can be represented as a sum

$$\delta_k = \sum_{i=1}^{n} p_{ki}|\phi_{ki}\rangle\langle\phi_{ki}|.$$

Thus, $\boldsymbol{\mu} = \sum_{ki} p_{ki}|\phi_{ki}\rangle\langle\phi_{ki}|$, with a computable sequence $p_{ki} \geqslant 0$, where $\sum_{k,i} p_{ki} < 1$. The vectors $|\phi_{ki}\rangle$ and the values p_{nk} can be chosen elementary. Noting $p_{ki} \stackrel{*}{<} \mathbf{m}(k,i) \stackrel{*}{<} \mathbf{m}(|\phi_{ki}\rangle)$ finishes the proof.

The statement of sum representations using projections and elementary density matrices is weaker than the statement about $\boldsymbol{\mu}'$. □

We will call $\boldsymbol{\mu}$ the *quantum universal (semi-) density matrix*. Thus, the quantum universal probability of a quantum state $|\psi\rangle$ is given by

$$\langle\psi|\boldsymbol{\mu}|\psi\rangle.$$

A representation analogous to (2.1) holds also for the quantum universal probability $\boldsymbol{\mu}$. It is not necessary to introduce a quantum Turing machine in place of a classical Turing machine, since instead of outputting an elementary quantum state $|\psi\rangle$, we can just output the probabilities themselves, leaving the preparation of the state itself to whatever device we want, which might as well be a quantum Turing machine. The output of $U(Z)$ classically is a probability distribution over the set of strings: string x comes out with probability $\mathbf{m}(x)$. When the outputs are quantum states $|\phi\rangle$ with probability $\mathbf{m}(|\phi\rangle)$, then the relevant output is not the distribution $|\phi\rangle \mapsto \mathbf{m}(|\phi\rangle)$: by far not all this information is available. The actual physical output is just the density matrix $\boldsymbol{\mu}'$ as given in (2.3). Thus, we take the projection associated with each possible output $|\phi\rangle$, multiply it with its probability and add up all these terms. Indeed, assume that A is any self-adjoint operator expressing some property. The expected value of A over $U(Z)$ is given by $\operatorname{Tr} A\boldsymbol{\mu}'$. In particular, suppose that for some quantum state $|\psi\rangle$ we measure whether $U(Z) = |\psi\rangle$. The measurement will give a "yes" answer with probability

$$\begin{aligned}\sum_{|\phi\rangle} \mathbf{m}(|\phi\rangle)|\langle\phi|\psi\rangle|^2 &= \sum_{|\phi\rangle} \mathbf{m}(|\phi\rangle)\langle\psi|(|\phi\rangle\langle\phi|)|\psi\rangle \\ &= \langle\psi|\boldsymbol{\mu}'|\psi\rangle = \operatorname{Tr}|\psi\rangle\langle\psi|\boldsymbol{\mu}'.\end{aligned}$$

These analogies suggest to us to define complexity also as a self-adjoint operator:

$$\boldsymbol{\kappa} = -\log\boldsymbol{\mu}. \tag{2.4}$$

Proposition 2.4. *The operator function $A \mapsto \log A$ is monotonic.*

For a proof, see [2]. This implies the upper semicomputability of $(-\log \boldsymbol{\mu})$. For some readers to appreciate that the proposition is nontrivial, we mention that for example $A \mapsto e^A$ is not monotonic (see the same references). We will also use the following theorem, which could be called the "quantum Jensen inequality":

Proposition 2.5. *If $f(x)$ is a convex function in an interval $[a,b]$ containing the eigenvalues of operator A then for all $|\psi\rangle$ we have*

$$f(\langle\psi|A|\psi\rangle) \leqslant \langle\psi|f(A)|\psi\rangle. \tag{2.5}$$

Proof. Easy, see [7]. □

This implies:

Lemma 2.6. *Let f be a function concave in the interval $[a,b]$, and $|\psi\rangle$ a vector. Then the function $A \mapsto \langle\psi|f(A)|\psi\rangle$ is concave for self-adjoint operators A whose spectrum is contained in $[a,b]$.*

We have now two alternative definitions for quantum complexity of a pure state, depending on the order of taking the logarithm and taking the expectation:

$$\underline{H}(|\psi\rangle) = -\log\langle\psi|\boldsymbol{\mu}|\psi\rangle, \tag{2.6}$$

$$\overline{H}(|\psi\rangle) = -\langle\psi|(\log\boldsymbol{\mu})|\psi\rangle = \langle\psi|\boldsymbol{\kappa}|\psi\rangle. \tag{2.7}$$

An inequality in one direction can be established between them easily:

Theorem 3.

$$\underline{H}(|\psi\rangle) \leqslant \overline{H}(|\psi\rangle).$$

Proof. Use (2.5). □

The difference between the two quantities can be very large, as shown by the following example.

Example 2.7. Let $|1\rangle,\dots,|N\rangle$ be the eigenvectors of $\boldsymbol{\mu}$ ordered by decreasing eigenvalues p_i. Then $p_1 \stackrel{*}{=} 1$ and $p_N \stackrel{*}{=} N^{-1}$. For vector $|\psi\rangle = 2^{-1/2}(|1\rangle + |N\rangle)$ we have

$$\underline{H}(|\psi\rangle) = -\log\langle\psi|\boldsymbol{\mu}|\psi\rangle = -\log(p_1/2 + p_N/2) \stackrel{+}{=} 0,$$
$$\overline{H}(|\psi\rangle) = \langle\psi|\boldsymbol{\kappa}|\psi\rangle = (-\log p_1 - \log p_N)/2 \stackrel{+}{=} (\log N)/2.$$

♢

Which one of the two definitions is more appropriate? We prefer $\overline{H}$ since we like the idea of a complexity operator; however, in the present paper, we try to study both.

The complexity Kq introduced in [6] can be viewed as the formula resulting from $\underline{H}(|\psi\rangle)$ when the sum in (2.3) is replaced with supremum. In classical algorithmic information theory, the result does not change by more than a multiplicative constant after replacement, but Theorem 1 shows that it does in the quantum case.

Remark 2.8. It seems natural to generalize $\overline{H}(|\psi\rangle)$ and $\underline{H}(|\psi\rangle)$ to density matrices ρ by

$$\overline{H}(\rho) = \operatorname{Tr}\boldsymbol{\kappa}\rho, \quad \underline{H}(\rho) = -\log\operatorname{Tr}\boldsymbol{\mu}\rho,$$

but we do not explore this path in the present paper, and are not even sure that this is the right generalization. ♢

3. Properties of algorithmic entropy

3.1. **Relation to classical description complexity.** It was one of the major attractions of the original Kolmogorov complexity that it could be defined without reference to probability and then it could be used to characterize randomness. Unfortunately, we do not have any characterization, even to good approximation, of $\overline{H}(|\psi\rangle)$ or $\underline{H}(|\psi\rangle)$ in terms avoiding probability. As a generalization of classical complexity, it has the properties of classical complexity in the original domain, just as Kq and qubit complexity.

Theorem 4. *Let $|1\rangle, |2\rangle, \dots$ be a computable orthogonal sequence of states. Then for $H = \overline{H}$ or $\underline{H}$, we have*

$$H(|i\rangle) \stackrel{+}{=} K(i), \tag{3.1}$$

where the constant in $\stackrel{+}{=}$ depends on the definition of the sequence.

Proof. The function $f(i) = \langle i|\boldsymbol{\mu}|i\rangle$ is lower semicomputable with $\sum_i f(i) \leqslant 1$, hence it is dominated by $\mathbf{m}(i)$. This shows $K(i) \stackrel{+}{<} \underline{H}(|i\rangle)$.

On the other hand, the semi-density matrix $\rho = \sum_i \mathbf{m}(i)|i\rangle\langle i|$ is lower semicomputable, so $\rho \stackrel{*}{<} \boldsymbol{\mu}$, $-\log\rho \stackrel{+}{>} \boldsymbol{\kappa}$, hence

$$K(i) = \langle i|(-\log\rho)|i\rangle \stackrel{+}{>} \langle i|\boldsymbol{\kappa}|i\rangle = \overline{H}(|i\rangle).$$

□

3.2. **Upper and lower bounds in terms of small simple subspaces.** The simple upper bound follows immediately from the domination property of universal probability.

Theorem 5. *Assume that $|\psi\rangle \in \mathcal{H}_N$. Then*

$$\boldsymbol{\kappa} \stackrel{+}{<} (\log N)\mathbf{1}.$$

In particular, if $|\psi\rangle \in \mathcal{Q}_n$ then $\overline{H}(|\psi\rangle) \stackrel{+}{<} n$.

Proof. Let $\rho = N^{-1}\mathbf{1}$, then $\rho \stackrel{*}{<} \boldsymbol{\mu}$, hence $\boldsymbol{\kappa} \stackrel{+}{<} (\log N)\mathbf{1}$. □

Remark 3.1. N is an implicit parameter here, so it is more correct to write $\boldsymbol{\kappa}(\cdot \mid N) \stackrel{+}{<} (\log N)\mathbf{1}$. We do not have any general definition of quantum conditional complexity (just as no generally accepted notion of quantum conditional entropy is known), but conditioning on a classical parameter is not problematic. ♢

There is a more general theorem for classical complexity. For a finite set A let $K(A)$ be the length of the shortest program needed to enumerate the elements of A. Then for all $x \in A$ we have

$$K(x) \stackrel{+}{<} K(A) + \log \#A + 2\log \#A.$$

What may correspond to a simple finite set A is a projector P that is lower semicomputable as a nonnegative operator. What corresponds to $\#A$ is the dimension $\operatorname{Tr} P$ of the subspace to which P projects. What corresponds to $x \in A$ is measuring the angle between $|\psi\rangle$ and the space to which P projects.

Theorem 6. *Let P be a lower semicomputable projection with $d = \operatorname{Tr} P$. We have*

$$\underline{H}(|\psi\rangle) \stackrel{+}{<} K(P) + \log d - \log\langle\psi|P|\psi\rangle, \tag{3.2}$$

$$\overline{H}(|\psi\rangle) \stackrel{+}{<} K(P) + \log d + (1 - \langle\psi|P|\psi\rangle)\log N. \tag{3.3}$$

Proof. Let ρ be the semi-density matrix

$$\frac{1}{2}\left(\frac{P}{d} + \frac{\mathbf{1} - P}{N}\right) = \frac{1}{2}(\mathbf{1}/N + P(1/d - 1/N))$$

From the first form, it can be seen that it is semi-density, from the second form, it can be seen that it is lower semicomputable. By Theorem 2, we have $2^{K(\rho)}\rho \stackrel{*}{<} \boldsymbol{\mu}$. Since $K(\rho) \stackrel{+}{=} K(P)$, we have

$$\underline{H}(|\psi\rangle) = -\log\langle\psi|\boldsymbol{\mu}|\psi\rangle \stackrel{+}{<} K(P) + \log\langle\psi|(P/d)|\psi\rangle \stackrel{+}{=} K(P) + \log d - \log\langle\psi|P|\psi\rangle.$$

On the other hand,

$$\begin{aligned}\overline{H}(|\psi\rangle) &= \langle\psi|(-\log\boldsymbol{\mu})|\psi\rangle \\ &\stackrel{+}{<} K(P) + \langle\psi|P|\psi\rangle\log d + (1 - \langle\psi|P|\psi\rangle)\log N.\end{aligned}$$

□

This theorem points out again the difference between $\underline{H}$ and $\overline{H}$. If $|\psi\rangle$ has a small angle with a small-dimensional subspace this makes $\underline{H}(|\psi\rangle)$ small. For $\overline{H}(|\psi\rangle)$, the size of the angle gets multiplied by $\log N$, so if nothing more is known about $|\psi\rangle$ then not only the dimension of P counts but also the dimension of the whole space we are in.

Above, we defined what it means for a program to recursively "enumerate a subspace" by saying that it approximates the projector from below as a nonnegative operator: call this "weak enumeration". There is a simpler possible definition: let the program just list a sequence of orthogonal vectors that generate the subspace: call this "strong enumeration".

Remarks 3.2.

(1) The rest of the paper makes no use of the discussion of strong and week enumeration, so this part can be skipped.
(2) What is important is not only that the sequence of vectors in question can be enumerated, since this is in some sense trivially true for any finite sequence of elementary vectors. A recursively enumerable finite-dimensional subspace is always elementary. What matters is that the enumeration is done with a short program (which can use the dimension N as input). Without this remark, there is clearly no difference between an elementary subspace and a strongly enumerable one.

♢

Proposition 3.3. *The strong and weak kinds of enumeration of a subspace are equivalent. In other words, there is a program of length k enumerating a subspace in the weak sense if and only if there is a program of length $\stackrel{+}{=} k$ enumerating it in the strong sense.*

Proof. Given a strong enumeration $|\phi_1\rangle, |\phi_2\rangle, \dots$, the sum $\sum_i |\phi_i\rangle\langle\phi_i|$ clearly defines the projector in a form from which the possiblity of approximating it from below is seen.

Assume now that P is a projector and $\rho_1 \leqslant \rho_2 \leqslant \cdots$ is a sequence of elementary nonnegative operators approximating it.

Note that for a nonnegative operator A, we have $\langle\psi|A|\psi\rangle = 0$ iff $A|\psi\rangle = 0$. Now for any of the ρ_i, and any vector $|\psi\rangle$, if $P|\psi\rangle = 0$ then $\langle\psi|P|\psi\rangle = 0$, which implies $\langle\psi|\rho_i|\psi\rangle = 0$ and thus $\rho_i|\psi\rangle = 0$. Hence the kernel of ρ_i contains the kernel of P and hence the space of eigenvectors of ρ_i with nonnegative eigenvalues is contained in $P\mathcal{H}$. This shows that from ρ_i, $i = 1, 2, \dots$ we will be able to build up a sequence $|\phi_1\rangle, |\phi_2\rangle, \dots$ of orthogonal vectors spanning $P\mathcal{H}$. □

Theorem 7 below is analogous to the simple lower bound on classical description complexity. That lower bound says that the number of objects x with $K(x) < k$ is at most 2^k. What corresponds here to "number of objects" is dimension, and the statement is approximate: if $|\psi\rangle$ has complexity $< k$ then it is within a small angle from a certain fixed 2^{k+1}-dimensional space. The angle is really small for $\overline{H}$; it is not so small for $\underline{H}$ but it is still small enough that the whole domain within that angle makes up only a small portion of the Hilbert space.

Let $|u_1\rangle, |u_2\rangle, \dots$ be the sequence of eigenvectors of $\boldsymbol{\mu}$ with eigenvalues $\mu_1 \geqslant \mu_2 \geqslant \cdots$. (Since our space is finite-dimensional, the sequence exists.) Let $\kappa_i = -\log \mu_i$. Let E_k be the projector to the subspace generated by $|u_1\rangle, \dots, |u_k\rangle$.

Remark 3.4. The universal density matrix $\boldsymbol{\mu}$ is an object with an impressive invariance property: for any other universal density matrix ν we have $\nu \stackrel{*}{=} \boldsymbol{\mu}$. On the other hand, the individual eigenvectors $|u_i\rangle$ probably do not have any invariant significance. It is currently not clear whether even the projectors E_k enjoy any approximate invariance property. ♢

Theorem 7 (Lower bounds)**.** *Let $|\psi\rangle$ be any vector and let $\lambda > 1$. If $\overline{H}(|\psi\rangle) < k$ then we have*

$$\langle\psi|E_{2^{\lambda k}}|\psi\rangle > 1 - 1/\lambda. \tag{3.4}$$

If $\underline{H}(|\psi\rangle) < k$ *then we have*

$$\langle\psi|E_{\lambda 2^k}|\psi\rangle > 2^{-k}(1-1/\lambda). \tag{3.5}$$

Proof. Assume $\overline{H}(|\psi\rangle) < k$ and expand $|\psi\rangle$ in the basis $\{|u_i\rangle\}$ as $|\psi\rangle = \sum_i c_i|u_i\rangle$. By the assumption, we have $\sum_i \kappa_i|c_i|^2 < k$. Let m be the first i with $\kappa_i > \lambda k$. Since $\sum_i 2^{-\kappa_i} < 1$ we have $m \leqslant 2^{\lambda k}$. Also,

$$\lambda k \sum_{i\geqslant m} |c_i|^2 < \sum_{i\geqslant m} \kappa_i|c_i|^2 < k,$$

hence $\sum_{i\geqslant m}|c_i|^2 < 1/\lambda$, which proves (3.4).

Now assume $\underline{H}(|\psi\rangle) < k$, then we have $\sum_i \mu_i|c_i|^2 \geqslant 2^{-k}$. Let m be the first i with $\mu_i < 2^{-k}/\lambda$. Since $\sum_i \mu_i < 1$ we have $m \leqslant 2^k\lambda$. Also,

$$\sum_{i\geqslant m} \mu_i|c_i|^2 < 2^{-k}/\lambda \sum_i |c_i|^2 = 2^{-k}/\lambda,$$

hence

$$\begin{aligned}\langle\psi|E_m|\psi\rangle = \sum_{i<m}|c_i|^2 &> \sum_{i<m}\mu_i|c_i|^2 \geqslant 2^{-k} - \sum_{i\geqslant m}\mu_i|c_i|^2 \\ &> 2^{-k}(1-1/\lambda).\end{aligned} \tag{3.6}$$

□

The defect of this theorem is that the operators E_k are uncomputable. I do not know whether the above properties can be claimed for some lower semicomputable operators F_k.

3.3. **Quantum description complexities.**

3.3.1. *Vitányi's complexity.* Theorem 8 says that the complexity Kq from [6], (defined in Section 1) is not too much larger than $\underline{H}$, so we do not lose too much in replacing the sum (2.3) with a supremum: if the sum is $> 2^{-k}$ then the supremum is $> 2^{-4k}/k^2$.

Theorem 8 (Relation to Kq)**.**

$$\underline{H} \overset{+}{<} \mathrm{Kq} \overset{+}{<} 4\underline{H} + 2\log\underline{H}. \tag{3.7}$$

Proof. We start from the end of the proof of Theorem 7. We use (3.6) with $\lambda = 2$, and note that one term, say, $|c_r|^2$ of the sum $\sum_{i\leqslant m}|c_i|^2$ must be at least 2^{-2k-2}. We would be done if we could upperbound $K(|u_r\rangle)$ appropriately. It would seem that $K(|u_r\rangle)$ can be bounded approximately by k since $m \leqslant 2^{k+1}$. But unfortunately, neither the vectors $|u_i\rangle$ nor their sequence are computable; so, an approximation is needed. Let r be the largest binary number of length $\leqslant k$ smaller than $\operatorname{Tr}\boldsymbol{\mu}$. Then there is a program p of length $\leqslant k + 2\log k$ computing a lower approximation $\hat{\boldsymbol{\mu}}$ of $\boldsymbol{\mu}$ such that $\operatorname{Tr}\boldsymbol{\mu} - \operatorname{Tr}\hat{\boldsymbol{\mu}} \leqslant 2^{-k}$. Indeed, let p specify the binary digits of r and then compute an approximation of $\operatorname{Tr}\boldsymbol{\mu}$ that exceeds r.

The condition $\langle\psi|\boldsymbol{\mu}|\psi\rangle \geqslant 2^{-k}$ implies $\langle\psi|\hat{\boldsymbol{\mu}}|\psi\rangle \geqslant 2^{-k+1}$. We can now proceed with $\hat{\boldsymbol{\mu}}$ as with $\boldsymbol{\mu}$. We compute eigenvectors $|\hat{u}_i\rangle$ for $\hat{\boldsymbol{\mu}}$, and find an elementary vector $|\hat{u}_r\rangle$ with

$$K(|\hat{u}_r\rangle) \overset{+}{<} 2k + 2\log k, \quad |\langle\psi|\hat{u}_r\rangle|^2 \overset{*}{>} 2^{-2k}.$$

The extra $k + 2\log k$ in $K(|\hat{u}_r\rangle)$ is coming from the program p above. □

3.3.2. *Qubit complexity.* Let us define the qubit complexity introduced in [1]. We refer to that paper for further references on quantum Turing machines and detailed specifications of the quantum Turing machine used. Our machine starts from an input (on the input tape) consisting of a qubit program and a rational number $\varepsilon > 0$. On the output tape, an output appears, preceded by a 0/1 symbol telling whether the machine is considered halted. The halting symbol as well as the content of the output tape does not change after the halting symbol turns 1. (The input tape, which is also the work tape, keeps changing.) We can assume that input and output strings of different lengths can always be padded to the same length at the end by 0's, or if this is inconvenient, by some special "blank", or "vacuum" symbol. The input of the machine is a density matrix ρ. For any segment of some length n of the output, and any given time t there is a completely positive operator $\Phi_{k,t}$ such that the n symbols of the output at time t are described by a density matrix $\sigma = \Phi_{k,t}\rho$. We only want to consider the output state when the machine halted. If H is a projection to the set of those states then the semi-density matrix $H\sigma H$ is the output we are interested in. The operation $\Psi_{n,t} : \rho \mapsto H\sigma H$ is a completely positive operator but it is not trace-preserving, it may decrease the trace. It is also monotonically increasing in t.

For a state $|\psi\rangle$, let $\mathrm{QC}^{\varepsilon}(|\psi\rangle)$ be the length k of the smallest qubit program (an arbitrary state in $\mathcal{Q}_k$, or more precisely the density matrix corresponding to this pure state) which, when given as input along with ε, results in an output density matrix σ with $\langle\psi|\sigma|\psi\rangle \geqslant 1-\varepsilon$. The paper [1] shows that this quantity has the same machine-independence properties as Kolmogorov complexity, so we also assume that a suitable universal quantum Turing machine has been fixed. For the following theorem, we will compute complexities of strings in $\mathcal{H}_N = \mathcal{Q}_n$, so $N = 2^n$.

Lemma 3.5. *If for a semi-density matrix ρ and a state $|\psi\rangle$ we have $\langle\psi|\rho|\psi\rangle \geqslant 1-\varepsilon$ and ρ has the eigenvalue decomposition $\sum_i p_i|i\rangle\langle i|$ where $p_1 \geqslant p_2 \geqslant \cdots$, then*

$$p_1 \geqslant 1-\varepsilon, \quad |\langle 1|\psi\rangle|^2 \geqslant 1-2\varepsilon.$$

Proof. Let $c_i = \langle i|\psi\rangle$, then $\langle\psi|\rho|\psi\rangle = \sum_i p_i|c_i^2| \geqslant 1-\varepsilon$. Hence $p_1 \geqslant 1-\varepsilon$, therefore

$$|c_1|^2 + \varepsilon \geqslant \sum_i p_i|c_i^2| \geqslant 1-\varepsilon,$$

giving $|c_1^2| \geqslant 1-2\varepsilon$. □

Theorem 9. *For $\varepsilon < 0.5$, if $\mathrm{QC}^{\varepsilon}(|\psi\rangle) \leqslant k$ then*

$$\overline{H}(|\psi\rangle) \stackrel{+}{<} k + K(k) + 2\varepsilon n.$$

Proof. For each k, let I_k be the projection to the space $\mathcal{Q}_k$ of k-length inputs. The operator

$$\lambda = \sum_k \mathbf{m}(k)2^{-k}I_k$$

is a semicomputable semi-density matrix on the set of all inputs. For each time t, the semi-density matrix $\Psi_{n,t}\lambda$ is semicomputable. As it is increasing in t, the limit $\nu = \lim_t \Psi_{n,t}\lambda$ is a semicomputable semi-density matrix, and therefore $\nu \stackrel{*}{<} \boldsymbol{\mu}$. Let $|\phi\rangle \in \mathcal{Q}_k$, then $|\phi\rangle\langle\phi| \leqslant I_k$, hence $\mathbf{m}(k)2^{-k}|\phi\rangle\langle\phi| \leqslant \lambda$, hence for each t we have

$$\mathbf{m}(k)2^{-k}\Psi_{t,k}|\psi\rangle\langle\psi| \leqslant \nu \stackrel{*}{<} \boldsymbol{\mu}.$$

Since also $2^{-n}I_n \stackrel{*}{<} \boldsymbol{\mu}$, we can assert, with $\rho_{t,k} = \Psi_{t,k}|\phi\rangle\langle\phi|$, that

$$\sigma = \mathbf{m}(k)2^{-k}\rho_{t,k} + 2^{-n}I_n \stackrel{*}{<} \boldsymbol{\mu}.$$

Assume that $\langle\psi|\rho_{t,k}|\psi\rangle \geqslant 1-\varepsilon$. Then by Lemma 3.5, if $\rho_{t,k}$ has the eigenvalue decomposition $\sum_i p_i|i\rangle\langle i|$ then $p_1 \geqslant 1-\varepsilon$ and $|\langle 1|\psi\rangle|^2 \geqslant 1-2\varepsilon$. The matrix $(-\log\sigma)$ can be written as

$$-\sum_i \log(\mathbf{m}(k)2^{-k}p_i + 2^{-n})|i\rangle\langle i|.$$

Hence, with $c_i = \langle i|\psi\rangle$, and using Lemma 3.5 and $\varepsilon < 0.5$

$$\begin{aligned} -\langle\psi|\log\boldsymbol{\mu}|\psi\rangle &\stackrel{+}{<} -\langle\psi|\log\sigma|\psi\rangle \\ &= \sum_i \log(\mathbf{m}(k)2^{-k}p_i + 2^{-n})|c_i|^2 \\ &\leqslant k + K(k) + \log(1-\varepsilon) + 2\varepsilon n. \end{aligned}$$

In the last inequality, the first two terms come from the first term of the previous sum, while $2\varepsilon n$ comes from the rest of the terms. □

Using the definitions of [1], we write $\mathrm{QC}(|\psi\rangle) \leqslant k$ if there is a $|\phi\rangle$ such that for all ε of the form $1/m$, when $|\phi\rangle$ is given as input along with ε, we get an output density matrix σ with $\langle\psi|\sigma|\psi\rangle \geqslant 1-\varepsilon$. The above theorem implies that in this case,

$$\overline{H}(|\psi\rangle) \stackrel{+}{<} k + K(k). \tag{3.8}$$

Let x be a bit string, then we know from (3.1) that

$$\overline{H}(|x\rangle) \stackrel{+}{=} K(x). \tag{3.9}$$

It has been shown in [1] that $\mathrm{QC}(|x\rangle) \stackrel{+}{<} C(x)$ where $C(x)$ is the (not prefix-free) Kolmogorov complexity. We can show directly that also $C(x) \stackrel{+}{<} \mathrm{QC}(|x\rangle)$, but we will not do it in this paper. It follows from (3.8) and (3.9) that $K(x) \stackrel{+}{=} \overline{H}(|x\rangle) \stackrel{+}{<} \mathrm{QC}(|x\rangle) + K(\mathrm{QC}(|x\rangle))$. This is in some way stronger, since another interesting quantity, $\overline{H}(|x\rangle)$ is interpolated, and in another way it seems slightly weaker. But only very slightly, since one can bound $K(x)$ by $C(x)$ in general only via $K(x) \stackrel{+}{<} C(x) + K(C(x))$.

Just as we obtained an upper bound on Kq using (3.5) combined with an approximation of the uncomputable $\boldsymbol{\mu}$, we may hope to obtain an upper bound on QC using (3.4) combined with a suitable approximation of the uncomputable $\boldsymbol{\mu}$ or $(-\log\boldsymbol{\mu})$. But we did not find an approximation in this case for a reasonable price in complexity: the best we can say replaces $\overline{H}(|\psi\rangle)$ with $\langle\psi|(-\log\mu)|\psi\rangle$ for any computable density matrix μ. Or, we can upperbound not $\mathrm{QC}(|\psi\rangle)$ but $\mathrm{QC}(|\psi\rangle \mid \chi)$ where χ is an encoding of the halting problem into a suitable infinite binary string. The concept of an oracle quantum computation with a read-only classical oracle tape presents no difficulties.

Theorem 10. *For each rational ε and any computable density matrix μ we have*

$$\mathrm{QC}^{\varepsilon}(|\psi\rangle) \stackrel{+}{<} \langle\psi|(-\log\mu)|\psi\rangle/\varepsilon + K(\mu).$$

Similarly,

$$\mathrm{QC}^{\varepsilon}(|\psi\rangle \mid \chi) \stackrel{+}{<} \overline{H}(|\psi\rangle)/\varepsilon.$$

Proof. For the second inequality, we can use (3.4) with $k = \overline{H}(|\psi\rangle)$ and $\lambda = 1/\varepsilon$. The oracle χ allows us to compute the space $E_{2^{\lambda k}}$ with arbitrary precision. Then our quantum Turning machine can simply map the space of λk-length qubit strings into the (approximate) $E_{2^{\lambda k}}$.

Similarly, for the first inequality, if μ is computable then we can compute the subspaces corresponding to $E_{2^{\lambda k}}$ with arbitrary precision. □

3.4. **Invariance under computable transformations.**

Theorem 11. *Let U be any computable unitary transformation. Then we have*

$$\overline{H}(U|\psi\rangle) \stackrel{+}{=} \overline{H}(|\psi\rangle), \quad \underline{H}(U|\psi\rangle) \stackrel{+}{=} \underline{H}(|\psi\rangle).$$

Proof. Straightforward. □

This theorem needs to be generalized: it should be understood how complexity changes under a completely positive operator.

4. Complexity and entropy

In classical algorithmic information theory, if ρ is a discrete computable probability distribution then its entropy is equal, to a good approximation, to the average complexity. In the quantum case, entropy is defined as

$$S(\rho) = -\operatorname{Tr}\rho\log\rho.$$

There is a quantity corresponding to the Kullback information distance, and called *relative entropy* in [7]: it is defined as

$$S(\rho \parallel \sigma) = \operatorname{Tr}\rho(\log\rho - \log\sigma),$$

where ρ and σ are density matrices.

Proposition 4.1.

$$S(\rho \parallel \sigma) \geqslant 0. \tag{4.1}$$

Proof. See [7]. □

The following theorem can be interpreted as saying that entropy is equal to average complexity:

Theorem 12. *For any lower semicomputable semi-density matrix ρ we have*

$$S(\rho) \stackrel{+}{=} \operatorname{Tr}\rho\boldsymbol{\kappa} \tag{4.2}$$

Proof. Let $\Omega = \operatorname{Tr}\boldsymbol{\mu}$, then $\sigma = \boldsymbol{\mu}/\Omega$ is a density matrix, and hence by (4.1), $S(\rho \parallel \sigma) \geqslant 0$. It follows that $S(\rho) \stackrel{+}{<} \operatorname{Tr}\rho\boldsymbol{\kappa}$.

On the other hand, since $\rho \stackrel{*}{<} \boldsymbol{\mu}$, the monotonicity of logarithm gives $\boldsymbol{\kappa} \stackrel{+}{<} -\log\rho$ which gives the other inequality. □

For what follows the following property of logarithm is useful:

Lemma 4.2. *If A and B are nonnegative operators over X and Y respectively, then*

$$\log A\otimes B = (\log A)\otimes\mathbf{1}_Y + \mathbf{1}_X\otimes(\log B). \tag{4.3}$$

Proof. Direct computation. □

Some properties of complexity that can be deduced from its universal probability formulation will carry over to the quantum form. As an example, take subadditivity:

$$K(x,y) \stackrel{+}{<} K(x) + K(y).$$

What corresponds to this in the quantum formulation is the following:

Theorem 13 (Subadditivity)**.** *We have*

$$\boldsymbol{\mu}_X\otimes\boldsymbol{\mu}_Y \stackrel{*}{<} \boldsymbol{\mu}_{XY}. \tag{4.4}$$

For $|\phi\rangle, |\psi\rangle \in \mathcal{H}_N$ and $H = \overline{H}$ or $\underline{H}$ we have

$$H(|\phi\rangle|\psi\rangle) \stackrel{+}{<} H(|\phi\rangle) + H(|\psi\rangle). \tag{4.5}$$

Proof. The density matrix $\boldsymbol{\mu}_X\otimes\boldsymbol{\mu}_Y$ over the space $\mathcal{H}_{XY} = \mathcal{H}_X\otimes\mathcal{H}_Y$ is lower semicomputable, therefore (4.4) follows. Hence

$$\begin{aligned}(\langle\phi|\boldsymbol{\mu}_X|\phi\rangle)(\langle\psi|\boldsymbol{\mu}_Y|\psi\rangle) &= \langle\phi|\langle\psi|(\boldsymbol{\mu}_X\otimes\boldsymbol{\mu}_Y)|\phi\rangle|\psi\rangle\\ &\stackrel{*}{<} \langle\phi|\langle\psi|\boldsymbol{\mu}_{XY}|\phi\rangle|\psi\rangle.\end{aligned}$$

which gives (4.5) for $H = \underline{H}$. For $H = \overline{H}$ note that by the monotonicity of logarithm, identity (4.3) and (4.4) implies

$$(\log\boldsymbol{\mu}_X)\otimes\mathbf{1}_Y + \mathbf{1}_X\otimes(\log\boldsymbol{\mu}_Y) = \log\boldsymbol{\mu}_X\otimes\boldsymbol{\mu}_Y \stackrel{+}{<} \log\boldsymbol{\mu}_{XY}.$$

Taking the expectation (multiplying by $\langle\psi|$ on left and $|\psi\rangle$ on right) gives the desired result. □

The analogous subadditivity property also holds for the quantum entropy $S(\rho)$.

For classical complexity we have $K(x) \stackrel{+}{<} K(x,y)$, and the corresponding property also holds for classical entropy. This monotonicity property can also be proved for quantum complexity.

Theorem 14 (Monotonicity)**.** *We have*

$$\operatorname{Tr}_Y \boldsymbol{\mu}_{XY} \stackrel{*}{=} \boldsymbol{\mu}_X, \tag{4.6}$$

$$\boldsymbol{\kappa}_{XY} \stackrel{+}{>} \boldsymbol{\kappa}_X \otimes \mathbf{1}_Y. \tag{4.7}$$

For $|\phi\rangle, |\psi\rangle \in \mathcal{H}_N$, and $H = \overline{H}$ or $\underline{H}$ we have

$$H(|\phi\rangle) \stackrel{+}{<} H(|\phi\rangle|\psi\rangle). \tag{4.8}$$

Proof. Let $\rho_X = \operatorname{Tr}_Y \boldsymbol{\mu}_{XY}$. Then ρ_X is a semicomputable semi-density matrix over $\mathcal{H}_X$ and thus $\rho_X \stackrel{*}{<} \boldsymbol{\mu}_X$. At the same time, for any fixed vector $|\psi\rangle$, the matrix $\sigma_{XY} = \boldsymbol{\mu}_X \otimes |\psi\rangle\langle\psi|$ is a lower semicomputable semi-density matrix, hence $\boldsymbol{\mu}_{XY} \stackrel{*}{>} \sigma_{XY}$. Taking the partial trace gives

$$\boldsymbol{\mu}_X = \operatorname{Tr}_Y \sigma_{XY} \stackrel{*}{<} \operatorname{Tr}_Y \boldsymbol{\mu}_{XY} = \rho_X.$$

This proves (4.6), which implies the inequality for $\underline{H}$.

Let $\{|\psi_i\rangle\}$ be any orthogonal basis of $\mathcal{H}_Y$ with $|\psi_1\rangle = |\psi\rangle$. Then we have

$$\begin{aligned}\langle\phi|\langle\psi|\boldsymbol{\mu}_X \otimes \mathbf{1}_Y|\phi\rangle|\psi\rangle &= \langle\phi|\boldsymbol{\mu}_X|\phi\rangle \\ &\stackrel{*}{=} \langle\phi|\operatorname{Tr}_Y \boldsymbol{\mu}_{XY}|\phi\rangle = \sum_i \langle\phi|\langle\psi_i|\boldsymbol{\mu}_{XY}|\phi\rangle|\psi_i\rangle \geqslant \langle\phi|\langle\psi|\boldsymbol{\mu}_{XY}|\phi\rangle|\psi\rangle,\end{aligned}$$

which proves $\boldsymbol{\mu}_X \otimes \mathbf{1}_Y \stackrel{*}{>} \boldsymbol{\mu}_{XY}$. Taking logarithms and noting that $\log \mathbf{1}_Y = 0$, we get (4.7) which proves the inequality for $\overline{H}$. □

The quantum entropy analog of this monotonicity fails in a spectacular way. It is not true in general that $S(\rho_X) \leqslant S(\rho_{XY})$. Indeed, ρ_{XY} could be the density matrix of a pure state, and then $S(\rho_{XY}) = 0$. At the same time, if this pure state is an entangled state, a state that cannot be represented in the form of $|\phi\rangle|\psi\rangle$, only as the linear combination of such states, then $S(\rho_X) > 0$. This paradox does not contradict to the possibility that entropy is "average complexity". It just reminds us that Theorem 14 says nothing about entangled states. An entangled state can be simple even if it is a big sum, but in this case it will contain a lot of complex components.

5. The cloning problem

5.1. Maximal complexity of cloned states.

For classical description complexity, the relation

$$K(x,x) \stackrel{+}{=} K(x)$$

holds and is to be expected: once we have x we can copy it and get the pair (x,x). But there is a "no cloning theorem" [4] in quantum mechanics saying that there is no physical way to get $|\psi\rangle|\psi\rangle$ from $|\psi\rangle$. It is interesting to see that a much stronger form of this theorem also holds, saying that sometimes $\overline{H}(|\psi\rangle|\psi\rangle)$ is much larger than $\overline{H}(|\psi\rangle)$ (of course, at most twice as large). Moreover, we can determine the maximum complexity of states of the form $|\psi\rangle^{\otimes k}$. Our results in this are very similar in form to those of [1], and the proof method is also similar.

For $|\psi\rangle \in \mathcal{H}_N$, let $|\psi\rangle^{\otimes m}$ denote the m-fold tensor product of $|\psi\rangle$ with itself, an element of $\mathcal{H}^{\otimes m}$.

Let

$$\mathcal{S}_{N,m} = \mathcal{H}^{\vee m} \subset \mathcal{H}_N^{\otimes m}$$

be the subspace of elements of $\mathcal{H}_N^{\otimes m}$ invariant under the orthogonal transformations arising from the permutations

$$|\phi_1\rangle \dots |\phi_m\rangle \mapsto |\phi_{\pi(1)}\rangle \dots |\phi_{\pi(m)}\rangle.$$

Lemma 5.1 (see [8])**.**

(1) $\dim \mathcal{S}_{N,m} = \binom{m+N-1}{m}$.
(2) *$\mathcal{S}_{N,m}$ is invariant under unitary transformations of the form $U^{\otimes m}$.*
(3) *If a density matrix over $\mathcal{S}_{N,m}$ commutes with all such transformations then it is a multiple of unity.*

Let

$$\overline{C}_{N,m} = \max_{|\psi\rangle \in \mathcal{H}_N} \overline{H}(|\psi\rangle^{\otimes m}), \tag{5.1}$$

and let $\underline{C}_{N,m}$ be defined the same way with $\underline{H}$ in place of $\overline{H}$.

Theorem 15. *We have*

$$\overline{C}_{N,m} \stackrel{+}{<} K(m) + \log \binom{m+N-1}{m},$$
$$\underline{C}_{N,m} \geqslant \log \binom{m+N-1}{m}.$$

Proof. The upper bound follows from the fact that $|\psi\rangle \in \mathcal{S}_{N,m}$ and from (3.3).

For simplicity, let us write for the moment, $|\psi\rangle^m = |\psi\rangle^{\otimes m}$. For the lower bound, let us first set $c = \underline{C}_{N,m}$. We have

$$\operatorname{Tr} \boldsymbol{\mu} |\psi\rangle^m \langle\psi|^m = \langle\psi|^m \boldsymbol{\mu} |\psi\rangle^m \geqslant 2^{-c} \tag{5.2}$$

for all states $|\psi\rangle \in \mathcal{H}_N$. Let P_S be the projection to $\mathcal{S}_{N,m}$. Let Λ be the uniform distribution on the unit sphere in $\mathcal{H}_N$. Then

$$\rho = \int |\psi\rangle^m \langle\psi|^m \, d\Lambda$$

is a density matrix over $\mathcal{S}_{N,m}$. It commutes with all unitary transformations of the form $U^{\otimes m}$, and therefore according to Lemma 5.1,

$$\rho = \binom{m+N-1}{m}^{-1} P_S.$$

Integrating (5.2) by $d\Lambda$ we get

$$2^{-c} \leqslant \operatorname{Tr} \boldsymbol{\mu} \rho = \binom{m+N-1}{m}^{-1} \operatorname{Tr} \boldsymbol{\mu} P_S \leqslant \binom{m+N-1}{m}^{-1}.$$

Taking negative logarithm, we get the lower bound on $\underline{C}$. □

5.2. **An algebraic consequence.** This subsection says nothing new about quantum complexities, it only draws some technical inferences from the previous subsection.

The problem of estimating $\overline{H}(|\psi\rangle|\psi\rangle)$ can be reformulated into an algebraic problem for which we are not aware of any previous solution. The results obtained above solve the problem: maybe such a solution will also have some independent interest. For any $N \times N$ matrix A, let

$$u(A) = \frac{\|A^\dagger A\|}{\operatorname{Tr} A^\dagger A} = \max_i \frac{\alpha_i}{\sum_j \alpha_j}$$

where α_j are the eigenvalues of $A^\dagger A$. The function $u(A)$ measures the "unevenness" of the distribution of eigenvalues of $A^\dagger A$. It can vary between $1/N$ for $A = \mathbf{1}$ and 1 (when $A^\dagger A$ has rank 1). For a subspace F of the vector space of symmetric (not necessarily self-adjoint!) matrices, let $u(F) = \max_{A \in F} u(A)$. Let $N' = N(N+1)/2$. For $0 < d < N'$, we are interested in the quantity

$$u(d, N) = \min\{\, u(F) : \dim F \geqslant d \,\}.$$

Theorem 16. *We have $u(d, N) \geqslant d/N'$.*

Remark 5.2. This theorem has been strengthened from its preprint version. ♢

Before the proof, we give some lemmas setting up the connection with cloning.

Lemma 5.3. *Let A be a symmetric $N \times N$ matrix (a_{ij}) and let*

$$\alpha = \sum_{ij} a^*_{ij} |\beta_i\rangle|\beta_j\rangle.$$

Then

$$\sup_{|\phi\rangle \in \mathcal{H}_N} |\langle\alpha|(|\phi\rangle|\phi\rangle)|^2 = u(A). \tag{5.3}$$

Proof. We can restrict ourselves to matrices A with $\mathrm{Tr}\, A^\dagger A = \langle\alpha|\alpha\rangle = 1$. Then with $|\psi\rangle = |\phi\rangle|\phi\rangle$, $|\phi\rangle = \sum_i x_i|\beta_i\rangle$,

$$|\langle\alpha|\psi\rangle|^2 = |\sum_{ij} a_{ij}x_ix_j|^2 = |x^T A x|^2,$$

where x^T is the transpose of x (without conjugation).

By singular value decomposition (see [2]), every matrix can be written in the form VDU where D is a nonnegative diagonal matrix and U, V are unitary transformations. If the elements of D are all distinct, positive and in decreasing order then U, V are unique. In this case, clearly if A is symmetric then $V = U^T$. This can be generalized to the case when the elements of D are not all positive and distinct, using for example limits. Thus, $A = U^T DU$. This gives $x^T Ax = x^T U^T DUx = (Ux)^T D(Ux)$. As x runs through all possible vectors with $\sum_i |x_i|^2 = 1$, so does Ux. Let d_1 be the largest element on the diagonal of D, then $d_1^2 = \|A^\dagger A\|$.

$$|(Ux)^T DUx| = |\sum_i d_i (Ux)_i^2| \leqslant \sum_i d_i |(Ux)_i|^2 \leqslant d_1$$

since $\sum_i |(Ux)_i|^2 = 1$. The maximum of $|(Ux)^T D(Ux)|^2$ is achieved by the element $x = U^{-1}|\beta_1\rangle$, and then it is $d_1^2 = u(A)$. □

Lemma 5.4. *For $0 < d < N'$, there is a computable semi-density matrix ρ with*

$$\sup_{|\psi\rangle = |\phi\rangle|\phi\rangle} -\log\langle\psi|\rho|\psi\rangle \leqslant \log(N' - d) - \log(1 - u(d, N)).$$

Proof. Using the notation of Lemma 5.3, let F be the subspace of dimension d of vectors α on which the minimum $u(d, N)$ is achieved. Witn $P = 1 - F$, let ρ be the semi-density matrix defined in the proof of Theorem 6. Similarly to (3.2) we have, for any $\psi = |\phi\rangle|\phi\rangle$:

$$-\log\langle\psi|\rho|\psi\rangle \leqslant \log(N' - d) - \log(1 - \langle\psi|F|\psi\rangle).$$

Note that $\langle\psi|F|\psi\rangle = |\langle\alpha|\psi\rangle|^2$ for some $\alpha \in F$, hence by (5.3) we have $\langle\psi|F|\psi\rangle \leqslant u$, hence the last term of the right-hand side is $\leqslant -\log(1 - u)$. □

Proof of Theorem 16. The reasoning of Theorem 15 implies that $\log N'$ lower-bounds the left-hand side in the above lemma. Thus,

$$\log N' \leqslant \log(1 - d/N') + \log N' - \log(1 - u),$$
$$u \geqslant d/N'.$$

□

6. Randomness tests

6.1. **Universal tests.** In classical algorithmic information theory (see for example [3]), description complexity helps clarify what experimental outcomes should be called random with respect to a hypothetical probability distribution. If the set of possible outcomes is a discrete one, say the set of natural numbers, then, given a probability distribution ν, we call a lower semicomputable function $f(x)$ a *randomness test* if $\sum_x f(x)\nu(x) \leqslant 1$. It is known that there is a universal test $t_\nu(x)$, a test

that dominates all other tests to within a multiplicative constant. An outcome is considered non-random with respect to ν when $t_\nu(x)$ is large. In case of a computable distribution ν, we have

$$t_\nu(x) \stackrel{*}{=} \frac{\mathbf{m}(x)}{\nu(x)}, \tag{6.1}$$

where the multiplicative constant in the $\stackrel{*}{=}$ depends on ν. (The general case is more complicated.) The deficiency of randomness is defined as $d_\nu(x) = \log t_\nu(x)$. In case of a computable distribution ν it is known to be

$$\stackrel{+}{=} -\log \nu(x) + \log \mathbf{m}(x) \stackrel{+}{=} -\log \nu(x) - K(x) \tag{6.2}$$

Thus, for a computable distribution, the universal test measures the difference betwen the logarithm of the probability and the complexity.

In the quantum setting, what corresponds to a probability distribution is a computable density matrix ρ. What corresponds to a function is a self-adjoint operator. So, let us say that a *randomness test* is a lower semicomputable self-adjoint operator F_ρ with

$$\operatorname{Tr} F_\rho \rho \leqslant 1.$$

Remark 6.1. In the theorem below, the expression

$$T' = \rho^{-1/2} \boldsymbol{\mu} \rho^{-1/2}$$

appears, which does not make sense if ρ is not invertible. However, let us write $\sigma = \mu^{1/2}\rho^{-1/2}$; this expression makes sense on the subspace V orthogonal to the kernel of ρ, and therefore $T' = \sigma^\dagger \sigma$ also makes sense there. Therefore we define $\langle \psi | T' | \psi \rangle$ as ∞ for any $|\psi\rangle \notin V$, and there is no problem for $|\psi\rangle \in V$. ♢

Theorem 17 (Universal test)**.** *There is a test T_ρ which is universal in the sense that it dominates each other test R: we have $R \stackrel{*}{<} T_\rho$, where the multiplicative constant in $\stackrel{*}{<}$ may depend on R and ρ. We have $T_\rho \stackrel{*}{=} T'_\rho \stackrel{*}{=} T''_\rho$ where*

$$T'_\rho = \sum_{|\phi\rangle} \frac{\mathbf{m}(|\phi\rangle)|\phi\rangle\langle\phi|}{\langle\phi|\rho|\phi\rangle},$$
$$T''_\rho = \rho^{-1/2} \boldsymbol{\mu} \rho^{-1/2}.$$

Proof. The proof of the existence of a universal test is similar to the proof of Proposition 2.1. The proof of $T \stackrel{*}{=} T'$ is similar to the one showing $\boldsymbol{\mu}' \stackrel{*}{=} \boldsymbol{\mu}$ in Theorem 2.

Let us prove $T \stackrel{*}{=} T''$. To see that T'' is lower semicomputable, note that as direct computation shows, for any operator C the function $A \mapsto C^\dagger A C$ is monotonic on the set of self-adjoint operators A with respect to the relation $\leqslant$. By the cyclic property of the trace, we also have $\operatorname{Tr} T''\rho = \operatorname{Tr} \boldsymbol{\mu} \leqslant 1$. This proves $T'' \stackrel{+}{<} T$, it remains to prove that $T \stackrel{*}{<} T''$. This is equivalent to

$$\rho^{1/2} T \rho^{1/2} \leqslant \rho^{1/2} T'' \rho^{1/2} = \boldsymbol{\mu}.$$

But the left-hand side is a lower semicomputable nonnegative definite matrix whose trace is $\leqslant 1$, again due to the cyclic property of trace. Therefore by the defining property of $\boldsymbol{\mu}$, it is $\stackrel{*}{<} \boldsymbol{\mu}$. □

The expression for T''_ρ is similar to (6.1), but it does not separate the roles of the density matrix ρ and of the universal probability $\boldsymbol{\mu}$ as neatly, certainly not in the typical cases when $\boldsymbol{\mu}$ and ρ do not commute. Assume that the eigenvalues of ρ are $p_1 \geqslant p_2 \geqslant \cdots$, with the corresponding eigenvectors $|v_i\rangle$ (these exist since our space is finite-dimensional). Let (m_{ij}) be the matrix of the operator $\boldsymbol{\mu}$ when expressed in this basis. For a certain state $|\psi\rangle = \sum_i c_i |v_i\rangle$, we can express the value of the test on

$|\psi\rangle$ as follows. If there is any i with $p_i = 0$ and $c_i \neq 0$ then according to Remark 6.1, the value is ∞. Otherwise, it is

$$\langle\psi|T''_\rho|\psi\rangle = \sum_{i,j} m_{ij}(p_i p_j)^{-1/2} c_i^* c_j. \tag{6.3}$$

The term $(p_i p_j)^{-1/2} c_i^* c_j$ is defined to be 0 if $c_i^* c_j = 0$, and we excluded the case when $p_i p_j = 0$ but $c_i^* c_j \neq 0$. The roles of $\boldsymbol{\mu}$ and ρ do not seem to be separable in the same way as in the classical case. However, if ρ is the uniform distribution then the expression simplifies to

$$N^{-1} \sum_{i,j=1}^{N} m_{ij} c_i^* c_j = N^{-1} \langle\psi|\boldsymbol{\mu}|\psi\rangle,$$

which is the classical comparison of the probability to the universal probability.

6.2. **Relation to Martin-Löf tests.** The sum for T'_ρ in Theorem 17 is similar to $\boldsymbol{\mu}'$ in Theorem 2. In the classical case and with a computable ρ, just like there, it can be replaced with a supremum. In the quantum case it cannot: indeed, the expression of $\boldsymbol{\mu}'$ is a special case of T', and we have shown in Section 3 that the sum in $\boldsymbol{\mu}'$ cannot be replaced with supremum. We do not know whether there is still an approximate relation like in Theorem 8: the proof does not carry over.

It is worth generalizing the sum for T'_ρ as

$$\sum_F \frac{\mathbf{m}(F) F}{\operatorname{Tr} F\rho}$$

where F runs through all elementary nonnegative self-adjoint operators. An interesting kind of self-adjoint operator is a projection P to some subspace. Such a term looks like

$$\frac{\mathbf{m}(P)}{\operatorname{Tr} P\rho} P.$$

This term is analogous to a Martin-Löf test. An outcome x would be caught by a Martin-Löf test in the discrete classical case if it falls into some simple set S with small probability. The fact that S is simple means that $K(S)$ is small, in other words $\mathbf{m}(S)$ is large. Altogether, we can say that x is caught if the expression

$$\frac{\mathbf{m}(S)}{\rho(S)} 1_S(x)$$

is large, where $1_S(x)$ is the indicator function of the set S. In the quantum case, for state $|\psi\rangle$, what corresponds to this is the expression

$$\frac{\mathbf{m}(P)}{\operatorname{Tr} P\rho} \langle\psi|P|\psi\rangle.$$

The probability of S translates to $\operatorname{Tr} P\rho$, and $1_S(x)$ translates to $\langle\psi|P|\psi\rangle$. Thus, a quantum Martin-Löf test catches a state $|\psi\rangle$ if it is "not sufficiently orthogonal" to some simple low-probability subspace. Compare this with Theorem 6.

As we see, the universal quantum randomness test contains the natural generalizations of the classical randomness tests, but on account of the possible non-commutativity between ρ and $\boldsymbol{\mu}$, it may also test $|\psi\rangle$ in some new ways that do not correspond to anything classical. It would be interesting to find what these ways are.

7. Proof of Theorem 1

Let us denote

$$K_m(|\psi\rangle) = \min\{\, l(p) : U(p) = |\phi\rangle,\ -\log|\langle\phi|\psi\rangle|^2 \leqslant m \,\}.$$

The first lemma lowerbounds $K_\infty(|\psi\rangle)$, the later ones lowerbound $K_m(|\psi\rangle)$ for finite m.

Lemma 7.1. *For each k there is a subspace V of $\mathcal{Q}_n$, of dimension $2^n - 2^k$ with the property that for all $|\psi\rangle \in V$ we have $K_\infty(|\psi\rangle) \geqslant k$.*

Proof. Let $p_1, \dots, p_r$ be all programs of length $< k$ for which $U(p_m) \in \mathcal{Q}_n$. Then $r < 2^k$. Let V be the set of elements of $\mathcal{Q}_n$ orthogonal to all vectors of the form $U(p_i)$. □

Let b_n denote the volume of the unit ball in an n-dimensional Euclidean space. Then for the surface volume s_n of this ball we have

$$b_{n-1} < s_n = nb_n. \tag{7.1}$$

For an angle α, let $s_n(\alpha)$ be the surface volume of a subset of the surface cut out by a cone of half-angle α: for some vector $|u\rangle$, this is the set of all vectors $|x\rangle$ of unit length with $\langle u|x\rangle \geqslant \cos\alpha$. Thus, we have $s_n = s_n(\pi)$. We are interested in how fast $s_n(\alpha)$ decreases from $s_n/2$ to 0 as α moves from $\pi/2$ to 0.

Lemma 7.2. *Let $\alpha = \pi/2 - y$. Then*

$$s_n(\alpha)/s_n \stackrel{*}{<} \exp(-ny^2/2 + \ln n). \tag{7.2}$$

Proof. We have, for $k \geqslant 2$:

$$s_k(\alpha) = s_{k-1} \int_0^\alpha \sin^{k-2} x\, dx \leqslant s_{k-1}\alpha \sin^{k-2}\alpha. \tag{7.3}$$

So, we need to estimate $\int_0^\alpha \sin^n x\, dx$. The method used (also called "Laplace's" method), works for any twice differentiable function with a single maximum. Let $g(x) = \ln\sin x$, then it can be checked that $g'(\pi/2) = 0$, $g''(\pi/2) = -1$, $g'''(x) > 0$ for $x < \pi/2$. The Taylor expansion around $\pi/2$ gives, for $y > 0$:

$$g(\pi/2 - y) = -y^2/2 - y^3 g'''(\pi/2 - z)/6 < -y^2/2.$$

where $0 < z < y$. Hence, since $\sin x$ is increasing, we have for $x < \pi/2 - y$,

$$\sin^n(x) < e^{-ny^2/2}.$$

On the other hand, by (7.1), $s_k \geqslant b_{k-1} = s_{k-1}/(k-1)$, showing $s_{n-1} < (n-1)s_n$. Hence

$$s_n(\alpha) < \frac{\pi}{2} e^{-(n-2)y^2/2} s_{n-1} < \frac{(n-1)\pi}{2} e^{-(n-2)y^2/2} s_n \stackrel{*}{<} s_n e^{-ny^2/2 + \ln n}.$$

□

Lemma 7.3. *In any Hilbert space $\mathcal{H}$ of dimension 2^n (it may be a subspace of some $\mathcal{Q}_r$), the volume fraction of the set of unit vectors $|\psi\rangle$ in $\mathcal{H}$ with the property that $K_m(|\psi\rangle) < k$ is*

$$\stackrel{*}{<} \exp(-2^{n-m} + k\ln 2 + n).$$

Proof. We view $\mathcal{Q}_n$ as a 2^{n+1}-dimensional Euclidean space. Assume $-\log|\langle\phi|\psi\rangle|^2 \leqslant m$. If α is the angle between $|\phi\rangle$ and $|\psi\rangle$ then this means

$$2^{-m/2} < |\langle\phi|\psi\rangle| = \cos\alpha = \sin(\pi/2 - \alpha) \leqslant \pi/2 - \alpha,$$

giving $\alpha < \pi/2 - 2^{-m/2}$. For a fixed $|\phi\rangle$, the relative volume (with respect to $s_{2^{n+1}}$) of the set of vectors with $-\log|\langle\phi|\psi\rangle|^2 \leqslant m$ is therefore by (7.2)

$$\stackrel{*}{<} \exp(-2^{n+1}2^{-m}/2 + n) = \exp(-2^{n-m} + n).$$

Let $p_1, \dots, p_r$ be all programs of length $< k$ for which $U(p_m) \in \mathcal{Q}_n$. Then $r < 2^k$. The volume of all vectors $|\psi\rangle$ that are close in the above sense to at least one of the vectors $U(p_i)$ is thus

$$\stackrel{*}{<} 2^k \exp(-2^{n-m} + n) = \exp(-2^{n-m} + k\ln 2 + n).$$

□

Proof of Theorem 1. According to Lemma 7.1, there is a subspace V of $\mathcal{Q}_n$, of dimension $2^n - 2^{n-1} = 2^{n-1}$ with the property that for all $|\psi\rangle \in V$, for all m we have $K_m(|\psi\rangle) \geqslant n-1$. Let $m = n - 2\log n$. We can apply Lemma 7.3 to this subspace V of dimension 2^{n-1}, and obtain that for a certain constant c, the volume fraction of vectors with $K_m(|\psi\rangle) < 2n$ is

$$\begin{aligned} &\leqslant \exp(-2^{(n-1)-(n-2\log n)} + 2n\ln 2 + (n-1) + c) \\ &= \exp(-n^2/2 + n(2\ln 2 + 1) + c - 1). \end{aligned}$$

If n is large this is smaller than 1, so there are states $|\psi\rangle$ with $K_\infty(|\psi\rangle) \geqslant n-1$ and $K_{n-2\log n}(|\psi\rangle) > 2n$. For these, clearly

$$\mathrm{Kq}(|\psi\rangle) \geqslant (n-1) + (n - 2\log n + 1) = 2n - 2\log n.$$

□

8. Conclusions

We advanced a new proposal to extend the theory of descriptional complexity to the quantum setting. The approach starting from the universal density matrix appears to be fruitful and leads to some attractive relations. However, the theory is still very incomplete. The following tasks seem to be the most urgent.

(1) Strengthen Theorem 10 in a way that the smallness of $\overline{H}(|\psi\rangle)$ allows a direct inference on the smallness of $QC(|\psi\rangle)$ (or find a counterexample). For this, it seems to us that behavior of a monotonically increasing sequence of density functions needs to be understood better: namely, whether some approximate monotonicity can be stated about the subspaces E_k. Even if such a monotonicity will be found, even if Thoerem 10 can be proved for $\boldsymbol{\mu}$ instead of just computable density matrices, the result is too weak. To strengthen it, probably the theory of indeterminate-length quantum codes (the quantum analog of variable-length codes) will be needed, as developed in [5].

(2) Find the proper generalization to the quantum setting of the classical theorem saying that information cannot increase under the effect of any probabilistic computable transformation.

(3) What kind of addition theorems can be expected for quantum description complexity? The question is unsolved even for the von Neumann entropy. Also, the translation between the results on quantum description complexity and those on the von Neumann entropy will not be straightforward. As we remarked, the relation $\overline{H}(|\phi\rangle|\psi\rangle) \stackrel{+}{>} \overline{H}(|\phi\rangle)$ holds while $S(\rho_X) \leqslant S(\rho_{XY})$ does not. Still, maybe the study of the problem for quantum description complexity helps with the understanding of the problem for von Neumann entropy, and its relation to coding tasks of quantum information theory.

Despite all the caveats, let us ask the question (risking that somebody finds a trivial answer): does $\overline{H}$ obey strong superadditivity?

Acknowledgement. The author is grateful to Paul Vitányi, Harry Buhrman and Ronald de Wolf for discussions.

References

[1] André Berthiaume, Wim van Dam, and Sophie Laplante. Quantum Kolmogorov complexity. Preprint, quant-ph/0005018, 2000. See also Proc. 15th IEEE Conf. Computational Complexity, 2000 pp 240-249. 1, 1.2, 1.3, 3.3.2, 3.3.2, 3.3.2, 5.1

[2] Rajendra Bhatia. *Matrix Analysis*. Graduate Texts in Mathematics. Springer, New York, 1996. 2, 5.2

[3] M. Li and P. M. B. Vitányi. *Introduction to Kolmogorov Complexity and its Applications*. Springer Verlag, New York, 1993. 1.1, 2, 6.1
[4] Asher Peres. *Quantum Theory: Concepts and Methods*. Fundamental Theories of Physics. Kluwer Academic Publishers, Boston, 1995. 5.1
[5] Benjamin Schumacher and Michael D. Westmoreland. Indeterminate-Length Quantum Coding. Preprint quant-ph/0011014, 2000. 1
[6] Paul M. B. Vitányi. Three approaches to the quantitative definition of information in an individual pure quantum state. Preprint, quant-ph/9907035, 1999. See also Proc. 15th IEEE Conf. Computational Complexity, 2000 pp 263-270. 1, 1.2, 2, 3.3.1
[7] Alfred Wehrl. General properties of entropy. *Reviews of Modern Physics*, 50(2):221–260, April 1978. 2, 4, 4
[8] Hermann Weyl. *The Classical Groups, Their Invariants and Representations*. Princeton Univ. Press, 1946. 5.1

CWI AND BOSTON UNIVERSITY
E-mail address: gacs@bu.edu