Strong Converse for Identification via Quantum Channels

Rudolf Ahlswede* and Andreas Winter•

Abstract— **In this paper we present a simple proof of the strong converse for identification via discrete memoryless quantum channels, based on a novel covering lemma. The new method is a generalization to quantum communication channels of Ahlswede's recently discovered appoach to classical channels. It involves a development of explicit large deviation estimates to the case of random variables taking values in selfadjoint operators on a Hilbert space. This theory is presented separately in an appendix, and we illustrate it by showing its application to quantum generalizations of classical hypergraph covering problems.**

Keywords— **Identification, covering hypergraphs, quantum channels, large deviations.**

I. Introduction

Ahlswede and Dueck [4] found the identification capacity of a discrete memoryless channel by establishing the optimal (second order) rate via a so–called *soft converse*. Subsequently, the *strong converse*, conjectured by them, was proved by Han and Verdú [11]. Even their second, simplified proof [12] uses rather involved arguments.

In [3] it is shown how simple ideas regarding coverings of hypergraphs (formalized in lemma 8) can be used to obtain the approximations of output statistics needed in the converse.

Formally, we investigate the following situation: consider a discrete memoryless channel $W^n : \mathcal{X}^n \to \mathcal{Y}^n$ $(n \geq 1)$, i.e. for $x^n = x_1 \dots x_n \in \mathcal{X}^n$, $y^n = y_1 \dots y_n \in \mathcal{Y}^n$

$$W^n(y^n|x^n) = W(y_1|x_1) \cdots W(y_n|x_n),$$

with a channel $W : \mathcal{X} \to \mathcal{Y}$ which we identify with the DMC. It is well known [23] that the transmission capacity if this channel (with the strong converse proven by Wolfowitz [28]) is

$$C(W) = \max_{P \text{ p.d. on } \mathcal{X}} I(P;W).$$

Here $I(P;W) = H(PW) - H(W|P)$ is Shannon's mutual information, where $PW = \sum_{x\in\mathcal{X}} P(x)W(\cdot|x)$ is the output distribution on $\mathcal{Y}$, and $H(W|P) = \sum_{x\in\mathcal{X}} P(x)H(W(\cdot|x))$ is the conditional entropy of the channel for the input distribution P.

Ahlwede and Dueck, considering not the problem that the receiver wants to recover a message (*transmission problem*), but wants to decide whether or not the sent message is identical to an arbitrarily chosen one (*identification problem*), defined a $(n, N, \lambda_1, \lambda_2)$ identification (ID) code to be a collection of pairs

$$\{(P_i, \mathcal{D}_i) : i = 1, \dots, N\},$$

with probability distributions P_i on $\mathcal{X}^n$ and $\mathcal{D}_i \subset \mathcal{Y}^n$, such that the error probabilities of first resp. second kind satisfy

$$P_i W^n(\mathcal{D}_i^c) = \sum_{x^n\in\mathcal{X}^n} P_i(x^n)W^n(\mathcal{D}_i^c|x^n) \leq \lambda_1,$$
$$P_j W^n(\mathcal{D}_i) = \sum_{x^n\in\mathcal{X}^n} P_j(x^n)W^n(\mathcal{D}_i|x^n) \leq \lambda_2,$$

for all $i, j = 1, \dots, N$, $i \neq j$. Here $\mathcal{D}_i^c = \mathcal{Y}^n \setminus \mathcal{D}_i$ is the set complement of $\mathcal{D}_i$ in $\mathcal{Y}^n$, and

$$W^n(\mathcal{A}|x^n) = \sum_{y^n\in\mathcal{A}} W^n(y^n|x^n)$$

is a convenient shortcut for the probability of an event $\mathcal{A} \subset \mathcal{Y}^n$ conditional on x^n. Define $N(n, \lambda_1, \lambda_2)$ to be the maximal N such that a $(n, N, \lambda_1, \lambda_2)$ ID code exists.

With these definitions one has

Theorem 1 (Ahlswede, Dueck [4]) For every $\lambda_1, \lambda_2 > 0$ and $\delta > 0$, and for every sufficiently large n

$$N(n, \lambda_1, \lambda_2) \geq \exp(\exp(n(C(W) - \delta))).$$

■

The work [3] is devoted to a comparably short, and conceptually simple proof of

Theorem 2: Let $\lambda_1, \lambda_2 > 0$ such that $\lambda_1 + \lambda_2 < 1$. Then for every $\delta > 0$ and every sufficiently large n

$$N(n, \lambda_1, \lambda_2) \leq \exp(\exp(n(C(W) + \delta))).$$

■

Note that for $\lambda_1 + \lambda_2 \geq 1$ no upper bound on $N(n, \lambda_1, \lambda_2)$ can hold: a successful strategy would be that the receiver ignores the actual signal, and to identify i guesses YES with probability $1 - \lambda_1$, NO with probability $\lambda_1 \geq 1 - \lambda_2$.

The first proof of theorem 2 was given in [11], the method to be further extended in [12]. In [3] it is returned to the very first idea from [4], essentially to replace the distributions P_i by uniform distributions on "small" subsets of $\mathcal{X}^n$, namely with cardinality slightly above $\exp(nC(W))$.

Löber [18] began the study of identification via quantum channels. Following his work, and after Holevo [14], we define a (discrete memoryless) classical–quantum channel (quantum channel for short) to be a map

$$W : \mathcal{X} \longrightarrow \mathcal{S}(\mathcal{H}),$$

*Fakultät für Mathematik, Universität Bielefeld, Postfach 100131, 33501 Bielefeld, Germany.

•Department of Computer Science, University of Bristol, Merchant Venturers Building, Woodland Road, Bristol BS8 1UB, United Kingdom. Email: winter@cs.bris.ac.uk.

Version of October 19, 2001.

with $\mathcal{X}$ a finite set, as before, and $\mathcal{S}(\mathcal{H})$ the set of quantum states of the complex Hilbert space $\mathcal{H}$, which we assume to be finite dimensional. In the sequel, we shall use $a = |\mathcal{X}|$ and $d = \dim \mathcal{H}$. We identify $\mathcal{S}(\mathcal{H})$, as usual, with the set of density operators, i.e. the selfadjoint, positive semidefinite, linear operators on $\mathcal{H}$ with unit trace[1]:

$$\mathcal{S}(\mathcal{H}) = \{\rho : \rho = \rho^* \geq 0, \mathrm{Tr}\rho = 1\}.$$

In the sequel we will write W_x for the images $W(x)$ of the channel map.

Associated to W is the channel map on n–blocks

$$W^n : \mathcal{X}^n \longrightarrow \mathcal{S}(\mathcal{H}^{\otimes n}),$$

with

$$W^n_{x^n} = W_{x_1} \otimes \cdots \otimes W_{x_n}.$$

One can use quantum channels to transmit classical information, and Holevo [16] showed that the capacity is

$$C(W) = \max_{P \text{ p.d. on } \mathcal{X}} I(P; W).$$

Here $I(P;W) = H(PW) - H(W|P)$ is the von Neumann mutual information, with the output state $PW = \sum_{x\in\mathcal{X}} P(x)W_x$ on $\mathcal{H}$, and $H(W|P) = \sum_{x\in\mathcal{X}} P(x)H(W_x)$ the conditional entropy of the channel for the input distribution P. The only difference to Shannon's result is that here H denotes the *von Neumann* entropy which is defined, for a state ρ, as

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

The strong converse for this situation was proved (independently) in [20] and [25].

Quantum channels are a generalization of classical channels in the following sense: choose any orthonormal basis $(e_y : y \in \mathcal{Y})$ of the $|\mathcal{Y}|$–dimensional Hilbert space $\mathcal{H}$, and define for the classical channel $W : \mathcal{X} \to \mathcal{Y}$ the corresponding quantum channel $\widetilde{W} : \mathcal{X} \to \mathcal{S}(\mathcal{H})$ by

$$\widetilde{W}_x = \sum_y W(y|x)|e_y\rangle\langle e_y|.$$

Obviously for another channel V one has $\widetilde{V\times W} = \widetilde{V} \otimes \widetilde{W}$.

Regarding the decoding sets let $\mathcal{D} \subset \mathcal{Y}$, then the corresponding operator $D = \sum_{y\in\mathcal{D}} |e_y\rangle\langle e_y|$ satisfies for all x

$$W(\mathcal{D}|x) = \mathrm{Tr}(\widetilde{W}_x D).$$

Observe that by this translation rule a partition of $\mathcal{Y}$ corresponds to a projection valued measure (PVM) on $\mathcal{H}$, i.e. a collection of mutually orthogonal projectors which sum to $\mathbb{1}$. Conversely, given any operator D on $\mathcal{H}$ with $0 \leq D \leq \mathbb{1}$, define the function $\delta : \mathcal{Y} \to [0,1]$ by $\delta(y) = \langle e_y|D|e_y\rangle$. Then for all x

$$\mathrm{Tr}(\widetilde{W}_x D) = \sum_{y\in\mathcal{Y}} \delta(y)W(y|x),$$

which implies that every quantum observation, i.e. a positive operator valued measure (POVM), of the states $\widetilde{W}_x$ can be simulated by a classical randomized decision rule on $\mathcal{Y}$. One consequence of this is that the transmission capacities of W and of $\widetilde{W}$ are equal: $C(W) = C(\widetilde{W})$. Equally, also the identification capacities (whose definition in the quantum case is given below) coincide. For randomization at the decoder cannot improve either minimum error probability.

Abstractly, just given the states W_x, this situation occurs if they pairwise commute: for then they are simultaneously diagonalizable, hence the orthonormal basis $(e_y : y \in \mathcal{Y})$ arises.

According to [18] a $(n, N, \lambda_1, \lambda_2)$ *quantum identification (QID) code* is a collection of pairs

$$\{(P_i, D_i) : i = 1, \ldots, N\},$$

with probability distributions P_i on $\mathcal{X}^n$, and operators D_i on $\mathcal{H}^{\otimes n}$ satisfying $0 \leq D_i \leq \mathbb{1}$, such that the error probabilities of first resp. second kind satisfy

$$\begin{aligned}
&\mathrm{Tr}\left(P_i W^n(\mathbb{1} - D_i)\right) \\
&\quad = \mathrm{Tr}\left(\left(\sum_{x^n\in\mathcal{X}^n} P_i(x^n)W_{x^n}\right)(\mathbb{1} - D_i)\right) \leq \lambda_1, \\
&\mathrm{Tr}\left(P_j W^n \cdot D_i\right) \\
&\quad = \mathrm{Tr}\left(\left(\sum_{x^n\in\mathcal{X}^n} P_j(x^n)W_{x^n}\right)D_i\right) \leq \lambda_2,
\end{aligned}$$

for all $i, j = 1, \ldots, N$, $i \neq j$. Again, define $N(n, \lambda_1, \lambda_2)$ to be the maximal N such that a $(n, N, \lambda_1, \lambda_2)$ QID code exists.

This definition has a subtle problem: since the D_i need not commute, it is possible that identifying for a message i prohibits identification for j, as the corresponding POVMs $(D_i, \mathbb{1} - D_i)$ and $(D_j, \mathbb{1} - D_j)$ may be incompatible. To allow simultaneous identification of all messages we have to assue that the D_i have a common refinement, i.e. there exists a POVM $(E_k : k = 1, \ldots, K)$ and subsets I_i of $\{1, \ldots, K\}$ such that

$$D_i = \sum_{k\in I_i} E_k,$$

for all i. In this case the QID code is called *simultaneous*, and $N_{\mathrm{sim}}(n, \lambda_1, \lambda_2)$ is the maximal N such that a simultaneous $(n, N, \lambda_1, \lambda_2)$ quantum identification code exists. Clearly

$$N_{\mathrm{sim}}(n, \lambda_1, \lambda_2) \leq N(n, \lambda_1, \lambda_2).$$

In analogy to the above theorems it was proved:

Theorem 3 (Löber [18]) For every $\lambda_1, \lambda_2 > 0$ and $\delta > 0$, and for every sufficiently large n

$$N_{\mathrm{sim}}(n, \lambda_1, \lambda_2) \geq \exp(\exp(n(C(W) - \delta))).$$

On the other hand, let $\lambda_1, \lambda_2 > 0$ such that $\lambda_1 + \lambda_2 < 1$. Then for every $\delta > 0$ and every sufficiently large n

$$N_{\mathrm{sim}}(n, \lambda_1, \lambda_2) \leq \exp(\exp(n(C(W) + \delta))).$$

[1] See Davies [7] for the mathematics to describe quantum systems.

■

Looking at the examples given in [4] the simultaneity condition seems completely natural. But this need not always be the case.

Example 4: Modify the "sailors' wives" situation (Example 1 form [4]) as follows: the N sailors are not married each to one wife but instead are all in love with a single girl. One day in a storm one sailor drowns, and his identity should be communicated home. The girl however is capricious to the degree that it is impossible to predict who is her sweetheart at a given moment: when the message about the drowned sailor arrives, she will only ask for her present sweetheart, and only she will ask.

With our present approach we can get rid of the simultaneity condition in the converse (whereas by the above theorem identification codes approaching the capacity can be designed to be simultaneous — namely, by [4] for any sort of channel and a transmission code of rate R for it, one can construct an ID code "on top" of the transmission code, and with identification rate R, asymptotically):

Theorem 5: Let $\lambda_1, \lambda_2 > 0$ such that $\lambda_1 + \lambda_2 < 1$. Then for every $\delta > 0$ and every sufficiently large n

$$N(n, \lambda_1, \lambda_2) \leq \exp(\exp(n(C(W) + \delta))).$$

The rest of the paper is divided into two major blocks: first, after a short review of the ideas from [3] in section II, the rest of the main text will be devoted to the proof of theorem 5 (as explained, this contains theorem 2 indeed as a special case), in section III.

The other block is the appendix, containing the fundamentals of a theory of (selfadjoint) operator valued random variables. There the large deviation bounds to be used in the main text are derived.

II. The classical case

The core of the proof of theorem 2 in [3] is the following result about hypergraphs. Recall that a *hypergraph* is a pair $\Gamma = (\mathcal{V}, \mathcal{E})$ with a finite set $\mathcal{V}$ of vertices, and a finite set $\mathcal{E}$ of (hyper–) edges $E \subset \mathcal{V}$. We call Γ e–uniform, if all its edges have cardinality e. For an edge $E \in \mathcal{E}$ denote the characteristic function of $E \subset \mathcal{V}$ by 1_E.

The starting point is a result from large deviation theory:

Lemma 6: For an i.i.d. sequence $Z_1, \ldots, Z_L$ of random variables with values in $[0, 1]$ with expectation $\mathbb{E}Z_i = \mu$, and $0 < \epsilon < 1$

$$\Pr\left\{\frac{1}{L}\sum_{i=1}^{L} Z_i > (1+\epsilon)\mu\right\} \leq \exp(-LD((1+\epsilon)\mu\|\mu)),$$

$$\Pr\left\{\frac{1}{L}\sum_{i=1}^{L} Z_i < (1-\epsilon)\mu\right\} \leq \exp(-LD((1-\epsilon)\mu\|\mu)),$$

where $D(\alpha\|\beta)$ is the information divergence of the binary distributions $(\alpha, 1-\alpha)$ and $(\beta, 1-\beta)$. Since for $-\frac{1}{2} \leq x \leq \frac{1}{2}$ $D((1+x)\mu\|\mu) \geq \frac{1}{2\ln 2}x^2\mu$, it follows that

$$\Pr\left\{\frac{1}{L}\sum_{i=1}^{L} Z_i \notin [(1-\epsilon)\mu, (1+\epsilon)\mu]\right\} \leq 2\exp\left(-L\cdot\frac{\epsilon^2\mu}{2\ln 2}\right).$$

■

Lemma 7: Let $\Gamma = (\mathcal{V}, \mathcal{E})$ be an e–uniform hypergraph, and P a probability distribution on $\mathcal{E}$. Define the probability distribution Q on $\mathcal{V}$ by

$$Q(v) = \sum_{E\in\mathcal{E}} P(E)\frac{1}{e}1_E(v),$$

and fix $\epsilon, \tau > 0$. Then there exist vertices $\mathcal{V}_0 \subset \mathcal{V}$ and edges $E_1, \ldots, E_L \in \mathcal{E}$ such that with

$$\bar{Q}(v) = \frac{1}{L}\sum_{i=1}^{L}\frac{1}{e}1_{E_i}(v)$$

the following holds:

$$Q(\mathcal{V}_0) \leq \tau,$$

$$\forall v \in \mathcal{V}\setminus\mathcal{V}_0 \quad (1-\epsilon)Q(v) \leq \bar{Q}(v) \leq (1+\epsilon)Q(v),$$

$$L \leq 1 + \frac{|\mathcal{V}|}{e}\frac{2\ln 2\log(2|\mathcal{V}|)}{\epsilon^2\tau}.$$

Proof: See [3]. ■

For ease of application we formulate a slightly more general version of this:

Lemma 8: Let $\Gamma = (\mathcal{V}, \mathcal{E})$ be a hypergraph, with a measure Q_E on each edge E, such that $Q_E(v) \leq \eta$ for all E, $v \in E$. For a probability distribution P on $\mathcal{E}$ define

$$Q = \sum_{E\in\mathcal{E}} P(E)Q_E,$$

and fix $\epsilon, \tau > 0$.Then there exist vertices $\mathcal{V}_0 \subset \mathcal{V}$ and edges $E_1, \ldots, E_L \in \mathcal{E}$ such that with

$$\bar{Q} = \frac{1}{L}\sum_{i=1}^{L} Q_{E_i}$$

the following holds:

$$Q(\mathcal{V}_0) \leq \tau,$$

$$\forall v \in \mathcal{V}\setminus\mathcal{V}_0 \quad (1-\epsilon)Q(v) \leq \bar{Q}(v) \leq (1+\epsilon)Q(v),$$

$$L \leq 1 + \eta|\mathcal{V}|\frac{2\ln 2\log(2|\mathcal{V}|)}{\epsilon^2\tau}.$$

■

The interpretation of this result is as follows: Q is the expectation measure of the measures Q_E, which are sampled by the Q_{E_i}. The lemma says how close the sampling average $\bar{Q}$ can be to Q. In fact, assuming $Q_E(E) = q \leq 1$ for all $E \in \mathcal{E}$, one easily sees that

$$\|Q - \bar{Q}\|_1 \leq 2\epsilon + 2\tau.$$

The idea for the proof of theorem 2 is now: to replace the (in principle) arbitrary distributions P_i on $\mathcal{X}^n$ of a $(n, N, \lambda_1, \lambda_2)$ ID code $\{(P_i, D_i) : i = 1, \ldots, N\}$, by uniform distributions on subsets of $\mathcal{X}^n$, with cardinality bounded essentially by $\exp(nC(W))$. The condition is that the corresponding *output* distributions are close, so the resulting ID code will be a bit worse, but still nontrivial. This is done with the help of the covering lemma 8, applied to typical sequences in $\mathcal{Y}^n$ as vertices, and sets of induced typical sequences as edges. For details see [3].

III. Proof of theorem 5

It was already pointed out in the previous section that the main idea of the converse proof is to replace the arbitrary code distributions P_i by regularized approximations, the quality of approximation being measured by the $\|\cdot\|_1$–distance of the output distributions.

Hence, to extend the method to quantum channels we have to use the $\|\cdot\|_1$–distance of the output *quantum states*, and we have to find quantum versions of the lemmas 6 and 8.

Define a *quantum hypergraph* to be a pair $(\mathcal{V},\mathcal{E})$ with a finite dimensional Hilbert space $\mathcal{V}$ and a (finite) collection $\mathcal{E}$ of operators E on $\mathcal{V}$, $0\leq E\leq\mathbb{1}$ (see the discussion in subsection F of the appendix).

Analogous to lemma 6 is theorem 19 (in the appendix), the analog of lemma 8 is

Lemma 9: Let $(\mathcal{V},\mathcal{E})$ be a quantum hypergraph such that $E\leq\eta\mathbb{1}$ for all $E\in\mathcal{E}$. For a probability distribution P on $\mathcal{E}$ define

$$\rho=\sum_{E\in\mathcal{E}}P(E)E,$$

and fix $\epsilon,\tau>0$. Then there exists a subspace $\mathcal{V}_0<\mathcal{V}$ and $E_1,\ldots,E_L\in\mathcal{E}$ such that with

$$\bar{\rho}=\frac{1}{L}\sum_{i=1}^{L}E_i,$$

and denoting by Π_0 and Π_1 the orthogonal projections onto $\mathcal{V}_0$ and its complement, respectively, the following holds:

$$\mathrm{Tr}\rho\Pi_0\leq\tau,$$

$$(1-\epsilon)\Pi_1\rho\Pi_1\leq\Pi_1\bar{\rho}\Pi_1\leq(1+\epsilon)\Pi_1\rho\Pi_1,$$

$$L\leq 1+\eta(\dim\mathcal{V})\frac{2\ln 2\log(2\dim\mathcal{V})}{\epsilon^2\tau}.$$

Proof: Diagonalize ρ, i.e. $\rho=\sum_j r_j\pi_j$, and let

$$\Pi_0=\sum_{j:\ r_j<\tau/\dim\mathcal{V}}\pi_j,\quad\Pi_1=\mathbb{1}-\Pi_0.$$

Clearly $\mathrm{Tr}(\rho\Pi_o)\leq\tau$.

Now consider the quantum hypergraph $(\mathcal{V}_0^\perp,\Pi_1\mathcal{E}\Pi_1)$, whose edges also obey the upper bound $\eta\mathbb{1}$, and which has edge average $\Pi_1\rho\Pi_1\geq\tau\mathbb{1}/\dim\mathcal{V}$.

Now let $X_1,\ldots,X_L$ i.i.d. with $\Pr\{X_i=\Pi_1E\Pi_1\}=P(E)$. Then we can estimate with theorem 19, applied to the variables $\eta^{-1}X_i$:

$$\Pr\left\{\frac{1}{L}\sum_{i=1}^{L}X_i\notin[(1-\epsilon)\Pi_1\rho\Pi_1,(1+\epsilon)\Pi_1\rho\Pi_1]\right\}\leq 2(\dim\mathcal{V})\exp\left(-L\cdot\frac{\epsilon^2\tau}{2\eta\dim\mathcal{V}\ln 2}\right),$$

which is smaller than 1 if

$$L>\eta(\dim\mathcal{V})\frac{2\ln 2\log(2\dim\mathcal{V})}{\epsilon^2\tau},$$

in which case the desired covering exists. ∎

For application, assume $\mathrm{Tr}E=q\leq 1$ for all $E\in\mathcal{E}$. Then a consequence of the estimates of the lemma is

$$\|\rho-\bar{\rho}\|_1\leq(\epsilon+\tau)+\sqrt{8(\epsilon+\tau)}.$$

To see this observe

$$\begin{aligned}\|\rho-\bar{\rho}\|_1&\leq\|\rho-\Pi_1\rho\Pi_1\|_1+\|\Pi_1\rho\Pi_1-\Pi_1\bar{\rho}\Pi_1\|_1\\&\quad+\|\bar{\rho}-\Pi_1\bar{\rho}\Pi_1\|_1\\&\leq\tau+\epsilon+\sqrt{8(\epsilon+\tau)},\end{aligned}$$

where the three terms are estimated as follows: for the first note $\rho-\Pi_1\rho\Pi_1=\Pi_0\rho\Pi_0$, and apply $\mathrm{Tr}\rho\Pi_0\leq\tau$. For the second use the lemma, and for the third use $\mathrm{Tr}\bar{\rho}\Pi_0\leq\epsilon+\tau$ in lemma V.9 of [25].

We collect here a number of standard facts about types and typical sequences (cf. [6]):

*Empirical distributions (*aka *types)*: For a probability distribution P on $\mathcal{X}$ define

$$\mathcal{T}_P^n=\{x^n\in\mathcal{X}^n:\forall x\in\mathcal{X}\ N(x|x^n)=nP(x)\},$$

where $N(x|x^n)$ counts the number of x's in x^n. If this set is nonempty, we call P an *n–distribution*, or *type*, or *empirical distribution*. Notice that the number of types is

$$\binom{n+a-1}{a-1}\leq(n+1)^a.$$

Typical sequences: For $\alpha\geq 0$ and any distribution P on $\mathcal{X}$ define the following set of *typical sequences*:

$$\begin{aligned}\mathcal{T}_{P,\alpha}^n&=\Big\{x^n:\forall x\ |N(x|x^n)-nP(x)|\\&\qquad\leq\alpha\sqrt{n}\sqrt{P(x)(1-P(x))}\Big\}\\&=\bigcup_{Q\text{ s.t. }|Q(x)-P(x)|\leq\sqrt{\frac{P(1-P)}{n}}}\mathcal{T}_Q^n\end{aligned}$$

Note that by Chebyshev inequality $P^{\otimes n}(\mathcal{T}_{P,\alpha}^n)\geq 1-\frac{a}{\alpha^2}$.

From [25] recall the following facts about the quantum version of the previous constructions:

Typical subspace: For ρ on $\mathcal{H}$ and $\alpha\geq 0$ there exists an orthogonal subspace projector $\Pi_{\rho,\alpha}^n$ commuting with $\rho^{\otimes n}$, and satisfying

$$\begin{aligned}\mathrm{Tr}(\rho^{\otimes n}\Pi_{\rho,\alpha}^n)&\geq 1-\frac{d}{\alpha^2},\qquad(1)\\\mathrm{Tr}\Pi_{\rho,\alpha}^n&\leq\exp(nH(\rho)+Kd\alpha\sqrt{n}),\\\Pi_{\rho,\alpha}^n\rho^{\otimes n}\Pi_{\rho,\alpha}^n&\geq\exp(-nH(\rho)-Kd\alpha\sqrt{n})\Pi_{\rho,\alpha}^n.\end{aligned}$$

Conditional typical subspace: For $x^n\in\mathcal{T}_P^n$ and $\alpha\geq 0$ there exists an orthogonal subspace projector $\Pi_{W,\alpha}^n(x^n)$ commut-

ing with $W^n_{x^n}$, and satisfying

$$\mathrm{Tr}(W^n_{x^n}\Pi^n_{W,\alpha}(x^n)) \geq 1 - \frac{ad}{\alpha^2}, \tag{2}$$

$$\mathrm{Tr}\Pi^n_{W,\alpha}(x^n) \leq \exp(nH(W|P) + Kda\alpha\sqrt{n}),$$

$$\Pi^n_{W,\alpha}(x^n)W^n_{x^n}\Pi^n_{W,\alpha}(x^n) \leq \exp(-nH(W|P) + Kda\alpha\sqrt{n})\cdot \cdot\Pi^n_{W,\alpha}(x^n),$$

$$\mathrm{Tr}(W^n_{x^n}\Pi^n_{PW,\alpha\sqrt{a}}) \geq 1 - \frac{ad}{\alpha^2}. \tag{3}$$

Proof of theorem 5: We follow the strategy of the proof for theorem 2: consider a $(n, N, \lambda_1, \lambda_2)$ QID code $\{(P_i, D_i) : i = 1, \dots, N\}$, $\lambda_1 + \lambda_2 = 1 - \lambda < 1$, and concentrate on one P_i for the moment. Introduce, for empirical distributions T on $\mathcal{X}$, the probability distributions

$$P_i^T(x^n) = \frac{P_i(x^n)}{P_i(\mathcal{T}^n_T)} \text{ for } x^n \in \mathcal{T}^n_T,$$

extended by 0 to $\mathcal{X}^n$. For $x^n \in \mathcal{T}^n_T$ and with

$$\alpha = \frac{\sqrt{600ad}}{\lambda},$$

construct the conditional typical projector $\Pi^n_{W,\alpha}(x^n)$, and the typical projector $\Pi^n_{TW,\alpha\sqrt{a}}$.

Define the operators

$$Q_{x^n} = \Pi^n_{TW,\alpha\sqrt{a}}\Pi^n_{W,\alpha}(x^n)W^n_{x^n}\Pi^n_{W,\alpha}(x^n)\Pi^n_{TW,\alpha\sqrt{a}},$$

and note that

$$\|Q_{x^n} - W^n_{x^n}\|_1 \leq \frac{\lambda}{6},$$

by equations (2) and (3), and lemma V.9 of [25].

Now we apply lemma 9 with $\epsilon = \tau = \lambda^2/1200$ to the quantum hypergraph with the range of $\Pi^n_{TW,\alpha\sqrt{a}}$ as vertex space and edges

$$\Pi^n_{TW,\alpha\sqrt{a}}\Pi^n_{W,\alpha}(x^n)W^n_{x^n}\Pi^n_{W,\alpha}(x^n)\Pi^n_{TW,\alpha\sqrt{a}},\ x^n \in \mathcal{T}^n_T.$$

Combining we get a L–distribution $\bar{P}_i^T$ with

$$\|P_i^T Q - \bar{P}_i^T Q\|_1 \leq \frac{\lambda}{6},$$

$$L \leq \exp(nI(T;W) + O(\sqrt{n})) \leq \exp(nC(W) + O(\sqrt{n})),$$

where the constants depend explicitly on α, δ, τ. By construction we get

$$\|P_i^T W^n - \bar{P}_i^T W^n\|_1 \leq \frac{\lambda}{3}.$$

By the proof of lemma 9 we can choose $L = \exp(nC(W) + O(\sqrt{n}))$, independent of i and T.

Now choose a K–distribution R on the set of all empirical distributions such that

$$\sum_{T \text{ emp. distr.}} |P_i(\mathcal{T}^n_T) - R(T)| \leq \frac{\lambda}{3},$$

which is possible for

$$K = \lceil 3(n+1)^{|\mathcal{X}|}/\lambda \rceil.$$

Defining then

$$\bar{P}_i = \sum_{T \text{ emp. distr.}} R(T)\bar{P}_i^T$$

we deduce finally

$$\frac{1}{2}\|P_i W^n - \bar{P}_i W^n\|_1 \leq \frac{\lambda}{3}.$$

Since for every operator D on $\mathcal{H}^{\otimes n}$, $0 \leq D \leq \mathbb{1}$

$$|\mathrm{Tr}(P_i W^n \cdot D) - \mathrm{Tr}(\bar{P}_i W^n \cdot D)| \leq \frac{1}{2}\|P_i W^n - \bar{P}_i W^n\|_1$$

the collection $\{(\bar{P}_i, D_i) : i = 1, \dots, N\}$ is indeed a $(n, N, \lambda_1 + \lambda/3, \lambda_2 + \lambda/3)$ QID code.

The proof is concluded by two observations: because of $\lambda_1 + \lambda_2 + 2\lambda/3 < 1$ we have $\bar{P}_i \neq \bar{P}_j$ for $i \neq j$. Since the $\bar{P}_i$ however are KL–distributions, we find

$$\begin{aligned} N &\leq |\mathcal{X}^n|^{KL} = \exp(n \log|\mathcal{X}| \cdot KL) \\ &\leq \exp(\exp(n(C(W) + \delta))), \end{aligned}$$

the last if only n is large enough. ∎

We note that we actually proved the upper bound $C(W)$ to the resolution of a discrete memoryless quantum channel, in the following sense:

Definition 10: Let $\mathbf{W}$ be any quantum channel, i.e. a family $\mathbf{W} = (W^1, W^2, \dots)$ of maps

$$W^n : \mathcal{X}^n \to \mathcal{S}(\mathcal{H}^{\otimes n}).$$

A number R is called ϵ–*achievable resolution rate* if for all $\delta > 0$ there is n_0 such that for all $n \geq n_0$ and all probability distributions P^n on $\mathcal{X}^n$ there is an M–distribution Q^n on $\mathcal{X}^n$ with the properties

$$M \leq \exp(n(R + \delta)) \text{ and } \|P^n W^n - Q^n W^n\|_1 \leq \epsilon.$$

Define

$$S_\epsilon = \inf\{R : R \text{ is } \epsilon\text{–achievable resolution rate}\},$$

the channel's ϵ–*resolution.*

Observe that this goes beyond the definition of [18], where the resolution was a function of a measurement process $\mathbf{E}$:

Definition 11 (Löber [18]) Let $\mathbf{E} = (E^1, E^2, \dots)$ be a sequence of POVMs E_n on $\mathcal{H}^{\otimes n}$, and adopt the notations of definition 10. A number R is called ϵ–*achievable resolution rate for* $\mathbf{E}$ if for all $\delta > 0$ there is n_0 such that for all $n \geq n_0$ and all probability distributions P^n on $\mathcal{X}^n$ there is an M–distribution Q^n on $\mathcal{X}^n$ with the properties

$$M \leq \exp(n(R + \delta)) \text{ and } d_{E^n}(P^n W^n, Q^n W^n) \leq \epsilon,$$

where $d_E(\rho, \sigma)$ is the total variational distance of the two output distributions generated by applying E to states ρ, σ, respectively.

Define

$$S_\epsilon(\mathbf{E}) = \inf\{R \text{ is } \epsilon\text{–achievable resolution rate for } \mathbf{E}\},$$

the channel's ϵ–*resolution for* $\mathbf{E}$.
In general

$$S_\epsilon(\mathbf{E}) \leq S_\epsilon.$$

We do not know if there is an example of a channel such that

$$\sup_{\mathbf{E}} S_\epsilon(\mathbf{E}) < S_\epsilon.$$

Acknowledgements

Conversations with Alexander S. Holevo, Friedrich Götze, and Peter Eichelsbacher about the theory of nonreal random variables are acknowledged.

The research of AW was partially supported by SFB 343 "Diskrete Strukturen in der Mathematik" of the Deutsche Forschungsgemeinschaft.

Appendix

Operator valued random variables

A. Introduction

The theory of real random variables provides the framework of much of modern probability theory, such as laws of large numbers, limit theorems, and probability estimates for 'deviations', when sums of independent random variables are involved. However several authors have started to develop analogous theories for the case that the algebraic structure of the reals is substituted by more general structures such as groups, vector spaces, etc., see for example [10].

In the present work we focus on a structure that has vital interest in quantum probability theory, namely the algebra of operators on a (complex) Hilbert space, and in particular the real vector space of selfadjoint operators therein which can be regarded as a partially ordered generalization of the reals (as embedded in the complex numbers). In particular it makes sense to discuss probability estimates as the Markov and Chebyshev inequality (subsection C), and in fact one can even generalize the exponentially good estimates for large deviations by the so–called Bernstein trick which yield the famous Chernoff bounds (subsection D).

Otherwise the plan of this appendix is as follows: subsection B collects basic definitions and notation we employ, and some facts from the theory of operator and trace inequalities, after the central subsections C and D we collect a number of plausible conjectures (subsection E), and close with an application to the noncommutative generalization of the covering problem for hypergraphs, in subsection F.

B. Basic facts and definitions

We will study random variables $X : \Omega \longrightarrow \mathfrak{A}_s$, where $\mathfrak{A}_s = \{A \in \mathfrak{A} : A = A^*\}$ is the selfadjoint part of the C*–algebra $\mathfrak{A}$, which is a real vector space. Usually we will restrict our attention to the most interesting and in a sense generic case of the full operator algebra $\mathfrak{L}(\mathcal{H})$ of the complex Hilbert space $\mathcal{H}$. Throughout the paper we denote $d = \dim \mathcal{H}$, which we assume to be finite. In the general case $d = \mathrm{Tr}\mathbb{1}$, and $\mathfrak{A}$ can be embedded into $\mathfrak{L}(\mathbb{C}^d)$ as an algebra, preserving the trace.

The real cone $\mathfrak{A}_+ = \{A \in \mathfrak{A} : A = A^* \geq 0\}$ induces a partial order $\leq$ in $\mathfrak{A}_s$, which will be the main object of interest in what follows. Let us introduce some convenient notation: for $A, B \in \mathfrak{A}_s$ the *closed interval* $[A, B]$ is defined as

$$[A, B] = \{X \in \mathfrak{A}_s : A \leq X \leq B\}.$$

(Similarly open and halfopen intervals (A, B), $[A, B)$, etc.).

For simplicity we will assume that the space Ω on which the random variables live is discrete.

Some remarks on the operator order:
(A) $\leq$ is not a total order unless $\mathfrak{A} = \mathbb{C}$, in which case $\mathfrak{A}_s = \mathbb{R}$. Thus in this case (which we will refer to as the *classical case*) the theory developed below reduces to the study of real random variables.
(B) $A \geq 0$ is equivalent to saying that all eigenvalues of A are nonnegative. These are d nonlinear inequalities. However from the alternative characterization

$$\begin{aligned} A \geq 0 &\Longleftrightarrow \forall \rho \text{ density operator } \mathrm{Tr}(\rho A) \geq 0 \\ &\Longleftrightarrow \forall \pi \text{ one–dim. projector } \mathrm{Tr}(\pi A) \geq 0 \end{aligned}$$

we see that this is equivalent to infinitely many *linear* inequalities, which is better adapted to the vector space structure of $\mathfrak{A}_s$.
(C) The operator mappings $A \mapsto A^s$ (for $s \in [0, 1]$) and $A \mapsto \log A$ are defined on $\mathfrak{A}_+$, and both are operator monotone and operator concave.
In contrast, the mappings $A \mapsto A^s$ (for $s > 2$) and $A \mapsto \exp A$ are neither operator monotone nor operator convex. Interestingly, A^s for $s \in [1, 2]$ *is* operator convex (though not operator monotone).
All this follows from Löwner's theorem [19], a good account of which is given in Donoghue's book [8], and a characterization of operator convex functions due to Hansen and Pedersen [13].
(D) Note however that the mapping $A \mapsto \mathrm{Tr}\exp A$ *is* monotone and convex: see Lieb [17].
(E) Golden–Thompson–inequality ([9], [24]): for $A, B \in \mathfrak{A}_s$

$$\mathrm{Tr}\exp(A + B) \leq \mathrm{Tr}\left((\exp A)(\exp B)\right).$$

C. Markov and Chebyshev inequality

Theorem 12 (Markov inequality) Let X a random variable with values in $\mathfrak{A}_+$ and expectation $M = \mathbb{E}X = \sum_x \Pr\{X = x\}x$, and $A \geq 0$ (i.e. $A \in \mathfrak{L}(\mathcal{H})_+$). Then

$$Pr\{X \not\leq A\} \leq \mathrm{Tr}\left(MA^{-1}\right).$$

Proof: We may assume that the support of A contains the support of M, otherwise the theorem is trivial.

Consider the positive random variable

$$Y = A^{-1/2} X A^{-1/2},$$

which has expectation $\mathbb{E}Y = N = A^{-1/2} M A^{-1/2}$. Since the events $\{X \leq A\}$ and $\{Y \leq \mathbb{1}\}$ coincide we have to show that

$$Pr\{Y \not\leq \mathbb{1}\} \leq \mathrm{Tr}N.$$

This is seen as follows:

$$N = \sum_y \Pr\{Y = y\} y \geq \sum_{y \not\leq \mathbb{1}} \Pr\{Y = y\} y.$$

Taking traces, and observing that a positive operator which is not less than or equal $\mathbb{1}$ must have trace at least 1, we find

$$\begin{aligned} \mathrm{Tr} N &\geq \sum_{y \not\leq \mathbb{1}} \Pr\{Y = y\} \mathrm{Tr} y \\ &\geq \sum_{y \not\leq \mathbb{1}} \Pr\{Y = y\} = \Pr\{Y \not\leq \mathbb{1}\}, \end{aligned}$$

which is what we wanted. ∎

Remark 13: In the case of $\mathcal{H} = \mathbb{C}$ the theorem reduces to the well known Markov inequality for nonnegative real random variables. One can easily see that like in this classical case the inequality of the theorem is optimal in the sense that there are examples when it is assumed with equality. If we assume knowledge about the second moment of X we can prove

Theorem 14 (Chebyshev inequality) Let X a random variable with values in $\mathfrak{A}_s$, expectation $M = \mathbb{E}X$, and variance $\mathrm{Var}X = S^2 = \mathbb{E}\left((X-M)^2\right) = \mathbb{E}(X^2) - M^2$. For $\Delta \geq 0$

$$\Pr\{|X - M| \not\leq \Delta\} \leq \mathrm{Tr}\left(S^2 \Delta^{-2}\right).$$

Proof: Observing

$$|X - M| \leq \Delta \Longleftarrow (X-M)^2 \leq \Delta^2$$

(because $\sqrt{\ }$ is operator monotone, see section B, *(C)*) we find

$$\begin{aligned} \Pr\{|X - M| \not\leq \Delta\} &\leq \Pr\{(X-M)^2 \not\leq \Delta^2\} \\ &\leq \mathrm{Tr}\left(S^2 \Delta^{-2}\right) \end{aligned}$$

(by theorem 12). ∎

Remark 15: If X, Y are independent, then $\mathrm{Var}(X+Y) = \mathrm{Var}X + \mathrm{Var}Y$. The calculation is the same as in the classical case, but one has to take care of the noncommutativity.

Corollary 16 (Weak law of large numbers) Let X, X_1, $\ldots$, X_n i.i.d. random variables with $\mathbb{E}X = M$, $\mathrm{Var}X = S^2$, and $\Delta \geq 0$. Then

$$\Pr\left\{\frac{1}{n}\sum_{i=1}^n X_i \notin [M-\Delta, M+\Delta]\right\} \leq \frac{1}{n}\mathrm{Tr}\left(S^2\Delta^{-2}\right),$$

$$\Pr\left\{\sum_{i=1}^n X_i \notin [nM - \Delta\sqrt{n}, nM + \Delta\sqrt{n}]\right\} \leq \mathrm{Tr}\left(S^2\Delta^{-2}\right).$$

Proof: Observe that $Y \notin [M-\Delta, M+\Delta]$ is equivalent to $|Y - M| \not\leq \Delta$, and apply the previous theorem. ∎

D. Large deviations and Bernstein trick

Lemma 17: For a random variable Y, $B \in \mathfrak{A}_s$, and $T \in \mathfrak{A}$ such that $T^*T > 0$

$$\Pr\{Y \not\leq B\} \leq \mathrm{Tr}\left(\mathbb{E}\exp(TYT^* - TBT^*)\right).$$

Proof: A direct calculation:

$$\begin{aligned} \Pr\{Y \not\leq B\} &= \Pr\{Y - B \not\leq 0\} \\ &= \Pr\{TYT^* - TBT^* \not\leq 0\} \\ &= \Pr\{\exp(TYT^* - TBT^*) \not\leq \mathbb{1}\} \\ &\leq \mathrm{Tr}\left(\mathbb{E}\exp(TYT^* - TBT^*)\right). \end{aligned}$$

Here the second line is because the mapping $X \mapsto TXT^*$ is bijective and preserves the order, the third because for *commuting* operators A, B, $A \leq B$ is equivalent to $\exp A \leq \exp B$, and the last line by theorem 12. ∎

Theorem 18: Let $X, X_1, \ldots, X_n$ i.i.d. random variables with values in $\mathfrak{A}_s$, $A \in \mathfrak{A}_s$. Then for $T \in \mathfrak{A}$, $T^*T > 0$

$$\Pr\left\{\sum_{i=1}^n X_i \not\leq nA\right\} \leq d \cdot \left\|\mathbb{E}\exp\left(TXT^* - TAT^*\right)\right\|^n.$$

Proof: Using the previous theorem with $Y = \sum_{i=1}^n X_i$ and $B = nA$ we find

$$\begin{aligned} \Pr\left\{\sum_{i=1}^n X_i \not\leq nA\right\} &\leq \mathrm{Tr}\left(\mathbb{E}\exp\left(\sum_{i=1}^n T(X_i - A)T^*\right)\right) \\ &= \mathbb{E}\mathrm{Tr}\exp\left(\sum_{i=1}^n T(X_i - A)T^*\right) \\ &\leq \mathbb{E}\mathrm{Tr}\left[\exp\left(\sum_{i=1}^{n-1} T(X_i - A)T^*\right)\right. \\ &\qquad \left. \cdot \exp\left(T(X_n - A)T^*\right)\right] \\ &= \mathbb{E}_{1\ldots n-1}\mathrm{Tr}\left[\exp\left(\sum_{i=1}^{n-1} T(X_i - A)T^*\right)\right. \\ &\qquad \left. \cdot \mathbb{E}\exp\left(T(X_n - A)T^*\right)\right] \\ &\leq \left\|\mathbb{E}\exp\left(T(X_n - A)T^*\right)\right\| \cdot \\ &\quad \cdot \mathbb{E}_{1\ldots n-1}\mathrm{Tr}\exp\left(\sum_{i=1}^{n-1} T(X_i - A)T^*\right) \\ &\leq \ldots \leq d \cdot \left\|\mathbb{E}\exp\left(T(X_n - A)T^*\right)\right\|^n. \end{aligned}$$

Here everything is straightforward, except for the third line which is by the Golden–Thompson–inequality (section B, *(E)*). ∎

The problem is now to minimize $\|\mathbb{E}\exp\left(TXT^* - TAT^*\right)\|$ with respect to T. Observe that without loss of generality we may assume that T is selfadjoint, because of the polar decomposition $T = U \cdot |T|$, with a unitary U. The case we will pursue further is that of a bounded random variable. Introducing the binary I–divergence

$$D(u\|v) = u(\log u - \log v) + (1-u)\left(\log(1-u) - \log(1-v)\right)$$

we find

Theorem 19 (Chernoff) Let $X, X_1, \ldots, X_n$ i.i.d. random variables with values in $[0, \mathbb{1}] \subset \mathfrak{A}_s$, $\mathbb{E}X \leq m\mathbb{1}$, $A \geq a\mathbb{1}$,

$1 \geq a \geq m \geq 0$. Then

$$\Pr\left\{\sum_{i=1}^{n} X_i \not\leq nA\right\} \leq d \cdot \exp\left(-nD(a\|m)\right).$$

Similarly, if $\mathbb{E}X \geq m\mathbb{1}$, $A \leq a\mathbb{1}$, $0 \leq a \leq m \leq 1$. Then

$$\Pr\left\{\sum_{i=1}^{n} X_i \not\geq nA\right\} \leq d \cdot \exp\left(-nD(a\|m)\right).$$

As a consequence we get, for $\mathbb{E}X = M \geq \mu\mathbb{1}$ and $0 \leq \epsilon \leq \frac{1}{2}$,

$$\Pr\left\{\frac{1}{n}\sum_{i=1}^{n} X_i \notin [(1-\epsilon)M, (1+\epsilon)M]\right\} \leq 2d \cdot \exp\left(-n \cdot \frac{\epsilon^2\mu}{2\ln 2}\right).$$

Proof: The second part follows from the first by considering $Y_i = \mathbb{1} - X_i$, and the observation that $D(a\|m) = D(1-a\|1-m)$.

To prove it we apply theorem 18 with $T = \sqrt{t}\mathbb{1}$:

$$\Pr\left\{\sum_{i=1}^{n} X_i \not\leq nA\right\} \leq \Pr\left\{\sum_{i=1}^{n} X_i \not\leq na\mathbb{1}\right\} \leq d \cdot \left\|\mathbb{E}\exp(tX)\exp(-ta)\right\|^n.$$

Now using

$$\exp(tX) - \mathbb{1} \leq X(\exp(t) - 1)$$

(which follows from the validity of the estimate for real x, $x \in (0,1)$:

$$\frac{\exp(tx)-1}{x} \leq \frac{\exp(t)-1}{1},$$

which in turn is just the convexity of exp) we find

$$\begin{aligned}\mathbb{E}\exp(tX) &\leq \mathbb{1} + \mathbb{E}X(\exp(t)-1)\\ &\leq (1 - m + m\exp t)\mathbb{1}.\end{aligned}$$

Hence

$$\left\|\mathbb{E}\exp(tX)\exp(-ta)\right\| \leq (1 - m + m\exp t)\exp(-at),$$

and choosing

$$t = \log\left(\frac{a}{m} \cdot \frac{1-m}{1-a}\right) > 0$$

the right hand side becomes exactly $\exp\left(-D(a\|m)\right)$.

To prove the last claim of the theorem consider the variables $Y_i = \mu M^{-1/2} X_i M^{-1/2}$ with expectation $\mathbb{E}Y_i = \mu$ and $Y_i \in [0, \mathbb{1}]$, by hypotheses. Because of

$$\frac{1}{n}\sum_{i=1}^{n} X_i \in [(1-\epsilon)M, (1+\epsilon)M] \iff \frac{1}{n}\sum_{i=1}^{n} Y_i \in [(1-\epsilon)\mu\mathbb{1}, (1-\epsilon)\mu\mathbb{1}]$$

we can apply what we just proved to obtain

$$\begin{aligned}&\Pr\left\{\frac{1}{n}\sum_{i=1}^{n} X_i \notin [(1-\epsilon)M, (1+\epsilon)M]\right\}\\ &\quad\leq d\big[\exp(-nD((1-\epsilon)\mu\|\mu)) + \exp(-nD((1+\epsilon)\mu\|\mu))\big]\\ &\quad\leq 2d\cdot\exp\left(-n\cdot\frac{\epsilon^2\mu}{2\ln 2}\right),\end{aligned}$$

the last line by the already used inequality $D((1+x)\mu\|\mu) \geq \frac{1}{2\ln 2}x^2\mu$. ∎

E. Conjectures

We have the feeling that in the estimates of the previous section we waste too much. In particular the theorems become useless in the infinite dimensional case, because in the traces we could only account for the supremum of the involved eigenvalues, multiplied by the dimension of the underlying space.

Conjecture 20: Under the assumptions of theorem 18 it even holds that

$$\Pr\left\{\sum_{i=1}^{n} X_i \not\leq nA\right\} \leq \mathrm{Tr}\left[\left(\mathbb{E}\exp\left(TXT^* - TAT^*\right)\right)^n\right],$$

since we conjecture that for i.i.d. random variables Z, $Z_1, \ldots, Z_n \in \mathfrak{A}_s$

$$\mathrm{Tr}\mathbb{E}\exp\left(\sum_{i=1}^{n} Z_i\right) \leq \mathrm{Tr}\left(\left(\mathbb{E}\exp(Z)\right)^n\right).$$

Note that this is indeed true for $n = 2$, thanks to the Golden–Thompson inequality!

For larger n there seems to be no applicable generalization of the Golden–Thompson inequality, so a different approach is needed. We propose to take logarithms in the above conjecture instead of traces: by the monotonicity of $\mathrm{Tr}\exp A$ the conjecture is true if

$$\log\mathbb{E}\exp\left(\sum_{i=1}^{n} Z_i\right) \leq n\log\mathbb{E}\exp(Z).$$

Thus by induction and monotonicity of $\mathrm{Tr}\exp A$ we can indeed prove conjecture 20 if the following is true:

Conjecture 21: For finite families of selfadjoint operators A_i and B_j

$$\log\left(\sum_{ij}\exp(A_i + B_j)\right) \leq \log\left(\sum_i \exp A_i\right) + \log\left(\sum_j \exp B_j\right).$$

Note that if all A_i, B_j commute then equality holds!
It may be that taking this for granted one can prove the following conjecture (compare with theorem 19):

Conjecture 22: Let $X, X_1, \ldots, X_n$ i.i.d. random variables with values in $[0, \mathbb{1}]$, $\mathbb{E}X \leq M \leq A \leq \mathbb{1}$. Then

$$\Pr\left\{\sum_{i=1}^{n} X_i \not\leq nA\right\} \leq \operatorname{Tr}\exp\left(-n\mathcal{D}(A\|M)\right),$$

where $\mathcal{D}$ is the operator version of the binary I–divergence:

$$\mathcal{D}(A\|M) = \sqrt{A}(\log A - \log M)\sqrt{A} + \sqrt{\mathbb{1}-A}\left(\log(\mathbb{1}-A) - \log(\mathbb{1}-M)\right)\sqrt{\mathbb{1}-A}.$$

Again, given conjecture 21, it would suffice to compare $\log \mathbb{E}\exp(TXT^* - TAT^*)$ and $\mathcal{D}(A\|M)$, for a clever choice of T.

F. An application

In this last subsection we want to discuss one application of our estimates in "noncommutative combinatorics", namely as a tool in applying the probabilistic method to the noncommutative analogue of covering hypergraphs. Apart from the application in the main text, we would like to point out two other ones: an approximation problem in quantum estimation theory [27], and its generalization to asymptotic convex decompositions of POVMs [26].

F.1 Noncommutative hypergraphs

We will define noncommutative hypergraphs as generalizations of the usual ones. To understand the following definition one has to recall the correspondence between a compact space X and the C*–algebra $C(X)$ of its continuous $\mathbb{C}$–valued functions, provided by the Gelfand–Naimark theorem (see [5], ch. 2.3). In the case of a finite discrete set this is summarized in the fact that the positive idempotents of the function algebra are exactly the characteristic functions of subsets. Thus we can talk about hypergraphs $(\mathcal{V}, \mathcal{E})$ — $\mathcal{V}$ is the finite *vertex* set, and $\mathcal{E} \subset 2^{\mathcal{V}}$ the set of *hyperedges* (or *edges* for short) — in the language of finite dimensional commutative C*–algebras and certain of their idempotents.

A *noncommutative hypergraph* Γ is a pair $(\mathfrak{V}, \mathcal{E})$ with a finite dimensional C*–algebra $\mathfrak{V}$ and a set $\mathcal{E} \subset [0, \mathbb{1}]$ (usually finite). We call Γ *strict* if all elements of $\mathcal{E}$ are idempotents. Finally Γ is a *quantum hypergraph* if $\mathfrak{V}$ is the full operator algebra of a finite dimensional complex Hilbert spave $\mathcal{V}$, in which case we denote Γ as $(\mathcal{V}, \mathcal{E})$. From the theory of finite dimensional C*–algebras it is known that $\mathfrak{V}$ can be embedded into the full operator algebra of a Hilbert space of dimension $\operatorname{Tr}\mathbb{1}$, preserving the trace. Thus we will in the sequel always assume that we deal with quantum hypergraphs.

For a finite edge set $\mathcal{E}$ the *degree* is defined as the operator

$$\deg_{\mathcal{E}} = \sum_{E\in\mathcal{E}} E.$$

A *covering* of $\Gamma = (\mathcal{V}, \mathcal{E})$ is a finite family $\mathcal{C}$ of edges such that $\deg_{\mathcal{C}} \geq \mathbb{1}$.

F.2 Covering theorems

Now we come to our first covering theorem (for the classical case compare [1]):

Theorem 23: Let Γ a quantum hypergraph with

$$\deg_{\Gamma} \geq \delta\mathbb{1}.$$

Then there exists a covering of Γ with $k \leq 1 + \frac{8|\mathcal{E}|\ln 2}{\delta}\log d$ many edges.

This is the special case of the uniform distribution in the following

Theorem 24: Let Γ a quantum hypergraph and P a probability distribution on $\mathcal{E}$, such that

$$\sum_{E\in\mathcal{E}} P(E)E \geq \mu\mathbb{1}.$$

Then there exists a covering of Γ with $k \leq 1 + 8(\ln 2 \log d)\mu^{-1}$ many edges.

Proof: Draw edges at random, i.e. consider i.i.d. random variables $X, X_1, \ldots, X_k$ with $\Pr\{X = E\} = P(E)$. Then we obtain, using theorem 19:

$$\begin{aligned}\Pr\left\{\sum_{i=1}^{k} X_i \not\geq \mathbb{1}\right\} &= \Pr\left\{\sum_{i=1}^{k} X_i \not\geq k\cdot\frac{1}{k}\mathbb{1}\right\}\\ &\leq d\exp\left(-k\cdot D\left(\frac{1}{k}\|\mu\right)\right)\\ &\leq d\exp\left(-k\cdot D\left(\frac{1}{2}\mu\|\mu\right)\right)\\ &\leq d\exp\left(-k\cdot\frac{1}{8\ln 2}\mu\right),\end{aligned}$$

the third line only if we have $k \geq 2\mu^{-1}$. Now the last expression is smaller than 1 for

$$k > 8(\ln 2\log d)\mu^{-1},$$

justifying ex post our estimates. Hence for k as in the theorem there exists a covering with k edges. ∎

We apply this result to a generalization of a result on covering numbers of hypergraphs, due to Posner and McEliece [21], obtained independently, but a little bit later, by Ahlswede and reported in [2]. For a quantum hypergraph $\Gamma = (\mathcal{V}, \mathcal{E})$ define $\Gamma^n = (\mathcal{V}^{\otimes n}, \mathcal{E}^n)$, with

$$\mathcal{E}^n = \{E^n = E_1 \otimes \cdots \otimes E_n : E_1, \ldots, E_n \in \mathcal{E}\}.$$

We are interested in the covering number $c(n)$ of Γ^n, i.e. the minimum cardinality of a covering of Γ^n. Finally define

$$\tilde{c}(n) = \min\left\{\sum_{E^n} v(E^n) : v \geq 0,\ \sum_{E^n} v(E^n)E^n \geq \mathbb{1}^{\otimes n}\right\}.$$

(The v can be seen as a continous weight version of coverings, and will be called *generalized coverings*). It is immediate that $c(n) \geq \tilde{c}(n)$.

Theorem 25: With

$$C = -\log\left(\max_{P \text{ p.d. on } \mathcal{E}} \min\sum_{E\in\mathcal{E}} P(E)E\right),$$

where the min means the minimal eigenvalue, one has

$$\tilde{c}(n) \geq \exp(Cn),$$
$$c(n) \leq 1 + 8(\ln 2 \log d) \cdot n \exp(Cn).$$

In particular

$$\lim_{n\to\infty} \frac{1}{n} \log c(n) = \lim_{n\to\infty} \frac{1}{n} \log \tilde{c}(n) = C.$$

Proof: The second estimate follows by applying theorem 24 with the distribution $P^{\otimes n}$.

The first is proved by induction on n. The case $n = 0$ is trivial, and the case $n = 1$ is seen as follows: let v^* be a minimal weight generalized covering of Γ, i.e.

$$\tilde{c}(1) = \sum_E v^*(E) \text{ and } \sum_E V^*(E)E \geq \mathbb{1}.$$

What we have to show is that

$$\tilde{c}(1) \geq \left(\max_P \min \sum_E P(E)E \right)^{-1},$$

which means we have to find a distribution P such that

$$sum_E P(E)E \geq \tilde{c}(1)^{-1}\mathbb{1}.$$

With $P(E) = v^*(E)\tilde{c}(1)^{-1}$ this is obviously satisfied.

Now assume $n > 0$, and let v^* a minimal weight generalized covering of Γ^n. Define a probability distribution Q on $\mathcal{E}$ by

$$Q(E) = \frac{1}{\tilde{c}(n)} \sum_{E^n \in \mathcal{E}^n, E_n = E} v^*(E^n).$$

Multiplying the relation

$$\sum_{E^n} v^*(E^n)E^n \geq \mathbb{1}^{\otimes n}$$

by $\mathbb{1}^{\otimes(n-1)} \otimes \pi$ (for a one–dimensional projector π on $\mathcal{V}$) from both sides and taking the trace over the last factor we find

$$\begin{aligned}\mathbb{1}^{\otimes(n-1)} &\leq \sum_{E_n \in \mathcal{E}} \mathrm{Tr}(\pi E_n) \sum_{E^{n-1} \in \mathcal{E}^{n-1}} v^*(E^n)E^{n-1} \\ &= \sum_{E^{n-1} \in \mathcal{E}^{n-1}} \left(\sum_{E_n \in \mathcal{E}} \mathrm{Tr}(\pi E_n) v^*(E^n) \right) E^{n-1}.\end{aligned}$$

This means that we have a generalized covering of Γ^{n-1} and hence

$$\begin{aligned}\tilde{c}(n-1) &\leq \sum_{E_n \in \mathcal{E}} \mathrm{Tr}(\pi E_n) \sum_{E^{n-1} \in \mathcal{E}^{n-1}} v^*(E^n) \\ &= \sum_{E_n \in \mathcal{E}} \mathrm{Tr}(\pi E_n) Q(E_n) \tilde{c}(n).\end{aligned}$$

Thus for all π

$$\sum_{E \in \mathcal{E}} \mathrm{Tr}(\pi E) Q(E) \geq \frac{\tilde{c}(n-1)}{\tilde{c}(n)},$$

which implies

$$\min \sum_{E \in \mathcal{E}} Q(E)E \geq \frac{\tilde{c}(n-1)}{\tilde{c}(n)},$$

which in turn implies

$$\exp(-C) = \max_P \min \sum_{E \in \mathcal{E}} P(E)E \geq \frac{\tilde{c}(n-1)}{\tilde{c}(n)}.$$

■

References

[1] R. Ahlswede, "Coloring Hypergraphs: A New Approach to Multi–user Source Coding — I", J. Combinatorics, Information & System Sciences, vol. 4, no. 1, pp. 76–115, 1979.

[2] R. Ahlswede, "On set coverings in cartesian product spaces", Preprint E92–005, Sonderforschungsbereich 343 "Diskrete Strukturen in der Mathematik", Universität Bielefeld, 1992.

[3] R. Ahlswede, "On concepts of performance parameters for channels", to appear in IEEE Trans. Inf. Theory, Special issue in memory of A. D. Wyner.

[4] R. Ahlswede, G. Dueck, "Identification via channels", IEEE Trans. Inf. Theory, vol. 35, pp. 15–29, 1989.

[5] O. Bratteli, D. W. Robinson, *Operator Algebras and Quantum Statistical Mechanics I*, Springer–Verlag, 1979.

[6] I. Csiszár, J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*, Academic Press, New York, 1981.

[7] E. B. Davies, *Quantum Theory of Open Systems*, Academic Press, London, 1976.

[8] W. F. Donoghue, Jr., *Monotone Matrix Functions and Analytic Continuation*, Springer, 1974.

[9] S. Golden, "Lower Bounds for the Helmholtz Function", Physical Review, vol. 137B, no. 4, pp. B1127–1128, 1965.

[10] U. Grenander, *Probabilities on Algebraic Structures*, Wiley & Sons, New York, London, 1963.

[11] T. S. Han, S. Verdú, "New results in the theory of identification via channels", IEEE Trans. Inf. Theory, vol. 38, no. 1, pp. 14–25, 1992.

[12] T. S. Han, S. Verdú, "Approximation theory of output statistics", IEEE Trans. Inf. Theory, vol. 39, no. 3, pp. 752–772, 1993.

[13] F. Hansen, G. K. Pedersen, "Jensen's Inequality for Operators and Löwner's Theorem", Math. Ann., vol. 258, pp. 229–241, 1982.

[14] A. S. Holevo, Problemy Peredachi Informatsii, vol. 9, no. 3, pp. 3–11, 1973 (english translation: "Bounds for the quantity of information transmitted by a quantum channel", Probl. Inf. Transm., vol. 9, no. 3, pp. 177–183, 1973).

[15] A. S. Holevo, "Problems in the mathematical theory of quantum communication channels", Rep. Math. Phys., vol. 12, no. 2, pp. 273–278, 1977.

[16] A. S. Holevo, "The Capacity of the Quantum Channel with General Signal States", IEEE Trans. Inf. Theory, vol. 44, no. 1, pp. 269–273, 1998.

[17] E. H. Lieb, "Convex trace functions and the Wigner–Yanase–Dyson conjecture", Adv. Math., vol. 11, pp. 267–288, 1973.

[18] P. Löber, *Quantum Channels and Simultaneous ID Coding*, doctoral dissertation, Universität Bielefeld, 1999. WWW at `http://archiv.ub.uni-bielefeld.de/disshabi/mathe.htm`.

[19] K. Löwner, "Über monotone Matrixfunktionen", Math. Z., vol. 38, pp. 177–216, 1934.

[20] T. Ogawa, H. Nagaoka, "Strong Converse to the Quantum Channel Coding Theorem", IEEE Trans. Inform. Theory, vol. 45, no. 7, pp. 2486–2489, 1999.

[21] R. J. McEliece, E. C. Posner, "Hide and seek, data storage and entropy", Annals Math. Statistics, vol. 42, pp. 1706–1716, 1971.

[22] B. Schumacher, "Quantum Coding", Phys. Rev. A, vol. 51, no. 4, pp. 2738–2747, 1995.

[23] C. E. Shannon, "A mathematical theory of communication", Bell System Tech. J., vol. 27, pp. 379–423; *ibid.* pp. 623–656, 1948.

[24] C. J. Thompson, "Inequality with Applications in Statistical Mechanics", J. Math. Phys., vol. 6, no. 11, pp. 1812–1823, 1965.

[25] A. Winter, "Coding theorem and strong converse for quantum channels", IEEE Trans. Inform. Theory, vol. 45, no. 7, pp. 2481–2485, 1999.

[26] A. Winter, "Extrinsic and intrinsic data in quantum measurements: asymptotic convex decomposition of positive operator valued measures", e–print `quant-ph/0109050` at `http://arXiv.org`, 2001.
[27] A. Winter, S. Massar, "Compression of quantum measurement operations", Phys. Rev. A, vol. 64, 012311, 2001.
[28] J. Wolfowitz, "The coding of messages subject to chance errors", Illinois J. Math., vol. 1, no. 4, pp. 591–606, 1957.