Coding Theorem and Strong Converse for Quantum Channels

Andreas Winter*
SFB 343, Fakultät für Mathematik
Universität Bielefeld, Postfach 100131, 33501 Bielefeld

September 7, 1998

Abstract

In this communication we present a new proof of Holevo's channel coding theorem for quantum channels, and its strong converse. The technique is largely inspired by Wolfowitz's combinatorial approach using types of sequences.

1 Introduction

After the recent achievements in quantum information theory, most notably Schumacher's quantum data compression [16], and the determination of the quantum channel capacity by Holevo [12] (and independently by Schumacher and Westmoreland [17]) building on ideas of Hausladen et.al. [7], we feel that one should try to convert other and stronger techniques of classical information theory than those used in the cited works to the quantum case. The present work will do this for the method of types, as it is called in [3] (and which constitutes a manifestly combinatorial approach to information theory), by rephrasing it in the operator language of quantum theory. This we apply to the coding problem for discrete memoryless quantum channels: we will give a new proof of the quantum channel coding theorem, and prove the strong converse. We point out that our techique is also suited to the situations of encoding under linear constraints, and with arbitrary signal states.

*Electronic address: winter@mathematik.uni-bielefeld.de

2 Prerequisites and notational conventions

We will use the definitions and notation of [18], in particular finite sets will be $\mathcal{A}$, $\mathcal{B}$, $\mathcal{M}$, $\mathcal{X}$, ..., quantum states (density operators) ρ, σ, ..., probability distributions P, Q, ..., and classical–quantum operations V, W, ... (also stochastic matrices), whereas general quantum operations (trace preserving, and completely positive maps) of C*–algebras $\mathfrak{X}$, $\mathfrak{Y}$, ... are denoted as adjoint maps φ_*, ψ_*, ... Our algebras will be of finite dimension, and apart from commutative ones we will confine ourselves to the C*–algebra $\mathfrak{L}(\mathcal{H})$, the algebra of (bounded) linear operators of the complex Hilbert space $\mathcal{H}$, even though everything works equally well for the general case.

The exponential function exp is always understood to basis 2, as well as the logarithm log. The same symbol H denotes the Shannon and the von Neumann entropy (as Shannon's is the commutative case of von Neumann's).

We will use two different norms on operators: the supremum norm $\|\cdot\|_\infty$ which is the largest absolute value of an eigenvalue, and (for states etc.) the trace norm $\|\cdot\|_1$ which is the sum of the absolute values of the eigenvalues. It is known (and not difficult to prove, using the polar decomposition of α, see Arveson [1]) that

$$\|\alpha\|_1 = \sup_{B:\|B\|_\infty \leq 1} |\mathrm{Tr}\,(\alpha B)|$$

If α is selfadjoint we may write it as the difference $\alpha_+ - \alpha_-$ of its positive and its negative part. Observe that then

$$\|\alpha\|_1 = \mathrm{Tr}\,(\alpha_+) + \mathrm{Tr}\,(\alpha_-) = \sup_{-\mathbb{1}\leq B\leq \mathbb{1}} \mathrm{Tr}\,(\alpha B)$$

Another useful formula is

$$\mathrm{Tr}\,(\alpha_+) = \sup_{0\leq B\leq \mathbb{1}} \mathrm{Tr}\,(\alpha B)$$

3 Quantum sources and channels

The following definition is from [9]: a (discrete memoryless) *classical–quantum channel* (cq–DMC) is a mapping W from a finite set $\mathcal{X}$ into the states on a Hilbert space $\mathcal{H}$, taking $x \in \mathcal{X}$ to W_x (by linear extension we may view this as a quantum operation from $(\mathbb{C}\mathcal{X})_*$ to $\mathfrak{L}(\mathcal{H})_*$).

From [18] we recall: for a probability distribution P on $\mathcal{X}$ let $PW = \sum_{x\in\mathcal{X}} P(x)W_x$

the average state of the channel W. $H(W|P) = \sum_{x\in\mathcal{X}} P(x)H(W_x)$ is the conditional von Neumann entropy, the mutual information between a distribution and the channel is $I(P,W) = H(PW) - H(W|P)$. Note that these notions still make sense for infinite $\mathcal{X}$ if only W is required to be measurable (so $\mathcal{X}$ has to carry some measurable structure): then also $H(W)$ is measurable, and PW, $H(W|P)$ are expectations over the probability measure P.

A *n–block code* for a quantum channel W is a pair (f,D), where f is a mapping from a finite set $\mathcal{M}$ into $\mathcal{X}^n$, and D is an observable on $\mathcal{H}^{\otimes n}$ indexed by $\mathcal{M}' \supset \mathcal{M}$, i.e. a partition of $\mathbb{1}$ into positive operators D_m, $m \in \mathcal{M}'$. The (maximum) error probability of the code is defined as

$$e(f,D) = \max\{1 - \mathrm{Tr}\,(W_{f(m)}D_m)|m \in \mathcal{M}\}$$

We call (f,D) an (n,ϵ)–code, if $e(f,D) \le \epsilon$. The *rate* of an n–block code is defined as $\frac{1}{n}\log|\mathcal{M}|$.

4 Techniques

4.1 Types, shells, η–images

Let n a positve integer, and consider words $\mathbf{x}$ of length n from $\mathcal{X}^n$. The *type* of $\mathbf{x}$ is the probability distribution on $\mathcal{X}$ of the relative frequencies of $x \in \mathcal{X}$ in $\mathbf{x}$, denoted $P_{\mathbf{x}}$. Given a probability distribution P on $\mathcal{X}$ the set of words of type P will be denoted by $\mathcal{T}_P = \mathcal{T}_P^n$ (often we will suppress the superscript n), cf. [3], whose notion of type and typicality we employ, instead of the even stronger concept of Wolfowitz [19]. This is mainly due to greater convenience with the operator calculus.

Fix diagonalizations of the W_x, say

$$W_x = \sum_{k\in\mathcal{K}} W(k|x)|\psi_{xk}\rangle\langle\psi_{xk}|, \qquad |\mathcal{K}| = \dim\mathcal{H}$$

(where $W(\cdot|\cdot)$ is a stochastic matrix, the double meaning of W should be no serious ambiguity). For words $\mathbf{x}$ this is extended by considering tensor powers, and the tensor products of the $|\psi_{xk}\rangle$ appearing there will be denoted $|\psi_{\mathbf{xk}}\rangle$.

For a cq–DMC V let us say (following [3], Definition 1.2.4) that a sequence of unit vectors $(|\psi_1\rangle, \dots, |\psi_n\rangle)$ (in the sequel identified with their tensor product — phase is unimportant, as will be seen) has *conditional type V given* $\mathbf{x}$, if for all x, and $n_x = |\{i|x_i = x\}| \neq 0$:

$$\frac{1}{n_x} \sum_{i:x_i=x} |\psi_i\rangle\langle\psi_i| = V_x$$

The V–*shell* $T_V^n(\mathbf{x})$ of $\mathbf{x}$ is defined as the set of all $\mathbf{k}$ such that $\psi_{\mathbf{xk}}$ has conditional type V given $\mathbf{x}$. Assume that the V–shell is not empty: notice that for those x with $n_x \neq 0$ V_x commutes with W_x, and we assume this also for the x with $n_x = 0$.

But since there is no natural way of comparing the diagonalizations of different states we introduce the projection operators $\Pi_V(\mathbf{x}) = \Pi_V^n(\mathbf{x})$, also called the V–shell of $\mathbf{x}$, defining

$$\Pi_V(\mathbf{x}) = \sum_{\mathbf{k} \in T_V(\mathbf{x})} |\psi_{\mathbf{xk}}\rangle\langle\psi_{\mathbf{xk}}|$$

For the special case of $|\mathcal{X}| = 1$, i.e. a single state ρ, we obtain the *typical projectors* Π_σ. Note: Π_σ, $\Pi_V(\mathbf{x})$ depend only on ρ, W, respectively, not on the particular diagonalization chosen.

As in [3] we introduce now the set of words of approximate type P with constant δ, denoted $T^n_{[P]_\delta}$, as the (disjoint) union of all T^n_Q with $|Q(x) - P(x)| \leq \delta$ for all $x \in \mathcal{X}$, i.e. $\|Q - P\|_\infty \leq \delta$ (with the supremum norm on distributions), and $\operatorname{supp} Q \subset \operatorname{supp} P$.

Similarly the projector of approximate conditional type W given $\mathbf{x}$ with constant δ, $\Pi^n_{[W]_\delta}(\mathbf{x})$, is the (orthogonal!) sum of all V-shells $\Pi^n_V(\mathbf{x})$ with $\|P_{\mathbf{x}}(x)V_x - P_{\mathbf{x}}(x)W_x\|_\infty \leq \delta$ for all $x \in \mathcal{X}$ (with the operator supremum norm: since V_x and W_x commute this can be calculated from the eigenvalues of their common diagonalizations) and $\operatorname{supp} V_x \subset \operatorname{supp} W_x$.

Again for the special case of $|\mathcal{X}| = 1$ we get the approximate typical projector with constant δ, denoted $\Pi^n_{[\rho]_\delta}$.[1]

Note 1 (Delta–convention) *Given $\mathcal{X}$ and $\mathcal{H}$ fix a sequence δ_n with $\delta_n \longrightarrow 0$, $\sqrt{n}\delta_n \longrightarrow \infty$ as $n \longrightarrow \infty$. In the typical sets and typical projectors with constant δ it is understood that δ_n is meant, and we usually suppress this in the notation.*

We will at times employ several different δ–sequences, and will then require a few simple relations between the various sequences (see note after lemma 10).

For $0 < \eta \leq 1$ a positive operator B on $\mathcal{H}$ with $B \leq \mathbb{1}$ will be called an η–*image* of $\mathcal{A} \subset \mathcal{X}$ if for all $a \in \mathcal{A}$: $\operatorname{Tr}(W_a B) \geq \eta$. The *size* of the operator B is defined as $\operatorname{Tr} B$,

[1] Distinguish from the typical projector introduced by Schumacher [16] (which we would like to call *entropy typical*), though there are similarities, as will be seen below.

and $g_W(\mathcal{A},\eta)$ is the minimal size of η–images of $\mathcal{A}$.

The following basic results are in the classical situation from [3].

Lemma 2 (Type counting) *There are at most $(n+1)^{|\mathcal{X}|}$ types of sequences from $\mathcal{X}$. There are at most $(n+1)^{|\mathcal{X}|\dim\mathcal{H}}$ nonzero V–shells $\Pi_V(\mathbf{x})$.*

Proof. By definition this reduces to classical type counting. □

Lemma 3 (Shell size) *For every $\mathbf{x}\in\mathcal{X}$ and every stochastic matrix V with $\Pi_V(\mathbf{x})\neq 0$:*

$$(n+1)^{-|\mathcal{X}|\dim\mathcal{H}}\exp(nH(V|P_{\mathbf{x}}))\leq \operatorname{Tr}\Pi_V(\mathbf{x})\leq\exp(nH(V|P_{\mathbf{x}}))$$

and in particular for a state ρ, and $\Pi_\sigma\neq 0$

$$(n+1)^{-\dim\mathcal{H}}\exp(nH(\sigma))\leq\operatorname{Tr}\Pi_\sigma\leq\exp(nH(\sigma))$$

Proof. Observe that $\operatorname{Tr}\Pi_V(\mathbf{x})=|T_V(\mathbf{x})|$ and use the classical estimate (cf. [3], lemma 1.2.5). □

Lemma 4 (Continuity) *Let ρ,σ states with $\|\rho-\sigma\|_1\leq\theta\leq\frac{1}{2}$. Then*

$$|H(\rho)-H(\sigma)|\leq-\theta\log\frac{\theta}{\dim\mathcal{H}}$$

Proof. See [15]. □

Lemma 5 (Typical subspace) *There exists a sequence $\epsilon_n\longrightarrow 0$ depending only on $|\mathcal{X}|$ and $\dim\mathcal{H}$ such that for every distribution P and channel W*

$$\left|\frac{1}{n}\log(\operatorname{Tr}\Pi^n_{[W]}(\mathbf{x}))-H(W|P)\right|\leq\epsilon_n \textit{ for all } \mathbf{x}\in T^n_{[P]}$$

In particular for a state ρ

$$\left|\frac{1}{n}\log(\operatorname{Tr}\Pi^n_{[\rho]})-H(\rho)\right|\leq\epsilon_n$$

Proof. With $\operatorname{Tr}\Pi^n_{[W]}(\mathbf{x})=|T^n_{[W]}(\mathbf{x})|$ this is a direct consequence of the shell size lemma 3 with continuity and type counting. □

Lemma 6 (Large weight) *For any cq–DMC W, integral n, $\delta > 0$, and $\mathbf{x} \in \mathcal{X}^n$:*

$$\mathrm{Tr}\,(W_{\mathbf{x}}\Pi_{[W]_\delta}(\mathbf{x})) \geq 1 - \frac{|\mathcal{X}| \dim \mathcal{H}}{4n\delta^2}$$

In particular for a state ρ

$$\mathrm{Tr}\,(\rho^{\otimes n}\Pi_{[\rho]_\delta}) \geq 1 - \frac{\dim \mathcal{H}}{4n\delta^2}$$

Proof. Just as in the proof of the lemma 1.2.12 of [3] this is an application of Chebyshev's inequality. □

The previous lemmas show that the typical projection operators are fairly good η–images whose size is controlled by certain (conditional) entropies. The following lemmas show that arbitrary η–images cannot essentially beat these bounds:

Lemma 7 (Considerable mass size) *Given $0 < \eta < 1$ there exists a sequence $\epsilon_n \longrightarrow 0$ (depending only on η, $|\mathcal{X}|$ and $\dim \mathcal{H}$) such that: if B is a positive operator, upper bounded by $\mathbb{1}$, on $\mathcal{H}^{\otimes n}$, with $\mathrm{Tr}\,(W_{\mathbf{x}}B) \geq \eta$ then*

$$\frac{1}{n} \log \mathrm{Tr}\, B \geq H(W|P_{\mathbf{x}}) - \epsilon_n$$

In particular, if ρ is a state on $\mathcal{H}$, and $0 \leq A \leq \mathbb{1}$, then $\mathrm{Tr}\,(\rho^{\otimes n}A) \geq \eta$ implies

$$\frac{1}{n} \log \mathrm{Tr}\, A \geq H(\rho) - \epsilon_n$$

Proof. For n large enough (depending on $|\mathcal{X}|$, $\dim \mathcal{H}$ and η) we have (by the large weight lemma 6) $\mathrm{Tr}\,(\Pi_{[W]}(\mathbf{x})W_{\mathbf{x}}\Pi_{[W]}(\mathbf{x})) \geq 1 - \eta/2$, thus

$$\mathrm{Tr}\,(\Pi_{[W]}(\mathbf{x})W_{\mathbf{x}}\Pi_{[W]}(\mathbf{x})B) \geq \eta/2$$

Now observe that $\Pi_{[W]}(\mathbf{x})$ is an orthogonal sum of certain V-shells $\Pi_V(\mathbf{x})$, and all the V-shells commute with $W_{\mathbf{x}}$ (since they are constructed from eigenvectors). So

$$\begin{aligned}
\eta/2 &\leq \mathrm{Tr}\,(\Pi_{[W]}(\mathbf{x})W_{\mathbf{x}}\Pi_{[W]}(\mathbf{x})B) \\
&= \sum_V \mathrm{Tr}\,(\Pi_V(\mathbf{x})W_{\mathbf{x}}\Pi_V(\mathbf{x})B) \\
&= \sum_V \mathrm{Tr}\,(W_{\mathbf{x}}\Pi_V(\mathbf{x})B\Pi_V(\mathbf{x})) \\
&= \sum_V W(T_V(\mathbf{x})|\mathbf{x}) \frac{\mathrm{Tr}\,(\Pi_V(\mathbf{x})B\Pi_V(\mathbf{x}))}{|T_V(\mathbf{x})|}
\end{aligned}$$

and since the $W(T_V(\mathbf{x})|\mathbf{x})$ sum to at most 1 there is one V with

$$\eta/2 \cdot \mathrm{Tr}\,(\Pi_V(\mathbf{x})) \leq \mathrm{Tr}\,(\Pi_V(\mathbf{x})B\Pi_V(\mathbf{x})) \leq \mathrm{Tr}\,B$$

and the assertion follows from the shell size lemma 3 and the continuity of entropy. $\square$

Lemma 8 (Weak law) *Let $\rho_1, \dots, \rho_n$ be states and $\delta, \delta' > 0$ such that $\|\frac{1}{n}\sum_i \rho_i - \rho\|_\infty \leq \delta'$. Then*

$$\mathrm{Tr}\,(\rho_1 \otimes \cdots \otimes \rho_n \Pi^n_{[\rho]_{\delta+\delta'}}) \geq 1 - \frac{\dim \mathcal{H}}{4n\delta^2}$$

Proof. Consider the diagonalization $\rho = \sum_l q_l |\varphi_l\rangle\langle\varphi_l|$, and the *corset* map

$$\kappa_* : \sigma \longmapsto \sum_l |\varphi_l\rangle\langle\varphi_l|\sigma|\varphi_l\rangle\langle\varphi_l|$$

Defining $\rho_i' = \kappa_*(\rho_i)$ we claim that

$$\Pi^n_{[\frac{1}{n}\sum_i \rho_i']_\delta} \leq \Pi^n_{[\rho]_{\delta+\delta'}}$$

Indeed observe that we have

$$\|\frac{1}{n}\sum_i \rho_i' - \rho\|_\infty \leq \|\frac{1}{n}\sum_i \rho_i - \rho\|_\infty \leq \delta'$$

Thus for $\mathbf{l} = (l_1, \dots, l_n)$ with

$$\|\frac{1}{n}\sum_i |\varphi_{l_i}\rangle\langle\varphi_{l_i}| - \frac{1}{n}\sum_i \rho_i'\|_\infty \leq \delta$$

we have by triangle inequality

$$\|\frac{1}{n}\sum_i |\varphi_{l_i}\rangle\langle\varphi_{l_i}| - \rho\|_\infty \leq \delta + \delta'$$

So we can estimate

$$\begin{aligned}
\mathrm{Tr}\,(\rho_1 \otimes \cdots \otimes \rho_n \Pi^n_{[\rho]_{\delta+\delta'}}) &\geq \mathrm{Tr}\,(\rho_1 \otimes \cdots \otimes \rho_n \Pi^n_{[\frac{1}{n}\sum_i \rho_i']_\delta}) \\
&= \mathrm{Tr}\,(\rho_1' \otimes \cdots \otimes \rho_n' \Pi^n_{[\frac{1}{n}\sum_i \rho_i']_\delta}) \\
&\geq 1 - \frac{\dim \mathcal{H}}{4n\delta^2}
\end{aligned}$$

the last line by Chebyshev's inequality. $\square$

We will apply this lemma in two important situations.

The first gives precise meaning to the intuition that for typical $\mathbf{x}$ the conditional approximate W–shell $\Pi_{[W]}(\mathbf{x})$ is close to being contained in the typical subspace $\Pi_{[PW]}$:

Lemma 9 (Induced typicality) *(Compare [3], Lemma 1.2.10) Let* $\mathbf{x} \in T_{[P]_\delta}$, $\mathbf{k} \in T_{[W]_{\delta'}}(\mathbf{x})$. *Then*

$$\mathrm{Tr}\left(|\psi_{\mathbf{xk}}\rangle\langle\psi_{\mathbf{xk}}|\Pi_{[PW]_{\delta^*}}\right) \geq 1 - \frac{\dim\mathcal{H}}{4n\delta''^2}$$

for $\delta^* = \delta'' + (\delta+\delta')|\mathcal{X}|\dim\mathcal{H}$. *Hence*

$$\sum_{\mathbf{k}\in T_{[W]_{\delta'}}(\mathbf{x})} |\psi_{\mathbf{xk}}\rangle\langle\psi_{\mathbf{xk}}|\Pi_{[PW]_{\delta^*}}|\psi_{\mathbf{xk}}\rangle\langle\psi_{\mathbf{xk}}| \geq \left(1 - \frac{\dim\mathcal{H}}{4n\delta''^2}\right)\Pi_{[W]_{\delta'}}(\mathbf{x})$$

And in particular for the corset map κ *of an eigenbasis of* $W_{\mathbf{x}}$

$$\kappa(\Pi_{[PW]_{\delta^*}}) \geq \left(1 - \frac{\dim\mathcal{H}}{4n\delta''^2}\right)\Pi_{[W]_{\delta'}}(\mathbf{x})$$

Proof. W.l.o.g. $|\psi_{\mathbf{xk}}\rangle = \bigotimes_{x\in\mathcal{X},k\in\mathcal{K}} |\psi_{xk}\rangle^{\otimes n_{xk}}$ with

$$\left|\frac{n_x}{n} - P(x)\right| \leq \delta, \quad \left|\frac{n_{xk}}{n_x} - W(k|x)\right| \leq \delta'$$

and so

$$\left|\frac{n_{xk}}{n} - P(x)W(k|x)\right| \leq \delta + \delta'$$

(with the additional requirement that $n_x = 0$ if $P(x) = 0$ and $n_{xk} = 0$ if $P(x)W(k|x) = 0$). Then obviously

$$\begin{aligned}\|\frac{1}{n}\sum_i |\psi_{x_ik_i}\rangle\langle\psi_{x_ik_i}| - PW\|_\infty &= \|\sum_{x,k}\frac{n_{xk}}{n}|\psi_{xk}\rangle\langle\psi_{xk}| - \sum_{x,k}P(x)W(k|x)|\psi_{xk}\rangle\langle\psi_{xk}|\|_\infty \\ &\leq (\delta+\delta')|\mathcal{X}|\dim\mathcal{H}\end{aligned}$$

Now apply the weak law lemma 8. □

This lemma implies that not only the relatively typical projector $\Pi_{[W]}(\mathbf{x})$ carries much weight of $W_{\mathbf{x}}$ but also the global typical projector $\Pi_{[PW]}$, at least for $\mathbf{x}$ with approximate type P. The bound for this weight can be deduced from the lemma, but we can improve on it:

Lemma 10 (Global typicality) *If* $\mathbf{x} \in T^n_{[P]_{\delta'}}$ *then* $\mathrm{Tr}\left(W_{\mathbf{x}}\Pi_{[PW]_{\delta+\delta'|\mathcal{X}|}}\right) \geq 1 - \dfrac{\dim\mathcal{H}}{4n\delta^2}$.

Proof. Apply the weak law lemma 8 with $\rho_i = W_{x_i}$. □

If we use global typicality or induced typicality it will be implicit that the corresponding δ–sequences satisfy the relations of the preceding lemmas, respectively.

4.2 Coding and decoding

Lemma 11 *Let ρ be a state, and X a positive operator with $X \leq \mathbb{1}$ and $1 - \mathrm{Tr}\,(\rho X) \leq \epsilon < 1$. Then*

$$\|\rho - \sqrt{X}\rho\sqrt{X}\|_1 \leq \sqrt{8\epsilon}$$

Proof. Let $Y = \sqrt{X}$ and write $\rho = \sum_k p_k \pi_k$ with one–dimensional projectors π_k and weights $p_k \geq 0$. Now

$$\begin{aligned}
\|\rho - Y\rho Y\|_1^2 &\leq \left(\sum_k p_k \|\pi_k - Y\pi_k Y\|_1\right)^2 \\
&\leq \sum_k p_k \|\pi_k - Y\pi_k Y\|_1^2 \\
&\leq 4\sum_k p_k (1 - \mathrm{Tr}\,(\pi_k Y \pi_k Y)) \\
&\leq 8\sum_k p_k (1 - \mathrm{Tr}\,(\pi_k Y)) \\
&= 8(1 - \mathrm{Tr}\,(\rho Y)) \\
&\leq 8(1 - \mathrm{Tr}\,(\rho X)) = 8\epsilon
\end{aligned}$$

by triangle inequality, convexity of $x \mapsto x^2$, lemma A.2, $1 - x^2 \leq 2(1-x)$, and $X \leq Y$. □

Lemma 12 (Tender measurement) *Let ρ_a ($a \in \mathcal{A}$) be a set of states on $\mathcal{H}$, and D an observable with values in $\mathcal{B}$. Let further $\varphi : \mathcal{A} \longrightarrow \mathcal{B}$ be a map and $\epsilon > 0$ such that*

$$\forall a \in \mathcal{A} \qquad 1 - \mathrm{Tr}\,(\rho_a D_{\varphi(a)}) \leq \epsilon$$

(I.e. the observable identifies $\varphi(a)$ from ρ_a with maximal error probability ϵ). Then the canonically corresponding (trace preserving) quantum operation

$$D_{\mathrm{int}*} : \rho \longmapsto \sum_{b\in\mathcal{B}} \sqrt{D_b}\rho\sqrt{D_b}$$

disturbes the states ρ_a only a little:

$$\forall a \in \mathcal{A} \qquad \|\rho_a - D_{\mathrm{int}*}\rho_a\|_1 \leq \sqrt{8\epsilon} + \epsilon$$

Furthermore the total observable operation

$$D_{\mathrm{tot}*} : \rho \longmapsto \sum_{b\in\mathcal{B}} [b] \otimes \sqrt{D_b}\rho\sqrt{D_b}$$

satisfies

$$\forall a \in \mathcal{A} \qquad \|[\varphi(a)] \otimes \rho_a - D_{\mathrm{tot}*}\rho_a\|_1 \leq \sqrt{8\epsilon} + \epsilon$$

Proof. An easy calculation:

$$\begin{aligned}
\|\rho_a - D_{\mathrm{int}*}\rho_a\|_1 &\leq \|\rho_a - \sqrt{D_{\varphi(a)}}\rho_a\sqrt{D_{\varphi(a)}}\|_1 + \sum_{b\neq\varphi(a)} \|\sqrt{D_b}\rho_a\sqrt{D_b}\|_1 \\
&= \|\rho_a - \sqrt{D_{\varphi(a)}}\rho_a\sqrt{D_{\varphi(a)}}\|_1 + \sum_{b\neq\varphi(a)} \mathrm{Tr}\,(\rho_a D_b) \\
&\leq \sqrt{8(1-\mathrm{Tr}\,(\rho_a D_{\varphi(a)}))} + (1-\mathrm{Tr}\,(\rho_a D_{\varphi(a)})) \\
&\leq \sqrt{8\epsilon} + \epsilon
\end{aligned}$$

using triangle inequality and lemma 11. The second part (which actually implies the first) is similar. □

Note 13 *If we modify the statement of the lemma to that the average error in identifying $\varphi(a)$ from ρ_a should be at most $\bar{\epsilon}$ (relative a distribution on $\mathcal{A}$), then also the distortion bound of the lemma holds — on average: see appendix A.*

Lemma 14 (Maximal code) *(Compare [3], Lemma 2.1.3) For $0 < \tau < \epsilon < 1$, cq–DMC W, distribution P on $\mathcal{X}$ and $\mathcal{A} \subset T^n_{[P]}$ there exists a (n,ϵ)–code (f,D) such that for every $m \in \mathcal{M}$*

$$f(m) \in \mathcal{A}, \text{ and } \mathrm{Tr}\, D_m \leq \mathrm{Tr}\,\Pi^n_{[W]}(f(m))$$

and (letting $\eta = \min\{1-\epsilon, \frac{1}{8}(\epsilon-\tau)^2\}$) with rate

$$\frac{1}{n}\log|\mathcal{M}| \geq \frac{1}{n}\log g_{W^n}(\mathcal{A},\eta) - H(W|P) - \tau$$

provided $n \geq n_0(|\mathcal{X}|, \dim\mathcal{H}, \tau)$.

Proof. Let (f,D) a maximal code with the stated properties and define $B = \sum_{m\in\mathcal{M}} D_m$. We claim that with η as defined

$$\forall \mathbf{x} \in \mathcal{A} \quad \mathrm{Tr}\,(W_{\mathbf{x}}B) \geq \eta$$

This is clear for codewords, and true for the others because otherwise we could extend our code by the codeword $\mathbf{x}$ with corresponding observable operator

$$D_{\mathbf{x}} = \sqrt{\mathbb{1}-B}\,\Pi^n_{[W]}(\mathbf{x})\sqrt{\mathbb{1}-B}$$

which clearly satisfies the trace bound (note that $B + D_{\mathbf{x}} \leq \mathbb{1}$): to see this apply lemma 11 to obtain

$$\|W_{\mathbf{x}} - \sqrt{\mathbb{1}-B}W_{\mathbf{x}}\sqrt{\mathbb{1}-B}\|_1 \leq \sqrt{8\eta} \leq \epsilon - \tau$$

Thus

$$\begin{aligned}\mathrm{Tr}\,(W_{\mathbf{x}}\sqrt{\mathbb{1}-B}\Pi^n_{[W]}(f(m))\sqrt{\mathbb{1}-B}) &= \mathrm{Tr}\,(W_{\mathbf{x}}\Pi_{[W]}(\mathbf{x})) - \\ &\quad - \mathrm{Tr}\,((W_{\mathbf{x}} - \sqrt{\mathbb{1}-B}W_{\mathbf{x}}\sqrt{\mathbb{1}-B})\Pi_{[W]}(\mathbf{x})) \\ &\geq (1-\tau) - (\epsilon-\tau) = 1-\epsilon\end{aligned}$$

So B is an η–image of $\mathcal{A}$, and consequently $\mathrm{Tr}\,B \geq g_{W^n}(\mathcal{A},\eta)$. On the other hand

$$\begin{aligned}\mathrm{Tr}\,B = \sum_{m\in\mathcal{M}} \mathrm{Tr}\,D_m &\leq \sum_{m\in\mathcal{M}} \mathrm{Tr}\,\Pi_{[W]}(f(m)) \\ &\leq |\mathcal{M}|\cdot\exp(n(H(W|P)+\tau))\end{aligned}$$

the last inequality by the typical subspace lemma 5, and we are done. □

Note 15 *It is interesting to note from the proof that the decoder may be chosen a von Neumann observable (i.e. all its operators are mutually orthogonal projectors). This is because if (f,D) is of this type, then B is a projector, and this means that we may instead of the constructed $D_{\mathbf{x}} \leq \mathbb{1} - B$ use the projector $D'_{\mathbf{x}} = \mathrm{supp}\,D_{\mathbf{x}}$: this is still bounded by $\mathbb{1} - B$, only decreases the error probability, and obeys the size condition:* $\mathrm{Tr}\,\mathrm{supp}\,D_{\mathbf{x}} = \dim\mathrm{im}D_{\mathbf{x}} \leq \dim\mathrm{im}\Pi^n_{[W]}(\mathbf{x}) = \mathrm{Tr}\,\Pi^n_{[W]}(\mathbf{x})$.

Lemma 16 (Code bound) *(Compare [3], Lemma 2.1.4) For $\epsilon,\tau > 0$, $\epsilon+\tau \leq 1$, if (f,D) is an (n,ϵ)–code for W such that all codewords belong to $\mathcal{A}\subset T^n_{[P]}$, then*

$$\frac{1}{n}\log|\mathcal{M}| \leq \frac{1}{n}\log g_{W^n}(\mathcal{A}, 1-2(1-\sqrt{2-\epsilon-\tau})^2) - H(W|P) + \tau$$

provided $n \geq n_0(|\mathcal{X}|,\dim\mathcal{H},\tau)$.

Proof. Let $\epsilon' = 2(1-\sqrt{2-\epsilon-\tau})^2$ and B a $(1-\epsilon')$–image of $\mathcal{A}$ with $\mathrm{Tr}\,B = g_{W^n}(\mathcal{A},1-\epsilon')$. From $1 - \mathrm{Tr}\,(W_{f(m)}B) \leq \epsilon'$ we get by lemma 11

$$\|W_{f(m)} - \sqrt{B}W_{f(m)}\sqrt{B}\|_1 \leq \sqrt{8\epsilon'}$$

and consequently

$$\begin{aligned}\mathrm{Tr}\,(W_{f(m)}\sqrt{B}D_m\sqrt{B}) &= \mathrm{Tr}\,(\sqrt{B}W_{f(m)}\sqrt{B}D_m) \\ &= \mathrm{Tr}\,(W_{f(m)}D_m) - \mathrm{Tr}\,\left((W_{f(m)} - \sqrt{B}W_{f(m)}\sqrt{B})D_m\right) \\ &\geq 1-\epsilon - \mathrm{Tr}\,\left((W_{f(m)} - \sqrt{B}W_{f(m)}\sqrt{B})_+\right) \\ &\geq 1-\epsilon - (\sqrt{2\epsilon'} + \frac{\epsilon'}{2}) = \tau\end{aligned}$$

(Observe that for $\rho = W_{f(m)} - \sqrt{B} W_{f(m)} \sqrt{B}$ we have $\operatorname{Tr} \rho \leq \epsilon'$ and $\|\rho\|_1 \leq \sqrt{8\epsilon'}$). Thus $\operatorname{Tr}(BD_m) = \operatorname{Tr}(\sqrt{B} D_m \sqrt{B}) \geq \exp(n(H(W|P) - \tau))$ by the considerable mass size lemma 7 for large enough n. Hence

$$\operatorname{Tr} B \geq \sum_{m \in \mathcal{M}} \operatorname{Tr}(BD_m) \geq |\mathcal{M}| \cdot \exp(n(H(W|P) - \tau))$$

which proves the lemma. □

The preceding lemmata reduce maximal code size (at least asymptotically) to the minimal size of η–images.

5 Coding theorem and strong converse

Given a cq–DMC W, n, $\mathcal{A} \subset \mathcal{X}^n$, and $\epsilon > 0$ define the ϵ–capacity $C_W(\mathcal{A}, \epsilon)$ to be the maximal rate of (n, ϵ)–codes all whose codewords belong to $\mathcal{A}$.

Theorem 1 (Type constrained code) *For all $\eta, \epsilon \in (0,1)$, $\tau > 0$, input distribution P and cq–DMC W: if $P^n(\mathcal{A}) \geq \eta$ then $C_W(\mathcal{A}, \epsilon) \geq I(P, W) - 2\tau$ whenever $n \geq n_0(|\mathcal{X}|, \dim \mathcal{H}, \tau, \epsilon, \eta)$.*

Proof. Let $\tilde{\mathcal{A}} = \mathcal{A} \cap T_{[P]}$ and n large enough so that $P^n(\tilde{\mathcal{A}}) \geq \eta/2$. In view of the maximal code lemma 4.14 it suffices to prove

$$\frac{1}{n} \log g_{W^n}(\tilde{\mathcal{A}}, \frac{1}{8}(\epsilon - \tau)^2) \geq H(PW) - \tau$$

To see this let B be a $(\epsilon - \tau)^2/8$–image of $\tilde{\mathcal{A}}$. Then $\operatorname{Tr}((PW)^{\otimes n} B) \geq (\epsilon - \tau)^2 \eta / 16$ and hence by the considerable mass size lemma 4.7 the desired inequality. □

Corollary 2 *For any channel W and $\epsilon > 0$ the ϵ–capacity of W, defined as $C_W(\epsilon) = \limsup_n C_W(\mathcal{X}^n, \epsilon)$, satisfies*

$$C_W(\epsilon) \geq \liminf_n C_W(\mathcal{X}^n, \epsilon) \geq \max_P I(P, W)$$

□

Theorem 3 (Strong Converse) *For any $\epsilon \in (0,1)$ and cq–DMC W:*

$$C_W(\epsilon) \leq \max_P I(P, W)$$

Proof. Splitting the code into equitype components, we have, in view of the code bound lemma 4.16 and type counting, to prove that

$$\frac{1}{n}\log g_{W^n}(T^n_{[P]}, 1-2(1-\sqrt{2-\epsilon-\tau})^2) \leq H(PW)+\tau$$

for $\epsilon+\tau < 1$ and n large enough (uniform in P).[2] But it follows by combining the induced typicality lemma 4.9 and the large weight lemma 4.6 (or alternatively by the global typicality lemma 4.10) that for any $\eta \in (0,1)$ $\Pi_{[PW]}$ is an η–image for $T^n_{[P]}$, if n is large enough. We conclude observing that by the typical subspace lemma 4.5 $\frac{1}{n}\log \operatorname{Tr}\Pi_{[PW]} \leq H(PW)+\tau$. Observe that the bound on n in the lemmas used does not depend on the distribution P. □

Note 4 *We have actually proved that $C_W(\epsilon)-\tau$ is achievable for all sufficiently large blocklengths n (not just for infinitely many, as some authors require in their definition of capacity), and $C_W(\epsilon)+\tau$ is never achievable but for finitely many exceptions. I.e. we showed identity of the "pessimistic" and the "optimistic" capacity concept (cf. [3], p.111). This remark also applies to the results of the following two subsections.*

Note 5 *It is interesting to note that the capacity may be achieved by using only codes with von Neumann observables (as was not clear from the constructions in [7], [12], [17]): see note 4.15.*
It would be nice on the other hand if we could decide if the decoder may consist of separable operators. It is clear that a product observable cannot do, as was pointed out by Holevo (see [10]): otherwise larger capacities could not be reached using block decoding. But it may be that nonlocality as in the recent work of Bennett et.al. [2] is sufficient, and genuine entanglement is not needed (as was proposed in the cited work of Holevo).

To conclude we want to point out that the same techique yields the coding theorem and strong converse under linear constraints (see Holevo's [11] for definitions and capacity formula): simply because satisfying the linear constraint is a property of whole types, not just individual sequences.
Also we can prove the coding theorem and strong converse in the case of arbitrary (product) signal states in a general (discrete memoryless) *quantum–quantum channel* (qq–DMC), see [9]: this is a completely positive and unit preserving map $\varphi : \mathfrak{A}_2 \to \mathfrak{A}_2$ between

[2]Observe that $1-2(1-\sqrt{2-\epsilon-\tau})^2 \in (0,1)$.

finite dimensional C*–algebras $\mathfrak{A}_1, \mathfrak{A}_2$ (or rather its adjoint $\varphi_* : \mathfrak{S}(\mathfrak{A}_1) \rightarrow \mathfrak{S}(\mathfrak{A}_2)$). This includes the cq–DMC as the special case $\mathfrak{A}_1 = \mathbb{C}\mathcal{X}$ and $\mathfrak{A}_2 = \mathfrak{L}(\mathcal{H})$.

An (n, ϵ)–code for this channel is a pair (F, D) with a map $F : \mathcal{M} \rightarrow \mathfrak{S}(\mathfrak{A}_1^{\otimes n})$ and an observable D on $\mathfrak{A}_2^{\otimes n}$ indexed by $\mathcal{M}' \supset \mathcal{M}$, with error probability

$$e(F, D) = \max\{1 - \mathrm{Tr}\,(\varphi_*^{\otimes n}(F(m)) \cdot D_m) | m \in \mathcal{M}\}$$

being at most ϵ. We will consider only the case that the $F(m)$ are product states (such codes we call *1–separable*, after [17], and the corresponding operational capacity *product state capacity* $C^{(1)}_{\varphi_*}(\epsilon)$), so the channel and all its possible codes are determined by the image of φ_* in $\mathfrak{S}(\mathfrak{A}_2)$, a compact convex set. Thus we are back in our original situation, with X now a compact convex set of states in $\mathfrak{L}(\mathcal{H})$ and $W = \mathrm{id}_X$. The capacity we denote $C_X(\epsilon)$, it was for $\epsilon \rightarrow 0$ determined Schumacher and Westmoreland [17] (an improved argument for their weak converse may be found in [18]).

Theorem 6 *With the notations above and for $\epsilon \in (0, 1)$: $C_X(\epsilon) = \sup_{\textit{finite } W \subset X} C_W$. Furthermore the* sup *is in fact a* max*, which is achieved by a finite $W \subset X$ with cardinality at most* $\dim_{\mathbb{C}} \mathfrak{Y})$*, and the W being extremal points of X.*

The details of the proof are relegated to a future, more extensive exposition: the basic idea is to replace the finiteness of $\mathcal{X}$ by the finite volume of X, and use approximation. The second part of the statement is from [6].

Acknowledgements

Thanks to Peter Löber for discussions on various aspects of the present work, and to Prof. Rudolf Ahlswede for his teaching of classical information theory.

A Distortion and fidelity

This appendix compares the distortion measure D coming from the trace norm $\|\cdot\|_1$

$$D(\rho, \sigma) = \frac{1}{2}\|\rho - \sigma\|_1$$

on the states of a quantum system with the (pure state) fidelity F. First note that $\|\cdot\|_1$ is indeed a distortion measure:

$$\|\rho_1 \otimes \rho_2 - \sigma_1 \otimes \sigma_2\|_1 \leq \|\rho_1 - \sigma_1\|_1 + \|\rho_2 - \sigma_2\|_1$$

In contrast, the fidelity has a multiplicative law under tensor products, so we might consider the log of it:

$$-\log F(\rho_1 \otimes \rho_2, \sigma_1 \otimes \sigma_2) = -\log F(\rho_1, \sigma_1) - \log F(\rho_2, \sigma_2)$$

But since $-\log F$ is unbounded as $F \to 0$ we prefer to use $1 - F$. Obviously

$$1 - F(\rho_1 \otimes \rho_2, \sigma_1 \otimes \sigma_2) \leq 1 - F(\rho_1, \sigma_1) + 1 - F(\rho_2, \sigma_2)$$

so $1 - F$ is also a distortion measure.

Lemma 1 (Pure state) *Let $\rho = |\psi\rangle\langle\psi|$ and $\sigma = |\phi\rangle\langle\phi|$ be pure states. Then*

$$1 - F(\rho, \sigma) = D(\rho, \sigma)^2$$

Proof. W.l.o.g. we may assume $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$ and $|\phi\rangle = \alpha|0\rangle - \beta|1\rangle$ ($|\alpha|^2 + |\beta|^2 = 1$). A straightforward calculation shows $F = (|\alpha|^2 - |\beta|^2)^2$, and $D = 2|\alpha\beta|$. Now

$$\begin{aligned} 1 - F &= 1 - (|\alpha|^2 - |\beta|^2)^2 = (1 + |\alpha|^2 - |\beta|^2)(1 - |\alpha|^2 + |\beta|^2) \\ &= 4|\alpha\beta|^2 = D^2 \end{aligned}$$

□

Lemma 2 (Mixed state) *Let σ an arbitrary mixed state (and ρ pure as above). Then*

$$D \geq 1 - F \geq D^2$$

Proof. Write $\sigma = \sum_j q_j |j\rangle\langle j|$. Then

$$\begin{aligned} 1 - F(\rho, \sigma) &= \sum_j q_j \left(1 - F(\rho, |j\rangle\langle j|)\right) = \sum_j q_j D(\rho, |j\rangle\langle j|)^2 \\ &\geq \left(\sum_j q_j D(\rho, |j\rangle\langle j|)\right)^2 \geq D(\rho, \sigma)^2 \end{aligned}$$

Conversely: extend $|\psi\rangle$ to an orthonormal basis $(|r\rangle)_r$ ($|0\rangle = |\psi\rangle$), and consider the trace preserving quantum operation

$$\kappa_* : \sigma \longmapsto \sum_{r \geq 0} |r\rangle\langle r|\sigma|r\rangle\langle r|$$

Then (with monotonocity of $\|\cdot\|_1$ under quantum operations)

$$2D = \|\rho - \sigma\|_1 \geq \|\kappa_* \rho - \kappa_* \sigma\|_1 = \|\rho - \kappa_* \sigma\|_1$$

(since $\rho = \kappa_* \rho$). Hence with $F = \langle\psi|\sigma|\psi\rangle$

$$\begin{aligned} 2D &\geq \|(1-F)|\psi\rangle\langle\psi| - \sum_{r\geq 1} \langle r|\sigma|r\rangle \cdot |r\rangle\langle r|\|_1 \\ &= (1-F) + \sum_{r\geq 1} \langle r|\sigma|r\rangle \\ &= 2(1-F) \end{aligned}$$

□

Observe that the inequalities of this lemma still hold if only $\sum_j q_j \leq 1$.

Lemma 3 (Ensembles) *Let ρ_i pure states, σ_i arbitrary states, and p_i probabilities. Then the average fidelity $\bar{F} = \sum_i p_i F(\rho_i, \sigma_i)$ and the average distortion $\bar{D} = \sum_i p_i D(\rho_i, \sigma_i)$ satisfy*

$$\bar{D} \geq 1 - \bar{F} \geq \bar{D}^2$$

Proof. From lemma 2 the inequality $\bar{D} \geq 1 - \bar{F}$ is immediate, and also the inequality $1 - \bar{F} \geq \bar{D}^2$ (with the convexity of the square function). □

As a consequence, the average (pure state) fidelity is close to one, if and only if the average trace norm distortion is small. In particular the conditions that $1 - \bar{F}$, $\bar{D}$ tend to zero, resp., are equivalent. Thus if comparing pure states with mixed states we may freely use either concept, but we will prefer D, as it easily generalizes to mixed states (whereas F is more involved), and has some nice properties. E.g. it does not increase under trace preserving quantum operations, as we already used above.

References

[1] W. Arveson, *An Invitation to C*–Algebras*, New York, Heidelberg, Berlin 1976

[2] C. H. Bennett, D. P. DiVincenzo, Ch. A. Fuchs, T. Mor, E. Rains, P. W. Shor, J. A. Smolin, "Nonlocality without Entanglement", LANL eprint `quant-ph/9804053`

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

[4] E. B. Davies, *Quantum Theory of Open Systems*, London 1976

[5] E. B. Davies, "Information and Quantum Measurement", IEEE Trans. Inf. Theory **24**,5(1978), 596–599

[6] A. Fujiwara, H. Nagaoka, "Operational Capacity and Pseudoclassicality of a Quantum Channel", IEEE Trans. Inf. Theory **44**,3(1998), 1071–1086

[7] P. Hausladen, R. Jozsa, B. Schumacher, M. Westmoreland, W. K. Wootters, Phys. Rev. A **54**,3(1996), 1869–1876

[8] A. S. Holevo, Problemy Peredachi Informatsii **9**,3(1973), 3–11 (english translation: "Bounds for the quantity of information transmitted by a quantum channel", Probl. Inf. Transm. **9**,3(1973), 177–183)

[9] A. S. Holevo, "Problems in the mathematical theory of quantum communication channels", Rep. Math. Phys. **12**,2(1977), 273–278

[10] A. S. Holevo, "Coding Theorems for Quantum Communication Channels", LANL eprint `quant-ph/9708046`

[11] A. S. Holevo, "On Quantum Communication Channels with Constrained Inputs", LANL eprint `quant-ph/9705054`

[12] A. S. Holevo, "The Capacity of the Quantum Channel with General Signal States", IEEE Trans. Inf. Theory **44**,1(1998), 269–273 (also available as LANL eprint `quant-ph/9611023`)

[13] R. Jozsa, M. Horodecki, P. Horodecki, R. Horodecki, "Universal Quantum Data Compression", LANL eprint `quant-ph/9805017`

[14] K. Kraus, *States, Effects, and Operations*, Springer Lecture Notes in Physics 190, Berlin 1983

[15] M. Ohya, D. Petz, *Quantum Entropy and its Use*, Berlin, Heidelberg, New York 1993

[16] B. Schumacher, "Quantum Coding", Phys. Rev. A **51**,4(1995), 2738–2747

[17] B. Schumacher, M. Westmoreland, “Sending classical information via noisy quantum channels”, Phys. Rev. A **56**,1(1997), 131–138

[18] A. Winter, “Languages of Quantum Information Theory”, LANL eprint `quant-ph/9807008`

[19] J. Wolfowitz, *Coding Theorems of Information Theory*, Berlin 1964