Commun. Math. Phys. 305, 85–97 (2011)
Digital Object Identifier (DOI) 10.1007/s00220-010-1172-y

Communications in
Mathematical
Physics

Hastings's Additivity Counterexample via Dvoretzky's Theorem

Guillaume Aubrun[1], **Stanisław Szarek**[2,3], **Elisabeth Werner**[3,4]

[1] Institut Camille Jordan, Université Claude Bernard Lyon 1, 43 Boulevard du 11 Novembre 1918, 69622 Villeurbanne Cedex, France. E-mail: aubrun@math.univ-lyon1.fr
[2] Equipe d'Analyse Fonctionnelle, Institut de Mathématiques de Jussieu, Université Pierre et Marie Curie-Paris 6, 4 Place Jussieu, 75252 Paris, France. E-mail: szarek@math.jussieu.fr
[3] Department of Mathematics, Case Western Reserve University, Cleveland, Ohio 44106, USA. E-mail: elisabeth.werner@case.edu
[4] Université de Lille 1, UFR de Mathématique, 59655 Villeneuve d'Ascq, France

Received: 21 June 2010 / Accepted: 11 August 2010
Published online: 26 November 2010 –

Abstract: The goal of this note is to show that Hastings's counterexample to the additivity of minimal output von Neumann entropy can be readily deduced from a sharp version of Dvoretzky's theorem.

1. Introduction

A fundamental problem in Quantum Information Theory is to determine the capacity of a quantum channel to transmit classical information. The seminal Holevo–Schumacher–Westmoreland theorem expresses this capacity as a regularization of the so-called Holevo χ-quantity (which gives the one-shot capacity) over multiple uses of the channel; see, e.g., [1]. This extra step could have been skipped if the χ-quantity had been additive, i.e., if

$$\chi(\Phi \otimes \Psi) = \chi(\Phi) + \chi(\Psi) \tag{1}$$

for every pair (Φ, Ψ) of quantum channels. It would have then followed that the χ-quantity and the capacity coincide, yielding a single-letter formula for the latter. Determining the veracity of (1) had been a major open problem for at least a decade (we refer, e.g., to the survey [2]). A substantial progress was made by Shor [3] who showed that (1) was formally equivalent to the additivity of the minimal output von Neumann entropy of quantum channels — a much more tractable quantity. Using this equivalence, the equality (1) was eventually shown to be false by Hastings [4], with appropriate randomly constructed channels as a counterexample.

In this note, we revisit Hastings's counterexample from the viewpoint of Asymptotic Geometric Analysis (AGA). This field — originally an offspring of Functional Analysis — aims at studying geometric properties of convex bodies (or equivalently, norms) in spaces of high (but finite) dimension. More specifically, our goal is to show that (a variant of) Hastings's analysis can be rephrased in the language of AGA, and his result deduced with only minor effort from a sharp version of Dvoretzky's theorem [5] on

almost spherical sections of convex bodies — a fundamental result of AGA. This makes the argument much more transparent and will hopefully lead to a better understanding of the problem of capacity. Our approach is largely inspired by Brandao–Horodecki [6], who were able to reformulate Hastings's analysis in the framework of concentration of measure.

2. Notation

Throughout the paper, the letters C, c, C', ... denote absolute positive constants, independent of the instance of the problem (most notably of the dimensions involved), whose values may change from occurrence to occurrence. The values of these constants can be computed by reverse-engineering the argument, but we will not pursue this task. We also use the following convention: whenever a formula is given for the dimension of a (sub)space, it is tacitly understood that one should take the integer part.

Let $\mathcal{M}_{k,d}$ be the space of $k \times d$ matrices (with complex entries), and $\mathcal{M}_d = \mathcal{M}_{d,d}$. More generally, $\mathcal{B}(\mathcal{H})$ will stand for the space of (bounded) linear operators on the Hilbert space $\mathcal{H}$. We will write $\|\cdot\|_p$ for the Schatten p-norm $\|A\|_p = \left(\mathrm{Tr}(A^\dagger A)^{p/2}\right)^{1/p}$. The limit case $\|\cdot\|_\infty$ is the operator (or "spectral") norm, while $\|\cdot\|_{HS} = \|\cdot\|_2$ is the Hilbert–Schmidt (or Frobenius) norm. Let $\mathcal{D}(\mathbf{C}^d)$ be the set of *density matrices* on $\mathbf{C}^d$, i.e., positive semi-definite trace one operators on $\mathbf{C}^d$ (or *states* on $\mathbf{C}^d$). If ρ is a state on $\mathbf{C}^d$, its *von Neumann entropy* $S(\rho)$ is defined as $S(\rho) = -\mathrm{Tr}\,\rho \log \rho$. If $\Phi : \mathcal{M}_m \to \mathcal{M}_k$ is a *quantum channel* (completely positive trace preserving map), its *minimal output entropy* is

$$S_{\min}(\Phi) = \min_{\rho \in \mathcal{D}(\mathbf{C}^m)} S(\Phi(\rho)).$$

Concavity of S implies that the minimum is achieved on a pure state.

3. Channels as Subspaces

The crucial insight allowing to relate analysis of quantum channels to high-dimensional convex geometry is the observation that there is an essentially one-to-one correspondence between channels and linear subspaces of composite Hilbert spaces. Specifically, let $\mathcal{W}$ be a subspace of $\mathbf{C}^k \otimes \mathbf{C}^d$ of dimension m. Then $\Phi : \mathcal{B}(\mathcal{W}) \to \mathcal{M}_k$ defined by $\Phi(\rho) = \mathrm{Tr}_{\mathbf{C}^d}(\rho)$ is a quantum channel; here $\mathrm{Tr}_{\mathbf{C}^d}$ is the *partial trace* with respect to the second factor in $\mathbf{C}^k \otimes \mathbf{C}^d$. Alternatively, and perhaps more properly, we could identify $\mathcal{W}$ with $\mathbf{C}^m$ via an isometry $V : \mathbf{C}^m \to \mathbf{C}^k \otimes \mathbf{C}^d$ whose range is $\mathcal{W}$ and define, for $\rho \in \mathcal{M}_m$, the corresponding channel $\Phi : \mathcal{M}_m \to \mathcal{M}_k$ by

$$\Phi(\rho) = \mathrm{Tr}_{\mathbf{C}^d}(V \rho V^\dagger). \tag{2}$$

It is now easy to define a natural family of random quantum channels. They will be associated, via the above scheme, to random m-dimensional subspaces $\mathcal{W}$ of $\mathbf{C}^k \otimes \mathbf{C}^d$, distributed according to the Haar measure on the corresponding Grassmann manifold (for some fixed positive integers m, d, k that will be specified later). Note that all reasonable parameters of a channel defined by (2) such as $S_{\min}(\Phi)$ depend only on the subspace $\mathcal{W} = V(\mathbf{C}^m)$ and not on a particular choice of the isometry V (this will be also obvious from what follows). In particular, the language of "random m-dimensional subspaces of $\mathbf{C}^k \otimes \mathbf{C}^d$" is equivalent to that of "random isometries from $\mathbf{C}^m$ to $\mathbf{C}^k \otimes \mathbf{C}^d$."

4. The Additivity Conjectures and the Main Theorem

The following question has attracted considerable attention in the last few years: if Φ and Ψ are two quantum channels, is it true that

$$S_{\min}(\Phi \otimes \Psi) = S_{\min}(\Phi) + S_{\min}(\Psi) \ ? \tag{3}$$

Shor [3] showed it to be formally equivalent to a number of central questions in quantum information theory, including the additivity of the χ-quantity mentioned in the Introduction.

Note that the inequality "$\leq$" always holds (consider product input states). However, as was first proved by Hastings using random constructions [4], the reverse inequality is false in general. The exegesis of Hastings's argument has subsequently been carried out in [6] and [7]. We will show here that the analysis of (a variant of) Hastings's example essentially amounts to applying the right version of Dvoretzky's theorem and leads to the conclusion that high-dimensional random channels typically violate (3).

Theorem 1. *Let $k \in \mathbf{N}$, $m = ck^2$ and $d = Ck^2$ (c and C being appropriate absolute constants). Let $V : \mathbf{C}^m \to \mathbf{C}^k \otimes \mathbf{C}^d$ be a random isometry and $\Phi : \mathcal{M}_m \to \mathcal{M}_k$ be the corresponding random channel given by (2). Then for k large enough, with large probability,*

$$S_{\min}(\Phi \otimes \bar{\Phi}) < S_{\min}(\Phi) + S_{\min}(\bar{\Phi}).$$

The expression "with large probability" in Theorem 1 and in what follows may be understood as "with probability $> \theta$, where $\theta \in (0, 1)$ is arbitrary but fixed in advance" (note that, in particular, the threshold value of k could then depend on θ). However, much stronger assertions are in fact true, for example the probability of the exceptional set in Theorem 1 can be majorized by $\exp(-c'm)$. Another comment: one only uses in the proof that m and d are comparable, and larger than ck^2.

The proof will be based on separately majorizing $S_{\min}(\Phi \otimes \bar{\Phi})$, which is done via a well-known and relatively simple trick, and on minorizing $S_{\min}(\Phi) = S_{\min}(\bar{\Phi})$, which is the main point of the argument.

A question analogous to (3) can be asked for the minimal output p-Rényi entropy ($p > 1$). For the additivity of Rényi entropy, random counterexamples were constructed earlier by Hayden–Winter [8]. It was shown in [9] that the Hayden–Winter analysis can also be simplified (at least conceptually) by appealing to Dvoretzky's theorem. Working with the von Neumann entropy, however, requires more effort. First, while [9] relied on a straightforward instance of Milman's "tangible" version [10,11] of Dvoretzky's theorem for Schatten classes that was documented in the literature already in the 1970's, we now need a more subtle, sharp version (which appears in the literature only implicitly). Second, this sharp version is not applied in the most direct way and requires additional preparatory work (for which we mostly follow the approach of Brandao–Horodecki [6]).

5. Lower Bound for $S_{\min}(\Phi)$: the Approach

Since we are going to consider channels with near-maximal minimal output entropy, the following simple inequality (Lemma III.1 in [6], or formula (40) in [4]) will allow to replace the analysis of the von Neumann entropy S by that of a smoother quantity.

Lemma 2. *For every state $\sigma \in \mathcal{D}(\mathbf{C}^k)$,*

$$S(\sigma) \geq S\left(\frac{\mathrm{Id}}{k}\right) - k \left\| \sigma - \frac{\mathrm{Id}}{k} \right\|_{HS}^2 .$$

Consequently, for every quantum channel $\Phi : \mathcal{M}_m \to \mathcal{M}_k$,

$$S_{\min}(\Phi) \geq \log(k) - k \cdot \max_{\rho \in \mathcal{D}(\mathbf{C}^m)} \left\| \Phi(\rho) - \frac{\mathrm{Id}}{k} \right\|_{HS}^2 . \tag{4}$$

It will be convenient to identify $\mathbf{C}^k \otimes \mathbf{C}^d$ (or, to be more precise, $\mathbf{C}^k \otimes \overline{\mathbf{C}^d}$ — a distinction we will ignore) with $\mathcal{M}_{k,d}$ via the canonical map induced by $u \otimes v \to |u\rangle\langle v|$. If $x \in \mathbf{C}^k \otimes \mathbf{C}^d$ is so identified with a matrix $M \in \mathcal{M}_{k,d}$, then

$$\mathrm{Tr}_{\mathbf{C}^d} |x\rangle\langle x| = MM^\dagger . \tag{5}$$

Via this identification, Schmidt coefficients of $|x\rangle$ coincide with singular values of M. While the tensor and matrix formalisms are equivalent, the matrix formalism is arguably more transparent, which sometimes leads to simpler arguments.

Denote by $\mathcal{W} \subset \mathbf{C}^k \otimes \mathbf{C}^d$ the subspace inducing Φ. Note that the maximum in (4) is necessarily attained on pure states which, in this identification, correspond to unit vectors $x \in \mathcal{W}$. For such states the action of Φ is given — in the matrix formalism — by (5), and so the inequality (4) can be rewritten as

$$S_{\min}(\Phi) \geq \log(k) - k \cdot \max_{M \in \mathcal{W},\, \|M\|_{HS}=1} \left\| MM^\dagger - \frac{\mathrm{Id}}{k} \right\|_{HS}^2 . \tag{6}$$

The idea will be to show that, for a random subspace $\mathcal{W}$, the maximum on the right is very small; this will be formalized in the next proposition.

6. The Main Proposition and the Derivation of the Main Theorem

The heart of the argument is the following proposition

Proposition 3. *There are absolute constants $c, C, C' > 0$ so that for every k, for $d = Ck^2$ and $m = cd$, a random Haar-distributed subspace $\mathcal{W}$ of dimension m in $\mathcal{M}_{k,d}$ satisfies*

$$\max_{M \in \mathcal{W},\, \|M\|_{HS}=1} \left\| MM^\dagger - \frac{\mathrm{Id}}{k} \right\|_{HS} \leq \frac{C'}{k} \tag{7}$$

with large probability (tending to 1 when k tends to ∞).

From the proposition one quickly deduces that the pair $(\Phi, \bar{\Phi})$ is a counterexample to the additivity of minimum output von Neumann entropy. Indeed, a straightforward calculation shows that applying $\Phi \otimes \bar{\Phi}$ to the maximally entangled state yields an output state with one eigenvalue greater than or equal to $\frac{\dim \mathcal{W}}{\dim \mathcal{M}_{k,d}} = \frac{m}{kd} = \frac{c}{k}$ ([8], Lemma III.3; see also Sect. 6 in [12]). Then, a simple argument using just concavity of $S(\cdot)$ reduces

the problem to calculating the entropy of the state with one eigenvalue *equal* to $\frac{c}{k}$ and all the remaining ones identical, which yields

$$S_{\min}(\Phi \otimes \bar{\Phi}) \leq 2\log k - \frac{c\log k}{k} + \frac{1}{k}.$$

On the other hand, Eq. (6) together with Proposition 3 implies

$$S_{\min}(\Phi) \geq \log(k) - \frac{C'^2}{k}.$$

Since $S_{\min}(\bar{\Phi}) = S_{\min}(\Phi)$, the inequality of Theorem 1 follows if k is large enough, as required.

7. Dvoretzky's Theorem: Take One

We wish to point out that while Proposition 3 will be *derived from* a Dvoretzky-like theorem for Lipschitz functions (Theorem 4 below), it can be *rephrased* in the language of the standard Dvoretzky's theorem. Indeed, its assertion says that for every $M \in \mathcal{W}$ with $\|M\|_{HS} = 1$ we have

$$\frac{C^2}{k^2} \geq \left\| MM^\dagger - \frac{\mathrm{Id}}{k} \right\|_{HS}^2 = \mathrm{Tr}\,|M|^4 - \frac{2\,\mathrm{Tr}\,MM^\dagger}{k} + \frac{\mathrm{Tr}\,\mathrm{Id}}{k^2} = \mathrm{Tr}\,|M|^4 - \frac{1}{k} \geq 0. \quad (8)$$

Consequently,

$$k^{-1/4}\|M\|_{HS} \leq \|M\|_4 \leq k^{-1/4}\left(1 + \frac{C^2}{k}\right)^{1/4}\|M\|_{HS} \leq k^{-1/4}\left(1 + \frac{C^2}{4k}\right)\|M\|_{HS} \quad (9)$$

for all $M \in \mathcal{W}$. In other words, $\mathcal{W}$ is $(1+\delta)$-Euclidean, with $\delta = \frac{C^2}{4k}$, when considered as a subspace of the normed space $\left(\mathcal{M}_{k,d}, \|\cdot\|_4\right)$, the Schatten 4-class.

In our prior work [9] we similarly observed that the crucial technical step of the Hayden-Winter proof of non-additivity of p-Rényi entropy for $p > 1$ can be restated as an instance of Dvoretzky's theorem for the Schatten $2p$-class. There is an important difference, however. While in the case of p-Rényi entropy the needed Dvoretzky-type statement was known since the 1970s, for the statement of the type (9) needed in the present context, the "off the shelf" methods seem to yield only $\delta = O(k^{-1/4})$ as opposed to $\delta = O(k^{-1})$ above. This also suggests that while for the p-Rényi entropy derandomization of the example — i.e., supplying *explicit* channels for which the additivity fails — may be a feasible project (see Sect. IX in [9] and references therein), the analogous task for the von Neumann entropy is likely to be much harder.

8. Dvoretzky's Theorem: Take Two

We use the following definitions: if f is a function from a metric space (X, d) to $\mathbf{R}$, and $\mu \in \mathbf{R}$, the *oscillation* of f around μ on a subset $A \subset X$ is

$$\mathrm{osc}(f, A, \mu) = \sup_A |f - \mu|.$$

A function f defined on the unit sphere $S_{\mathbf{C}^n}$ is called *circled* if $f(e^{i\theta}x) = f(x)$ for any $x \in S_{\mathbf{C}^n}$, $\theta \in [0, 2\pi]$. If X is a real random variable, we will say that μ is a *central value* of X if μ is either the mean of X, or any number between the 1st and the 3rd quartile of X (i.e., if $\min\{\mathbf{P}(X \geq \mu), \mathbf{P}(X \leq \mu)\} \geq \frac{1}{4}$; this happens in particular if μ is the median of X).

We will need the following variant of Milman's "tangible" version of Dvoretzky's theorem.

Theorem 4 (Dvoretzky's theorem for Lipschitz functions). *If $f : S_{\mathbf{C}^n} \to \mathbf{R}$ is a 1-Lipschitz circled function, then for every $\varepsilon > 0$, if $E \subset \mathbf{C}^n$ is a random subspace (Haar-distributed) of dimension $c_0 n\varepsilon^2$, we have with large probability*

$$\operatorname{osc}(f, S_{\mathbf{C}^n} \cap E, \mu) \leq \varepsilon,$$

where μ is any central value of f (with respect to the normalized Lebesgue measure on $S_{\mathbf{C}^n}$) and c_0 is an absolute constant. If the function is L-Lipschitz, the dimension changes to $c_0 n(\varepsilon/L)^2$.

A striking application of the theorem above is to the case when f is the gauge function of a convex body, or a norm: it leads to the fact that any high-dimensional convex body has almost spherical sections.

At the heart of Dvoretzky-like phenomena lies the concentration of measure, which in our framework is expressed by

Lemma 5 (Lévy's lemma [13]). *If $f : S^{n-1} \to \mathbf{R}$ is a 1-Lipschitz function, then for every $\varepsilon > 0$,*

$$\mathbf{P}(|f(x) - \mu| > \varepsilon) \leq C_1 \exp(-c_1 n\varepsilon^2),$$

where x is uniformly distributed on S^{n-1}, μ is any central value of f, and $C_1, c_1 > 0$ are absolute constants.

Results such as Theorem 4 or Lévy's lemma are usually stated with μ equal to the median or the mean of f. However, once we know that the result is true for *some* central value (or, for that matter, for *any* $\mu \in \mathbf{R}$), it holds *a posteriori* for *any* such value (up to changes in the constants) as, for 1-Lipschitz functions, all central values differ at most by $C/\sqrt{n}$.

The obvious idea to prove Theorem 4 is to use Lévy's lemma and an ε-net argument — using the fact that an ε-net in $S_{\mathbf{C}^n} = S^{2n-1}$ can be chosen to have cardinality $\leq (1+2/\varepsilon)^{2n}$ (see [14], Lemma 4.10). Indeed, this was essentially Milman's original argument in [10]. However, one only obtains this way a subspace E of dimension $cn\varepsilon^2/\log(1/\varepsilon)$. For many applications (including our previous paper [9]), this extra logarithmic factor is not an issue. However, in the present case, having the optimal dependence on ε is crucial.

The classical framework of convex geometry is the real case (with or without the assumption "circled," which in that context just means then that the function is even). In that setting, Theorem 4 was proved by Gordon [15] who used comparison inequalities for Gaussian processes. A proof based on concentration of measure was later given by Schechtman [16]. The complex case does not seem to appear in the literature. Actually, at the face of it, Gordon's proof does not extend to the complex setting, while Schechtman's proof does. We sketch Schechtman's proof of Theorem 4 in Appendix A. It is not clear whether the assumption "f circled" in Theorem 4 can be completely removed; we do know that it is needed at most for very small values of ε.

9. Proof of the Main Proposition

Let S_{HS} be the Hilbert–Schmidt sphere in $\mathcal{M}_{k,d}$ and let M be a random matrix uniformly distributed on S_{HS}. Let $\tilde{g}(\cdot)$ be the function defined on S_{HS} by

$$\tilde{g}(M) = \left\| MM^\dagger - \frac{\mathrm{Id}}{k} \right\|_{HS}.$$

The next well-known lemma asserts that the singular values of a very rectangular random matrix are very concentrated. This is a familiar phenomenon in random matrix theory that goes back to [17]. Versions of this lemma appeared in the QIT literature under the tensor formalism (see for example Lemma III.4 in [18]). However, these versions typically introduce an unnecessary logarithmic factor which would imply that the main proposition holds with $d = Ck^2 \log k$ instead of $d = Ck^2$. For completeness, we include a proof of Lemma 6 in Appendix B.

Lemma 6. *There exist absolute constants $C, c > 0$ such that, if M is uniformly distributed on the Hilbert–Schmidt sphere in $\mathcal{M}_{k,d}$ ($d \geq C^2 k$), then with probability larger than $1 - \exp(-ck)$,*

$$\mathrm{spec}(MM^\dagger) \subset \left[\left(\frac{1}{\sqrt{k}} - \frac{C}{\sqrt{d}} \right)^2, \left(\frac{1}{\sqrt{k}} + \frac{C}{\sqrt{d}} \right)^2 \right]. \tag{10}$$

We note that inclusion (10) can be reformulated as follows: all singular values of M differ from $1/\sqrt{k}$ by less than $C/\sqrt{d}$. (Recall that the singular values of M correspond to the Schmidt coefficients of a random pure state in $\mathbf{C}^k \otimes \mathbf{C}^d$.)

We will use in the sequel the following immediate corollary of Lemma 6.

Corollary 7. *Under the hypotheses of Lemma 6 and denoting $C_0 = 3C$*

(a) *with probability larger than $1 - \exp(-ck)$, all eigenvalues of $MM^\dagger$ differ from $1/k$ by less than $C_0/\sqrt{kd}$; consequently, the median (or any fixed quantile) of $\tilde{g}$ is bounded by $C_0/\sqrt{d}$ for k large enough.*

(b) *if $d \geq C^2 k$, the median (or any fixed quantile) of $\|M\|_\infty$ is bounded by $2/\sqrt{k}$ for k large enough.*

We point out that while we chose to present statements (a) and (b) above as consequences of Lemma 6 for clarity and for "cultural" reasons (the lemma being familiar to the QIT community), more precise versions of these statements are available in (or can be readily deduced from) the random matrix literature. Re (a), the study of the distribution of $\tilde{g}$ is, by (8), equivalent to that of $\mathrm{Tr}\,|M|^4$, and a closed formula for the expected value of the latter is known (up to terms of smaller order, its value is $1/k + 1/d$); see, e.g., [19] (Sect. 8) and its references. Re (b), sharp estimates on the tail of $\|M\|_\infty$ can also be found in [19] (proof of Lemma 7.3), in particular every fixed quantile is $1/\sqrt{k} + 1/\sqrt{d}$ up to terms of smaller order. This result can also be retrieved via methods of earlier papers [20,21], which focused on the real case.

The function $\tilde{g}$ is 2-Lipschitz on S_{HS}, and Corollary 7(a) implies that the median of $\tilde{g}$ is as small as we want for large d. However, a direct application of Theorem 4 yields only a bound of order $1/\sqrt{k}$ in (7). The trick — already present in the previous approaches — is to exploit the fact that $\tilde{g}$ has a much smaller Lipschitz constant when

restricted to a certain large subset of S_{HS}. As we will see, this bootstrapping argument is equivalent to applying Theorem 4 *twice*.

The following lemma appears in [6] with a rather long proof, but using the matrix formalism completely demystifies it.

Lemma 8. *The function $\tilde{g}$ is $6/\sqrt{k}$-Lipschitz when restricted to the set*

$$\Omega = \{M \in S_{HS} \text{ s.t } \|M\|_\infty \leq 3/\sqrt{k}\}.$$

Proof. The lemma is a consequence of the following chain of matrix inequalities

$$\begin{aligned}\left\|MM^\dagger - \frac{\mathrm{Id}}{k}\right\|_{HS} - \left\|NN^\dagger - \frac{\mathrm{Id}}{k}\right\|_{HS} &\leq \|MM^\dagger - NN^\dagger\|_{HS} \\ &\leq \|M(M^\dagger - N^\dagger) + (M-N)N^\dagger\|_{HS} \\ &\leq \|M\|_\infty \|M^\dagger - N^\dagger\|_{HS} + \|M-N\|_{HS}\|N^\dagger\|_\infty \\ &\leq (\|M\|_\infty + \|N\|_\infty)\|M-N\|_{HS}.\end{aligned}$$

□

The function $\|\cdot\|_\infty$ is 1-Lipschitz on S_{HS}. By Corollary 7(b), its median is bounded by $2/\sqrt{k}$ for $d \geq C^2 k$. (Note that Lévy's lemma shows that the measure of the complement of Ω is very small.) An application of the standard Dvoretzky's theorem (i.e.,Theorem 4 for norms) to $f = \|\cdot\|_\infty$ with μ equal to the median of $\|\cdot\|_\infty$ and with $\varepsilon = 1/\sqrt{k}$ (note that the dimension of the ambient space is $n = kd$) shows that the intersection of S_{HS} with a random subspace of dimension cd in $\mathcal{M}_{k,d}$ is contained in Ω with large probability.

Let g be a $6k^{-1/2}$-Lipschitz extension of $\tilde{g}_{|\Omega}$ to S_{HS} — in any metric space X, it is possible to extend any L-Lipschitz function $\tilde{h}$ defined on a subset Y without increasing the Lipschitz constant; use, e.g., the formula

$$h(x) = \inf_{y \in Y} \left[\tilde{h}(y) + L\,\mathrm{dist}(x,y)\right].$$

This formula also guarantees that the extended function g is circled. Since $g = \tilde{g}$ on most of S_{HS}, the median of g (resp., $\tilde{g}$) is a central value of $\tilde{g}$ (resp., g). We apply Theorem 4 to g with $\varepsilon = 1/k$ and $L = 6k^{-1/2}$ to get (μ being the median of $\tilde{g}$)

$$\mathrm{osc}(g, S_{HS} \cap E, \mu) \leq 1/k$$

on a random subspace $E \subset \mathcal{M}_{k,d}$ of dimension $m = c_0 \cdot kd \cdot (k^{-1}/(6k^{-1/2}))^2 = cd$. Using Corollary 7(a), we obtain that $\mu \leq 1/k$ for $d \geq (C_0 k)^2$. We then have

$$\mathrm{osc}(g, S_{HS} \cap E, 0) \leq 2/k.$$

If $S_{HS} \cap E \subset \Omega$ (which, as noticed before, holds with large probability), g and $\tilde{g}$ coincide on $S_{HS} \cap E$ and therefore $\mathrm{osc}(\tilde{g}, S_{HS} \cap E, 0) \leq 2/k$. This completes the proof of Proposition 3 and hence that of Theorem 1.

Acknowledgements. The research of the first named author was partially supported by the *Agence Nationale de la Recherche* grant ANR-08-BLAN-0311-03. The research of the second and third named authors was partially supported by their respective grants from the *National Science Foundation* (U.S.A.) and from the *U.S.-Israel Binational Science Foundation*. The authors would like to thank M. B. Hastings and M. Horodecki for valuable comments, and *MF Oberwolfach* – where insights crucial to this project were crystallized – for their hospitality.

Appendix A. Proof of Theorem 4 (après Schechtman)

We sketch here a proof of Theorem 4, essentially following Schechtman [16]. As we already mentioned, a simple use of a ε-net argument gives a parasitic factor $\log(1/\varepsilon)$. This can be improved by a *chaining* argument, which goes back (at least) to Kolmogorov — a way to use η-nets for all values of η simultaneously.

Consider the canonical inclusion $\mathbf{C}^m \subset \mathbf{C}^n$, and let $U \in \mathcal{U}(n)$ be a random Haar-distributed unitary matrix. Then $F := U(\mathbf{C}^m)$ is distributed according to the Haar measure on the Grassmann manifold of m-dimensional subspaces. If $f : S_{\mathbf{C}^n} \to \mathbf{R}$ is a 1-Lipschitz circled function with mean μ, we need to show that $\operatorname{osc}(f \circ U, S_{\mathbf{C}^m}, \mu) \leq \varepsilon$ with large probability provided $m \leq c_0 n\varepsilon^2$. We first prove a lemma.

Lemma 9. *Let $f : S_{\mathbf{C}^n} \to \mathbf{R}$ be a 1-Lipschitz circled function and $U \in \mathcal{U}(n)$ be a Haar-distributed random unitary matrix. Then for any $x, y \in S_{\mathbf{C}^n}$ with $x \neq y$ and for any $\lambda > 0$,*

$$\mathbf{P}(|f(Ux) - f(Uy)| > \lambda) \leq C \exp\left(-cn\frac{\lambda^2}{|x-y|^2}\right).$$

Proof. Fix $x, y \in S_{\mathbf{C}^n}$. Since f is circled (and U is $\mathbf{C}$-linear), we may replace y by $e^{i\theta}y$ and choose θ so that $\langle x|y\rangle$ is real nonnegative; note that this choice of θ minimizes $|x-y|$ and assures that $x+y$ and $y-x$ are orthogonal. (This is the only *really* new point needed to accommodate the complex setting.) Set $z = \frac{x+y}{2}$ and $w = \frac{y-x}{2}$, then $x = z+w$ and $y = z-w$. Further, set $\beta = |w| = \frac{1}{2}|x-y|$ (we may assume that $\beta \neq 0$) and $w' = \beta^{-1}w$. Then, conditionally on $u = U(z)$, $U(w')$ is distributed uniformly on the sphere $S_{u^\perp} := S_{\mathbf{C}^n} \cap u^\perp$. Since $U(x) = u + \beta U(w')$ and $U(y) = u - \beta U(w')$, it follows that the conditional (on $u = U(z)$) distribution of $f(Ux) - f(Uy)$ is the same as that of $f_u : S_{u^\perp} \to \mathbf{R}$ defined by

$$f_u(v) = f(u + \beta v) - f(u - \beta v).$$

As is readily seen, f_u is 2β-Lipschitz and its mean is 0. From Lévy's lemma, applied to f_u and to the $(2n-3)$-dimensional sphere $S_{u^\perp}$, we deduce that, conditionally on $u = U(z)$,

$$\mathbf{P}(|f(Ux) - f(Uy)| > \lambda) \leq C_1 \exp(-c_1(2n-2)\lambda^2/|x-y|^2),$$

and hence the same inequality holds also without the conditioning. □

The end of the proof (the actual chaining argument) is identical to that in Schechtman's paper, so — rather than copying it — we present the general principle on which it is based. Let (S, ρ) be a compact metric space and let $(X_s)_{s\in S}$ be a family of mean 0 random variables (a stochastic process indexed by S). We say that (X_s) is *subgaussian* if there are $A, \alpha > 0$ such that, for all $s, t \in S$ with $s \neq t$ and for all $\lambda \geq 0$,

$$\mathbf{P}(|X_s - X_t| \geq \lambda) \leq A \exp\left(-\alpha\frac{\lambda^2}{\rho(s,t)^2}\right). \tag{11}$$

Proposition 10 (Dudley's inequality). *If $(X_s)_{s\in S}$ satisfies (11) and some mild regularity conditions, then*

$$\mathbf{E}\sup_{s,t\in S}|X_s - X_t| \leq C' A\alpha^{-1/2}\int_0^\infty \sqrt{\log N(S,\eta)}\,d\eta,$$

where $N(S,\eta)$ is the minimal cardinality of a η-net of S (in particular the integrand is 0 if η is larger than the radius of S).

See [22] for the original article, [23] for a generalization to the subgaussian case that is relevant here, and [24] for a book exposition; we also sketch a proof further below for the reader's convenience.

In our case we choose $S = S_{\mathbf{C}^m} \cup \{0\}$ (with the usual Euclidean metric), $X_s = f(Us) - \mu$ if $s \in S_{\mathbf{C}^m}$ and $X_0 = 0$; then

$$\operatorname{osc}(f \circ U, S_{\mathbf{C}^m}, \mu) = \sup_{s\in S} |X_s|.$$

The underlying probability space is $\mathcal{U}(n)$, and the subgaussian property is given by Lemma 9 if $s, t \in S_{\mathbf{C}^m}$ and directly by Lévy's lemma if s or t equals 0. Next, the bound $N(S_{\mathbf{C}^m}, \eta) = N(S^{2m-1}, \eta) \leq (1+2/\eta)^{2m}$ mentioned in the comments following Lemma 5 leads to an estimate $2\sqrt{m}$ for the integral and to the bound

$$E := \mathbf{E} \sup_{s\in S} |X_s| \leq \mathbf{E} \sup_{s,t\in S} |X_s - X_t| \leq C'C(cn)^{-1/2} \cdot 2\sqrt{m} = C''\sqrt{\frac{m}{n}}.$$

(For readers confused by different quantities appearing on the left side in different forms of Dudley's inequality, we point out that the first inequality above uses the fact that one of the variables X_t equals 0, and that we always have $\sup_{s,t} |X_s - X_t| = \sup_s X_s + \sup_t (-X_t)$.) The assertion of Theorem 4 follows now from Markov's inequality if ε is sufficiently larger than E, which is assured by choosing c_0 small enough. A slightly more careful argument (such as that given in [16], or see [24]) or an application of the appropriate concentration inequality (for functions on $\mathcal{U}(n)$) yields a bound of the form $\exp(-c'\varepsilon^2 n)$ on the probability of the exceptional set $\sup_{s\in S} |X_s| > C''\sqrt{\frac{m}{n}} + \varepsilon$ (hence for the exceptional set from Theorem 4).

Let us comment here that the value of the constant c_0 given by the proof of Theorem 4 is probably the single most important obstacle to showing Theorem 1 for "reasonable" values of k, m. An adaptation of the proof from [15] (which yields good constants) to the complex case could be helpful here.

Proof of Dudley's inequality. For every $k \in \mathbf{Z}$, let $\mathcal{N}_k$ be a 2^{-k}-net of minimal cardinality for (S, ρ). Let $k_0 \in \mathbf{Z}$ be such that the radius of S lies between $2^{-(k_0+1)}$ and 2^{-k_0}; the net $\mathcal{N}_{k_0}$ consists of a single element s_0. For every $s \in S$ and $k \in \mathbf{Z}$, let $\pi_k(s)$ be an element of $\mathcal{N}_k$ satisfying $\rho(s, \pi_k(s)) \leq 2^{-k}$. The *chaining equation* reads for every $s \in S$,

$$X_s = X_{s_0} + \sum_{k\geq k_0} X_{\pi_{k+1}(s)} - X_{\pi_k(s)}. \tag{12}$$

(It is here where some regularity of (X_s) – path continuity – is used.) It follows that

$$\sup_{s,t\in S} |X_s - X_t| \leq 2 \sum_{k\geq k_0} \sup_{s\in S} |X_{\pi_{k+1}(s)} - X_{\pi_k(s)}| \leq 2 \sum_{k\geq k_0} \sup_{u,u'} |X_u - X_{u'}|, \tag{13}$$

where the last supremum is taken over couples $(u, u') \in \mathcal{N}_{k+1} \times \mathcal{N}_k$ satisfying $\rho(u, u') \leq 2^{-k} + 2^{-(k+1)} < 2^{-k+1}$. It remains to bound the expectation of each term in the sum, using the following fact

Fact 11. *If $N \geq 2$ and $Y_1, \ldots, Y_N$ are nonnegative random variables satisfying the tail estimate $\mathbf{P}(Y_i \geq t) \leq A \exp(-t^2/2\beta^2)$ for all $t \geq 0$, then*

$$\mathbf{E} \max Y_i \leq C A \beta \sqrt{\log N}.$$

To bound $\mathbf{E} \sup |X_u - X_{u'}|$, we apply the above fact with $\beta = 2^{-k+1}\alpha^{-1/2}$ and $N = \mathrm{card}(\mathscr{N}_k) \cdot \mathrm{card}(\mathscr{N}_{k+1}) \leq N(S, 2^{-(k+1)})^2$. This gives

$$\mathbf{E} \sup_{s,t \in S} |X_s - X_t| \leq C' A \alpha^{-1/2} \sum_{k \geq k_0} 2^{-k} \sqrt{\log N(S, 2^{-(k+1)})}.$$

The result now follows by relating the last series to the integral in Proposition 10 (a version of the integral test from calculus). □

Proof of Fact 11. We may assume $\beta = 1$ by working with Y_i/β. Then simply write

$$\begin{aligned} \mathbf{E} \max Y_i &= \int_0^\infty \mathbf{P}(\max Y_i \geq t) dt \\ &\leq \sqrt{2 \log N} + AN \int_{\sqrt{2 \log N}}^\infty \exp(-t^2/2) dt \leq \sqrt{2 \log N} + A. \end{aligned}$$

The last inequality follows from $\int_{\sqrt{2 \log N}}^\infty \exp(-t^2/2) dt \leq \int_{\sqrt{2 \log N}}^\infty t \exp(-t^2/2) dt = 1/N$. Note that the hypotheses force $A \geq 1$. □

Appendix B. Proof of Lemma 6

The lemma will follow if we show that with large probability,

$$\|\Delta\|_\infty \leq \frac{C}{\sqrt{kd}},$$

where $\Delta = MM^\dagger - \mathrm{Id}/k \in \mathcal{M}_k$ and $\|\cdot\|_\infty$ is the operator (or spectral) norm. Let $\mathscr{N}$ be a $\frac{1}{4}$-net of $S_{\mathbf{C}^k}$ with cardinality bounded by $(C_0)^k$. One checks that if $x \in S_{\mathbf{C}^k}$ and $\bar{x} \in \mathscr{N}$ satisfy $|x - \bar{x}| \leq 1/4$, then

$$|\langle x|\Delta|x\rangle| \leq |\langle \bar{x}|\Delta|\bar{x}\rangle| + |\langle x - \bar{x}|\Delta|\bar{x}\rangle| + |\langle x|\Delta|x - \bar{x}\rangle| \leq |\langle \bar{x}|\Delta|\bar{x}\rangle| + 2 \cdot \frac{1}{4}\|\Delta\|_\infty,$$

so that taking supremum over $x \in S_{\mathbf{C}^k}$, we get

$$\|\Delta\|_\infty \leq 2 \sup_{\bar{x} \in \mathscr{N}} |\langle \bar{x}|\Delta|\bar{x}\rangle|.$$

An application of the union bound gives

$$\begin{aligned} \mathbf{P}\left(\|\Delta\|_\infty \geq \frac{C}{\sqrt{kd}}\right) &\leq (C_0)^k \cdot \mathbf{P}\left(\langle x_0|\Delta|x_0\rangle \geq \frac{C}{2\sqrt{kd}}\right) \\ &= (C_0)^k \cdot \mathbf{P}\left(|M^\dagger x_0|^2 \geq \frac{1}{k} + \frac{C}{2\sqrt{kd}}\right) \\ &\leq (C_0)^k \cdot \mathbf{P}\left(|M^\dagger x_0| \geq \frac{1}{\sqrt{k}} + \frac{C}{5\sqrt{d}}\right), \end{aligned}$$

where $x_0 \in \mathbf{C}^k$ is any fixed unit vector (remember that $d \geq C^2 k$). The probabilities above can be expressed in terms of Beta-type integrals, but it's easier to estimate them using Lévy's lemma. The function $M \mapsto |M^\dagger x_0|$ is 1-Lipschitz on the Hilbert–Schmidt sphere (if x_0 is the first vector of the canonical basis, then $M^\dagger x_0$ is essentially the first row of M) and

$$\mathbf{E}\,|M^\dagger x_0| \leq \left(\mathbf{E}\,|M^\dagger x_0|^2\right)^{1/2} = \sqrt{1/k}.$$

Hence, by Lévy's lemma (with $n = 2kd$ and $\varepsilon = \frac{C}{5\sqrt{d}}$), we get

$$\mathbf{P}\left(\|\Delta\|_\infty \geq \frac{C}{\sqrt{kd}}\right) \leq \exp(-ck)$$

for some choice of the constants $C, c > 0$, as required.

References

1. Nielsen, M. A., Chuang, I. L.: *Quantum computation and quantum information*. Cambridge: Cambridge University Press (2000)
2. Holevo, A.S.: *The additivity problem in quantum information theory*. In: "Proceedings of the International Congress of Mathematicians (Madrid, 2006)," Vol. III, Zürich: Eur. Math. Soc., 2006, pp. 999–1018
3. Shor, P.W.: Equivalence of additivity questions in quantum information theory. Commun. Math. Phys. **246**(3), 453–472 (2004)
4. Hastings, M.B.: Superadditivity of communication capacity using entangled inputs. Nature Phys. **5**, 255 (2009)
5. Dvoretzky, A.: *Some Results on Convex Bodies and Banach Spaces.* In: "Proc. Internat. Sympos. Linear Spaces (Jerusalem, 1960)," Jerusalem: Jerusalem Academic Press, Oxford: Pergamon, 1961, pp. 123–160
6. Brandao, F., Horodecki, M.: On Hastings' counterexamples to the minimum output entropy additivity conjecture. Open Syst. Inf. Dyn. **17**, 31 (2010)
7. Fukuda, M., King, C., Moser, D.: Comments on Hastings' Additivity Counterexamples. Commun. Math. Phys. **296**, 111 (2010)
8. Hayden, P., Winter, A.: Counterexamples to the maximal p-norm multiplicativity conjecture for all $p > 1$. Commun. Math. Phys. **284**, 263–280 (2008)
9. Aubrun, G., Szarek, S., Werner, E.: Non-additivity of Rényi entropy and Dvoretzky's theorem. J. Math. Phys. **51**, 022102 (2010)
10. Milman, V.: A new proof of the theorem of A. Dvoretzky on sections of convex bodies. Funct. Anal. Appl. **5**, 28–37 (1971) (English translation)
11. Figiel, T., Lindenstrauss, J., Milman, V.D.: The dimension of almost spherical sections of convex bodies. Acta Math. **139**(1-2), 53–94 (1977)
12. Collins, B., Nechita, I.: Gaussianization and eigenvalue statistics for random quantum channels (III), Ann. Appl. Probab., to appear; http://arxiv.org/abs/0910.1768v2 [quant-ph], 2009
13. Lévy, P.: *Problémes concrets d'analyse fonctionnelle*, 2nd ed. Paris: Gauthier-Villars, 1951
14. Pisier, G.: *The volume of convex bodies and Banach space geometry.* Cambridge Tracts in Mathematics, **94**. Cambridge: Cambridge University Press, 1989
15. Gordon, Y.: *On Milman's inequality and random subspaces which escape through a mesh in* $\mathbf{R}^n$. In: "Geometric aspects of functional analysis (1986/87)," Lecture Notes in Math., **1317,** Berlin: Springer, 1988, pp. 84–106
16. Schechtman, G.: *A remark concerning the dependence on* ε *in Dvoretzky's theorem*. In: "Geometric aspects of functional analysis (1987–88)," Lecture Notes in Math., **1376,** Berlin: Springer, 1989, pp. 274–277
17. Marchenko, V.A., Pastur, L.A.: The distribution of eigenvalues in certain sets of random matrices. Mat. Sb. **72**, 507–536 (1967)
18. Hayden, P., Leung, D., Winter, A.: Aspects of generic entanglement. Commun. Math. Phys. **265**, 95–117 (2006)
19. Haagerup, U., Thorbjørnsen, S.: Random matrices with complex Gaussian entries. Expos. Math. **21**, 293–337 (2003)
20. Geman, S.: A limit theorem for the norm of random matrices. Ann. Probab. **8**, 252–261 (1980)

21. Silverstein, J.W.: The smallest eigenvalue of a large-dimensional Wishart matrix. Ann. Probab **13**, 1364–1368 (1985)
22. Dudley, R.M.: The sizes of compact subsets of Hilbert space and continuity of Gaussian processes. J. Funct. Anal. **1**, 290–330 (1967)
23. Jain, N. C., Marcus, M. B.: *Continuity of subgaussian processes.* In: "Probability on Banach Spaces," Advances in Probability, Vol. **4**, New York: Dekker, 1978, pp. 81–196
24. Talagrand, M.: *The generic chaining. Upper and Lower bounds of Stochastic Processes.* Berlin-Heidelberg-New York: Springer, 2005

Communicated by M.B. Ruskai