



A remainder term for Hölder's inequality for matrices and quantum entropy inequalities

ERIC A. CARLEN 

Abstract. We prove a sharp remainder term for Hölder's inequality for traces as a consequence of the uniform convexity properties of the Schatten trace norms. We then show how this implies a novel family of Pinsker type bounds for the quantum Rényi entropy. Finally, we show how the sharp form of the usual quantum Pinsker inequality for relative entropy may be obtained as a fairly direct consequence of uniform convexity.

Mathematics Subject Classification. 26B25, 94A17.

Keywords. Density matrix, Entropy, Uniform convexity.

1. Introduction. For any complex $n \times n$ matrix A , define $|A| = (A^*A)^{1/2}$, and for $1 \leq p < \infty$, $\|A\|_p = (\text{Tr}|A|^p)^{1/p}$. If $\sigma_1 \geq \dots \geq \sigma_n$ are the singular values of A , then $\|A\|_p = \left(\sum_{j=1}^n \sigma_j^p\right)^{1/p}$. For $p = \infty$, $\|A\|_\infty$ is simply the operator norm of A , which is also the largest singular value of A . It is well-known that $\|\cdot\|_p$ is a norm, the Schatten p norm, on M_n , the space of $n \times n$ matrices. The space C_p is the space M_n of $n \times n$ complex matrices equipped with this norm.

The Schatten norms are in many ways close analogs of the ℓ_p norms. In particular, one has the analog of Hölder's inequality

$$|\text{Tr}[AB]| \leq \|A\|_p \|B\|_{p'}$$

where $1/p + 1/p' = 1$, $1 \leq p \leq \infty$. Whenever p and p' appear together below, it is assumed that $1/p + 1/p' = 1$. For all $1 \leq p \leq \infty$, a simple argument using the singular value decomposition shows that

$$\|A\|_p = \sup \{ \Re(\text{Tr}[AB]) : \|B\|_{p'} = 1 \}, \quad (1.1)$$

and in fact, the supremum is achieved. The Minkowski inequality; i.e., the fact that our norms are norms, follows in the usual way. For $1 < p < \infty$, and non-zero $A \in C_p$, define

$$\mathcal{D}_p(A) = \|A\|_p^{1-p} |A|^{p-1} U^* \tag{1.2}$$

where $A = U|A|$ is the polar decomposition of A . Then one readily checks that

$$\|\mathcal{D}_p(A)\|_{p'} = 1 \quad \text{and} \quad \text{Tr}[\mathcal{D}_p(A)A] = \|A\|_p. \tag{1.3}$$

Thus, for $1 < p < \infty$, the supremum in (1.1) is a maximum, and the maximum is attained at $\mathcal{D}_p(A)$. It then follows from (1.1) that for all $A, B \in M_n$ and all $t \in \mathbb{R}$,

$$\|A + tB\|_p \geq \text{Tr}[\mathcal{D}_p(A)(A + tB)] = \|A\|_p + t \text{Tr}[\mathcal{D}_p(A)B].$$

Likewise, writing $A = (A + tB) - tB$,

$$\|A\|_p \geq \text{Tr}[\mathcal{D}_p(A + tB)(A)] = \|A + tB\|_p - t \text{Tr}[\mathcal{D}_p(A + tB)B].$$

Thus, provided that $t\|\mathcal{D}_p(A + tB) - \mathcal{D}_p(A)\|_p = o(|t|)$, we have that

$$\left| \|A + tB\|_p - \|A\|_p - t \text{Tr}[\mathcal{D}_p(A)B] \right| = o(|t|),$$

and this says that the norm function $A \mapsto \|A\|_p$ is (Fréchet) differentiable for $1 < p < \infty$, and that $\mathcal{D}_p(A)$ is the derivative at $A \in M_n$. In fact, for $1 < p < \infty$, the map $A \mapsto \mathcal{D}_p(A)$ is Hölder continuous, and the modulus of continuity has been given in [4]. Thus, $A \mapsto \mathcal{D}_p(A)$ is the gradient of the norm function $A \mapsto \|A\|_p$ for $1 < p < \infty$, and this is the reason for the notation using \mathcal{D} .

The map $A \mapsto \mathcal{D}_p(A)$ is closely related to the *non-commutative Mazur map* studied in [1] and [9]. For $1 \leq p, q \leq \infty$, the *Mazur map* $\mathcal{M}_{p,q}$ is defined on M_n by $\mathcal{M}_{p,q}(A) = A|A|^{(p-q)/q}$. For $q = p'$, $(p - q)/q = p - 2$, and hence

$$\mathcal{M}_{p,p'}(A) = \|A\|_p^{p-1} (\mathcal{D}_p(A))^*.$$

Sharp Hölder continuity bounds on $\mathcal{M}_{p,q}$ in a very general von Neumann algebra setting are proved in [9], which can be consulted for further references.

The norm gradient maps, which are the Mazur maps for $q = p'$, normalized to be homogeneous of degree zero, are the focus of this note which concerns another setting in which they arise. Our first result is a quantitative remainder term for the tracial Hölder inequality. From this we shall deduce several quantum entropy inequalities.

The next theorem is a non-commutative analog of a theorem proved in [3] in the commutative context of L_p spaces for Lebesgue integration. The proof simply uses the sharp uniform convexity properties of the C_p norms proved in [2] in place of the corresponding sharp uniform convexity properties of the C_p norms that were used in [3].

1.1 Theorem. (Hölder’s inequality with remainder) *Let $1 < p \leq 2$. Let A be a unit vector in C_p , and let B be a unit vector in $C_{p'}$. Let $\theta \in [0, 2\pi)$ be chosen such that $e^{i\theta}\text{Tr}[AB]$ is non-negative. Then we have both*

$$|\text{Tr}[AB]| \leq 1 - \frac{p-1}{4} \|\mathcal{D}_{p'}(B) - e^{i\theta}A\|_p^2, \tag{1.4}$$

and

$$|\mathrm{Tr}[AB]| \leq 1 - \frac{1}{p' 2^{p'-1}} \|e^{i\theta} B - \mathcal{D}_p(A)\|_{p'}^{p'}. \tag{1.5}$$

The exponents 2 and p' on the right sides of (1.4) and (1.5) are best possible.

Proof. By (1.2) and the choice of θ , $1 + e^{i\theta}\mathrm{Tr}[AB] = \mathrm{Tr}[(\mathcal{D}_{p'}(B) + e^{i\theta}A)B]$. Therefore, by Hölder’s inequality and the choice of θ ,

$$1 + |\mathrm{Tr}[AB]| \leq \|\mathcal{D}_{p'}(B) + e^{i\theta}A\|_p \leq 2 \left\| \frac{\mathcal{D}_{p'}(B) + e^{i\theta}A}{2} \right\|_p. \tag{1.6}$$

Now apply the optimal 2-uniform convexity inequality [2], valid for $1 < p \leq 2$, and unit vectors $X, Y \in C_p$:

$$\left\| \frac{X + Y}{2} \right\|_p \leq 1 - \frac{p-1}{2} \left\| \frac{X - Y}{2} \right\|_p^2. \tag{1.7}$$

This leads directly to (1.4). The proof of (1.5) is similar except that one uses

$$\left\| \frac{X + Y}{2} \right\|_p \leq 1 - \frac{1}{p} \left\| \frac{X - Y}{2} \right\|_p^p$$

valid for $2 \leq p$, and unit vectors $X, Y \in C_p$ [2]. The fact that the exponents are the best possible follows from the fact that this is true in the commutative case, and the proof of this may be found in [3, Theorem 3.1]. \square

2. Application to entropy. Recall that for $\alpha \in (0, 1)$, the Rényi α -relative entropy for ρ with respect to σ is the quantity

$$D_\alpha(\rho||\sigma) = \frac{1}{\alpha - 1} \log (\mathrm{Tr}[\rho^\alpha \sigma^{1-\alpha}]). \tag{2.1}$$

Recall also that

$$\lim_{\alpha \rightarrow 1} D_\alpha(\rho||\sigma) = D(\rho||\sigma) := \mathrm{Tr}[\rho(\log \rho - \log \sigma)],$$

the von Neumann relative entropy. Pinsker’s inequality for the von Neumann relative entropy states that

$$D(\rho||\sigma) \geq \frac{1}{2} \|\rho - \sigma\|_1^2. \tag{2.2}$$

We now show that Theorem 1.1 gives a Pinsker type inequality for the Rényi entropy from which (2.2) can be derived in the limit $\alpha \rightarrow 1$. By the definition (2.1), for $\alpha \in (0, 1)$, an upper bound on $\mathrm{Tr}[\rho^\alpha \sigma^{1-\alpha}]$ implies a lower bound on $D_\alpha(\rho||\sigma)$.

2.1 Theorem. *Let ρ and σ be density matrices in M_n for some n , and let $1 < p \leq 2$. Then*

$$\mathrm{Tr}[\sigma^{1-1/p} \rho^{1/p}] \leq 1 - \frac{p-1}{4} \|\rho^{1/p} - \sigma^{1/p}\|_p^2, \tag{2.3}$$

and

$$\mathrm{Tr}[\sigma^{1-1/p} \rho^{1/p}] \leq 1 - \frac{1}{p' 2^{p'-1}} \|\rho^{1/p'} - \sigma^{1/p'}\|_{p'}^{p'}. \tag{2.4}$$

Proof. Define $A = \rho^{1/p}$ and $B = \sigma^{1/p'}$ so that A and B are unit vectors in C_p and $C_{p'}$ respectively. First note that $\mathcal{D}_{p'}(B) = B^{1/(p-1)} = \sigma^{1/p}$. Hence (2.3) follows directly from (1.4). Next, note that $\mathcal{D}_p(A) = A^{p-1} = \rho^{1-1/p}$. Hence (2.4) follows directly from (1.5). \square

2.2 Corollary. For all $\alpha \in [1/2, 1)$,

$$D_\alpha(\rho||\sigma) \geq \frac{1}{4\alpha} \|\rho^\alpha - \sigma^\alpha\|_{1/\alpha}^2. \tag{2.5}$$

Proof. Take $p := 1/\alpha$ and $\alpha \in [1/2, 1)$ so that $p \in (1, 2]$. Then (2.3) yields (2.5). \square

We could of course use (1.5) to treat the cases $\alpha \in (0, 1/2)$ in an analogous way; the result would be similar, but the exponent on the right would be $1/\alpha$ in place of 2.

Lower bounds on $D_\alpha(\rho||\sigma)$ in terms of $\|\rho - \sigma\|_1$ are known in the classical case, and easily generalize to the quantum case, but these bounds are weaker than the bounds provided by Corollary 2.2. It is known [6] in the classical case (ρ and σ commuting) that

$$D_\alpha(\rho||\sigma) \geq \frac{\alpha}{2} \|\rho - \sigma\|_1^2. \tag{2.6}$$

As a quite direct consequence of the Lieb concavity theorem [7] which says that $(\rho, \sigma) \mapsto \text{Tr}[\rho^\alpha \sigma^{1-\alpha}]$ is concave for $\alpha \in [0, 1]$, this is also valid in the quantum case. Indeed, let P be the projector onto the range of $(\rho - \sigma)_+$. Let U be any unitary that commutes with P . Then

$$\text{Tr}[(U\rho U^*)^\alpha (U\sigma U^*)^{1-\alpha}] = \text{Tr}[\rho^\alpha \sigma^{1-\alpha}].$$

By a theorem of Uhlmann [11], there is a finite set of such unitaries such that if we define $\hat{\rho}$ and $\hat{\sigma}$ to be the averages of $U\rho U^*$ and $U\sigma U^*$, respectively, over all unitaries U in our set, then $\hat{\rho}$ and $\hat{\sigma}$ both belong to the algebra generated by P , and hence for some $p, q \in [0, 1]$,

$$\hat{\rho} = \frac{p}{\text{Tr}[P]}P + \frac{(1-p)}{\text{Tr}[I-P]}(I-P) \quad \text{and} \quad \hat{\sigma} = \frac{q}{\text{Tr}[P]}P + \frac{(1-q)}{\text{Tr}[I-P]}I - P.$$

Then the Lieb concavity theorem implies that $\text{Tr}[\hat{\rho}^\alpha \hat{\sigma}^{1-\alpha}] \geq \text{Tr}[\rho^\alpha \sigma^{1-\alpha}]$. Hence $D_\alpha(\rho||\sigma) \geq D_\alpha(\hat{\rho}||\hat{\sigma})$. However, since $\hat{\rho}$ and $\hat{\sigma}$ commute, the classical bound (2.6) applies to yield $D_\alpha(\hat{\rho}||\hat{\sigma}) \geq \frac{\alpha}{2} \|\hat{\rho} - \hat{\sigma}\|_1^2$, and one easily sees that $\|\rho - \sigma\|_1 = \|\hat{\rho} - \hat{\sigma}\|_1$. Hence, (2.6) is valid in the quantum setting as well.

We now show that (2.5) improves upon (2.6): the ratio of $\|\rho^\alpha - \sigma^\alpha\|_{1/\alpha}$ to $\|\rho - \sigma\|_1$ can be arbitrarily large, and the ratio of $\|\rho - \sigma\|_1$ to $\|\rho^\alpha - \sigma^\alpha\|_{1/\alpha}$ is bounded above by a finite constant.

For the first of these points, an example suffices. Let $\epsilon \in (0, 1/2)$, and define

$$\rho = \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{bmatrix} \quad \text{and} \quad \sigma = \begin{bmatrix} \frac{1}{2} + \epsilon & 0 \\ 0 & \frac{1}{2} - \epsilon \end{bmatrix}.$$

Then $\|\rho - \sigma\|_1 = 2\epsilon$, while $\|\rho^\alpha - \sigma^\alpha\|_{1/\alpha} = (2\alpha\epsilon)^\alpha + \mathcal{O}(\epsilon^{2\alpha})$. It follows that

$$\frac{\|\rho^\alpha - \sigma^\alpha\|_{1/\alpha}}{\|\rho - \sigma\|_1} = \frac{\alpha^\alpha}{2\epsilon^{1-\alpha}} + \mathcal{O}(\epsilon^{2\alpha-1}) \tag{2.7}$$

This shows that (2.5) can provide a much stronger bound than (2.6).

Lemma 2.3 in [9] says (in particular) that for all positive $A, B \in M_n$, and all $\alpha \in (0, 1)$,

$$\frac{\alpha}{2} \|A - B\|_1 \leq \|A^\alpha - B^\alpha\|_{1/\alpha} \max\{\|A^\alpha\|_{1/\alpha}, \|B^\alpha\|_{1/\alpha}\}. \tag{2.8}$$

Applying this with $A = \rho$ and $B = \sigma$, $(\alpha/3)\|\rho - \sigma\|_1 \leq \|\rho^\alpha - \sigma^\alpha\|_{1/\alpha}$. Combining this with (2.5) yields, for $\alpha \in [1/2, 1)$,

$$D_\alpha(\rho|\sigma) \geq \frac{\alpha}{36} \|\rho - \sigma\|_1^2, \tag{2.9}$$

which is (2.6) apart from a constant that is worse by a factor of 18. Thus apart from the constant, (2.5) implies (2.6). Part of the discrepancy in the constants is due to the constant in (2.8), but part also is due to the fact that we have not yet made optimal use of the uniform convexity bounds.

As we now show, more can be gleaned from the argument that we used to deduce a remainder term for Hölder’s inequality from uniform convexity bounds. We now prove a variant of Theorem 1.1 and show that using this variant, we may obtain the full non-commutative Pinsker inequality; i.e., the $\alpha \rightarrow 1$ limit of (2.6), with the exact constants. This derivation shows that the sharp form of Pinsker’s inequality is actually a fairly direct consequence of the uniform convexity properties of the C_p spaces.

3. Pinsker’s inequality and uniform convexity. Since

$$\lim_{\alpha \uparrow 1} D_\alpha(\rho|\sigma) = D(\rho|\sigma) = \text{Tr}[\rho(\log \rho - \log \sigma)],$$

taking the limit $\alpha \uparrow 1$ in (2.2) yields $D(\rho|\sigma) \geq \frac{1}{4} \|\rho - \sigma\|_1^2$. This is Pinsker’s inequality [8, 10], except that it is not in the sharp form which has a factor of 1/2 in place of the 1/4 on the right, which is what one obtains from (2.6) in the limit $\alpha \uparrow 1$. However, one can recover the sharp form of Pinsker’s inequality from the optimal 2-uniform convexity inequality by going back to the proof of Theorem 1.1 and noting that we gave something up arriving at (1.6) by applying the usual Hölder inequality without taking the remainder into account.

3.1 Definition. Let \mathcal{P} be the set of functions $A(p)$ from $[1, 2]$ into the positive $n \times n$ matrices such that $\lim_{p \rightarrow 1} A(p) = A(1)$ in C_1 and such that $\|A(p)\|_p = 1$ for each $p \in [1, 2]$.

For example, let ρ be any density matrix in M_n . Then $A(p) := \rho^{1/p} \in \mathcal{P}$. Moreover, if $A(p)$ and $B(p)$ belong to \mathcal{P} , then so does $(A(p) + B(p))/\|A(p) + B(p)\|_p$.

3.2 Theorem. (Variant of Hölder’s inequality with remainder) *Let $1 < p \leq 2$. Then for all $A(p)$ and $B(p)$ in \mathcal{P} , and any constant $K < 1/2$,*

$$\text{Tr}[A(p)B^{p-1}(p)] \leq 1 - K(p - 1)\|A(p) - B(p)\|_p^2 + o(p - 1). \tag{3.1}$$

3.3 Corollary. (Pinsker’s inequality for density matrices) *For all density matrices ρ and σ in M_n ,*

$$D(\rho||\sigma) \geq \frac{1}{2} \|\rho - \sigma\|_1^2.$$

Proof. Take $A(p) = \rho^{1/p}$ and $B(p) = \sigma^{1/p}$. By Theorem 3.2, for all $K < 1/2$,

$$\text{Tr}[\sigma^{1-1/p} \rho^{1/p}] \leq 1 - K(p-1) \|\rho^{1/p} - \sigma^{1/p}\|_p^2 + o(p-1).$$

Rearranging terms as above, and taking $p \rightarrow 1$, we obtain

$$D(\rho||\sigma) \geq K \|\rho - \sigma\|_1^2.$$

Since $K < 1/2$ can be arbitrarily close to $1/2$, the inequality is proved. □

Proof of Theorem 3.2. We note that for $K = 1/4$, (3.1) is valid by Theorem 1.1. Next, supposing that (3.1) is valid for some constant K , we show that it is also valid when K is replaced by $(K + 1/2)/2$. Iterating this yields the claimed result.

Therefore, let us make the inductive assumption that (3.1) is valid for some constant K . We have

$$\begin{aligned} 1 + \text{Tr} [A(p)B^{p-1}(p)] &= \text{Tr}[(A(p) + B(p))B^{p-1}(p)] \\ &= \|A(p) + B(p)\|_p \text{Tr}[C(p)B^{p-1}(p)] \end{aligned} \tag{3.2}$$

where

$$C(p) = \|A(p) + B(p)\|_p^{-1} (A(p) + B(p)) \in \mathcal{P}.$$

By hypothesis,

$$\text{Tr}[C(p)B^{p-1}(p)] \leq 1 - K(p-1) \|B(p) - C(p)\|_p^2 + o(p-1),$$

and since $\lim_{p \downarrow 1} \|A(p) + B(p)\|_p = 2$,

$$\|B(p) - C(p)\|_p = \left\| B(p) - \frac{A(p) + B(p)}{2} \right\|_p + o(1) = \frac{1}{2} \|A(p) - B(p)\|_p + o(1).$$

Combining this with the previous bound,

$$\text{Tr}[B^{p-1}(p)C(p)] \leq 1 - (p-1) \frac{K}{4} \|A(p) - B(p)\|_p^2 + o(p-1). \tag{3.3}$$

By the 2-uniform convexity inequality,

$$\|A(p) + B(p)\|_p \leq 2 - \frac{p-1}{4} \|A(p) - B(p)\|_p^2.$$

Using this and (3.3) in (3.2), we obtain

$$\begin{aligned} 1 + \text{Tr} [B^{p-1}(p)A(p)] &\leq 2 \left(1 - \frac{p-1}{8} \|A(p) - B(p)\|_p^2 \right) \\ &\quad \times \left(1 - (p-1) \frac{K}{4} \|A(p) - B(p)\|_p^2 + o(1) \right) \\ &\leq 2 - (p-1) \left(\frac{1}{4} + \frac{K}{2} \right) \|A(p) - B(p)\|_p^2 + o(p-1). \end{aligned}$$

Thus, in (3.1), we may replace K by $(K+1/2)/2$, and the validity is maintained. □

Acknowledgements. These results in this paper were obtained while the author was visiting at the I.M.A. in Minnesota during Spring 2015. The author thanks the anonymous referee for a careful reading of the paper.

References

- [1] A. ALEKSANDROV AND V. PELLER, Functions of operators under perturbations of class S_p , *J. Funct. Anal.* **258** (2010), 3675–3724.
- [2] K. BALL, E. A. CARLEN, AND E. H. LIEB, Sharp uniform convexity and smoothness inequalities for trace norms, *Invent. Math.* **115** (1994), 463–482.
- [3] E. A. CARLEN, R. FRANK, AND E. H. LIEB, Stability estimates for the lowest eigenvalue of a Schrödinger operator, *Geom. Funct. Anal.* **24** (2014), 63–84.
- [4] E. A. CARLEN AND E. H. LIEB, Optimal hypercontractivity for Fermi fields and related noncommutative integration inequalities, *Comm. Math. Phys.* **155** (1993), 27–46.
- [5] E. H. LIEB AND W. THIRRING, Inequalities for the moments of the eigenvalues of the Schrödinger Hamiltonian and their relation to Sobolev inequalities, *Stud. math. Phys.*, *Essays Honor Valentine Bargmann*, (1976), 269–303.
- [6] G. GILARDONI, On Pinskers and Vajdas type inequalities for Csiszár's f -divergences, *IEEE Trans. Inform. Theory* **56** (2010), 5377–5386.
- [7] E. H. LIEB, Convex trace functions and the Wigner-Yanase-Dyson conjecture, *Adv. Math.* **11** (1973), 267–288.
- [8] M. S. PINSKER, *Information and Information Stability of Random Variables and Processes*, Translated and edited by Amiel Feinstein Holden-Day, Inc., San Francisco, Calif.-London-Amsterdam, 1964.
- [9] E. RICARD, Hölder estimates for the noncommutative Mazur maps, *Arch. Math.* **104** (2015), 37–45.
- [10] A. E. RASTEGIN, Bounds of the Pinsker and Fannes types on the Tsallis relative entropy, *Math. Phys. Anal. Geom.* **16** (2013), 213–228.
- [11] A. UHLMANN, Sätze über Dichtematrizen, *Wiss. Z. Karl-Marx Univ. Leipzig Math.-Natur. Reihe* **20** (1971), 633–637.

ERIC A. CARLEN
Department of Mathematics, Hill Center
Rutgers University
110 Frelinghuysen Road
Piscataway NJ 08854-8019
USA
e-mail: carlen@math.rutgers.edu

Received: 2 May 2017