QUANTITATIVE UNIFORM CONVEXITY

Eric A. Carlen

Rutgers University and IMA

B.I.R.S., Feb. 24, 2015

Uniform convexity

Let V be a normed vector space. Minkowski's inequality says that if x and y are unit vectors in V,

$$\left\| \frac{x+y}{2} \right\| \leq 1 \, .$$

If the unit ball has faces, this is all that can be said. However, if the unit ball is strictly convex, then for unit vectors x and y in V with $x \neq y$,

$$\left\| \frac{x+y}{2} \right\| < 1 \, .$$

The unit ball is *uniformly convex* is case for all $\epsilon > 0$, there is a $\delta > 0$ so that for unit vectors x and y in V,

$$\|x - y\| \geq 2\epsilon \quad \Rightarrow \quad \left\|\frac{x+y}{2}\right\| \leq 1 - \delta \; .$$

The *modulus of convexity* of a normed vector space V is the function δ_V defined on $(0, 1]$ by

$$\delta_V(\epsilon) = \inf\left\{1 - \left\|\frac{u+v}{2}\right\| \;:\; \|u\| = \|v\| = 1 \;\text{ and }\; \|u - v\| \geq 2\epsilon \right\} \; .$$

Evidently, V is *uniformly convex* in case $\delta_V(\epsilon) > 0$ for all $\epsilon > 0$.

The notion of uniform convexity was introduced by James Clarkson in 1936. He proved that L^p, for any measure space, is uniformly convex for $1 < p < \infty$, and he also discussed some of the consequences of uniform convexity.

The function $\delta_{L^p}(\epsilon)$ was *exactly computed* by Olaf Hanner in 1955, though he reported that he had been told that this result was presented in unpublished lectures of Arne Beurling at Uppsala in 1945.

The Schatten trace norms

In addition to the usual L^p norms, we will be concerned with the *Schatten trace norms* on the the space M_n of $n \times n$ complex matrices.
For any $n \times n$ matrix A, $n \geq 2$, we define

$$|A| = (A^*A)^{1/2} ,$$

and for $1 \leq p < \infty$,

$$\|A\|_p = (\mathrm{Tr}\,|A|^p)^{1/p} .$$

If $\sigma_1 \geq \cdots \geq \sigma_n$ are the singular values of A, then it is readily checked that

$$\|A\|_p = \left(\sum_{j=1}^{n} \sigma_j^p\right)^{1/p} .$$

Uniform convexity for the Schatten p-norms $1 < p < \infty$ was proved by Nicole Tomczak-Jagermann in 1974 who obtained a sharp result when either p or its dual index p' is an even integer. She then obtained a result for all p by an interpolation method. This gives a worse bound than for L^p when neither p not p' is an even integer.

The Schatten trace norms have much in common with the corresponding L^p norms. For example, one has Hölder's inequality: For $1/p + 1/p' = 1$,

$$\|AB\|_1 \leq \|A\|_p \|B\|_{p'} \ .$$

There are, however, differences, as we shall see.

One might ask, for example: For $4/3 < p < 4$, $p \neq 2$, a range Tomczak-Jagermann covered by interpolation, does non-commutativity really result in a worse modulus of convexity?

This question is still partially open, though a theorem of Ball, C, and Lieb, says “not by anything significant” and moreover, that outside this range; i.e., in all of the other gaps, the answer is: “No, the modulus of convexity for S_p is that same as that of L_p outside this range.

One way to prove convexity is to differentiate. The non-commutativity of matrix multiplication introduces new features in formulas for derivatives of powers, and in fact, the function $A \mapsto A^p$ on the positive $n \times n$ matrices (with its usual order) is *not* convex for $p > 2$.

To calculate with powers of matrices, it is often convenient to use certain integral representation formulas. For example, for $1 < p < 2$, and A is a positive matrix,

$$A^{p-1} = c_p \int_0^\infty t^{p-1} \left[\frac{1}{t} - \frac{1}{1+A} \right] \mathrm{d}t \ ,$$

where

$$\frac{1}{c_p} = \int_0^\infty t^{p-1} \left[\frac{1}{t} - \frac{1}{1+1} \right] \mathrm{d}t \ .$$

This is useful since one may differentiate $(A + tB)^{-1}$ directly from the definition of the derivative, finding

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{1}{A + tB} = -\frac{1}{A + tB}B\frac{1}{A + tB}\ ,$$

where we assume $A > 0$ B self adjoint, and t such that $A + tB > 0$.

However, before going further into the difficulties of working with the Schatten norms, let us return to the L^p case, and introduce the inequalities that concern us, and their applications.

Uniform convexity in L^2

In any Hilbert space, in particular L^2, we have the parallelogram identity:

$$\left\|\frac{f+g}{2}\right\|_2^2 + \left\|\frac{f-g}{2}\right\|_2^2 = \frac{\|f\|_2^2 + \|g\|_2^2}{2} .$$

In particular, if f and g are unit vectors, this yields

$$\left\|\frac{f+g}{2}\right\|_2 \le \left(1 - \left\|\frac{f-g}{2}\right\|_2^2\right)^{1/2} \le 1 - \frac{1}{2}\left\|\frac{f-g}{2}\right\|_2^2 .$$

Thus we see that $\delta_{L^2}(\epsilon) \ge \frac{1}{2}\epsilon^2$ and an easy argument shows this is, in fact, equality. Since S^2 is also a Hilbert space (with the Hilbert-Schmidt inner product, $\delta_{S^2} = \delta_{L^2}$.

Uniform convexity in L^p, $p > 2$

Recall that for counting measure, $\|f\|_p \geq \|f\|_q$ for $p < q$, while for any probability measure, $\|f\|_p \leq \|f\|_q$ for $p < q$. For all $a, b > 0$,

$$\begin{aligned}
\left(\left|\frac{a+b}{2}\right|^p + \left|\frac{a-b}{2}\right|^p\right)^{1/p} &\leq \left(\left|\frac{a+b}{2}\right|^2 + \left|\frac{a-b}{2}\right|^2\right)^{1/2} \\
&= \left(\frac{a^2+b^2}{2}\right)^{1/2} \\
&\leq \left(\frac{a^p+b^p}{2}\right)^{1/p} .
\end{aligned}$$

Next, for all $z, w \in \mathbb{C}$, and all $p \geq 2$,

$$|z+w|^p + |z-w|^p \leq ||z| + |w||^p + ||z| - |w||^p \ .$$

Combining, we obtain,

$$\left|\frac{f(x)+g(x)}{2}\right|^p + \left|\frac{f(x)-g(x)}{2}\right|^p \leq \frac{|f(x)|^p + |g(x)|^p}{2}$$

Integrating in x yields *Clarkson's inequality*:

$$\left\|\frac{f+g}{2}\right\|_p^p + \left\|\frac{f-g}{2}\right\|_p^p \leq \frac{\|f\|_p^p + \|g\|_p^p}{2} \ .$$

If f and g are unit vectors, this yields

$$\left\| \frac{f+g}{2} \right\|_p \leq \left(1 - \left\| \frac{f-g}{2} \right\|_p^p \right)^{1/p} \leq 1 - \frac{1}{p} \left\| \frac{f-g}{2} \right\|_p^p .$$

Thus we see that

$$\delta_{L^p}(\epsilon) \geq \frac{1}{p}\epsilon^p ,$$

and an easy argument shows this is, in fact, equality.

Uniform convexity in L^p, $1 < p < 2$

Uniform convexity for $1 < p < 2$ is more subtle, and the result is somewhat surprising: It turns out that for $1 < p \leq 2$,

$$\delta_{L^p}(\epsilon) \geq \frac{p-1}{2}\epsilon^2 \; .$$

Notice that the exponent is 2, as in the Hilbert space case. However, as p decreases towards 1, the constant $(p-1)/2$ decreases to zero.

Both the exponent 2, and the constant $(p-1)/2$ are best possible, and both have significant implications, as we shall explain.

The result is in some sense due to Hanner, who exactly computed $\delta_{L^p}(\epsilon)$. The final remark in his 1955 paper is (in slightly different notation) that

$$\delta_{L^p}(\epsilon) = \frac{p-1}{2}\epsilon^2 + \mathcal{O}(\epsilon^3) \ ,$$

which certinaly shows that $\delta_{L^p}(\epsilon)$ is bounded below by some multiple of ϵ^2. The fact that the remainder terms is positive and may be droped, yielding the asserted lower bound, may be folklore, but appears in work by Ball and Pisier in the 1990's. They used Hanner's exact computationof δ_{L^p} and Gross's 2-point inequality, as we shall explain. (This gives a first hint at the connections with hypercontractivity.) However, the sharp bound may be proved directly, and this is crucila for the Schatten norms.

Direct proof of $\delta_{L^p}(\epsilon) \geq \frac{p-1}{2}\epsilon^2$

Let f and g be simple functions of the form

$$f(x) = \sum_{j=1}^{n} z_j 1_{A_j}(x) \qquad \text{and} \qquad g(x) = \sum_{j=1}^{n} w_j 1_{A_j}(x) \ ,$$

where for each j, $z_j w_j^*$ is not real. This guarantees that $z_j + tw_j \neq 0$ for any real t, and thus for all $x \in \cup_{j=1}^{n} A_j$, and all $t \in \mathbb{R}$, $f(x) + tg(x) \neq 0$. Define

$$Y(t) = \|f + tg\|_p^p \qquad \text{and} \qquad q = \frac{p}{2} \ ,$$

so that $\|f + tg\|_p^2 = Y^{1/q}(t)$.

Differentiating twice,

$$\begin{aligned} \frac{\mathrm{d}^2}{\mathrm{d}t^2}\|f+tg\|_p^2 &= \frac{1}{q}\left(\frac{1}{q}-1\right)Y^{1/q-2}(Y')^2+\frac{1}{q}Y^{1/q-1}Y'' \\ &\geq \frac{1}{q}Y^{1/q-1}Y'' \end{aligned}$$

Next, a simple calculation yields

$$Y''(t) \geq p(p-1)\int |f+tg|^{2q-2}|g|^2\mathrm{d}\mu \ .$$

To this we apply the *reverse Hölder inequality*, which says that for $0 < r < 1$ and $s = r/(r-1)$, whenever $a_j \geq 0$ for $j = 1, \ldots, n$, and $b_j > 0$ for $j = 1, \ldots, n$,

$$\sum_{j=1}^{n} a_j b_j \geq \left(\sum_{j=1}^{n} a_j^r \right)^{1/r} \left(\sum_{j=1}^{n} b_j^s \right)^{1/s} .$$

The result is that, for all t,

$$\frac{\mathrm{d}^2}{\mathrm{d}t^2} \|f + tg\|_p^2 \geq 2(p-1) \|g\|_p^2 .$$

Let $\psi''(t) \geq 2c$ for all t, and define

$$\varphi(t) := \psi(t) + ct(1-t) \ .$$

Then φ is convex, and thus

$$\varphi(1/2) \leq \frac{\varphi(0)+\varphi(1)}{2} \ , \qquad \text{that is,} \qquad \psi(1/2) + \frac{c}{4} \leq \frac{\psi(0)+\psi(1)}{2} \ .$$

We conclude that with f and g as above,

$$\|f+g/2\|_p^2 + \frac{p-1}{4}\|g\|_p^2 \leq \frac{\|f\|_p^2 + \|f+g\|_p^2}{2} \ .$$

The simple function approximation is now easily removed.

Now let u and v be vectors in L^p space, $1 < p \leq 2$, and let $f = u$ and $g = v - u$. Then

$$\left\|\frac{u+v}{2}\right\|_p^2 + (p-1)\left\|\frac{u-v}{2}\right\|_p^2 \leq \frac{\|u\|_p^2 + \|v\|_p^2}{2} \; .$$

If u and v are unit vectors, the right hand side is 1, and this implies

$$\left\|\frac{u+v}{2}\right\|_p \leq 1 - \frac{p-1}{2}\left\|\frac{u-v}{2}\right\|_p^2 \; ,$$

which proves that

$$\delta_{L^p}(\epsilon) \geq \frac{p-1}{2}\epsilon^2 \; .$$

Stability for Hölder's inequality

We give a first application of Uniform convexty to obtain a "remainder term" or "stability bound" for Hölder's inequality.

The *duality map* $\mathcal{D}_p$ on functions from L^p to the unit sphere in $L^{p'}$ is given by

$$\mathcal{D}_p(f)(x) = \|f\|_p^{1-p}|f|^{p-2}(x)\overline{f(x)}$$

The map has the property that $\int_X \mathcal{D}_p(f) f \mathrm{d}\mu = \|f\|_p$, and $\mathcal{D}_p(f)$ is the unique unit vector in the dual space $L^{p'}(X,\mu)$ to $L^p(X,\mu)$ that has this property.

The next theorem, due to C., Frank and Lieb, was appiled to prove stability bounds for eigenvalues of Schrödinger operators.

Theorem 0.1 (Hölder's inequality with remainder)**.** *Let* $1 < p \leq 2$*. Let* f *be a unit vector in* $L^{p'}(X,\mu)$*, and let* g *be a unit vector in* $L^p(X,\mu)$*. Then we have both*

$$\left|\int_X fg\,\mathrm{d}\mu\right| \leq 1 - \frac{p-1}{4}\|\mathcal{D}_{p'}(f) - e^{i\theta}g\|_p^2\,, \tag{1}$$

and

$$\left|\int_X fg\,\mathrm{d}\mu\right| \leq 1 - \frac{1}{p'\,2^{p'-1}}\|e^{i\theta}f - \mathcal{D}_p(g)\|_{p'}^{p'}\,. \tag{2}$$

where $\theta \in [0,2\pi)$ *is such that* $e^{i\theta}\int_X fg\,\mathrm{d}\mu$ *is positive. The exponents* 2 *and* p' *on the right sides of* (1) *and* (2) *are best possible.*

Proof of Theorem 0.1. Let $\|f\|_{p'} = \|g\|_p = 1$. Let θ be such that $e^{i\theta} \int_X f g \, \mathrm{d}\mu$ is positive. By Hölder's inequality,

$$1 + e^{i\theta} \int_X f g \, \mathrm{d}\mu = \int_X f(\mathcal{D}_{p'}(f) + e^{i\theta} g) \, \mathrm{d}\mu \leq \|\mathcal{D}_{p'}(f) + e^{i\theta} g\|_p \, .$$

Dividing by 2, and using the definition of θ,

$$\frac{1}{2} + \frac{1}{2} \left| \int_X f g \, \mathrm{d}\mu \right| \leq \left\| \frac{\mathcal{D}p'(f) + e^{i\theta} g}{2} \right\|_p \, ,$$

Now use $\left\| \frac{u+v}{2} \right\|_p \leq 1 - \frac{p-1}{2} \left\| \frac{u-v}{2} \right\|_p^2$, to obtain

$$\frac{1}{2} + \frac{1}{2} \left| \int_X f g \, \mathrm{d}\mu \right| \leq 1 - \frac{p-1}{8} \|\mathcal{D}_p(f) - e^{i\theta} g\|_{p'}^2 \, .$$

One can strengthen the bound by iterating, since in the first step of the proof we used the ordinary Hölder inequality, without taking into account the correction. Here is an application in which the strengthening is of interest.

Let ρ and σ be probability densities on some measure space (X, μ). For $1 < p \leq 2$, define $f = \rho^{1/p'}$ and $g = \sigma^{1/p}$ so that f and g are unit vectors in $L^{p'}$ and L^p respectively. In this case, $\mathcal{D}_{p'}(f) = f^{1/(p-1)} = \rho^{1/p}$. Hence our inequality implies

$$\int_X \rho^{1-1/p} \sigma^p \mathrm{d}\mu \leq 1 - \frac{p-1}{4} \|\rho^{1/p} - \sigma^{1/p}\|_p^2 \ .$$

Rearranging terms, we have

$$\frac{1}{1/p-1}\int_X \left[\rho^{1-1/p}\sigma^{1/p} - \sigma\right] \mathrm{d}\mu \geq \frac{p}{4}\|\rho^{1/p} - \sigma^{1/p}\|_p^2 \ .$$

In the limit $p \downarrow 1$, we obtain

$$D(\sigma||\rho) \geq \frac{1}{4}\|\rho - \sigma\|_1^2 \ .$$

This is Pinsker's inequality, apart from the factor of $1/4$, in place of $1/2$. Iterating corrects this. For this application, the constant $p-1$ and the power 2 are both essential. Pinsker's inequality is simply a differential expression of the optimal 2-uniform convexity of L^p, $1 < p \leq 2$.

The connection with hypercontractivity

Earlier we proved that for $1 < p \leq 2$, and $u, v \in L^p$, not necessarily unit vectors,

$$\left\|\frac{u+v}{2}\right\|_p^2 + (p-1)\left\|\frac{u-v}{2}\right\|_p^2 \leq \frac{\|u\|_p^2 + \|v\|_p^2}{2} \ .$$

Let $u = f + g$ and $v = f - g$, and an equivalent form is

$$\|f\|_p^2 + (p-1)\|g\|_p^2 \leq \frac{\|f+g\|_p^2 + \|f-g\|_p^2}{2} \ .$$

More is true:

$$\|f\|_p^2 + (p-1)\|g\|_p^2 \leq \left(\frac{\|f+g\|_p^p + \|f-g\|_p^p}{2}\right)^{2/p} \ .$$

The formally stronger inequality follows from what we proved earlier by a simple doubling argument. Replace the underlying measure space (Ω, μ) with $(\Omega \times \{-1, 1\}, \mu \otimes \nu)$ where ν is the fair Bernouili measure. Let 1 be the unit constant function on the 2-point space, and let Z be the identity function on it.

Define

$$\widetilde{f} = 1 \otimes f \quad \text{and} \quad \widetilde{g} = Z \otimes g \ .$$

Then

$$\|\widetilde{f} + \widetilde{g}\|_p^p = \|\widetilde{f} - \widetilde{g}\|_p^p = \frac{\|f + g\|_p^p + \|f - g\|_p^p}{2} \ .$$

Then since $\|\widetilde{f}\|_p = \|f\|_p$ and $\|\widetilde{g}\|_p = \|g\|_p$,

$$\begin{aligned} \|f\|_p^2 + (p-1)\|g\|_p^2 &= \|\widetilde{f}\|_p^2 + (p-1)\|\widetilde{g}\|_p^2 \\ &\le \frac{\|\widetilde{f}+\widetilde{g}\|_p^2 + \|\widetilde{f}-\widetilde{g}\|_p^2}{2} \\ &= \left(\frac{\|f+g\|_p^p + \|f-g\|_p^p}{2}\right)^{2/p} . \end{aligned}$$

Specialized to the 2-point space with $f = a1$ and $g = bZ$, $a, b \in \mathbb{C}$, we get Gross's 2-point inequality:

$$\left(a^2 + (p-1)b^2\right)^{1/2} \le \left(\frac{|a+b|^p + |a-b|^p}{2}\right)^{1/p} .$$

Now let $X_N := \{-1,1\}^N$, the N dimensional discrete cube. Let ν_N be normalized counting measure on X_N. For $x = (x_1, \ldots, x_N) \in X_N$, let

$$\pi_j(x) = x_j \ ,$$

$j = 1, \ldots, N$. Let $\alpha \in \{0,1\}^N$, and define

$$\varphi_\alpha(x) = \prod_{j=1}^{N} \pi_j^{\alpha_j}(x) \ .$$

The 2^N functions φ_α, $\alpha \in \{0,1\}^N$, are easily seen to be an orthonormal basis of

$$L^2(X_N, \nu_N) \ .$$

Define a semigroup $\{P_t\}_{t>0}$ on $L^2(X_N, \nu_N)$ by

$$P_t\varphi_\alpha = e^{-t|\alpha|}\varphi_\alpha \ ,$$

where $|\alpha| = \sum_{j=1}^{N} \alpha_j$.

Let h be any function on X_N. Then we can write

$$h = f + \pi_N g$$

for uniquely determined functions f and g on X_{N-1}. Using the obvious notation,

$$P_t h = P_t f + e^{-t}\pi_N P_t g \ .$$

Now $h(x) \geq 0$ for all x iff $f + g > 0$ and $f - g > 0$ everywhere on X_{N-1}. Suppose P_t preserves positivity on X_{N-1}. Then

$$\begin{aligned} h \geq 0 \quad &\Rightarrow \quad f + g \geq 0\ ,\ f - g \geq 0 \\ &\Rightarrow \quad P_t f + P_t g \geq 0\ ,\ P_t f - P_t g \geq 0 \\ &\Rightarrow \quad P_t h \geq 0\ . \end{aligned}$$

Since P_t clearly preserves positivity for functions on X_1, we see inductively that P_t preserves positivity for functions on X_N for all N.

Since evidently $P_t 1 = 1$, P_t is a contraction on L^∞. Since it is self-adjoint contraction on L^2, by interpolation and duality, it is a contraction on L^p for all p.

Theorem 0.2 (Gross's discrete hypercontractivity theorem)**.** *For all* $1 < p \leq 2$*, and all* f *on* X_N*,*

$$e^{-t} \leq \sqrt{p-1} \quad \Rightarrow \quad \|P_t f\|_2 \leq \|f\|_p ,$$

and the condition is sharp.
More generally, for all $1 < p < q < \infty$*,*

$$e^{-t} \leq \sqrt{\frac{p-1}{q-1}} \quad \Rightarrow \quad \|P_t f\|_q \leq \|f\|_p ,$$

and again, the condition is sharp.

Gross showed that this, together with the Central Limit Theorem, implies Nelson's optimal hypercontractivity theorem for Gauss space.

Proof: Consider the case $N = 1$. Any h on X_1 has the form $h = a1 + b\pi_1$ for some numbers a and b, and

$$P_t h = a + e^{-t} b\pi_1 \ .$$

Let $f = a1$ and $g = b\pi_1$, and apply

$$\left(\|f\|_p^2 + (p-1)\|g\|_p^2\right)^{1/2} \leq \left(\frac{\|f+g\|_p^p + \|f-g\|_p^p}{2}\right)^{1/p} \ .$$

This is the same as

$$\|P_{(-\ln\sqrt{p-1})} h\|_2 \leq \|h\|_p \ ,$$

which is the $N = 1$ case of the first inequality in the theorem.

In fact if one writes $\|P_{(-\ln\sqrt{p-1})}h\|_2 \leq \|h\|_p$ explicitly for $N=1$ and $h=a+b\pi_!$ in terms of a and b, one get's Gross's 2-point inequality, which Gross proved by direct calculation.

What Gross did then was to proceed directly to general N by using:

(1) On X_N, P_t is the N-fold tensor product of the two-point semigroup,

(2) *Segal's Lemma*, which say that if P and Q are positivity preserving operators from L^p to L^q for some measure space, then the L^p to L^q norm of $P \otimes Q$ is the product of the L^p to L^q norms of P and Q. In particular, if P and Q are contractice form L^p to L^q, so is $P \otimes Q$.

In the noncommutative setting, there is no general analog of Segal's lemma. Even for *completely positive* maps, the norm of the tensor product can be larger than the product of the norms, as was shown in work of Winter on the channel additivity conjecture that was finally disproved by Hastings.

It turns out that uniform convexity inequalities provide the means to get from $N = 1$ to general N, and that this approach also works in the non-commutative setting, as we shall explain.

First, we complete the proof of Gross' Theorem from the uniform convexity point of view.

We treat the case of general N by induction. We write $h = f + \pi_N g$ with f and g on X_{N-1}. As before,

$$P_t h = P_t f + e^{-t} \pi_N P_t g \ .$$

Let $e^{-t} = \sqrt{p-1}$, $\widetilde{f} = P_t f$ and $\widetilde{g} = \pi_N P_t g$. Then

$$\begin{aligned}
\|P_t h\|_2^2 &= \|\widetilde{f}\|_2^2 + (p-1)\|\widetilde{g}\|_2^2 \\
&= \|P_t f\|_2^2 + (p-1)\|P_t g\|_2^2 \\
&= \|f\|_p^2 + (p-1)\|g\|_p^2 \\
&\leq \left(\frac{\|f+g\|_p^p + \|f-g\|_p^p}{2} \right)^{2/p} = \|h\|_p^2 \ .
\end{aligned}$$

Mehler's formula

Any function f on X_N is a polynomial in the variables $x_1, \ldots, x_N$, at most linear in each variable. It is easy to see that

$$P_t f(x_1, \ldots, x_N) =$$

$$\int_{X_N} f(e^{-t}x_1 + \sqrt{1-e^{-2t}}\pi_1(y), \ldots, e^{-t}x_N + \sqrt{1-e^{-2t}}\pi_N(y)) \mathrm{d}\nu_N(y) \, .$$

If $z := (x_1 + \cdots + x_N)/\sqrt{N}$, and $f(x_1, \ldots, x_N) = \psi(z)$, for some bounded continuous functions ψ on $\mathbb{R}$,

$$P_t f(x_1, \ldots, x_N) = \int_{X_N} \psi(e^{-t}z + \sqrt{1-e^{-2t}}w) \mathrm{d}\nu_N(y)$$

where $w := (y_1 + \cdots + y_N)/\sqrt{N}$.

Thus, if f only depends on $x_1, \ldots, x_N$ through z, so does $P_t f$, and we may write the last equation as

$$P_t\psi(z) = \int_{X_N} \psi(e^{-t}z + \sqrt{1 - e^{-2t}}w)\mathrm{d}\nu_N(y)$$

As $N \to \infty$, the CLT says that the law of w tends to

$$\gamma = \frac{`1}{\sqrt{2\pi}} e^{-w^2/2}\mathrm{d}w \ .$$

Taking the limit $N \to \infty$, we obtain the *Mehler formula*

$$P_t\psi(z) = \int_{\mathbb{R}} \psi(e^{-t}z + \sqrt{1 - e^{-2t}}w)\mathrm{d}\gamma \ .$$

If we deine $\psi(x,t) = P_t\psi_0(x)$, then $\psi(x,t)$ solves

$$\frac{\partial}{\partial t}\rho(t,x) = (\nabla - x)\cdot\nabla\rho(x,t)\ .$$

Since our hypercontractivity bound for the cube is dimension-free, we obtain the following hypercontractivity bound for the Mehler semigroup: For all $1 < p \leq 2$,

$$e^{-t} \leq \sqrt{p-1} \quad \Rightarrow \quad \|P_t\psi\|_2 \leq \|\psi\|_p\ .$$

which is sharp.

Hanner's inequality

Theorem 0.3. *Let (X, μ) be a measure space, and let $1 < p \leq 2$. For $f, g \in L^p$,*

$$\|f + g\|_p^p + \|f - g\|_p^p \geq (\|f\|_p + \|g\|_p)^p + |\|f\|_p - \|g\|_p|^p \, .$$

For $2 \leq p < \infty$, the inequality reverses.

From this and Gross' 2-point inequality,

$$\left(\frac{|a+b|^p + |a-b|^p}{2} \right)^{1/p} \geq \left(a^2 + (p-1) b^2 \right)^{1/2} \; ,$$

we obtain

$$\left(\frac{\|f + g\|_p^p + \|f - g\|_p^p}{2} \right)^{1/p} \geq (\|f\|_p^2 + (p-1) \|g\|_p^2)^{1/2} \, .$$

This derivation of the 2-uniform convexity bound from Hanner's iequality is due to Ball and Pisier.

In fact, Hanner's inequality gives the optimal lower bound on $\|f+g\|_p^p + \|f-g\|_p^p$ for given values of $\|f\|_p$ and $\|g\|_p$, and as he showed, it goves the *exact* modulus of convexity for L^p, $1 < p`2$.

The proof is not difficult. We discuss it here, largley following Hanner's own treatment. Unfortunately, it does not carry over to the Schatten norm setting.

The basic fact behind Hanner's inequality is that the function

$$\varphi_p(t) = (1 + t^{1/p})^p + (1 - t^{1/p})^p \ ,$$

defined $[0,1]$, is convex for $1 \le p \le 2$, and is concave for $2 \le p < \infty$. Indeed:

$$\varphi_1(t) = 1 \ , \quad \varphi_2(t) = 2 + 2t \ , \quad \varphi_4(t) = 2 + 12\sqrt{t} + 2t \ .$$

We compute:

$$\begin{aligned} \varphi_p'(t) &= [(1 + t^{1/p})^{p-1} - (1 - t^{1/p})^{p-1}]t^{(1-p)/p} \\ &= (t^{-1/p} + 1)^{p-1} - (t^{-1/p} - 1)^{p-1} \end{aligned}$$

Then

$$\varphi_p''(t) = \frac{p-1}{p} t^{-1/p-1} \left[(t^{-1/p}+1)^{p-2} + (t^{-1/p}-1)^{p-2} \right] \geq 0 \ .$$

Hence,

$$\varphi_p(t) \geq \varphi_p(t_0) + \varphi_p'(t_0)(t-t_0) = [\varphi_p(t_0) - t_0\varphi_p'(t_0)] + \varphi_p'(t_0)t$$

for all t, with equality at $t = 0$.

Since $\varphi_p(t_0) - t_0\varphi_p'(t_0) = (1+t_0^{1/p})^{p-1} + (1-t_0^{1/p})^{p-1}$. we deine

$$\alpha(r) = (1+r)^{p-1} + |1-r|^{p-1}\mathrm{sgn}(1-r) \ ,$$

and then

$$\varphi_p(t) \geq \alpha(t_0^{1/p}) + \alpha(t_0^{-1/p})t \ .$$

Now let x, y be any real numbers. Consider the quantity

$$|x+y|^p + |x-y|^p$$

Assume that $0 < y \le x$. Define $t = y^p/x^p$. Then

$$\begin{aligned} |x+y|^p + |x-y|^p &= |x|^p \varphi_p(t) \ge |x|^p \left(\alpha(t_0^{1/p}) + \alpha(t_0^{-1/p}) t \right) \\ &= \alpha(t_0^{1/p})|x|^p + \alpha(t_0^{-1/p})|y|^p \end{aligned}$$

and there is equality at $t_0 = y^p/x^p$. By symmetry in x and y, we conclude:

Lemma 0.4. *For $1 < p \leq 2$, all $r \in (0, \infty)$, let*

$$\alpha(r) := |1 + r|^{p-1} + |1 - r|^{p-1}\mathrm{sgn}(1 - r) \ .$$

Then for all $x, y \in \mathbb{R}$,

$$|x + y|^p + |x - y|^p = \sup_{r>0}\{\alpha(r)|x|^p + \alpha(1/r)|y|^p\} \ .$$

The analogous result for $2 \leq p < \infty$ is true if one replaces supremum by infimum.

The lemma immediately yields Hanner's inequality since

$$\begin{aligned}
\|f+g\|_p^p + \|f-g\|_p^p &= \int_X \left[|f+g|^p + |f-g|^p\right] \mathrm{d}\mu \\
&= \int_X \left[\sup_{r>0}\{\alpha(r)|f|^p + \alpha(1/r)|g|^p\}\right] \mathrm{d}\mu \\
&\leq \sup_{r>0} \left\{ \int_X \left[\alpha(r)|f|^p + \alpha(1/r)|g|^p\right] \mathrm{d}\mu \right\} \\
&= \sup_{r>0}\{\alpha(r)|f|_p^p + \alpha(1/r)|g|_p^p\} \\
&= (\|f\|_p + \|g\|_p)^p + |\|f\|_p - \|g\|_p|^p \ .
\end{aligned}$$

Uniform convexity for trace norms

For the Schatten trace norms we have:

Theorem 0.5. *For all $n \times n$ matrices X and Y, and all $1 < p \leq 2$,*

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

The inequality reverses for $p > 2$.

Theorem 0.6. *For $n \times n$ matrices X and Y,*

$$\|X+Y\|_p^p + \|X-Y\|_p^p \geq (\|X\|_p + \|Y\|_p)^p + |\|X\|_p - \|Y\|_p|^p$$

provided $1 < p \leq 4/3$, or $1 < p \leq 2$ and $X \pm Y$ are positive matrices. The reverse inequality is valid proved $p \geq 4$ or $p \geq 2$ and $X \pm Y$ are positive matrices.

Unfortunately, in the matrix norm case, we are not in a position to prove the 2-uniform convexity theorem using the analog of Hanner's inequality, as we have done in our second proof for the L^p norms. Hence, we are led to try a differentiation argument.
The key is to differentiate

$$\|X + sY\|_p^p$$

twice, as we did in our proof for L^p.

To focus on the main ideas, let us consider the case $X > 0$ and Y self adjoint. Then at least for small t, $X + tY > 0$, and we may use the integral representation formula.

$$A^{p-1} = c_p \int_0^\infty t^{p-1} \left[\frac{1}{t} - \frac{1}{1+A}\right] \mathrm{d}t \ ,$$

For instance, for X strictly positive and Y self adjoint

$$\frac{\mathrm{d}}{\mathrm{d}s}(X+sY)^{p-1}\bigg|_{s=0} = c_p \int_0^\infty t^{p-1}\left[\frac{1}{1+X}Y\frac{1}{1+X}\right]\mathrm{d}t\ ,$$

and hence

$$\frac{\mathrm{d}^2}{\mathrm{d}^2 s}\,\mathrm{Tr}(X+sY)^{p}\bigg|_{s=0} = pc_p \int_0^\infty t^{p-1}\,\mathrm{Tr}\left[\frac{1}{1+X}Y\frac{1}{1+X}Y\right]\mathrm{d}t\ ,$$

It may be shown that for fixed Y,

$$X \mapsto \mathrm{Tr}\left[\frac{1}{1+X}Y\frac{1}{1+X}Y\right] := F(X)$$

is a convex function of X.

The convexity may be exploited as follows: Let U be any unitary matrix.Then

$$\begin{aligned} F(X) &= \mathrm{Tr}\left[U^* \frac{1}{1+X} Y \frac{1}{1+X} YU\right] \\ &= \mathrm{Tr}\left[U^* \frac{1}{1+X} UU^*YUU^* \frac{1}{1+X} UU^*YU\right] \\ &= \mathrm{Tr}\left[\frac{1}{1+U^*XU} U^*YU \frac{1}{1+U^*XU} U^*YU\right] \end{aligned}$$

If $U^*YU = Y$; i.e, $YU = UY$, then we have

$$F(X) = F(U^*XU) \ .$$

Let $\widehat{A}$ be the average of U^*AU over all unitaries commuting with B. Then

$$F(A) \geq F(\widehat{A}) \ .$$

Since $\widehat{A}$ is obtained from A by erasing all of the off-diagonal entries of X in a basis in which Y is diagonal, we make contact with the ℓ_p case.
Of course one first has to reduce to the case $X > 0$, and Y self adjoint, which would correspond to $f > 0$ and g real in the function case. This involves more arguments of the same flavor, and an adaptation of the "doubling trick" to matrices. In this way, one obtains the proof of the first theorem

A Matrix rearrangement conjecture

Let A be an $n \times n$ matrix. Let $\Sigma(A)$ be the diagonal matrix with the singular values of A on the diagonal, arranged in decreasing order from the upper left. C. and Lieb conjectured that for $1 \le p \le 2$,

$$\|X+Y\|_p^p + \|X-Y\|_p^p \ge \|\Sigma(X)+\Sigma(Y)\|_p^p + \|\Sigma(X)-\Sigma(Y)\|_p^p \ ,$$

wit the reverse inequality holding for $p > 2$.

Were this conjecture true, it would immediately reduce the proof of Hanner's inequality for matrices to the the known case of ℓ_p.

C. and Lieb proved is under the additions assumption that X and Y are self adjoint and $X \ge |Y|$.

Recall that for complex numbers, and $1 \leq p \leq 2$,

$$|z+w|^p + |z-w|^p \geq ||z|+|w||^p + ||z|-|w||^p \ .$$

Hence

$$\|f+g\|_p^p + \|f-g\|_p^p \geq \||f|+|g|\|_p^p + \||f|-|g|\|_p^p$$

and we may therefore restrict consideration to positive functions.
For matrix norms, things are not so nice: There is an inequality going the opposite way:

Theorem 0.7. *Let A, B be $n \times n$ self adjoint matrices with $A \geq |B| \geq 0$. Then for $1 \leq p \leq 2$,*

$$\|A+B\|_p^p + \|A-B\|_p^p \leq \||A|+|B|\|_p^p + \||A|-|B|\|_p^p \ .$$

Fermion hypercontractivity

Fortunately, while the full analog of Hanner's inequality is open in the non-commutative case, it is the optimal 2-uniform convexity inequality that is directly relevant to hypercontractivity, and this we have.
A non-commutative analog of functions on the discrete cube arising in quantum mechanics is the algebra generated by N self-adjoint operators $Q_1, \ldots, Q_N$ such that

$$Q_i Q_j + Q_j Q_i = 2\delta_{i,j} I \ .$$

This is a Clifford algebra, and various realizations as matrices on $\mathbb{C}^{2n}$ can be given.

Choosing one such representation, let τ denote the normalized trace. For A in the algebra generated by the Q's, we define

$$\|A\|_p = \left(\tau(|A|^p)\right)^{1/p} .$$

The normalization does not affect the uniform convexity properties.

Again, for $\alpha = (\alpha_1, \ldots, \alpha_N) \in \{0,1\}^N$ define

$$Q_\alpha = Q_1^{\alpha_1} \cdots Q_N^{\alpha_N} .$$

These are orthonormal in $L^2(\tau)$; i.e., the normalized Hilbert-Schmidt inner product, and are a basis.

Define a semigroup P_t by

$$P_t Q_\alpha = e^{-t|\alpha|} Q_\alpha \ ,$$

in exact analogy with what we did for functions on the discrete cube. Then since for any A in our algebra,

$$A = X + Q_N Y$$

where X and Y are in the algebra generated by the first $N-1$ Q's, the same inductive argument lead to

$$e^{-t} \leq \sqrt{q-1} \quad \to \quad \|P_t A\|_2 \leq \|A\|_p \ ,$$

which is how C. and Lieb proved Gross's conjecture for Fermion hypercontractivity.