Sharp estimates for conditionally centred moments and for compact operators on L^p spaces

Eugene Shargorodsky* and Teo Sharia**

* King's College London and Technische Universität Dresden
** Royal Holloway, University of London

Prologue: motivation from Analysis

Question from the theory of Hardy spaces / Toeplitz operators: what is the norm of the backward shift operator?

Equivalently,

what is the norm of the operator

$$f \quad \longmapsto \quad f - \frac{1}{2\pi} \int_{-\pi}^{\pi} f\left(e^{i\theta}\right) \, d\theta$$

on the Hardy space $H^p(\mathbb{T})$? Still unknown for $p \neq 2$!

Mathematical Surveys and Monographs
Volume 79

The Backward Shift on the Hardy Space

Joseph A. Cima
William T. Ross

American Mathematical Society

What is the norm of the operator

$$f \quad \longmapsto \quad f - \frac{1}{2\pi}\int_{-\pi}^{\pi} f\left(e^{i\theta}\right)\, d\theta$$

on $L^p(\mathbb{T})$? The (not widely!) known answer has led to the research presented here.

Estimates for centred moments

Notation:

$(\Omega, \mathcal{F}, \mathbf{P})$ – probability space

ξ – (real valued) random variable (r.v.)

$\mathbf{E}\xi := \int_\Omega \xi(\omega)\, d\mathbf{P}(\omega)$ – expectation of ξ,

We will always assume that Ω consists of more than one element and that $\mathcal{F}$ is nontrivial, i.e. $\mathcal{F} \neq \{\emptyset, \Omega\}$.

Question:

What is the optimal constant $c_p = c_p(\Omega, \mathcal{F}, \mathbf{P})$ in the following estimate?

$$\|\xi - \mathbf{E}\xi\|_p = \left(\mathbf{E}|\xi - \mathbf{E}\xi|^p\right)^{1/p} \le c_p \left(\mathbf{E}|\xi|^p\right)^{1/p} = c_p\|\xi\|_p, \quad 1 \le p < \infty$$

$$\|\xi - \mathbf{E}\xi\|_\infty = \operatorname*{ess\,sup}_{\omega\in\Omega} |\xi(\omega) - \mathbf{E}\xi| \le c_\infty \operatorname*{ess\,sup}_{\omega\in\Omega} |\xi(\omega)| = c_\infty\|\xi\|_\infty.$$

Preliminary analysis

Textbook case: $c_2 = 1$. Indeed,

$$\mathbf{E}|\xi - \mathbf{E}\xi|^2 = \mathbf{E}(\xi - \mathbf{E}\xi)^2 = \mathbf{E}\xi^2 - (\mathbf{E}\xi)^2 \le \mathbf{E}\xi^2.$$

If ξ is not a constant r.v., then $\eta := \xi - \mathbf{E}\xi \ne 0$, but $\mathbf{E}\eta = 0$.
Hence $\|\eta - \mathbf{E}\eta\|_p = \|\eta\|_p$, and $\boxed{c_p \ge 1 \text{ for all } p \in [1, \infty].}$

Hölder's inequality $\implies$ $|\mathbf{E}\xi| \le \mathbf{E}|\xi| = \|\xi\|_1 \le \|\xi\|_p$.

Hence

$$\|\xi - \mathbf{E}\xi\|_p \le \|\xi\|_p + \|\mathbf{E}\xi\|_p = \|\xi\|_p + |\mathbf{E}\xi| \le 2\|\xi\|_p.$$

So, $\boxed{c_p \le 2 \text{ for all } p \in [1, \infty].}$

Suppose that for every $\alpha \in (0,1)$, there exists $A \in \mathcal{F}$ such that $\mathbf{P}(A) = \alpha$.

Let $\xi := \mathbb{1}_A$. Then $\mathbf{P}(\xi = 1) = \alpha,\ \mathbf{P}(\xi = 0) = 1 - \alpha$,

$\mathbf{E}|\xi| = \mathbf{E}\xi = \alpha,\ \ \mathbf{E}|\xi - \mathbf{E}\xi| = 2\alpha(1-\alpha),$ and

$$\frac{\|\xi - \mathbf{E}\xi\|_1}{\|\xi\|_1} = 2(1-\alpha).$$

Sending α to 0, one concludes that $\boxed{c_1 = 2.}$

Similarly, if $\xi := \mathbb{1}_A - \mathbb{1}_{\Omega \setminus A}$, then

$\mathbf{P}(\xi = 1) = \alpha,\ \mathbf{P}(\xi = -1) = 1 - \alpha$,

$\mathbf{E}\xi = 2\alpha - 1$, $\|\xi\|_\infty = 1$, and $\|\xi - \mathbf{E}\xi\|_\infty = \max\{2\alpha, 2(1-\alpha)\}$.
Sending α to 1 or to 0, one concludes that $\boxed{c_\infty = 2.}$

Putting together

$$c_1 = 2 = c_\infty, \qquad c_2 = 1, \qquad 1 \le c_p \le 2 \quad \text{for all} \quad p \in (1, \infty),$$

where the first equality holds if there are $A \in \mathcal{F}$ with arbitrarily small positive $\mathbf{P}(A)$, e.g., if $\mathbf{P}$ is nonatomic.

c_p is the norm of the operator $\mathbf{C} : L^p(\Omega, \mathbf{P}) \to L^p(\Omega, \mathbf{P})$,

$$\xi \mapsto \mathbf{C}\xi := \xi - \mathbf{E}\xi.$$

Applying the Riesz-Thorin interpolation theorem to this operator, one deduces from the above equalities that

$$c_p \le 2^{\left|1 - \frac{2}{p}\right|}, \quad 1 < p < \infty$$

(S. Rolewicz, 1990).

Let $(\Omega, \mathcal{F}, \mathbf{P}) = ([0,1], \mathcal{L}, \lambda)$, where λ is the standard Lebesgue measure on $[0,1]$ and $\mathcal{L}$ is the σ-algebra of Lebesgue measurable subsets of $[0,1]$. (Equivalently, one can assume that Ω is a complete separable metric space, $\mathcal{F} = \mathcal{B}$ is the Borel σ-algebra of Ω, and $\mathbf{P}$ is nonatomic.)

C. Franchetti (1990):

$$c_p = c_p([0,1], \mathcal{L}, \lambda) = \max_{0<\alpha<1} C_p(\alpha) =: C_p \quad \text{for all} \quad 1 < p < \infty,$$

where

$$C_p(\alpha) := \left(\alpha^{p-1} + (1-\alpha)^{p-1}\right)^{\frac{1}{p}} \left(\alpha^{\frac{1}{p-1}} + (1-\alpha)^{\frac{1}{p-1}}\right)^{1-\frac{1}{p}}.$$

$$C_1 := \lim_{p\to 1+0} C_p = 2 = c_1, \qquad C_\infty := \lim_{p\to\infty} C_p = 2 = c_\infty,$$

$$C_2 = 1 = c_2, \qquad C_{p'} = C_p \quad \text{for} \quad p' = \frac{p}{p-1}\,, \qquad C_p \le 2^{\left|1-\frac{2}{p}\right|}$$

Explicit values:

$$C_3 = \frac{1}{3}\left(17 + 7\sqrt{7}\right)^{1/3} = 1.0957\ldots,$$

$$C_4 = \left(1 + \frac{2}{3}\sqrt{3}\right)^{1/4} = 1.21156\ldots$$

T.F. Móri (2009)

$$c_p(\Omega, \mathcal{F}, \mathbf{P}) \leq C_p \qquad (*)$$

holds for all probability spaces $(\Omega, \mathcal{F}, \mathbf{P})$.

A simple calculation shows that $c_p(\Omega, \mathcal{F}, \mathbf{P}) \geq C_p$ if there exist $A \in \mathcal{F}$ such that $\mathbf{P}(A) = \alpha_p$, where α_p is a point at which $C_p(\alpha)$ attains its global maximum. Obviously, this is satisfied if $\mathbf{P}$ is nonatomic.

Móri's proof of $(*)$ relies on the observation that every zero mean probability distribution on $\mathbb{R}$ is a mixture of distributions concentrated on two points and having zero mean. This allows one to reduce the proof of $(*)$ to showing that

$$\frac{(\mathbf{E}|\xi - \mathbf{E}\xi|^p)^{1/p}}{(\mathbf{E}|\xi|^p)^{1/p}} \leq C_p$$

holds for every r.v. ξ that takes only two values. The latter is an elementary although not an entirely trivial calculation.

G. Lewicki and L. Skrzypek (2016)

$\Omega = \{1, \ldots, n\}$ and $\mathbf{P}$ is the uniform distribution: $\mathbf{P}(k) = \frac{1}{n}$, $k = 1, \ldots, n$.

For $n = 3, 4$ and for all sufficiently large n, one has

$$c_p = \max\left\{ C_p\left(\frac{k_1}{n}\right),\ C_p\left(\frac{k_2}{n}\right)\right\}, \qquad (**)$$

where

$$k_1 := \max\left\{k \in \mathbb{N} :\ \frac{k}{n} \le \alpha_p\right\}, \quad k_2 := \min\left\{k \in \mathbb{N} :\ \alpha_p \le \frac{k}{n} < \frac{1}{2}\right\},$$

and $\alpha_p \in (0, 1/6)$ is the unique point at which $C_p(\alpha)$ attains its global maximum in $[0, 1/2]$.

Lewicki and Skrzypek showed that c_p in $(**)$ tends to Franchetti's c_p as $n \to \infty$ and obtained an alternative proof of Franchetti's result.

$$1 \leq c_p(\Omega, \mathcal{F}, \mathbf{P}) \leq C_p \quad \text{for all probability spaces} \quad (\Omega, \mathcal{F}, \mathbf{P}).$$

For every $c \in [1, C_p]$, there exist a probability space $(\Omega, \mathcal{F}, \mathbf{P})$ such that $c_p(\Omega, \mathcal{F}, \mathbf{P}) = c$.
Indeed, let $\Omega = \{-1, 1\}$, $\mathbf{P}(-1) = 1 - \alpha$, $\mathbf{P}(1) = \alpha$. Then $c_p = C_p(\alpha)$, and one can choose α in such a way that $c_p = c$.

All the above results remain true for complex valued random variables.

Question: What is the norm of the operator

$$L^p(\mathbb{T}) \ni f \mapsto A_n f := \Big(f - n\text{-th Fejér mean of } f\Big) \in L^p(\mathbb{T})?$$

We know that

$$1 \leq \|A_n\|_{L^p \to L^p} \leq 2^{\left|1-\frac{2}{p}\right|}, \quad 1 \leq p \leq \infty, \quad n \in \mathbb{N},$$

but this is unlikely to be sharp.

Conditional expectation operator

Let $(\Omega, \mathcal{F}, \mathbf{P})$ be a probability space, $\mathcal{G}$ be a sub-σ-algebra of $\mathcal{F}$, and let $\mathbf{E}^{\mathcal{G}} = \mathbf{E}(\cdot|\mathcal{G})$ be the corresponding conditional expectation operator.

$L^2(\Omega, \mathcal{G}, \mathbf{P})$ is a closed linear subspace of $L^2(\Omega, \mathcal{F}, \mathbf{P})$
and
$\mathbf{E}^{\mathcal{G}} : L^2(\Omega, \mathcal{F}, \mathbf{P}) \to L^2(\Omega, \mathcal{G}, \mathbf{P})$ is the orthogonal projection.

Example

Let $\{\Omega_j\}_{j=1}^N$, $N \in \mathbb{N} \cup \{\infty\}$ be a measurable partition of Ω:

$$\Omega_j \in \mathcal{F}, \qquad \Omega = \bigcup_{j=1}^N \Omega_j, \qquad \Omega_j \cap \Omega_k = \emptyset,\ j \neq k,$$

and let $\mathcal{G}$ be the σ-algebra generated by $\{\Omega_j\}_{j=1}^N$.

Then

$$\mathbf{E}^{\mathcal{G}}\xi = \sum_{n=1}^N \left(\frac{1}{\mathbf{P}(\Omega_n)} \int_{\Omega_n} \xi \, d\mathbf{P} \right) \mathbb{1}_{\Omega_n}.$$

$\mathbf{E}^{\mathcal{G}} : L^p(\Omega, \mathcal{F}, \mathbf{P}) \to L^p(\Omega, \mathcal{G}, \mathbf{P})$, $1 \leq p \leq \infty$ is a contractive projection that preserves constants, i.e.

$$\left\|\mathbf{E}^{\mathcal{G}}\xi\right\|_p \leq \|\xi\|_p, \quad \mathbf{E}^{\mathcal{G}}\left(\mathbf{E}^{\mathcal{G}}\xi\right) = \mathbf{E}^{\mathcal{G}}\xi \quad \text{for all} \quad \xi \in L^p(\Omega, \mathcal{F}, \mathbf{P}),$$
$$\mathbf{E}^{\mathcal{G}}\mathbb{1} = \mathbb{1},$$

where $\mathbb{1}(\omega) = 1$ a.s.

Every contractive projection on $L^p(\Omega, \mathcal{F}, \mathbf{P})$, $p \in [1, \infty) \setminus \{2\}$ that preserves constants is the conditional expectation operator $\mathbf{E}^{\mathcal{G}}$ for a certain sub-σ-algebra $\mathcal{G} \subseteq \mathcal{F}$ (T. Ando, 1966).

Question:

What is the optimal constant $c_p = c_p(\Omega, \mathcal{F}, \mathcal{G}, \mathbf{P})$ in the following estimate?

$$\left\| \xi - \mathbf{E}^{\mathcal{G}} \xi \right\|_p \leq c_p \| \xi \|_p, \quad 1 \leq p \leq \infty. \tag{1}$$

In other words, what is the norm of the operator $I - \mathbf{E}^{\mathcal{G}}$,

$$c_p(\Omega, \mathcal{F}, \mathcal{G}, \mathbf{P}) = \left\| I - \mathbf{E}^{\mathcal{G}} \right\|_{L^p(\Omega, \mathcal{F}, \mathbf{P}) \to L^p(\Omega, \mathcal{F}, \mathbf{P})},$$

where I is the identity operator?

Preliminary analysis

$$\|\mathbf{E}^{\mathcal{G}}\| = 1 \implies \boxed{\|I - \mathbf{E}^{\mathcal{G}}\| \leq 2.}$$

If $\mathcal{G} \neq \mathcal{F}$, then there exists a r.v. ξ such that $\eta := \xi - \mathbf{E}^{\mathcal{G}}\xi \neq 0$. Since $\mathbf{E}^{\mathcal{G}}\eta = 0$, one has $\|\eta - \mathbf{E}^{\mathcal{G}}\eta\|_p = \|\eta\|_p$, and $\boxed{\|I - \mathbf{E}^{\mathcal{G}}\| \geq 1}$ for all $p \in [1, \infty]$.

For $p = 2$, it follows from

$$\mathbf{E}|\xi - \mathbf{E}^{\mathcal{G}}\xi|^2 = \mathbf{E}|\xi|^2 - \mathbf{E}|\mathbf{E}^{\mathcal{G}}\xi|^2 \leq \mathbf{E}|\xi|^2$$

that $\boxed{c_2(\Omega, \mathcal{F}, \mathcal{G}, \mathbf{P}) = 1.}$

Theorem

$$c_p(\Omega, \mathcal{F}, \mathcal{G}, \mathbf{P}) \le C_p, \quad 1 \le p \le \infty.$$

Remark. In general, the inequality $c_p(\Omega, \mathcal{F}, \mathcal{G}, \mathbf{P}) \le c_p(\Omega, \mathcal{F}, \mathbf{P})$ does not hold.

Let $1 < p < \infty$, $\alpha_p \in (0, 1)$ be a point at which $C_p(\alpha)$ attains its maximum, $\Omega = \{-1, 0, 1\}$, $\mathbf{P}(-1) = \tau(1 - \alpha_p)$, $\mathbf{P}(1) = \tau\alpha_p$, $\mathbf{P}(0) = 1 - \tau$, $0 < \tau < 1$, and

$$\mathcal{G} = \Big\{\emptyset, \{0\}, \{-1, 1\}, \Omega\Big\}.$$

Then

$$c_p(\Omega, \mathcal{F}, \mathbf{P}) \le 1 + \tau^{1/p} + \tau^{1-1/p} \to 1 \quad \text{as} \quad \tau \to 0,$$
$$c_p(\Omega, \mathcal{F}, \mathcal{G}, \mathbf{P}) = C_p.$$

Theorem

For every $p \in [1, \infty]$ and every $c \in [1, C_p]$, there exists a sub-σ-algebra $\mathcal{G} \subset \mathcal{L}$ such that

$$c_p([0,1], \mathcal{L}, \mathcal{G}, \lambda) = c.$$

Remark. If a sub-σ-algebra $\mathcal{G}$ is much smaller than $\mathcal{L}$, then $c_p([0,1], \mathcal{L}, \mathcal{G}, \lambda) = C_p$. More precisely, if $(\Omega, \mathcal{F}, \mathbf{P})$ is a separable nonatomic probability space and there exists a r.v. ξ on $(\Omega, \mathcal{F}, \mathbf{P})$, which is independent of a sub-σ-algebra $\mathcal{G} \subset \mathcal{F}$ and has a nontrivial Gaussian distribution, then $c_p(\Omega, \mathcal{F}, \mathcal{G}, \mathbf{P}) \geq C_p$, $1 \leq p < \infty$ (A. Dorogovtsev and M. Popov, 2008 + C. Franchetti, 1992).

Estimates for compact operators

For a Banach space X, let $\mathcal{K}(X)$ denote the set of compact linear operators on X.

Definition

A Banach space X is said to have the bounded compact approximation property (BCAP) if there exists a constant $M \in (0, +\infty)$ such that given any $\varepsilon > 0$ and any finite set $F \subset X$, there exists an operator $T \in \mathcal{K}(X)$ such that

$$\|I - T\| \leq M \quad \text{and} \quad \|x - Tx\| < \varepsilon \quad \text{for all} \quad x \in F.$$

We denote by $M(X)$ the infimum of the constants M for which the above conditions are satisfied.

Many authors have the condition $\|T\| \leq M$ in place of $\|I - T\| \leq M$ in the definition of BCAP and of related approximation properties. Let $m(X)$ be the infimum of the constants M for which the conditions in this alternative definition of BCAP are satisfied. It is clear that

$$m(X) - 1 \leq M(X) \leq m(X) + 1.$$

If one is not interested in sharp constants, then it usually does not matter whether one knows $m(X)$ or $M(X)$.

It is well known that $m(L^p([0,1])) = 1$, $1 \leq p < \infty$. The next result answers the question about the value of $M(L^p([0,1]))$.

Theorem

$$M(L^p([0,1])) = C_p, \quad 1 \leq p < \infty.$$

The estimate

$$M(L^p([0,1])) \leq C_p$$

is proved by constructing a suitable conditional expectation operator $T = \mathbf{E}^{\mathcal{G}}$ and using the estimate $c_p(\Omega, \mathcal{F}, \mathcal{G}, \mathbf{P}) \leq C_p$.

The estimate

$$M(L^p([0,1])) \geq C_p$$

is a corollary of the following result.

Theorem

Let $1 \leq p < \infty$ and $T \in \mathcal{K}(L^p)$. Then

$$\|I - T\|_{L^p \to L^p} + \inf_{\|u\|_{L^p}=1} \|(I - T)u\|_{L^p} \geq C_p.$$

Theorem

Let $1 \leq p < \infty$, $T \in \mathcal{K}(L^p)$, and suppose $I - T$ is not invertible (i.e. 1 is an eigenvalue of T). Then

$$\|I - T\|_{L^p \to L^p} \geq C_p.$$

The next result shows how an arbitrary distribution with mean zero can be expressed as a mixture of centered two-point distributions. For any $a \le 0 \le b$, let $\nu_{a,b}$ denote the unique probability measure on $\{a, b\}$ with mean zero. Clearly, $\nu_{a,b} = \delta_0$ when $ab = 0$; otherwise,

$$\nu_{a,b} = \frac{b\delta_a - a\delta_b}{b - a}, \quad a < 0 < b.$$

It is easy to verify that ν is a probability kernel from $\mathbb{R}_- \times \mathbb{R}_+$ to $\mathbb{R}$. For mappings between two measure spaces, measurability is defined in terms of the σ-fields generated by all evaluation maps $\pi_B : \mu \mapsto \mu B$, where B is an arbitrary set in the underlying σ-field.

Lemma 12.4 *(randomization) For any distribution μ on $\mathbb{R}$ with mean zero, there exists a distribution μ^* on $\mathbb{R}_- \times \mathbb{R}_+$ with $\mu = \int \mu^*(dx\,dy)\nu_{x,y}$. Here we may choose μ^* to be a measurable function of μ.*

Proof (Chung): Let $\mu_\pm$ denote the restrictions of μ to $\mathbb{R}_\pm \setminus \{0\}$, define $l(x) \equiv x$, and put $c = \int l d\mu_+ = -\int l d\mu_-$. For any measurable function $f : \mathbb{R} \to \mathbb{R}_+$ with $f(0) = 0$, we get

$$\begin{aligned} c \int f d\mu &= \int l d\mu_+ \int f d\mu_- - \int l d\mu_- \int f d\mu_+ \\ &= \int\int (y - x)\mu_-(dx)\mu_+(dy) \int f d\nu_{x,y}, \end{aligned}$$

and so we may take

$$\mu^*(dx\,dy) = \mu\{0\}\delta_{0,0}(dx\,dy) + c^{-1}(y - x)\mu_-(dx)\mu_+(dy).$$

The measurability of the mapping $\mu \mapsto \mu^*$ is clear by a monotone class argument if we note that $\mu^*(A \times B)$ is a measurable function of μ for arbitrary $A, B \in \mathcal{B}(\mathbb{R})$. □