Stein's maximal principle

Suppose one has a measure space $X = (X, \mathcal{B}, \mu)$ and a sequence of operators $T_n : L^p(X) \to L^p(X)$ that are bounded on some $L^p(X)$ space, with $1 \le p < \infty$. Suppose that on some dense subclass of functions f in $L^p(X)$ (e.g. continuous compactly supported functions, if the space X is reasonable), one already knows that $T_n f$ converges pointwise almost everywhere to some limit Tf, for another bounded operator $T : L^p(X) \to L^p(X)$ (e.g. T could be the identity operator). What additional ingredient does one need to pass to the limit and conclude that $T_n f$ converges almost everywhere to Tf for *all* f in $L^p(X)$ (and not just for f in a dense subclass)?

One standard way to proceed here is to study the *maximal operator*

$$T_* f(x) := \sup_n |T_n f(x)|$$

and aim to establish a *weak-type maximal inequality*

$$\|T_* f\|_{L^{p,\infty}(X)} \le C \|f\|_{L^p(X)} \qquad (1)$$

for all $f \in L^p(X)$ (or all f in the dense subclass), and some constant C, where $L^{p,\infty}$ is the weak L^p norm

$$\|f\|_{L^{p,\infty}(X)} := \sup_{t>0} t\mu(\{x \in X : |f(x)| \ge t\})^{1/p}.$$

A standard approximation argument using (1) then shows that $T_n f$ will now indeed converge to Tf pointwise almost everywhere for all f in $L^p(X)$, and not just in the dense subclass. See for instance these lecture notes of mine, in which this method is used to deduce the Lebesgue differentiation theorem from the Hardy-Littlewood maximal inequality. This is by now a very standard approach to establishing pointwise almost everywhere convergence theorems, but it is natural to ask whether it is strictly necessary. In particular, is it possible to have a pointwise convergence result $T_n f \mapsto Tf$ without being able to obtain a weak-type maximal inequality of the form (1)?

In the case of *norm* convergence (in which one asks for $T_n f$ to converge to Tf in the L^p norm, rather than in the pointwise almost everywhere sense), the answer is no, thanks to the uniform boundedness principle, which among other things shows that norm convergence is only possible if one has the uniform bound

$$\sup_n \|T_n f\|_{L^p(X)} \le C \|f\|_{L^p(X)} \qquad (2)$$

for some $C > 0$ and all $f \in L^p(X)$; and conversely, if one has the uniform bound, and one has

already established norm convergence of $T_n f$ to Tf on a dense subclass of $L^p(X)$, (2) will extend that norm convergence to all of $L^p(X)$.

Returning to pointwise almost everywhere convergence, the answer in general is "yes". Consider for instance the rank one operators

$$T_n f(x) := 1_{[n,n+1]} \int_0^1 f(y) \, dy$$

from $L^1(\mathbf{R})$ to $L^1(\mathbf{R})$. It is clear that $T_n f$ converges pointwise almost everywhere to zero as $n \to \infty$ for any $f \in L^1(\mathbf{R})$, and the operators T_n are uniformly bounded on $L^1(\mathbf{R})$, but the maximal function T_* does not obey (1). One can modify this example in a number of ways to defeat almost any reasonable conjecture that something like (1) should be necessary for pointwise almost everywhere convergence.

In spite of this, a remarkable observation of Stein, now known as *Stein's maximal principle*, asserts that the maximal inequality *is* necessary to prove pointwise almost everywhere convergence, if one is working on a compact group and the operators T_n are translation invariant, and if the exponent p is at most 2:

> **Theorem 1 (Stein maximal principle)** Let G be a compact group, let X be a homogeneous space of G with a finite Haar measure μ, let $1 \le p \le 2$, and let $T_n : L^p(X) \to L^p(X)$ be a sequence of bounded linear operators commuting with translations, such that $T_n f$ converges pointwise almost everywhere for each $f \in L^p(X)$. Then (1) holds.

This is not quite the most general vesion of the principle; some additional variants and generalisations are given in the original paper of Stein. For instance, one can replace the discrete sequence T_n of operators with a continuous sequence T_t without much difficulty. As a typical application of this principle, we see that Carleson's celebrated theorem that the partial Fourier series $\sum_{n=-N}^{N} \hat{f}(n) e^{2\pi i n x}$ of an $L^2(\mathbf{R}/\mathbf{Z})$ function $f : \mathbf{R}/\mathbf{Z} \to \mathbf{C}$ converge almost everywhere is in fact equivalent to the estimate

$$\| \sup_{N>0} | \sum_{n=-N}^{N} \hat{f}(n) e^{2\pi i n \cdot} | \|_{L^{2,\infty}(\mathbf{R}/\mathbf{Z})} \le C \|f\|_{L^2(\mathbf{R}/\mathbf{Z})}. \qquad (3)$$

And unsurprisingly, most of the proofs of this (difficult) theorem have proceeded by first establishing (3), and Stein's maximal principle strongly suggests that this is the optimal way to try to prove this theorem.

On the other hand, the theorem does fail for $p > 2$, and almost everywhere convergence results in L^p for $p > 2$ can be proven by other methods than weak (p,p) estimates. For instance, the convergence of Bochner-Riesz multipliers in $L^p(\mathbf{R}^n)$ for any n (and for p in the range predicted by the Bochner-Riesz conjecture) was verified for $p > 2$ by Carbery, Rubio de Francia, and Vega, despite the fact that the weak (p,p) of even a *single* Bochner-Riesz multiplier, let alone the maximal function, has still not been completely verified in this range. (Carbery, Rubio de Francia and Vega use weighted L^2 estimates for the maximal Bochner-Riesz operator, rather than L^p type estimates.) For $p \le 2$, though, Stein's principle (after localising to a torus) does

apply, though, and pointwise almost everywhere convergence of Bochner-Riesz means is equivalent to the weak (p, p) estimate (1).

Stein's principle is restricted to compact groups (such as the torus $(\mathbf{R}/\mathbf{Z})^n$ or the rotation group $SO(n)$) and their homogeneous spaces (such as the torus $(\mathbf{R}/\mathbf{Z})^n$ again, or the sphere S^{n-1}). As stated, the principle fails in the noncompact setting; for instance, in $\mathbf{R}$, the convolution operators $T_n f := f * 1_{[n,n+1]}$ are such that $T_n f$ converges pointwise almost everywhere to zero for every $f \in L^1(\mathbf{R}^n)$, but the maximal function is not of weak-type $(1, 1)$. However, in many applications on non-compact domains, the T_n are "localised" enough that one can transfer from a non-compact setting to a compact setting and then apply Stein's principle. For instance, Carleson's theorem on the real line $\mathbf{R}$ is equivalent to Carleson's theorem on the circle $\mathbf{R}/\mathbf{Z}$ (due to the localisation of the Dirichlet kernels), which as discussed before is equivalent to the estimate (3) on the circle, which by a scaling argument is equivalent to the analogous estimate on the real line $\mathbf{R}$.

Stein's argument from his 1961 paper can be viewed nowadays as an application of the probabilistic method; starting with a sequence of increasingly bad counterexamples to the maximal inequality (1), one randomly combines them together to create a single "infinitely bad" counterexample. To make this idea work, Stein employs two basic ideas:

1. The *random rotations (or random translations) trick*. Given a subset E of X of small but positive measure, one can randomly select about $|G|/|E|$ translates $g_i E$ of E that cover most of X.
2. The *random sums trick* Given a collection $f_1, \ldots, f_n : X \to \mathbf{C}$ of signed functions that may possibly cancel each other in a deterministic sum $\sum_{i=1}^n f_i$, one can perform a random sum $\sum_{i=1}^n \pm f_i$ instead to obtain a random function whose magnitude will usually be comparable to the square function $(\sum_{i=1}^n |f_i|^2)^{1/2}$; this can be made rigorous by concentration of measure results, such as Khintchine's inequality.

These ideas have since been used repeatedly in harmonic analysis. For instance, I used the random rotations trick in a recent paper with Jordan Ellenberg and Richard Oberlin on Kakeya-type estimates in finite fields. The random sums trick is by now a standard tool to build various counterexamples to estimates (or to convergence results) in harmonic analysis, for instance being used by Fefferman in his famous paper disproving the boundedness of the ball multiplier on $L^p(\mathbf{R}^n)$ for $p \neq 2, n \geq 2$. Another use of the random sum trick is to show that Theorem 1 fails once $p > 2$; see Stein's original paper for details.

Another use of the random rotations trick, closely related to Theorem 1, is the *Nikishin-Stein factorisation theorem*. Here is Stein's formulation of this theorem:

> **Theorem 2 (Stein factorisation theorem)** Let G be a compact group, let X be a homogeneous space of G with a finite Haar measure μ, let $1 \leq p \leq 2$ and $q > 0$, and let $T : L^p(X) \to L^q(X)$ be a bounded linear operator commuting with translations and obeying the estimate
>
> $$\|Tf\|_{L^q(X)} \leq A\|f\|_{L^p(X)}$$
>
> for all $f \in L^p(X)$ and some $A > 0$. Then T also maps $L^p(X)$ to $L^{p,\infty}(X)$, with
>
> $$\|Tf\|_{L^{p,\infty}(X)} \leq C_{p,q}A\|f\|_{L^p(X)}$$
>
> for all $f \in L^p(X)$, with $C_{p,q}$ depending only on p,q.

This result is trivial with $q \geq p$, but becomes useful when $q < p$. In this regime, the translation invariance allows one to freely "upgrade" a strong-type (p,q) result to a weak-type (p,p) result. In other words, bounded linear operators from $L^p(X)$ to $L^q(X)$ automatically factor through the inclusion $L^{p,\infty}(X) \subset L^q(X)$, which helps explain the name "factorisation theorem". Factorisation theory has been developed further by many authors, including [Maurey](#) and [Pisier](#).

Stein's factorisation theorem (or more precisely, a variant of it) is useful in the theory of Kakeya and restriction theorems in Euclidean space, as first observed [by Bourgain](#).

In 1970, [Nikishin](#) obtained the following generalisation of Stein's factorisation theorem in which the translation-invariance hypothesis can be dropped, at the cost of excluding a set of small measure:

> **Theorem 3 (Nikishin-Stein factorisation theorem)** Let X be a finite measure space, let $1 \leq p \leq 2$ and $q > 0$, and let $T : L^p(X) \to L^q(X)$ be a bounded linear operator commuting with translations and obeying the estimate
>
> $$\|Tf\|_{L^q(X)} \leq A\|f\|_{L^p(X)}$$
>
> for all $f \in L^p(X)$ and some $A > 0$. Then for any $\epsilon > 0$, there exists a subset E of X of measure at most ϵ such that
>
> $$\|Tf\|_{L^{p,\infty}(X \setminus E)} \leq C_{p,q,\epsilon}A\|f\|_{L^p(X)} \qquad (4)$$
>
> for all $f \in L^p(X)$, with $C_{p,q,\epsilon}$ depending only on p,q,ϵ.

One can recover Theorem [2](#) from Theorem [3](#) by an averaging argument to eliminate the exceptional set; we omit the details.

— 1. Sketch of proofs —

We now sketch how Stein's maximal principle is proven. We may normalise $\mu(X) = 1$. Suppose the maximal inequality [(1)](#) fails for any C. Then, for any $A \geq 1$, we can find a non-zero function $f \in L^p(X)$ such that

$$\|T_*f\|_{L^{p,\infty}(X)} \geq A\|f\|_{L^p(X)}.$$

By homogeneity, we can arrange matters so that

$$\mu(E) \geq A^p \|f\|_{L^p(X)}^p,$$

where $E := \{x \in X : |T_* f(x)| \geq 1\}$.

At present, E could be a much smaller set than X: $\mu(E) \ll 1$. But we can amplify E by using the random rotations trick. Let m be a natural number comparable to $1/\mu(E)$, and let $g_1, \ldots, g_m$ be elements of G, chosen uniformly at random. Each element x of X has a probability $1 - (1 - \mu(E))^m \sim 1$ of lying in at least one of the translates $g_1 E, \ldots, g_m E$ of E. From this and the first moment method, we see that with probability ~ 1, the set $g_1 E \cup \ldots \cup g_m E$ has measure ~ 1.

Now form the function $F := \sum_{j=1}^m \epsilon_j \tau_{g_j} f$, where $\tau_{g_j} f(x) := f(g_j^{-1} x)$ is the left-translation of f by g_j, and the $\epsilon_j = \pm 1$ are randomly chosen signs. On the one hand, an application of moment methods (such as the [Paley-Zygmund inequality](#)), one can show that each element x of $g_1 E \cup \ldots \cup g_m E$ will be such that $|T_* F(x)| \gtrsim 1$ with probability ~ 1. On the other hand, an application of Khintchine's inequality shows that with high probability F will have an $L^p(X)$ norm bounded by

$$\lesssim \|(\sum_{j=1}^m |\tau_{g_j} f|^2)^{1/2}\|_{L^p(X)}.$$

Now we crucially use the hypothesis $p \leq 2$ to replace the ℓ^2-summation here by an ℓ^p summation. Interchanging the ℓ^p and L^p norms, we then conclude that with high probability we have

$$\|F\|_{L^p(X)} \lesssim m^{1/p} \|f\|_{L^p(X)} \lesssim 1/A.$$

To summarise, using the probabilistic method, we have constructed (for arbitrarily large A) a function $F = F_A$ whose L^p norm is only $O(1/A)$ in size, but such that $|T_* F(x)| \gtrsim 1$ on a subset of X of measure ~ 1. By sending A rapidly to infinity and taking a suitable combination of these functions F, one can then create a function G in L^p such that $T_* G$ is infinite on a set of positive measure, which contradicts the hypothesis of pointwise almost everywhere convergence.

Stein's factorisation theorem is proven in a similar fashion. For Nikishin's factorisation theorem, the group translation operations τ_{g_j} are no longer available. However, one can substitute for this by using the failure of the hypothesis [(4)](#), which among other things tells us that if one has a number of small sets $E_1, \ldots, E_i$ in X whose total measure is at most ϵ, then we can find another function f_{i+1} of small L^p norm for which $T f_{i+1}$ is large on a set E_{i+1} outside of $E_1 \cup \ldots \cup E_i$. Iterating this observation and choosing all parameters carefully, one can eventually establish the result.

> **Remark 1** A systematic discussion of these and other maximal principles is given in [this book of Guzman](#).
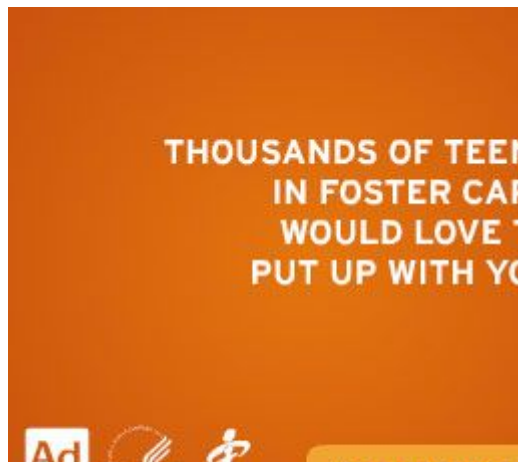
Share this:

Like this: Loading...

8 comments

Comments feed for this article

12 May, 2011 at 10:45 pm

Ming Wang

Dear Professor Tao,
Could you give some useful examples of X included in $\latex R^n$? Since I am not familar
with the concepts compact group and homogeneous space. Thank you very much!

[Some material added in this direction - T.]

    2   0   Rate This
Reply
13 May, 2011 at 10:58 am

shannon7774

In paragraph 4 or 5 you said, "Returning to pointwise almost everywhere convergence, the
answer in general is "no". " Since the question was whether there is a counterexample,
shouldn't the answer here be "yes"?

[Corrected, thanks - T.]

3   0   Rate This

Reply
13 May, 2011 at 2:03 pm

Anonymous

Dear Prof. Tao

In the first paragraph I believe it should read "conclude that T_n f converges"
instead of T f_n

[Corrected, thanks - T.]

0   2   Rate This

Reply
24 May, 2011 at 4:16 pm

Sixth Linkfest

[...] Tao: Stein's maximal principle, Stein's spherical maximal theorem, Locally compact
topological vector [...]

0   0   Rate This

Reply
4 August, 2011 at 4:32 am

Andrew Bailey

Small typo when you talk about Bochner-Riesz multipliers in the second paragraph after the
statement of Theorem 1: "depsite". *[Corrected, thanks - T.]*

0   0   Rate This

Reply
26 August, 2011 at 2:00 am

Anonymous

Dear Prof. Tao,
1)Minor typo: general version instead of "general vesion" in the first line after Theorem 1.

2)I don't understand this sentence:
On the one hand, an application of moment methods (such as the Paley-Zygmund
inequality), one can show that each element $\{x\}$ of $\{g_1 E \cup \ldots \cup g_m E\}$ will
be such that $\{|T_* F(x)| \gtrsim 1\}$ with probability $\{\sim 1\}$.
If someone (Prof Tao?) could give one or two lines on this... Thanks

0   0   Rate This

Reply

26 August, 2011 at 8:16 am

Terence Tao

Thanks for the correction. The Paley-Zygmund inequality can be used to show that if one has a sequence $x_1, \ldots, x_k$ of real numbers with $|x_1| \gtrsim 1$, and one forms the random sum $X := \varepsilon_1 x_1 + \ldots + \varepsilon_k x_k$, then one has $|X| \gtrsim 1$ with probability $\gtrsim 1$ (this is basically because $\mathbf{E}|X|^4 \lesssim (\mathbf{E}|X|^2)^2$, which is a special case of Khintchine's inequality; it also reflects the intuitive fact that random walks cannot converge to the origin). If one applies this fact to the numbers $x_j = T_n \tau_{g_j} f(x)$, where n is chosen to make x_1 large, one obtains the claim.

0   0   Rate This

Reply

28 August, 2011 at 10:11 pm

Anonymous

Thank you very much!

0   0   Rate This

Reply