ON RANDOM MULTILINEAR OPERATOR INEQUALITIES

MICHAEL CHRIST

1. Introduction

A venerable principle holds that the Fourier transform of a measure is “small” in a meaningful sense when linear structure is absent, in certain circumstances. For instance:

1. If \( \mu \) is supported on an appropriately curved submanifold of \( \mathbb{R}^d \), then \( \hat{\mu}(\xi) \to 0 \) at a certain rate as \( |\xi| \to \infty \).

2. If \( \mu_\omega \) is a random measure, with appropriate properties, then for typical \( \omega \), \( \hat{\mu}_\omega(\xi) \) has small supremum norm; or in other contexts, \( \hat{\mu}_\omega(\xi) \) tends to zero at an appropriate rate as \( |\xi| \to \infty \).

3. Let \( p \) be a large prime, and for \( x \in \mathbb{Z}_p \) let \( \mu_p(x) = 1 \) if \( x \) is a quadratic residue modulo \( p \), and \( \mu_p(x) = 0 \) otherwise. Then with a natural normalization of the Fourier transform, \( |\hat{\mu}_p(\xi)| \leq C p^{-1/2} \) for all \( \xi \neq 0 \), whereas \( \hat{\mu}_p(0) \approx 1 \).

Smallness of the Fourier transform may be reformulated in terms of a bilinear expression via the identity \( \|\hat{\mu}\|_\infty = \sup_{f,g \neq 0} \|\int \int f(x)g(y) d\mu(x-y)\|/\|f\|_2\|g\|_2 \). This formulation suggests multilinear extensions, involving e.g. \( \int \int \int f(x)g(y)h(x+y) d\mu(x-y) \). While various possible inequalities can be considered, we are primarily interested in bounds in terms of \( \|f\|_p\|g\|_q\|h\|_r \) with \( p^{-1} + q^{-1} + r^{-1} = 1 \); such quantities scale naturally from the perspective of ergodic theory.

If \( G \) is a finite Abelian group and \( \mu : G \to \mathbb{C} \) has \( \|\hat{\mu}\|_\infty \ll \|\mu\|_1 \), under appropriate normalizations, then \( \mu \) is sometimes said to be uniform [7]. There are higher-order notions of uniformity, due to Gowers [7], which have a multilinear character. Gowers uniformity is closely related to the type of smallness studied in this paper, but here we are dealing with rather singular measures.

In this paper we investigate the extension of this smallness principle to higher-degree multilinear expressions, for natural families of random measures. In §2 we give an example which demonstrates that linear structure is no longer the natural consideration. Indeed, for one of the most canonical (deterministic) examples of all, the natural trilinear extension satisfies no smallness condition, due to the presence of quadratic structure. In §3 we state our main results, which concern two classes of random measures. For one of these classes, our results are quite satisfactory, but for the other they apply only for a certain range of parameters which may not be optimal.

The author is indebted to Patrick LaVictoire for useful discussions.

2. A Nonlinear Obstruction

For convenience, the following example is given in the context of certain finite groups, rather than \( \mathbb{Z} \); there are no essential differences. Let \( d \geq 1 \) and let \( p \in \mathbb{N} \) be any prime. Let \( \mathbb{Z}_p \) be the finite cyclic group \( \mathbb{Z}/p\mathbb{Z} \). Let \( G_p = \mathbb{Z}_p^d \times \mathbb{Z}_p \). For \( x = (x_1, \cdots, x_d) \in \mathbb{Z}_p^d \), we

\[ Date: \text{March 9, 2010. Revised August 18, 2011.} \]
\[ The author was supported in part by NSF grant DMS-0901569. \]
write \(|x|^2 = \sum_{j=1}^{d} x_j^2\). Write \(G_p \ni x = (x', x_{d+1}) \in \mathbb{Z}_p^d \times \mathbb{Z}_p\). Let \(\mu_p\) be the function on \(G_p\) defined by

\[
\mu_p(x', x_{d+1}) = \begin{cases} p^{-d} & \text{if } x_{d+1} = |x'|^2, \\ 0 & \text{otherwise} \end{cases}
\]

and \(\nu_p = \mu_p - m_p\). \(m_p\) satisfies \(\hat{m}_p(\xi) = \mu_p(\xi)\) for \(\xi = 0\), and \(\hat{m}_p(\xi) = 0\) for all \(\xi \neq 0\).

Define the Fourier transform \(\hat{f}(\xi) = \sum_{x \in G_p} f(x) e^{-2\pi i \xi \cdot x/p}\), where \(\xi \cdot x = \sum_{j=1}^{d+1} x_j \xi_j\) and \(\xi \in G_p\). By a well-known identity for Gauss sums,

\[
\max_{\xi \neq 0} |\hat{\mu}_p(\xi)| = p^{-d/2},
\]

and consequently

\[
\max_{\xi} |\hat{\nu}_p(\xi)| = p^{-d/2}.
\]

Therefore by Plancherel’s identity, there is a bilinear inequality

\[
(2.1) \quad |\sum_{x,y \in G_p} f(x) g(y) \nu_p(x - y)| \leq p^{-d/2} \|f\|_2 \|g\|_2
\]

where \(\|f\|_q\) denotes the \(\ell^q(G_p)\) norm.

Does \((2.1)\) extend to a trilinear inequality

\[
(2.2) \quad |\sum_{x,y \in G_p} f(x) g(y) h(x + y) \nu_p(x - y)| \leq C p^{-\rho} \|f\|_2 \|g\|_2 \|h\|_\infty
\]

for some \(\rho = \rho(d) > 0\) independent of \(p\)?

**Observation 2.1.** No inequality of the form \((2.2)\) is valid.

To prove this, set

\[
\begin{align*}
h(x) &= e^{2\pi i |x'|^2/p} \\
f(x) &= e^{2\pi i [x_{d+1}^2 - 2|x'|^2]/p} \\
g(x) &= e^{2\pi i [-x_{d+1}^2 - 2|x'|^2]/p}.
\end{align*}
\]

Then \(f(x)g(y)h(x + y) = e^{2\pi i \Phi(x,y)/p}\) where

\[
\Phi(x,y) = |x' + y'|^2 + x_{d+1} - y_{d+1} - 2|x'|^2 - 2|y'|^2.
\]

For \((x,y)\) in the support of \(\mu_p(x - y)\), \(x_{d+1} - y_{d+1} \equiv |x' - y'|^2\) and consequently

\[
\Phi(x,y) = |x' + y'|^2 + |x' - y'|^2 - 2|x'|^2 - 2|y'|^2 \equiv 0.
\]

Therefore the contribution of \(\mu_p\) to our trilinear form equals

\[
\sum_{x,y \in G_p} f(x) g(y) h(x + y) \mu_p(x - y) = \sum_{x,y \in G_p: x_{d+1} - y_{d+1} = |y' - x'|^2} p^{-d} = p^{d+1},
\]

while

\[
\|f\|_2 \|g\|_2 \|h\|_\infty = (p^{d+1})^{1/2} \cdot (p^{d+1})^{1/2} \cdot 1 = p^{d+1}.
\]

On the other hand,

\[
\sum_{x,y \in G_p} f(x) g(y) h(x + y) m_p(x - y) = p^{-d-1} \sum_{x,y \in G_p} e^{2\pi i \Phi(x,y)/p}.
\]
For fixed $y, x', \Phi((x', t), y)$ takes the form $c(x', y) + t$, and
\[
\sum_{t \in \mathbb{Z}_p} e^{2\pi i (t+c(x', y))/p} = e^{2\pi i c(x', y)/p} \sum_{t \in \mathbb{Z}_p} e^{2\pi i t/p} = 0.
\]
Thus in all,
\[
\sum_{x,y \in G_p} f(x)g(y)h(x+y)\nu_p(x-y) = \sum_{x,y \in G_p} f(x)g(y)h(x+y)\mu_p(x-y) = \|f\|_2\|g\|_2\|h\|_{\infty};
\]
there is no cancellation in the sum.

3. Results

Our setting is the set $\mathbb{Z}$ of all integers, and we will work in terms of norms $L^p(\mathbb{Z}) = \ell^p$. Let $\Omega$ be a probability space, equipped with jointly independent, identically distributed, $\{0, 1\}$-valued selector variables $\{s(\omega, x) : x \in \mathbb{Z}\}$, such that $s(\omega, x) = 1$ with probability $p$ and $= 0$ with probability $1 - p$. Let $N$ be any large positive integer. Let $r(\omega, x) = (Np)^{-1}s(\omega, x) - N^{-1}$ for integers $x \in [-N, N]$, and $r(\omega, x) = 0$ otherwise. Thus $\mathbb{E}_s r(\omega, x) = p(Np)^{-1} - N^{-1} = 0$ for $x \in [-N, N]$.

Let $\{L_j : 0 \leq j \leq M\}$ be $\mathbb{Z}$-linear mappings from $\mathbb{Z}$ to $\mathbb{Z}$. Assume none of the $L_j$ are scalar multiples, over $\mathbb{Q}$, of $(x,y) \mapsto x$, that none are scalar multiples of $(x, y) \mapsto y$, and no $L_i$ is a scalar multiple of $L_j$.

In Theorem 3.1 we study multilinear operators
\[
T_\omega(f, g_1, \cdots, g_M)(x) = \sum_y f(y)r(\omega, L_0(x, y)) \prod_{j=1}^M g_j(L_j(x, y)).
\]
These depend also on $N$, and we are interested in their properties as $N \to \infty$. Define the operator norm
\[
\|T_\omega\|_{op} = \sup_{f,g_1,\cdots,g_M} \|T_\omega(f, g_1, \cdots, g_M)\|_2
\]
where the supremum is taken over all functions satisfying $\|f\|_2 \leq 1$ and $\|g_j\|_{\infty} \leq 1$ for all $j$.

**Theorem 3.1.** Suppose that $M \geq 1$ and $0 \leq \gamma < 2^{-M}$. There exist $\varepsilon > 0$ and $C < \infty$ such that for all $N \geq 1$ and $p \geq N^{-\gamma}$,
\[
\mathbb{E}_\omega \|T_\omega\|_{op} \leq CN^{-\varepsilon}.
\]
The constant $C$ is independent of $N$. We do not know whether the conclusion may hold for a larger range of exponents $\gamma$.

Of course
\[
\mathbb{E}_\omega \sup_{f,g_1,\cdots,g_M} \|T_\omega(f, g_1, \cdots, g_M)\|_2 \geq \sup_{g_1,\cdots,g_M} \mathbb{E}_\omega \sup_f \|T_\omega(f, g_1, \cdots, g_M)\|_2.
\]
The latter quantity is easier to analyze; see Proposition 4.1, which gives a satisfactory bound for all $\gamma < 1$, for all $M$.

An ergodic-theoretic consequence is as follows. Let $T$ be an invertible measure-preserving transformation on a probability space $(X, \mu)$. For each $\omega \in \Omega$, specify a subsequence $(n_k(\omega) : k = 1, 2, \cdots)$ of the natural numbers, as follows. Let $\gamma \in (0, 1)$. Let $\Omega$ be a probability space equipped with a family of jointly independent random variables $\{s_n(\omega) : n \in \mathbb{N}\}$ such that $s_n(\omega) = 1$ with probability $n^{-\gamma}$, and $s_n(\omega) = 0$ otherwise. For each $\omega \in \Omega$,
specify the random subsequence \((n_k(\omega))_{k \in \mathbb{N}}\) to consist of all \(n \in \mathbb{N}\) for which \(s_n(\omega) = 1\), listed in increasing order.

It has been proved \([1],[3],[6]\) that for all \(f_1, \ldots, f_M \in L^\infty(X)\),

\[
(3.3) \quad \lim_{N \to \infty} N^{-1} \sum_{k=1}^{N} f_1(T^k(x)) f_2(T^{2k}(x)) \cdots f_M(T^{Mk}(x)) \text{ exists in } L^1(X, d\mu(x)).
\]

This fundamental result, together with Theorem 3.1, give

**Theorem 3.2.** If \(0 \leq \gamma < 2^{-M+1}\) then for almost every \(\omega \in \Omega\), for all \(f_1, \ldots, f_M \in L^\infty(X)\),

\[
(3.4) \quad \lim_{N \to \infty} N^{-1} \sum_{k=1}^{N} f_1(T^{nk}(x)) f_2(T^{2nk}(x)) \cdots f_M(T^{Mnk}(x)) \text{ exists in } L^1(X, d\mu(x)).
\]

A generalization of Theorem 3.1 is natural, and of interest. \(e_\xi\) will denote the function \(y \mapsto e^{-i\xi y}\). With the above notations, define

\[
T^*_\omega(f, g_1, \ldots, g_M)(x) = \sup_{\xi \in \mathbb{Z}} \left| \sum_y e^{-i\xi y} f(y) r(\omega, L_0(x, y)) \prod_{j=1}^{M} g_j(L_j(x, y)) \right|
= \sup_{\xi \in \mathbb{Z}} \left| T_\omega(e_\xi f, g_1, \ldots, g_M)(x) \right|
\]

Multiplying each function \(g_j(z)\) by a factor \(e^{-i\xi_j z}\), and taking the supremum over all \((\xi, \xi_1, \ldots, \xi_M)\), introduces no additional generality since each \(e^{-i\xi_j L_j(x,y)}\) can be factored as \(e^{ia_j x} e^{-ib_j y}\xi_j\) for appropriate coefficients \(a_j, b_j\).

**Theorem 3.3.** For each \(0 \leq \gamma < 2^{-M-1}\) there exist \(\delta > 0\) and \(C < \infty\) such that for all \(N \geq 1\) and \(p \geq N^{-\gamma}\),

\[
(3.5) \quad \mathbb{E}_\omega \|T^*_\omega\|_{op} \leq CN^{-\delta}.
\]

The case \(M = 0\) has an ergodic-theoretic consequence, for return times of sparse random subsequences.

**Theorem 3.4 (Return Times).** Let \((X, A, \mu, \tau)\) be any dynamical system, such that \(\mu\) is a probability measure and \((X, A, \mu)\) is isomorphic to \([0,1]\) equipped with Lebesgue measure and the Lebesgue \(\sigma\)-algebra. Let \(0 \leq \gamma < \frac{1}{2}\). Let \(\{n_k(\omega)\}\) be a random sequence, constructed as in Theorem 3.2. Let \(p \in (1, \infty)\) and \(q \geq 2\). Then for almost every \(\omega \in \Omega\), the following holds. For each \(f \in L^p(X)\) there exists a subset \(X_0 \subset X\) of full measure such that for every dynamical system \((Y, F, \nu, \sigma)\), every \(g \in L^q(Y)\), and every \(x \in X_0\),

\[
\lim_{N \to \infty} N^{-1} \sum_{k=1}^{N} f(\tau^{n_k(\omega)}(x)) g(\sigma^{n_k(\omega)}(y)) \text{ exist for } \nu\text{-almost every } y \in Y.
\]

Thus far we have considered random variables which depend only on \(L(x,y)\) for some linear function \(L\). Next, we consider analogous results for random matrices \(r_\omega(x,y)\), with all entries mutually independent. Consider jointly independent random selector variables \(s_\omega(x,y)\) for \((x,y) \in [-N, \ldots, N]^2\), satisfying \(s_\omega(x,y) = 1\) with probability \(p\), and \(= 0\) otherwise. Then \(\mathbb{E}(\sum_x s_\omega(x,y)) \asymp Np\) and \(\mathbb{E}(\sum_y s_\omega(x,y)) \asymp Np\). Define \(r_\omega(x,y) = (Np)^{-1}(s_\omega(x,y) - p)\) so that \(\mathbb{E}_\omega r_\omega(x,y) = 0\).
Theorem 3.5. Let $M \geq 2$ and $0 \leq \gamma < 1$. For any $\{L_j : 0 \leq j \leq M\}$ satisfying the hypotheses of Theorem 3.1 and for any $\varepsilon > 0$ there exists $C_{M, \varepsilon} < \infty$ such that for all $N \geq 1$ and all $p \geq N^{-\gamma}$, the multilinear forms

$$T_{\omega}(f_1, \cdots, f_M) = \sum_{x,y} r_{\omega}(x,y) \prod_{j=1}^{M} f_j(L_j(x,y))$$

satisfy

$$E_{\omega} \| T_{\omega} \|_{op} \leq C_{M, \varepsilon} N^\varepsilon N^{-(1-\gamma)/2}.$$ (3.6)

In this formulation, $T_{\omega}(f_1, \cdots, f_M)$ is a complex number, not a function. It is possible to generalize Theorem 3.5 by incorporating factors $e^{-i\xi y}$, with a supremum over all $\xi$, parallel to Theorem 3.3.

The conclusion of Theorem 3.5 fails to hold for $\gamma > 1$. The method of proof of Theorem 3.3 applies only in the restricted range $\gamma < 2^{-(M-2)}$, and with some added complications since the Fourier transform cannot be applied directly. However, our proof for the full range $\gamma < 1$ proceeds along quite different lines, relying on entropy considerations along with large deviations bounds.

4. A Preliminary Bound

The order of quantifiers in Theorem 3.1 is significant. In this preliminary section we discuss a variant in which the supremum in the definition (3.1) of $T_{\omega}$ is taken only over $f$, with $g_1, \cdots, g_M$ fixed. For this variant, and even for a substantial generalization, more complete results can be obtained, by a simpler method.

Generalize $T_{\omega}$ by considering linear operators

$$(4.1) \quad L_{\omega,h}(f)(x) = \sum_y f(y)r(\omega, x - y)h(x, y),$$

where $h \in \ell^\infty(\mathbb{Z}^2)$ is an arbitrary bounded function of two variables. In particular, this includes the case where $h(x, y) = \prod_{j=1}^{M} g_j(L_j(x, y))$, for arbitrary $M$ and $g_j \in \ell^\infty$.

Let $\Omega$, $p$, $N$, $s(\omega, \cdot)$, $r(\omega, x)$ be as in Theorem 3.1. Regard $L_{\omega,h}$ as a linear operator on $L^2([-N, N])$.

Proposition 4.1. For any $\varepsilon > 0$, there exists $C_\varepsilon < \infty$ such that for every $h \in \ell^\infty$,

$$E_{\omega} \| L_{\omega,h} \|_{op} \leq C_\varepsilon N^\varepsilon (Np)^{-1/2} \| h \|_{\ell^\infty}.$$ (4.2)

Proof. Fix $N$. Denote by trace the trace of a self-adjoint linear operator on $\ell^2([-N, N])$. Fix $h$, and write $L_{\omega} = L_{\omega,h}$. Since

$$E_{\omega} \| L_{\omega} \|_{op} \leq \left( E_{\omega} \| L_{\omega} \|_{op} \right)^{1/2q} \leq \left( E_{\omega} \text{trace} \left( (L_{\omega}^* L_{\omega})^q \right) \right)^{1/2q},$$

it therefore suffices to show that for any positive integer $q$,

$$(4.3) \quad E_{\omega} \text{trace} \left( L_{\omega}^* L_{\omega} \right)^q \leq C_q N \cdot (Np)^{-q} \| h \|_{\ell^\infty}^{2q}.$$

Write $\vec{n} = (n_1, \cdots, n_{2q})$ where $n_j \in [-N, N]$ are arbitrary. Define $n_{2q+1} \equiv n_1$. All sums over $\vec{n}$ written below are understood to be taken over all such vectors $\vec{n} \in [-N, N]^{2q}$. We say that $\vec{n}$ is admissible if in the vector

$$\vec{m} = (n_2 - n_1, n_2 - n_3, n_4 - n_3, n_4 - n_5, n_6 - n_5, n_6 - n_7, \cdots, n_{2q} - n_{2q+1}),$$

no integer appears as a coordinate exactly once. We write $\sum_{\vec{n}}^+$ to denote the sum over all admissible $\vec{n}$. 

With this notation, the trace can be expanded in the form
\[
\text{trace} \left( L_\omega^* L_\omega \right)^q = \sum_{\vec{n}} H(\vec{n}) \prod_{i=1}^{2q} r^*(\omega, n_{i+1} - n_i)
\]
where \( r^*(\omega, n_{i+1} - n_i) = r(\omega, n_{i+1} - n_i) \) if \( i \) is odd, and \( = r(\omega, n_i - n_{i+1}) \) if \( i \) is even. Here \( H(\vec{n}) \) is a product of \( 2q \) factors of \( h \), so \( \|H\|_{\ell^\infty} \leq \|h\|_{\ell^\infty}^{2q} \). Moreover,
\[
(4.4) \quad E_\omega \text{trace} \left( L_\omega^* L_\omega \right)^q = \sum_{\vec{n}} H(\vec{n}) E_\omega \prod_{i=1}^{2q} r^*(\omega, n_{i+1} - n_i),
\]
\[
E_\omega \prod_{i=1}^{2q} r^*(\omega, n_{i+1} - n_i) = 0
\]
by independence whenever \( \vec{n} \) is not admissible.

If \( \vec{n} \) is admissible, then the number \( K \) of pairwise distinct coordinates of \( \vec{m}(n) = (n_2 - n_1, n_2 - n_3, \cdots) \) satisfies \( K \leq q \). Fix any \( K \in [1, q] \). The number of \( \vec{m} = (m_1, \cdots, m_{2q}) \in [-2N, 2N]^{2q} \) having exactly \( K \) pairwise distinct coordinates is \( \leq C_q N^K \). The number of such \( \vec{m} \) possessing the additional property that \( m_1 - m_2 + m_3 - m_4 + \cdots = 0 \) is of course no greater. The number of \( \vec{n} \in [-N, N] \) for which \( \vec{m}(n) \) has exactly \( K \) distinct coordinates is therefore \( \leq C_q N^{K+1} \); one additional power of \( N \) arises, because \( \vec{n} \) is determined by \( \vec{m}(n) \) together with \( n_1 \), though not by \( \vec{m}(n) \) alone.

If \( \vec{m}(n) \) has \( K \) pairwise distinct coordinates, then
\[
E_\omega \prod_{i=1}^{2q} |r^*(\omega, n_{i+1} - n_i)| \leq C_q (Np)^{-2q} p^K.
\]
Therefore the total contribution made to (4.4) by all admissible indices \( \vec{n} \) having \( K \) pairwise distinct coordinates is
\[
\leq C_q N \cdot (Np)^{-2q} N^K p^K \leq C_q N \cdot (Np)^q - 2q = C_q N \cdot (Np)^q
\]
since \( Np \geq 1 \) and \( K \leq q \). Summing over all \( K \) gives (4.3). \( \square \)

5. Reduction of degree of multilinearity

The proof of Theorem 3.1 will proceed by descending induction on the degree of multilinearity, \( M \). In this section we set up a simple lemma which implements the inductive step.

It will be useful to reformulate and to modestly generalize the operators \( T_\omega \). Consider a scalar-valued multilinear form
\[
(5.1) \quad T(f_1, \cdots, f_M, \rho) = \sum_{(x,y) \in [-AN,AN]^2} \prod_{j=1}^M f_j(L_j(x,y)) \rho(L_0(x,y))
\]
where \( M \geq 2 \), \( f_j : \mathbb{Z} \to \mathbb{R} \), each \( L_j : \mathbb{Z} \to \mathbb{Q} \) is \( \mathbb{Q} \)-linear, \( L_i \) is not a scalar multiple, over \( \mathbb{Q} \), of \( L_j \) if \( i \neq j \), and \( \rho : \mathbb{Z} \to \mathbb{R} \). We operate under the convention that \( f_j(L_j(x,y)) \) is to be interpreted as 0 whenever \( L_j(x,y) \in \mathbb{Q} \setminus \mathbb{Z} \), and likewise for \( \rho(L_0(x,y)) \). Moreover, all \( f_j \) and \( \rho \) are supported in \([-AN, AN]\). Here \( A \geq 1 \) is any positive integer, which is initially 1 but will increase in a controlled manner with each inductive step. We seek to bound \( T(f_1, \cdots, f_M, \rho) \) by a suitable constant times \( \|f_1\|_2 \|f_2\|_2 \cdots \|f_j\|_\infty \). This suitable constant will depend on \( A \), in a manner which will not be specified. In the application, \( \rho \) will depend on \( \omega \in \Omega \) and will be constructed from \( r(\omega, \cdot) \) in a recursive manner.
By assumption, \(\mathbb{Z}^2 \ni (x, y) \mapsto (L_1(x, y), L_2(x, y))\) is injective, and has range equal to a lattice of rank 2. Make a linear “change of variables” \((x, y) \mapsto (u, v) = \lambda(L_1(x, y), L_2(x, y))\) where \(0 \neq \lambda \in \mathbb{Z}\) is chosen so that \(\lambda L_i(x, y) \in \mathbb{Z}\) for \(i = 1, 2\) for all \((x, y) \in \mathbb{Z}^2\). The range of \(\lambda L_1\) need not be arranged to be all of \(\mathbb{Z}\); set \(f_1 \equiv 0\) at all integers not in this range, and likewise \(f_2 \equiv 0\) at all integers not in the range of \(\lambda L_2\), and for \(j > 2\), \(f_j \equiv 0\) and \(\rho \equiv 0\) at all appropriate points so that \(\mathcal{T}(f_1, \ldots, f_M, \rho)\) may be rewritten as

\[
\mathcal{T}(f_1, \ldots, f_M, \rho) = \sum_{x, y} \rho(L_0(x, y)) f_1(x) f_2(y) \prod_{3 \leq j \leq M} f_j(L_j(x, y));
\]

\(\prod_{j \geq 3} f_j(L_j(x, y))\) is interpreted as 1 if \(M = 2\). The sum is now over \((x, y) \in [-AN, AN]^2\) for a possibly increased value of \(A\). The functions \(f_j\) and linear functionals \(L_j, L_0\) appearing here are not the same as those in (5.1), but the new functionals continue to satisfy all hypotheses, and the new functions \(f_j\) have all \(L^p\) norms equal to the corresponding norms of the old functions \(f_j\).

By Cauchy-Schwarz,

\[
|\mathcal{T}(f_1, \ldots, f_M, \rho)|^2 \leq \|f_1\|_2^2 \sum_{x, y} f_2(y) f_2(y') \rho(L_0(x, y)) \rho(L_0(x, y')) \prod_{j > 2} f_j(L_j(x, y)) f_j(L_j(x, y'))
\]

where \(x, y, y'\) are all restricted to \([-AN, AN]\). Substitute \(y' = y + z\) to reexpress the triple sum as

\[
\sum_z \sum_{x, y} \rho^z(L(x, y)) \prod_{j \geq 2} f_j^z(L_j(x, y))
\]

where \(L_2(x, y) = y, z \in [-2AN, A2N]\),

\[
f_j^z(u) = f_j(u) f_j(u + L_j(0, z)),
\]

and

\[
\rho^z(u) = \rho(u) \rho(u + L_0(0, z)).
\]

Thus

\[
|\mathcal{T}(f_1, \ldots, f_M, \rho)|^2 \leq \|f_1\|_2^2 \sum_z |\mathcal{T}^z(f_2^z, \ldots, f_M^z, \rho^z)|
\]

where

\[
\mathcal{T}^z(f_2^z, \ldots, f_M^z, \rho^z) = \sum_{x, y} \rho^z(L(x, y)) \prod_{j = 2}^M f_j^z(L_j(x, y))
\]

takes the same form as did \(\mathcal{T}\), with the primary change that the number of functions \(f_j\) has been reduced by one.

Now \(f_2^z(L_2(x, y)) \equiv f_2(y) f_2(y + z)\), so

\[
(5.2) \quad \sum_z \|f_2^z\|_2 \leq CN^{1/2} \|f_2\|_2^2
\]

by Cauchy-Schwarz. Likewise since all functions are supported in \([-AN, AN]\),

\[
(5.3) \quad \|f_3^z\|_2 \leq CN^{1/2} \|f_3\|_2^2.
\]

Certain values of the parameter \(z\) are exceptional, and will be treated as follows. By Cauchy-Schwarz,

\[
\|f_2^z\|_1 \leq \|f_2\|_2^2
\]
for all \( z \). Likewise \( \|\rho^z\|_1 \leq \|\rho\|_2^z \). Therefore for any \( z \),
\[
|T^z(f_2, \cdots, f_M, \rho^z)| \leq \|f_2\|_1 \prod_{j > 2} \|f_j\|_\infty \sup_y \sum_x |\rho^z(L_0(x,y))|
\leq \|f_2\|_2^z \prod_{j > 2} \|f_j\|_\infty^2 \sup_y \sum_x |\rho^z(L_0(x,y))|
\leq \|f_2\|_2^z \prod_{j > 2} \|f_j\|_\infty^2 \|\rho^z\|_1
\leq \|f_2\|_2^z \prod_{j > 2} \|f_j\|_\infty^2 \|\rho\|_2^z.
\]

Define
\[
\|T(\rho)\|_{\text{op}} = \sup |T(f_1, \cdots, f_M, \rho)|
\]
where the supremum is taken over all functions satisfying \( \|f_j\|_2 \leq 1 \) for \( j \in \{1, 2\} \) and \( \|f_j\|_\infty \leq 1 \) for \( j > 2 \). Similarly
\[
\|T^z(\rho^z)\|_{\text{op}} = \sup |T(f_2, \cdots, f_M, \rho^z)|
\]
where the supremum is taken over all functions satisfying \( \|f_j\|_2 \leq 1 \) for \( j \in \{2, 3\} \) and \( \|f_j\|_\infty \leq 1 \) for \( j > 3 \).

Write \( |B| \) to denote the cardinality of a set \( B \). We have shown:

**Lemma 5.1.** For any set \( B \subset \mathbb{Z} \),
\[
(5.4) \quad \|T(\rho)\|_{\text{op}} \leq C N^{1/2} \max_{z \notin B} \|T^z(\rho^z)\|_{\text{op}}^{1/2} + |B|^{1/2} \cdot \|\rho\|_2.
\]

**Remark 5.1.** It may be helpful to understand the role of the different terms here, and the question of whether there is any essential loss when Lemma 5.1 is applied. The factors of \( N^{1/2} \) in (5.2) and (5.3) are natural, cannot be improved, and represent no loss. Indeed, when \( \rho(x) \asymp N^{-1} \) for all \( x \in [-N,N] \), \( \|T^z(\rho^z)\|_{\text{op}} = O(N^{-1}) \) for all \( z \), compensating exactly for the leading factor of \( N^{1/2} \) in (5.2); thus this factor does not in and of itself represent any loss. As the support of \( \rho \) becomes sparser, \( \|\rho\|_\infty \) becomes large in order to maintain the normalization \( \|\rho\|_1 \asymp 1 \). Since \( \rho^z \) is a product of two factors of \( \rho \), \( \|\rho^z\|_\infty \) becomes larger for many values of \( z \), essentially by a factor of \( N^2 \|\rho\|_\infty^2 \) relative to the non-sparse averaging case. But for this loss there is also compensation; the support of \( \rho^z(x) = \rho(x)\rho(x+z) \) is, on the average with respect to \( z \), correspondingly smaller than the support of \( \rho \). For natural classes of random probability measures \( \rho_\omega \), a simple back-of-the-envelope calculation gives heuristically that \( N^{1/2} \|T^z(\rho^z)\|_{\text{op}} = O(1) \) for typical \( z,\omega \), provided that the support of \( \rho \) has cardinality \( \gg N^{1/2} \). Thus one may expect to have no essential loss in applying Lemma 5.1 when dealing with random \( \rho \) whose supports are not too small.

However, if the support of \( \rho \) has cardinality \( \ll N \), then \( \rho_z \) will vanish identically for most values of \( z \), and Lemma 5.1 must yield poor bounds. It is this issue which leads to the restrictions on \( \gamma \) in our main theorems.

In our application of Lemma 5.1 \( \rho \) will take the form
\[
(5.5) \quad \rho(x) = \rho_\omega(x) = \prod_{i \in I} r(\omega, x + z_i)
\]
where \( I \) is some finite index set, and it will always be the case that \( z_i \neq z_j \) whenever \( i \neq j \).
Then
\[ \rho^z(x) = \prod_{i \in I} r(\omega, x + z_i) r(\omega, x + z_i + L(0, z)). \]

Here \( z \mapsto L(0, z) \) is injective. In this situation, we define the set \( B \) of exceptional values of the parameter \( z \) to be
\[ (5.6) \quad B = \{ z : \text{there exist } i, j \in I \text{ such that } z_i = z_j + L(0, z) \}. \]

Then
\[ |B| \leq |I|^2, \]

since \( z \mapsto L(0, z) \) is injective. Moreover, if \( z \notin B \), \( \rho^z \) takes the same form as did \( \rho \), with the size of \( I \) increased; \( B \) is defined so that the condition (5.6) is inherited from \( \rho \) by \( \rho^z \).

With this definition of \( B \), then,
\[ (5.7) \quad \|T^z(\rho)\|_{\text{op}} \leq C N^{1/2} \max_{z \notin B} \|T^z(\rho^z)\|_{\text{op}} + C \|\rho\|_2 \]

where \( C \) depends only on \( A, |I| \).

### 6. Proof of Theorem 3.1

Let \( \rho \) be of the form (5.5). Then
\[ \mathbb{E}_\omega(\|\rho_\omega\|_2^2) = \sum_x \mathbb{E}\left( \prod_{i \in I} r(\omega, x + z_i)^2 \right). \]

For each \( x \), the \( |I| \) factors \( r^2(\omega, x + z_i) \) are jointly independent since \( \{z_i\} \) are distinct. By definition,
\[ \mathbb{E}(r(\cdot, x)^2) \leq C(Np)^{-2} p = C N^{-2} p^{-1} \]

for some constant \( C < \infty \). Therefore
\[ \mathbb{E}(\prod_{i \in I} r(\omega, x + z_i)^2) \leq C^{|I|} N^{-2|I|} p^{-|I|} \]

and hence
\[ (6.1) \quad \mathbb{E}_\omega(\|\rho_\omega\|_2^2) \leq C^{|I|} N^{-2|I|} p^{-|I|}. \]

A stronger result will be required. The supremum \( \sup_z \) in the next lemma is taken over all \( |I| \)-tuples \( z = (z_1, \cdots, z_{|I|}) \) satisfying \( z_i \neq z_j \) whenever \( i \neq j \).

**Lemma 6.1.** Let \( \rho_{\omega,z}(x) = \prod_{i \in I} r(\omega, x + z_i) \) where \( z_i \neq z_j \) whenever \( i \neq j \). Then for any \( q < \infty \) there exists \( C_q < \infty \) independent of \( z \), such that for every \( \xi \in \mathbb{T}, \)
\[ (6.2) \quad \mathbb{E}_\omega(|\rho_{\omega,z}(\xi)|^q) \leq C_q \left( N^{-|I|+\frac{1}{2}} p^{-|I|/2} \right)^q. \]

Moreover, for any \( \epsilon > 0 \) there exists \( C_\epsilon < \infty \) such that
\[ (6.3) \quad \mathbb{E}_\omega\left( \sup_z \|\rho_{\omega,z}\|_{L^\infty(\mathbb{T})} \right) \leq C_\epsilon N^{-|I|+\frac{1}{2}+\epsilon} p^{-|I|/2}. \]

The proof will be given below. By Parseval’s theorem, (6.3) implies
\[ (6.4) \quad \mathbb{E}_\omega(\sup_z \|\rho_{\omega,z}\|_2) \leq C_\epsilon C^{|I|} N^{-|I|+\frac{1}{2}+\epsilon} p^{-|I|/2}. \]

We are now in a position to argue by induction on the degree of multilinearity \( M \). Some additional notation is required, because the base case in the induction depends on \( M \). Let \( \mathcal{T}(\omega) = \mathcal{T}_M(\omega) \) be the \( M \)-linear scalar-valued form to be analyzed; thus \( \rho = r \). Define
\(B_{M+1} = \emptyset\), and \(B_M = \{0\} \subseteq \mathbb{Z}^1\). For \(z \notin B_M\) define \(T_{M-1}(\omega, z)\) to be the associated \(M-1\)-linear scalar form, as discussed above. Define \(B_{M-1}\) to be the set of all \((z_1, z_2) \in \mathbb{Z}^2\) such that \(z_1 \notin B_M\) and \(z_2\) does not lie in the finite exceptional set \(B\) associated to \(z_1\) in the above discussion. For \((z_1, z_2) \notin B_{M-1}\) let \(T_{M-2}(\omega, (z_1, z_2))\) be the associated \(M-2\)-linear scalar form. Continue, constructing \(T_{M-k}(\omega, z)\) for \(k = 0, 1, 2, \cdots, M-2\) for (most) \(z \in \mathbb{Z}^k\), and exceptional sets \(B_{M-k} \subseteq \mathbb{Z}^{k+1}\). For each \(z \notin B_{M-k}\), \(\{\zeta : (z, \zeta) \in B_{M-k}\}\) is a finite set whose cardinality is bounded by a constant which depends only on \(k\). By (5.7),

\[
\|T_k(\omega, (z_1, \cdots, z_{M-k}))\|_{op} \leq C N^{1/2} \max_{(z, \zeta) \notin B_k} \|T_{k-1}(\omega, (z_1, \cdots, z_{M-k}, \zeta))\|_{op} + C \|\rho_{\omega, (z_1, \cdots, z_{M-k})}\|_2.
\]

**Lemma 6.2.** Suppose that \(p \geq N^{-1}\). Then for any \(\varepsilon > 0\) and \(k \in \{2, 3, \cdots, M\}\),

\[
E_{\omega} \left( \sup_{z \notin B_{k+1}} \|T_k(\omega, z)\|_{op} \right) \leq C_{\varepsilon} N^{1+\varepsilon} N^{-\frac{2}{M-k}} p^{-\frac{2}{M-k}}.
\]

Specializing this conclusion to \(k = M\) yields the sought-for bound.

**Corollary 6.3.** Provided that \(p \geq N^{-1}\),

\[
E_{\omega} \left( \|T(\omega)\|_{op} \right) \leq C_{\varepsilon} N^{1+\varepsilon} N^{-\frac{2}{M}} p^{-\frac{2}{M}}.
\]

If \(p \geq N^{-\gamma}\) and if \(\gamma < 2^{-\frac{M-2}{2}}\) then

\[
E_{\omega} \left( \|T(\omega)\|_{op} \right) = O(N^{-\delta}) \text{ for any } \delta < \frac{1}{2}(2^{-\frac{M-2}{2}} - \gamma).
\]

**Proof of Lemma 6.2.** We proceed by ascending induction on \(k\). \(T_k(\omega, z)\) is associated to an index set \(I\) of cardinality \(|I| = 2^{M-k}\). In the base case \(k = 2\), \(T_0(\omega, z)\) is the bilinear form associated to a linear operator defined, in appropriate coordinates, by convolution with \(\rho_{\omega, z}\). \(\|T_0(\omega, z)\|_{op}\) is simply the \(L^2(\mathbb{Z}) \rightarrow L^2(\mathbb{Z})\) operator norm of this convolution operator, which is the \(L^\infty\) norm of the Fourier transform \(\widehat{\rho_{\omega, z}}\). Therefore by Lemma 6.1

\[
E_{\omega} \sup_{z \notin B_3} \|T_0(\omega, z)\|_{op} \leq C_{\varepsilon} N^{1+\varepsilon} N^{-\frac{2}{M}} p^{-\frac{2}{M}}
\]

which is the bound stated for \(k = 2\).

For the inductive step,

\[
E_{\omega} \max_{z \notin B_{k+1}} \|T_k(\omega, z)\| \leq C N^{1/2} E_{\omega} \max_{(z, \zeta) \notin B_{k}} \|T_{k-1}(\omega, (z, \zeta))\|_{op}^{1/2} + C E_{\omega} \max_{z \notin B_{k+1}} \|\rho_{\omega, z}\|_2
\]

\[
\leq C_{\varepsilon} N^{1+\varepsilon} N^{-\frac{2}{M-k}} p^{-\frac{2}{M-k}} + C E_{\omega} \max_{z} \|\rho_{\omega, z}\|_2
\]

\[
= C_{\varepsilon} N^{1+\varepsilon} N^{-\frac{2}{M-k}} p^{-\frac{2}{M-k}} + C E_{\omega} \max_{z} \|\rho_{\omega, z}\|_2.
\]

The first term on the final line is of the desired form. By Lemma 6.1

\[
E_{\omega} \max_{z} \|\rho_{\omega, z}\|_2 \leq C_{\varepsilon} N^{1/2} N^{-|I|+\varepsilon} p^{-|I|/2}
\]
where $|I| = 2^{M-k}$. Thus

\[
\mathbb{E}_\omega \max_z \|\rho_{\omega,z}\|_2 \leq C_\varepsilon N^{\frac{1}{2} + \varepsilon} N^{-2^{M-k}} \frac{p}{p - 2^{M-k-1}}
\]

\[
= C_\varepsilon N^{1+\varepsilon} N^{-1/2} N^{-2^{M-k}} \frac{p}{p - 2^{M-k-1}}
\]

\[
\leq C_\varepsilon N^{1+\varepsilon} N^{-2^{1-k}} N^{-2^{M-k}} \frac{p}{p - 2^{M-k-1}}
\]

since $k \geq 2$. This completes the inductive step. \qed

**Proof of Lemma 6.1.** It suffices to treat the case where $q$ is an even positive integer. Thus we may replace $q$ by $2q$. For any $\xi \in \mathbb{T}$,

\[
\mathbb{E}_\omega(\|\rho_{\omega,z}(\xi)\|_{2q}^2) = \mathbb{E}_\omega \left( \sum_{n_1, \ldots, n_{2q}} \prod_{\alpha=1}^{2q} \rho_{\omega,z}(n_\alpha) \prod_{\beta=1}^{2q} \rho_{\omega,z}(n'_\beta) e^{-i\xi(\sum_\alpha n_\alpha - \sum_\beta n'_\beta)} \right)
\]

\[
\leq \sum_{n_1, \ldots, n_{2q}} \mathbb{E}_\omega \left( \prod_{\alpha=1}^{2q} \rho_{\omega,z}(n_\alpha) \right),
\]

where $n_\alpha = n'_\alpha - q$ for $q > \alpha$.

For $m \in \mathbb{Z}$ and $\tilde{n} = (n_1, \ldots, n_{2q}) \in \mathbb{Z}^{2q}$, define $\nu(m, \tilde{n})$ to be the number of indices $(\alpha, i) \in \{1, 2, \ldots, 2q\} \times I$ which satisfy $n_\alpha + z_i = m$. Since $z_i \neq z_j$ whenever $i \neq j$, there can be at most one such pair with a given value of $\alpha$. For fixed $k$, the $|I|$ random variables $\rho_{\omega,k+z_i}$ are jointly independent and $\mathbb{E}_\omega(\rho_{\omega,k+z_i}) = 0$ for each $i$. Therefore

\[
\mathbb{E}_\omega \prod_{\alpha=1}^{2q} \rho_{\omega,z}(n_\alpha) = 0 \text{ unless for every } m \in \mathbb{Z}, \nu(m, \tilde{n}) \neq 1.
\]

We say that $\tilde{n} \in \mathbb{Z}^{2q}$ is negligible if there exists at least one $m \in \mathbb{Z}$ satisfying $\nu(m, \tilde{n}) = 1$. The number of nonnegligible multi-indices $\tilde{n}$ is $\leq C_{q,|I|}(AN)^q$. To prove this, given $\tilde{n}$, partition the indices $1, 2, \ldots, 2q$ into equivalence classes, by saying that $n_\alpha$ is equivalent to $n_\beta$ if there exist indices $i, j$ such that $n_\alpha + z_i = n_\beta + z_j$, and forming the smallest transitive relation $\equiv$ generated by these relations. Each $\tilde{n}$ is thereby associated to a unique equivalence relation on $\{1, 2, \ldots, 2q\}$. The number of such relations is a finite quantity, for each $q$. Consider all $\tilde{n}$ associated to a given relation, with $C$ distinct equivalence classes. $C \leq q$, since each equivalence class contains at least two elements. If $\{\beta\}$ is a collection of indices $\alpha$, with exactly one chosen from each equivalence class, then for every $\alpha \notin \{\beta\}$, $n_\alpha$ is determined from some $n_\beta$ by an equation $n_\alpha + z_j = n_\beta + z_i$. Therefore at most $(AN)^C$ values of $\tilde{n}$ remain undetermined. Therefore there are at most $(AN)^C \leq (AN)^q$ indices $\tilde{n}$ associated to any given equivalence relation.

If $\tilde{n}$ is not negligible then

\[
\mathbb{E}_\omega \prod_{\alpha=1}^{2q} \rho_{\omega,z}(n_\alpha) \leq C_q(Np)^{-2|I|} \prod_{m: \nu(m, \tilde{n}) \geq 2} p = C_q(Np)^{-2|I|} p^{\mu(\tilde{n})}
\]

where $\mu(\tilde{n})$ is the number of $m \in \mathbb{Z}$ satisfying $\nu(m, \tilde{n}) \geq 2$. Plainly $\mu(\tilde{n}) \leq 2q|I|/2 = q|I|$, so

\[
\mathbb{E}_\omega \prod_{\alpha=1}^{2q} \rho_{\omega,z}(n_\alpha) \leq C_q(Np)^{-2|I|} p^{q|I|},
\]
Summing over all nonnegligible \( n \) gives
\[
\mathbb{E}_\omega |\hat{\rho}_{\omega,z}(\xi)|^{2q} \leq C_{q,|I|}(Np)^{-2q|I|p|I|}(AN)^q = C_{q,|I|}A^q N^{(1-2|I|)q/p|I|},
\]
as was to be proved.

To derive (6.3) from (6.2), temporarily fix any \( \xi \in \mathbb{T} \). For any \( q < \infty \),
\[
\left( \mathbb{E}_\omega \sup_z |\hat{\rho}_{\omega,z}(\xi)| \right)^q \leq \mathbb{E}_\omega \left( \sup_z |\hat{\rho}_{\omega,z}(\xi)|^q \right) \leq \mathbb{E}_\omega \left( \sum_z |\hat{\rho}_{\omega,z}(\xi)|^q \right)
\]
\[
= \sum_z \mathbb{E}_\omega (|\hat{\rho}_{\omega,z}(\xi)|^q) \leq C_{A,|I|} N^{|I|} \mathbb{E}_\omega \left( |\hat{\rho}_{\omega,z}(\xi)|^q \right) \leq C_{A,|I|} N^{|I|} C_q C_{|q|q^{-q}|q|} N^{2q/p - |I|/2},
\]
since at most \( C_{A,|I|} N^{|I|} \) values of \( z \) arise. Choosing \( q = |I|/\varepsilon \) yields
\[
(6.9) \quad \mathbb{E}_\omega \left( \sup_w |\hat{\rho}_{\omega,w}(\xi)| \right) \leq C_{\varepsilon,|I|} N^{-|I| + \frac{1}{2} + \varepsilon} p^{-|I|/2}.
\]

This is weaker than (6.4), in which \( |\hat{\rho}_{\omega,z}(\xi)| \) is replaced by \( \|\hat{\rho}_{\omega,z}\|_\infty \). But since \( \rho_{\omega,z} \) is supported on an interval \([-AN, AN]\), by the Shannon sampling theorem
\[
\|\hat{\rho}_{\omega,z}\|_\infty \leq \max_j |\hat{\rho}_{\omega,z}(\xi_j)|
\]
where \( \{\xi_j\} \subset \mathbb{T} \) is an arithmetic progression consisting of \( KAN \) points with spacing \( K^{-1}A^{-1}N^{-1} \), where \( K \) is an absolute constant. Since such a progression consists of \( O(N) \) points, the same reasoning used to introduce the supremum over \( z \) in (6.9) also suffices to introduce the supremum over all \( \xi_j \), at the expense of another factor of \( N^\varepsilon \). Thus (6.3) follows from (6.9). \( \square \)

7. Extensions

In this section we present an extension of Theorem 3.1, then show how Theorem 3.3 is an almost immediate consequence of this extension. Finally, we show how the application to return times of random subsequences is deduced from Theorem 3.3.

For \( K \geq 1 \) let \( S_K \subset \mathbb{Z}^K \) be the set of all \( z = (z_1, \cdots, z_K) \in \mathbb{Z}^K \) satisfying
\[
(7.1) \quad i \neq j \Rightarrow z_i \neq z_j.
\]
Define \( \rho(\omega, z) : \mathbb{Z} \to \mathbb{R} \) by
\[
(7.2) \quad \rho(\omega, z)(x) = N^{K-1} \prod_{i=1}^K r(\omega, x + z_i).
\]
Consider multilinear operators
\[
(7.3) \quad T_{\omega,z}(f_1, \cdots, g_M)(x) = \sum_y f(y) \prod_{j=1}^M g_j(L_j(x, y))
\]
where \( \{L_j\} \) satisfy ... Define
\[
|||T|||_{\text{op}} = \sup_{f, g_1, \cdots, g_M} |||T_{\omega}(f, g_1, \cdots, g_M)|||_2
\]
where the supremum is taken over all \( f, \{g_j\} \) satisfying \( ||f||_2 \leq 1 \) and \( ||g_j||_\infty \leq 1 \). Since
\[
\mathbb{E}_\omega ||\rho(\omega, z)||_1 \approx N^{K-1}(Np)^{-K} \cdot N \cdot p^K \equiv 1,
\]
the factor $N^{K-1}$ in the definition of $\rho(\omega, z)$ is the natural normalization here.

The same analysis as above establishes:

**Theorem 7.1.** There exist $\gamma = \gamma(M, K) > 0$ and $\delta > 0$, $C < \infty$ such that for all $p \geq N^{-\gamma}$,

(7.4) \[ E_\omega \sup_{z \in S_K} \| T_{\omega, z} \|_{op} \leq C N^{-\delta} \]

uniformly for all $N$.

**Proof of Theorem 7.1.** Introduce a function $\xi(x, z)$ so that

\[ |T_{\omega, z}(\xi(\xi(x, z), f, g_1, \cdots, g_M)(x))| \geq \frac{1}{2} \sup_{x, y} |T_{\omega, z}(\xi f, g_1, \cdots, g_M)(x)| \]

for all $x \in \mathbb{Z}$, where $\xi(x, z)$ denotes the function $x \mapsto e^{-\xi(x, z)}$. Now

(7.5) \[ \|T_{\omega, z}(\xi(x, z), f, g_1, \cdots, g_M)(x)\|^2 \]

\[ = \sum_w \sum_{x, y} \rho(\omega, z)(L(x, y)) \rho(\omega, z)(L(x, y) + L(0, w)) e^{-\xi(x, z)w} f_w(y) \prod_j g_{j, w}(L_j(x, y)) \]

with the same notations for $f_w, g_{j, w}$ as in the beginning of the discussion of Theorem 3.1.

Now in (7.5), set $g_0(w) = e^{-\xi(x, z)w}$. Then $\|g_0, w\|\infty = 1$. Regarding this expression as a multilinear form in $f_w, \{g_{j, w} : 0 \leq j \leq M\}$, it is in a form to which Theorem 7.1 applies, yielding the desired bound.

**Sketch of proof of Theorem 3.4.** For $j \in \mathbb{N}$ let $N_j$ be the number of indices $k$ such that $n_k \in [1, 2^{j+1}]$. With probability $\geq 1 - e^{-\delta 2^j}$ for some $\epsilon, \delta > 0$, $N_j \sim 2^{(1-\epsilon)j}$. Moreover, $N_j - N_{j-1} \sim N_j$, with similarly high probability.

We will sketch the proof of a weaker result, namely:

(7.6) \[ \lim_{j \to \infty} N_j^{-1} \sum_{k : n_k \in [1, 2^{j+1}]} f(\sigma_{n_k}(x))g(\sigma_{n_k}(y)) \]

exists, with the same quantifiers as in Theorem 3.4. The only distinction is that we average here only over initial segments $n_k \in [1, 2^j]$, rather than over arbitrary initial segments $n_k \in [1, N]$. The stronger conclusion stated in the theorem is proved by modifying the proof below, as follows: Partition $[2^{j-1}, 2^j]$ into subintervals of lengths $2^{(1-\eta)j}$ for sufficiently small $\eta > 0$, and augment the sequence $(2^j : j \in \mathbb{N})$ in the argument below by adjoining all endpoints of these subintervals. Details are left to the reader.

For $f \in \ell^p = L^p(\mathbb{E})$ and $g \in \ell^q$ introduce the differences

\[ \Delta^\omega_j(f, g)(x, y) = (N_j - N_{j-1})^{-1} \sum_{n_k \in [2^j, 2^{j-1}]} f(x + n_k)g(y + n_k) - 2^{-j} \sum_{n \in [2^j, 2^{j+1}]} f(x + n)g(y + n). \]

We will show momentarily that

(7.7) \[ E_\omega \sup_{x, y} \| \sup_{g} \| \Delta^\omega_j(f, g)(x, y) \|_{\ell^q} \|_{\ell^p} \leq C 2^{-j\delta} \]

for some $\delta > 0$ and $C < \infty$, where the suprema are taken over all $f, g$ satisfying $\|f\|_{\ell^p} \leq 1$ and $\|g\|_{\ell^q} \leq 1$, respectively.

For $g \in \ell^\infty$, the same bound holds with $\delta$ replaced by 0. A corresponding inequality with $\ell^2$ replaced by $\ell^q$ then follows for all $q \in (2, \infty)$, by interpolation between the endpoints $q = 2$ and $q = \infty$. It then follows by transference that a corresponding maximal inequality holds for arbitrary dynamical systems $(X, \tau)$ and $(Y, \sigma)$. This maximal inequality, together
with the almost everywhere existence of the full averages \( (8.3) \) for \( L^\infty \) functions, yields \((7.6)\).

To establish \((7.7)\), write \( g \) in terms of its Fourier transform to represent

\[
\Delta_j^\omega (f, g)(x, y) = c \int_{\xi \in \mathbb{T}} m_{\omega, j}(\xi, x)e^{i\xi y\hat{g}(\xi)}
\]

where \( \mathbb{T} = \mathbb{R}/2\pi\mathbb{Z} \) and

\[
m_{\omega, j}(\xi, x) = N_j^{-1} \sum_{n_k \in (2^j, 2^{j+1}]} f(x + n_k)e^{i\xi n_k}.
\]

Then

\[
\|\Delta_j^\omega (f, g)(x, y)\|_{\ell^p_g} \leq C \sup_{\xi} |m_{\omega, j}(\xi, x)|
\]

for every \( g \) satisfying \( \|g\|_{\ell^p_g} \leq 1 \), uniformly for every \( \omega \).

Now

\[
\sup_{\xi} |m_{\omega, j}(\xi, x)| = |T_{\omega, j}^*(f)(x)|
\]

where \( T_{\omega, j}^* \) is a maximal function of the type treated in Theorem \( 3.3 \) associated to the random set \( \{ n_k \in (2^j, 2^{j+1}) \} \). Recall that this set was specified using independent random selector variables \( s_n(\omega) \), such that \( s_n(\omega) = 1 \) with probability \( \asymp n^{-\gamma} \); for \( n \in (2^j, 2^{j+1}] \) this probability is \( \asymp 2^{-j\gamma} \). Theorem \( 3.3 \) therefore applies, and asserts that \( E_\omega \|T_{\omega, j}^*(f)\|_{\ell^p_g} \leq 2^{-j\delta} \) for a certain \( \delta > 0 \).

\[\square\]

8. A Variant

In this section we discuss the variant in which the random variables are independent for distinct values of \((x, y)\), rather than depending only on some scalar-valued linear functional \( L(x, y) \). Consider jointly independent random selector variables \( s_n(\omega) \) for \((x, y) \in [-N, \cdots, N]^2\), satisfying \( s_n(\omega, x, y) = 1 \) with probability \( p \), and \( = 0 \) otherwise. Then \( E(\sum_n s_n(x, y)) \asymp Np \) and \( E(\sum_n s_n(x, y)) \asymp Np \). Define \( r_\omega(x, y) = (Np)^{-1}(s_n(x, y) - p) \) so that \( E_\omega r_\omega(x, y) = 0 \). Let \( T_\omega \) be the associated multilinear operators. We will sometimes write \( T_\omega^{(M)} \) to indicate the degree of multilinearity of \( T_\omega \).

The factor \((Np)^{-1}\) in the definition of \( r_\omega \) represents the natural normalization, so that the expected value of the norm of the linear operator \( f \mapsto \sum_y |r_\omega(x, y)|f(y) \), on \( L^2([-N, N]^2) \), is uniformly bounded. More precisely:

**Lemma 8.1.** For any \( M \geq 2 \) \( A < \infty \), \( \gamma_0 \in [0, 1) \), and family \( \{ L_j \} \) satisfying the hypotheses of Theorem \( 8.3 \) there exists \( C < \infty \) such that for any \( \gamma \in [0, \gamma_0] \) and any index \( i \in \{1, 2, \cdots, M\} \), for any \( N \geq 1 \),

\[
E_\omega \left( \sup_{f_1, \cdots, f_M} \sum_{(x, y) \in [-AN, AN]^2^2} |r_\omega(x, y)| \prod_{j=1}^M |f_j(L_j(x, y))|^2 \right) \leq C_{A, \gamma} \log(N)^2,
\]

where the supremum is taken over all functions satisfying \( \|f_i\|_1 \leq 1 \) and \( \|f_j\|_\infty \leq 1 \) for all \( j \neq i \).
An application of Chernoff’s inequality (see below for a similar argument) yields
\[ \sup_{x \in [-AN, AN]} \sum_y |r_\omega(x, y)| \leq C \log(2 + AN), \]
and the same for the expectation of the square.

Proof. Let \( \lambda \) denote the characteristic function \( \chi_{E_\lambda} \) of a set \( E_\lambda \subset [-N, N] \). We will simplify notation by writing \( T_\omega(E_1, \cdots, E_M) \) for \( T_\omega(\chi_{E_1}, \cdots, \chi_{E_M}) \). Introduce the restricted weak type norm
\[ \|T\|_{\text{weak}} = \sup_{E_1, \cdots, E_M} |E_1|^{-1/2} |E_2|^{-1/2} |T(E_1, \cdots, E_M)|. \] (8.1)

Suppose now that the theorem has been proved for \( M - 1 \). Therefore for \( E_M = [-N, N] \),
\[ \sup_{E_1, \cdots, E_{M-1}} |E_1|^{-1/2} |E_2|^{-1/2} |T_\omega(E_1, \cdots, E_{M-1}, [-N, N])| \leq C N^{\varepsilon (1-\eta)/2} \] (8.2)
where \( C \) depends on \( \varepsilon, M, \gamma \).

Lemma 8.2. For any \( \varepsilon > 0 \), \( E_\omega \|T_\omega^2\|_{\text{op}} \leq C_\varepsilon N^{\varepsilon N (1-\gamma)/2} \).

We turn to the proof of Theorem 3.5. Let \( M \geq 3 \), and \( \gamma \in [0, 1) \). By a simple interpolation, it suffices to prove the inequality under the assumption that each function \( f_j \) equals the characteristic function \( \chi_{E_j} \) of a set \( E_j \subset [-N, N] \). We will simplify notation by writing \( T_\omega(E_1, \cdots, E_M) \) for \( T_\omega(\chi_{E_1}, \cdots, \chi_{E_M}) \). Introduce the restricted weak type norm
\[ \|T\|_{\text{weak}} = \sup_{E_1, \cdots, E_M} |E_1|^{-1/2} |E_2|^{-1/2} |T(E_1, \cdots, E_M)|. \] (8.1)

Theorem 3.5 will be proved by induction on the degree \( M \) of multilinearity. The following base result will be proved later.

Lemma 8.3. For any \( \eta \in (0, 1) \) and for any \( \omega \),
\[ \|T_\omega^M\|_{\text{weak}} \leq C N^{-\eta/2} + \|T_\omega^{M-1}\|_{\text{op}} + \sup_{\{E_m\}} |E_1|^{-1/2} |E_2|^{-1/2} |T_\omega^M(E_1, \cdots, E_M)| \] (8.3)
where \( \sup_{\{E_m\}} \) denotes the supremum over all \( M \)-tuples of sets \( E_j \) satisfying
\[ |E_1| \cdot |E_2| \geq N^{2-\eta}. \]

Proof. Denote by \( \mathbf{1} \) the constant function \( \mathbf{1}(x) = 1 \) for all \( x \in [-N, N] \). Define the nonrandom averaging forms
\[ \mathcal{A}(f_1, \cdots, f_M) = N^{-1} \sum_{x, y} \prod_{j=1}^M f_j(L_j(x, y)). \]

As for \( T \), write \( \mathcal{A}(E_1, \cdots, E_M) \) when each \( f_j \) is the characteristic function of a set \( E_j \). Then
\[ |\mathcal{A}(E_1, \cdots, E_M)| \leq C |E_1|^{1/2} |E_2|^{1/2} = \|f_1\|_2 \|f_2\|_2 \prod_{k>2} \|f_k\|_\infty, \]
but the trivial bound
\[ \mathcal{A}(E_1, \cdots, E_M) \leq \mathcal{A}(E_1, E_2, 1, \cdots, 1) \leq CN^{-1} |E_1| \cdot |E_2| = C (\leq 1/2) |E_2|^{1/2} / |E_1|^{1/2} / |E_2|^{1/2} \] (8.4)
expresses a significant improvement unless $|E_1| \cdot |E_2| \asymp N$.

$\mathcal{A}(f_1, \ldots, f_M)$ never decreases if all functions are replaced by their absolute values; nor does it decrease if some $f_j$ increases, provided that all $f_i$ are nonnegative. The same holds for $(\mathcal{T}_\omega - \mathcal{A})(f_1, \ldots, f_M) = (Np)^{-1} \sum_{x,y} s_\omega(x,y) \prod_j f_j(L_j(x,y))$. Therefore if $\|f_j\|_\infty \leq 1$ for all $j \notin \{1, 2\}$, then

$$|\mathcal{T}_\omega(f_1, \ldots, f_M)| \leq |\mathcal{A}(f_1, \ldots, f_M)| + |(\mathcal{T}_\omega - \mathcal{A})(f_1, \ldots, f_M)|$$

$$\leq \mathcal{A}(|f_1|, \ldots, |f_M|) + (\mathcal{T}_\omega - \mathcal{A})(|f_1|, |f_2|, 1, \ldots, 1)$$

$$\leq 2\mathcal{A}(|f_1|, |f_2|, 1, \ldots, 1) + \mathcal{T}_\omega(|f_1|, |f_2|, 1, \ldots, 1).$$

Write 1 to denote the characteristic function of $[-N, N]$, as well as this set itself. Let $\rho < (1 - \gamma)/2$. Then by induction,

$$\mathbb{E}_\omega \sup_{\{E_m: m \leq M-1\}} |E_1|^{-1/2} |E_2|^{-1/2} |\mathcal{T}_\omega(E_1, \ldots, E_{M-1}, 1)| \leq C N^{-\rho}.$$ 

Therefore by replacing $E_M$ by its complement $[-AN, AN] \setminus E_M$ if necessary, we may assume without loss of generality that

$$\sum_{x,y} \prod_{j=1}^M \chi_{E_j}(L_j(x,y)) \geq \frac{1}{2} \sum_{x,y} \prod_{j=1}^{M-1} \chi_{E_j}(L_j(x,y)).$$

Applying this argument to the indices $m = M-1, M-2, \ldots$ in sequence, we reduce to the case where the set

$$\mathcal{E} = \{(x,y) : L_j(x,y) \in E_j \text{ for all } j \in [1,M]\} \subset E_1 \times E_2$$

satisfies

$$|E_1| \cdot |E_2| \leq 2^M |\mathcal{E}|.$$  

For any set $\mathcal{E}$ consider the random variable

$$X_{\mathcal{E}}(\omega) = Np \sum_{(x,y) \in \mathcal{E}} r_\omega(x,y) = \sum_{(x,y) \in \mathcal{E}} (s_\omega(x,y) - p).$$

$\mathbb{E}_\omega X_{\mathcal{E}}(\omega) = 0$. The summands $s_\omega(x,y) - p$ are jointly independent, with values in $[-1,1]$. $X_{\mathcal{E}}$ has standard deviation $\sigma \asymp p^2 |\mathcal{E}|^{1/2}$, with implicit constants depending on $\gamma_0$ but not on $N$.

Chernoff’s inequality \cite{7} asserts that $\Pr (|X_{\mathcal{E}}(\omega)| > \lambda \sigma) \leq C e^{-c \min(\lambda^2, \lambda \sigma)}$. Set

$$\lambda = Np \cdot N^{-\rho} |\mathcal{E}|^{1/2} \sigma^{-1} \asymp N^{1-\rho} p^{1/2} \asymp N^{1-\rho-\gamma/2}.$$ 

Then

$$\min(\lambda^2, \lambda \sigma) = \min(N^{2-2\rho-\gamma}, N^{1-\rho-\gamma}) |\mathcal{E}|^{1/2} \geq c \min(N^{2-2\rho-\gamma}, N^{2-\rho-\gamma-\nu/2}).$$

Moreover

$$(Np)^{-1} \lambda \sigma = N^{-\rho} |\mathcal{E}|^{1/2} \leq C_M N^{-\rho} |E_1|^{1/2} |E_2|^{1/2}.$$ 

Consider the exceptional event

$$\Omega_{M}^*(\mathcal{E}) = \{\omega \in \Omega : |X_{\mathcal{E}}(\omega)| > \lambda \sigma\}.$$ 

By the definition of $\lambda$,

$$|\mathcal{T}_\omega(E_1, \ldots, E_M)| \leq N^{-\rho} |E_1|^{1/2} |E_2|^{1/2} \text{ for all } \omega \notin \Omega_{M}^*(\mathcal{E}).$$
Choose $\nu = 2\rho$. Since $\rho < (1 - \gamma)/2$, we conclude that
\begin{equation}
\Pr(\Omega^*_M(\mathcal{E})) \leq Ce^{-cN^{1+\delta}}
\end{equation}
for some $\delta > 0$.
Define
\[\Omega^*_M = \Omega^*_{M-1} \cup \cup_{E} \Omega^*_M(\mathcal{E}).\]
The total number of sets $\mathcal{E}$, of all cardinalities, is at most $2^{CMN}$, because $\mathcal{E}$ is uniquely determined by $E_1 \times \cdots \times E_M$. So
\[\Pr(\cup_{E} \Omega^*_M(\mathcal{E})) \leq C2^{CMN}e^{-cN^{1+\delta}},\]
and consequently $\Pr(\Omega^*_M) \leq CN^{-\delta}$ for another $\delta > 0$. It follows from Lemma 8.1 and Hölder’s inequality that
\begin{equation}
\int_{\Omega^*_M} \|T^{(M)}\|_{op} \, d\omega \leq Ce^{-cN^{1+\delta}}
\end{equation}
for some $C, c, \delta \in \mathbb{R}^+$. Since $\eta = 2\rho$, (8.12), and (8.10), and Lemma 8.3 together give $\mathbb{E}_\omega(\|T^{(M)}\|_{weak}) \leq CN^{-\rho}$. This completes the inductive step. \[\square\]

Proof of Lemma 8.4. For any linear operator $Tf(x) = \sum_y K(x,y)f(y)$,
\[\|T\|_{op}^{2K} \leq \text{trace} \left((T^{*}T)^{K}\right)\]
where the sum is taken over all $2K$-tuples $(x_1, y_1, \cdots, x_K, y_K)$. Apply this with $K(x,y) = K_\omega(x,y) = (Np)r_\omega(x,y)$. Fix $(x_1, y_1, x_2, y_2, \cdots, x_K, y_K)$. Define the multiplicity of $(s,t)$ to be the number of factors $K_\omega(x_i, y_j)$ in this product for which $(x_i, y_j) = (s,t)$; here $j = i$, or $j = i - 1$, or $j = K - 1$ and $i = 1$.

To $(x_1, y_1, x_2, y_2, \cdots, x_K, y_K)$ is associated a nonincreasing partition of $2K$, namely the ordered tuple of all nonzero multiplicities of elements $(s,t)$ of $\mathbb{Z}^2$, written in nonincreasing order. We denote such a partition by $(m_1, \cdots, m_j)$, where $\sum_{j=1}^{J} m_j = 2K$.

The expectation of $K_\omega(x_1, y_1)K_\omega(x_2, y_1)K_\omega(x_2, y_2)\cdots K_\omega(x_1, y_K)$ vanishes unless no $(s,t) \in [1, \cdots, N]^2$ has multiplicity equal to one. Therefore only partitions with all $m_j \geq 2$ contribute to the expectation. The number $J$ of summands $m_j$ is then $\leq K$.

Lemma 8.4. The number of points $(x_1, y_1, x_2, y_2, \cdots, x_K, y_K) \in [1, N]^{2K}$ which give rise to any particular partition $(m_1, \cdots, m_j)$ is $\leq C_K N^{J+1}$.

This will be proved below.

The number of possible partitions is a function of $2K$.
\[\mathbb{E}_\omega \left( K_\omega(x_1, y_1)K_\omega(x_2, y_1)K_\omega(x_2, y_2)\cdots K_\omega(x_1, y_K) \right) \leq C_K \prod_j p = C^K p^J.\]
The product $C^K p^J \cdot N^{J+1}$ is $\leq C_K N(Np)^K$ since $J \leq K$. Summing these upper bounds for expected values over all $(x_1, y_1, x_2, y_2, \cdots, x_K, y_K)$ associated to a given partition, then summing over all partitions, yields
\begin{equation}
\mathbb{E}_\omega \text{trace} \left((T^{*}T)^{K}\right) \leq C_K N(Np)^K,
\end{equation}
whence $T_\omega = T^{(2)}_\omega$ satisfies $\mathbb{E}_\omega\|T_\omega\|_{op} \leq C_K (Np)^{-1/2} N^{1/2} (Np)^{1/2} = C_K N^{1/2} (Np)^{-1/2}$.
Since $K$ may be taken to be arbitrarily large, this establishes Lemma 8.4. \[\square\]
Proof of Lemma 3.4. Write $x_{K+1} = x_1$ to facilitate the discussion. If $J = 1$ then $(x_1, y_1) = (x_i, y_i)$ for all $2 \leq i \leq K$, and there are $N^2 = N^{J+1}$ possible values of $(x_1, y_1)$. If $J > 1$, set $z_1 = (x_1, y_1)$, $z_2 = (x_2, y_1)$, $z_3 = (x_2, y_2)$, $z_4 = (x_3, y_2)$, \ldots $z_{2K-1} = (x_K, y_K)$, $z_{2K} = (x_1, y_K)$. To the partition $(m_1, m_2, \ldots, m_J)$ of $2K$ we associate all possible equivalence relations on $\{z_k : 1 \leq k \leq 2K\}$ such that there are $J$ equivalence classes, with $m_1, m_2, \ldots, m_J$ elements. Such an equivalence relation is said to be feasible if there exist values of the $z_k \in [1, N]$ such that $z_l = z_k$ if and only if $z_l, z_k$ belong to the same equivalence class. The number of equivalence relations is a function of $K$ alone, so it suffices to bound the number of points $(x_1, y_1, x_2, y_2, \ldots, x_K, y_K) \in [1, N]^{2K}$ which give rise to one equivalence relation.

Consider any feasible equivalence relation associated to the partition $(m_1, \ldots, m_J)$. Choose some equivalence class with $m_1$ elements $z_k$. Choose two coordinates, $x_i$ and $y_i$ or $x_{i+1}$ and $y_i$, which determine all $z_k$ in this class. These coordinates are said to be free, while any $x_l$ or $y_l$ which is one of the two coordinates of some $z_k$ in this class, is said to be bound. Thus the first equivalence class accounts for exactly two free coordinates.

There must exist either $z_k = (x_i, y_i)$ in this class such that $(x_{i+1}, y_i)$ does not belong to this class, or $z_k = (x_{i+1}, y_i)$ such that $(x_{i+1}, y_i)$ does not belong to this class; otherwise the class would include every $z_k$, which is impossible since $J > 1$. In the first case, $(x_{i+1}, y_i)$ belongs to a second equivalence class. The coordinates of any other $z_k$ in this second class are determined by $x_{i+1}, y_i$. $y_i$ is a coordinate of some element of the first class. $x_{i+1}$ cannot be a coordinate of some element of the first class, since $(x_{i+1}, y_i)$ would belong to that class. Designate $x_{i+1}$ to be a free coordinate, all coordinates of all other $z_l$ in the second class are determined by $x_{i+1}$ and $y_i$, hence by $x_{i+1}$ together with the two free coordinates associated to the first class. Thus three free coordinates (together with the equivalence relation itself) are required to determine all coordinates of all points in the union of the first two classes. Repeating this reasoning, we obtain if $J > 2$ a third class and one additional free coordinate, and so on. Proceeding through all $J$ classes, a total of $J + 1$ free coordinates are obtained. Each of these coordinates can take on $N$ values, so in all there are $N^{J+1}$ possible points associated to an individual equivalence relation associated to a partition with $J$ elements. □

References

[1] T. Austin, Pleasant extensions retaining algebraic structure. arXiv:0905.0518
[2] J. Bourgain, Temps de retour pour les systèmes dynamiques, C. R. Acad. Sci. Paris Sér. I Math. 306 (1988), 483 – 485.
[3] ______, Pointwise ergodic theorems for arithmetic sets, with appendix “Appendix on return-time sequences” by J. Bourgain, H. Furstenberg, Y. Katznelson, and D. S. Ornstein, Inst. Hautes ’Etudes Sci. Publ. Math. 69 (1989), 5 – 45.
[4] C. Demeter, M. Lacey, T. Tao, and C. Thiele, Breaking the duality in the return times theorem, Duke Math. J. 143 (2008), no. 2, 281–355.
[5] B. Host and B. Kra, Nonconventional ergodic averages and nilmanifolds. (English summary) Ann. of Math. (2) 161 (2005), no. 1, 397–488.
[6] T. Tao, Norm convergence of multiple ergodic averages for commuting transformations. (English summary) Ergodic Theory Dynam. Systems 28 (2008), no. 2, 657–688.
[7] T. Tao and V. Vu, Additive Combinatorics, Cambridge University Press, 2010.

Michael Christ, Department of Mathematics, University of California, Berkeley, CA 94720-3840, USA
E-mail address: mchrist@math.berkeley.edu