ORTHONORMAL SYSTEMS IN LINEAR SPANS

ALLISON LEWKO AND MARK LEWKO

We show that any $N$-dimensional linear subspace of $L^2(\mathbb{T})$ admits an orthonormal system such that the $L^2$ norm of the square variation operator $V^2$ is as small as possible. When applied to the span of the trigonometric system, we obtain an orthonormal system of trigonometric polynomials with a $V^2$ operator that is considerably smaller than the associated operator for the trigonometric system itself.

1. Introduction

Let $(\mathbb{T}, \mathcal{B}, \mu)$ denote a probability space and $\Phi := \{\phi_n\}_{n=1}^N$ an orthonormal system (ONS) of ($\mu$-measurable) functions from $\mathbb{T}$ to $\mathbb{R}$. Motivated by questions regarding almost everywhere convergence, one is often interested in the behavior of the maximal function

$$Mf := \max_{\ell \leq N} \left| \sum_{n=1}^{\ell} a_n \phi_n \right|.$$ 

Here we let $f := \sum_{n=2}^{N} a_n \phi_n$. For an arbitrary ONS, the Rademacher–Menshov theorem states that $\|Mf\|_{L^2} \ll \log(N)\|f\|_{L^2}$, where the $\log(N)$ factor is known to be sharp. However, one can do much better for many classical systems; for instance one can replace $\log(N)$ with an absolute constant in the case of the trigonometric system (the Carleson–Hunt inequality). More recently, there has been interest in variational refinements of these maximal results. Define the $r$-th variation operator by

$$V^r f := \left( \max_{\pi \in \mathcal{P}_N} \sum_{I \in \pi} \left| \sum_{n \in I} a_n \phi_n \right|^r \right)^{1/r},$$

where $\mathcal{P}_N$ denotes the set of partitions of $[N]$ into subintervals. Clearly, $|Mf| \leq |V^r f|$ for all $r < \infty$. In the case of the trigonometric system, strengthening the Carleson–Hunt theorem, Oberlin, Seeger, Tao, Thiele, and Wright [Oberlin et al. 2012] have shown that $\|V^r f\|_{L^2} \ll \|f\|_{L^2}$ for $r > 2$. When $r = 2$, it has been shown that $\|V^2 f\|_{L^2} \ll \sqrt{\log(N)}\|f\|_{L^2}$ [Lewko and Lewko 2012a], where the factor $\sqrt{\log(N)}$ is optimal. This later inequality has some applications to sieve theory [Lewko and Lewko 2012c]. The factor $\sqrt{\log(n)}$ is rather unfortunate, leading to inefficiencies in these applications. It is likely that this factor can be improved for the functions arising in the applications, for instance, if the Fourier support of $f$ is contained in certain arithmetic sets. This is a potential route towards improving the estimates

The first author was supported by a Microsoft Research PhD Fellowship. The second author was supported by an NSF postdoctoral fellowship, DMS-1204206.

MSC2010: 42A61, 42B05, 42C05.

Keywords: orthogonal systems, square variation, maximal operator, Fourier analysis.
in [Lewko and Lewko 2012c]. Some results in this direction can be found in section 7 of [Lewko and Lewko 2012a].

In a different direction, it seems that the $\sqrt{\log(n)}$ factor might also be an eccentricity of the standard ordering of the trigonometric system. In [Lewko and Lewko 2012a] the following problem was posed.

**Problem 1.** Is there a permutation $\sigma : [N] \to [N]$ such that the reordering of the trigonometric system

$$\Phi := \{\phi_n = e(\sigma(n)x)\} \quad (\text{where } e(x) := e^{2\pi i x})$$

satisfies

$$\|\gamma^2 f\|_{L^2} \ll o(\sqrt{\log(N)})\|f\|_{L^2}$$

for all $f$ in the span of the system$^1$?

This problem can be thought of as a variational variant of Garsia’s conjecture. A longstanding problem in the theory of orthonormal systems, often called Kolmogorov’s rearrangement problem, asks if every (infinite) ONS can be reordered such that the expansion of every $L^2$ function converges almost everywhere. Garsia’s conjecture is the stronger assertion (see [Garsia 1970] for a proof of this implication) that any finite ONS can be reordered to satisfy $\|Mf\|_{L^2} \ll \|f\|_{L^2}$ where the implicit constant is absolute. Towards Garsia’s conjecture, Bourgain [1989] proved that one can rearrange a uniformly bounded ONS such that $\|Mf\|_{L^2} \ll \log \log(N)\|f\|_{L^2}$. His proof proceeds by showing that this holds for a uniformly randomly selected permutation with high probability. Unfortunately this is the best estimate one can obtain from a purely probabilistic approach. Bourgain showed that if one is allowed to select a new ONS with the same span as $\Phi$ (which allows more freedom than just reordering the system), one can obtain $\|Mf\|_{L^2} \ll \|f\|_{L^2}$ for the new system with the same span.

In this paper, we will study the analogous linear span version of Problem 1. Given an ONS $\Phi := \{\phi_n(x)\}_{n=1}^N$ and an $N \times N$ orthogonal matrix $O = \{o_{i, n}\}_{\substack{1 \leq i \leq N \\leq N}}$, we define a new ONS, $\Psi := \{\psi_n(x)\}_{n=1}^N$, by

$$\psi_n(x) := \sum_{i=1}^N o_{i, n}\phi_i(x).$$

This new system will span the same space as the original system. Conversely, every such ONS can be obtained from some element of the orthogonal group, $\mathcal{O}(N)$. Let us write $\Phi(O) := \Psi$. Furthermore, in what follows $Q$ will denote a measurable subset of $\mathcal{O}(N)$ and $\mathbb{P}[Q]$ will denote the Haar measure of $Q$.

**Theorem 2.** Given an $N$-dimensional subspace of $L^2(\mathbb{T})$, there exists an ONS $\Psi$ that satisfies

$$\|\gamma^2 f\|_{L^2} \ll \sqrt{\log \log(N)}\|f\|_{L^2}$$

for all $f$ in the span. In fact, if we take an arbitrary basis $\Phi$ for $F$, the conclusion holds for the ONS $\Phi(O)$ for all $O \in Q$ for some $Q \subset \mathcal{O}(N)$ with $\mathbb{P}[Q] \geq 1 - Ce^{-cN^{2/5}}$ (for some absolute positive constants $C, c$).

---

$^1$We have recently proved [Lewko and Lewko 2012b] that there exists a rearrangement such that $\|\gamma^2 f\|_{L^2} \ll \epsilon \log^{9/22+\epsilon}(N)\|f\|_{L^2}$ for $\epsilon > 0$ (for all uniformly bounded ONS). This is likely far from best possible.
If we take $\Phi := \{e(nx)\}_{n=1}^N$ (on the circle with the Lebesgue measure), this produces an ONS of trigonometric polynomials (spanning the same space as the first $N$ elements of the trigonometric system) with much smaller square variation than the trigonometric system. Strictly speaking, Theorem 2 is stated for real valued ONS, but the result for the trigonometric system can be obtained by splitting into real and imaginary parts and noting that the corresponding result holds on each with large probability. We note that Problem 1 asks for a similar conclusion where $O$ is restricted to be a permutation matrix instead of just an orthogonal matrix.

Theorem 2 is sharp. Consider an ONS of independent, mean zero, variance on Gaussians, $\{g_i\}_{i=1}^N$. Notice that applying an orthogonal transformation to this system leaves it metrically unchanged. On the other hand, we have (almost surely) that $\max_{\pi \in \mathfrak{S}_N} \sum_{i \in \pi} |\sum_{n \in i} g_n|^2 \sim 2N \log \log(N)$ from the variational law of the iterated logarithm [Lewko and Lewko 2011].

Let us briefly outline the key idea in the proof of Theorem 2. In [Lewko and Lewko 2012a], we proved an estimate of the form (1) for systems of bounded independent random variables; see Theorem 9. The key ingredient in that case is that for every $f$ in the span of the system we have the subgaussian tail estimate $\|f\|_q \ll \|f\|_{L^2}$ (where $\|\cdot\|_q$ is the Orlicz space norm associated to $e^{x^2} - 1$). This clearly cannot hold in the setting of Theorem 2, since any $L^2$ function can be in the span of the system. However, we will show that a function $f$ in the span of a generic basis $\Phi(O)$ can be split as $f = G + E$, where $G$ satisfies a subgaussian tail inequality and $E$ has small $L^2$ norm (decreasing with the size of the Fourier support of $f$). More precisely, we will prove the following (note that we abuse the notation $c$ below to denote multiple distinct constants):

**Proposition 3.** For $N$ fixed, let $\Phi = \{\phi_n(x)\}_{n=1}^N$ be an ONS such that $\sum_{n=1}^N |\phi_n(x)|^2 \leq N$ holds (pointwise). There exists $Q \subset \mathfrak{C}(N)$ with $\mathbb{P}[Q] \geq 1 - C e^{-cN^{2/3}}$ such that for $O \in Q$, we have that the associated ONS $\Phi(O) = \{\psi_n\}_{n=1}^N$ satisfies the following property. For any $f = \sum a_n \psi_n$, letting $m$ denote support($\{a_i\}$) (the number of nonzero $a_i$ values), we have that the function defined by

$$f := \sum a_n \psi_n(x)$$

can be decomposed as $f := G + E$ where $\|G\|_{\mathcal{G}} \ll \|f\|_{L^2}$ and $\|E\|_{L^2} \ll (m/N)^c \|f\|_{L^2}$ for some universal constant $c > 0$.

See Proposition 15 below, which gives a stronger maximal form of this statement. The condition $\sum_{n=1}^N |\phi_n(x)|^2 \leq N$ can usually be removed in applications (such as Theorem 2) by a change of measure argument (see Lemma 6). It seems likely that this decomposition has other applications.

### 2. Preliminaries

We fix the probability space $(\mathbb{T}, \mathfrak{B}, \mu)$. We define several different norms on the space of functions from $\mathbb{T}$ to $\mathbb{R}$. First, for a positive constant $c$, let $\|\cdot\|_{\mathcal{G}(c)}$ denote the norm of the Orlicz space associated to the convex function $e^{c x^2} - 1$. That is,

$$\|f\|_{\mathcal{G}(c)} := \inf_{\lambda \in \mathbb{R}^+} \left\{ \int_\mathbb{T} e^{c|f(\lambda)|^2} d\mu - 1 \leq 1 \right\}.$$
When we write $\| \cdot \|_q$ with the specification of $c$ omitted, we mean $c = 1$.

We next define the convex function

$$\Gamma_K(t) := \begin{cases} e^{t^2} - 1, & |t| \leq K, \\ e^{K^2} t^2 + e^{K^2} (1 - K^2) - 1, & |t| \geq K \end{cases}$$

and denote the associated Orlicz norm by $\| \cdot \|_{\Gamma_K}$.

**Lemma 4.** When $K \geq 1$, for all $t$, we have that

$$\Gamma_K(t) \leq e^{t^2} - 1, \quad \Gamma_K(t) \leq e^{K^2} t^2.$$ 

It follows that for $f : \mathbb{T} \to \mathbb{R}$ we have $\| f \|_{\Gamma_K} \leq \| f \|_q$ and $\| f \|_{\Gamma_K} \leq e^{K^2/2} \| f \|_{L^2}$.

**Proof.** We first prove $\Gamma_K(t) \leq e^{t^2} - 1$ for all $t$. For $t$ such that $|t| \leq K$, this is clear since $\Gamma_K(t) = e^{t^2} - 1$.

We consider $t$ such that $|t| \geq K$. Then $\Gamma_K(t) = e^{K^2} t^2 + e^{K^2} (1 - K^2) - 1$, so we must show that $e^{K^2} t^2 + e^{K^2} (1 - K^2) \leq e^{t^2}$. We note that for all real $x \geq 0$, $1 + x \leq e^x$. Applying this to the quantity $t^2 - K^2 + 1 > 0$, we have

$$e^{K^2} t^2 + e^{K^2} (1 - K^2) = e^{K^2} (t^2 - K^2 + 1) \leq e^{K^2} e^{t^2 - K^2} = e^{t^2},$$

as required.

We let $f$ be a function from $\mathbb{T}$ to $\mathbb{R}$. For any fixed positive real number $\lambda$ such that $\int_{\mathbb{T}} e^{[f/\lambda]^2} \, d\mu - 1 \leq 1$ (that is, $\lambda \geq \| f \|_q$), we have

$$\int_{\mathbb{T}} \Gamma_K(f/\lambda) \, d\mu \leq \int_{\mathbb{T}} e^{[f/\lambda]^2} \, d\mu - 1 \leq 1,$$

since $\Gamma_K(t) \leq e^{t^2} - 1$ for all $t$. This shows that $\lambda \geq \| f \|_{\Gamma_K}$. Hence $\| f \|_{\Gamma_K} \leq \| f \|_q$.

Next we prove $\Gamma_K(t) \leq e^{K^2} t^2$. We first consider $t$ such that $|t| \geq K$. $\Gamma_K(t) = e^{K^2} t^2 + e^{K^2} (1 - K^2) - 1$ in this case. Since $K \geq 1$, we see that $e^{K^2} (1 - K^2) < 0$, so $\Gamma_K(t) \leq e^{K^2} t^2$ follows. For $t$ such that $|t| \leq K$, we have $\Gamma_K(t) = e^{t^2} - 1$, so we must show that $e^{t^2} - 1 \leq e^{K^2} t^2$ for $|t| \leq K$.

We consider $(e^{t^2} - 1)/t^2$ as a function of $t$ for $t \geq 0$. Its derivative is

$$2(t^{-1} e^{t^2} - t^{-3} e^{t^2} + t^{-3}).$$

We observe that this is always nonnegative. To see this, consider multiplying the quantity by $t^3$ to obtain $2(t^2 e^{t^2} - e^{t^2} + 1)$. Nonnegativity then follows from the inequality $1 + x e^x \geq e^x$ for all real $x \geq 0$. (This inequality can be proved by noting that $x e^x \geq \int_0^x e^u \, du$.) Hence $(e^{t^2} - 1)/t^2$ is a nondecreasing function of $t$ in the range $0 \leq t \leq K$. So it suffices to consider the value at $t = K$, which is $K^{-2} (e^{K^2} - 1)$. Since $K \geq 1$, this is $< e^{K^2}$, as required.

For $f : \mathbb{T} \to \mathbb{R}$, we consider $\lambda := e^{K^2/2} \| f \|_{L^2}$. Then

$$\int_{\mathbb{T}} \Gamma_K(f/\lambda) \, d\mu \leq \int_{\mathbb{T}} e^{K^2} \frac{f^2}{\lambda^2} \, d\mu = \frac{e^{K^2}}{\lambda^2} \| f \|_{L^2}^2 = 1,$$

since $\Gamma_K(t) \leq e^{K^2} t^2$. Thus, $\| f \|_{\Gamma_K} \leq e^{K^2/2} \| f \|_{L^2}$. \qed
Lemma 5. For any (measurable) $f : \mathbb{T} \to \mathbb{R}$, we can decompose $f = f_1 + f_2$ such that

$$
\| f_1 \|_\gamma \ll \| f \|_{\Gamma_K} \quad \text{and} \quad \| f_2 \|_{L^2} \ll e^{-cK^2} \| f \|_{\Gamma_K}
$$

for some universal constant $c > 0$.

Proof. Given $f$, we define $\gamma := 2\| f \|_{\Gamma_K}$ to simplify our notation. We then set

$$
f_1 := f \cdot \mathbb{1}_{|f|/\gamma \leq K} \quad \text{and} \quad f_2 := f \cdot \mathbb{1}_{|f|/\gamma \geq K},
$$

where $\mathbb{1}_S$ for a set $S \subset \mathbb{T}$ denotes the indicator function for that set. By definition of $\gamma = 2\| f \|_{\Gamma_K} > \| f \|_{\Gamma_K}$, we have that

$$
\int_\mathbb{T} \Gamma_K (f/\gamma) \, d\mu = \int_\mathbb{T} \left( e^{\| f/\gamma \|^2} - 1 \right) \cdot \mathbb{1}_{|f/\gamma| \leq K} \, d\mu + \int_\mathbb{T} \left( e^{K^2 f^2/\gamma^2} + e^{K^2 (1 - K^2) - 1} \right) \cdot \mathbb{1}_{|f/\gamma| \geq K} \, d\mu \leq 1. \quad (2)
$$

Since this is a sum of two nonnegative quantities, this implies

$$
\int_\mathbb{T} \left( e^{\| f/\gamma \|^2} - 1 \right) \cdot \mathbb{1}_{|f/\gamma| \leq K} \, d\mu \leq 1.
$$

This is equivalent to

$$
\int_\mathbb{T} e^{\| f/\gamma \|^2} \, d\mu - 1 \leq 1,
$$

and so $\| f_1 \|_\gamma \leq \gamma \ll \| f \|_{\Gamma_K}$.

Again considering (2), we also have

$$
\int_\mathbb{T} \left( e^{K^2 f^2/\gamma^2} + e^{K^2 (1 - K^2) - 1} \right) \cdot \mathbb{1}_{|f/\gamma| \geq K} \, d\mu \leq 1.
$$

We let $\mu(|f/\gamma| \geq K)$ denote the measure of the set in $\mathbb{T}$ on which $|f/\gamma| \geq K$. We can then rewrite the above as

$$
\mu \left( \left| \frac{f}{\gamma} \right| \geq K \right) \left( e^{K^2 (1 - K^2)} - 1 \right) + \int_\mathbb{T} e^{K^2 f^2/\gamma^2} \, d\mu \leq 1. \quad (3)
$$

Now, since $\int_\mathbb{T} \Gamma_K (f/\gamma) \, d\mu \leq 1$ and $\Gamma_K (f/\gamma) \geq e^{K^2} - 1$ whenever $|f/\gamma| \geq K$, we must have

$$
\mu \left( \left| \frac{f}{\gamma} \right| \geq K \right) (e^{K^2} - 1) \leq 1.
$$

Thus, $\mu(|f/\gamma| \geq K) \leq 1/(e^{K^2} - 1)$. Combining this with (3), we have

$$
\int_\mathbb{T} e^{K^2 f^2/\gamma^2} \, d\mu \leq 1 + \mu \left( \left| \frac{f}{\gamma} \right| \geq K \right) (e^{K^2 (K^2 - 1)} + 1) \ll K^2,
$$

and hence

$$
\| f_2 \|_{L^2}^2 \ll K^2 e^{-K^2} \gamma^2,
$$

implying that $\| f_2 \|_{L^2} \ll e^{-cK^2} \| f \|_{\Gamma_K}$ for some universal constant $c > 0$. \qed

Finally, we note the following.
Lemma 6. It suffices to prove (the second formulation of) Theorem 2 with the restriction \( \sum_{n=1}^{N} |\phi_n|^2 \leq N \).

Proof. Consider an arbitrary ONS \( \Phi := \{\phi_n\}_{n=1}^{N} \) and define \( \nu(x) = N^{-1} \sum_{n=1}^{N} |\phi_n(x)|^2 \). Fix \( O \in \mathcal{O}(N) \). Define \( \Phi := \Phi(O) \). Furthermore, consider the ONS \( \Psi \) defined on \( \mathbb{T} \) (with the measure induced by integration against \( \nu(x) \, d\mu \)) by \( \psi_n(x) := \nu^{-1/2}(x)|\phi_n(x)| \). Furthermore, define \( \tilde{\Psi} = \Psi(O) \). We have the trivial identity

\[
\int_{\mathbb{T}} \max_{I \in \mathcal{I}} \sum_{n \in I} |a_n \tilde{\phi}_n(x)|^2 \, d\mu = \int_{\mathbb{T}} \max_{I \in \mathcal{I}} \sum_{n \in I} \sum_{l \in I} |a_n \tilde{\phi}_n(x)|^2 \nu(x) \, d\mu.
\]

Thus, the conclusion of Theorem 2 holds for \( \Phi \) if and only if it holds for \( \Psi \). However, \( \sum_{n=1}^{N} |\psi_n|^2 \leq N \) by construction. \( \square \)

3. Probabilistic Methods

In this section we establish the following result.

Proposition 7. For \( N \) fixed, let \( \{\phi_n(x)\}_{n=1}^{N} \) be an ONS such that \( \sum_{n=1}^{N} |\phi_n(x)|^2 \leq N \). Define for each \( 1 \leq m \leq N \) the function

\[
\Gamma_* := \Gamma \sqrt{(2/5) \log((N/m) \log(N/m+1))}
\]

(the dependence on \( m \) is implicit in this notation). There exists a subset \( Q \subset \mathcal{O}(N) \) with \( \mathbb{P}[Q] \geq 1 - C(e^{-C/2}) \) such that for all \( O = \{o_{i,n}\}_{1 \leq i, n \leq N} \in Q \) the corresponding base change of \( \{\phi_n\}_{n=1}^{N} \), that is

\[
\psi_n(x) := \sum_{i=1}^{N} o_{i,n} \phi_i(x),
\]

satisfies the following. For each \( m \) in the range \( 1 \leq m \leq N \),

\[
\left\| \sum_{n=1}^{N} a_n \psi_n \right\|_{\Gamma_*} \ll \left( \sum_{n=1}^{N} a_n^2 \right)^{1/2}
\]

for all vectors \( a \in \mathbb{R}^N \) such that \( \text{support}(a) \leq m \). (We use \( \text{support}(a) \) to denote the number of nonzero coordinates of \( a \).)

The proof will build on arguments from [Bourgain 1989], although the estimates we obtain are substantially stronger. We start by establishing a weaker result. For a fixed \( m \) in the range \( 1 \leq m \leq N \), we let \( \mathcal{S}_m \subset \mathbb{R}^N \) denote the subset of vectors \( b \) such that \( \|b\|_{L^2} \leq 1 \) and \( \text{support}(b) \leq m \). We then define

\[
B(m, \mathcal{C}) := \sup_{a \in \mathcal{S}_m} \left\| \sum_{n=1}^{N} a_n \psi_n \right\|_{\Gamma_*}.
\]

Note that both the set \( \mathcal{S}_m \) and the function \( \Gamma_* := \Gamma \sqrt{(2/5) \log((N/m) \log(N/m+1))} \) depend on \( m \). Our first step will be to establish the following.
Proposition 8. For any $1 \leq m \leq N$ we have that
\[ \mathbb{E}_{\mathcal{O}}(N) B(m, O) \ll 1, \]
where the implied constant is independent of $m$ and $N$.

This does not quite give Proposition 7, since there the claim is made with large probability and we require the estimates to hold for all $m$ simultaneously. The stronger claim, however, will be deduced later from the weaker statement using the concentration of measure phenomenon on the orthogonal group.

We will need the following result, which is Lemma 5.5 from [Bourgain 1989], where it is attributed to [Benyamini and Gordon 1981]. The result is a concatenation of Lemmas 1.10 and 1.12 in [Benyamini and Gordon 1981]. These are due to [Chevet 1978] and [Marcus and Pisier 1981], respectively.

Lemma 9. Let $X$ and $Y$ be Banach spaces, and consider the operator
\[ T_O := \sum_{i,j=1}^{N} o_{ij} (x_i^* \otimes y_j) \]
for $O := (o_{ij})_{1 \leq i,j \leq N} \in \mathcal{O}(N)$, and where $\{x_i^*\}_{i=1}^{N}$ (respectively $\{y_j\}_{j=1}^{N}$) are sequences in $X^*$ (respectively $Y$). Then
\[ \int_{\mathcal{O}} \|T_O\| \leq \frac{C \alpha(\{x_i^*\}_{i=1}^{N})}{\sqrt{N}} \int \left\| \sum_{j=1}^{N} g_j(\omega) y_j \right\| d\omega + \frac{C \alpha(\{y_j\}_{j=1}^{N})}{\sqrt{N}} \int \left\| \sum_{i=1}^{N} g_i(\omega)x_i^* \right\| d\omega, \tag{4} \]
where
\[ \alpha(\{x_i^*\}) := \sup \left\{ \left( \sum |\langle x_i^*, x \rangle|^2 \right)^{1/2} : x \in X, \|x\| \leq 1 \right\}, \]
\[ \alpha(\{y_j\}) := \sup \left\{ \left( \sum |\langle y_j, y^* \rangle|^2 \right)^{1/2} : y^* \in Y^*, \|y^*\| \leq 1 \right\}, \]
and $\{g_i\}_{i=1}^{N}$ is a system of independent Gaussians with mean zero and variance one. Note that the norms in (4) refer, respectively, to the Banach spaces $B(X, Y)$, $Y$, and $X^*$.

Let $\ell^2[N]$ denote the set of real sequences $a := \{a_n\}_{n=1}^{N}$. We will denote by $X$ the Banach space obtained by considering this set with the norm $\| \cdot \|_{[2]}$ defined as follows. For a vector $a$, we define $\|a\|_{[2]}$ to be the infimum of positive $c \in \mathbb{R}$ such that scaling the convex hull of $S_m$ by $c$ results in a set containing $a$. We take $Y$ to be the space of real-valued functions on $T$ equipped with the Orlicz norm associated to $\Gamma_*$.

Let $x_i^*$ ($1 \leq i \leq N$) denote the canonical unit vectors in $\mathbb{R}^N$ (which is naturally identified with the dual space $X^*$). We have, from Lemma 9, that
\[ \mathbb{E}B(m, O) \ll \frac{\alpha(\{x_i^*\}_{i=1}^{N})}{\sqrt{N}} \left\| \sum_{i=1}^{N} g_i \phi_i \right\|_{\Gamma_*} + \frac{\alpha(\{\phi_i\}_{i=1}^{N})}{\sqrt{N}} \left\| \sum_{i=1}^{N} g_i x_i^* \right\|_{X^*}. \]
In order to establish Proposition 8, we need to show the above is $\ll 1$. This follows from the estimates
\[ \alpha(\{x_i^{*1}\}_{i=1}^N) \ll 1, \quad \alpha(\{|\phi_i\}_i^N) \ll \left( \frac{N}{m} \log \left( \frac{N}{m} + 1 \right) \right)^{1/5}, \]
\[ \mathbb{E} \left\| \sum_{i} g_i \phi_i \right\|_{\Gamma_*} \leq \sqrt{N}, \quad \mathbb{E} \left\| \sum_{i} g_i x_i^{*} \right\|_{\mathcal{X}^{*}} \leq \sqrt{m} \sqrt{\log \left( \frac{N}{m} + 1 \right)} . \]

The first estimate above follows from the observation that the convex hull of $\mathbb{S}_m$ is contained in the $\ell^2$ unit ball in $\mathbb{R}^N$. We will prove the others in the following lemmas.

**Lemma 10.** We have that $\mathbb{E} \| \sum_{i} g_i \phi_i \|_{\Gamma_*} \ll \sqrt{N}$.

**Proof.**
Letting $C$ be a positive constant, by Fubini’s theorem we have that
\[ \mathbb{E} \int_{\mathbb{T}} e^{\sum_{i} g_i \phi_i(x)^2/(CN)} d\mu = \int_{\mathbb{T}} \mathbb{E} e^{\sum_{i} g_i \phi_i(x)^2/(CN)} d\mu . \]

Now, for each fixed $x$, we recall that $\sum_{i} |\phi_i(x)|^2 \leq N$, so $(1/\sqrt{CN}) \sum_{i} g_i \phi_i(x)$ is a Gaussian random variable with mean 0 and variance at most $1/C$. Thus, $\int_{\mathbb{T}} \mathbb{E} e^{(\sum_{i} g_i \phi_i(x)^2)/(CN)} d\mu \ll 1$ for an appropriate choice of $C$.

Since $e^{f^2/\lambda} \leq 1 + e^{f^2}/\lambda$ for $\lambda \geq 1$, we have that $\inf_{\lambda \in \mathbb{R}^+} \{ \int_{\mathbb{T}} e^{f^2/\lambda} d\mu \leq 2 \} \ll 1 + \int_{\mathbb{T}} e^{f^2} d\mu$. Applying this to $f = (1/\sqrt{CN}) \sum_{i} g_i \phi_i$, we have
\[ \left\| \frac{1}{\sqrt{CN}} \sum_{i} g_i \phi_i \right\|_{\Gamma_*} \ll \int_{\mathbb{T}} e^{(\sum_{i} g_i \phi_i(x)^2)/(CN)} d\mu . \]

Taking expectations on both sides, we have $\mathbb{E} \| \sum_{i} g_i \phi_i \|_{\Gamma_*} \ll \sqrt{N}$, as required. \qed

**Lemma 11.** We have that $\alpha(\{|\phi_i\}_i^N) \ll ((N/m) \log(N/m + 1))^{1/5}$.

**Proof.** From Lemma 4 it follows that $\| f \|_{\Gamma_*} \leq ((N/m) \log(N/m + 1))^{1/5} \| f \|_{L^2}$. Now
\[ \| g \|_{\Gamma_*} = \sup_{f \in \Gamma_*} \frac{\langle f, g \rangle}{\| f \|_{\Gamma_*}} \geq \frac{\langle g, g \rangle}{\| g \|_{\Gamma_*}} \gg \frac{\| g \|^2}{((N/m) \log(N/m + 1))^{1/5} \| g \|_{L^2}} \gg \left( \frac{N}{m} \log \left( \frac{N}{m} + 1 \right) \right)^{-1/5} \| g \|_{L^2} . \]

Here we have used that each element of the dual space $\Gamma_*^{\mathcal{X}^{*}}$ can be represented as integration against a measurable function. This follows from standard properties of Orlicz spaces. In particular, see Theorem 14.2 of [Krasnosel'skiĭ and Rutickiĭ 1961], since the modulus $\Gamma_*$ satisfies the $\Delta^2_2$ condition.

It now follows that if $\| g \|_{\Gamma_*} \leq 1$, then $\| g \|_{L^2} \ll ((N/m) \log(N/m + 1))^{1/5}$. Thus by Bessel’s inequality we have
\[ \alpha(\{|\phi_i\}) := \sup \left\{ \left( \sum_{g \in \Gamma_*^{\mathcal{X}^{*}}} \langle \phi_i, g \rangle \right)^2 : \| g \|_{\Gamma_*} \leq 1 \right\} \ll \left( \frac{N}{m} \log \left( \frac{N}{m} + 1 \right) \right)^{1/5} , \]
which completes the proof. \qed
Lemma 12. \[ E\|\sum g_i x_i^*\|_{X^*} \leq \sqrt{m} \sqrt{\log(N/m + 1)}. \]

Proof. It follows from the definition of \( X^* \) that

\[ E\|\sum g_i x_i^*\|_{X^*} = E \sup_{a \in S_m} \left| \sum g_i a_i \right|. \]

(Note that taking the supremum over the convex hull of \( S_m \) would yield the same result.)

The latter quantity is well studied in the theory of Gaussian processes. Recall that Dudley’s bound [1967] gives

\[ \ll \int_0^\infty \sqrt{\log(N(S_m, \epsilon))} \, d\epsilon, \]

where \( N(S_m, \epsilon) \) denotes the number of \( \ell^2 \) balls of radius \( \epsilon \) needed to cover \( S_m \). Now, clearly \( S_m \) is a subset of the \( n \)-dimensional \( \ell^2 \) unit ball. Thus \( \log(N(S_m, \epsilon)) = 0 \) for \( \epsilon \geq 1 \), and the above quantity is equal to

\[ \int_0^1 \sqrt{\log(N(S_m, \epsilon))} \, d\epsilon. \]

Lemma 12 now follows from the following.

Lemma 13. For \( 0 < \epsilon \leq 1 \), we have that

\[ N(S_m, \epsilon) \ll \left( \frac{N}{m} \right) \left( \frac{3}{\epsilon} \right)^m, \]

and thus

\[ \log N(S_m, \epsilon) \ll m \log \left( \frac{N}{m} + 1 \right) + m \log \left( \frac{3}{\epsilon} \right). \]

Proof. We prove the first inequality (the second follows by taking logarithms). We let \( K \) denote the unit \( \ell^2 \) ball in \( \mathbb{R}^m \). Then \( N(K, \epsilon K) \leq (3/\epsilon)^m \), where \( N(K, \epsilon K) \) denotes the number of translates of \( \epsilon K \) needed to cover \( K \). To see this, consider a maximal set of disjoint balls of radius \( \epsilon/2 \) with centers in \( K \). Let \( T \) denote the set of their centers. By maximality, taking balls of radius \( \epsilon \) around each point in \( T \) yields a cover of \( K \), and hence the cardinality of \( T \) is an upper bound on \( N(K, \epsilon K) \). Now, the union of all the disjoint balls of radius \( \epsilon/2 \) with centers in \( T \) is a set with volume equal to \( |T| \, \text{vol}((\epsilon/2)K) \), where \( |T| \) denotes the cardinality of \( T \) and \( \text{vol}((\epsilon/2)K) \) denotes the volume of the ball of radius \( \epsilon/2 \). Since this set is contained in \( (1 + \epsilon/2)K \), we have

\[ N(K, \epsilon K) \leq \frac{\text{vol}((1 + \epsilon/2)K)}{\text{vol}((\epsilon/2)K)} = \frac{(1 + \epsilon/2)^m}{(\epsilon/2)^m} = \left(1 + \frac{\epsilon}{2}\right)^m \leq \left(\frac{3}{\epsilon}\right)^m \]

whenever \( 0 < \epsilon \leq 1 \).

Fix \( m \) coordinates and consider the associated \( m \)-dimensional \( \ell^2 \) ball. We have shown that this can be covered by \( (3/\epsilon)^m \) balls of radius \( \epsilon \). Summing over all \( \binom{N}{m} \) such balls completes the proof. \( \square \)

This completes the proof of Lemma 12 and hence the proof of Proposition 8. \( \square \)
3.1. Concentration of measure on $\mathbb{O}(n)$. In the prior section, we proved that for any $1 \leq m \leq N$ we have $E_{O(N)} B(m, O) \ll 1$. It follows from Markov’s inequality that, for some large universal $C$, we have $\nu(\mathcal{A}(m)) > \frac{1}{2}$, where

$$\mathcal{A}(m) := \{ O \in \mathbb{O}(N) : B(m, O) \leq C \}$$

and $\nu(\mathcal{A}(m))$ denotes the measure of the set $\mathcal{A}(m)$ in $\mathbb{O}(N)$.

Consider the Hilbert–Schmidt norm on the set of $N \times N$ matrices, $\|A\|_{HS} := (\sum_{1 \leq i, j \leq N} |A_{i,j}|^2)^{1/2}$. We recall the concentration of measure inequality on the Orthogonal group; see [Milman and Schechtman 1986].

**Lemma 14.** Let $\nu$ denote the Haar measure on the orthogonal group $O(N)$ and $A \subset O(N)$ such that $\nu(A) > \frac{1}{2}$. Then

$$\mathbb{P}[A \in O(N) : \inf_{B \in A_c} \|A - B\|_{HS} > \epsilon] \ll e^{-c\epsilon^2 N}$$

for some absolute positive constant $c$.

For any $N \times N$ matrix $M = \{m_{i,j}\}$, using the bounds from Lemma 4, we have

$$\left\| \sum_{1 \leq i, n \leq N} m_{i,n} a_i \phi_n \right\|_{\Gamma_*} \ll \left( \frac{N}{m} \log \left( \frac{N}{m} \right) \right)^{1/5} \left( \sum_n \left( \sum_i m_{i,n} a_i \right)^2 \right)^{1/2}
\ll \left( \frac{N}{m} \log \left( \frac{N}{m} \right) \right)^{1/5} \|M\|_{HS} \|a\|_{\ell^2}$$

(5)

for all $a \in \mathbb{R}^N$. The final inequality follows from Cauchy–Schwarz.

Now consider $\mathcal{A}(m, \epsilon) \subset \mathbb{O}(N)$, defined to be the set of all orthogonal matrices that differ from an element of $\mathcal{A}(m)$ by a matrix with Hilbert–Schmidt norm at most $\epsilon$. Using (5), we have that for $O \in \mathcal{A}(m, (m/(N \log(N/m)))^{1/5})$, $B(m, O) \leq C'$, where $C'$ is a new absolute constant. On the other hand, denoting the complement of $\mathcal{A}(m, (m/(N \log(N/m)))^{1/5})$ by $\mathcal{A}^c(m, (m/(N \log(N/m)))^{1/5})$, by Lemma 14, we have

$$\mathbb{P}\left[ O \in \mathcal{A}^c \left( m, \left( \frac{m}{N \log(N/m)} \right)^{1/5} \right) \right] \ll e^{-cN^{2/5}}$$

for some positive constant $c$.

Now to conclude the proof of Proposition 7, it suffices to find a sufficiently high probability set of elements $O \in \mathbb{O}(N)$ such that for every $1 \leq m \leq N$ we have $O \in \mathcal{A}\left( m, \left( \frac{m}{N \log(N/m)} \right)^{1/5} \right)$. However, for sufficiently large $N$, we see from the union bound that

$$\nu \left( \bigcup_{1 \leq m \leq N} \mathcal{A}^c \left( m, \left( \frac{m}{N \log(N/m)} \right)^{1/5} \right) \right) \leq Ne^{-cN^{2/5}} \ll e^{-c_2 N^{2/5}}.$$ 

This completes the proof of Proposition 7.
We also define $I_k$ in an interval $Q$ to be empty. By construction we have that $I_k$ satisfies the following property. For any $f = \sum a_n \psi_n$, letting $m$ denote $\text{support}(\{a_n\})$, we have that the maximal function defined by

$$Mf := \sup_{I \subseteq [N]} \left| \sum_{n \in I} a_n \psi_n \right|$$

can be decomposed as $Mf = \widetilde{G} + \widetilde{E}$, where $\|\widetilde{G}\|_\infty \ll \|f\|_{L^2}$ and $\|\widetilde{E}\|_{L^2} \ll (m/N)^c \|f\|_{L^2}$ for some universal constant $c > 0$.

To prove this, we fix $Q \subset \mathcal{C}(N)$ from Proposition 7. We now decompose $[N]$ into a family of subintervals according to a concept of mass defined with respect to the $a_i$ values. We define the mass of a subinterval $I \subseteq [N]$ as $M(I) := \sum_{n \in I} |a_n|^2$. By normalization, we may assume that $M([N]) = 1$. We define $I_{0,1} := [N]$ and we iteratively define $I_{k,s}$, for $1 \leq s \leq 2^k$, as follows. Assuming we have already defined $I_{k-1,s}$ for all $1 \leq s \leq 2^{k-1}$, we will define $I_{k,2s-1}$ and $I_{k,2s}$, which are subintervals of $I_{k-1,s}$. $I_{k,2s-1}$ begins at the left endpoint of $I_{k-1,s}$ and extends to the right as far as possible while covering strictly less than half the mass of $I_{k-1,s}$, while $I_{k,2s}$ ends at the right endpoint of $I_{k-1,s}$ and extends to the left as far as possible while covering at most half the mass of $I_{k-1,s}$. More formally, we define $I_{k,2s-1}$ as the maximal subinterval of $I_{k-1,s}$ which contains the left endpoint of $I_{k-1,s}$ and satisfies $M(I_{k,2s-1}) < \frac{1}{2}M(I_{k-1,s})$. We also define $I_{k,2s}$ as the maximal subinterval of $I_{k-1,s}$ which contains the right endpoint of $I_{k-1,s}$ and satisfies $M(I_{k,2s}) \leq \frac{1}{2}M(I_{k-1,s})$. We note that these subintervals are disjoint. We may express $I_{k-1,s}$ as $I_{k,2s-1} \cup I_{k,2s} \cup i_{k,s}$, where $i_{k,s} \in I_{k-1,s}$. In other words, $i_{k,s}$ denotes the single element which lies between $I_{k,2s-1}$ and $I_{k,2s}$ (note that such a point always exists because we have required that $I_{k,2s-1}$ contains strictly less than half of the mass of the interval). Here it is acceptable, and in many instances necessary, for some choices of the intervals in this decomposition to be empty. By construction we have that

$$M(I_{k,s}) \leq 2^{-k}.$$  \hfill (6)

We call an interval $J \subseteq [N]$ admissible if it is an element of the decomposition given above. We denote the collection of admissible intervals by $\mathcal{A}$. We additionally refer to the subset $\{I_{k,s} : 1 \leq s \leq 2^k\}$ of $\mathcal{A}$ as the admissible intervals on level $k$ and the subset $\{i_{k,s} : 1 \leq s \leq 2^k\}$ as the admissible points on level $k$. We note that every point in $[N]$ is an admissible point on some level. (Eventually, we have subdivided all intervals down to being single elements.)

Now we write $\mathcal{J}_k := \{I_{k,s} : 1 \leq s \leq 2^k\}$. We decompose this as $\mathcal{J}_k^a := \{I \in \mathcal{J}_k : |I| \leq 2^{-k/2}N\}$ and its complement, $\mathcal{J}_k^b := \{I \in \mathcal{J}_k : |I| > 2^{-k/2}N\}$. Here, $|I|$ denotes the number of nonzero $a_i$ values contained in an interval $I$.

For $J \subseteq [N]$, we define

$$S_J(x) = \sum_{n \in J} a_n \psi_n(x).$$
We also define
\[ \tilde{S}_J(x) := \max_{I \subseteq J} \left| \sum_{n \in I} a_n \psi_n(x) \right|. \]

From Lemma 5 and Proposition 7, we deduce that \( S_J = G_J + E_J \), where \( \| G_J \|_{\ell^q} \ll \| S_J \|_{L^2} \) and \( \| E_J \|_{L^2} \ll (|J|/N)^c \| S_J \|_{L^2} \) for some positive constant \( c' \). Our purpose now is to show a similar decomposition for \( \tilde{S}_J(x) \). Clearly, it suffices to show such a decomposition for a pointwise majorant. Denote the decomposition of \( S_{k,s} \) by \( S_{k,s} := G_{k,s} + E_{k,s} \), and the decomposition of \( S_{i,k,s} \) by \( S_{i,k,s} := G_{i,k,s} + E_{i,k,s} \). Setting \( r = 3 \), for an interval \( J \) we have the following bound, where the sums below are restricted to values of \( k, s \) such that \( I_{k,s}, i_{k,s} \subseteq J \):

\[
\tilde{S}_J(x) \ll \sum_k \left( \sum_s |G_{k,s} + E_{k,s}|^r \right)^{1/r} + \sum_k \left( \sum_s |G_{i,k,s} + E_{i,k,s}|^r \right)^{1/r} 
\ll \left( \sum_k \left( \sum_s |G_{k,s}|^r \right)^{1/r} + \sum_k \left( \sum_s |G_{i,k,s}|^r \right)^{1/r} \right) 
\ll \left( \sum_k \left( \sum_s |E_{k,s}|^r \right)^{1/r} + \sum_k \left( \sum_s |E_{i,k,s}|^r \right)^{1/r} \right) 
= : \tilde{G}_J + \tilde{E}_J. \tag{7}
\]

This follows from the observation that, for each point \( x \), the maximizing subinterval \( I \subseteq J \) can be decomposed as a union of admissible intervals and points with at most two intervals and points on each level. The contribution on each level can then be bounded by a constant times the contribution from the "worst" interval/point, which is in turn bounded by the quantity inside the sum over \( k \) above for each level \( k \).

For an admissible interval \( J \), we let \( k^* \) denote the level of \( J \). We note that the sums over \( k \) in (7) range only over \( k \geq k^* \) (and the sums over \( s \) are also appropriately restricted). Next we show that \( \| \tilde{G}_J \|_{\ell^q(c)} \ll \| S_J \|_{L^2} \) for some absolute constant \( c \) and \( \| \tilde{E}_J \|_{L^2} \ll (|J|/N)^c \| S_J \|_{L^2} \).

Now let us estimate \( \| \tilde{E}_J \|_{L^2} \). We first estimate the contribution from the admissible points \( i_{k,s} \in J \). We observe that

\[
\left\| \sum_k \left( \sum_s |E_{i,k,s}|^r \right)^{1/r} \right\|_{L^2} \leq \sum_k \left\| \left( \sum_s |E_{i,k,s}|^r \right)^{1/r} \right\|_{L^2}.
\]

Since \( r > 2 \), this is at most

\[
\sum_k \left( \sum_s \| E_{i,k,s} \|_{L^2}^2 \right)^{1/2} \ll \left( \frac{1}{N} \right)^{c'} \sum_k \left( \sum_s \| S_{i,k,s} \|_{L^2}^2 \right)^{1/2},
\]

where the latter inequality follows from the definition of \( E_{i,k,s} \).

Now since these sums only range over values of \( k, s \) such that \( i_{k,s} \in J \), we may split the sum over \( k \) into two portions as

\[
\sum_k \left( \sum_s \| S_{i,k,s} \|_{L^2}^2 \right)^{1/2} = \sum_{k=k^*}^{k^*+10 \log(N)} \left( \sum_s \| S_{i,k,s} \|_{L^2}^2 \right)^{1/2} + \sum_{k>k^*+10 \log(N)} \left( \sum_s \| S_{i,k,s} \|_{L^2}^2 \right)^{1/2}. \tag{8}
\]
To bound the first quantity in (8), it suffices to observe that the inner quantity for each \( k \) is at most \( \| S_J \|_{L^2} \), and hence its contribution is \( \ll \log(N) \| S_J \|_{L^2} \ll N^\epsilon \| S_J \|_{L^2} \), for a constant \( 0 < c' < 1 \). (Thus we will adjust the value of \( c' \) for our final estimate by subtracting \( \epsilon \).)

To bound the second quantity in (8), we note that, for any \( i_{k,s} \in J \) with \( k > k^* + 10 \log(N) \), we have \( \| S_{i_{k,s}} \|_{L^2} \leq N^{-10} \| S_J \|_{L^2} \). There are at most \( N \) points \( i_{k,s} \) in the sum, and thus

\[
\sum_{k > k^* + 10 \log(N)} \left( \sum_s \| S_{i_{k,s}} \|_{L^2}^2 \right)^{1/2} \ll N^{-4} \| S_J \|_{L^2}.
\]

To estimate the contribution from the admissible intervals, we proceed as follows. For each \( k \geq k^* \), we define \( I_k^a(J) \) to be the set of admissible intervals \( I \) on level \( k \) contained in \( J \) such that \( |I| < 2^{-(k-k^*)/2} |J| \) and we let \( I_k^b(J) \) denote the set of remaining admissible intervals on level \( k \) contained in \( J \). Note that \( I_k^a(J) \) and \( I_k^b(J) \) are disjoint, and their union is the set of all admissible intervals on level \( k \) contained in \( J \). It thus suffices to estimate

\[
\tilde{E}_J^a + \tilde{E}_J^b := \sum_{k \geq k^*} \left( \sum_{I_{k,s} \in I_k^a(J)} |E_{i_{k,s}}|^r \right)^{1/r} + \sum_k \left( \sum_{I_{k,s} \in I_k^b(J)} |E_{i_{k,s}}|^r \right)^{1/r}.
\]

Now \( |I_k^b(J)| \leq 2^{(k-k^*)/2} \), and we also have

\[
\| E_{i_{k,s}} \|_{L^2} \ll \left( \frac{|J|}{N} \right)^{c'} \| S_{i_{k,s}} \|_{L^2} \ll \left( \frac{|J|}{N} \right)^{c'} 2^{-(k-k^*)/2} \| S_J \|_{L^2}.
\]

Since \( r > 2 \), we have

\[
\left\| \sum_{k \geq k^*} \left( \sum_{s \in I_k^b(J)} |E_{i_{k,s}}|^r \right)^{1/r} \right\|_{L^2} \leq \sum_{k \geq k^*} \left( \sum_{s \in I_k^b(J)} \| E_{i_{k,s}} \|_{L^2}^2 \right)^{1/2} \ll \left( \frac{|J|}{N} \right)^{c'} \| S_J \|_{L^2} \sum_{j \geq 0} 2^{-j/4} \ll \left( \frac{|J|}{N} \right)^{c'} \| S_J \|_{L^2}.
\]

Next, we recall that \( I \in I_k^a(J) \) implies \( |I| \leq 2^{-(k-k^*)/2} |J| \). We have \( \| S_{i_{k,s}} \|_{L^2} \ll 2^{-(k-k^*)/2} \| S_J \|_{L^2} \).

Thus \( \| E_{i_{k,s}} \|_{L^2} \ll (|J|/N)^{c'} 2^{-c'(k-k^*)/2} \| S_{i_{k,s}} \|_{L^2} \ll (|J|/N)^{c'} 2^{-(c'+1)(k-k^*)/2} \| S_J \|_{L^2} \).

We then have

\[
\left\| \sum_{k \geq k^*} \left( \sum_{I_{k,s} \in I_k^a(J)} |E_{i_{k,s}}|^r \right)^{1/r} \right\|_{L^2} \leq \sum_{k \geq k^*} \left( \sum_{I_{k,s} \in I_k^a(J)} \| E_{i_{k,s}} \|_{L^2}^2 \right)^{1/2} \ll \left( \frac{|J|}{N} \right)^{c'} \| S_J \|_{L^2} \sum_{k \geq k^*} 2^{k-k^*} 2^{-(c'+1)(k-k^*)} \ll \left( \frac{|J|}{N} \right)^{c'} \| S_J \|_{L^2}.
\]

Here we have used the fact that there are at most \( 2^{k-k^*} \) values of \( s \) such that \( I_{k,s} \subseteq J \) for each \( k \geq k^* \). We can apply this for \( J = [N] \) in particular, recalling that \( |J| \) denotes the number of nonzero \( a_i \) values contained in \( J \), which in this case is \( m \). This completes the proof that \( \| \tilde{E} \|_{L^2} \ll (m/N)^{c'} \| f \|_{L^2} \) for some positive constant \( c' \).

To show that \( \| \tilde{G} \|_{S_{c_0}(c)} \ll \| f \|_{L^2} \) for some universal constant \( c > 0 \), we will use the following lemma. These implications and arguments are well known, however, we include a proof for completeness.
Lemma 16. Let A denote a fixed, positive constant. For positive constants c, C, we define the following sets of measurable functions:

\[ S_1(c) := \{ f : \mathbb{T} \to \mathbb{R} \mid \| f \|_{L^p} \leq c \sqrt{p} A \text{ for all } p \geq 2 \}, \]

\[ S_2(c, C) := \{ f : \mathbb{T} \to \mathbb{R} \mid \mu(|f| \geq \lambda) \leq Ce^{-ck^2/A^2} \text{ for all } \lambda \geq 0 \}, \]

\[ S_3(c) := \{ f : \mathbb{T} \to \mathbb{R} \mid \| f \|_{\|g\|_{C_k}} \leq A \}, \]

where \( \mu(|f| \geq \lambda) \) denotes the measure of the subset of \( x \in \mathbb{T} \) such that \( |f(x)| \geq \lambda \). Then, for any \( c > 0 \), there exist positive constants \( c', C', c'' \) (depending only on \( c \)) such that \( S_1(c) \subseteq S_2(c', C') \) and \( S_1(c) \subseteq S_3(c'') \). Similarly, for any \( c, C > 0 \), there exist positive constants \( c', C', c'' \) (depending only on \( c, C \)) such that \( S_2(c, C) \subseteq S_1(c') \) and \( S_2(c, C) \subseteq S_3(c'') \). Finally, for any \( c > 0 \), there exist positive constants \( c', C', c'' \) (depending only on \( c \)) such that \( S_3(c) \subseteq S_2(c', C') \) and \( S_3(c) \subseteq S_1(c'') \).

Proof. Fixing \( c, C \), we will determine \( c' \) such that \( S_2(c, C) \subseteq S_3(c') \) (for every \( A \)). We consider an \( f \in S_2(c, C) \). We consider \( c' := d_1d_2 \) as a product of two variables \( d_1, d_2 \) whose values will be set later. We assume \( d_1 \leq 1 \). We have

\[ \int_{\mathbb{T}} e^{c'|f|^2/A^2} d\mu = \int_{\mathbb{T}} e^{d_1d_2|f|^2/A^2} d\mu \leq 1 + d_1 \int_{\mathbb{T}} e^{d_1|f|^2/A^2} d\mu, \]  

(9)

using the inequality \( e^{x/a} \leq (1/a)e^x + 1 \) for all \( a \geq 1 \) and nonnegative \( x \) (this can be seen by considering the Taylor expansion of \( e^x \)).

Now, we observe that

\[ \int_{\mathbb{T}} e^{d_2|f|^2/A^2} d\mu \leq \sum_{k \geq 0} \int_{\mathbb{T}} e^{d_2|f|^2/A^2} \un_{A^2k \leq |f|^2 < A^2(k+1)} d\mu \leq \sum_{k \geq 0} \mu(|f| \geq A^2k)e^{d_2(k+1)}, \]

where \( \un_{A^2k \leq |f|^2 < A^2(k+1)} \) denotes the characteristic function of the set on which \( |f|^2 \) takes values between \( A^2k \) and \( A^2(k+1) \). Since \( f \in S_2(c, C) \), we have \( \mu(|f|^2 \geq A^2k) \leq Ce^{-ck} \) for all \( k \geq 0 \). Thus, we conclude

\[ \int_{\mathbb{T}} e^{d_2|f|^2/A^2} d\mu \leq \sum_{k \geq 0} Ce^{-ck+d_2(k+1)} = Ce^{d_2} \sum_{k \geq 0} e^{-(c-d_2)k} = \frac{Ce^c}{e^{c-d_2} - 1}, \]

whenever \( d_2 < c \). Setting \( d_2 = c/2 \), we obtain the above quantity is \( \leq Ce^c/(e^{c/2} - 1) \). Letting \( d_1 = \min\{1, (e^{c/2} - 1)/(Ce^c)\} \), we have

\[ d_1 \int_{\mathbb{T}} e^{d_1|f|^2/A^2} d\mu \leq 1, \]

and hence \( \int_{\mathbb{T}} e^{c'|f|^2/A^2} d\mu - 1 \leq 1 \) for \( c' = d_1d_2 \), showing that \( f \in S_3(c') \). Note that \( c' = d_1d_2 \) depends only on \( c \) and \( C \).

Conversely, we observe that, for every \( c > 0 \), \( S_3(c) \subseteq S_2(c, 2) \). To see this, consider \( f \in S_3(c) \). Then we have

\[ \int_{\mathbb{T}} e^{c'|f|^2/A^2} d\mu - 1 \leq 1 \Rightarrow \int_{\mathbb{T}} e^{c|f|^2/A^2} d\mu \leq 2. \]
Thus, for any \( \lambda > 0 \),
\[
\mu(|f| \geq \lambda) e^{c\lambda^2/A^2} \leq \int_\mathbb{R} e^{c|f|^2/A^2} \, d\mu \leq 2.
\]

It follows that \( f \in S_2(c, 2) \).

For any \( c > 0 \), we will now show there exist \( c', C \) such that \( S_1(c) \subseteq S_2(c', C) \) (for every \( A \)). We consider an \( f \in S_1(c) \). This means that \( \|f\|_p^p \leq c^p p^{p/2} A^p \) for all \( p \geq 2 \). Thus, for every \( \lambda > 0 \),
\[
\mu(|f| \geq \lambda) \lambda^p \leq (cA)^p p^{p/2},
\]
which implies
\[
\mu(|f| \geq \lambda) \leq \frac{(cA)^p p^{p/2}}{\lambda^p}.
\]

For a fixed \( \lambda \), we may minimize this quantity over the choices of \( p \geq 2 \). In the case that \( \lambda^2/(ec^2 A^2) \geq 2 \), we may set \( p \) equal to this value, and the quantity in (10) then becomes
\[
\left( \frac{cA}{\lambda} \right)^{\lambda^2/(ec^2 A^2)} \left( \frac{\lambda^2}{ec^2 A^2} \right)^{\lambda^2/(2ec^2 A^2)} = e^{-\lambda^2/(2ec^2 A^2)}.
\]

Hence, by setting \( c' = 1/(2ec^2) \), we achieve \( \mu(|f| \geq \lambda) \leq e^{-c' \lambda^2/A^2} \) in these cases.

Now, when \( \lambda^2/(ec^2 A^2) < 2 \), we note that \( e^{-c' \lambda^2/A^2} \geq e^{-c'(2ec^2)} = e^{-1} \). Thus, setting \( C = e \), we have \( \mu(|f| \geq \lambda) \leq 1 \leq Ce^{-c' \lambda^2/A^2} \) in these cases. Hence, in all cases, we have that
\[
\mu(|f| \geq \lambda) \leq Ce^{-c' \lambda^2/A^2},
\]
so \( f \in S_2(c', C) \).

Conversely, for any \( c, C > 0 \), we will show that there exists \( c' \) such that \( S_2(c, C) \subseteq S_1(c') \) for every \( A \). We consider an \( f \in S_2(c, C) \). Then, for every \( \lambda \geq 0 \), we have \( \mu(|f| \geq \lambda) \leq Ce^{-c' \lambda^2/A^2} \). We fix \( p \geq 2 \). We observe that
\[
\|f\|_p^p = p \int_0^\infty \lambda^{p-1} \mu(|f| > \lambda) \, d\lambda \ll p \int_0^\infty \lambda^{p-1} e^{-c\lambda^2/A^2} \, d\lambda.
\]
Substituting \( \lambda = t^{1/p} \), we see this equals
\[
\int_0^\infty e^{-ct^{2/p}/A^2} \, dt.
\]

We note the identity \( (p/2) \Gamma(p/2) = \int_0^\infty e^{-x^{2/p}} \, dx \) where \( \Gamma \) denotes the function \( \Gamma(z) := \int_0^\infty y^{z-1} e^{-y} \, dy \). Setting \( s = (c/A^2)^{p/2} t \), we see that the quantity in (11) is equal to
\[
c^{-p/2} A^p \int_0^\infty e^{-s^{2/p}} \, ds = c^{-p/2} A^p \left( \frac{p}{2} \right) \Gamma \left( \frac{p}{2} \right).
\]

By Sterling’s formula, \( \Gamma(p/2) \ll p^{-1/2} (p/(2e))^{p/2} \). Hence,
\[
\|f\|_p \ll A \sqrt{p} (p^{1/(2p)}) \ll A \sqrt{p},
\]
as required. □
Appealing to Lemma 16, we see that we may bound the quantity \( \| \tilde{G}_J \|_{q(c)} \) by considering the \( p \) norm. We recall that

\[
\tilde{G}_J = \sum_k \left( \sum_s |G_{k,s}|^r \right)^{1/r} + \sum_k \left( \sum_s |G_{k,s}|^r \right)^{1/r},
\]

where the sums are restricted to values of \( k, s \) such that \( I_{k,s}, i_{k,s} \subseteq J \). We let \( k^* \) again denote the level of \( J \), so we are only summing over values \( k \geq k^* \).

By the triangle inequality, we have

\[
\left\| \sum_k \left( \sum_s |G_{k,s}|^r \right)^{1/r} + \sum_k \left( \sum_s |G_{k,s}|^r \right)^{1/r} \right\|_{L^p} \leq \sum_k \left\| \left( \sum_s |G_{k,s}|^r \right)^{1/r} \right\|_{L^p} + \sum_k \left\| \left( \sum_s |G_{k,s}|^r \right)^{1/r} \right\|_{L^p},
\]

which, by another application of the triangle inequality, is equal to

\[
\sum_k \left( \sum_s |G_{k,s}|^r \right)^{1/r} + \sum_k \left( \sum_s |G_{k,s}|^r \right)^{1/r} \leq \sum_k \left( \sum_s \|G_{k,s}^r\|_{L^p}^{1/r} \right) + \sum_k \left( \sum_s \|G_{k,s}^r\|_{L^p}^{1/r} \right) \leq \sum_k \left( \sum_s \|G_{k,s}^r\|_{L^p} \right)^{1/r} + \sum_k \left( \sum_s \|G_{k,s}^r\|_{L^p} \right)^{1/r} = \sum_k \left( \sum_s \|G_{k,s}^r\|_{L^p} \right)^{1/r} + \sum_k \left( \sum_s \|G_{k,s}^r\|_{L^p} \right)^{1/r}.
\]

Now, using that \( \|G_{k,s}\|_{L^p} \ll \sqrt{p} \|S_{i_{k,s}}\|_{L^2} \) and \( \|G_{k,s}\|_{L^p} \ll \sqrt{p} \|S_{i_{k,s}}\|_{L^2} \), by Lemma 16 and \( \|S_{i_{k,s}}\|_{L^2} \ll \|S_{i_{k,s}}\|_{L^2} 2^{-(k-k^*)/2} \) and \( \|S_{i_{k,s}}\|_{L^2} \ll \|S_{i_{k,s}}\|_{L^2} 2^{-(k-k^*)/2} \), we have

\[
\|\tilde{G}_J\|_{L^p} \leq \sum_{k \geq k^*} \left( \sum_s \|G_{k,s}\|_{L^p} \right)^{1/r} + \sum_{k \geq k^*} \left( \sum_s \|G_{k,s}\|_{L^p} \right)^{1/r} \ll \sqrt{p} \|S_J\|_{L^2} \sum_{k \geq k^*} \left( \sum_s 2^{-(k-k^*)/2} \right)^{1/r}.
\]

Since the sum of \( s \) ranges over at most \( 2^{k-k^*} \) values (recall we only include values of \( s \) such that \( I_{k,s} \subseteq J \)) and \( r > 2 \), this is

\[
\ll \sqrt{p} \|S_J\|_{L^2} \sum_{k \geq k^*} 2^{(k-k^*)(r-2)/2} \ll \sqrt{p} \|S_J\|_{L^2}.
\]

It thus follows from Lemma 16 that

\[
\|\tilde{G}_J\|_{q(c)} \ll \|S_J\|_{L^2}
\]

for some positive constant \( c \). Lastly, we have that \( \|\tilde{G}_J\|_q \ll \|\tilde{G}_J\|_{q(c)} \) from the definition of the Orlicz norm.

5. Proof of the main result

We are now ready to prove the following.

Theorem 17. Let \( \Phi := \{ \phi_n(x) \}_{n=1}^N \) be an ONS such that \( \sum_{n=1}^N |\phi_n(x)|^2 \leq N \). Then there exists \( Q \subseteq \mathcal{C}(N) \) with \( \mathbb{P}[Q] \geq 1 - Ce^{-cN^{2/5}} \) such that, for \( O \in Q \), the alternate ONS \( \Phi(O) \) satisfies

\[
\|\mathcal{V}f\|_{L^2} \ll \sqrt{\log \log(N)} \|f\|_{L^2}.
\]
**Proof.** Here we use the mass decomposition (into dyadic subintervals \(I_{k,s}\)) stated previously. We use the following easily verified fact; see Lemma 29 of [Lewko and Lewko 2012a].

**Lemma 18.** For every \(J \subseteq \{N\} \setminus \emptyset\), there exist \(J_\ell, J_r \in A\) and \(i_j \in \{N\}\) such that \(J := J_\ell \cup i_j \cup J_r\) is an interval (that is, \(J_\ell, i_j, J_r\) are adjacent), \(J \subseteq \bar{J}\), and \(M(\bar{J}) \leq 2M(J)\).

Without loss of generality, we set \(||f||_{L^2} = 1\) and we have the pointwise inequality

\[
|\nu^2 f(x)|^2 < \sum_{k,s} |\tilde{S}_{I_{k,s}}\|_{B(I_{k,s})}|^2 + \sum_{k,s} |S_{I_{k,s}}|^2 + \log \log(N),
\]

where \(B(I_{k,s}) \subseteq \mathbb{T}\) is the set such that \(|\tilde{S}_{I_{k,s}}(x)|^2 \geq C \log \log(N)M(I_{k,s})\), for a fixed constant \(C\) whose value will be chosen to be sufficiently large. Appealing to Proposition 15, for each \(I_{k,s}\) we can decompose \(\tilde{S}_{I_{k,s}} = \tilde{G}_{I_{k,s}} + \tilde{E}_{I_{k,s}}\). We then define \(B_G(I_{k,s}) \subseteq \mathbb{T}\) by \(|\tilde{G}_{I_{k,s}}(x)|^2 \geq (C/10) \log \log(N)M(I_{k,s})\) and \(B_E(I_{k,s}) \subseteq \mathbb{T}\) by \(|\tilde{E}_{I_{k,s}}(x)|^2 \geq (C/10) \log \log(N)M(I_{k,s})\).

Clearly \(\sum_{k,s} |S_{I_{k,s}}|^2 \leq 1\) is acceptable, so it suffices to show that

\[
\int_{\mathbb{T}} \sum_{k,s} |\tilde{S}_{I_{k,s}}\|_{B(I_{k,s})}|^2 d\mu \ll 1.
\]

Now, appealing to the decomposition above, we have

\[
\int_{\mathbb{T}} \sum_{k,s} |\tilde{S}_{I_{k,s}}\|_{B(I_{k,s})}|^2 d\mu \ll \int_{\mathbb{T}} \sum_{k,s} |\tilde{G}_{I_{k,s}}\|_{B_G(I_{k,s})}|^2 d\mu + \int_{\mathbb{T}} \sum_{k,s} |\tilde{E}_{I_{k,s}}\|_{B_E(I_{k,s})}|^2 d\mu.
\]

First we estimate

\[
\int_{\mathbb{T}} \sum_{k,s} |\tilde{E}_{I_{k,s}}\|_{B_E(I_{k,s})}|^2 d\mu \ll \int_{\mathbb{T}} \sum_{k,s} |\tilde{E}_{I_{k,s}}\|^2 d\mu.
\]

Employing notation used above, we let \(I_k^a := \{I_{k,s} \ s.t. \ |I_{k,s}| \leq 2^{-k/2}N\}\) and \(I_k^b := \{I_{k,s} \ s.t. \ |I_{k,s}| > 2^{-k/2}N\}\). Thus \(I \in I_k^a\) implies \(|I| \leq 2^{-k/2}N\) and \(|I_k| \leq 2^{k/2}\). We then have

\[
\int_{\mathbb{T}} \sum_{k,s} |\tilde{E}_{I_{k,s}}|^2 d\mu = \int_{\mathbb{T}} \sum_{I_{k,s} \in I_k^a} |\tilde{E}_{I_{k,s}}|^2 d\mu + \int_{\mathbb{T}} \sum_{I_{k,s} \in I_k^b} |\tilde{E}_{I_{k,s}}|^2 d\mu.
\]

Using that \(I \in I_k^a\) implies \(|I| \leq 2^{-k/2}N\), we have \(\int |\tilde{E}_{I_{k,s}}|^2 \ll 2^{-c_k/2} \|S_{I_{k,s}}\|^2 \ll 2^{-c_k/2}. \) Thus

\[
\int_{\mathbb{T}} \sum_{I_{k,s} \in I_k^a} |\tilde{E}_{I_{k,s}}|^2 d\mu \ll \sum_k 2^{-c_k/2} \ll 1.
\]

Next, using that \(|I_k^b| \leq 2^{k/2}\) and \(\int |\tilde{E}_{I_{k,s}}|^2 \ll 2^{-k}, \) we have

\[
\int_{\mathbb{T}} \sum_{I_{k,s} \in I_k^b} |\tilde{E}_{I_{k,s}}|^2 \ll \sum_k 2^{-k/2} \ll 1.
\]

Finally, we estimate

\[
\int_{\mathbb{T}} \sum_{k,s} |\tilde{G}_{I_{k,s}}\|_{B_G(I_{k,s})}|^2 d\mu.
\]
We can choose $C$ sufficiently large so that $|B_G(I_{k,s})| \ll 1/\log^{10}(N)$ for all $k$, $s$ (here, $|B_G(I_{k,s})|$ denotes the $\mu$-measure). To see this, recall that $\|\tilde{G}_{I_{k,s}}\|_{\ell_b(c)} \ll \sqrt{M(I_{k,s})}$. By Lemma 16, there exists a constant $c' > 0$ such that

$$\mu(|\tilde{G}_{I_{k,s}}| \geq \lambda) \ll e^{-c'\lambda^2/M(I_{k,s})}$$

for all $\lambda \geq 0$. Setting $\lambda^2 = (C/10) \log \log(N)M(I_{k,s})$, we obtain

$$|B_G(I_{k,s})| \ll \log(N)^{-c'C/10}.$$

We can then choose $C$ sufficiently large with respect to $c'$ to make this estimate $\ll 1/\log^{10}(N)$.

Now we split the sum at $k = 100 \log(N)$ so

$$\int_T \sum_{k,s} \tilde{G}_{I_{k,s},I_{k,s}} d\mu = \int_T \sum_{k,s, k \geq 100 \log(N)} |\tilde{G}_{I_{k,s},I_{k,s}}|^2 d\mu + \int_T \sum_{k,s, k < 100 \log(N)} |\tilde{G}_{I_{k,s},I_{k,s}}|^2 d\mu.$$

By the Cauchy–Schwarz inequality,

$$\int_T \sum_{k,s, k < 100 \log(N)} |\tilde{G}_{I_{k,s},I_{k,s}}|^2 d\mu \ll \sum_{k,s, k < 100 \log(N)} \|\tilde{G}_{I_{k,s},I_{k,s}}\|^2_4 \|1_{B_G(I_{k,s})}\|_4^2.$$

Now, by Lemma 16, we have $\|\tilde{G}_{I_{k,s},I_{k,s}}\|^2_4 \ll \|S_{I_{k,s}}\|^2_2 \ll 2^{-k}$, and, by the previous estimate, $\|1_{B_G(I_{k,s})}\|_4^2 \ll 1/\log^5(N)$. Thus we have shown that the quantity above is

$$\ll \frac{1}{\log^5(N)} \sum_{k,s, k < 100 \log(N)} \|\tilde{G}_{I_{k,s},I_{k,s}}\|^2_4 \ll \frac{1}{\log^4(N)} \ll 1.$$

Lastly, let $T \subset \{N\}$ denote the set of indices appearing in some $I_{k,s}$ for $k \geq 100 \log(N)$. Note that any index will appear in at most $N$ such intervals, and that $M(I_{k,s}) \leq N^{-100}$ if $k \geq 100 \log(N)$. Thus $|a_n|^2 \ll N^{-100}$ for $n \in T$. Thus we have

$$\int_T \sum_{k,s, k \geq 100 \log(N)} |\tilde{G}_{I_{k,s},I_{k,s}}|^2 d\mu \ll N^2 \int_T \sum_{n \in T} |a_n\phi_n(x)|^2 d\mu \ll N^{-98} \int_T \sum_{n \in T} |\phi_n(x)|^2 d\mu \ll 1.$$

This completes the proof. \qed

References


Received 10 May 2012. Revised 26 Feb 2013. Accepted 3 Apr 2013.

ALLISON L EWKO: alewko@cs.columbia.edu

Department of Computer Science, Columbia University, 1214 Amsterdam Avenue, MC 0401, New York, NY 10027, United States

MARK L EWKO: mlewko@math.ucla.edu

Department of Mathematics, University of California, Los Angeles, 520 Portola Plaza, Math Sciences Building 6363, Los Angeles, CA 90095, United States
**Global well-posedness of slightly supercritical active scalar equations**

Michael Dabkowski, Alexander Kiselev, Luis Silvestre and Vlad Vicol

**The nonlinear Schrödinger equation ground states on product spaces**

Susanna Terracini, Nikolay Tzvetkov and Nicola Visciglia

**Orthonormal systems in linear spans**

Allison Lewko and Mark Lewko

**A partial data result for the magnetic Schrödinger inverse problem**

Francis J. Chung

**Sharp polynomial decay rates for the damped wave equation on the torus**

Nalini Anantharaman and Matthieu Léautaud

**The $J$-flow on Kähler surfaces: a boundary case**

Hao Fang, Mejia Lai, Jian Song and Ben Weinkove

**A priori estimates for complex Hessian equations**

Sławomir Dinew and Sławomir Kołodziej

**The Aharonov–Bohm effect in spectral asymptotics of the magnetic Schrödinger operator**

Gregory Eskin and James Ralston