

FOUR VARIANTS OF THE FOURIER-ANALYTIC TRANSFERENCE PRINCIPLE

SEAN PRENDIVILLE

ABSTRACT. We survey four instances of the Fourier analytic ‘transference principle’ or ‘dense model lemma’, which allows one to approximate an unbounded function on the integers by a bounded function with similar Fourier transform. Such a result forms a component of a general method pioneered by Green to count solutions to a single linear equation in a sparse subset of integers.

CONTENTS

1. Introduction	1
2. Green’s L^∞ -bounded approximation lemma	6
3. Helfgott and De Roton’s L^2 -bounded approximation lemma	8
4. Naslund’s L^k -bounded approximation lemma	10
5. The Hahn–Banach approach	13
Appendix A. The large spectrum and Bohr sets	16
Appendix B. The supporting hyperplane theorem	17
Appendix C. The semi-finite minimax theorem	21
Appendix D. The Weierstrass polynomial approximation theorem	23
References	24

1. INTRODUCTION

1.1. **Aim.** There has been much recent work on counting arithmetic configurations in a sparse set of integers, such as the set of primes [Gre05, GT08, GT10], smooth numbers [Har], random sets [CG, Sch], pseudorandom sets [CFZ], or dense subsets thereof. Given such a sparse set, it is often useful to be able to construct a dense subset of integers whose arithmetic properties resemble those of the sparse set, the theory being much more developed in the dense regime, with recourse to powerful results such as Szemerédi’s theorem and affiliated techniques.

When counting solutions to a single linear equation, the arithmetic closeness of the dense model set to our original sparse set can be measured by the level of similarity in their Fourier transform, provided that we weight the characteristic function of our sparse set suitably. The sparseness of our set forces this weight function to grow asymptotically, so we are left with the problem of approximating an unbounded function by a bounded function, with the closeness of approximation

Date: July 14, 2017.

The author was supported by Tim Browning’s ERC grant 306457. Thanks to Tim for many useful conversations.

measured by the L^∞ -norm of their Fourier transform. The purpose of this note is to survey four variants of such a *bounded approximation lemma*, also called a *transference principle* or *dense model lemma* in the literature: the original found in Green [Gre05], a quantitative improvement due to Helfgott–De Roton [HDR11], a further quantitative refinement due to Naslund [Nas15], and finally a much more general technique due (independently) to Gowers [Gow10] and Reingold et al. [RTTV]. Our focus is on the quantitative strength of each of these results. We give a complete account of the required background in the appendices.

The Fourier-analytic transference principle is particularly powerful when combined with the Hardy–Littlewood circle method. Traditionally, the circle method is performed with respect to a function defined on the integers, whose Fourier transform is then defined on the circle group $\mathbb{T} = \mathbb{R}/\mathbb{Z}$. In the majority of the references we survey, the Fourier analysis is performed with respect to functions defined on the integers modulo a large prime number. This has the expositional advantage that both physical and phase space are discrete, and in fact isomorphic. However, we believe this reduction is artificial, and in order to highlight the utility of the transference principle within the traditional number-theoretic circle method, we opt to give all proofs with respect to Fourier analysis on the integers.

1.2. Motivation: a sparse version of Roth’s theorem. A theorem of Roth [Rot53, Rot54] quantifies the density required of a set of integers to ensure that it contains a non-trivial solution to a single linear equation

$$c_1x_1 + \cdots + c_sx_s = 0. \tag{1}$$

Assuming the coefficients sum to zero, a variant of this theorem due to Bloom [Blo12] states that for any $\delta > 0$ there exists $c(\delta) > 0$ such that if A is a subset of $[N]$ of density at least δ (i.e. $|A| \geq \delta N$), then A contains many solutions to the equation, in that

$$\sum_{\mathbf{c} \cdot \mathbf{x} = 0} 1_A(x_1) \cdots 1_A(x_s) \geq c_0(\delta)N^{s-1}, \tag{2}$$

where one may take

$$c_0(\delta) \gg_{\mathbf{c}} \exp\left(-C/\delta^{\frac{1}{s-2-\varepsilon}}\right) \tag{3}$$

for some absolute constant $C = C(s, \varepsilon)$ and any $\varepsilon > 0$.

Roth’s method for proving such a result proceeds by exploiting the orthogonality relation

$$\sum_{\mathbf{c} \cdot \mathbf{x} = 0} 1_A(x_1) \cdots 1_A(x_s) = \int_{\mathbb{T}} \hat{1}_A(c_1\alpha) \cdots \hat{1}_A(c_s\alpha) d\alpha, \tag{4}$$

where we define the Fourier transform of a function $f : \mathbb{Z} \rightarrow \mathbb{C}$ of finite support by

$$\hat{f}(\alpha) := \sum_n f(n)e(\alpha n). \tag{5}$$

If we know the distribution of A in arithmetic progressions, then the classical circle method allows us predict the behaviour of $\hat{1}_A$. Roth’s argument says that either A is equidistributed in arithmetic progressions, in which case we can calculate $\hat{1}_A$ and therefore (4), or alternatively A is biased towards at least one arithmetic progression. Exploiting this bias then forms the so-called ‘density increment’ argument. We

refrain from the details here, but hope to convey to the reader the sense that if one knows how the Fourier transform $\hat{1}_A$ behaves, then one can count solutions to a linear equation in A .

Suppose that we wish to prove an analogue of Roth's theorem for subsets of the integers which are not dense in the interval $[N]$, but are dense in some fixed sparse subset $S \subset [N]$, so that $|S| = o(N)$. For example, one may take S to be of arithmetic nature, such as the set of primes or the set of squares, or even the set of squares of primes. Alternatively, one could take S to be a random subset of $[N]$.

Given a subset A of our sparse set S which is *relatively dense*, in the sense that $|A| \geq \delta|S|$, we wish to prove a lower bound of the form (2). Notice that if we could construct a *dense* subset $B \subset [N]$, with $|B| \geq \delta^c N$ say, and such that we have the Fourier approximation

$$\hat{1}_A \approx \hat{1}_B, \tag{6}$$

then we can employ Roth's theorem to obtain a lower bound for the number of solutions to (1) in A as follows

$$\begin{aligned} \sum_{\mathbf{c} \cdot \mathbf{x} = 0} 1_A(x_1) \cdots 1_A(x_s) &= \int_{\mathbb{T}} \hat{1}_A(c_1 \alpha) \cdots \hat{1}_A(c_s \alpha) d\alpha \\ &\approx \int_{\mathbb{T}} \hat{1}_B(c_1 \alpha) \cdots \hat{1}_B(c_s \alpha) d\alpha \\ &= \sum_{\mathbf{c} \cdot \mathbf{x} = 0} 1_B(x_1) \cdots 1_B(x_s) \\ &\geq c(\delta^c) N^{s-1}. \end{aligned} \tag{7}$$

Obtaining an approximation such as (6) is the strategy of the Fourier-analytic transference principle, or dense model lemma, originating in Green [Gre05]. However, as stated, such an approximation is too much to hope for. Looking at the Fourier transform evaluated at $\alpha = 0$, we would deduce that

$$|S| \geq |A| = \hat{1}_A(0) \approx \hat{1}_B(0) \geq \delta^c N. \tag{8}$$

This would then imply that S is itself a dense subset of the interval $[N]$. To get around this, we weight the indicator function of A in order to ensure that we have some hope of approximating its Fourier transform by the Fourier transform of a dense set. In order to deal with arbitrary relatively dense subsets of S , it makes sense to choose this weighting independently of the set A itself.

Definition 1.1 (Majorant). Given $S \subset [N]$ define a *majorant on S* to be a non-negative function $\nu : \mathbb{Z} \rightarrow [0, \infty)$ with support contained in S and such that

$$\sum_n \nu(n) = (1 + o(1))N. \tag{9}$$

Given a majorant ν on S and $A \subset S$ with relative density $|A| \geq \delta|S|$, define

$$f := 1_A \nu.$$

Provided that we choose our majorant sensibly, we should be able to prove that

$$\sum_n f(n) \geq \delta^c N \tag{10}$$

for some $c > 0$. We therefore hope to obtain a Fourier approximation of the form

$$\hat{f} \approx \hat{1}_B,$$

for some $B \subset [N]$ which is suitably dense $|B| \geq \delta^c N$. In fact, we do not need our dense approximant to be the characteristic function of a set; it suffices for the function to have bounded L^2 -norm, an observation first recorded by Helfgott and De Roton [HDR11].

Lemma 1.2 (L^2 -boundedness suffices). *Let $c_1 + \dots + c_s = 0$. Then for any $\delta > 0$ and any constant C there exists $c(\delta, C) > 0$ such that the following holds. Suppose that $g : \mathbb{Z} \rightarrow [0, \infty)$ is a non-negative function supported on $[N]$ which has bounded¹ L^2 -norm*

$$\sum_n g(n)^2 \leq CN. \quad (11)$$

Then the density assumption

$$\sum_n g(n) \geq \delta N \quad (12)$$

implies that

$$\sum_{\mathbf{c} \cdot \mathbf{x} = 0} g(x_1) \cdots g(x_s) \geq c(\delta, C) N^{s-1}.$$

In fact, one may take

$$c(\delta, C) = (\delta/2)^s c_0(\delta^2/(4C)), \quad (13)$$

where $c_0(\delta)$ is the constant appearing in Roth's theorem (2).

Proof. Define

$$B := \{x \in [N] : g(x) \geq \delta/2\}.$$

Then, employing the Cauchy–Schwarz inequality, we have

$$\begin{aligned} \delta N &\leq \sum_x g(x) = \sum_{x \notin B} g(x) + \sum_{x \in B} g(x) \\ &\leq \frac{1}{2} \delta N + |B|^{1/2} \left(\sum_x g(x)^2 \right)^{1/2} \\ &\leq \frac{1}{2} \delta N + (|B|CN)^{1/2}. \end{aligned} \quad (14)$$

Therefore

$$|B| \geq \frac{\delta^2}{4C} N. \quad (15)$$

Applying Bloom's variant of Roth's theorem, we deduce that

$$\sum_{\mathbf{c} \cdot \mathbf{x} = 0} 1_B(x_1) \cdots 1_B(x_s) \geq c_0(\delta^2/(4C)) N^{s-1}.$$

Hence

$$\sum_{\mathbf{c} \cdot \mathbf{x}} g(x_1) \cdots g(x_s) \geq (\delta/2)^s \sum_{\mathbf{c} \cdot \mathbf{x}} 1_B(x_1) \cdots 1_B(x_s) \geq (\delta/2)^s c_0(\delta^2/(4C)) N^{s-1}.$$

□

¹Although this estimate for the L^2 -norm appears to grow with N , it is the same estimate one would obtain for a bounded function on $[N]$.

Let us sketch how this result, when combined with a transference principle, allows one to extract a quantitative bound on the relative density of a subset $A \subset S \subset [N]$ lacking non-trivial solutions to (1). Write $\delta := |A|/|S|$ for the relative density of A in S . Then provided that one has made a sensible choice for the weighted majorant ν on S , one should have

$$\sum_n 1_A(n)\nu(n) \geq \delta^c N,$$

for some absolute $c > 0$. Applying a transference principle to the function $f = 1_A \nu$, one obtains an approximant g supported on $[N]$ with bounded L^2 -norm of the form (11) and such that $\hat{f} \approx \hat{g}$ uniformly on \mathbb{T} . Performing an approximation similar to (7) and applying Lemma 1.2 yields

$$\sum_{\mathbf{c} \cdot \mathbf{x} = 0} f(x_1) \cdots f(x_s) \geq c(\delta^c, C) N^{s-1}. \quad (16)$$

Yet if A contains only trivial solutions to (1), we have

$$\sum_{\mathbf{c} \cdot \mathbf{x} = 0} f(x_1) \cdots f(x_s) \leq \sum_{\substack{\mathbf{c} \cdot \mathbf{x} = 0 \\ \mathbf{x} \text{ trivial}}} \nu(x_1) \cdots \nu(x_s). \quad (17)$$

There are various possible candidates for what should constitute a trivial solution to (1), one such choice being that \mathbf{x} belongs to one of a finite collection of proper subspaces of the hyperplane $\mathbf{c} \cdot \mathbf{x} = 0$. Whatever choice of triviality one makes, one would expect that the trivial solutions should be a sparse subset of the solution space, so that

$$\sum_{\substack{\mathbf{c} \cdot \mathbf{x} = 0 \\ \mathbf{x} \text{ trivial}}} 1_{[N]}(x_1) \cdots 1_{[N]}(x_s) \leq \frac{N^{s-1}}{\omega(N)}$$

for some function $\omega(N) \rightarrow \infty$. Moreover, a sensible choice of majorant should respect this sparseness, so that

$$\sum_{\substack{\mathbf{c} \cdot \mathbf{x} = 0 \\ \mathbf{x} \text{ trivial}}} \nu(x_1) \cdots \nu(x_s) \ll \frac{N^{s-1}}{\omega(N)}.$$

Combining this with (16) and (17) yields

$$c(\delta, C) \ll \frac{1}{\omega(N)}.$$

Using the lower bounds (3) and (13) then allows us to extract an upper bound on δ in terms of $\omega(N)^{-1}$. For instance, if $C = O(1)$ then one has

$$c(\delta, C) \gg_c \exp(-C_{s,\varepsilon} \delta^{-\frac{2}{s-2-\varepsilon}}),$$

which implies that

$$\delta \ll_{c,\varepsilon} (\log \omega(N))^{-\frac{s}{2} + 1 + \varepsilon}. \quad (18)$$

In view of Lemma 1.2 and the discussion which precedes it, our aim in the remainder of this note is to provide sufficient conditions a majorant ν should satisfy to ensure that if $0 \leq f \leq \nu$ with $\sum_n f(n) \geq \delta N$ then there exists a function g which is dense (as in (12)), which has bounded L^2 -norm (as in (11)), and such that

$\|\hat{f} - \hat{g}\|_\infty$ is small. By comparing Fourier transforms at the zero frequency, one sees that non-negative functions which are close in the L^∞ -Fourier norm are also close in the L^1 -norm (an observation previously employed in (8)), so that

$$\sum_n g(n) = \sum_n f(n) + O\left(\|\hat{f} - \hat{g}\|_\infty\right).$$

Hence the density of f automatically implies the density of g . We may therefore drop the requirement that our approximant g is dense, as this follows from the Fourier approximation. Our aim is therefore to answer the following question.

Question. *What conditions does a majorant ν on $[N]$ need to satisfy in order to ensure that any function $0 \leq f \leq \nu$ has a non-negative approximant g with bounded L^2 -norm and such that the difference $\|\hat{f} - \hat{g}\|_\infty$ is small?*

Any result which provides conditions answering this question we call a *bounded approximation lemma*, since we are attempting to approximate our unbounded function f by a function g which exhibits less growth, as measured by the L^2 -norm.

1.3. Notation. In order to be consistent with the normalisation of our Fourier transform (5), we define the L^p -norm of a function on the integers $f : \mathbb{Z} \rightarrow \mathbb{C}$ with respect to counting measure, so that

$$\|f\|_p := \left(\sum_n |f(n)|^p \right)^{1/p}.$$

For functions on \mathbb{T} , all L^p -norms are taken with respect to the Haar probability measure, so that for finitely supported $f : \mathbb{Z} \rightarrow \mathbb{C}$ we have

$$\|\hat{f}\|_2 = \|f\|_2$$

Notice that if ν is a majorant then we also have the identity

$$\|\hat{\nu}\|_\infty = \|\nu\|_1. \tag{19}$$

2. GREEN'S L^∞ -BOUNDED APPROXIMATION LEMMA

In this section we give a proof of perhaps the simplest bounded approximation lemma, originating in Green [Gre05]. Not only does this yield an approximant with bounded L^2 -norm, but also bounded L^∞ -norm, so in some sense this approximant has the best possible boundedness properties. The price to be paid for such good boundedness is the quality of our final Fourier approximation $\hat{f} \approx \hat{g}$.

Definition 2.1 (Fourier decay). We say that a majorant ν on $[N]$ has *Fourier decay of level θ* if

$$\left\| \hat{\nu} - \hat{1}_{[N]} \right\|_\infty \leq \theta N.$$

Notice from (19), that if a majorant has Fourier decay of level θ then

$$\|\nu\|_1 = N + O(\theta N).$$

Definition 2.2 (Restriction at p). We say that a majorant ν supported on $[N]$ satisfies a restriction estimate at exponent p if

$$\sup_{|\phi| \leq \nu} \int_{\mathbb{T}} |\hat{\phi}(\alpha)|^p d\alpha \ll_p \|\nu\|_1^p N^{-1}.$$

Theorem 2.1 (Green [Gre05]). Suppose that the majorant ν has Fourier decay of level θ and satisfies a restriction estimate at exponent p . Then for any $0 \leq f \leq \nu$ there exists $0 \leq g \ll 1_{[N]}$ such that

$$\|\hat{f} - \hat{g}\|_{\infty} \ll_p \log(\theta^{-1})^{-\frac{1}{p+2}} N.$$

As the function g delivered by this theorem is genuinely bounded, we call this an L^{∞} -bounded approximation lemma.

We begin the proof of Theorem 2.1 by defining the large spectrum of f to be the set

$$\text{Spec}(f, \eta \|\nu\|_1) := \left\{ \alpha \in \mathbb{T} : |\hat{f}(\alpha)| \geq \eta \|\nu\|_1 \right\}.$$

Define the Bohr set with frequency set $S := \text{Spec}(f, \eta \|\nu\|_1)$ and width $\varepsilon \leq 1/2$ by

$$B(S, \varepsilon) := \{n \in [-\varepsilon N, \varepsilon N] : \|n\alpha\| \leq \varepsilon \quad (\forall \alpha \in S)\}. \quad (20)$$

In this definition we have used $\|\cdot\|$ to denote the distance to the nearest integer (not to be confused with the plethora of norms employed in this paper). Write σ for the normalised characteristic function of $B := B(S, \varepsilon)$, so that

$$\sigma := |B|^{-1} 1_B.$$

Then we define

$$g := f * \sigma * \sigma, \quad (21)$$

where, for finitely supported f_i , we set

$$f_1 * f_2(n) := \sum_{m_1 + m_2 = n} f_1(m_1) f_2(m_2).$$

We first estimate $|\hat{f} - \hat{g}|$. The key identity is

$$\widehat{f_1 * f_2} = \hat{f}_1 \hat{f}_2.$$

If $\alpha \notin \text{Spec}(f, \eta N)$ then we have

$$|\hat{f}(\alpha) - \hat{g}(\alpha)| = |\hat{f}(\alpha)| |1 - \hat{\sigma}(\alpha)^2| \leq 2\eta \|\nu\|_1 \ll \eta N.$$

If $\alpha \in \text{Spec}(f, \eta N)$, then for each $n \in B$ we have $e(\alpha n) = 1 + O(\varepsilon)$. Hence $\hat{\sigma}(\alpha) = 1 + O(\varepsilon)$, and consequently

$$|\hat{f}(\alpha) - \hat{g}(\alpha)| = |\hat{f}(\alpha)| |1 + \hat{\sigma}(\alpha)| |1 - \hat{\sigma}(\alpha)| \ll \|\nu\|_1 \varepsilon \ll \varepsilon N.$$

Combining both cases gives

$$\|\hat{f} - \hat{g}\|_{\infty} \ll (\varepsilon + \eta) N. \quad (22)$$

It remains to show that g is bounded. By positivity and orthogonality, we have

$$\begin{aligned} g(n) &= \sum_{x+y+z=n} f(x)\sigma(y)\sigma(z) \leq \sum_{x+y+z=n} v(x)\sigma(y)\sigma(z) \\ &= \int_{\mathbb{T}} \hat{v}(\alpha)\hat{\sigma}(\alpha)^2 e(-\alpha n) d\alpha. \end{aligned}$$

It therefore suffices to show that

$$\int_{\mathbb{T}} \hat{v}(\alpha)\hat{1}_B(\alpha)^2 e(-\alpha n) d\alpha \ll |B|^2. \quad (23)$$

Inserting our Fourier decay assumption and using Parseval, we have

$$\begin{aligned} \int_{\mathbb{T}} \hat{v}(\alpha)\hat{1}_B(\alpha)^2 e(-\alpha n) d\alpha &\leq \int_{\mathbb{T}} \hat{1}_{[N]}(\alpha)\hat{1}_B(\alpha)^2 e(-\alpha n) d\alpha + \theta N \int_{\mathbb{T}} |\hat{1}_B(\alpha)|^2 d\alpha \\ &= \sum_{x+y+z=n} 1_{[N]}(x)1_B(y)1_B(z) + \theta N|B| \\ &\leq |B|^2 + \theta N|B|. \end{aligned}$$

We therefore obtain (23) provided that

$$\theta N \ll |B|. \quad (24)$$

By Lemma A.2 we have $|B| \geq \varepsilon^{O_p(\eta^{-p-1})}N$, so (24) follows provided that $\theta \leq \varepsilon^{C_p\eta^{-1-p}}$. In view of (22), let us take $\varepsilon = \eta$ with $\theta = \varepsilon^{C_p\eta^{-1-p}}$. Then

$$\log(\theta^{-1}) \leq C_p \log(\varepsilon^{-1})\varepsilon^{-1-p} \ll_p \varepsilon^{-2-p}.$$

This implies that

$$\|\hat{f} - \hat{g}\|_{\infty} \ll \varepsilon N \ll_p \left(\log(\theta^{-1})\right)^{-\frac{1}{2+p}} N,$$

which completes the proof of Theorem 2.1.

3. HELFGOTT AND DE ROTON'S L^2 -BOUNDED APPROXIMATION LEMMA

For quantitative applications, a drawback of Green's bounded approximation lemma is the dependence of the final Fourier bound $\|\hat{f} - \hat{g}\|_{\infty}$ on the level of Fourier decay θ exhibited by the majorant v . Typically our majorant satisfies a Fourier decay assumption of the form

$$\|\hat{v} - \hat{1}_{[N]}\|_{\infty} \ll N(\log N)^{-c}. \quad (25)$$

This results in a final Fourier bound of the form

$$\|\hat{f} - \hat{g}\|_{\infty} \ll N(\log \log N)^{-\frac{1}{p+2}}. \quad (26)$$

Notice that this approximation is one logarithmic iteration worse than our assumed Fourier bound (25), even when $f = v$, where we may take $g = 1_{[N]}$.

In the process of improving Green's bound [Gre05] for Roth's theorem in the primes, Helfgott and De Roton [HDR11] developed a new variant of the bounded approximation lemma which removes this logarithmic loss from the final Fourier bound. There is a price to be paid for this improvement. The first is that the approximant may no longer be an L^{∞} -bounded function, but instead has the weaker property of being bounded in the L^2 -norm. However, as Helfgott and De Roton

observed in Lemma 1.2, this is not really an impediment. A more serious price must be paid in making a stronger assumption on their majorant ν than Fourier decay.

Definition 3.1 (Two-point correlation estimate). Let us say that a majorant satisfies a *two point correlation estimate* if for any non-zero m we have

$$\sum_n \nu(n)\nu(n+m) \ll N. \quad (27)$$

Definition 3.2 (L^2 -boundedness of level θ). We say that a majorant ν on $[N]$ has L^2 -boundedness of level θ if

$$\sum_n \nu(n)^2 \leq \theta N^2. \quad (28)$$

Notice that if a majorant satisfies the L^∞ -bound $\nu \leq \theta N$, then the L^1 -assumption (9) gives L^2 -boundedness of level θ .

Theorem 3.1 (Helfgott and De Roton [HDR11]). *Suppose that the majorant ν satisfies a restriction estimate at exponent p , a two-point correlation estimate and has L^2 -boundedness of level θ . Then for any $0 \leq f \leq \nu$ there exists $g \geq 0$ such that $\sum_n g(n)^2 \ll N$ and*

$$\|\hat{f} - \hat{g}\|_\infty \ll_p \log(\theta^{-1})^{-\frac{1}{p+2}} N.$$

In applications the θ parameter resulting from the level of L^2 -boundedness (31) is of the form N^{-c} for some absolute constant $c > 0$. In practice, this is much smaller than the Fourier decay parameter θ that one might hope to obtain for ν , which is usually of the form $(\log N)^{-c}$. This results in a final Fourier approximation of the form

$$\|\hat{f} - \hat{g}\|_\infty \ll N(\log N)^{-\frac{1}{p+2}},$$

which saves a logarithm over the estimate given in (26).

The proof of Theorem 3.1 is similar to that given in §2. Adopting the notation of §2, we define g as in (21), albeit with one less convolution

$$g := f * \sigma.$$

The same argument given in §2 gives the Fourier bound

$$\|\hat{f} - \hat{g}\|_\infty \ll (\varepsilon + \eta)N,$$

so we take $\varepsilon = \eta$ to yield $\|\hat{f} - \hat{g}\|_\infty \ll \varepsilon N$.

Now our treatment departs from that given previously as we are aiming to prove the L^2 -bound $\sum_n g(n)^2 \ll N$, which is equivalent to

$$\sum_{n_1 - n_2 = m_1 - m_2} f(n_1)f(n_2)1_B(m_1)1_B(m_2) \ll N|B|^2.$$

Utilising $f \leq \nu$, this equals

$$\sum_m \left(\sum_{m_1 - m_2 = m} 1_B(m_1)1_B(m_2) \right) \left(\sum_n \nu(n)\nu(n+m) \right).$$

Incorporating our assumptions (27) and (31), this is at most

$$\theta N^2|B| + O(N|B|^2).$$

We have therefore obtained L^2 -boundedness provided that $\theta N \ll |B|$. Recalling Lemma A.2, it suffices to have

$$\theta \leq \varepsilon^{O_p(\varepsilon^{-p-1})},$$

or equivalently

$$C_p \log(\theta^{-1})^{-\frac{1}{p+2}} \leq \varepsilon.$$

Taking the smallest permissible value of ε then yields Theorem 3.1.

4. NASLUND'S L^k -BOUNDED APPROXIMATION LEMMA

As is apparent in the deduction of the density bound (18), if one is interested in quantitative bounds for sets lacking solutions to (1), then the quantitative dependence in (13) is important. Ideally, one would hope not to lose too much by passing from the constant $c(\delta)$ available for the characteristic function of a dense set, to the constant $c(\delta, C)$ available for a function with bounded L^2 -norm. In a perfect world, this loss would take the form, say

$$c(\delta, C) = c\left(\frac{\delta}{100C}\right),$$

whereas the proof of Lemma 1.2 yields

$$c(\delta, C) = (\delta/2)^s c\left(\frac{\delta^2}{4C}\right).$$

The occurrence of the factor δ^s in (13) seems unavoidable. Fortunately, this factor is not too costly, since it is much larger than the lower bound (3) for $c(\delta)$. A more significant loss is the appearance of δ^2 within the function $c\left(\frac{\delta^2}{4C}\right)$, which ultimately stems from the lower bound (15).

As observed by Naslund [Nas15], one may replace the use of Cauchy–Schwarz in (14) by Hölder's inequality in order to replace the occurrence of δ^2 by, essentially, $\delta^{1+\varepsilon}$. This improvement ultimately stems from aiming for an L^k -bounded approximant for some large k (depending on ε), rather than the weaker L^2 -approximant of Helfgott and De Roton. Since the L^k -norm of a finitely supported function tends to the L^∞ -norm with k , one may think of L^k -boundedness as a half-way house between the weak notion of L^2 -boundedness and the strong L^∞ -notion.

Lemma 4.1 (L^k -boundedness suffices). *Let $c_1 + \dots + c_s = 0$. Then for any $\delta > 0$, any constant C and any $k \geq 2$ there exists $c(\delta, C, k) > 0$ such that the following holds. Suppose that $g : \mathbb{Z} \rightarrow [0, \infty)$ is a non-negative function supported on $[N]$ which has bounded L^k -norm*

$$\sum_n g(n)^k \leq CN.$$

Then the density assumption $\sum_n g(n) \geq \delta N$ implies that

$$\sum_{\mathbf{c} \cdot \mathbf{x} = 0} g(x_1) \cdots g(x_s) \geq c(\delta, C, k) N^{s-1},$$

Moreover, one may take

$$c(\delta, C, k) = (\delta/2)^s c_0 \left(\left(\frac{\delta}{2C} \right)^{1 + \frac{1}{k-1}} \right), \quad (29)$$

where $c_0(\delta)$ is the constant appearing in Roth's theorem (2).

Proof. We proceed as in Helfgott and De Roton's argument for Lemma 1.2, albeit using Hölder's inequality to give the upper bound

$$\sum_{x \in B} g(x) \leq |B|^{1-\frac{1}{k}} \left(\sum_x g(x)^k \right)^{\frac{1}{k}}.$$

This results in the lower bound

$$|B| \geq \left(\frac{\delta}{2C} \right)^{1+\frac{1}{k-1}} N,$$

from which (29) follows. □

The price to paid for obtaining an approximant with the stronger notion of L^k -boundedness is that one's majorant must now satisfy a more stringent correlation condition.

Definition 4.2 (*k*-point correlation estimates). Let us say that a majorant satisfies the *k*-point correlation estimates if for any distinct m_1, \dots, m_l with $l \leq k$ we have

$$\sum_n v(n + m_1) \cdots v(n + m_l) \ll N. \tag{30}$$

Definition 4.3 (L^∞ -boundedness of level θ). We say that a majorant v on $[N]$ has L^∞ -boundedness of level θ if for all n we have

$$v(n) \leq \theta N. \tag{31}$$

By assumption a majorant satisfies $\sum_n v(n) = (1 + o(1))N$, so that the level of L^∞ -boundedness is at worst $O(1)$, and unless v is concentrated on a bounded set, will be $o(1)$ in applications.

Theorem 4.1 (Naslund [Nas15]). Suppose that v is a majorant on $[N]$ satisfying a restriction estimate at exponent p , with L^∞ -boundedness of level θ and satisfying the *k*-point correlation estimates with

$$k \leq \frac{1}{2} \sqrt{\log(\theta^{-1})}. \tag{32}$$

Then for any $0 \leq f \leq v$ there exists $g \geq 0$ such that $\sum_n g(n)^k \ll N$ and

$$\|\hat{f} - \hat{g}\|_\infty \ll_p \log(\theta^{-1})^{-\frac{1}{p+2}} N.$$

In order to employ this result in conjunction with Lemma 4.1, one might hope, in view of (29), to take $k = \lceil 1 + \varepsilon^{-1} \rceil$. As mentioned previously, in applications we expect to be able to obtain L^∞ -boundedness of level N^{-c} . Hence (32) certainly follows if $N \geq \exp(C\varepsilon^{-2})$. Provided that one can prove the $(1 + \varepsilon^{-1})$ -point correlation estimates, one may then deduce a lower bound in (29) of the form

$$c(\delta, C) \gg_\varepsilon (\delta/2)^s c \left(\left(\frac{\delta}{2C} \right)^{1+\varepsilon} \right).$$

Proof. The construction is the same as in the proof of Theorem 3.1. Just as in that proof we take $\eta = \varepsilon$ to obtain an approximant $g \geq 0$ with $\|\hat{f} - \hat{g}\|_\infty \ll \varepsilon N$. Our task then reduces to determining a permissible value of ε which allows one to show that

$$\sum_n \left(\sum_m \nu(n-m) 1_B(m) \right)^k \ll N|B|^k.$$

Expanding out the k th power and noting that (by definition (20) of the Bohr set) $B = -B$, this is equivalent to the estimate

$$\sum_{m_1, \dots, m_k \in B} \sum_n \nu(n+m_1) \cdots \nu(n+m_k) \ll N|B|^k. \quad (33)$$

Fix a choice of $(m_1, \dots, m_k) \in B^k$ and let (m'_1, \dots, m'_l) denote the distinct values occurring in this choice, written in the order in which they appear in the tuple, and with respective multiplicities k_1, \dots, k_l . Then by the level of L^∞ -boundedness and the k -point correlation estimate, we have

$$\begin{aligned} \sum_n \nu(n+m_1) \cdots \nu(n+m_k) &= \sum_n \nu(n+m'_1)^{k_1} \cdots \nu(n+m'_l)^{k_l} \\ &\leq (\theta N)^{k-l} \sum_n \nu(n+m'_1) \cdots \nu(n+m'_l) \\ &\ll (\theta N)^{k-l} N. \end{aligned}$$

By choosing one of the symbols '=' or '≠' for each pair of indices $1 \leq i < j \leq k$, we see that for each choice of tuple $(m'_1, \dots, m'_l) \in B^l$ with distinct entries, there are at most $2^{\binom{k}{2}}$ choices of $(m_1, \dots, m_k) \in B^k$ giving rise to (m'_1, \dots, m'_l) . It follows that

$$\begin{aligned} \sum_{m_1, \dots, m_k \in B} \sum_n \nu(n+m_1) \cdots \nu(n+m_k) &\ll \sum_{l=1}^k 2^{\binom{k}{2}} |B|^l (\theta N)^{k-l} N \\ &\leq |B|^k N \max_{1 \leq l \leq k} \left(\frac{k 2^{\binom{k}{2}} \theta N}{|B|} \right)^l \end{aligned}$$

The required bound (33) then follows on ensuring that $|B| \geq k 2^{\binom{k}{2}} \theta N$, which from Lemma A.2, follows if

$$\varepsilon^{O_p(\varepsilon^{-p-1})} \geq k 2^{\binom{k}{2}} \theta.$$

This in turn follows if

$$\log(\theta^{-1}) \geq \log k + \binom{k}{2} \log 2 + C_p \varepsilon^{-p-2}.$$

By (32) and the inequality $\log k + \binom{k}{2} \log 2 \leq k^2$, it suffices to take

$$\varepsilon = \left(\frac{2C_p}{\log(\theta^{-1})} \right)^{\frac{1}{p+2}}.$$

□

5. THE HAHN–BANACH APPROACH

The fact that a majorant satisfies a restriction estimate at some exponent p is essential in applications of the transference principle to the circle method, see for instance [GT06, Har, BP]. In general, if a function f is efficiently bounded by a majorant ν , one can count solutions to a linear equation in s variables weighted by f provided that one can obtain a restriction estimate for ν for some $p < s$.

The limited use of the restriction estimate in the proof of theorems 2.1, 3.1 and 4.1 suggests that it may not be necessary for a majorant ν to satisfy such an estimate in order for $f \leq \nu$ to have a bounded approximation. This was first shown by Gowers [Gow10] and, independently, by Reingold et al [RTTV]. It turns out that removing the quantitative dependence of the final Fourier approximation on the restriction parameter gives a marginally stronger bound. Their method extends to give a bounded approximation lemma for norms other than the L^∞ -Fourier norm, giving an alternative derivation of the transference principle found in [GT08, TZ08], and which is essential for applications to systems of linear equations such as [GT10, Mat12]. In this section we give an exposition of their argument limited to the simpler Fourier-analytic context.

In common with Green’s transference principle, the approximation theorem assumes some level of Fourier decay. Although quantitatively weaker than the assumption of a correlation condition, this is in some sense a more useful assumption for applications, such as Roth-type theorems in k th powers [BP, Cho], where the correlation estimates (27) and (30) do not necessarily hold.

Theorem 5.1 ([Gow10, RTTV]). *Suppose that the majorant ν has Fourier decay of level θ . Then for any $0 \leq f \leq \nu$ there exists a bounded function $0 \leq g \leq 1_{[N]}$ such that*

$$\|\hat{f} - \hat{g}\|_\infty \ll \log(1/\theta)^{-3/2} N.$$

Both [Gow10] and [RTTV] follow similar lines in proving this result, employing either the supporting hyperplane theorem or the minimax theorem to give the existence of g , rather than the explicit construction of §§2–4. Both of these subsidiary results are closely related to the finite dimensional Hahn–Banach theorem. We give a complete account of the necessary background in the appendices.

We identify the set of functions $f : \mathbb{Z} \rightarrow \mathbb{C}$ whose support is contained in $[N]$ with the finite dimensional space \mathbb{C}^N . Then the functional

$$\|f\| := \|\hat{f}\|_\infty$$

forms a norm on this space. Recall that we define the dual norm by

$$\|\phi\|^* := \sup_{\|f\| \leq 1} |\langle f, \phi \rangle|,$$

where

$$\langle f, \phi \rangle := \sum_{n=1}^N f(n) \overline{\phi(n)}.$$

One can check that this is itself a norm on \mathbb{C}^N , and it follows directly from the definition that for any $f, \phi \in \mathbb{C}^N$ we have the inequality

$$|\langle f, \phi \rangle| \leq \|f\| \|\phi\|^*. \quad (34)$$

Lemma 5.1 (Properties of the dual of $\|\hat{f}\|_\infty$).

(i) (Algebra property)

$$\|\phi_1 \phi_2\|^* \leq \|\phi_1\|^* \|\phi_2\|^*. \quad (35)$$

(ii) (L^∞ -compatibility)

$$\|\phi\|_\infty \leq \|\phi\|^*.$$

(iii) (Real compatibility)

$$\|\operatorname{Re} \phi\|^* \leq \|\phi\|^*.$$

(iv) (Duality) For any $f \in \mathbb{C}^N$ there exists $\phi \in \mathbb{C}^N$ with $\|\phi\|^* = 1$ such that

$$\|f\| = \operatorname{Re} \langle f, \phi \rangle.$$

Proof. Let $f \in \mathbb{C}^N$ with $\|f\| \leq 1$. Then by (34) we have

$$|\langle f, \phi_1 \phi_2 \rangle| = |\langle f \bar{\phi}_1, \phi_2 \rangle| \leq \|f \bar{\phi}_1\| \|\phi_2\|^*.$$

For $\alpha \in \mathbb{T}$ write $e_\alpha(n) := e(\alpha n)$. Then

$$\begin{aligned} |\widehat{f \bar{\phi}_1}(\alpha)| &= |\langle f e_\alpha, \phi_1 \rangle| \\ &\leq \|f e_\alpha\| \|\phi_1\|^*. \end{aligned}$$

By a change of variables we have $\|f e_\alpha\| = \|f\| \leq 1$. Thus $\|f \bar{\phi}_1\| \leq \|\phi_1\|^*$, which establishes (i).

To prove (ii) it suffices, by homogeneity, to show that the ball $B^* := \{\phi \in \mathbb{C}^N : \|\phi\|^* \leq 1\}$ is contained in the ball $B_\infty := \{\phi \in \mathbb{C}^N : \|\phi\|_\infty \leq 1\}$. By equivalence of norms on finite dimensional spaces, B^* is a bounded subset of \mathbb{C}^N . Suppose that ϕ is an element of $B^* \setminus B_\infty$, so that $|\phi(n)| > 1$ for some $n \in [N]$. By the algebra property (35), $\phi^k \in B^*$ for all $k \in \mathbb{N}$. Yet $|\phi^k(n)| \rightarrow \infty$ as $k \rightarrow \infty$, contradicting boundedness.

For (iii), we first note that $\|\cdot\|$ is invariant under complex conjugation, since

$$\|\bar{f}\| = \sup_\alpha \left| \sum_n \overline{f(n)} e(\alpha n) \right| = \sup_\alpha \left| \sum_n f(n) e(-\alpha n) \right| = \|f\|.$$

It follows that

$$\|\bar{\phi}\|^* = \sup_{\|f\| \leq 1} |\langle f, \bar{\phi} \rangle| = \sup_{\|f\| \leq 1} |\langle \bar{f}, \phi \rangle| = \sup_{\|g\| \leq 1} |\langle g, \phi \rangle| = \|\phi\|^*.$$

Hence by the triangle inequality and homogeneity

$$\|\operatorname{Re} \phi\|^* = \left\| \frac{\phi + \bar{\phi}}{2} \right\|^* \leq \frac{\|\phi\|^* + \|\bar{\phi}\|^*}{2} = \|\phi\|^*.$$

To prove (iv) it suffices to prove that for $f \neq 0$ there exists $\phi \neq 0$ such that $\operatorname{Re} \langle f, \phi \rangle \geq \|f\| \|\phi\|^*$, as the reverse inequality follows from (34), and (iv) then follows by homogeneity. Consider the convex set $C = \{g : \|g\| \leq \|f\|\}$. Since

$f \notin \text{int}(C)$, the complex supporting hyperplane theorem (Corollary B.8) gives the existence of $\phi \neq 0$ such that for any $\|g\| \leq \|f\|$ we have

$$\text{Re}\langle f, \phi \rangle \geq \text{Re}\langle g, \phi \rangle.$$

For each g with $\|g\| \leq \|f\|$ there exists $|\theta| = 1$ such that

$$|\langle g, \phi \rangle| = \theta \langle g, \phi \rangle = \langle \theta g, \phi \rangle = \text{Re}\langle \theta g, \phi \rangle.$$

Notice that $\|\theta g\| \leq \|f\|$ also, therefore

$$\text{Re}\langle f, \phi \rangle \geq \text{Re}\langle \theta g, \phi \rangle = |\langle g, \phi \rangle|.$$

Hence by homogeneity

$$\text{Re}\langle f, \phi \rangle \geq \sup_{\|g\| \leq \|f\|} |\langle g, \phi \rangle| = \|f\| \sup_{\|g\| \leq 1} |\langle g, \phi \rangle| = \|f\| \|\phi\|^*.$$

□

Proof of Theorem 5.1. We prove the contrapositive, supposing there exists $0 \leq f \leq \nu$ such that for any $0 \leq g \leq 1_{[N]}$ we have

$$\|f - g\| > \varepsilon N.$$

Our aim is to deduce that $\|\nu - 1_{[N]}\| > \exp(-C\varepsilon^{-2/3})N$. If $\|\nu - 1_{[N]}\| > N$ we are done, so we may assume that $\|\nu - 1_{[N]}\| \leq N$. In particular, it is useful to note for later that

$$\|\nu\|_1 = \|\hat{\nu}\|_\infty \leq \|\hat{1}_{[N]}\|_\infty + \|\hat{\nu} - \hat{1}_{[N]}\|_\infty \leq 2N. \quad (36)$$

By Lemma 5.1 (iv), for each $0 \leq g \leq 1_{[N]}$ there exists ϕ_g with $\|\phi_g\|^* = 1$ such that

$$\text{Re}\langle f - g, \phi_g \rangle > \varepsilon N. \quad (37)$$

Consider the subsets of \mathbb{C}^N given by

$$A := \{g - f : 0 \leq g \leq 1_{[N]}\} \quad \text{and} \quad B := \{\phi : \|\phi\|^* \leq 1\}.$$

One can check that both A and B are convex, compact and non-empty. Moreover, A is the convex hull of the finite set $\{1_S - f : S \subset [N]\}$. Applying the minimax theorem (Corollary C.2), there exists $0 \leq g_0 \leq 1_{[N]}$ and $\|\phi_0\|^* \leq 1$ such that for any $0 \leq g \leq 1_{[N]}$ and $\|\phi\|^* \leq 1$ we have

$$\text{Re}\langle g_0 - f, \phi \rangle \geq \text{Re}\langle g - f, \phi_0 \rangle.$$

In particular, using (37) we see that for any $0 \leq g \leq 1_{[N]}$ we have

$$\text{Re}\langle f - g, \phi_0 \rangle \geq \text{Re}\langle f - g_0, \phi_{g_0} \rangle > \varepsilon N.$$

Set $\psi := \text{Re}\phi_0$ and write ψ_+ for the positive part of ψ . Taking $g := 1_{\psi \geq 0}$, non-negativity gives that

$$\langle \nu, \psi_+ \rangle \geq \langle f, \psi_+ \rangle \geq \langle f, \psi \rangle = \text{Re}\langle f, \phi_0 \rangle > \text{Re}\langle g, \phi_0 \rangle + \varepsilon N = \langle 1_{[N]}, \psi_+ \rangle + \varepsilon N.$$

Therefore

$$\langle \nu - 1_{[N]}, \psi_+ \rangle > \varepsilon N.$$

By L^∞ -compatibility (Lemma 5.1 (ii)) we have

$$\|\psi\|_\infty \leq \|\phi_0\|_\infty \leq \|\phi_0\|^* \leq 1.$$

Hence by the Weierstrass polynomial approximation theorem (Lemma D) there exists a polynomial P of degree at most $C\varepsilon^{-2/3}$ and height at most $\exp(C\varepsilon^{-2/3})$ such that

$$\|P \circ \psi - \psi_+\|_\infty \leq \frac{1}{4}\varepsilon$$

Using this and the observation (36), we see that

$$\begin{aligned} \langle \nu - 1_{[N]}, P \circ \psi \rangle &= \langle \nu - 1_{[N]}, \psi_+ \rangle + \langle \nu - 1_{[N]}, P \circ \psi - \psi_+ \rangle \\ &\geq \varepsilon N - \left\| \nu - 1_{[N]} \right\|_1 \|P \circ \psi - \psi_+\|_\infty \\ &\geq \frac{1}{2}\varepsilon N. \end{aligned}$$

By (34) it follows that

$$\left\| \nu - 1_{[N]} \right\| \|P \circ \psi\|^* \geq \frac{1}{2}\varepsilon N. \quad (38)$$

By real compatibility (Lemma 5.1 (iii)), we have $\|\psi\|^* \leq \|\phi_0\|^* \leq 1$. Hence by the algebra property (Lemma 5.1 (i)) and the triangle inequality, we deduce that

$$\|P \circ \psi\|^* \ll \exp(C\varepsilon^{-2/3}).$$

Combining this with (38) finally yields the required bound. \square

APPENDIX A. THE LARGE SPECTRUM AND BOHR SETS

As in §2 we define the $(\eta \|\nu\|_1)$ -large spectrum of f to be the set

$$\text{Spec}(f, \eta \|\nu\|_1) := \left\{ \alpha \in \mathbb{T} : |\hat{f}(\alpha)| \geq \eta \|\nu\|_1 \right\}.$$

Notice that this set is empty unless $\eta \leq 1$, which we assume throughout what follows.

Lemma A.1. *Suppose that ν is a majorant on $[N]$ satisfying a restriction estimate at exponent p . Then for any $0 \leq f \leq \nu$ we have*

$$\text{meas}(\text{Spec}(f, \eta \|\nu\|_1)) \ll_p \eta^{-p} N^{-1}.$$

Proof. We have

$$\begin{aligned} \text{meas}(\text{Spec}(f, \eta N)) &\leq (\eta \|\nu\|_1)^{-p} \int_{\text{Spec}(f, \eta N)} |f(\alpha)|^p d\alpha \\ &\leq (\eta \|\nu\|_1)^{-p} \int_{\mathbb{T}} |f(\alpha)|^p d\alpha. \end{aligned}$$

By the restriction estimate we have

$$\int_{\mathbb{T}} |f(\alpha)|^p d\alpha \ll_p \|\nu\|_1^p N^{-1}.$$

\square

Define the Bohr set with frequency set $S \subset \mathbb{T}$ and width $\varepsilon \leq 1/2$ by

$$B(S, \varepsilon) := \{n \in [-\varepsilon N, \varepsilon N] : \|n\alpha\| \leq \varepsilon \quad (\forall \alpha \in S)\}.$$

Lemma A.2. *Suppose that v is a majorant on $[N]$ satisfying a restriction estimate at exponent p . Then for $0 \leq f \leq v$ and $S = \text{Spec}(f, \eta \|v\|_1)$ we have*

$$|B(S, \varepsilon)| \geq \varepsilon^{O_p(\eta^{-p-1})} N.$$

Proof. Set

$$M := \lceil 4\pi N \eta^{-1} \rceil$$

and partition \mathbb{T} into M half-open intervals of length M^{-1} . Let I_1, \dots, I_r denote those intervals which intersect $S = \text{Spec}(f, \eta N)$. We claim that

$$\bigcup_{i=1}^r I_i \subset \text{Spec}(f, \frac{1}{2}\eta \|v\|_1).$$

To see this, let us fix a choice of $\alpha_i \in I_i \cap S$ for each i . If $\alpha \in I_i$ then $\|\alpha - \alpha_i\| \leq \eta/(4\pi N)$ so that

$$\begin{aligned} |\hat{f}(\alpha)| &\geq |\hat{f}(\alpha_i)| - |\hat{f}(\alpha) - \hat{f}(\alpha_i)| \\ &\geq \eta \|v\|_1 - \|f\|_1 N 2\pi \|\alpha - \alpha_i\| \\ &\geq \frac{1}{2}\eta \|v\|_1. \end{aligned}$$

By Lemma A.1 we therefore have

$$r\eta/N \ll \text{meas}\left(\bigcup_{i=1}^r I_i\right) \ll_p \eta^{-p}/N,$$

so that

$$r \ll_p \eta^{-1-p}.$$

One can check that

$$B(\{\alpha_1, \dots, \alpha_r\}, \varepsilon/2) \subset B(S, \varepsilon).$$

Therefore

$$|B(S, \varepsilon)| \geq |B(\{\alpha_1, \dots, \alpha_r\}, \varepsilon/2)|.$$

Set $T := \lceil 2/\varepsilon \rceil$ and partition \mathbb{T}^r into T^r half-open cubes of side-length T^{-1} . By the pigeon-hole principle, some such cube C contains the point $n(\alpha_1, \dots, \alpha_r)$ for at least $\frac{1}{2}\varepsilon N T^{-r}$ values of $n \in [0, \frac{1}{2}\varepsilon N]$. Then $C - C \subset [-T^{-1}, T^{-1}]^r$ contains at least $\frac{1}{2}\varepsilon N T^{-r}$ values of $n \in [-\frac{1}{2}\varepsilon N, \frac{1}{2}\varepsilon N]$. In conclusion, we have shown that

$$|B(\{\alpha_1, \dots, \alpha_r\}, \varepsilon/2)| \geq \frac{1}{2}\varepsilon N \lceil 2/\varepsilon \rceil^{-r} \geq \lceil 2/\varepsilon \rceil^{-(r+1)} N.$$

The lemma now follows. □

APPENDIX B. THE SUPPORTING HYPERPLANE THEOREM

In this appendix we give an account of the supporting hyperplane theorem, employed in §5, and also needed in the proof of the minimax theorem given in Appendix C. The result is itself a weak version of the finite dimensional Hahn–Banach theorem and is standard. However, we have not found a satisfactory reference for the version of the result we require.

Definition B.1 (Affine independence). We say $x_0, x_1, \dots, x_k \in \mathbb{R}^n$ are *affinely dependent* if there exist $\lambda_i \in \mathbb{R}$ not all zero such that

$$\sum_{i=1}^k \lambda_i x_i = 0 \quad \text{and} \quad \sum_{i=1}^k \lambda_i = 0.$$

Equivalently, the differences $x_1 - x_0, \dots, x_k - x_0$ are linearly dependent.

Lemma B.2. *If $x_0, x_1, \dots, x_n \in \mathbb{R}^n$ are affinely independent, then the simplex*

$$\left\{ \sum_{i=0}^n \lambda_i x_i : \lambda_i > 0, \sum_{i=0}^n \lambda_i = 1 \right\} \quad (39)$$

is a non-empty open set.

Proof. The set

$$\Delta := \left\{ \mu \in \mathbb{R}^n : \mu_i > 0, \sum_{i=1}^n \mu_i < 1 \right\}$$

is the finite intersection of $n + 1$ open sets each containing $(1/2, \dots, 1/2)$, so is itself a non-empty open set.

The simplex (39) is equal to

$$\left\{ \sum_{i=1}^n \mu_i (x_i - x_0) : \mu_i > 0, \sum_{i=1}^n \mu_i < 1 \right\},$$

which is the image of Δ under a map with continuous inverse. Hence (39) is open and non-empty. \square

Given $x \in \mathbb{R}^n$, write

$$|x|_\infty := \max_i |x_i| \quad \text{and} \quad B_\varepsilon^\infty(x) := \{y \in \mathbb{R}^n : |x - y|_\infty < \varepsilon\}.$$

Lemma B.3. *For $x_0, x_1, \dots, x_n \in \mathbb{R}^n$ affinely independent, there exists $\varepsilon > 0$ such that for any $y \in B_\varepsilon^\infty(x_n)$ the vectors x_0, \dots, x_{n-1}, y are also affinely independent.*

Proof. Let T denote the invertible linear map $\lambda \mapsto \sum_i \lambda_i (x_i - x_0)$. Then there exists $C = C(x_i) > 0$ such that for any $v \in \mathbb{R}^n$ we have

$$\left| T^{-1}v \right|_\infty \leq C |v|_\infty.$$

Suppose that $x_0, \dots, x_{n-1}, x_n + v$ are affinely dependent. Then there exist λ_i with $\lambda_n = 1$ such that

$$\sum_{i=1}^n \lambda_i (x_i - x_0) = -v.$$

Therefore $|T^{-1}v|_\infty \geq 1$, which in turn implies that $|v|_\infty \geq C^{-1}$. The lemma now follows on taking $\varepsilon = C^{-1}$. \square

Given a subset C of a topological space, write \bar{C} for its closure and $\text{int}(C)$ for its interior.

Lemma B.4. *Let $C \subset \mathbb{R}^n$ be a convex set. Then $\text{int}(C) = \emptyset$ if and only if $\text{int}(\bar{C}) = \emptyset$.*

Proof. It suffices to prove the contrapositive of the ‘only if’ direction. Let $x \in \text{int}(\overline{C})$, so that there exists $\varepsilon > 0$ such that

$$B_\varepsilon^\infty(x) \subset \overline{C}.$$

Taking $x_0 = x$ and $x_i = x + (\varepsilon/2)e_i$, one sees that the set $B_\varepsilon^\infty(x)$ contains $n + 1$ affinely independent points.

By Lemma B.3 there exists $\delta > 0$ such that $B_\delta^\infty(x_n) \subset B_\varepsilon^\infty(x)$ and for any $y \in B_\delta(x_n)$ the vectors $x_0, x_1, \dots, x_{n-1}, y$ are affinely independent. Since $x_n \in \overline{C}$, there exists $x'_n \in B_\delta(x_n) \cap C$, so that $x_0, x_1, \dots, x_{n-1}, x'_n$ are affinely independent elements of $B_\varepsilon^\infty(x)$.

Repeating the above argument with x_i in place of x_n , we see that we can find affinely independent $x'_0, x'_1, \dots, x'_n \in B_\varepsilon^\infty(x) \cap C$. It then follows from convexity and Lemma B.2 that the set

$$\left\{ \sum_{i=0}^n \lambda_i x'_i : \lambda_i > 0, \sum_{i=0}^n \lambda_i = 1 \right\}$$

is a non-empty open subset of C . □

Lemma B.5. *Let C be a convex subset of \mathbb{R}^n with $x \in \text{int}(C)$ and $y \in \overline{C}$. Then $\text{int}(C)$ contains the line segment*

$$[x, y) := \{(1 - \lambda)x + \lambda y : \lambda \in [0, 1)\}.$$

Proof. Since $x \in \text{int}(C)$ there exists an open set $U \subset C$ with $x \in U$. Let $z \in (x, y)$, so that there exists $\lambda \in (0, 1)$ with

$$z = (1 - \lambda)x + \lambda y.$$

Taking $\mu = \lambda^{-1}$ we have

$$y = \mu z + (1 - \mu)x,$$

so that y is an element of the open set

$$V := \bigcup_{\mu > 1} (\mu z + (1 - \mu) \cdot U).$$

Since $y \in \overline{C}$, there exists $y_1 \in V \cap C$. Hence there exists $\mu_1 > 1$ and $u_1 \in U$ such that

$$y_1 = \mu_1 z + (1 - \mu_1)u_1.$$

Taking $\lambda_1 = \frac{1}{\mu_1}$, we have

$$z = \lambda_1 y_1 + (1 - \lambda_1)u_1,$$

so that z is an element of the open set

$$W := \bigcup_{0 \leq \lambda < 1} (\lambda y_1 + (1 - \lambda) \cdot U).$$

By convexity $W \subset C$, hence $z \in \text{int}(C)$. □

Lemma B.6. *If $C \subset \mathbb{R}^n$ is convex then*

$$\text{int}(C) = \text{int}(\overline{C}).$$

Proof. By Lemma B.4 we may assume that $\text{int}(C)$ is non-empty. It suffices to show that if $B_\varepsilon(x) \subset \bar{C}$ for some $\varepsilon > 0$ then $x \in \text{int}(C)$. Let $x_0 \in \text{int}(C)$. Choosing $\delta > 0$ sufficiently small, one can ensure that

$$y := x + \delta(x - x_0) \in B_\varepsilon(x) \subset \bar{C}.$$

Hence by the previous lemma $[x_0, y) \subset \text{int}(C)$. Taking $\lambda = \frac{1}{1+\delta}$ we see that

$$x = (1 - \lambda)x_0 + \lambda y \in (x_0, y) \subset \text{int} C.$$

□

In order to distinguish between the complex inner product on \mathbb{C}^n and the real inner product on \mathbb{R}^{2n} , we write $\langle x, y \rangle$ for the former and $x \cdot y$ for the latter.

Lemma B.7 (Supporting hyperplane theorem). *Let C be a convex subset of \mathbb{R}^n and $x \notin \text{int} C$. Then there exists a non-zero vector $\phi \in \mathbb{R}^n \setminus \{0\}$ such that for all $y \in \bar{C}$ we have*

$$y \cdot \phi \leq x \cdot \phi.$$

Proof. Let us first prove the result under the assumption that $x \notin \bar{C}$. The result is trivial if $C = \emptyset$, so we may assume that $C \neq \emptyset$. Using absolute values to denote the L^2 -norm on \mathbb{R}^n , it follows that there exists $y_0 \in \bar{C}$ such that

$$|y_0 - x| = \inf_{y \in \bar{C}} |y - x|.$$

Heuristically, we expect that for any $y \in \bar{C}$, the angle formed in the plane between $x - y_0$ and $y - y_0$ should be obtuse. If this angle were acute, then there should exist a point y_1 on the line segment between y and y_0 such that x is closer to y_1 than y_0 (draw a picture). Since \bar{C} is convex we have $y_1 \in \bar{C}$ and we have contradicted our choice of y_0 .

More rigorously, we show that for any $y \in \bar{C}$ we have

$$(x - y_0) \cdot (y - y_0) \leq 0. \quad (40)$$

Suppose this is not the case. Then we claim that there exists $t \in (0, 1]$ such that

$$|(1 - t)y_0 + ty - x| < |y_0 - x|,$$

and hence obtain our desired contradiction. Write $(1 - t)y_0 + ty - x = y_0 - x + t(y - y_0)$, then square, expand out and divide through by t to deduce that this is equivalent to the existence of $t \in (0, 1]$ such that

$$t|y - y_0|^2 < 2(x - y_0) \cdot (y - y_0).$$

Since we are assuming that (40) does not hold, we may take

$$t := \min \left\{ 1, \frac{(x - y_0) \cdot (y - y_0)}{|y - y_0|^2} \right\}.$$

Assuming that $x \notin \bar{C}$, we may take $\phi := x - y_0 \neq 0$ to deduce that for any $y \in \bar{C}$ we have

$$y \cdot \phi \leq y_0 \cdot \phi \leq x \cdot \phi,$$

the latter inequality following from the fact that $(x \cdot \phi) - (y_0 \cdot \phi) = (\phi \cdot \phi) \geq 0$.

It remains to prove the result when $x \in \overline{C} \setminus \text{int}(C)$. Since C is convex, it follows from Lemma B.6 that $x \notin \text{int}(\overline{C})$, so that for any $m \in \mathbb{N}$ there exists $x_m \notin \overline{C}$ such that

$$|x - x_m| \leq 1/m.$$

By our previous argument, there exists $\phi_m \neq 0$ such that for any $y \in \overline{C}$ we have

$$y \cdot \phi_m \leq x_m \cdot \phi_m. \quad (41)$$

Normalising so that $|\phi_m| = 1$, we have a sequence in a compact set, so there exists a convergent subsequence $\phi_{k(m)} \rightarrow \phi$ with $|\phi| = 1$. Taking limits in (41) then gives the desired inequality. \square

Corollary B.8 (Complex supporting hyperplane theorem). *Let C be a convex subset of \mathbb{C}^n and $x \notin \text{int}(C)$. Then there exists $\phi \in \mathbb{C}^n \setminus \{0\}$ such that for all $y \in \overline{C}$ we have*

$$\text{Re} \langle y, \phi \rangle \leq \text{Re} \langle x, \phi \rangle.$$

Proof. This follows from the observation that for $x, y \in \mathbb{C}^n \cong \mathbb{R}^{2n}$ we have

$$\text{Re} \langle x, y \rangle = x \cdot y.$$

\square

APPENDIX C. THE SEMI-FINITE MINIMAX THEOREM

We have not been able to find a reference for the variant of the minimax theorem employed in §5.

Proposition C.1 (Semi-finite minimax). *Let A and B be non-empty compact convex subsets of \mathbb{R}^n at least one of which is equal to the convex hull of finitely many points. Then there exist $a_0 \in A$ and $b_0 \in B$ such that for any $a \in A$ and any $b \in B$ we have*

$$a \cdot b_0 \leq a_0 \cdot b.$$

Proof. We may assume that A is the convex hull of finitely many points, otherwise we re-label, taking $A' := -B$ and $B' := A$ to obtain $b_0 \in B$ and $a_0 \in A$ such that for any $b \in B$ and $a \in A$ we have

$$-b \cdot a_0 \leq -b_0 \cdot a,$$

which yields the claimed result.

Define

$$L := \sup_{a \in A} \inf_{b \in B} (a \cdot b) \quad \text{and} \quad U := \inf_{b \in B} \sup_{a \in A} (a \cdot b).$$

In order to prove the proposition, it suffices to establish that

(i) There exists $a_0 \in A$ and $b_0 \in B$ such that

$$L = \inf_{b \in B} (a_0 \cdot b) \quad \text{and} \quad U = \sup_{a \in A} (a \cdot b_0).$$

(ii) $U \leq L$.

We begin by showing that

$$-\infty < L \leq U < \infty. \quad (42)$$

For any $a_1 \in A$ and $b_1 \in B$ we have

$$\inf_{b \in B} (a_1 \cdot b) \leq a_1 \cdot b_1 \leq \sup_{a \in A} (a \cdot b_1).$$

Since a_1 and b_1 are arbitrary, it follows that $L \leq U$. Since B is non-empty, there exists $b_1 \in B$. Thus

$$U \leq \sup_{a \in A} (a \cdot b_1)$$

By compactness, there exists $a_1 \in A$ such that

$$\sup_{a \in A} (a \cdot b_1) = a_1 \cdot b_1 < \infty.$$

We conclude that $U < \infty$. Similarly, compactness of B and non-emptiness of A yields $L > -\infty$. This establishes (42).

Since U is finite, for any $k \in \mathbb{N}$ there exists $b_k \in B$ such that

$$U \leq \sup_{a \in A} (a \cdot b_k) \leq U + \frac{1}{k}.$$

By compactness of B , there exists a convergent subsequence $b_{k_m} \rightarrow b \in B$. Continuity of the map $(a, b) \mapsto a \cdot b$ then ensures that for any $a \in A$ we have

$$a \cdot b = \lim_{m \rightarrow \infty} a \cdot b_{k_m} \leq U.$$

Thus $\sup_{a \in A} (a \cdot b) \leq U$, which by definition of U implies that $\sup_{a \in A} (a \cdot b) = U$. A similar argument holds for L . This proves (i).

Finally, we show that for any $\alpha \in \mathbb{R}$ we either have $L \geq \alpha$ or $U \leq \alpha$. Combining this with the fact that $L \leq U$, it follows that $L = U$ (if not, any $\alpha \in (L, U)$ leads to a contradiction).

Since A is the convex hull of finitely many points, there exist $a_1, \dots, a_k \in A$ such that

$$A = \left\{ \sum_{i=1}^k \lambda_i a_i : \lambda_i \geq 0 \text{ and } \sum_i \lambda_i = 1 \right\}.$$

Given $b \in B$ let us write

$$v_b := \left((a_1 \cdot b) - \alpha, \dots, (a_k \cdot b) - \alpha \right) \in \mathbb{R}^k.$$

Define C to be the convex hull of the set

$$\{v_b : b \in B\} \cup \{e_1, \dots, e_k\}.$$

Let us first suppose that $0 \in C$. Then there exist $b_1, \dots, b_m \in B$, $\lambda_1, \dots, \lambda_m, \mu_1, \dots, \mu_k \geq 0$ such that $1 = \sum_i \lambda_i + \sum_j \mu_j$ and

$$0 = \sum_i \lambda_i v_{b_i} + (\mu_1, \dots, \mu_k). \quad (43)$$

It follows that for each $j = 1, \dots, k$ we have

$$\sum_{i=1}^m \lambda_i ((a_j \cdot b_i) - \alpha) \leq 0. \quad (44)$$

By (43) we cannot have all λ_i equal to zero. We may therefore re-normalise, to conclude that there exist $\lambda_i \geq 0$ with $\sum_i \lambda_i = 1$ satisfying (44). We deduce that for each $j = 1, \dots, k$ we have

$$a_j \cdot \sum_{i=1}^m \lambda_i b_i \leq \alpha.$$

Convexity then shows that for $b = \sum_i \lambda_i b_i \in B$ and for any $a \in A = \text{ConvexHull}(a_1, \dots, a_k)$ we have $a \cdot b \leq \alpha$. Hence $U \leq \alpha$.

Next suppose that $0 \notin C$. By the supporting hyperplane theorem (Lemma B.7, and the remark which follows it), there exists $\phi \in \mathbb{R}^k \setminus \{0\}$ such that for all $b_1, \dots, b_m \in B$ and $\lambda_1, \dots, \lambda_m, \mu_1, \dots, \mu_k \geq 0$ with $\sum_i \lambda_i + \sum_j \mu_j = 1$ we have

$$\left(\sum_i \lambda_i v_{b_i} + \sum_j \mu_j e_j \right) \cdot \phi \geq 0.$$

In particular, we have

$$\phi_j = e_j \cdot \phi \geq 0 \quad (j = 1, \dots, k),$$

and for each $b \in B$ we have

$$v_b \cdot \phi \geq 0.$$

Since $\phi \neq 0$ we may re-normalise to conclude that there exists $\phi_j \geq 0$ with $\sum_j \phi_j = 1$ such that for any $b \in B$ we have

$$\left(\sum_{j=1}^k \phi_j a_j \right) \cdot b \geq \alpha.$$

Convexity of A then gives the existence of $a = \sum_j \phi_j a_j \in A$ such that for all $b \in B$ we have $a \cdot b \geq \alpha$, so that $L \geq \alpha$. \square

Recall that in order to distinguish between the complex inner product on \mathbb{C}^n and the real inner product on \mathbb{R}^{2n} , we write $\langle x, y \rangle$ for the former and $x \cdot y$ for the latter.

Corollary C.2 (Complex minimax). *Let A and B be non-empty compact convex subsets of \mathbb{C}^n at least one of which is equal to the convex hull of finitely many points. Then there exist $a_0 \in A$ and $b_0 \in B$ such that for any $a \in A$ and any $b \in B$ we have*

$$\text{Re} \langle a, b_0 \rangle \leq \text{Re} \langle a_0, b \rangle.$$

APPENDIX D. THE WEIERSTRASS POLYNOMIAL APPROXIMATION THEOREM

Given a real number x write

$$x_+ := \max \{x, 0\} = \frac{1}{2}(x + |x|).$$

Lemma D.1 (Weierstrass polynomial approximation). *There exists an absolute constant $C > 0$ such that for any $\varepsilon \in (0, 1)$ there exists a polynomial P of degree at most $C\varepsilon^{-2/3}$ and height at most $\exp(C\varepsilon^{-2/3})$ such that*

$$\sup_{|x| \leq 1} |P(x) - x_+| \leq \varepsilon.$$

Proof. By the Taylor series theorem, for any $t \in [0, 1)$ we have

$$(1 - t)^{1/2} = - \sum_{n=0}^N c_n t^n + O(c_{N+1}), \quad (45)$$

where

$$c_n = \frac{(2n)!}{(2n-1)2^{2n}(n!)^2}$$

Using Stirling's formula, one can check that there exists a constant C such that

$$c_n \sim Cn^{-3/2} \quad \text{as } n \rightarrow \infty.$$

In particular, by absolute convergence and continuity, the approximation (45) is valid for $t \in [0, 1]$.

For any $x \in [-1, 1]$ we see that

$$|x| = (1 - (1 - x^2))^{1/2} = \sum_{m=0}^N (-1)^m \left(\sum_{n=m}^N c_n \binom{n}{m} \right) x^{2m} + O(N^{-3/2}).$$

Using the crude bound

$$\left| \sum_{n=m}^N c_n \binom{n}{m} \right| \leq 2^N,$$

we deduce that for any $N \in \mathbb{N}$ there exists a real polynomial P_N of degree at most $2N$ and height at most $C \exp(N)$ such that

$$\sup_{x \in [-1, 1]} |P_N(x) - |x|| \ll N^{-3/2}.$$

The result now follows on taking $P(x) := \frac{1}{2}(P_N(x) + x)$ and ensuring that $N \geq C/\varepsilon^{2/3}$ for some absolute constant C . \square

REFERENCES

- [Blo12] T. F. Bloom, *Translation invariant equations and the method of Sanders*, Bull. Lond. Math. Soc. **44** (2012), no. 5, 1050–1067.
- [BP] T. D. Browning and S. M. Prendiville, *A transference approach to a Roth-type theorem in the squares*, IMRN **7** (2017), 2219–2248.
- [Cho] S. Chow, *Roth–Waring–Goldbach*, IMRN (2017).
- [CG] D. Conlon and W. T. Gowers, *Combinatorial theorems relative to a sparse random set*, Ann. of Math. **184** (2016), 367–454.
- [CFZ] D. Conlon, J. Fox and Y. Zhao, *A relative Szemerédi theorem*, GAFA **25** (2015), 733–762.
- [Gow10] W. T. Gowers, *Decompositions, approximate structure, transference, and the Hahn-Banach theorem*, Bull. Lond. Math. Soc. **42** (2010), no. 4, 573–606.
- [Gre05] B. J. Green, *Roth's theorem in the primes*, Ann. of Math. **161** (2005), 1609–1636.

- [GT06] B. J. Green and T. Tao , *Restriction theory of the Selberg sieve, with applications*, Journal de théorie des nombres de Bordeaux **18** (2006), 147–182.
- [GT08] ———, *The primes contain arbitrarily long arithmetic progressions*, Ann. of Math. **167** (2008), 481–547.
- [GT10] ———, *Linear equations in primes*, Ann. of Math. **171** (2010), 1753–1850.
- [Har] A. J. Harper, *Minor arcs, mean values, and restriction theory for exponential sums over smooth numbers*, Compos. Math. **152** (2016), 1121–1158.
- [HDR11] H. A. Helfgott and A. De Roton, *Improving Roth’s theorem in the primes*, IMRN (2011), no. 4, 767–783.
- [Mat12] L. Matthiesen, *Correlations of the divisor function*, Proc. London Math. Soc. **104** (2012), 827–858.
- [Nas15] E. Naslund, *On improving Roth’s theorem in the primes*, Mathematika **61** (2015), 49–62.
- [RTTV] O. Reingold, L. Trevisan, M. Tulsiani, and S. Vadhan, *New proofs of the Green–Tao–Ziegler dense model theorem: an exposition*, preprint available at [arXiv:0806.0381](https://arxiv.org/abs/0806.0381).
- [Rot53] K. F. Roth, *On certain sets of integers*, J. London Math. Soc. **28** (1953), 104–109.
- [Rot54] ———, *On certain sets of integers. II*, J. London Math. Soc. **29** (1954), 20–26.
- [Sch] M. Schacht, *Extremal results for random discrete structures*, Ann. of Math. **184** (2016), (333–365).
- [TZ08] T. Tao and T. Ziegler, *The primes contain arbitrarily long polynomial progressions*, Acta Math. **201** (2008), no. 2, 213–305.

SCHOOL OF MATHEMATICS, UNIVERSITY OF MANCHESTER, MANCHESTER, UK
E-mail address: sean.prendiville@gmail.com