

QUANTITATIVE INVERSE THEORY OF GOWERS UNIFORMITY NORMS

[after F. Manners]

by **Thomas F. Bloom**

ABSTRACT:

Gowers uniformity norms are the central objects of higher order Fourier analysis, one of the cornerstones of additive combinatorics, and play an important role in both Gowers' proof of Szemerédi's theorem and the Green–Tao theorem. The inverse theorem states that if a function has a large uniformity norm, which is a robust combinatorial measure of structure, then it must correlate with a nilsequence, which is a highly structured algebraic object. This was proved in a qualitative sense by Green, Tao, and Ziegler, but with a proof that was incapable of providing reasonable bounds. In 2018 Manners achieved a breakthrough by giving a new proof of the inverse theorem. Not only does this new proof contain a wealth of new insights but it also, for the first time, provides quantitative bounds, that are at worst only doubly exponential. This talk will give a high-level overview of what the inverse theorem says, why it is important, and the new proof of Manners.

INTRODUCTION

Let $s \geq 1$ be a fixed integer (in particular, all implicit constants may depend on s). We fix some large prime N and use \mathbb{Z}_N to denote the cyclic group $\mathbb{Z}/N\mathbb{Z}$. If $f : \mathbb{Z}_N \rightarrow \mathbb{C}$ and $h \in \mathbb{Z}_N$ we define the multiplicative derivative $\Delta_h f : \mathbb{Z}_N \rightarrow \mathbb{C}$ by $\Delta_h f(x) = f(x)\overline{f(x+h)}$. This definition has a natural extension to vectors $\mathbf{h} \in \mathbb{Z}_N^k$, so that

$$\Delta_{\mathbf{h}} f(x) = \Delta_{h_1} \cdots \Delta_{h_k} f.$$

We can now define our central object of study: the Gowers uniformity norm of degree s . This is defined for $f : \mathbb{Z}_N \rightarrow \mathbb{C}$ by

$$\|f\|_{U^{s+1}} = \left(\mathbb{E}_{\mathbf{h} \in \mathbb{Z}_N^{s+1}} \mathbb{E}_{x \in \mathbb{Z}_N} \Delta_{\mathbf{h}} f(x) \right)^{1/2^{s+1}}.$$

Here we have used the expectation notation, now standard within additive combinatorics, to denote the normalised sum (so that, for example, $\mathbb{E}_{x \in H}$ means $\frac{1}{|H|} \sum_{x \in H}$). Alternatively, one could use the inductive definition

$$\|f\|_{U^1} = \left| \mathbb{E}_x f(x) \right| \quad \text{and} \quad \|f\|_{U^{s+1}} = \left(\mathbb{E}_{h \in \mathbb{Z}_N} \|\Delta_h f\|_{U^s}^{2^s} \right)^{1/2^{s+1}}.$$

These norms⁽¹⁾ were introduced into additive combinatorics by GOWERS (2001) in his analytic proof of Szemerédi’s theorem. They provide a quantitative measure of the additive structure of f . For example, the U^2 norm

$$\|f\|_{U^2} = \left(\mathbb{E}_{x, h_1, h_2 \in \mathbb{Z}_N} f(x) \overline{f(x+h_1)} \overline{f(x+h_2)} f(x+h_1+h_2) \right)^{1/4}$$

measures f along 2-dimensional additive quadruples of the shape $(x, x+h_1, x+h_2, x+h_1+h_2)$. In general, the U^{s+1} norm measures f along 2^{s+1} -tuples which are (the projections of) $(s+1)$ -dimensional cubes.

These configurations are generic enough that if we can understand how f behaves on them then we can understand f on many other kinds of linear system. For example, the count of f along arithmetic progressions of length k is strongly related to $\|f\|_{U^{k-1}}$. As a result the uniformity norms play a central role in modern additive combinatorics. They often appear in proofs which use a ‘structure vs. randomness’ philosophy. Such proofs go along the following lines:

1. Find some function f such that behaviour of the objects one wishes to understand is governed by some $\|f\|_{U^{s+1}}$ (for example, when counting k -term arithmetic progressions inside some $A \subset \mathbb{Z}_N$ one would take $f = 1_A - \frac{|A|}{N}$ and $s = k - 2$);
2. Show that the theorem in question follows if $\|f\|_{U^{s+1}}$ is *small* – this is the random aspect of the ‘structure vs. randomness’ dichotomy, and usually follows from elementary counting methods;
3. Show that if $\|f\|_{U^{s+1}}$ is *large* then f has some algebraic structure;
4. Finally conclude the proof by showing that if f has this algebraic structure then the theorem also follows.

Landmark applications of this method are the analytic proof of Szemerédi’s theorem by GOWERS (2001), the first to deliver reasonable quantitative bounds, and the proof by GREEN and TAO (2008b) that the primes contain arbitrarily long arithmetic progressions.

The inverse theory of uniformity norms, which is the focus of this article, addresses the third point, and seeks to answer the following question.

⁽¹⁾It is not obvious from the definition that they are indeed norms, but this is true for $s \geq 1$. A proof can be found in the original paper of GOWERS (2001).

QUESTION 0.1 (The Inverse Question). — *If a 1-bounded⁽²⁾ function $f : \mathbb{Z}_N \rightarrow \mathbb{C}$ has large U^{s+1} norm then what can we deduce about f ?*

We will first explore this question in the simplest case $s = 1$. A brief experimentation shows that the function⁽³⁾ $x \mapsto e(\alpha x)$, where $\alpha \in \frac{1}{N}\mathbb{Z}$ (so that this does indeed give a well-defined function on \mathbb{Z}_N), satisfies $\|f\|_{U^2} = 1$, the maximum possible. Therefore (linear) characters provide examples of bounded functions with large U^2 norm. The inverse theorem for the U^2 norm says that the converse is also true, at least in an approximate sense. More precisely, if $\|f\|_{U^2} \geq \delta$ then f must correlate with a linear character, in that there exists some $\alpha \in \frac{1}{N}\mathbb{Z}$ such that

$$(1) \quad \left| \mathbb{E}_x f(x) \overline{e(\alpha x)} \right| > c(\delta)$$

for some constant $c(\delta) > 0$ depending only on δ . The proof of this inverse result is very short. We note first that the U^2 system of quadruples of the shape $(x, x + h_1, x + h_2, x + h_1 + h_2)$ are exactly those quadruples (x_1, x_2, x_3, x_4) such that $x_1 - x_2 = x_3 - x_4$. We may use linear characters and orthogonality to detect this linear condition so that, defining the Fourier transform by $\widehat{f}(a) = \mathbb{E}_x f(x) e(-ax/N)$ for $a \in \mathbb{Z}_N$, we have

$$\begin{aligned} \|f\|_{U^2}^4 &= \mathbb{E}_{x_1 - x_2 = x_3 - x_4} f(x_1) \overline{f(x_2)} \overline{f(x_3)} f(x_4) = \sum_{a \in \mathbb{Z}_N} |\widehat{f}(a)|^4 \\ &\leq \sup_a |\widehat{f}(a)|^2 \sum_{a \in \mathbb{Z}_N} |\widehat{f}(a)|^2 \\ &= \sup_a |\widehat{f}(a)|^2 \mathbb{E}_{x \in \mathbb{Z}_N} |f(x)|^2 \\ &\leq \sup_a |\widehat{f}(a)|^2, \end{aligned}$$

using Parseval's identity and the assumption that f is 1-bounded. It follows that if $\|f\|_{U^2} \geq \delta$ then (1) holds with the explicit lower bound $\geq \delta^2$.

For higher uniformity norms with $s \geq 2$ this simple argument fails, and it is less clear what functions have large U^{s+1} norm. We note that the uniformity norms are nested, in the sense that

$$\|f\|_{U^2} \leq \|f\|_{U^3} \leq \|f\|_{U^4} \leq \cdots,$$

and so in particular this inverse question becomes more difficult as s increases, since any example of a function with large U^{s+1} norm certainly also has large U^{s+2} norm, but the converse may not hold.

Considering what happens when $s = 1$, we observe that the reason exponentials with linear phases have large U^2 norm is because the second derivative of a linear function always vanishes (hence $\Delta_{h_1, h_2} f$ is an exponential with phase 0, and so is identically 1).

⁽²⁾A function f is 1-bounded if $|f(x)| \leq 1$ for all x .

⁽³⁾We use $e(x)$ to denote $e^{2\pi i x}$.

Generalising this observation shows that $f(x) = e(P(x))$ will have $\|f\|_{U^{s+1}} = 1$ whenever P is a polynomial of degree $\leq s$ (with coefficients in $\frac{1}{N}\mathbb{Z}$ so that it is well-defined on \mathbb{Z}_N).

At this point one may guess that, just as when $s = 1$, the approximate converse also holds, and conjecture something like: if $\|f\|_{U^{s+1}} \geq \delta$ then

$$(2) \quad \left| \mathbb{E}_x f(x) \overline{e(P(x))} \right| > c(\delta)$$

for some polynomial $P(x) \in \frac{1}{N}\mathbb{Z}[x]$ of degree $\leq s$ and some constant $c(\delta) > 0$ depending only on δ . It turns out this is not quite enough, and the exponentials with polynomial phases do not represent a full set of obstructions.⁽⁴⁾ To see why, observe that we don't require our 'phase functions' $P(x)$ to actually vanish after taking $s + 1$ derivatives. For the U^{s+1} norm of $e(P(x))$ to be large, it suffices for the derivative of $P(x)$ to be biased towards 0 (so that the multiplicative derivative of $e(P(x))$ is biased towards 1). Consider, for example, the function on \mathbb{Z}_N defined by⁽⁵⁾

$$x \mapsto \{\alpha x\}\{\beta x\}$$

for $\alpha, \beta \in \frac{1}{N}\mathbb{Z}$. This is not a true quadratic, since the third derivative is not identically zero. This failure arises because $x \mapsto \{\alpha x\}$ is not truly linear. We note, however, that it is linear a large proportion of the time, since we can write $\{\alpha(x + y)\} = \{\alpha x\} + \{\alpha y\} - \rho_{x,y}$, where $\rho_{x,y}$ is a 'carry bit' function that is 1 if $\{\alpha x\} + \{\alpha y\} \geq 1$ and 0 otherwise. In particular, with probability $1/2$, $\rho_{x,y} = 0$ and hence $\{\alpha x\}$ behaves like a linear function with probability $1/2$. By a similar calculation, the third derivative of a quadratic function like $P(x) = \{\alpha x\}\{\beta x\}$ vanishes a positive proportion of the time, and thus $e(P(x))$ has large U^3 norm. Thus we need to also include these generalised bracket polynomials (generated by repeated composition of, not just addition and multiplication, but also the fractional part operator $\{\cdot\}$) as possible obstructions.

The inverse theorem for the Gowers uniformity norms states that this expanded set of obstructions captures all the reasons that a function might have large uniformity norm. That is, if $\|f\|_{U^{s+1}} \geq \delta$ then there exists some bracket polynomial P of degree $\leq s$ such that (2) holds. Such a statement was conjectured by GREEN and TAO (2010b) in their work on linear equations in primes and first proved in a qualitative sense (that is, with no bounds on the function $c(\delta)$) by GREEN, TAO, and ZIEGLER (2012).

The focus of this article is a recent new proof of the inverse theorem by MANNERS (2018) which gives, for the first time, a quantitative version of this statement. Our aims are threefold:

1. To state precisely the inverse theorem as proved in MANNERS (2018), defining all the concepts required;

⁽⁴⁾The term obstructions here comes from thinking of functions g such that " $|\mathbb{E}_x f(x) \overline{g(x)}| \gg 1$ implies $\|f\|_{U^{s+1}} \gg 1$ " as 'obstructions' to having small uniformity norm.

⁽⁵⁾As usual in number theory, we use $\{\cdot\} : \mathbb{R} \rightarrow [0, 1)$ to denote the fractional part operator $x \mapsto x - \lfloor x \rfloor$, where $\lfloor x \rfloor$ is the largest integer n such that $n \leq x$.

2. To sketch some of the ideas used in the proof, giving a flavour of the kind of arguments used; and
3. To state precisely some of the new definitions and concepts introduced by Manners.

The paper of MANNERS (2018) itself is over 100 pages long, so we will not be able to even approach a proper proof of the inverse theorem. As a result of this intimidating length, however, some of the beautiful new ideas introduced by Manners may go otherwise unnoticed by those without the time to plumb the technical depths. We hope that this article helps popularise them, so that they can find many other applications.

ACKNOWLEDGEMENTS

I would like to thank Freddie Manners for many helpful comments and guidance on understanding his work, and Aled Walker for his notes and insights on Manners’ proof, which have helped me immensely. I would also like to thank Nicolas Bourbaki, Gabriel Conant, Ben Green, and Sarah Peluse for their comments on an earlier draft.

1. THE INVERSE THEOREM

The following quantitative inverse theorem for the Gowers uniformity norms is the main result of MANNERS (2018), and the focus of this article.

THEOREM 1.1 (Manners). — *Let $\delta > 0$ and $N \geq 2$ be prime. If $f : \mathbb{Z}_N \rightarrow \mathbb{C}$ is a 1-bounded function such that $\|f\|_{U^{s+1}} \geq \delta$ then there exists $\epsilon > 0$ and a 1-bounded, N -periodic, nilsequence $\psi : \mathbb{Z} \rightarrow \mathbb{C}$ with degree s , dimension D , parameter K , and complexity M , such that*

$$\left| \mathbb{E}_x f(x) \overline{\psi(x)} \right| \geq \epsilon,$$

where the parameters are bounded in terms of δ by

$$D = O(\delta^{-O(1)}) \text{ and } \epsilon^{-1}, K, M \leq \begin{cases} \exp(O(\delta^{-O(1)})) & \text{if } s \leq 3 \text{ and} \\ \exp(\exp(O(\delta^{-O(1)}))) & \text{if } s \geq 4. \end{cases}$$

We have followed Manners in stating the inverse result using nilsequences rather than bracket polynomials. These are qualitatively equivalent, as shown by BERGELSON and LEIBMAN (2007).⁽⁶⁾ Nilsequences are usually, however, easier to work with. We defer an explicit definition of nilsequences, along with the meaning of degree, dimension, parameter, and complexity, till Section 3. For now the reader should think of a nilsequence of

⁽⁶⁾BERGELSON and LEIBMAN (2007) do not give any quantitative form of this equivalence, which would be needed to make a version of Theorem 1.1 precise for bracket polynomials. Since nilsequences are more convenient to work with anyway we will not address this, and use bracket polynomials only as motivational examples.

degree s as a function of the shape $n \mapsto e(P(n))$, where $P(n)$ is a (bracket) polynomial of degree at most s .

Since it can also be shown that if f correlates with a nilsequence of degree s then it has large U^{s+1} norm (we sketch a proof of this in Section 3.4), this gives a complete characterisation of functions with large uniformity norm. A characterisation of this type was first conjectured by GREEN and TAO (2010b, Conjecture 8.3) in their work on linear equations in primes.

Before we explore Theorem 1.1 and its proof we briefly summarise other, related, results.

1. The case $s = 1$ is a trivial argument using Fourier analysis, which we have already seen in the introduction. Observe that this argument gives bounds which are polynomial in δ .
2. The case $s = 2$ was first established by GREEN and TAO (2008a), with bounds exponential in $\delta^{-O(1)}$. The inverse theory for the U^3 norm has a strong connection with Freiman-type inverse sumset problems. For example, it is known that polynomial bounds for the inverse U^3 problem are equivalent to the Polynomial Freiman-Ruzsa conjecture (as shown by GREEN and TAO (2010a) and LOVETT (2012)). Using this equivalence, quantitative improvements for inverse sumset problems by SANDERS (2012) imply a bound for the U^3 inverse theorem exponential in $(\log(1/\delta))^{O(1)}$.
3. The case for general s was proved in a qualitative sense by GREEN, TAO, and ZIEGLER (2012). This proof gives no explicit bounds at all, and uses the language of non-standard analysis. In theory this proof could be made explicit, but any bounds that could be extracted would be terrible, and far up in the Ackermann hierarchy (in particular, much worse than even a tower of exponentials of height $O(\delta^{-1})$).
4. An independent approach towards a proof of the inverse theorem for general s was initiated by SZEGEDY (2012) and further developed in a number of papers by CAMARENA and SZEGEDY (2010), CANDELA and SZEGEDY (2019), and GUTMAN, MANNERS, and VARJÚ (2019, 2020a,b). This approach is also qualitative, and would deliver similarly terrible bounds.
5. These inverse results are all ‘global’, in that they find a structured object with which f correlates on the entirety of \mathbb{Z}_N . A weaker local inverse statement, in which one just finds some large subset of \mathbb{Z}_N on which f has strong correlation with a structured object, was proved by GOWERS (2001, Theorem 18.1) (with even better bounds, only polynomial in δ). This local statement was sufficient to prove Szemerédi’s theorem, but for most applications of the uniformity norms (such as the work of GREEN and TAO (2010b) on linear equations in primes) a global inverse statement is required.
6. Similar uniformity norms and their inverse results have been studied in ergodic theory, where they were introduced independently of additive combinatorics and the

work of Gowers. The ergodic theory analogue of the inverse result was established by BERGELSON, HOST, and KRA (2005). This work was the first to recognise the importance of nilsequences as the proper set of obstructions, and shaped the formulation of the inverse conjecture by GREEN and TAO (2010b). Although we will not use ergodic-theoretic language in this article, insights and arguments from ergodic theory have been extremely influential in much of the work on uniformity norms. For an in-depth discussion of the relationship between uniformity norms in additive combinatorics and ergodic theory we refer to HOST and KRA (2009).

7. The discussion so far has been entirely for functions $f : \mathbb{Z}_N \rightarrow \mathbb{C}$, but one could define uniformity norms and ask the inverse question with \mathbb{Z}_N replaced by any finite abelian group. The principal alternative is \mathbb{F}_p^n , where the inverse theorem was first proved qualitatively by TAO and ZIEGLER (2010). Very recently a new proof with good quantitative bounds was given by GOWERS and MILIĆEVIĆ (2017, 2020). While there are some similarities between the methods used for \mathbb{Z}_N and \mathbb{F}_p^n , there are important differences that make it difficult to translate progress in one setting to the other. In particular, the strategy of Manners uses crucially that \mathbb{Z}_N has no non-trivial subgroups (when N is prime), a fact which fails dramatically for \mathbb{F}_p^n .

One of the remarkable aspects of Theorem 1.1 is its quantitative strength. There can be few greater indications that Manners has found the ‘right’ way to prove the inverse theorem than how his work moves us from no bounds at all to bounds that are, at worst, only doubly exponential in the parameter δ . The presence of the double exponential when $s \geq 4$ is, moreover, largely a technical artefact of one particular recursive argument in the proof. Most of the techniques used are much more efficient, and Manners believes that a more refined version of one part of the argument should be possible, which would bring the bounds down to exponential in δ for all s . The true bounds should probably be only polynomial in δ — as mentioned above this is equivalent to the polynomial Frieman–Ruzsa conjecture when $s = 2$, and we see no reason to suspect that they would not be polynomial in general, although to the best of our knowledge this conjecture has not appeared explicitly in the literature.

Our aim is, as stated above, to explain carefully what the inverse theorem says and sketch some of the ideas used by Manners. In limited space there is necessarily a great deal more that must be left unsaid. In particular, we omit any discussion of how the inverse theorem can be applied. The qualitative version has already been put to a variety of uses, and now that Manners has proved an effective quantitative version, its power has only grown. For the reader interested in seeing how inverse results can be applied to number theoretic problems we refer to the work of GREEN and TAO (2010b) on linear equations in primes, or the book by TAO (2012) on higher order Fourier analysis.

In Section 3 we will define carefully what a nilsequence is, completing the statement of Theorem 1.1. Before going into details, however, we give a very rough sketch of how

proofs of inverse theorems tend to go. The reader unfamiliar with nilsequences can, for now, read ‘a nilsequence of degree s ’ as ‘ $e(P(n))$ ’ where $P \in \frac{1}{N}\mathbb{Z}[x]$ is a polynomial of degree s ’.

2. A BIRD’S EYE VIEW

We fix some $\delta > 0$ and let $f : \mathbb{Z}_N \rightarrow \mathbb{C}$ be a 1-bounded function such that $\|f\|_{U^{s+1}} \geq \delta$. Our goal is to find some nilsequence ψ of degree s such that f ‘correlates’ with ψ , that is, $|\mathbb{E} f \overline{\psi}| \gg_\delta 1$.⁽⁷⁾ The case $s = 1$ has already been proved using Fourier analysis, allowing for an inductive approach.

We first sketch the approach of GREEN, TAO, and ZIEGLER (2012). They first use the identity

$$(3) \quad \|f\|_{U^{s+1}}^{2^{s+1}} = \mathbb{E}_h \|\Delta_h f\|_{U^s}^{2^s}$$

to deduce that if $\|f\|_{U^{s+1}} \geq \delta$ then there are $\gg_\delta N$ many h such that $\|\Delta_h f\|_{U^s} \gg_\delta 1$, and hence by induction there exist corresponding nilsequences ψ_h of degree $s - 1$ which correlate with $\Delta_h f$. By a Cauchy-Schwarz argument similar to one used by GOWERS (2001), they then show that the function $h \mapsto \psi_h$ is ‘approximately linear’, in that there are $\gg_\delta N^3$ many quadruples (h_1, h_2, h_3, h_4) such that

$$(4) \quad h_1 + h_2 = h_3 + h_4 \quad \text{and} \quad \psi_{h_1} \psi_{h_2} \approx \psi_{h_3} \psi_{h_4} \quad (\text{for some notion of } \approx).$$

This argument makes crucial use of the cocycle identity (valid for all $f : \mathbb{Z}_N \rightarrow \mathbb{C}$ such that $|f(x)| \equiv 1$)

$$(5) \quad \Delta_{h+k} f(x) = \Delta_k f(x) \Delta_h f(x+k),$$

which is a trivial consequence of the definition of Δ . They then use tools from additive combinatorics to further regularise the structure of the function $h \mapsto \psi_h$, after passing to a smaller set of h , so that on these h the map $h \mapsto \psi_h$ is genuinely linear.

After a ‘symmetrisation’ argument, which revolves around the symmetry property

$$(6) \quad \Delta_h \Delta_k f(x) = \Delta_k \Delta_h f(x),$$

one can explicitly write $\psi_h(x) \approx \Delta_h \Psi(x)$, where Ψ is a nilsequence of degree s , at least up to terms of degree $s - 2$. Since each $\Delta_h f$ correlates with ψ_h and ψ_h agrees with $\Delta_h \Psi$ up to a nilsequence of degree $s - 2$, it follows that $\Delta_h(f \overline{\Psi})$ correlates with a degree $s - 2$ nilsequence. By the converse to the inverse theorem it follows that $\|\Delta_h(f \overline{\Psi})\|_{U^{s-1}} \gg_\delta 1$. Since this is true for $\gg_\delta N$ many h , the recursive definition (3) implies that $\|f \overline{\Psi}\|_{U^s} \gg_\delta 1$ and so, by induction, there is a nilsequence Ψ' of degree $s - 1$ such that $f \overline{\Psi}$ correlates with Ψ' , and hence f correlates with $\Psi \Psi'$, a nilsequence of degree s , as required.

⁽⁷⁾Here we use the Vinogradov notation $\gg_\delta 1$ to mean $> c(\delta)$ for some function depending only on δ .

Putting all this into practice is highly non-trivial, and requires frequent recourse to technical results on the quantitative equidistribution of polynomial sequences on nilmanifolds developed by GREEN and TAO (2012).

Manners takes a different approach, which has more in common with the original paper of GOWERS (2001) (which proved a quantitatively reasonable *local* inverse theorem). Instead of using that for many $h \in \mathbb{Z}_N$ the derivative $\Delta_h f$ has large U^s norm, then using induction and working with degree $s - 1$ nilsequences, Manners differences all the way down to U^2 , finding many $\mathbf{h} \in \mathbb{Z}_N^{s-1}$ such that $\Delta_{\mathbf{h}} f$ has large U^2 norm, and then applies Fourier analysis there, before working all the way back up again.

Applying the U^2 inverse theorem to $\Delta_{\mathbf{h}} f$ produces a function $\mathbf{h} \mapsto \psi_{\mathbf{h}}$, say, which finds some linear character with which $\Delta_{\mathbf{h}} f$ correlates. Again by repeated use of the Cauchy–Schwarz inequality we may then find some structure within this ψ function. Rather than show that it is approximately a linear function, however, as in (4), we obtain something much weaker, and can only show that it is approximately a polynomial of degree $s - 1$.

One is then faced with the daunting task of understanding the structure of such approximate polynomials for higher degree, rather than just the linear case. This is the principal component of the proof of MANNERS (2018): a structural theorem for approximate polynomials of any degree $s \geq 1$. He proves that any such approximate polynomial must resemble a structured algebraic object called a nilpolynomial.

With such an inverse result for approximate polynomials in hand, the inverse result for uniformity norms is (relatively) straightforward to deduce. It is the proof of the inverse result for approximate polynomials that takes up the bulk of the paper of Manners, requiring an argument of extraordinary subtlety and care, especially since we need to keep careful track of all quantitative parameters, and ensure they do not grow out of control.

We will discuss the proof of the inverse result for approximate polynomials in Sections 5 and 6 of this article, and state it precisely (along with defining what a nilpolynomial is) in Section 4. Before then, however, we return to the inverse theorem itself, and define what a nilsequence actually is.

3. NILOBJECTS

In this section we will build towards a precise definition of a nilsequence, which are the obstructions to having small uniformity norm that appear in the statement of the inverse theorem. Nilsequences are functions $\psi : \mathbb{Z} \rightarrow \mathbb{C}$ that should certainly include the classical (phase) polynomials $x \mapsto e(P(x))$, where $P \in \mathbb{R}[x]$, but also need to allow for the possibility of more complicated ‘bracket’ structure, as alluded to in the introduction.

The reader may wonder why we are discussing nilsequences from $\mathbb{Z} \rightarrow \mathbb{C}$, and then obtaining well-defined functions on \mathbb{Z}_N by insisting on N -periodicity, rather than directly defining nilsequences from $\mathbb{Z}_N \rightarrow \mathbb{C}$. The latter is possible, using the machinery of Host–Kra cube groups, as described in MANNERS (2018, Appendix C). For this expository article we have adopted a less conceptual approach, which is easier to describe, and simply insist on N -periodicity by fiat.

To motivate the construction of nilsequences, we focus on one simple example: the function $n \mapsto e(\alpha n^s)$, which our informal discussion suggests should be a nilsequence of degree s . This function can be broken down into three steps, factoring first through \mathbb{R} and then the quotient space \mathbb{R}/\mathbb{Z} :

$$n \longmapsto \alpha n^s \longmapsto \{\alpha n^s\} \longmapsto e(\alpha n^s)$$

$$\mathbb{Z} \longrightarrow \mathbb{R} \longrightarrow \mathbb{R}/\mathbb{Z} \longrightarrow \mathbb{C}.$$

This factor construction is how we will define a general nilsequence, but we want to have the flexibility of allowing the intermediate group \mathbb{R} to be replaced by some other, possibly non-abelian, group G . The first map from $\mathbb{Z} \rightarrow G$ should still be a polynomial of degree s (whatever this means for a general group), and the second should be from this group to some kind of compact quotient space. The third will then be a nicely behaved (in our case, Lipschitz) function from this quotient space to \mathbb{C} .

Before we give the full definition of a nilsequence we need to first say exactly what we mean by ‘a polynomial of degree s from $\mathbb{Z} \rightarrow G$ ’ where G is some (not necessarily abelian) group, and then to say what kind of quotient structure is needed.

3.1. Polynomial maps

Intuitively, a polynomial map of degree s is one whose derivatives of order $s + 1$ all vanish. This suffices for a definition when, for example, considering maps from \mathbb{R} to \mathbb{R} . We need to generalise the concept to handle maps from arbitrary groups to other arbitrary groups, which means we need to be more careful both about what a ‘derivative of order $s + 1$ ’ means, and, indeed, what it means to ‘vanish’. The correct generalisation of polynomial maps to arbitrary groups was given by LEIBMAN (2002).

To properly define polynomial maps we need to consider not just groups, but groups equipped with an additional structure. A *filtered group* of degree s is a group G equipped with a filtration of subgroups⁽⁸⁾ (closed if we are considering topological groups):

$$G = G_0 = G_1 \geq G_2 \geq \cdots \geq G_s \geq \{e\} = G_{s+1} = G_{s+2} = \cdots$$

such that if $g_i \in G_i$ and $g_j \in G_j$ then $[g_i, g_j] = g_i^{-1}g_j^{-1}g_i g_j \in G_{i+j}$. Observe that a filtered group is, as we have defined it, necessarily nilpotent (of class at most s), and note that a filtered group of degree s is also one of any degree $s' \geq s$. If G is abelian

⁽⁸⁾Manners calls this a ‘proper filtration’, and allows $G_0 \neq G_1$ in a filtration. This has a few technical advantages, but for this exposition we will simplify matters by always assuming $G_0 = G_1$.

then it has a filtration of any degree $s \geq 1$, the standard degree s filtration, where $G_i = G$ for $0 \leq i \leq s$, which we denote by $G_{(s)}$.

The most important example is $\mathbb{R}_{(s)}$, the standard degree s filtration of the reals. Another illustrative example to bear in mind, and our principal non-abelian example, is the degree 2 filtration of the Heisenberg group (when we speak of the Heisenberg group it will always be with this filtration) defined by

$$\begin{pmatrix} 1 & \mathbb{R} & \mathbb{R} \\ 0 & 1 & \mathbb{R} \\ 0 & 0 & 1 \end{pmatrix} = G_0 = G_1 \geq G_2 = \begin{pmatrix} 1 & 0 & \mathbb{R} \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

We extend the definition of derivative to maps between arbitrary groups by defining $\Delta_h p : H \rightarrow G$, for $p : H \rightarrow G$ and $h \in H$, by

$$\Delta_h p(x) = p(x)p(hx)^{-1},$$

and when $\mathbf{h} \in H^m$ we define $\Delta_{\mathbf{h}} p = \Delta_{h_1} \cdots \Delta_{h_m} p$.

DEFINITION 3.1 (Polynomial map). — *If H and G are both filtered groups then a polynomial map $p : H \rightarrow G$ is a continuous map such that for all $x \in H$ and $\mathbf{h} \in H^m$, if $h_j \in H_{i_j}$ for $1 \leq j \leq m$ then*

$$\Delta_{\mathbf{h}} p(x) \in G_{i_1 + \cdots + i_m}.$$

If A is an abelian group then a polynomial from H to A of degree s is a polynomial map in the above sense from H to $A_{(s)}$, the standard degree s filtration of A , when $s \geq 1$, and a constant map when $s = 0$.

For example, if $G = \mathbb{R}_{(s)}$ then this condition is just saying that $\Delta_{h_1, \dots, h_{s+1}} p(x) = 0$ for all $x, h_1, \dots, h_{s+1} \in H$. This is therefore a generalisation of the standard definition that ‘a polynomial of degree s is a function whose derivatives of order $s + 1$ identically vanish’. Note that between filtered groups we do not need to specify the degree of the polynomial map, since this information is already present in the degree of the filtered groups H and G .

For our purposes, the domain H will usually be the filtered group $\mathbb{Z}_{(1)}$, in which case the definition becomes

$$\Delta_{h_1, \dots, h_m} p(x) \in G_m$$

for all $x, h_1, \dots, h_m \in \mathbb{Z}$. For example, polynomial maps from $\mathbb{Z}_{(1)}$ to $\mathbb{R}_{(s)}$ are exactly those polynomials in $\mathbb{Z}[x]$ of degree at most s , and polynomial maps from $\mathbb{Z}_{(1)}$ to the Heisenberg group are those of the form

$$n \mapsto \begin{pmatrix} 1 & c_1 n + c_2 & c_3 n^2 + c_4 n + c_5 \\ 0 & 1 & c_6 n + c_7 \\ 0 & 0 & 1 \end{pmatrix}$$

where $c_1, \dots, c_7 \in \mathbb{R}$ are any constants. It can be shown that, with the definition we have given, the only polynomial maps from a *finite* group to a torsion-free abelian group are constant maps. This is part of the reason that we need to define nilsequences with

domain \mathbb{Z} , rather than \mathbb{Z}_N . As mentioned above, there is a more abstract approach that allows for a non-trivial notion of non-constant polynomial maps on \mathbb{Z}_N , which is that used by Manners, but we will not pursue this here.

The set of polynomial maps as defined here is, just like the set of classical polynomials, closed under pointwise multiplication and composition. This is not at all obvious from the definition! This remarkable fact was proved by LEIBMAN (2002). (In fact the set of polynomial maps can itself be given the structure of a filtered group.)

3.2. Nilmanifolds

Having defined what we mean by a polynomial map from \mathbb{Z} to G , where G is a filtered group of degree s , we now need to say what kind of quotient structure is needed for a nilsequence. We would like to have structure sufficient to capture the prototypical example of a nilsequence discussed above, which factors through $\mathbb{Z} \rightarrow \mathbb{R} \rightarrow \mathbb{R}/\mathbb{Z}$. In general we will replace \mathbb{R} by some other \mathbb{R}^d , given the structure of a (filtered) Lie group, such as the normal additive group structure on \mathbb{R}^d , or the Heisenberg group. We would then like to quotient out by ‘the integer part’ of \mathbb{R}^d , in a way which is compatible with the Lie group structure. This leads us to the following definition of a nilmanifold.

DEFINITION 3.2 (Nilmanifold). — *Let $s \geq 1$ be an integer, and $d_1, \dots, d_s, M \geq 0$ be any integers. A nilmanifold G/Γ of degree s , dimension (d_1, \dots, d_s) , and complexity M , is the quotient space of a pair (G, Γ) where*

- G is a simply connected filtered Lie group of degree s , where the filtration subgroup G_i is a real manifold with dimension $d_i + \dots + d_s$, with ‘generators’ γ_{ij} for $1 \leq i \leq s$ and $1 \leq j \leq d_i$ (i.e. homomorphisms $\gamma_{ij} : \mathbb{R} \rightarrow G$ where we write $\gamma_{ij}^a = \gamma_{ij}(a)$),
- every $g \in G$ has a unique representation as

$$\prod_{1 \leq i \leq s} \prod_{1 \leq j \leq d_i} \gamma_{ij}^{a_{ij}}$$

for some ‘coordinates’ $a_{ij} \in \mathbb{R}$, and the associated bijection $\sigma : G \rightarrow \mathbb{R}^{d_1 + \dots + d_s}$ is a homeomorphism,

- the filtration subgroup G_r consists of those g such that $a_{ij} = 0$ for all $i < r$ and $1 \leq j \leq d_i$,
- for all i, j, k, l , the coordinates of the commutator $[\gamma_{ij}, \gamma_{kl}]$ are integers bounded by M , and
- $\Gamma \leq G$ is the subgroup consisting of all those elements with integral coordinates.

We also write $D = d_1 + \dots + d_s$ for the dimension of G/Γ .

We caution that G/Γ is not necessarily a group, but it is a compact space which has $[0, 1)^D$ as a fundamental domain. Heuristically, one should think of a nilmanifold of dimension D as the cube $[0, 1)^D$ acted upon by a D -dimensional Lie group (although, again, we stress that it is not necessarily actually a group itself). The complexity parameter M allows us to swap coordinates in a representation of an element of G at a ‘cost’ of at most M . It offers some control of precisely how non-abelian our nilmanifold

is. Furthermore, the coordinate homeomorphism $\sigma : G \rightarrow \mathbb{R}^D$ allows us to define a metric on G/Γ by

$$d(x_1\Gamma, x_2\Gamma) = \inf_{\gamma_1, \gamma_2 \in \Gamma} \|\sigma(x_1\gamma_1) - \sigma(x_2\gamma_2)\|_1,$$

and hence we can speak of Lipschitz functions on G/Γ .

This is the definition that Manners uses, which is quite explicit. The usual, more compact, definition, as used by GREEN, TAO, and ZIEGLER (2012) for example, is that a nilmanifold is a quotient G/Γ where G is any connected, simply connected, nilpotent Lie group, and Γ is a discrete subgroup such that G/Γ is compact. This can be shown to be equivalent to our explicit definition using the lower central series filtration and the machinery of Mal'cev coordinates, but we prefer to be explicit from the outset.

The simplest abelian example is the nilmanifold of degree s , dimension $(0, \dots, 0, 1)$, and complexity 0 given by $G = \mathbb{R}_{(s)}$ and $\Gamma = \mathbb{Z}$. A slightly less trivial abelian example is the nilmanifold of degree s , dimension $(1, \dots, 1)$, and complexity 0, where $G = \mathbb{R}^s$ is given a filtration of degree s by

$$G_1 = \mathbb{R}^s \geq \mathbb{R}^{s-1} \geq \dots \geq \mathbb{R} \geq \{0\}$$

and $\Gamma = \mathbb{Z}^s$. The simplest non-abelian example is the Heisenberg filtered group discussed above, which (quotiented out by the obvious integral subgroup) is a nilmanifold of degree 2, dimension $(2, 1)$, and complexity 1.

3.3. Nilsequences

Having defined a general polynomial map and an appropriate quotient object to factor through we can now give the definition of a nilsequence.

DEFINITION 3.3 (Nilsequence). — *A nilsequence $\psi : \mathbb{Z} \rightarrow \mathbb{C}$ of degree s , dimension D , complexity M , and parameter K , is a function that factors as*

$$\mathbb{Z} \xrightarrow{p} G \xrightarrow{\pi} G/\Gamma \xrightarrow{F} \mathbb{C}$$

where

1. G/Γ is a nilmanifold of degree s , dimension D , and complexity M ,
2. $p : \mathbb{Z} \rightarrow G$ is a polynomial map;
3. $\pi : G \rightarrow G/\Gamma$ is the projection map; and
4. $F : G/\Gamma \rightarrow \mathbb{C}$ is Lipschitz with constant K .

We say that ψ is N -periodic if $\pi \circ p : \mathbb{Z} \rightarrow G/\Gamma$ is.

For example, if we take the abelian nilmanifold $G = \mathbb{R}^s$ and $\Gamma = \mathbb{Z}^s$ as above, then $p : \mathbb{Z} \rightarrow \mathbb{R}^s$ will have the form

$$p(n) = (p_1(n), \dots, p_s(n))$$

where $p_j \in \mathbb{R}[x]$ is a polynomial of degree j . If we take, for example, the Lipschitz function from $(\mathbb{R}/\mathbb{Z})^s$ to \mathbb{C} defined by $(\alpha_1, \dots, \alpha_s) \mapsto e(\alpha_1 + \dots + \alpha_s)$, then we see that

any function of the form $n \mapsto e(P(n))$, where $P \in \mathbb{R}[x]$ is a polynomial of degree s , is indeed a nilsequence of degree s as expected.

One would also expect that the more complicated phase polynomials, where we allow the bracket operation also, are nilsequences. There is an unfortunate technical hitch, however, which means they are not precisely nilsequences as we have defined them. Consider, for example, the function $n \mapsto e(-\alpha n \lfloor \beta n \rfloor)$, which we would expect to be a nilsequence of degree 2. One would like to say it is, using the Heisenberg nilmanifold. Indeed, one can check that the map

$$n \mapsto \begin{pmatrix} 1 & \alpha n & 0 \\ 0 & 1 & \beta n \\ 0 & 0 & 1 \end{pmatrix}$$

is indeed a polynomial map from \mathbb{Z} to the Heisenberg group, and the projection map $\pi : G \rightarrow G/\Gamma$ for the Heisenberg nilmanifold has the form

$$\begin{pmatrix} 1 & x & z \\ 0 & 1 & y \\ 0 & 0 & 1 \end{pmatrix} \mapsto \begin{pmatrix} 1 & \{x\} & \{z - x\lfloor y \rfloor\} \\ 0 & 1 & \{y\} \\ 0 & 0 & 1 \end{pmatrix}$$

(because with this choice there is a unique $\gamma \in \Gamma = \begin{pmatrix} 1 & \mathbb{Z} & \mathbb{Z} \\ 0 & 1 & \mathbb{Z} \\ 0 & 0 & 1 \end{pmatrix}$ such that $g = \pi(g)\gamma$), so the first two steps of the definition work as expected. The problem is that, identifying G/Γ with the fundamental domain $[0, 1]^3$, the map $F : G/\Gamma \rightarrow \mathbb{C}$ defined by $F(a, b, c) = e(c)$ is not Lipschitz — indeed, it is not even continuous! This is because as y varies from 0.99 to 1.01 the floor function $\lfloor y \rfloor$ has a discontinuous jump, and hence the c coordinate of

$$\begin{pmatrix} 1 & x & z \\ 0 & 1 & y \\ 0 & 0 & 1 \end{pmatrix}$$

in the fundamental domain, which is $\{z - x\lfloor y \rfloor\}$, will jump discontinuously (assuming $x \neq 0$). It follows that F is not a continuous function. In their work on the U^4 norm GREEN, TAO, and ZIEGLER (2011) allow for F to be discontinuous, so that these bracket phase polynomials are genuine nilsequences, but the ensuing analytic complications are hard to deal with for $s \geq 4$. We therefore use the definition above, where the function F is analytically quite nice, the disadvantage being that the ‘non-abelian’ nilsequences do not quite look how one would expect if one is guided by bracket polynomials as the motivating examples. The reader should view this as a purely technical issue, however, and would not lose any intuition in continuing to view nilsequences as exponentials of bracket polynomials for all intents and purposes.

Finally, we remark that many of the analytic arguments when dealing with nilsequences become much simpler if we insist that F is a smooth function (with control on the associated Sobolev norms), rather than just Lipschitz. This alternative definition will hopefully be adopted in future work on the subject, but almost all of the literature

(including the work of Manners) works with F as Lipschitz, and so we have followed convention.

3.4. Nilsequences as obstructions

Finally, we sketch a proof that nilsequences are indeed obstructions to having small uniformity norm, which is the (easier) converse to the inverse theorem. This is a trivial repeated differencing procedure when the nilsequence has the simple form $n \mapsto e(P(n))$ for $P(x) \in \mathbb{R}[x]$ a genuine polynomial, but the more general case requires some work, and was first proved by GREEN and TAO (2008a). We sketch an alternative proof given by GREEN, TAO, and ZIEGLER (2011, Appendix G).

THEOREM 3.1 (Green–Tao–Ziegler). — *If $f : \mathbb{Z}_N \rightarrow \mathbb{C}$ is a 1-bounded function and there exists a 1-bounded, N -periodic, nilsequence $\psi : \mathbb{Z} \rightarrow \mathbb{C}$ of degree s , complexity M , parameter K , and dimension D such that*

$$\left| \mathbb{E}_x f(x) \overline{\psi(x)} \right| \geq \delta$$

then there exists $\epsilon > 0$, depending only on (s, δ, M, K, D) , such that

$$\|f\|_{U^{s+1}} > \epsilon.$$

The dependence of ϵ , once we have fixed s , is polynomial in all the other parameters.

Sketch proof. — We use induction on s . The case $s = 1$ is simple Fourier analysis coupled with the fact that all degree 1 nilsequences of bounded complexity, parameter, and dimension can be approximated by a short trigonometric sum.

For general $s \geq 2$, suppose the nilsequence factors as $\psi = F \circ \pi \circ p$. Since $G_s/(\Gamma \cap G_s)$ is a compact abelian group we can use Fourier analysis to expand F as the sum of F_χ as χ ranges over the dual group of $G_s/(\Gamma \cap G_s)$, where

$$F_\chi(x) = \int_{G_s/(\Gamma \cap G_s)} F(xg_s) \overline{\chi(g_s)} dg_s.$$

Furthermore, each F_χ has a vertical character, which means that for all $g_s \in G_s/(\Gamma \cap G_s)$ and all $x \in G/\Gamma$ we have $F_\chi(g_s x) = \chi(g_s) F_\chi(x)$. After proving suitable decay properties of this Fourier series, we may find a single character χ such that

$$(7) \quad \left| \mathbb{E}_x f(x) \overline{F_\chi \circ \pi \circ p(x)} \right| \gg_\delta 1.$$

We now consider the function $\tilde{F}_h : \mathbb{Z}_N \rightarrow \mathbb{C}$, defined for any $h \in \mathbb{Z}_N$ by

$$\tilde{F}_h(x) = F_\chi(p(x)\Gamma) \overline{F_\chi(p(x+h)\Gamma)}.$$

Squaring (7) and expanding out the brackets, then using the pigeonhole principle, we deduce that there exist $\gg_\delta N$ many $h \in \mathbb{Z}_N$ such that

$$\left| \mathbb{E}_x \Delta_h f(x) \tilde{F}_h(x) \right| \gg_\delta 1.$$

One can show (see GREEN, TAO, and ZIEGLER (2011, Appendix G) for details) that \tilde{F}_h is, for any fixed h , itself a nilsequence of degree $s-1$ (this is not entirely straightforward, and some care must be taken to ensure that the complexity does not increase by too much). The inductive hypothesis then implies that $\|\Delta_h f\|_{U^s} \gg_\delta 1$ for $\gg_\delta N$ many h , and thus

$$\|f\|_{U^{s+1}}^{2^{s+1}} = \prod_h \|\Delta_h f\|_{U^s}^{2^s} \gg_\delta 1$$

as required. \square

With all the terms used in the statement of Theorem 1.1 defined, we now turn to sketching the proof of Manners.

4. AN INVERSE THEOREM FOR APPROXIMATE POLYNOMIALS

The core of Manners' proof is another inverse result, but not for uniformity norms, rather for the more combinatorial notion of an approximate polynomial. We first introduce the additive discrete derivative. Let H be an abelian group, then for any $f : H \rightarrow \mathbb{R}$ and $h \in H$ we define $\partial f : H \rightarrow \mathbb{R}$ by

$$\partial_h f(x) = f(x) - f(x+h),$$

again extended to multiple derivatives $\partial_{\mathbf{h}}$ in the natural fashion.⁽⁹⁾ Recall that a polynomial of degree s from H to \mathbb{R} is one whose derivatives of order $s+1$ are all identically zero — that is, a function f such that $\partial_{\mathbf{h}} f(x) = 0$ for all $(x, \mathbf{h}) \in H^{s+2}$. To obtain an approximate polynomial we weaken this to just holding for some small proportion of H^{s+2} .

DEFINITION 4.1 (Approximate polynomial). — *Let H be a finite abelian group. Let $\epsilon > 0$ and $X \subset H$. A function $f : X \rightarrow \mathbb{R}$ is an ϵ -polynomial of degree s on X if*

$$\#\{(x, \mathbf{h}) \in H^{s+2} : \partial_{\mathbf{h}} f(x) = 0\} \geq \epsilon |H|^{s+2},$$

where the set ranges over those (x, \mathbf{h}) such that $\partial_{\mathbf{h}} f(x)$ is defined (i.e. $x + \omega \cdot \mathbf{h} \in X$ for all $\omega \in \{0, 1\}^{s+1}$). In particular, it follows that $|X| \gg_s \epsilon |H|$.

It is an instructive calculation to verify that all of the following are examples of approximate polynomials (for some $\epsilon > 0$ bounded away from zero, that is dependent on the precise construction):

1. bracket polynomials, such as $x \mapsto \{mx/N\}$ for some fixed $m \in \mathbb{Z}$;
2. any function taking any K arbitrary values (where K is independent of H , and ϵ can depend on K);
3. a function that agrees with a bracket polynomial on a positive density set and is otherwise random; and

⁽⁹⁾This is just the derivative operator Δ applied to functions taking values in the additive group of the reals, but we find it helpful to have a reminder of whether we're thinking additively or multiplicatively.

4. a function which switches between a few different bracket polynomials as we move across some arbitrary partition of \mathbb{Z}_N .

The inverse result mentioned above, and the keystone of Manners' work, is the fact that examples of these kinds are, essentially, the only types of approximate polynomials. Before stating it, we will define the concept of nilpolynomials, which are to bracket polynomials as nilsequences are to phase bracket polynomials. Again, this is because they are technically far more convenient to work with, although they are essentially equivalent, and the reader can think of them as bracket polynomials if they prefer.

4.1. Nilpolynomials

We need to work in a slightly more general context than just cyclic groups, so we give the definition for \mathbb{Z}^d for any $d \geq 1$.

DEFINITION 4.2 (Nilpolynomial). — *A nilpolynomial $f : \mathbb{Z}^d \rightarrow \mathbb{R}$ of degree s , dimension D , and complexity M , is a function that factors as*

$$\mathbb{Z}^d \xrightarrow{\rho} G \xrightarrow{F} \mathbb{R}$$

where

1. there is $\Gamma \leq G$ such that G/Γ is a nilmanifold of degree s , dimension D , and complexity M ,
2. there is some polynomial map $p : \mathbb{Z}^d \rightarrow G$ such that $\pi \circ p = \pi \circ \rho$, where π is the usual projection map $\pi : G \rightarrow G/\Gamma$,
3. ρ maps into the ball of radius M around the identity, and
4. $F : G \rightarrow \mathbb{R}$ is a polynomial map of degree s .

We say that f is N -periodic if the function ρ is (so that the nilpolynomial induces a function $f : \mathbb{Z}_N^d \rightarrow \mathbb{R}$).

The point is that we are not just limiting ourselves to compositions $F \circ p$, where $p : \mathbb{Z}^d \rightarrow G$ and $F : G \rightarrow \mathbb{R}$ are polynomials, but we also allow the flexibility of combining p with some arbitrary (but bounded) 'lift' from G/Γ to G . Roughly speaking, therefore, nilpolynomials are 'polynomials factored through nilmanifolds', although we stress that they are *not* (necessarily) polynomial maps from \mathbb{Z}^d to \mathbb{R} themselves. If one develops more machinery, as Manners does, one can define a notion of polynomial maps from \mathbb{Z}_N^d to nilmanifolds directly, but for our purposes, the definition of nilpolynomial that results is equivalent to what we have presented here.

For example, if G is the abelian nilmanifold \mathbb{R}^s , taking $\rho(n) = (0, \dots, 0, \{p(n)\})$ and $F : \mathbb{R}^s \rightarrow \mathbb{R}$ as the projection onto the last coordinate, where $p(n)$ is a polynomial of degree s , exhibits $n \mapsto \{p(n)\}$ as a nilpolynomial of degree s , as expected. It is N -periodic if $p(x) \in \frac{1}{N}\mathbb{Z}[x]$. Furthermore, the map $n \mapsto \{p(n)\} + r(n)$ is also a nilpolynomial, now of complexity M , for any arbitrary function $r : \mathbb{Z} \rightarrow \mathbb{Z}$ taking values in $(-M, M)$ (and is still N -periodic if r is). More general bracket polynomials, such as

$n \mapsto \{\alpha n\}\{\beta n\}$, are also nilpolynomials, and can be constructed through Heisenberg-type nilmanifolds as in the previous section.

The point of the parameter M , and in allowing ρ to take values outside the fundamental domain of G/Γ , is to give greater flexibility to the function ρ so that it does not itself need to exhibit strict polynomial behaviour, even though $\pi \circ \rho$ does. This means that we can construct functions that take on a bounded number of arbitrary values as genuine nilpolynomials in their own right.

For example, suppose that $A_1 \sqcup \cdots \sqcup A_r$ is a partition of \mathbb{Z} , and f takes the value $c_i \in \mathbb{R}$ on A_i . Let $G = \mathbb{R}^r$ with the standard degree 1 filtration, and let

$$\rho(x) = (1_{A_1}(x), \dots, 1_{A_r}(x)).$$

If we choose $F : \mathbb{R}^r \rightarrow \mathbb{R}$ to be the affine projection

$$(x_1, \dots, x_r) \mapsto c_1 x_1 + \cdots + c_r x_r,$$

then $F \circ \rho = f$ as required. This exhibits f as a nilpolynomial of degree 1, dimension r , and complexity 1.

4.2. The inverse result for approximate polynomials

One can show that any nilpolynomial is an approximate polynomial. Manners proves the following strong inverse theorem.

THEOREM 4.1 (Every approximate polynomial is approximately a nilpolynomial)

If $X \subset \mathbb{Z}_N^d$ and $f : X \rightarrow \mathbb{R}$ is an ϵ -polynomial of degree s then there is an N -periodic nilpolynomial $g : \mathbb{Z}^d \rightarrow \mathbb{R}$ of degree s , dimension $O(\epsilon^{-O(1)})$, and complexity M such that

$$f(x) = g(x) \quad \text{for at least } (\epsilon/2)^{O(1)} |X| \text{ many } x \in X$$

and

$$M \leq \begin{cases} \exp(O(\epsilon^{-O(1)})) & \text{if } s \leq 3 \text{ and} \\ \exp(\exp(O(\epsilon^{-O(1)}))) & \text{if } s \geq 4. \end{cases}$$

We now sketch how Manners deduces the inverse theorem for uniformity norms from this inverse theorem for approximate polynomials. After this, we will focus on the proof of Theorem 4.1, and make no further mention of uniformity norms.

Sketch proof of Theorem 1.1 from Theorem 4.1. — This part of the proof of Manners uses Fourier analysis combined with elaborate but elementary arguments involving repeated use of the Cauchy–Schwarz inequality. In this it has more in common with the work of GOWERS (2001) than the work of GREEN, TAO, and ZIEGLER (2012), which replaces the use of Fourier analysis with a far more complicated generalised theory of nilsequences of higher degree. This is inevitable in their approach as they have much less combinatorial information about the functions they are working with, and hence have to use as an inductive hypothesis the inverse theorem for the U^s norm quite early in the argument. This produces an ensemble of nilsequences of degree s , which must then be explicitly stitched together somehow to form a nilsequence of degree $s + 1$.

Instead of using the fact that $\|f\|_{U^{s+1}}^{2^{s+1}} = \mathbb{E}_{\mathbf{h}} \|\Delta_{\mathbf{h}} f\|_{U^s}^{2^s}$ to relate the U^{s+1} norm of f to U^s -type information for the 1-variable function $\Delta_{\mathbf{h}} f$, Manners uses the relationship

$$\|f\|_{U^{s+1}}^{2^{s+1}} = \mathbb{E}_{\mathbf{h} \in \mathbb{Z}_N^{s-1}} \|\Delta_{\mathbf{h}} f\|_{U^2}^4.$$

We can then use the simple inverse theorem for the U^2 norm (Manners does not use precisely this as a black box, but uses Fourier arguments of a similar flavour) to deduce that for many $\mathbf{h} \in \mathbb{Z}_N^{s-1}$ there exists some $\chi_{\mathbf{h}} \in \widehat{\mathbb{Z}_N}$ such that

$$\left| \mathbb{E}_x \Delta_{\mathbf{h}} f(x) \overline{\chi_{\mathbf{h}}(x)} \right| \gg_{\delta} 1.$$

Repeated use of the Cauchy–Schwarz inequality then implies that the function $\Phi : \mathbb{Z}_N^{s-1} \rightarrow \widehat{\mathbb{Z}_N}$ which selects this large character $\chi_{\mathbf{h}}$ for each \mathbf{h} is an approximate polynomial of degree $s - 1$, and we can apply Theorem 4.1 to find some nilpolynomial P which agrees with Φ for many $\mathbf{h} \in \mathbb{Z}_N^{s-1}$.

There are some serious difficulties at this point arising from the fact that the set of \mathbf{h} where Φ agrees with P is possibly quite sparse, which Manners overcomes by an elegant argument involving a partition of unity and arguing with a number of nilpolynomials at once. Ignoring such issues, we can show that if

$$S_s f(\mathbf{h}) = \mathbb{E}_x \Delta_{\mathbf{h}} f(x)$$

there exists some nilpolynomial P such that

$$\left| \mathbb{E}_{\mathbf{h} \in \mathbb{Z}_N^{s-1}} S_s f(\mathbf{h}) e(\overline{P(\mathbf{h})}) \right| \gg_{\delta} 1.$$

We have found correlation of the $(s - 1)$ -variable function $S_s f$ with a function that resembles a multi-variable nilsequence. It remains to deduce correlation of f itself with a nilsequence. This is done by projecting $e(P(h_1, \dots, h_{s-1}))$ down to a 1-variable function by replacing it by

$$x \mapsto e(P(h_1 + x, \dots, h_{s-1} + x))$$

for a randomly chosen \mathbf{h} . Further combinatorial arguments (again relying heavily on the Cauchy–Schwarz inequality) imply that f itself correlates with (essentially) this function, and using the nilpolynomial structure of P one can show that this 1-variable function is a nilsequence, thus completing the proof of Theorem 1.1.

While we have not discussed the behaviour of quantitative factors in this proof, the lack of iterative arguments and reliance on purely elementary arguments involving the Cauchy–Schwarz inequality should reassure the reader that the quantitative strength of Theorem 4.1 is transferred relatively undiminished to Theorem 1.1. \square

5. FROM 1% STRUCTURE TO 99% STRUCTURE

The rest of this article will discuss the proof of the inverse theorem for approximate polynomials, Theorem 4.1. In a high-level overview of an argument like this that uses many quantitative parameters it is convenient to use the 1% and 99% heuristic language. This means that we will describe a parameter as being 1% if it is something like $\epsilon > 0$, where ϵ is to be thought of as very small, and as being 99% if it is more on the scale of $1 - \epsilon$. We will continue to use precise notation in some places, but will not hesitate in using this rougher language where it clarifies the exposition, and refer the reader interested in seeing precise versions of what follows to MANNERS (2018).

The ideas employed in what follows are valid in any finite abelian group, so we will use H to denote an arbitrary finite abelian group (which, for our applications, will be \mathbb{Z}_N^d for some $d \leq s$, but that need not concern us here). Unless otherwise specified, X denotes some subset of H of 1% density (i.e. $|X| \geq \epsilon |H|$ for some small $\epsilon > 0$).

We are given some approximate polynomial $f : X \rightarrow \mathbb{R}$ of degree s , so that the function $\partial_{\mathbf{h}} f(x)$ vanishes for at least 1% of $(x, \mathbf{h}) \in H^{s+2}$. Our eventual goal is to show that there is a nilpolynomial P such that $P(x) = f(x)$ for 1% of $x \in H$. This will be done in three stages:

1. As f only has a 1% type of structure we will first need to find some g which has a 99% type of structure, such that $f(x) = g(x)$ for 1% of all $x \in H$;
2. We then show that a function which has this 99% type of structure must have some weak algebraic structure;
3. Finally, we show that a function with this weak algebraic structure must agree with a genuine nilpolynomial 1% of the time (and that this 1% can be arranged to be a subset of the 1% of values from (1), so that f itself agrees with a genuine nilpolynomial 1% of the time).

In this section we discuss the first of these steps. Heuristically, we aim to find some g which agrees with f for a 1% proportion of H , but such that g itself is a 99%-approximate polynomial. This cannot be established directly, but we will show that it is true, aside from possible obstructions of lower degree. We can then use the same argument to show that these obstructions themselves enjoy a 99% type of structure, modulo further obstructions of even lower degree, and so on. This cascading type of structure is captured by the notion of a ‘polynomial hierarchy’, and we will give the definition of this before stating the main result of this section.

5.1. Polynomial Hierarchies

We will be working with derivatives of different orders, for which the following language will be useful. A cube of dimension k is formally just an element of H^{k+1} , but we will generally write it as $c = (x, \mathbf{h})$ for some $x \in H$ and $\mathbf{h} \in H^k$. This should be thought of as a fixed basepoint x together with a k -dimensional direction vector \mathbf{h} . The coordinates of a cube are the h_1, \dots, h_k components, and the x component is referred

to as the basepoint. A d -dimensional face of a cube of dimension k is a d -dimensional cube obtained by removing any $k - d$ of the directional components and replacing the basepoint x by $x + \sum_{i \in I} h_i$ for some subset I of the removed $k - d$ components. When $c = (x, \mathbf{h})$ we will write

$$\partial f(c) = \partial_{\mathbf{h}} f(x),$$

so that in particular ∂f is a function on a set of cubes, the dimension of which is encoded in the cube c itself. Writing out the definition in full, we have

$$\partial f(c) = \sum_{\omega \in \{0,1\}^k} (-1)^{|\omega|} f(x + \omega \cdot \mathbf{h}),$$

where $|\omega| = \omega_1 + \dots + \omega_k$ is the Hamming weight of ω . Note that if c' is obtained from c by a permutation of the (directional) coordinates then $\partial f(c) = \partial f(c')$. Furthermore, if c' is obtained from c by a ‘coordinate reflection’, where $x \mapsto x + h_i$ and $h_i \mapsto -h_i$, then $\partial f(c') = -\partial f(c)$. This means that if we are considering the vanishing of ∂f on some set of cubes then it is natural to ask that such a set be closed under coordinate permutations and reflections.

Since we will also be moving between taking derivatives of different orders for different sets of cubes it is desirable to have some natural compatibility conditions between the sets of cubes at different levels. These conditions are captured by the following definition.

DEFINITION 5.1 (System of cubes). — *A system of cubes of degree s and density $\delta > 0$ is a family (S_0, \dots, S_{s+1}) of non-empty sets of cubes, with $S_i \subset H^{i+1}$, such that*

1. *each level is symmetric, i.e. each S_k is closed under permutations of the (directional) coordinates and under coordinate reflections, where $x \mapsto x + h_i$ and $h_i \mapsto -h_i$,*
2. *the system is closed under taking faces, so that if $c \in S_k$ and c' is a face of c of dimension d then $c' \in S_d$, and*
3. *for $0 \leq k \leq s$, any $(x, \mathbf{h}) \in S_k$ can be extended in at least $\delta |H|$ many ways to some $(x, \mathbf{h}, h_{k+1}) \in S_{k+1}$.*

Note that the second and third conditions ensure that $|S_0| \geq \delta |H|$, and hence by induction $|S_k| \geq \delta^{k+1} |H|^{k+1}$ for $0 \leq k \leq s + 1$. In our applications δ will usually be a 99% parameter (i.e. very close to 1) so a system of cubes will consist of 99% of all possible cubes.

The concept of a ‘derivatives condition’ underpins everything which follows, as it allows us to talk of the derivatives of some function being expressible in terms of simpler functions. For comparison, note that by the definition of derivative, namely

$$\partial g(c) = \sum_{\omega \in \{0,1\}^k} (-1)^{|\omega|} g(x + \omega \cdot \mathbf{h}),$$

we know that ∂g can be trivially expressed as a sum involving g and the projected cubes $x + \omega \cdot \mathbf{h}$. The (\mathbf{f}, k, t, M) -derivatives condition says that the k th derivative of g

can instead be expressed as a ‘low-complexity’ (measured by the parameter M) linear combination of functions from \mathbf{f} – moreover, only those of level $\leq t - 1$. In particular, if $t = 0$ this means that $\partial g \equiv 0$ on S . More precisely, we have the following.

DEFINITION 5.2 (Derivatives condition). — *Let $s \geq 1$, let (d_1, \dots, d_s) be positive integers, and let k and t be integers such that $0 \leq k, t \leq s + 1$. Let $X \subset H$ and $S \subset H^{k+1}$ be a set of cubes such that $x + \omega \cdot \mathbf{h} \in X$ for every $(x, \mathbf{h}) \in S$ and $\omega \in \{0, 1\}^k$. Let $M \geq 1$ be an integer. Let $\mathbf{f} = (f_i : X \rightarrow \mathbb{R}^{d_i})_{0 \leq i \leq s}$ be a family of functions and let $b : S \times \{0, 1\}^k \rightarrow \mathbb{Z}^{d_0 + \dots + d_{t-1}}$ be a function.*

A function $g : X \rightarrow \mathbb{R}$ satisfies the (\mathbf{f}, k, t, M) -derivatives condition on S with coefficients b if $\|b(c, \omega)\|_1 \leq M$ for all $c \in S$ and $\omega \in \{0, 1\}^k$, and for all $c = (x, \mathbf{h}) \in S$

$$\partial g(c) = \sum_{\omega \in \{0, 1\}^k} (-1)^{|\omega|} b(c, \omega) \cdot \mathbf{f}_{\leq t-1}(x + \omega \cdot \mathbf{h}).$$

Here $|\omega| = \omega_1 + \dots + \omega_k$ is the Hamming weight of ω , and $\mathbf{f}_{\leq t-1}$ is the family of functions $(f_i : X \rightarrow \mathbb{R}^{d_i})_{0 \leq i \leq t-1}$.

In isolation, this definition has very little content — indeed, every g satisfies the $(\mathbf{f}, k, 1, 1)$ -derivatives condition for any $k \geq 1$ if we let $f_0 = g$. Its importance is that it lets us talk about moving ‘between levels’, so that we can express the derivative of g in terms of simpler functions. It is important to note that we allow the function $b(c, \omega)$ to depend in some completely unspecified fashion on c and ω , and indeed its dependence may be quite unwieldy (which is a complication we will have to deal with later on). Crucially, however, it only takes bounded integer values, and so it is simpler in this sense than the less-controlled $\partial g : S \rightarrow \mathbb{R}$.

The following examples are instructive. Consider first the bracket linear function $g : \mathbb{Z}_N \rightarrow \mathbb{R}$ defined, for some fixed $a \in \mathbb{Z}$, by

$$g(x) = \{ax/N\}.$$

This is an approximate polynomial of degree 1. It is easy to check that $g(x + y) = g(x) + g(y) - b_{x,y}$ where $b_{x,y} \in \{0, 1\}$ (in fact, $b_{x,y} = 1$ if and only if $\{ax/N\} + \{ay/N\} \geq 1$). It follows that, for any $c = (x, h_1, h_2)$,

$$\begin{aligned} \partial g(c) &= g(x) - g(x + h_1) - g(x + h_2) + g(x + h_1 + h_2) \\ &= b_{x,h_2} - b_{x+h_1,h_2} \\ &= b(c) \end{aligned}$$

where $b(c) \in \mathbb{Z}$ and $|b(c)| \leq 1$. In particular, the second derivative of g is always a bounded integer. We do not have much control on how this integer depends on c , but the key point is that the derivative of g can be expressed as an integer linear combination of ‘simpler’ functions (in this case, just the constant function $\equiv 1$).

We now consider a more complicated example, the bracket quadratic function $g : \mathbb{Z}_N \rightarrow \mathbb{R}$ defined, for some fixed $a, b \in \mathbb{Z}$, by

$$g(x) = \{ax/N\}\{bx/N\},$$

which is an approximate polynomial of degree 2. An explicit calculation, similar to the above, shows that, for any $c = (x, h_1, h_2, h_3)$,

$$\partial g(c) = \sum_{\omega \in \{0,1\}^3} (-1)^{|\omega|} b(c, \omega) \cdot (1, \{ax/N\}, \{bx/N\}),$$

where $b(c, \omega) \in \mathbb{Z}^3$ and $\|b(c, \omega)\|_1 \leq 25$. In particular, the derivative of the bracket quadratic g , although it does not vanish completely like that of a genuine quadratic, can be expressed as a simple linear combination of the bracket linear functions 1 , $\{ax/N\}$, and $\{bx/N\}$. We encourage the reader to bear these examples in mind for what follows.

Having expressed the derivative of g in terms of simpler functions of a lower level, it is natural to attempt the same with these simpler functions. By applying the derivatives condition inductively, so the derivatives of these functions can themselves be simplified, and so on, we arrive at the following key definition.

DEFINITION 5.3 (Polynomial hierarchy). — *Let $X \subset H$ and (S_0, \dots, S_{s+1}) be a system of cubes such that, for $0 \leq i \leq s+1$, we have $x + \omega \cdot \mathbf{h} \in X$ for every $(x, \mathbf{h}) \in S_i$ and $\omega \in \{0, 1\}^i$.*

A system of functions $\mathbf{f} = (f_i : X \rightarrow \mathbb{R}^{d_i})_{0 \leq i \leq s}$ is a polynomial hierarchy of height s and complexity M on S_0, \dots, S_{s+1} if, for any $0 \leq i \leq s$, the function f_i satisfies the $(\mathbf{f}, i+1, i, M)$ -derivatives condition on S_{i+1} .

Furthermore, we say that a function $g : X \rightarrow \mathbb{R}$ sits at the top of a (D, M) -polynomial hierarchy of height s on (S_0, \dots, S_{s+1}) if there is a polynomial hierarchy \mathbf{f} of height $s-1$ and complexity M on (S_0, \dots, S_s) with $d_0 + \dots + d_s = D$ such that g satisfies the $(\mathbf{f}, s+1, s, M)$ -derivatives condition on S_{s+1} .

Technically we have only defined the derivatives condition for functions $g : X \rightarrow \mathbb{R}$, but this is easily generalised to functions $g : X \rightarrow \mathbb{R}^{d_i}$ by interpreting it as holding for all d_i individual component functions (where we allow different coordinate functions for each $1 \leq j \leq d_i$). Note that as above the condition at $i=0$ is equivalent to $\partial f_0(c) = 0$ for all $c \in S_1$, which (provided $|S_1| > \frac{1}{2}|H|^2$) is equivalent to $f_0 : X \rightarrow \mathbb{R}^{d_0}$ being a constant function.

Roughly speaking, this definition is saying that all the functions on level i can be reduced to functions on level $i-1$ after taking derivatives of order $i+1$. In applications we will think of functions on level i as resembling polynomials of degree i . This definition may then seem odd, since then one would expect to reduce to level $i-1$ after only a single derivative.

The way to correct one's intuition is to realise that the relationship between level i and the lower levels of a polynomial hierarchy does *not* correspond to that between polynomials of degree i and *all* polynomials of degree $\leq i-1$. Instead, the levels $\leq i-1$ are collections of obstructions that capture how objects on level i *fail* to be polynomials of degree i . Put another way, if a function f on level i is meant to resemble a polynomial of degree i then we expect $\partial^{i+1} f = 0$. The functions in $f_{\leq i-1}$ capture to

what extent this fails. We now give a more easily remembered ‘slogan form’ of the definition, summarising this discussion.

DEFINITION 5.4 (Polynomial hierarchy, slogan form). — *A system of functions $\mathbf{f} = (f_i : X \rightarrow \mathbb{R}^{d_i})_{0 \leq i \leq s}$ is a polynomial hierarchy if functions on level i behave like polynomials of degree i up to errors which are a bounded linear combination of functions on levels $\leq i - 1$.*

5.2. Building a polynomial hierarchy

The following is the first major step in the proof of Theorem 4.1. To avoid obscuring the content under technicalities we state it in rough form only. For the full, precise, statement, see MANNERS (2018, Corollary 2.5.4).

PROPOSITION 5.1. — *Suppose that $f : X \rightarrow \mathbb{R}$ is a 1%-approximate polynomial of degree s . There exists a system of cubes (S_0, \dots, S_{s+1}) which are 99% dense (i.e. of density $\delta = 1 - \epsilon$ for some small $\epsilon > 0$) and $g : S_0 \rightarrow \mathbb{R}$ such that*

- g agrees with f 1% of the time and
- g sits at the top of a polynomial hierarchy of height s on S_0, \dots, S_{s+1} .

Recalling the definition of a polynomial hierarchy, this proposition allows to pass from knowing that the $(s + 1)$ -derivative of f vanishes 1% of the time, to knowing that the $(s + 1)$ -derivative of g vanishes 99% of the time, modulo a collection of functions whose s -derivatives vanish 99% of the time, and so on, until we reach constant functions at the bottom of the hierarchy.

We will see how to make use of this hierarchical structure in the next section. For the remainder of this section, we will sketch how Proposition 5.1 is proved. The main ingredient is a recursive application of the following lemma, which says that a 1%-approximate polynomial of degree s agrees 1% of the time with a function that is a 99%-approximate polynomial of degree s , up to obstructions which are a bounded complexity combination of degree $\leq s - 1$ approximate polynomials. To get a flavour of the kind of difficulties involved the reader may find it instructive to calculate what $\partial_{h_1, h_2} f(x)$ is when $f(x) = \{\alpha x\}\{\beta x\}$.

LEMMA 5.2. — *Let $f : X \rightarrow \mathbb{R}$ be a 1%-approximate polynomial of degree s . There is a function f' defined on 99% of H such that*

1. $f(x) = f'(x)$ for 1% of $x \in X$;
2. there exists a collection of 1%-approximate polynomials g_1, \dots, g_K of degree $s - 1$ (for some reasonably bounded K); and
3. there exists a function $b : S \times \{0, 1\}^{s+1} \rightarrow \mathbb{Z}^K$, with $\|b(c, \omega)\|_1$ reasonably bounded for each $c \in S$ and $\omega \in \{0, 1\}^{s+1}$, such that for 99% of $s + 1$ -dimensional cubes $c = (x; \mathbf{h}) \in H^{s+2}$

$$\partial f'(c) = \sum_{\omega \in \{0, 1\}^{s+1}} (-1)^{|\omega|} b(c, \omega) \cdot \mathbf{g}(x + \omega \cdot \mathbf{h}).$$

Sketch proof. — Our starting observation is that if f is a 1%-approximate polynomial of degree s then, for at least 1% of $h \in H$, the derivative $\partial_h f$ is a 1%-approximate polynomial of degree $s - 1$. Indeed, this is a simple consequence of the pigeonhole principle, averaging over those cubes on which the derivative of f vanishes.

This suggests that we should take our auxiliary functions g_i to be derivatives of f . If we define $g_{a,b}$ by

$$x \mapsto f(x+a) - f(x+b) = \partial_{a-b} f(x+a)$$

then, for suitable choices of a and b (so that $a - b$ is one of the ‘good’ h from the previous paragraph), this is a 1%-approximate polynomial. This is tricky, however, since this will only be defined when both $x+a$ and $x+b$ are in X . Overcoming this difficulty is a subtle matter, and requires careful tracking of the pairs a, b and some initial combinatorial pruning of the set of cubes on which the derivative of f vanishes. We will not address these complications further here, and refer to Section 2 of MANNERS (2018).

To define f' we must somehow extend X , the domain of f which is 1% of the group, to a domain which is 99% of the group. For this we use a random sumset construction. Namely, if $A \subset H$ is a random subset, then we expect the difference set $X - A$ to expand quickly, since there will be few incidences of the form $x_1 + a_1 = x_2 + a_2$. In particular, we can ensure that $X - A$ is 99% of the entire group by taking A to be a random set of size $O(1)$. We will take the g_i to be the 1%-approximate polynomials of degree $s - 1$ of the form $g_{a,b}$ where $a, b \in A$ (note that there are $O(1)$ such g_i).

We then would like to define $f' : X - A \rightarrow \mathbb{R}$ by extending f in the natural way, so that if $x \in X - A$ then $f'(x) = f(x+a)$, where $a \in A$ is chosen so that $x+a \in X$. The problem is that, in general, there will be many such a (still $O(1)$, but certainly more than 1), and this is not an unambiguous definition of f' . We define f' by choosing arbitrarily an appropriate $a \in A$ for each $y \in X - A$, and note that the difference between possible choices, namely $f(x+a) - f(x+b)$, is precisely the function $g_{a,b}$. In particular, the arbitrary nature of our choice only affects the definition of f' ‘up to lower order terms’.

Finally, after suitable combinatorial pruning, one can show that for 99% of all possible cubes $c = (x, \mathbf{h}) \in H^{s+2}$ there exists some shift $c' = (x+a, \mathbf{h} + \mathbf{t})$ where a, t_1, \dots, t_{s+1} all lie in the random set A , and $\partial f(c')$ vanishes, since ∂f vanishes for at least 1% of all possible cubes (this is a multi-dimensional version of the fact alluded to above that if X has density 1% then, provided A is a random set of large but constant size, we expect $X - A$ to have density 99%). As above, by construction of f' and the auxiliary $g_{a,b}$, one can then show that $\partial f'(c) = \partial f(c')$ up to a ‘lower order error’ which is a bounded linear combination of the $g_{a,b}$ functions. Since $\partial f(c') = 0$, however, $\partial f'(c)$ itself is such a linear combination, and we are done. \square

6. FROM 99% STRUCTURE TO ALGEBRAIC STRUCTURE

We will now discuss how to pass from the 99% kind of structure of a polynomial hierarchy, which is inherently combinatorial, to a more algebraic kind of structure, and in particular agreeing often with a genuine nilpolynomial. Recall that we left the previous section with a function g sitting at the top of a polynomial hierarchy of height s , denoted by \mathbf{f} , which means that we can write the derivatives of functions on level t of the hierarchy at cubes of dimension t as

$$\partial f_t(c) = \sum_{\omega \in \{0,1\}^t} b(c, \omega) \cdot \mathbf{f}_{\leq t-1}(x + \omega \cdot \mathbf{h}).$$

The derivatives of the functions on level $t = 0$ must vanish, and hence these functions are constant. We would like to unwind the hierarchy, and propagate our knowledge of functions on level t to functions on level $t + 1$, until we arrive at a description of the function g itself. The primary obstacle in carrying this out are the mysterious coefficients $b(c, \omega)$ which, aside from being integer vectors of bounded size, we do not know much about. In particular, we have no information about how they vary with c and ω .

To get an idea what kind of behaviour we expect the coefficients $b(c, \omega)$ to exhibit, it is helpful to recall the earlier example of a bracket linear (or bracket quadratic) polynomial. For such functions the coordinate functions are given by linear combinations of ‘carry bit’ functions. For example, if $f(x) = \{ax/N\}$ then

$$\partial f(x, h_1, h_2) = b_{x, h_2} - b_{x+h_1, h_2}$$

where $b_{x,y} = 1_{\{ax/N\} + \{ay/N\} \geq 1}$.

The strategy of Manners is to show that, in general, the functions $b(c, \omega)$ behave structurally like the carry bit functions of bracket polynomial examples. Having done so, using the fact that the functions on level $t = 0$ are constant, one can propagate upwards through the levels, showing that the functions on each level resemble a bracket polynomial, eventually arriving at g on the top level.

We will focus, therefore, on the structure of the coefficient functions in the polynomial hierarchy produced in the previous section, and showing that they behave similarly to the carry bit coefficients of the bracket polynomial examples. The key property of carry bits that we require Manners calls a ‘generalised cocycle’ condition.

There are, overall, three main steps to complete the proof of Theorem 4.1: one needs to show that

- if g sits at the top of a polynomial hierarchy of height s then (after some quantitative loss) not only does g satisfy a $(\mathbf{f}, s + 1, s, M)$ -derivatives condition, but it also satisfies a stronger derivatives condition where the coefficients b themselves satisfy a ‘generalised cocycle’ condition;

- all functions satisfying a generalised cocycle condition arise in a similar fashion to carry bits (i.e. as the ‘integral residue’ of derivatives of functions whose derivatives vanish modulo \mathbb{Z});
- the polynomials whose carry bits produce the coefficients b can be reconstructed from b , and the fact that g sits at the top of a polynomial hierarchy with coefficients b means we can propagate this polynomial construction through the hierarchy to construct a genuine nilpolynomial which agrees with g 1% of the time.

While none of this is easy or straightforward, the first part is the most difficult — indeed, it is the most difficult part of the entire proof of MANNERS (2018), and is where the most significant quantitative loss occurs. The proofs of the second two parts are more algebraic, especially the third, which is where explicit nilpolynomials have to be constructed. These proofs are long and technical, with many new difficulties arising that Manners has to overcome, but they also have much in common with the kind of explicit nilmanifold constructions that have appeared before in the literature.

To keep the length of our exposition bounded, therefore, we will not discuss the proofs of the latter two points any further. This affords us space to focus on the new *combinatorial* ideas introduced by Manners, and in particular the definition of a ‘generalised cocycle’, which is a generalisation of the cocycle condition (5) heavily used in previous approaches to the inverse theorem, and which we expect to play an important role in the future of the field.

6.1. Cocycles

We begin by recalling the cocycle property of the derivative (5). We can rewrite this using the cube notation. For example, in dimension 2, if $c = (x, h_1, h_2)$ is a cube and $u \in H$ then

$$\begin{aligned} \partial f(c) &= f(x) - f(x + h_1) - f(x + h_2) + f(x + h_1 + h_2) \\ &= (f(x) - f(x + h_1) - f(x + h_2 + u) + f(x + h_1 + h_2 + u)) \\ &\quad - (f(x + h_2) - f(x + h_1 + h_2) - f(x + h_2 + u) + f(x + h_1 + h_2 + u)) \\ &= \partial f(c^u) - \partial f(c_u) \end{aligned}$$

where $c^u = (x, h_1, h_2 + u)$ and $c_u = (x + h_2, h_1, u)$. The relationship between the cubes c, c^u, c_u can be viewed geometrically as seen in Figure 6.1.

Generalising to arbitrary dimensions, if $c = (x, h_1, \dots, h_k)$ and $1 \leq i \leq k$ then we define the two related cubes

$$\begin{aligned} c^u &= (x; h_1, h_2, \dots, h_{i-1}, h_i + u, h_{i+1}, \dots, h_k) \\ c_u &= (x + h_i; h_1, h_2, \dots, h_{i-1}, u, h_{i+1}, \dots, h_k). \end{aligned}$$

That is, both are k -dimensional cubes, where the only changes have been to the base-point x and the component in direction i . Both are of the form $(x', h_1, \dots, h'_i, \dots, h_k)$ such that $(x + h_i) - (x' + h'_i) = u$. Note that this definition is dependent on the choice

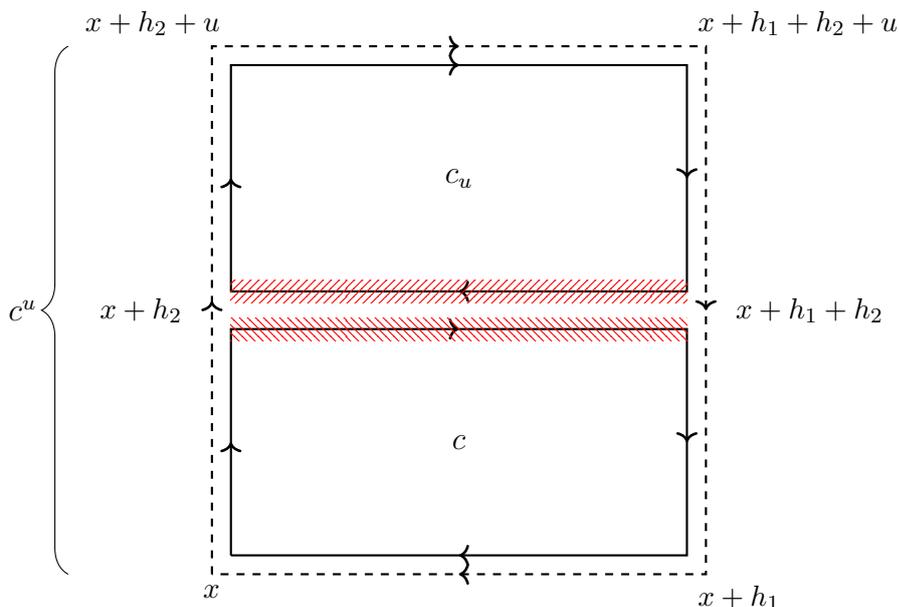


FIGURE 1. The relationship between the cubes c , c^u , and c_u . With the natural orientations we can ‘glue’ c and c_u together to form c^u , so that (informally) $c = c^u - c_u$. The cocycle property is that ∂ respects this relationship and $\partial f(c) = \partial f(c^u) - \partial f(c_u)$ for any function f .

of direction i , which we suppress for notational ease, and which will always be clear from context.

As in the 2-dimensional case it is a trivial matter of expanding out the definitions to check that, for any function f , cube $c \in H^{k+1}$, and $u \in H$, the cocycle identity $\partial f(c) = \partial f(c^u) - \partial f(c_u)$ holds. We call functions on the space of cubes which satisfy the cocycle identity for every cube $c \in H^{k+1}$ and $u \in H$ cocycles (so that, for any f , the derivative ∂f is always a cocycle). We will need a robust version of this concept, where the cocycle identity may only hold for 99% of all possible cubes, giving the following definition.

DEFINITION 6.1 (Cocycles). — *Let $\delta > 0$ and A be an abelian group. If $S \subset H^{k+1}$ is a set of cubes then a function $\rho : S \rightarrow A$ is a k -cocycle on S with loss δ if, for every direction $1 \leq i \leq k$, the cocycle identity*

$$\rho(c) = \rho(c^u) - \rho(c_u)$$

is true for all but at most $\delta|H|^{k+2}$ pairs $(c; u) \in H^{k+2}$ such that $c, c^u, c_u \in S$.

In particular, the previous discussion shows that any derivative ∂f (of order k) is a k -cocycle. Remarkably, the converse is also essentially true — one can view this ‘algebraic’ inverse result as the seed from which our titular ‘combinatorial’ inverse results grow.

LEMMA 6.1. — *If S_0, \dots, S_k is a system of cubes which are 99% dense and $\rho : S_k \rightarrow \mathbb{Z}^d$ is a k -cocycle with 1% loss on S_k then there exists $\lambda : H \rightarrow \mathbb{R}^d$ with $\partial\lambda : H^{k+1} \rightarrow \mathbb{Z}^d$ such that*

1. $\partial\lambda$ agrees with ρ for 99% of cubes in S_k and
2. $\sup_{x \in H} \|\lambda(x)\|_1 \ll_k \sup_{c \in S_k} \|\rho(c)\|_1 + d$.

It is important to note that, even where $\partial\lambda$ does not agree with ρ , the lemma guarantees that it is still an integer-valued function, and hence λ behaves like a polynomial of degree k ‘modulo \mathbb{Z} ’.

Sketch proof. — The key observation is that if we let

$$\lambda(x) = \mathbb{E}_{\mathbf{h} \in H^k} \rho(x; \mathbf{h}),$$

then a lengthy but routine calculation using the definitions proves that $\partial\lambda = \rho$, at least when ρ is a 100% cocycle. Some highly non-trivial combinatorial smoothing allows one to show that this identity still holds 99% of the time when ρ is a 99% cocycle. \square

6.2. Generalised cocycles

To motivate the introduction of generalised cocycles, we first sketch how the algebraic inverse result of the previous section may be applied to obtain a combinatorial inverse result when $s = 1$.

Recall that our starting point is some function f with 99% structure, quantified by being at the top of a (D, M) -polynomial hierarchy of height $s = 1$, say with coefficient function $b : H^3 \times \{0, 1\}^2 \rightarrow \mathbb{Z}^D$. Our goal is to show that the coefficient function $b(c, \omega)$ behaves like the carry bits from a bracket linear function, which is enough for an inverse result as mentioned in the introduction to this section.

Unpacking the definitions (and assuming all 99% objects are in fact 100% objects for simplicity), we have a constant function $f_0 : H \rightarrow \mathbb{R}^D$ such that for all $c \in H^3$,

$$\begin{aligned} \partial f(c) &= \sum_{\omega_1, \omega_2 \in \{0, 1\}} (-1)^{\omega_0 + \omega_1} f(x + \omega_1 h_1 + \omega_2 h_2) \\ &= (b(c, 00) - b(c, 01) - b(c, 10) + b(c, 11)) \cdot f_0. \end{aligned}$$

Since f_0 is constant, we can consolidate the first factor into a single function $B : H^3 \rightarrow \mathbb{Z}^D$, expressing the right-hand side as $B(c) \cdot f_0$. In particular, when $s = 1$, we can assume that the coordinate function does not depend on ω (note that the same consolidation takes place for the derivative of a bracket linear function). From the definition alone, the only knowledge we have about the function B is that it takes on values which are L^1 -bounded (by the parameter M).

Since $B(c) \cdot f_0$ is a derivative, the cocycle identity must hold. If we can ‘invert’ the $\cdot f_0$ then B must itself satisfy the cocycle identity. By the algebraic inverse result Lemma 6.1, B must itself, essentially, be a derivative, and so there is some $f' : H \rightarrow \mathbb{R}^D$ such that $\partial f' : H^3 \rightarrow \mathbb{Z}^D$ (so that f' is like a linear polynomial modulo \mathbb{Z}) and $\partial f = (\partial f') \cdot f_0$. This is still rather weak algebraic structure, but it is enough to be

bootstrapped into constructing a genuine nilpolynomial which agrees with f 1% of the time. At this point the arguments become more algebraic than combinatorial, and as mentioned above, will not be discussed any further in this article.

Instead, we will take that final algebraic step for granted, and view obtaining information like ‘the coefficients b of the polynomial hierarchy are themselves derivatives of some other function, which has some polynomial-like structure’ as our goal. We will now indicate how the previous discussion generalises to higher $s \geq 2$. The key input when $s = 1$ was that a function closely related to the coefficients b satisfied the cocycle condition, since it is a derivative, and so we could apply the algebraic inverse result Lemma 6.1.

For the more general argument, we must find some property of the coefficients b that follows from the fact that they are related to a derivative, which satisfies the cocycle condition. The problem is that when $s > 1$ the coefficients now depend on both c and ω , and there is no cheap way to remove this dependence. We therefore need a generalisation of the cocycle condition which somehow takes both the cube c and the ‘face’ ω into consideration. This property Manners calls a ‘generalised cocycle condition’, and can be derived heuristically as follows.

Suppose that f sits on top of a polynomial hierarchy of height s , so that for any $c = (x, \mathbf{h}) \in S_{s+1}$,

$$\partial f(c) = \sum_{\omega \in \{0,1\}^{s+1}} (-1)^{|\omega|} b(c, \omega) \cdot \mathbf{f}_{\leq s-1}(x + \omega \cdot \mathbf{h}).$$

We now see what this identity combined with the cocycle property of derivatives gives us. Let $(c, u) \in H^{s+3}$ be such that $c^u, c_u \in S_{s+1}$ (we fix an arbitrary direction $1 \leq i \leq s+1$). The cocycle identity implies that

$$\partial f(c) = \sum_{\omega \in \{0,1\}^{s+1}} (-1)^{|\omega|} (b(c^u, \omega) \cdot \mathbf{f}_{\leq s-1}(x + \omega \cdot \mathbf{h}^u) - b(c_u, \omega) \cdot \mathbf{f}_{\leq s-1}(x + h_i + \omega \cdot \mathbf{h}_u)).$$

If we let

$$(8) \quad \mathbf{b}(c, u, \omega) = \begin{cases} b(c^u, \omega) & \text{if } \omega_i = 0 \text{ and} \\ b(c_u, \omega \setminus \{i\}) & \text{if } \omega_i = 1 \end{cases}$$

then we can regroup this as

$$\begin{aligned} \partial f(c) &= \sum_{\omega \in \{0,1\}^{s+1}} (-1)^{|\omega|} \mathbf{b}(c, u, \omega) \cdot \mathbf{f}_{\leq s-1}(x + \omega \cdot \mathbf{h}) \\ &\quad + \sum_{\substack{\omega \in \{0,1\}^{s+1} \\ i \in \omega}} (-1)^{|\omega|} (b(c^u, \omega) - b(c_u, \omega)) \cdot \mathbf{f}_{\leq s-1}(x + \omega \cdot \mathbf{h}). \end{aligned}$$

If $b(c^u, \omega) = b(c_u, \omega)$, then the second sum vanishes and

$$\sum_{\omega \in \{0,1\}^{s+1}} (-1)^{|\omega|} b(c, \omega) \cdot \mathbf{f}_{\leq s-1}(x + \omega \cdot \mathbf{h}) = \sum_{\omega \in \{0,1\}^{s+1}} (-1)^{|\omega|} \mathbf{b}(c, u, \omega) \cdot \mathbf{f}_{\leq s-1}(x + \omega \cdot \mathbf{h}).$$

Comparing coefficients we may then hope that $b(c, \omega) = \mathbf{b}(c, u, \omega)$. This, roughly, is the idea of the ‘generalised cocycle’ condition.

There are two technical matters to clarify before we give the precise definition of a generalised cocycle. The first is that in the final step we illegitimately concluded from ‘comparing coefficients’ that $b(c, \omega) = \mathbf{b}(c, u, \omega)$ for fixed c, u, ω . The second is that we assumed that $b(c^u, \omega) = b(c_u, \omega)$. Neither of these are, in general, true, but can be recovered after some work if for each fixed cube c the functions $b(c, \cdot)$ enjoy a special kind of structure, namely being in ‘normal form’.

DEFINITION 6.2 (Normal form). — *Let $t \leq k$. A function $b : \{0, 1\}^k \rightarrow \mathbb{Z}^{d_0 + \dots + d_t}$ is in normal form if whenever $|\omega| = \omega_1 + \dots + \omega_k = j$ the first $d_0 + \dots + d_{j-1}$ coordinates of $b(\omega)$ are all zero.*

Although we can certainly not assume that the coefficient function b is in normal form itself, it can be transformed into one, incurring only a small cost. It is an easy exercise to check that, for any $b : \{0, 1\}^k \rightarrow \mathbb{Z}^d$, the function

$$(9) \quad b'(\omega') = \sum_{\omega \in \{0, 1\}^k} Z_r(\omega, \omega') b(\omega)$$

vanishes whenever $|\omega'| > r$, where⁽¹⁰⁾

$$Z_r(\omega, \omega') = \sum_{\substack{\eta \in \{0, 1\}^k \\ \omega \supset \eta \supset \omega' \\ |\eta| \leq r}} (-1)^{|\omega| - |\eta|}.$$

By applying this transformation to each level of the coefficient function in a polynomial hierarchy, and keeping track of the constants using the fact that $|Z_r| = O_s(1)$, we obtain the following lemma, which lets us assume that the coefficient function is in normal form.

LEMMA 6.2. — *If $g : X \rightarrow \mathbb{R}$ sits at the top of a polynomial hierarchy of height s with an associated coefficient function $b : S_{s+1} \times \{0, 1\}^{s+1} \rightarrow \mathbb{Z}^{d_0 + \dots + d_{s-1}}$ then we can assume that the function $b(c, \cdot)$ is in normal form for every $c \in S$, at the cost of replacing the M parameter by some $O_s(M)$.*

We can therefore assume that b is in normal form. We would still need to know that \mathbf{b} is in normal form also, but we can get around this by applying the transform (9) to \mathbf{b} , putting it into something like normal form.

Before we define a generalised cocycle we need to quantify the condition that $b(c^u, \omega) = b(c_u, \omega)$ for ‘most’ c and u that was used in our heuristic discussion.

DEFINITION 6.3. — *If $S \subset H^{k+1}$ is a set of k -dimensional cubes then $b : S \times \{0, 1\}^k \rightarrow \mathbb{Z}^d$ is δ -almost upper compatible on S if, for every direction $1 \leq i \leq k$ and $\omega \in \{0, 1\}^k$ such that $\omega_i = 1$, the identity*

$$b(c^u, \omega) = b(c_u, \omega)$$

holds for all but at most $\delta |H|^{k+2}$ pairs $(c, u) \in H^{k+2}$ such that $c^u, c_u \in S$.

⁽¹⁰⁾We identify vectors in $\{0, 1\}^k$ with subsets of $\{1, \dots, k\}$ when we say $\omega \supset \eta \supset \omega'$.

As mentioned above, generically we expect the coefficients b produced in the proof to be upper compatible, so we will not dwell on this issue here. The case when they are not upper compatible cannot be ruled out, and adds yet more technicalities, but it is somewhat pathological, and can be dealt with by other means.

We now come to the crucial definition, that of a generalised cocycle. This definition is informed by our heuristic discussion above, the idea being that it captures what the cocycle identity implies for the coefficient functions of a polynomial hierarchy.

DEFINITION 6.4 (Generalised cocycle). — *Let $b : S \times \{0, 1\}^k \rightarrow \mathbb{Z}^d$ be a function in normal form. If $0 \leq r < k$ and $S \subset H^{k+1}$ is a set of cubes then we say that b is a generalised k -cocycle of type r and loss $\delta > 0$ on S if*

1. b is δ -almost upper compatible on S and
2. in every direction $1 \leq i \leq k$, for all but at most $\delta |H|^{k+2}$ many $(c, u) \in H^{k+2}$ such that $c^u, c_u \in S$, for all $\omega \in \{0, 1\}^k$,

$$b(c, \omega) = \tilde{\mathbf{b}}_r(c, u, \omega')$$

where \mathbf{b} is given by (8) and the transform $\tilde{\mathbf{b}}_r$ was defined in (9).

There is a strong structural result available for generalised cocycles, just as there is for cocycles. Indeed, one can check that if λ is a ‘polynomial modulo \mathbb{Z} ’ of degree $k - r - 1$ (in that $\partial\lambda(c)$ vanishes modulo \mathbb{Z} for all cubes of dimension $k - r$) then, if $c = (x, \mathbf{h}) \in H^{k+1}$, the function $b(c, \omega) = \lambda(x + \omega \cdot \mathbf{h})$ is a generalised k -cocycle of type r (at least, after passing to a normal form of b as in (9)). The following lemma says that the converse holds.

LEMMA 6.3. — *Let S_0, \dots, S_k be a system of cubes with 99% density and $b : S_k \times \{0, 1\}^k \rightarrow \mathbb{Z}^d$ be a generalised k -cocycle of type r and 1% loss.*

There is a function $\lambda : H \rightarrow \mathbb{R}^d$ such that, if $\Lambda(c, \omega)$ is the transformation (as defined in (9)) of the function $(c, \omega) \mapsto \lambda(x + \omega \cdot \mathbf{h})$, then

1. $b(c, \omega) = \Lambda(c, \omega)$ for all $\omega \in \{0, 1\}^k$ for 99% of cubes $c \in S_k$,
2. $\partial\lambda(c) \in \mathbb{Z}^d$ for all $c \in H^{k-r+1}$, and
3. $\sup_{x \in H} \|\lambda(x)\|_1 \ll \sup_{c \in S_k} \|b(c)\|_1 + d$.

In particular, if the coefficient function $b(c, \omega)$ is a generalised cocycle then it agrees for 99% of values with a function that, in a certain sense, is a polynomial ‘modulo \mathbb{Z} ’. Thus in the expression

$$\partial f_t(c) = \sum_{\omega \in \{0, 1\}^t} b(c, \omega) \cdot \mathbf{f}_{\leq t-1}(x + \omega \cdot \mathbf{h})$$

all terms that appear on the right-hand side are essentially polynomial objects (the $\mathbf{f}_{\leq t-1}$ components by induction) and thus ∂f vanishes ‘up to lower degree polynomials’, which is enough to show that f itself is a nilpolynomial, at least 1% of the time. How do we convert the heuristic discussion above into knowledge that the coefficient function $b(c, \omega)$ genuinely is a generalised cocycle, however?

6.3. Concluding the proof

The following is a convenient package definition, which couples the derivatives condition of the previous section with the extra restriction that the coefficients are generalised cocycles.

DEFINITION 6.5 (Strong derivatives condition). — *Let $\mathbf{f} = (f_i : X \rightarrow \mathbb{R}^{d_i})_{0 \leq i \leq s}$ be a tuple of functions, let $0 \leq k, t \leq s+1$, and $b : S \times \{0, 1\}^k \rightarrow \mathbb{Z}^{d_0 + \dots + d_{t-1}}$ where $S \subset H^{k+1}$ is a set of cubes.*

A function $g : X \rightarrow \mathbb{R}$ satisfies the $(\mathbf{f}, k, t, M, \delta)$ -strong derivatives condition on S with coefficients b if

1. *g satisfies the (\mathbf{f}, k, t, M) -derivatives condition on S with coefficients b ;*
2. *b is in normal form;*
3. *if we write $b = (b_0, \dots, b_{t-1})$ where $b_i : S \times \{0, 1\}^k \rightarrow \mathbb{Z}^{d_i}$ then b_i is δ -almost upper compatible on S for $1 \leq i \leq t-1$; and*
4. *b_{t-1} is a generalised k -cocycle of type $t-1$ and loss δ on S .*

From our heuristic discussion (that is, combining the cocycle property of derivatives with the definition of the derivatives condition) we expect that the coefficients within a polynomial hierarchy should have coefficients that are generalised cocycles, and in particular it should satisfy the strong derivatives condition with respect to this hierarchy.

The following lemma is a precise statement of this fact. There are many aspects we have shrugged off or simply ignored in our heuristic discussion, such as the fact that b may not be even remotely upper compatible, or the fact that all our statements hold only 99% of the time rather than 100% of the time. Perhaps the most significant, however, is the hierarchical nature of our structure. Unless one proceeds extremely carefully, the 99% amount of structure enjoyed at the top level may only translate to a 1% amount of structure on the next level down, at which point all of our heuristics cease to be even approximately true.

To avoid this, one must take care to expand this 1% to 99% before we try to locate generalised cocycles. This ‘expansion’ is probably the most lengthy and difficult part of Manners’ proof. Indeed, it is also quantitatively the most costly, and is the reason there are different bounds between the cases $s \leq 3$ and $s \geq 4$ in Theorem 4.1 and thence Theorem 1.1.

LEMMA 6.4. — *Let $\epsilon > 0$ and let (S_0, \dots, S_{s+1}) be a system of cubes inside H of density $1 - \epsilon$. Suppose $g : S_0 \rightarrow \mathbb{R}$ sits at the top of a (D, M) -polynomial hierarchy of height s on S_0, \dots, S_{s+1} .*

There exists $M' \geq 1$ and a system of cubes (S'_0, \dots, S'_{s+1}) of density $1 - 2\epsilon$ such that $S'_k \subset S_k$ and $|S_k \setminus S'_k| \leq \epsilon |S_k|$ for each k and either

1. *H has a proper subgroup of size at least $|H|/M'$, or*

2. g sits at the top of a (D, M') -polynomial hierarchy \mathbf{f}' of height s on S'_0, \dots, S'_{s+1} which is an M' -bounded linear combination of \mathbf{f} , and further g satisfies the $(\mathbf{f}, s + 1, s, M', \epsilon)$ -strong derivatives condition on S'_{s+1} ,

where

$$M' \leq \begin{cases} O(M/\epsilon)^{O(D)} & \text{if } s \leq 2 \text{ and} \\ O(M/\delta)^{O(D)^{O(D)}} & \text{if } s \geq 3. \end{cases}$$

Note that if H has no non-trivial subgroups (as is the case for Theorem 1.1, when $H = \mathbb{Z}_N$ for N prime), then the first case cannot hold (for bounded M'), and hence we must be in the second case. If H does have large subgroups, then we cannot hope to iteratively apply Lemma 6.4, and hence the strategy of Manners breaks down. This is a significant restriction on the work of Manners, and in particular means that this approach cannot be used for inverse theorems over groups with many subgroups, such as \mathbb{F}_p^n (for which see the recent alternative approach of GOWERS and MILIĆEVIĆ (2017, 2020)).

We will attempt only a very brief sketch of the ideas involved. The reader is cautioned that many highly non-trivial difficulties lurk in the details. We may heuristically interpret the hypothesis as saying that ∂g vanishes (up to lower order terms) on 99% of all $(s + 1)$ -dimensional cubes. We then would like to argue as in our heuristic discussion above to deduce that the coefficients satisfy a generalised cocycle condition, and hence g satisfies a strong derivative condition as required.

Roughly speaking, this can be done *unless* ∂g vanishes on 1% of s -dimensional cubes, at which point too much of the potential information gained about the coefficients is annihilated. The idea is to then show that we can upgrade this 1% to a 99%, so that in fact ∂g vanishes on 99% of all s -dimensional cubes, and we can repeat the argument from the beginning, but with s reduced by 1, and so on. We must win somewhere, since the functions at the bottom level of the hierarchy are simply constants, and therefore at some level we can exit with a strong derivatives condition as required.

We conclude by sketching how we grow this 1% to 99%. The goal is to show that, assuming $\partial g = 0$ on 99% of $(s + 1)$ -dimensional cubes and $\partial g = 0$ on 1% of all s -dimensional cubes, either

1. H has a large subgroup, or
2. ∂g vanishes on 99% of s -dimensional cubes.

We assume that the first condition fails, and will attempt to transform the 1% of cubes on which ∂g vanishes to 99% of cubes. There are two operations on cubes that are used to do so: gluing (where from c^u and c_u we produce c) and translation (where from (x, \mathbf{h}) we produce $(x + u, \mathbf{h})$). These are based on the trivial observations that:

1. the cocycle identity $\partial g(c) = \partial g(c^u) - \partial g(c_u)$ implies that if ∂g vanishes on both c^u and c_u then it must also vanish on c and

2. the fact that by definition

$$\partial g(x, h_1, \dots, h_k, u) = \partial g(x, h_1, \dots, h_k) - \partial g(x + u, h_1, \dots, h_k)$$

implies that if ∂g vanishes on both (x, \mathbf{h}) and (x, \mathbf{h}, u) then it also vanishes on $(x + u, \mathbf{h})$.

Since we are working inside a system of cubes with good extension properties, the second in particular implies that if ∂g vanishes on (x, \mathbf{h}) then there are many u such that it vanishes for $(x + u, \mathbf{h})$. We call this the translation of (x, \mathbf{h}) by u .

These two observations, coupled with reasonable compatibility conditions on the system of cubes, combine to imply that the set of s -dimensional cubes for which ∂g vanishes is closed under both gluing and translation. Methods from additive combinatorics can then be used to show that this closure property of the set of cubes in H^{s+1} forces them to actually be 99% of H_0^{s+1} where $H_0 \leq H$ is some subgroup. Since we are assuming that the original set of cubes is at least 1% of all cubes, and also that there are no non-trivial large subgroups, it follows that $H_0 = H$, and hence in fact ∂g must vanish on 99% of all cubes.

We remark that this final part of the argument is, in some sense, quantitatively the most inefficient, and is the reason for the double exponential in Theorem 1.1 when $s \geq 4$. Manners suspects that a more refined argument could remove this loss, which would mean that Theorem 1.1 holds for all $s \geq 1$ with only singly exponential bounds.

7. SUMMARY OF THE PROOF

We conclude this article by giving a summary of the entire proof of Manners to reorient the reader. We are given some function $f : \mathbb{Z}_N \rightarrow \mathbb{C}$ such that $\|f\|_{U^{s+1}} \geq \delta$.

1. There is a set $X \subset \mathbb{Z}_N^{s-1}$, which is 1% of all tuples, such that if $\mathbf{h} \in X$ then $\Delta_{\mathbf{h}}f$ has large U^2 norm;
2. The U^2 inverse theorem (a simple consequence of orthogonality) implies that there is some function $\Psi : X \rightarrow \mathbb{R}$ such that each $\Delta_{\mathbf{h}}f$ correlates with $e(\Psi(\mathbf{h})x)$;
3. The function Ψ is an approximate polynomial of degree $s - 1$, in that $\partial\Psi$ vanishes for 1% of cubes $c \in (\mathbb{Z}_N^{s-1})^s$;
4. There is another function Ψ' , which agrees with Ψ 1% of the time, such that 99% of the time the s -derivatives of Ψ' vanish, up to lower order terms (i.e. Ψ' sits at the top of a polynomial hierarchy);
5. The lower order terms in the polynomial hierarchy are linear combinations of functions with coefficients that satisfy the generalised cocycle property;
6. Functions which satisfy the generalised cocycle property arise from the derivatives of functions which are polynomials modulo \mathbb{Z} ;
7. Functions which are at the top of a polynomial hierarchy, whose coefficients satisfy a generalised cocycle property, must agree some of the time with a nilpolynomial (that can be explicitly constructed from the coefficients);

8. Thus there is a nilpolynomial $P : \mathbb{Z}_N^{s-1} \rightarrow \mathbb{R}$, of degree $s - 1$, such that, for 1% of $\mathbf{h} \in \mathbb{Z}_N^{s-1}$, the multiplicative derivative $\Delta_{\mathbf{h}}f$ correlates with $e(P(\mathbf{h})x)$. From this via an ‘integration and projection’ argument, an explicit nilsequence Φ of degree s can be constructed, such that f correlates with Φ .

REFERENCES

- Vitaly BERGELSON, Bernard HOST, and Bryna KRA (2005). “Multiple recurrence and nilsequences”, *Invent. Math.* **160** (2). With an appendix by Imre Ruzsa, pp. 261–303.
- Vitaly BERGELSON and Alexander LEIBMAN (2007). “Distribution of values of bounded generalized polynomials”, *Acta Math.* **198** (2), pp. 155–230.
- Omar A. CAMARENA and Bálaazs SZEGEDY (2010). *Nilspaces, nilmanifolds and their morphisms*. <https://arxiv.org/abs/1009.3825>.
- Pablo CANDELA and Bálaazs SZEGEDY (2019). *Regularity and inverse theorems for uniformity norms on compact abelian groups and nilmanifolds*. <https://arxiv.org/abs/1902.01098>.
- Timothy GOWERS (2001). “A new proof of Szemerédi’s theorem”, *Geom. Funct. Anal.* **11** (3), pp. 465–588.
- Timothy GOWERS and Luka MILIĆEVIĆ (2017). *A quantitative inverse theorem for the U^4 norm over finite fields*. <https://arxiv.org/abs/1712.00241>.
- (2020). *An inverse theorem for Freiman multi-homomorphisms*. <https://arxiv.org/abs/2002.11667>.
- Ben GREEN and Terence TAO (2008a). “An inverse theorem for the Gowers $U^3(G)$ norm”, *Proc. Edinb. Math. Soc. (2)* **51** (1), pp. 73–153.
- (2008b). “The primes contain arbitrarily long arithmetic progressions”, *Ann. of Math. (2)* **167** (2), pp. 481–547.
- (2010a). “An equivalence between inverse sumset theorems and inverse conjectures for the U^3 norm”, *Math. Proc. Cambridge Philos. Soc.* **149** (1), pp. 1–19.
- (2010b). “Linear equations in primes”, *Ann. of Math. (2)* **171** (3), pp. 1753–1850.
- (2012). “The quantitative behaviour of polynomial orbits on nilmanifolds”, *Ann. of Math. (2)* **175** (2), pp. 465–540.
- Ben GREEN, Terence TAO, and Tamar ZIEGLER (2011). “An inverse theorem for the Gowers U^4 -norm”, *Glasg. Math. J.* **53** (1), pp. 1–50.
- (2012). “An inverse theorem for the Gowers $U^{s+1}[N]$ -norm”, *Ann. of Math. (2)* **176** (2), pp. 1231–1372.
- Yonatan GUTMAN, Freddie MANNERS, and Péter VARJÚ (2019). “The structure theory of nilspaces II: Representation as nilmanifolds”, *Trans. Amer. Math. Soc.* **371** (7), pp. 4951–4992.
- (2020a). “The structure theory of nilspaces I”, *J. Anal. Math.* **140** (1), pp. 299–369.

- (2020b). “The structure theory of nilspaces III: Inverse limit representations and topological dynamics”, *Adv. Math.* **365**, pp. 107059, 53.
- Bernard HOST and Bryna KRA (2009). “Uniformity seminorms on ℓ^∞ and applications”, *J. Anal. Math.* **108**, pp. 219–276.
- Alexander LEIBMAN (2002). “Polynomial mappings of groups”, *Israel J. Math.* **129**, pp. 29–60.
- Shachar LOVETT (2012). “Equivalence of polynomial conjectures in additive combinatorics”, *Combinatorica* **32** (5), pp. 607–618.
- Freddie MANNERS (2018). *Quantitative bounds in the inverse theorem for the Gowers U^{s+1} -norms over cyclic groups*. <https://arxiv.org/abs/1811.00718>.
- Tom SANDERS (2012). “On the Bogolyubov-Ruzsa lemma”, *Anal. PDE* **5** (3), pp. 627–655.
- Bálint SZEGEDY (2012). *On higher order Fourier analysis*. <https://arxiv.org/abs/1203.2260>.
- Terence TAO (2012). *Higher order Fourier analysis*. Vol. 142. Graduate Studies in Mathematics. American Mathematical Society, Providence, RI, pp. x+187.
- Terence TAO and Tamar ZIEGLER (2010). “The inverse conjecture for the Gowers norm over finite fields via the correspondence principle”, *Anal. PDE* **3** (1), pp. 1–20.

Thomas F. Bloom
 Centre for Mathematical Sciences
 Cambridge University
 UK
E-mail : tb634@cam.ac.uk