ACTA ARITHMETICA
179.1 (2017)

A higher-dimensional Siegel–Walfisz theorem

by

PIERRE-YVES BIENVENU (Bristol)

1. Introduction. The prime number theorem in arithmetic progressions says that

$$\sum_{n\le N}\Lambda(qn+b)=1_{(q,b)=1}\frac{q}{\varphi(q)}N+o(N),\tag{1.1}$$

where Λ is the von Mangoldt function and φ the Euler totient function, while b and q are two fixed integers. The Siegel–Walfisz theorem states that the asymptotic (1.1) holds uniformly in b and q in the regime where $q=q(N)$ is a function of N satisfying $q=\log^{O(1)}N$ and b varies as $N\log^{O(1)}N$.

Using the notation

$$\Lambda_q(n)=\frac{q}{\varphi(q)}1_{(n,q)=1}$$

for $n\in\mathbb{Z}$ or $n\in\mathbb{Z}/q\mathbb{Z}$ and multiplicativity, we can rewrite equation (1.1) as

$$\sum_{n\le N}\Lambda(qn+b)=N\Big(\prod_{p\mid q}\Lambda_p(b)+o(1)\Big).\tag{1.2}$$

A theorem of Green and Tao [5] which we now state, relying on two conjectures later resolved by Green, Tao and Ziegler [6], [7], may be seen as a higher-dimensional analogue of (1.2).

THEOREM 1.1. *Let L be a constant and $\Psi=(\psi_1,\dots,\psi_t):\mathbb{Z}^d\to\mathbb{Z}^t$ be a system of affine-linear forms no two of which are affinely related. Suppose that the coefficients of the linear part $\dot\Psi$ are bounded by L, while the constant coefficients $\psi_i(0)$ satisfy $|\psi_i(0)|\le LN$. Let $K\subset[-N,N]^d$ be a convex body.*

2010 *Mathematics Subject Classification*: Primary 11B30.
Key words and phrases: Green–Tao theorem, prime tuples.
Received 1 September 2016.
Published online 28 April 2017.

Then

$$\sum_{n\in\mathbb{Z}^d\cap K}\prod_{i=1}^{t}\Lambda(\psi_i(n))=\beta_\infty\prod_p\beta_p+o_{d,t,L}(N^d),$$

where

$$\beta_\infty=\mathrm{Vol}(K\cap\Psi^{-1}(\mathbb{R}_+^t)),\qquad \beta_p=\mathbb{E}_{a\in(\mathbb{Z}/p\mathbb{Z})^d}\prod_{i=1}^{t}\Lambda_p(\psi_i(a)).$$

Above we have denoted by

$$\mathbb{E}_{a\in A}f(a)=\frac{1}{|A|}\sum_{a\in A}f(a)$$

the averaging operator. We also agree that the letter p is reserved for primes, thus $\prod_p$ implicitly means $\prod_{p\in\mathcal{P}}$. The factors β_p are known as *local factors.*

An interesting extension was obtained by the same authors together with Ford and Konyagin [2]. They showed that Theorem 1.1 was still valid when the constant coefficients satisfied the condition $|\psi_i(0)|\le N\log^C N$. This relaxed condition recently allowed Tao and Ziegler [11, Theorem 1.3] to obtain an improvement of the error term $o(N^d)$ to $o(\mathrm{Vol}(K))$ in the case where $K=[N]\times[M]^{d-1}$ with $M\gg N\log^{-O(1)}N$ and $\psi_i(n)=n_1+P_i(n_2,\dots,n_d)$ for some affine-linear forms $P_1,\dots,P_t$ whose linear coefficients are bounded.

Here we prove a further extension, allowing unbounded linear coefficients. The impetus for the work came from a discussion on Tao's blog [9]. Before stating our theorem, we collect several definitions pertaining to systems of affine-linear forms.

Definition 1.1. A system $\Psi=(\psi_1,\dots,\psi_t):\mathbb{Z}^d\to\mathbb{Z}^t$ with $d,t\ge 2$ of affine-linear forms has *finite complexity* if no form ψ_i is in the affine-linear span of another one. It is called *admissible* if it is of finite complexity and $\beta_p\ne 0$ for all p. In particular, if Ψ is admissible, no prime p divides all coefficients of any form ψ_i. A system $\Psi'=(\psi_{i_1},\dots,\psi_{i_s})$ for some sequence $1\le i_1<\dots<i_s\le t$ is called a *subsystem* of Ψ.

The system Ψ is called *bounded* when all its linear coefficients are bounded (in terms of the asymptotic parameter N), and *unbounded* otherwise. Its *size at scale* (N,B) is defined as

$$\|\Psi\|_{N,B}=\frac{1}{\log^B N}\Bigg(\sum_{i\in[t]}\left|\frac{\psi_i(0)}{N}\right|+\sum_{i\in[t],\,j\in[d]}|\dot{\psi}_i(e_j)|\Bigg),$$

where $(e_1,\dots,e_d)$ is the canonical basis of the lattice $\mathbb{Z}^d$, $[t]=\{1,\dots,t\}$ and $\dot{\psi}_i$ denotes the linear part of ψ_i.

Furthermore, a prime p is called *exceptional* for Ψ (and we write $p\in P_\Psi$) if there exist $i\ne j$ such that ψ_i and ψ_j are affinely related modulo p. In particular, if a form ψ_i is a nonzero constant modulo p, then p is exceptional.

We highlight that our definition of exceptional prime is different (less restrictive) than that of Green and Tao [5, Theorem D.3] and that the size of a system in our sense differs from their definition by the initial factor $\log^{-B} N$.

We now check that $\prod_p \beta_p$ is still convergent in the setting where the coefficients are unbounded. Thus the next lemma plays the role of [5, Lemma 1.3].

LEMMA 1.2. *Let Ψ be a system of affine-linear forms. Then if p is not exceptional,*

$$\beta_p = 1 + O_{d,t}(p^{-2}).$$

In particular, if Ψ is admissible, the product $\prod_p \beta_p$ is convergent and nonzero.

Proof. If two forms ψ_i and ψ_j are not affinely related modulo p, then the probability as a ranges over $(\mathbb{Z}/p\mathbb{Z})^d$ that they vanish simultaneously at a is p^{-2}, by elementary linear algebra (see for instance [1, Proposition C.5]). Inclusion-exclusion then yields $\beta_p = 1 + O(p^{-2})$ for p unexceptional.

Now, if Ψ is admissible, only finitely many primes are exceptional. Indeed, if ψ_i and ψ_j are affinely related modulo p then all the 2×2 minors of the matrix $(\dot{\psi}_k(e_\ell))_{k\in\{i,j\},\ell\in[d]}$ are divisible by p, although at least one of these minors has to be nonzero because they are not affinely related as forms over $\mathbb{Z}$. Moreover, the hypothesis of admissibility implies that $\beta_p \neq 0$ for every prime p, so that the product is convergent and nonzero. ∎

We now state our main theorem.

THEOREM 1.3. *Let d, t be positive integers and A, B, L be positive constants. Assume $\Psi = (\psi_1, \dots, \psi_t) : \mathbb{Z}^d \to \mathbb{Z}^t$ is an admissible system. Suppose that $\|\Psi\|_{N,B} \leq L$ and that $K \subset [-N,N]^d$ is a convex body satisfying $\mathrm{Vol}(K) \gg N^d \log^{-A} N$ and $\Psi(K) \subset \mathbb{R}^t_+$. Then*

$$\sum_{n\in\mathbb{Z}^d\cap K} \prod_{i=1}^{t} \Lambda(\psi_i(n)) = \mathrm{Vol}(K) \prod_p \beta_p (1 + o_{d,t,A,B,L}(1)). \tag{1.3}$$

Unlike Theorem 1.1, this theorem is still meaningful when the convex body $K \subset [-N,N]^d$ satisfies $\mathrm{Vol}(K) = o(N^d)$, and applies even when the linear coefficients are of size polylogarithmic in N.

In some special cases, Theorem 1.3 follows easily from the work of Green and Tao. In fact, Theorem 1.1 is a consequence of an asymptotic for the unbounded system $W\Psi + b$ where Ψ is a bounded system, $W = \prod_{p\leq w} p = \log^{1+o(1)} N$ and $b = (b_1, \dots, b_t) \in [W]^t$ is a t-tuple of integers coprime to W. More generally, an unbounded system $q\Psi + b$ with $q = \log^{O(1)} N$ and Ψ bounded is tractable, via an asymptotic for the system $\widetilde{W}\Psi + c$ where

$\widetilde{W} = Wq$ (¹). By decomposing into residue classes, this extends to systems Ψ such that for each j, the coefficients $\dot{\psi}_i(e_j)$ are bounded multiples of a common coefficient q_j. We show an example, corresponding to the count of k-term progressions of primes whose common difference is a multiple of q. We have

$$\sum_{\substack{1\le n,d\\ n+(k-1)qd\le N}} \prod_{i=0}^{k-1} \Lambda(n+iqd) = \sum_{a\in[q]} \sum_{\substack{1\le n,d\\ n+(k-1)d\le (N-a)/q}} \prod_{i=0}^{k-1} \Lambda(q(n+id)+a)$$

and are thus left with a system of the form $q\Psi + b$ with Ψ bounded.

We now provide less immediate examples where Theorem 1.3 applies.

Example 1.1. What is the proportion of arithmetic progressions $n + d\mathbb{N}$ whose q_1th, …, q_kth terms are all primes? Assume that $q_i = \lfloor \log^i N \rfloor$. The answer is given by

$$\sum_{1\le n,d\le N} \prod_{i=1}^{k} \Lambda(n+q_i d).$$

For this system, the factors β_p can be easily expressed, using the notation $h(p)$ for the number of classes modulo p occupied by $q_1, \ldots, q_k$, as

$$\beta_p = \left(\frac{p}{p-1}\right)^k \frac{(p-1)(1+p-h(p))}{p^2}.$$

Example 1.2. We can also count k-term arithmetic progressions of primes up to N whose common difference is $q = \lfloor \log N \rfloor$ times a prime. This time the sum to consider is

$$\sum_{1\le n\le n+(k-1)qd\le N} \Lambda(d) \prod_{i=0}^{k-1} \Lambda(n+iqd).$$

To simplify the expression of the local factors, assume $\prod_{p\le k} p \mid q$. Then

$$\beta_p = \left(\frac{p}{p-1}\right)^{k+1} \frac{1}{p^2} \begin{cases} (p-1)^2 & \text{if } p \mid q, \\ (p-1)(p-k) & \text{if } p \nmid q. \end{cases}$$

Example 1.3. We provide the asymptotic count of solutions to linear equations in shifted squarefree primes, that is, primes p for which $p-1$ is squarefree. As it is not a direct application, we give the details in the final section.

(¹) This remark was already exploited by the author in a previous paper [1, Section 2.3].

In view of the Siegel–Walfisz theorem (1.2), one may hope to write

$$\sum_{n\in\mathbb{Z}^d\cap K}\prod_{i=1}^{t}\Lambda(\psi_i(n)) = \mathrm{Vol}(K)\Big(\prod_p \beta_p + o_{d,t,A,B,L}(1)\Big)$$

instead of (1.3), but unfortunately our method does not yield this. Such an estimate is genuinely stronger given that $\prod_p \beta_p$ might well tend to infinity with N (and possibly the linear coefficients as well). This weaker bound is ultimately due to the ineffectiveness of the Gowers norm estimate [5, Theorem 7.2].

To prove Theorem 1.3, we first get rid of the convex body by decomposing it into reasonably small boxes, so that the theorem simply needs to be proven on boxes. In this context, the variables all have the same range and are independent of each other, which makes it possible, after the introduction of the W-trick, to prove a suitable von Neumann theorem (2). The latter eliminates all but one form, say ψ_1, thanks to a pseudorandom majorant. The next step is to equalize all coefficients of the chosen form ψ_1, by decomposing the ranges of averaging into congruence classes. We are then left requiring a Gowers norm estimate which was proven by Green and Tao, conditionally on two conjectures later fully resolved by Green, Tao and Ziegler [6, 7].

We assume a certain amount of familiarity with the original arguments of Green and Tao [5]. Wherever only minor changes are needed to accommodate the unbounded coefficients, we will simply describe what must be changed in the existing arguments.

Asymptotic notation. The main asymptotic parameter throughout the paper is N, and all the standard asymptotic notation $o, O, \ll$ refers to the limit as $N \to \infty$. There is an exception: whenever we discuss local factors, the limit is when $p \to \infty$; see for instance Lemma 1.2. Several statements require, sometimes implicitly, that N be large enough, which we always assume. Many other asymptotic parameters are defined in terms of N, such as M, X, Z, Y, so that the limit as $N \to \infty$ is the same as, say, the limit as $M \to \infty$. Indices may be added to symbols such as O, in which case they indicate upon which parameters the implied constant depends.

2. First reductions. As in [5, Theorem 4.1], we show that we can assume that the affine forms not only satisfy $\psi_i(n) \geq 0$ but actually $\psi_i(n) \geq N^{9/10}$. We remark that the set

$$\{n \in K \cap \mathbb{Z}^d : \exists i \in [t]\ \ \psi_i(n) \leq N^{9/10}\}$$

(2) The paper of Green and Tao also proceeds via a destruction of the convex body K in Appendix C, but their method ceases to bear fruit as soon as $\mathrm{Vol}(K) = o((\mathrm{diam}\, K)^d)$.

contains only $O(N^{d-1/10}) = o(\mathrm{Vol}(K)N^{-1/20})$ elements, so that we can replace K by the convex body $K \cap \bigcap_{i=1}^{t} \psi_i^{-1}([N^{9/10}, \infty))$ in Theorem 1.3. Moreover, prime powers are so sparse that we can restrict Λ to primes, replacing Λ by $\Lambda' = 1_{\mathcal{P}} \log$. This leads to the first reduction.

PROPOSITION 2.1. *Let $\Psi = (\psi_1, \ldots, \psi_t) : \mathbb{Z}^d \to \mathbb{Z}^t$ be an admissible system. Suppose that $\|\Psi\|_{N,B} \le L$ and that $K \subset [-N,N]^d$ is a convex body satisfying* $\mathrm{Vol}(K) \gg N^d \log^{-A} N$ *and $\Psi(K) \subset ([N^{9/10}, \infty))^t$. Then*

$$\sum_{n \in \mathbb{Z}^d \cap K} \prod_{i=1}^{t} \Lambda'(\psi_i(n)) = \beta_\infty \prod_p \beta_p (1 + o_{d,t,A,B,L}(1)).$$

We perform one more elementary reduction, namely we reduce to normal form; for the definition of this notion, we refer to [5, Definition 4.2]. The existence of a normal form extension Ψ' for a system Ψ follows from [5, Lemma 4.4]. We inspect its proof to check that $\|\Psi\|_{N,B} \le L$ implies $\|\Psi'\|_{N,B} \le L'$ for some constant L' depending only on L, B, d, t. With respect to the paper [5], the change is that the vector $f_{d+1}, \ldots, f_{d'}$ introduced there will now be of size $O_L(\log^C N)$ for some constant $C = O_{B,d,t}(1)$. This is because these vectors are obtained by Cramer's formula, i.e. by computing determinants of matrices of bounded dimensions whose coefficients are $O_L(\log^B N)$. We claim this procedure reduces Theorem 1.3 to the following proposition, corresponding to [5, Theorem 4.5].

PROPOSITION 2.2. *Let d, t be positive integers and A, B, L be positive constants. Let $\Psi = (\psi_1, \ldots, \psi_t) : \mathbb{Z}^d \to \mathbb{Z}^t$ be an admissible system in normal form. Suppose that $\|\Psi\|_{N,B} \le L$ and that $K \subset [-N,N]^d$ is a convex body satisfying* $\mathrm{Vol}(K) \gg N^d \log^{-A} N$ *and $\Psi(K) \subset ([N^{9/10}, \infty))^t$. Then*

$$\sum_{n \in \mathbb{Z}^d \cap K} \prod_{i=1}^{t} \Lambda(\psi_i(n)) = \beta_\infty \prod_p \beta_p (1 + o_{d,t,A,B,L}(1)).$$

We now check that the arguments following [5, Theorem 4.5] yield the deduction of Theorem 1.3 from Proposition 2.2. Let Ψ be an admissible system and K a convex body satisfying the hypotheses of Theorem 1.3. Let $f_{d+1}, \ldots, f_{d'}$ be the vectors and Ψ' the system in normal form produced by the procedure given above. Suppose all f_i satisfy $f_i = O(\log^C N)$. Letting $M = N \log^{-C} N$, we define an auxiliary convex body K' by

$$K' = \Big\{ (n, m_{d+1}, \ldots, m_{d'}) \in \mathbb{R}^d \times [-M, M]^{d'-d} \;\Big|\; n + \sum_{i=d+1}^{d'} m_i f_i \in K \Big\},$$

which is included in $[-N', N']^{d'}$ where $N' = O(N)$. Moreover, we still have

$$\mathrm{Vol}(K') = \mathrm{Vol}(K)(2M)^{d'-d} \gg N'^d \log^{-D} N'$$

for some constant D. Finally, the local factors are left unchanged by this operation. Thus if the system Ψ is admissible, so is Ψ', so that Proposition 2.2 can be applied to it, which concludes the proof of the reduction.

3. Reduction to the case of a box. We show that the main theorem can be deduced from the following very particular case.

PROPOSITION 3.1. *Let d, t be positive integers and A, B, L be positive constants. Let $\Psi = (\psi_1, \ldots, \psi_t) : \mathbb{Z}^d \to \mathbb{Z}^t$ be an admissible system in normal form. Suppose that $\|\Psi\|_{M,B} \le L$ and $\Psi([M]^d) \subset ([M^{9/10}, \infty))^t$. Then*

$$\sum_{n\in[M]^d} \prod_{i=1}^{t} \Lambda(\psi_i(n)) = M^d \prod_p \beta_p(1 + o_{d,t,A,B,L}(1)).$$

Proof that Proposition 3.1 implies Proposition 2.2. Let $K \subset [-N, N]^d$ be a convex body satisfying $\mathrm{Vol}(K) \gg N^d \log^{-A} N$. Let

$$K' = \{x \in K \mid d(x, \partial K) \ge N \log^{-A-1} N\},$$
$$K'' = \{x \in \mathbb{R}^d \mid d(x, K) \le N \log^{-A-1} N\}.$$

These are two convex bodies. The arguments from elementary convex geometry displayed in [5, Appendix A] allow one to infer that

$$\mathrm{Vol}(K') = \mathrm{Vol}(K) + O(N^d \log^{-A-1} N) = \mathrm{Vol}(K)(1 + o(1))$$

and the same for K''. Now let $M = N \log^{-A-1} N/\sqrt{d}$ and consider the grid $(M\mathbb{Z})^d$. Let $\mathcal{B} = \{c + [M]^d \mid c \in J\}$ be the collection of boxes defined by this grid that are included in K, and $\mathcal{B}' = \{c + [M]^d \mid c \in J'\}$ be the collection of boxes defined by this grid that meet K. Note that

$$K' \subset \bigcup_{B\in\mathcal{B}} B \subset K \subset \bigcup_{B\in\mathcal{B}'} B \subset K''. \tag{3.1}$$

The first inclusion is because if a box B from the grid meets K', then it is included in K.

Now let $\Psi = (\psi_1, \ldots, \psi_t) : \mathbb{Z}^d \to \mathbb{Z}^t$ be a system of affine-linear forms of finite complexity. Suppose that $\|\Psi\|_{M,B} \le L$ and

$$\Psi([M]^d) \subset ([M^{9/10}, \infty))^t.$$

Then

$$\sum_{B\in\mathcal{B}} \sum_{n\in B\cap\mathbb{Z}^d} \prod_{i\in[t]} \Lambda(\psi_i(n)) \le \sum_{n\in K\cap\mathbb{Z}^d} \prod_{i\in[t]} \Lambda(\psi_i(n)) \le \sum_{B\in\mathcal{B}'} \sum_{n\in B\cap\mathbb{Z}^d} \prod_{i\in[t]} \Lambda(\psi_i(n)).$$

Now if $B = c + [M]^d$ with $c \in \mathbb{Z}^d$, letting $\Psi_c = \Psi + \dot{\Psi}(c)$ we can write

$$\sum_{n\in B\cap\mathbb{Z}^d} \prod_{i\in[t]} \Lambda(\psi_i(n)) = \sum_{n\in[M]^d} \prod_{i\in[t]} \Lambda(\psi_{c,i}(n)).$$

We check that the system Ψ_c satisfies $\|\Psi_c\|_{M,C} = O(1)$ for some constant C. Indeed, the linear coefficients are unchanged, so still of size $O(\log^B N) = O(\log^B M)$, and the constant coefficients are of size $O(N \log^B N)$, hence $O(M \log^{A+B+1} M)$. Thus $C = A+B+1$ is good enough. Moreover, we have $\Psi_c([M]^d) \subset ([N^{9/10}, \infty))^t$. Thus we can apply Proposition 3.1. We note that the β_p it produces for a box $B = c+[M]^d$ is in fact independent of c, because the translation invariance of $\mathbb{Z}/p\mathbb{Z}$ allows one to write

$$\begin{aligned}\mathbb{E}_{a\in(\mathbb{Z}/p\mathbb{Z})^d} \prod_{i\in[t]} \Lambda_p(\psi_i(a) + \dot{\psi}_i(c)) &= \mathbb{E}_{a\in(\mathbb{Z}/p\mathbb{Z})^d} \prod_{i\in[t]} \Lambda_p(\psi_i(a+c)) \\ &= \mathbb{E}_{a\in(\mathbb{Z}/p\mathbb{Z})^d} \prod_{i\in[t]} \Lambda_p(\psi_i(a)).\end{aligned}$$

Finally,

$$|\mathcal{B}|M^d \prod_p \beta_p(1+o(1)) \le \sum_{n\in K\cap\mathbb{Z}^d} \prod_{i\in[t]} \Lambda(\psi_i(n)) \le |\mathcal{B}'|M^d \prod_p \beta_p(1+o(1)).$$

Because of the inclusions (3.1), we see that

$$\begin{aligned}\mathrm{Vol}(K)(1+o(1)) = \mathrm{Vol}(K') \le |\mathcal{B}|M^d &\le |\mathcal{B}'|M^d \\ &\le \mathrm{Vol}(K'') = \mathrm{Vol}(K)(1+o(1)).\end{aligned}$$

This completes the proof of Proposition 2.2. ■

4. The W-trick. We perform the W-trick in the same spirit as in [5], so as to eliminate biases modulo small primes. We introduce

$$w = w(N) = \log\log N, \qquad W = \prod_{p\le w} p = \log^{1+o(1)} N.$$

However, a simple one-dimensional example shows that one has to adapt it to our situation where coefficients are unbounded. Indeed, consider the system made of one form in one variable, namely $n \mapsto qn + b$ with q of size roughly $\log N$. The W-trick consists in writing

$$\sum_{n\le N} \Lambda(qn+b) = \sum_{a\in[W],\,(qa+b,W)=1} \frac{W}{\varphi(W)} \sum_{n\le N/W} \frac{\varphi(W)}{W} \Lambda(Wqn + qa + b).$$

But imposing that $(qa+b, W) = 1$ does not ensure that the inner sum is $N/W(1+o(1))$, because $qa+b$ could well have a common factor greater than w with q: when the coefficients are bounded, their factors are all less than w for large enough N but this is not the case any more in our setting. Moreover the relevant average is not $W/\varphi(W)$ but $Wq/\varphi(Wq)$, which may be different if q has prime factors larger than w ([3]). This suggests that

([3]) Nevertheless, it is easy to check using Mertens' theorem that if $w = \log\log N$ and $q \le \log^B N$ then $Wq/\varphi(Wq) = (1+o_B(1))W/\varphi(W)$.

the coefficients of the system have to be taken into account to determine a suitable parameter $\widetilde{W}$ instead of W.

We fix an admissible system $\Psi_1 = (\psi_1, \dots, \psi_t) : \mathbb{Z}^{d_1} \to \mathbb{Z}^{t_1}$ in normal form satisfying $\|\Psi_1\|_{M,B} \le L$ for some constants $B, L > 0$. Let

$$(4.1)\quad Q = \prod_{\substack{i\in[t_1],\, j\in[d_1]\\ \dot\psi_i(e_j)\neq 0}} \dot\psi_i(e_j) \times \prod_{\substack{1\le i<k\le t_1\\ 1\le j<\ell\le d_1\\ \dot\psi_i(e_j)\dot\psi_k(e_\ell)-\dot\psi_i(e_\ell)\dot\psi_k(e_j)\neq 0}} \big(\dot\psi_i(e_j)\dot\psi_k(e_\ell) - \dot\psi_i(e_\ell)\dot\psi_k(e_j)\big)$$

be the product of the nonzero minors of size 1 and 2 in the matrix $(\dot\psi_i(e_j))_{i,j}$; thus $Q = O_L(\log^{O_{d,t,B}(1)} N)$. Moreover, if a prime p is exceptional for Ψ_1, it must divide Q (see the proof of Lemma 1.2).

We now introduce $\widetilde{W} = WQ = O(\log^{O(1)} N)$. The hypothesis that $\psi_i(n) > N^{9/10}$ means that if $\psi_i(n)$ is to be a prime number, it has to be coprime to $\widetilde{W}$. Writing

$$\Lambda'_{\widetilde{W},b}(n) = \frac{\varphi(\widetilde{W})}{\widetilde{W}} \Lambda'(\widetilde{W}n + b)$$

we get

$$(4.2)\quad \sum_{n\in[M]^{d_1}} \prod_{i=1}^{t_1} \Lambda'(\psi_i(n))$$
$$= \sum_{\substack{a\in[\widetilde{W}]^{d_1}\\ \forall i\in[t_1]\ (\psi_i(a),\widetilde{W})=1}} \bigg(\frac{\widetilde{W}}{\varphi(\widetilde{W})}\bigg)^{t_1} \sum_{n\in[M/\widetilde{W}]^{d_1}} \prod_{i=1}^{t_1} \Lambda'_{\widetilde{W},b_i(a)}(\tilde\psi_i(n)) + O(\log^{O(1)} N),$$

where $\tilde\psi_i$ differs from ψ_i only in the constant coefficient (by a multiple of $\widetilde{W}$) and $b_i(a) \in [\widetilde{W}]$ is the reduction mod $\widetilde{W}$ of $\psi_i(a)$.

Using the notion of subsystem introduced in Definition 1.1, we reduce Proposition 3.1 to the following one.

Proposition 4.1. *Let $\Psi_0 = (\psi_1^0, \dots, \psi_{t_0}^0) : \mathbb{Z}^{d_0} \to \mathbb{Z}^{t_0}$ be a subsystem of Ψ_1. Suppose that $\Psi_0([M]^{d_0}) \subset ([M^{8/10}, \infty))^{t_0}$ and that $b_i \in [\widetilde{W}]$ is coprime to $\widetilde{W}$ for any $i \in [t_0]$. Then*

$$(4.3)\quad \sum_{n\in[M/\widetilde{W}]^{d_0}} \prod_{i\in[t_0]} \big(\Lambda'_{\widetilde{W},b_i}(\psi_i^0(n)) - 1\big) = o((M/\widetilde{W})^{d_0})$$

We show how the reduction works, adapting the argument following [5, Proposition 5.1].

Applying successively the decomposition (4.2), the trivial identity $x = x - 1 + 1$, and Proposition 4.1 to systems $\tilde\Psi$ where Ψ is a subsystem of Ψ_1, we obtain

$$\sum_{n\in[M]^{d_1}}\prod_{i=1}^{t_1}\Lambda(\psi_i(n)) = \sum_{\substack{a\in[\widetilde{W}]^{d_1}\\ \forall i\in[t_1]\,(\psi_i(a),\widetilde{W})=1}} \left(\frac{\widetilde{W}}{\varphi(\widetilde{W})}\right)^{t_1}\left(\frac{M}{\widetilde{W}}\right)^{d_1}(1+o(1)) + O(\log^{O(1)} N)$$
$$= M^{d_1}(1+o(1))\,\mathbb{E}_{a\in[\widetilde{W}]^{d_1}}\prod_{i\in[t_1]}\Lambda_{\widetilde{W}}(\psi_i(a)) + O(\log^{O(1)} N).$$

By the Chinese remainder theorem, and the fact that $\Lambda_{p^k}(b) = \Lambda_p(b)$, we have

$$\mathbb{E}_{a\in[\widetilde{W}]^{d_1}}\prod_{i\in[t_1]}\Lambda_{\widetilde{W}}(\psi_i(a)) = \prod_{p\mid WQ}\mathbb{E}_{a\in(\mathbb{Z}/p\mathbb{Z})^{d_1}}\prod_{i\in[t_1]}\Lambda_p(\psi_i(a)) = \prod_{p\mid WQ}\beta_p.$$

Moreover, if a prime p does not divide Q, it is not exceptional for Ψ_1. Then Lemma 1.2 implies that $\beta_p = 1+O(p^{-2})$, so that

$$\prod_{p\nmid WQ}\beta_p = \prod_{p>w}(1+O(p^{-2})) = 1+O(w^{-1}).$$

This concludes the reduction.

5. Reduction to a Gowers norm estimate. Write $X = M/\widetilde{W}$ and fix a system Ψ_0 and a tuple $b_1,\dots,b_{t_0}$ satisfying the conditions of Proposition 4.1. If $t_0 = 1$, this proposition follows directly from the one-dimensional Siegel–Walfisz theorem (1.1), so we suppose $t_0 \geq 2$. Let Q_0 be the product of 2×2 minors for the system Ψ_0 as defined by (4.1). In particular, $Q_0 \mid Q$. We have to prove

$$\sum_{n\in[X]^d}\prod_{i\in[t_0]}F_i(\psi_i(n)) = o(X^d)$$

for $F_i = \Lambda'_{\widetilde{W},b_i} - 1$. At this point Green and Tao move to some cyclic group $\mathbb{Z}/N'\mathbb{Z}$ with $N' = O(X)$, but we cannot do this here without any wrap-around, because of the large (unbounded) coefficients.

5.1. A pseudorandom majorant. Recall that we have fixed an admissible system

$$\Psi_0 = (\psi_i^0)_{i\in[t_0]} : \mathbb{Z}^{d_0}\to\mathbb{Z}^{t_0}.$$

We introduce the notion of a *derived system.* This captures the important properties of the systems that arise from repeated applications of the Cauchy–Schwarz inequality.

Definition 5.1. A system $\Psi : \mathbb{Z}^d\to\mathbb{Z}^t$ of affine-linear forms is said to be *derived* from Ψ_0 if the following conditions are all satisfied:

- $d \le 2d_0$;
- $t \le 2^{d_0} t_0$;
- $\|\Psi\|_{N,B} \ll \|\Psi_0\|_{N,B}$;
- any exceptional prime for Ψ divides Q_0.

We now define pseudorandomness, based on the so-called linear forms condition; it is a fairly standard notion (see [5, Section 6]) but our definition differs slightly to allow unbounded coefficients.

Definition 5.2. We say that a function $\nu_Z : [Z] \to \mathbb{R}_+$ satisfies the Ψ_0-*linear forms condition* if for any system Ψ derived from Ψ_0 we have

$$\mathbb{E}_{n\in[Z]^d} \prod_{i\in[t]} \nu(\psi_i(n)) = 1 + o(1).$$

We also say that ν is a Ψ_0-*pseudorandom measure*.

The next proposition is about the existence of a pseudorandom majorant for a $\widetilde{W}$-tricked von Mangoldt function.

Proposition 5.1. *For any integers $b_1, \ldots, b_{t_0}$ in $[\widetilde{W}]$ coprime to $\widetilde{W}$, for $Z \gg N \log^{-O(1)} N$, there exists a Ψ_0-pseudorandom measure ν on $[Z]$ such that*

$$1 + \Lambda'_{\widetilde{W},b_1} + \cdots + \Lambda'_{\widetilde{W},b_{t_0}} \ll \nu \quad \text{on } [Z^{3/5}, Z]. \tag{5.1}$$

The construction of the majorant, identical to [5], will be explained in Section 6.

5.2. Generalised von Neumann theorem. In spite of the impossibility to move to a cyclic group, we attempt to prove an analogue of [5, Proposition 7.1]. Compared with the setting in a cyclic group, we cannot assume that some linear coefficients are 1, and the range $[X]$ is not translation invariant. Recall that $\Psi_0 : \mathbb{Z}^{d_0} \to \mathbb{Z}^{t_0}$ is a fixed system of affine-linear forms in s-normal form; thus without loss of generality, write its first form as

$$\psi_1(n_1, \ldots, n_{s+1}, y) = q_1 n_1 + \cdots + q_{s+1} n_{s+1} + \psi_1(0, y)$$

with $q_i \neq 0$ for all i and $\prod_{j\in[s+1]} \dot{\psi}_i(e_j) = 0$ for all $i > 1$. We have dropped the exponent 0 in the names of the form of the system Ψ_0 and shall always do so in what follows. We now state our variant of the von Neumann theorem.

Theorem 5.2. *Let $f_1, \ldots, f_{t_0} : \mathbb{Z} \to \mathbb{R}$ be functions and ν be a Ψ_0-pseudorandom measure such that $|f_i| \le \nu$ for all i. Then*

$$(5.2)\qquad \Big|\mathbb{E}_{n\in[X]^d}\prod_{i\in[t_0]} f_i(\psi_i(n))\Big|$$
$$\leq \Big|\mathbb{E}_{y\in[X]^{d-s-1}}\,\mathbb{E}_{n^{(0)},n^{(1)}\in[X]^{s+1}}\prod_{\omega\in\{0,1\}^{s+1}} f_1\Big(\sum_{i=1}^{s+1} q_i n_i^{(\omega_i)}+\psi_1(0,y)\Big)\Big|^{1/2^{s+1}}+o(1).$$

We adapt the proof of Proposition 7.1″ in [5] and, for brevity, use the notation from that proof without redefining it. With this notation, the left-hand side equals

$$(5.3)\qquad \mathbb{E}_{y\in[X]^{d-s-1},\,x\in[X]^{s+1}}\prod_{B\subseteq[s+1]} F_{B,y}(x_B).$$

We observe that

$$F_{[s+1],y}(x_{[s+1]}) = f_1\Big(\sum_{i\in[s+1]} q_i x_i+\psi(0,y)\Big).$$

We have the bounds

$$|F_{B,y}|\leq \nu_{B,y}.$$

The introduction of the functions $F_{B,y},\nu_{B,y}:[X]^B\to\mathbb{N}$ hides the arithmetic nature of the setting, blurring away the difference between cyclic groups and intervals of integers. Thus we can use [5, Corollary B.4] with $A=[s+1]$ and $X_\alpha=[X]$ for all $\alpha\in[s+1]$. Hence we bound (5.3) by

$$(5.4)\qquad \mathbb{E}_{y\in[X]^{d-s-1}}\|F_{[s+1],y}\|_{\square(\nu_{[s+1],y})}\prod_{B\subsetneq[s+1]}\|\nu_{B,y}\|^{2^{|B|-(s+1)}}_{\square(\nu_{B,y})},$$

where we recall that

$$\|F\|^{2^{|B|}}_{\square(\nu_{B,y})} = \mathbb{E}_{x^{(0)},x^{(1)}\in[X]^B}\prod_{\omega\in\{0,1\}^B} F(x^{(\omega)})\prod_{C\subsetneq B}\nu_C((x_i^{(\omega_i)})_{i\in C},y).$$

By Hölder's inequality, it suffices to show that

$$(5.5)\qquad \mathbb{E}_{y\in[X]^{d-s-1}}\|F_{[s+1],y}\|^{2^{s+1}}_{\square(\nu_{[s+1],y})}$$
$$= \mathbb{E}_{y\in[X]^{d-s-1}}\,\mathbb{E}_{n^{(0)},n^{(1)}\in[X]^{s+1}}\prod_{\omega\in\{0,1\}^{s+1}} f_1\Big(\sum q_i n_i^{(\omega_i)}+\psi_1(0,y)\Big)+o(1)$$

and that

$$\mathbb{E}_{y\in[X]^{d-s-1}}\|\nu_{B,y}\|^{2^{|B|}}_{\square(\nu_{B,y})} = 1+o(1)$$

for all nonempty $B\subseteq[s+1]$. To prove the latter, expand the left-hand side as

(5.6)
$$\mathbb{E}_{y\in[X]^{d-s-1}}\,\mathbb{E}_{n^{(0)},n^{(1)}\in[X]^B}\prod_{C\subseteq B}\prod_{i:\,\Omega(i)=C}\prod_{\omega\in\{0,1\}^C}\nu\big(\psi_i((n_j^{(\omega_j)})_{j\in\Omega(i)},y)\big),$$

which is an expression involving the average of ν on a system

$$\Psi=(\psi_{i,\omega})_{i\in[t_0],\,\omega\in\{0,1\}^{\Omega(i)}}:\mathbb{Z}^d\to\mathbb{Z}^t$$

of linear forms and it is easy to check that $d\le 2d_0$ and $t\le 2^{d_0}t_0$. It is also obvious that $\|\Psi\|_{M,B}\ll\|\Psi_0\|_{M,B}$. Let the prime p be exceptional for Ψ and let us check that it divides Q_0; this would mean that Ψ is derived from Ψ_0, and thus we could apply the linear forms condition. So let $\psi_{i,\omega}\ne\psi_{k,\alpha}$ be two forms that are affinely related modulo p. Then if $i\ne k$, we conclude that ψ_i and ψ_k are related, and thus the prime is exceptional for Ψ_0, which implies that it divides Q_0. Otherwise $i=k$ and thus $\omega\ne\alpha$, in other words there exists $j\in[d_0]$ such that $\dot\psi_i(e_j)\ne 0$ and $\omega_j\ne\alpha_j$. Thus p must divide $\dot\psi_i(e_j)$, and hence also Q_0. By applying the linear forms condition, the term (5.6) is $1+o(1)$.

Let us look at (5.5). At this point, Green and Tao use the translation invariance of $\mathbb{Z}/N'\mathbb{Z}$ to perform a change of variable which is not possible here, but we make do without it. As the system is in normal form and $t\ge 2$, the form ψ_1 does not genuinely use all the variables. Indeed, the form ψ_2 must also have its set of $s+1$ variables that it is the only one to use fully, in particular ψ_1 does not use them all. Let us thus assume that ψ_1 only uses $x_1,\dots,x_{d-k}$ with $k\ge 1$, which enables us, by a slight abuse of notation, to regard ψ_1 as a map from $\mathbb{Z}^{d-k}$ to $\mathbb{Z}$. Upon expanding the norm, the left-hand side of (5.5) becomes

$$\mathbb{E}_{\substack{x^{(0)},x^{(1)}\in[X]^{s+1}\\ y\in[X]^{d-k-s-1}}}\prod_{\omega\in\{0,1\}^{s+1}}f_1\Big(\sum_{i=1}^{s+1}q_ix_i^{(\omega_i)}+\psi_1(0,y)\Big)\times\mathbb{E}_{z\in[X]^k}\prod_{\omega\in\{0,1\}^{s+1}}\prod_{C\subsetneq[s+1]}\nu_{C,(y,z)}(x_C^{(\omega_C)}),$$

where (y,z) is the vector in $\mathbb{Z}^{d-s-1}$ obtained by concatenating y and z. We want to replace the inner expectation over z, which is a function of $(x^{(0)},x^{(1)},y)$ of average 1, by 1. To do that, by Cauchy–Schwarz, it is enough to prove

$$\mathbb{E}_{\substack{x^{(0)},x^{(1)}\in[X]^{s+1}\\ y\in[X]^{d-k-s-1}}}\prod_{\omega\in\{0,1\}^{s+1}}\nu\Big(\sum_{i=1}^{s+1}q_ix_i^{(\omega_i)}+\psi_1(0,y)\Big)=1+o(1)=O(1),$$

which follows directly from the linear forms condition, and

$$\mathbb{E}_{\substack{x^{(0)},x^{(1)}\in[X]^{s+1}\\ y\in[X]^{d-k-s-1}}} \prod_{\omega\in\{0,1\}^{s+1}} \nu\Big(\sum_{i=1}^{s+1} q_i x_i^{(\omega_i)} + \psi_1(0,y)\Big)|\mathbb{E}_z W(x,y,z)-1|^2 = o(1),$$

where $W(x,y,z) = \prod_{\epsilon\in\{0,1\}^{s+1}} \prod_{C\subsetneq[s+1]} \nu_{C,(y,z)}(x_C^{(\epsilon_C)})$. This amounts to

$$\mathbb{E}_{\substack{x^{(0)},x^{(1)}\in[X]^{s+1}\\ y\in[X]^{d-k-s-1}}} \prod_{\omega\in\{0,1\}^{s+1}} \nu\Big(\sum_{i=1}^{s+1} q_i x_i^{(\omega_i)} + \psi_1(0,y)\Big)(\mathbb{E}_z W(x,y,z))^j = 1+o(1)$$

for $j=0,1,2$. Let us inspect the left-hand side in the most intricate case, namely $j=2$. Upon expanding the square, we get an expectation over $x^{(0)}, x^{(1)}, y, z^{(0)}, z^{(1)}$, thus the system is in at most $2d_0$ variables. There are 2^{s+1} forms arising from ψ_1 and at most $2^{s+2}(t-1)$ other forms, which gives at most $2^{d_0}t_0$ forms altogether. Now the reasoning we used to analyse the average (5.6) also applies and shows that the system is derived from Ψ_0. Thus the linear forms condition applies and equation (5.5) is proven, hence also Theorem 5.2.

5.3. A Gowers-norm estimate. Together with the existence of a pseudorandom majorant provided by Proposition 5.1, Theorem 5.2 reduces Proposition 4.1 to the following.

Proposition 5.3. *Let $b\in[\widetilde{W}]$ be coprime to $\widetilde{W}$. Let $B>0$ and $d\in\mathbb{N}$ be constants. Suppose $q_1,\dots,q_d$ are divisors of Q satisfying $q_i = O(\log^B N)$ while $c = O(N\log^B N)$. Then*

$$\mathbb{E}_{x^{(0)},x^{(1)}\in[X]^d} \prod_{\omega\in\{0,1\}^d} \Big(\Lambda'_{\widetilde{W},b}\Big(\sum_{i=1}^{d} q_i x_i^{(\omega_i)} + c\Big) - 1\Big) = o(1). \tag{5.7}$$

The progress compared to Proposition 4.1 is that each variable $x_i^{(\epsilon)}$ for $i\in[d]$ and $\epsilon\in\{0,1\}$ is multiplied throughout the system by one and the same coefficient q_i. We now attempt to transform the system so that all variables have the same coefficient Q'; the price we pay is that the variables will not have the same ranges any more.

To this end, we introduce

$$Q_i = \prod_{j\neq i} q_j$$

and variables $n_i^{(\omega_i)}, m_i^{(\omega_i)}$ such that $x_i^{(\omega_i)} = Q_i n_i^{(\omega_i)} + m_i^{(\omega_i)}$. Then the left-hand side of (5.7) decomposes as

$$\mathbb{E}_{m_i^{(\omega_i)}\in[q_i]} \mathbb{E}_{n_i^{(\omega_i)}\in[X/q_i]} \prod_{\omega\in\{0,1\}^d} \Big(\Lambda'_{\widetilde{W},b}\Big(\sum_{i=1}^{d} Q' n_i^{(\omega_i)} + q_i m_i^{(\omega_i)} + c\Big) - 1\Big) + o(1),$$

where $Q' = q_i Q_i$ for any i.

We recognise the function

$$n \mapsto F_a(n) = \frac{\varphi(\widetilde{W})}{\widetilde{W}} \Lambda'(Q'\widetilde{W}n + a) = \Lambda'_{Q'\widetilde{W},a},$$

where the equality holds because $\prod_{p|Q'} p$ divides Q and hence $\widetilde{W}$. The parameters a occurring are

$$a_\omega = \widetilde{W}\Big(\sum_{i=1}^{d} q_i m_i^{(\omega_i)} + c\Big) + b,$$

and given that $(b, \widetilde{W}) = 1$, we also have $(a_\omega, \widetilde{W}) = 1$ and finally $(a_\omega, \widetilde{W}Q') = 1$.

We remark that for any tuple $a \in [\widetilde{W}Q']^{2^d}$ of integers coprime to $\widetilde{W}Q'$, we can create a common Ξ-pseudorandom majorant for the functions $1 + F_{a_\omega}$ where $\Xi = (\xi_\omega)_{\omega\in\{0,1\}^d}$ is defined by

$$\xi_\omega = (n_1^{(0)}, \ldots, n_d^{(0)}, n_1^{(1)}, \ldots, n_d^{(1)}) \mapsto \sum_{i=1}^{d} n_i^{(\omega_i)}.$$

In fact we can rewrite Proposition 4.1 with $\widetilde{W}Q'$ instead of $\widetilde{W}$, because Q' still satisfies $Q' = O(\log^{O(1)} N)$, a bound which allows us to control the effect of exceptional primes in Proposition 6.1 below.

Thus it remains to prove that

$$\mathbb{E}_{n_i^{(\omega_i)} \in [X_i]} \prod_{\omega\in\{0,1\}^d} \Big(F_{a_\omega}\Big(\sum_{i=1}^{d} n_i^{(\omega_i)}\Big) - 1\Big) = o(1)$$

where each X_i satisfies $N \log^{-C} N \ll X_i \le N$. Letting $Z = \max_i X_i$ and $K = \prod_i [X_i]$, we have $K \subset [Z]^d$ and $\mathrm{Vol}(K) \gg Z^d \log^{-C'} Z$. Thus we can apply the same reasoning as in Section 2 where we approximated such a convex body by a set of small boxes of equal sides ([4]), and it suffices to prove that

$$\mathbb{E}_{n^{(0)},n^{(1)} \in [Y]^d} \prod_{\omega\in\{0,1\}^d} \Big(F_{a_\omega}\Big(\sum_{i=1}^{d} n_i^{(\omega_i)}\Big) - 1\Big) = o(1) \tag{5.8}$$

for some $Y \gg N \log^{-D} N$. Now that the linear forms have bounded coefficients (namely 0 and 1), there is no more objection to the use of Green–Tao's generalised von Neumann theorem [5, Proposition 7.1], as long as the functions $1 + \Lambda'_{Q'\widetilde{W},a_\omega}$ are dominated by a pseudorandom measure, in the sense of Green–Tao [5, Definition 6.2]. Green and Tao proved the existence of such

([4]) The reader might object that we then used the positivity of the function to average, which is not available here, but we can just as well use the majorant and the linear forms condition to bound the contribution of the few boxes included in K'' but not in K'.

a majorant, except that they had W instead of $Q'\widetilde{W}$, but this makes no difference as $Q'\widetilde{W}$ is w-smooth and divisible by W. See Section 6 for a review of the construction of a majorant and a closer scrutiny of the role of the linear coefficients in the linear forms condition. Thus (5.8) follows from the claim

$$\|F_a - 1\|_{U^k([Y'])} = \|\Lambda'_{Q'\widetilde{W},a} - 1\|_{U^k([Y])} = o(1) \tag{5.9}$$

for any $a \in [Q'\widetilde{W}]$ coprime to $Q'\widetilde{W}$. This is almost [5, Proposition 7.2]. Compared with that proposition, we have $W' = Q'\widetilde{W} = O(\log^{O(1)} N)$ instead of W. One can inspect attentively the remainder of the argument of Green and Tao to notice that the required properties of W are

- that it be divisible by all primes $p \le w$ for some function $w = w(N)$ tending to infinity;
- that it be $O(\log^{O(1)} N)$; this is crucial when applying the Möbius–Nilsequence theorem [6], which comes with a saving of size an arbitrary power of $\log N$.

These properties are equally satisfied by $Q'\widetilde{W}$. Thus the claim (5.9) holds. To complete the proof of Theorem 1.3, it remains only to prove Proposition 5.1, which we shall do in the next section.

6. The linear forms condition. We recall the notation from [5],

$$\Lambda_{\chi,R}(n) = (\log R)\Big(\sum_{d\mid n} \mu(d)\chi\Big(\frac{\log d}{\log R}\Big)\Big)^2$$

for $R = N^\gamma$ a small power of N, a smooth function χ supported on $[-1,1]$ satisfying $\chi(0) = 1$ and $\int \chi'^2 = 1$. The function $\Lambda_{\chi,R}$ is positive, and if n is a prime larger than R, then $\Lambda_{\chi,R}(n) = \log R$, so that $\Lambda' \le \gamma^{-1}\Lambda_{\chi,R}$ on $[N^\gamma, N]$.

We need to extend the range of application of [5, Theorem D.3]. There it is stated only for forms whose linear coefficients are bounded, although it was then applied to other natural systems such as $\Phi = W\Psi + c$, as the extension was straightforward. In [2, Appendix A], the constant coefficients are allowed to be as large $N^{1.01}$.

We claim that the estimate can be pushed further.

Proposition 6.1. *Let L, B be positive constants and $\Psi = (\psi_1, \ldots, \psi_t)$ an admissible system of affine-linear forms satisfying $\|\Psi\|_{Z,B} \le L$. Then*

$$\sum_{n\in[Z]^d} \prod_{i\in[t]} \Lambda_{\chi,R}(\psi_i(n)) = Z^d \prod_p \beta_p(1 + o(1)).$$

We now carefully analyse what needs to be changed in the proof of Theorem D.3 of [5] when the linear coefficients are of size up to $\log^B N$.

As remarked in [2, Appendix A], the first place where the bound on the coefficients is used is page 1833, where it is said that $\alpha(p,B) = O(1/p)$ for p large enough. In fact, as we have assumed from the outset that no form of the system ψ_i is divisible by any prime p, this is always the case. This also means that $\beta_p = 1 + O(1/p)$ for all p. The next moment where Green and Tao invoke the size of the coefficients is to get the bound $\beta_p = 1 + O(p^{-2})$; but in fact it is valid as soon as p is not exceptional, whatever the size of the coefficients. As seen in the proof of Lemma 1.2, the set P_Ψ of exceptional primes is finite but its cardinality may increase to infinity with Z. What we have seen implies that the asymptotic formula

$$\sum_{n\in[Z]^d}\prod_{i\in[t]}\Lambda_{\chi,R}(\psi_i(n)) = Z^d\prod_p\beta_p(1+e^{O(X)}\log^{-1/20}R)$$

is still valid, where $X = \sum_{p\in P_\Psi} p^{-1/2}$. It remains to bound X.

If $p \in P_\Psi$, then as already seen in Lemma 1.2, p divides the parameter $Q = \log^{O(1)} N$ introduced in (4.1). Letting

$$\omega(Q) = O(\log Q) = O(\log\log N)$$

be the number of its prime factors, we have

$$X = \sum_{p\in P_\Psi} p^{-1/2} \le \sum_{p\le\omega(Q)} p^{-1/2} \le \sum_{n\ll\log\log N} n^{-1/2} \ll \sqrt{\log\log N}.$$

Thus $e^{O(X)} \ll \log^{1/30} N$ while $\log^{-1/20} R \ll \log^{-1/20} N$, from which we infer $e^{O(X)}\log^{-1/20} R = o(1)$. This completes the verification of Proposition 6.1.

From Proposition 6.1, the proof of Proposition 5.1, that is, the construction of a majorant satisfying the adequate linear forms condition, runs just as in the paper of Green and Tao. We provide it here. Let $b_1,\dots,b_{t_0}$ be integers in $[\widetilde{W}]$ coprime to $\widetilde{W}$. Let Z be an asymptotic parameter satisfying $Z \gg N\log^{-A} N$ for some constant $A > 0$. Then, writing

$$\nu(n) = \frac{1}{t_0+1}\biggl(1+\frac{\varphi(\widetilde{W})}{\widetilde{W}}\sum_{i\in[t_0]}\Lambda_{\chi,R}(\widetilde{W}n+b_i)\biggr),$$

we have the bound (5.1) for $n \in [Z^{3/5}, Z]$ if $\gamma < 3/5$. To show that ν is a Ψ_0-pseudorandom measure, it is enough to check that for any system $\Psi : \mathbb{Z}^d \to \mathbb{Z}^t$ derived from Ψ_0, any $s \le t$ and any sequence $1 \le j_1 < \dots < j_s \le t$, we have

$$(6.1)\qquad \biggl(\frac{\varphi(\widetilde{W})}{\widetilde{W}}\biggr)^s \mathbb{E}_{n\in[Z^d]}\prod_{i\in[s]}\Lambda_{\chi,R}(\widetilde{W}\psi_{j_i}(n)+b_{j_i}) = 1+o(1).$$

To this end, observe that the hypotheses ensure that the system $\Phi = (\widetilde{W}\psi_{j_i}+b_{j_i})_{i\in[s]}$ is admissible. Moreover, because $\widetilde{W} = O(\log^{O(1)} N)$, the bound $\|\Phi_{Z,B}\| = O(1)$ holds for some $B = O(1)$. So we can use Proposition 6.1. For $p \mid \widetilde{W}$, the local factor is simply $(p/(p-1))^s$, while if $p \nmid \widetilde{W}$, the prime p is not exceptional for Ψ_0 and hence for Φ, which implies $\beta_p = 1 + O(p^{-2})$ by Lemma 1.2. Thus

$$\prod_p \beta_p = \left(\frac{\widetilde{W}}{\varphi(\widetilde{W})}\right)^s \prod_{p>w}(1+O(p^{-2})) = \left(\frac{\widetilde{W}}{\varphi(\widetilde{W})}\right)^s (1+O(w^{-1})).$$

This compensates exactly for the factor $(\varphi(\widetilde{W})/\widetilde{W})^s$ and finishes the proof of (6.1), hence also of Proposition 5.1, and finally of Theorem 1.3.

In the next section, we give a nice application of our higher-dimensional Siegel–Walfisz theorem.

7. Application to the primes p such that $p-1$ is squarefree. The set of primes p such that $p-1$ is squarefree is a well-known dense subset of the primes of density $\sum_a \mu(a)/\varphi(a^2) = \prod_p (1 - 1/(p(p-1)))$; this is a theorem of Mirsky [8]. As any dense subset of the primes, it contains arbitrarily long arithmetic progressions, by the Green–Tao theorem [4]. However, no asymptotic was available so far for the count of k-term progressions in this set, nor in fact in any dense subset of the primes (except residue classes). As a consequence of Theorem 1.3, we now prove such an asymptotic; in fact, we obtain an asymptotic for the number of solutions in this set of primes to any finite complexity system of equations.

For convenience, let F be the von Mangoldt function restricted to the squarefree shifted primes, that is $F(n) = \Lambda(n+1)\mu^2(n)$. Also we denote by $\mathbb{N}$ the set of positive integers.

THEOREM 7.1. *Let $\Psi : \mathbb{Z}^d \to \mathbb{Z}^t$ be a system of affine-linear forms of finite complexity and $K \subset [-N,N]^d$ a convex body. Suppose that the linear coefficients are $O(1)$, the constant ones are $O(N)$ and $\Psi(K) \subset \mathbb{N}^t$. Then there exists a constant $C(\Psi)$ (possibly equal to 0) such that*

$$\sum_{n \in K \cap \mathbb{Z}^d} \prod_{i \in [t]} F(\psi_i(n)) = C(\Psi)\mathrm{Vol}(K) + o(N^d). \tag{7.1}$$

The constant $C(\Psi)$ will appear explicitly in the proof, but its expression is unpleasant, so we do not give it here. Throughout the proof of this theorem, we will need the notation

$$\alpha_\Psi(k_1,\ldots,k_t) = \mathbb{E}_{a \in (\mathbb{Z}/m\mathbb{Z})^d} \prod_{i \in [t]} 1_{k_i \mid \psi_i(a)},$$

where $m = \mathrm{lcm}(k_1,\ldots,k_t)$. Elementary convex geometry reveals that this is the density of points of the lattice $\{n \in \mathbb{Z}^d : \forall i \in [t]\ \ k_i \mid \psi_i(n)\}$ per unit volume, in the sense that for any convex body $K \subset [-B,B]^d$, we have

$$(7.2)\qquad \sum_{n\in K\cap\mathbb{Z}^d}\prod_{i\in[t]} 1_{d_i|\psi_i(n)} = \mathrm{Vol}(K)\alpha_{\Psi}(d_1,\dots,d_t) + O(B^{d-1}\mathrm{lcm}(d_1,\dots,d_t)).$$

This follows from simple volume packing arguments (see [5, Appendix A], [10, Appendix C]). We now prove Theorem 7.1.

Proof of Theorem 7.1. We insert the formula $\mu^2(n) = \sum_{a^2|n}\mu(a)$ into the left-hand side of (7.1). Thus

$$(7.3)\qquad \sum_{n\in K\cap\mathbb{Z}^d}\prod_{i\in[t]} F(\psi_i(n)) = \sum_{(a_1,\dots,a_t)\in\mathbb{N}^t}\prod_{i\in[t]}\mu(a_i)\sum_{\substack{n\in K\cap\mathbb{Z}^d\\ \forall i\in[t]\, a_i^2|\psi_i(n)}}\prod_{i\in[t]}\Lambda(\psi_i(n)+1).$$

Now for any $a=(a_1,\dots,a_t)\in\mathbb{N}^t$, we introduce the set

$$L_a = \{n\in\mathbb{Z}^d : \forall i\in[t]\ \ a_i^2 \mid \psi_i(n)\}.$$

Fix an a for which $L_a \neq \emptyset$ and let $n_0\in L_a$. Then

$$L_a = n_0 + \bigcap_{i=1}^{t}\ker g_i,$$

where $g_i : \mathbb{Z}^d \to \prod_{i\in[t]}\mathbb{Z}/a_i^2\mathbb{Z}$ is the affine-linear map obtained by applying $\dot{\psi}_i$ and then reducing modulo a_i^2. So L_a is an affine sublattice of full rank: indeed, its direction contains $\{\prod_i a_i^2 e_1,\dots,\prod_i a_i^2 e_d\}$.

As a lattice of full rank, the direction $\overrightarrow{L_a}$ of L_a has a $\mathbb{Z}$-basis: there exist $f_1,\dots,f_d$ such that

$$L_a = \Big\{n_0 + \sum_{i=1}^{d} m_i f_i \ \Big|\ (m_1,\dots,m_d)\in\mathbb{Z}^d\Big\}.$$

By a theorem of Mahler, we can assume that $\|f_i\| \le i\lambda_i$ for $i=1,\dots,d$, where $\lambda_1 \le \dots \le \lambda_d$ are the successive minima of the lattice $\overrightarrow{L_a}$ with respect to the Euclidean unit ball. Let R^a be the affine transformation of $\mathbb{R}^d$ defined by $R^a(0) = n_0$ and $\dot{R}^a(e_i) = f_i$ for each $i\in[d]$. Note that $L_a\cap K = R^a(\mathbb{Z}^d\cap K_a)$ where K_a is also a convex body. For the notions of the geometry of numbers alluded to here, see for instance the notes of Green [3].

Now if one of the a_i is larger than $\log^C N$ for some $C>0$, then K_a is small. Indeed, the set of $n\in K\cap\mathbb{Z}^d$ such that there exist $i\in[t]$ and $a_i > \log^C N$ satisfying $a_i^2 \mid \psi_i(n)$ has $O(N^d\log^{-C} N)$ elements. This follows from (7.2) combined with the bound $\alpha_{\psi_i}(a_i^2) \ll a_i^{-2}$, obtained by multiplicativity, linear algebra (for instance [1, Corollary C.4]) and the fact that the coefficients of ψ_i are bounded, and finally $\sum_{a>x} a^{-2} \ll x^{-1}$. Bounding the contribution to the left-hand side of (7.3) of this exceptional set of $n \in K \cap \mathbb{Z}^d$ using $F \ll \log$, and supposing that $C \geq 2t$, we obtain

$$\begin{aligned}\sum_{n\in K\cap\mathbb{Z}^d} \prod_{i\in[t]} F(\psi_i(n)) &= \sum_{\substack{n\in K\cap\mathbb{Z}^d \\ \forall i\in[t]\, \forall a>\log^C N\, a^2\nmid\psi_i(n)}} \prod_{i\in[t]} F(\psi_i(n)) + O(N^d \log^{-C/2} N) \\ &= \sum_{1\leq a_1,\dots,a_t\leq \log^C N} \prod_{i\in[t]} \mu(a_i) \sum_{n\in K\cap L_a} \prod_{i\in[t]} \Lambda(\psi_i(n)+1) \\ &\quad + O(N^d \log^{-C/2} N).\end{aligned}$$

Fix $(a_1,\dots,a_t) \in [\log^C N]^t$ where $C = 2t$. For each $i \in [t]$, the map $\psi_i^a : L_a \to \mathbb{Z}$ defined by

$$\psi_i^a(n) = \psi_i(n)/a_i^2$$

is an affine map. Then introduce $\phi_i^a = \psi_i^a \circ R^a$. This defines a system $\Phi^a : \mathbb{Z}^d \to \mathbb{Z}^t$ of affine-linear forms which is again of finite complexity. Thus the inner sum on the left-hand side of (7.3) may be written as

$$\sum_{n\in K\cap L_a} \prod_i \Lambda(\psi_i(n)+1) = \sum_{m\in K_a\cap\mathbb{Z}^d} \prod_i \Lambda(a_i^2\phi_i^a(m)+1). \tag{7.4}$$

We now apply Theorem 1.3 to the inner sum. One can check that the linear coefficients of Φ_a have size $O(\log^{O(1)} N)$. To do so, it is enough to examine the size of the basis vectors f_j of the lattice $\overrightarrow{L_a}$. Indeed,

$$a_i^2|\dot{\phi}_i^a(e_j)| = |\dot{\psi}_i(f_j)| \leq \|\dot{\psi}_i\|\,\|f_j\| \ll \|f_j\|.$$

Moreover, the constant coefficients are $O(N)$. As observed, if $n_0 \in L_a$, the lattice

$$\Big\{ n_0 + \sum_{i\in[d]} k_i a_i^2 e_i \;\Big|\; k \in \mathbb{Z}^d \Big\}$$

is a sublattice of L_a and its determinant is $\prod_i a_i^2 \leq \log^{2dC} N$. Hence using Minkowski's second theorem, one finds that

$$\prod_{i\in[d]} \|f_i\| \leq d! \prod_{i\in[d]} \lambda_i \ll_d |\det L_a| \leq \log^{2dC} N.$$

Similarly, we obtain the bound

$$\mathrm{Vol}(K_a) = \mathrm{Vol}(K)\det(R^a)^{-1} \geq \mathrm{Vol}(K)\log^{-2dC} N.$$

Now Theorem 1.3 tells us that the right-hand side of (7.4) is equal to $\mathrm{Vol}(K_a)\prod_p \beta_p(1+o(1))$ as soon as none of the local factors $\beta_p(a)$ corresponding to the system of the forms $a_i^2\phi_i^a + 1$ vanishes. Note that if any $\beta_p(a)$ is 0, then for all m there exists $i \in [t]$ such that $p \mid a_i^2\phi_i^a(m)+1$. Then it is easy to see that

$$\sum_{m\in K_a\cap\mathbb{Z}^d} \prod_i \Lambda(a_i^2\phi_i^a(m)+1) = O(N^{d-1}\log^{O(1)} N).$$

Moreover, (7.2) reveals that

$$\begin{aligned}\mathrm{Vol}(K_a) &= |K_a \cap \mathbb{Z}^d| + O(N^{d-1}) = |K \cap L_a| + O(N^{d-1})\\ &= \mathrm{Vol}(K)\alpha_\Psi(a_1^2,\dots,a_t^2) + O(N^{d-1}\log^{O(1)} N).\end{aligned}$$

Thus, up to an error term of size $O(N^d \log^{-C/2} N)$, the left-hand side of (7.1) equals

$$\mathrm{Vol}(K)(1+o(1)) \sum_{1\le a_1,\dots,a_t\le \log^C N} \alpha_\Psi(a_1^2,\dots,a_t^2)\prod_p \beta_p(a)\prod_{i\in[t]}\mu(a_i).$$

We claim the sum over a is absolutely convergent. To see this, first observe that the exceptional primes for the system of the forms $a_i^2\phi_i^a+1$ are divisors of a_i^2 or exceptional primes for the system Φ^a; in any case, they are divisors ([5]) of a parameter $Q(a) = O(\prod_i a_i^{O(1)})$. For all other primes, we have $\beta_p = 1 + O(p^{-2})$ by Lemma 1.2, so that

$$\prod_p \beta_p(a) \ll \prod_{p\mid Q(a)} \beta_p(a) \le \left(\frac{Q(a)}{\varphi(Q(a))}\right)^t \ll (\log\log Q(a))^t \ll \Big(\log\log \prod_i a_i\Big)^t.$$

Then note that the sum

$$\sum_{a_1,\dots,a_t} \Big(\log\log\prod_i a_i\Big)^t \alpha_\Psi(a_1^2,\dots,a_t^2)$$

is convergent. Indeed, we have the bound

$$\begin{aligned}\alpha_\Psi(a_1^2,\dots,a_t^2) &= \prod_p \alpha_\Psi(p^{2v_p(a_1)},\dots,p^{2v_p(a_t)})\\ &\ll \prod_p p^{-2v_p(\max_i a_i)} = \mathrm{lcm}(a_1,\dots,a_t)^{-2},\end{aligned}$$

where the inequality holds because the forms ψ_i have bounded linear coefficients and if $p \nmid \psi_i$, then $\alpha_{\psi_i}(p^k) \le p^{-k}$ (this is linear algebra and Hensel's lemma, see [1, Corollary C.4]). The convergence then follows from a trivial bound for the number of t-tuples a of prescribed least common multiple k,

([5]) See (4.1) and the remark following it.

namely $\tau(k)^t$. This convergence result implies that

$$\sum_{1\le a_1,\dots,a_t\le \log^C N} \alpha_\Psi(a_1^2,\dots,a_t^2)\prod_p \beta_p(a)\prod_{i\in[t]}\mu(a_i) = C(\Psi)+o(1),$$

where

$$C(\Psi) = \sum_{(a_1,\dots,a_t)\in\mathbb{N}^t} \alpha_\Psi(a_1^2,\dots,a_t^2)\prod_p \beta_p(a)\prod_{i\in[t]}\mu(a_i).$$

This concludes the proof. ∎

Acknowledgements. The author was supported by the EPSRC (grant EP/M506473/1) while this work was carried out.

References

[1] P.-Y. Bienvenu, *Asymptotics for some polynomial patterns in the primes*, arXiv: 1511.07317 (2015).

[2] K. Ford, B. Green, S. Konyagin, and T. Tao, *Large gaps between consecutive prime numbers*, Ann. of Math. (2) 183 (2016), 935–974.

[3] B. Green, *Notes on progressions and convex geometry*, http://people.maths.ox.ac.uk /greenbj/papers/convexnotes.pdf.

[4] B. Green and T. Tao, *The primes contain arbitrarily long arithmetic progressions*, Ann. of Math. (2) 167 (2008), 481–547.

[5] B. Green and T. Tao, *Linear equations in primes*, Ann. of Math. (2) 171 (2010), 1753–1850.

[6] B. Green and T. Tao, *The Möbius function is strongly orthogonal to nilsequences*, Ann. of Math. (2) 175 (2012), 541–566.

[7] B. Green, T. Tao, and T. Ziegler, *An inverse theorem for the Gowers $U^{s+1}[N]$-norm*, Ann. of Math. (2) 176 (2012), 1231–1372.

[8] L. Mirsky, *The number of representations of an integer as the sum of a prime and a k-free integer*, Amer. Math. Monthly 56 (1949), 17–19.

[9] T. Tao, *Concatenation theorems for anti-Gowers-uniform functions and Host–Kra characteristic factors; polynomial patterns in primes*, blog post, https://terrytao. wordpress.com/2016/03/27.

[10] T. Tao and T. Ziegler, *The primes contain arbitrarily long polynomial progressions*, Acta Math. 201 (2008), 213–305.

[11] T. Tao and T. Ziegler, *Polynomial patterns in the primes*, arXiv:1603.07817 (2016).

Pierre-Yves Bienvenu
School of Mathematics
University of Bristol
Bristol BS8 1TW, United Kingdom
E-mail: pb14917@bristol.ac.uk